PROCEEDINGS of the ICCS-X

Tenth Islamic Countries Conference on Statistical Sciences

Volume I

STATISTICS FOR DEVELOPMENT AND GOOD GOVERNANCE

الإحصاء من أجل التنمية و الحكم الرشيد

Editors
Zeinab Amin and Ali S. Hadi

The American University in Cairo

Proceedings of the ICCS-X

Tenth Islamic Countries Conference on Statistical Sciences

Statistics for Development and Good Governance

Volume I

Editors

Zeinab Amin and Ali S. Hadi

Department of Mathematics and Actuarial Science

The American University in Cairo

THE AMERICAN UNIVERSITY IN CAIRO

الجامعة الأمريكية بالقاهرة

The Islamic Countries Society of Statistical Sciences
Plot No. 44-A, Civic Centre,
Liaqat Chowk,
Sabzazar Scheme,
Multan Road
Lahore, Pakistan

Telephones: +92-42-37490670, +92-42-3587-8583

E-mail: secretary@isoss.net

ISBN: 978-977-416-365-8

Printed in Cairo, Egypt

Mir Maswood Ali (1929 - 2009)

DEDICATION

The Tenth Islamic Countries Conference on Statistical Sciences, including this Proceedings are dedicated to the memory of Mir Maswood Ali (photo above), Professor Emeritus of Statistics at the University of Western Ontario. The following is an Obituary, written by his brother Mir Masoom Ali, George and Frances Ball Distinguished Professor Emeritus of Statistics, Ball State University:

Mir Maswood Ali, 80, Professor of Statistics Emeritus, University of Western Ontario and a brilliant statistician of Bangladeshi origin, died August 18, 2009 in London, Ontario, Canada due to pulmonary complications. It is my great honor and privilege to write this obituary for my older brother who was very dear to me and who had tremendous influence on my career.

Ali received his B.Sc. degree in Mathematics in 1948 and his M.Sc. degree in Statistics in 1950 both from the University of Dhaka. He belonged to the first batch of graduate students in statistics and had obtained first class and secured the highest mark for which he was awarded a gold medal. He served as Lecturer in the Department of Statistics at Dhaka University from 1950 to 1952. He then worked from 1952 to 1957 as an Actuarial Assistant at Norwich Union Life and Canada Life. In 1958, he obtained a second Master's degree in Actuarial Science at the University of Michigan and worked there as a Teaching Fellow until 1959. He then went to the University of Toronto where he obtained his Ph.D. degree in Statistics in 1961 under the supervision of D. A. S Fraser after merely two years of studies. He then joined the Mathematics Department at the University of Western Ontario (UWO) in London, Ontario, Canada as assistant professor in 1961. He was the first faculty member in statistics in the department and was quickly promoted to the rank of associate professor in 1963 and to full professor in 1966 and he remained there until his retirement in 1994 when he was named Professor Emeritus. Ali had developed the graduate and undergraduate programs in statistics in his department and he was instrumental in the creation of a separate Department of Statistics and Actuarial Sciences at UWO.

He supervised 15 Ph.D. students, a number of whom are now well-known statisticians and 40 Master's theses. He published in leading statistical journals such as the Annals of Mathematical Statistics, the Journal of the Royal Statistical Society, the Journal of Multivariate Analysis, the Pacific Journal of Mathematics, and Biometrika, to name a few. His research interests encompass many areas of Statistics and Mathematics, including order statistics, distribution theory, characterizations, spherically symmetric and elliptically contoured distributions, multivariate statistics, and *n*-dimensional geometry and his two highly rated papers are in geometry which appeared in the Pacific Journal of Mathematics.

He was a man of strong principle. He was also a very decent and humble man who never sought recognition for anything that he did or achieved. He was a dedicated family man and he devoted lot of his time to his own family. He left behind his loving wife of 47 years Surayia, and eight grown children, Rayhan, Yasmin, Selina, Sharmeene, Sadek, Nasreen, Ayesha, and Adnan, and seven grandchildren. His youngest daughter Ayesha followed his father's footsteps and now teaches statistics at the University of Guelph in Canada.

Mir Maswood Ali was my immediate older brother and it was due to his influence that I got into statistics as a student in 1953. He was a great mentor, a great teacher and a friend and he was all that I wanted to be in life. I will miss him dearly.

In loving memory of my brother,
Mir Masoom Ali
George and Frances Ball Distinguished Professor Emeritus
Ball State University, Muncie, Indiana, USA
Excerpted from IMS Bulletin, Vol. 38, Issue 9, 2009

CONTENTS

VOLUME I

VOLUME II

PREFACE

The Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X) was held during the period December 20-23, 2009 at the brand new campus of The American University in Cairo (AUC), in New Cairo, Egypt. The ICCS-X was organized by the Islamic Countries Society of Statistical Sciences (ISOSS) and cosponsored by AUC and the Egyptian Cabinet Information and Decision Support Center (IDSC).

The collaboration between a governmental organization, represented in IDSC, and a private non-for-profit university, represented in AUC, in sponsoring such an international conference has proven to be a very effective and mutually beneficial joint effort. The conference, which brought together researchers and practitioners in statistical sciences from 32 countries all over the world, was open to all people interested in the development of statistics and its applications regardless of affiliation, origin, nationality, gender or religion.

The theme of the ICCS-X was *Statistics for Development and Good Governance*. As can be seen in this Proceedings, three Discussion Panels "*Public Opinion Polling and Good Governance*," by Prof. Magued Osman, IDSC, "*Measuring the Unmeasurable*," by Dr. Anis Yusoff, National University of Malaysia, and "*Indicators and Politics*," by Prof. Ali S. Hadi, AUC, have been devoted entirely to this theme. Other papers dealt with various broad topics in statistics theory and its applications. As a result, ICCS-X has attracted a distinguished team of speakers giving more than 190 presentations.

These proceedings do not contain all articles that have been presented at the conference. Only articles that have undergone and passed peer reviews are included. The reviews have taken into consideration both the quality of the paper and the quality of the presentation at the conference. These Proceedings contain 29 abstracts and 85 complete papers. The papers were arranged alphabetically according to the first author. Due to the large size, these proceedings are split into two volumes; Volume I and Volume II.

Organizing a conference requires a lot of effort by many people, collaboration, coordination, and paying attention to very small details. We would like to thank all organizers and participants of the conference. We are particularly grateful to Prof. Jef Teugels, Catholic University of Leuven, Belgium and the current President of the International Statistical Institute for giving the opening keynote talk despite his very busy schedule. We are also thankful to Prof. Kaye E. Basford, University of Queensland, Australia, Prof. Jim Berger, Duke University, USA and former President of the Institute of Mathematical Statistics, and Prof. Edward J. Wegman, George Mason University, USA for giving the other three keynote talks. In addition to the four keynote talks and the three panel discussions, the program included nine invited sessions and 20 contributed sessions. According to feedback from participants, the conference was a great success.

We are also grateful for the following referees who devoted the time and effort to review these articles: Dr. Mina Abdel Malek, Dr. Maged George, Dr. Mohamed Gharib, Dr. Ramadan Hamed, Dr. Mohamed Ismail, Dr. Hafiz Khan, Dr. Mohamed Mahmoud, Dr. Nadia Makary, Dr. Amani Moussa, Dr. Abdel Nasser Saad, Dr. Kamal Selim, Dr. Tarek Selim, Dr. Zeinab Selim, and Dr. Mark Werner.

The Conference was organized by three Committees: the Scientific and Program Committee (Chair: Ali S. Hadi and Co-Chair: Zeinab Amin), the International Organizing Committee (Chair: Shahjahan Khan), and the Local Organizing Committee (Chair: Magued Osman and Co-

Chair: Zeinab Amin). The members of these committees are given on page xv. Each of these committees has worked tirelessly for the organization of this conference. We are indebted to each and every one of them. We have also benefited from the contributions of the ISOSS Headquarters and in particular the President of ISOSS Prof. Shahjahan Khan. Prof. Mohamed Ibrahim has generously shared with us his valuable experience in the organization of the ICCS-IX Conference that was held in Malaysia in 2007.

Prof. Wafik Younan helped in putting together the Local Organizing Committee (LOC), which consists of members from several Egyptian universities and government agencies including Ain Shams University, Al-Azhar University, The American University in Cairo, Cairo University, CAPMAS, Helwan University, and IDSC. Over the 15 months prior to the conference, the LOC has held monthly meetings at Cairo University's Institute of Statistical Studies and Research, where Dr. Amani Moussa and Dr. Mahmoud Riyad were the primary hosts.

Dr. Wafik Younan also served as the Treasurer. We are very grateful for the following organizations for their financial and other support: AUC, the Egyptian Ministry of Tourism, IDSC, the Islamic Development Bank. Mr. Amr Agamawi of IDSC was instrumental in fund raising and administrative activities and has so ably taken care of various logistics and attention to details. Dr. Mostafa Abou El-Neil, Mr. Waleed Gadow, and the multimedia team of IDSC, were responsible of the design and printing of various publications including posters, brochure, and the Book of Abstracts. Eng. Medhat El Bakry, Eng. Ibrahim Hamdy, Eng. Ahmed Khalifa, the information system and communication team of IDSC, and Eng. Mai Farouk, of AUC, were responsible of maintaining and updating the website of the conference. Lamyaa Mohamed Sayed prepared the list of contributors. Finally, the staff at the Office of AUC's Vice Provost (Samah Abdel-Geleel, Basma Al-Maabady, Sawsan Mardini, Dahlia Saad, and Nancy Wadie) helped with the correspondence and various other organizational details. We apologize if we left out some of the people who have provided us with help. This omission is, of course, not intentional.

Zeinab Amin and Ali S. Hadi, Joint Editors
Cairo, Egypt
July 2010

ICCS-X Organizing Committees

Scientific Program Committee (SPC)
Ahmed, Munir - Pakistan
Ahmed, Nasar U - USA
Ahmed, S Ejaz - Canada
Ahsanullah, Mohammad - USA
Aly, Emad Eldin - Kuwiat
AlZayed, Abdulhamid – KSA
Amin, Zeinab - Egypt (Co-Chair)
Arnold, Barry C. - USA
Castillo, Maria-Carmen - Spain
Djauhari, Maman - Indonesia
ElShaarawi, Abdulhameed - Canada
Ghosh, Malay - USA
Hadi, Ali S - Egypt (Chair)
Horova, Ivan - Czech Republic
Ibrahim, Noor Akma - Malaysia
Imon, Rahmatullah - USA
Khan, Shahjahan - Australia
Mengersen, Kerrie - Australia
Mian, Mohammad Hanif - Pakistan
Nyquist, Hans - Sweden
Osman, Magued - Egypt
Parsian, Ahmad - Iran
Provost, Serge B.- Canada
Puntanen, Simo - Finland
Raqab, Mohammad - Jordan
Rezakhak, Syed - Iran
Sinha, Bikas K - India
Yanagawa, Takashi - Japan

International Organizing Committee (IOC)
Abuammoh, Abdulrahman - KSA
Ageel, Mohammed I. - KSA
Ahmad, Munir - Pakistan
Ahmed, Kazi Saleh - Bangladesh
Ahmed, Nasar U - USA
Ahsanullah, Abd Mohammad - USA
Ahsanullah, Mohammad - USA
AlBiyyat, Hilal - Jordan
Al-Awati, Shafiqa - Kuwait
Al-Saleh, Mohammad - Qatar
Ali, Abdunnabi - Libya
Amin, Zeinab - Egypt
Basci, Sidika - Turkey
Bellout, Djamel - UAE
Chattopadhaya, Asis K - India
Djauhari, Maman - Indonesia
El-Bassouni, M Yahya - UAE
El-Shaarawi, Abdulhameed - Canada
Erçetin, Sefika Sule – Turkey
Grover, Gurpirit - India
Gupta, Pushpa L. - USA
Gupta, R.C. - USA
Hadi, Ali S - Egypt/USA
Hasan, Baktiar - Belgium
Hussain, Tayyab - Libya
Kabir, M - Bangladesh
Khan, Abdul Hamid - India
Khan, Bashir U, Canada
Khan, Hafiz Abdullah - UK
Khan, Shahjahan - Australia (Chair)
King, Max - Australia
Mian, Abul Basher - Bangladesh
Mian, Mohammad Hanif - Pakistan
Mohamed, Ibrahim - Malaysia
Ojikutu, Rasheed Kola - Nigeria
Osman, Magued - Egypt
Provost, Serge B - Canada
Rahman, Mohammad Shafiqur - Oman
Rejali, Ali - Iran
Rezakhah, Syed - Iran
Saefuddin, Asep - Indonesia
Samb, LO Gane - Senegal
Sinha, Bimal K. - USA
Soleiman, Hana - UAE
Srivastava, Muno - Canada
Srivastava, R.C. - USA
Sufian, Abu - Bahrain
Wegman, Edward - USA
Younan, Wafik – Egypt

Local Organizing Committee (LOC)
Abou El-Ela, Esmat - Al Azhar University

Abou El-Kassem, Essam - Helwan University
Amin, Zeinab (Co-Chair) - AUC
El-Agamawi, Amr - IDSC
El-Batrawy, Rawya - CAPMAS
El-Sheneity, Sahar - Cairo University
El-Tawila, Sahar - IDSC
Farouk, Mai - AUC
George, Maged - AUC
Gharib, Mohamed - Ain Shams University
Hadi, Ali S - AUC
Hamed, Ramadan - Cairo University
Ismail, Mohamed - Cairo University
Magdy, Dina - Cairo University
Mahmoud, Mohamed - Ain Shams University
Mousa, Amani - ISSR
Nour Eldien, Mohamed - ISSR
Osman, Magued (Chair) - IDSC
Riad, Mahmoud - ISSR
Saad, Abdel Naser – Cairo University
Werner, Mark - AUC
Younan, Wafik (Treasurer) - AUC
Zaky, Hassan - Cairo University

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 1–3.

KEYNOTE SESSIONS

CONSTRUCTION AND THREE-WAY ORDINATION OF THE WHEAT PHENOME ATLAS

V.N. Arief[1], P.M. Kroonenberg[2], I.H. Delacy[1,3], M.J. Dieters[1], J. Crossa[4], and K.E. Basford[1,3]
[1] The University of Queensland, School of Land, Crop and Food Sciences, Brisbane 4072, Australia
[2] Department of Education and Child Studies, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands
[3] Australian Centre for Plant Functional Genomics, Brisbane 4072, Australia
[4] International Maize and Wheat Improvement Center (CIMMYT), APDo. Postal 6-641, 06600 México, D.F., Mexico

ABSTRACT

Long-term plant breeding programs generate large quantities of genealogical, genotypic, phenotypic and environment characterization data. Marker-trait association studies are commonly used to integrate and represent such data, but the results are context dependent, as trait associated markers depend on the germplasm investigated, the environments in which they are studied, and the interaction of genotype and environment. The concept of a trait-associated marker block, defined as markers in a linkage disequilibrium block that show significant association with a trait, can be used to reduce the number of trait-associated markers, control family-wise error rate and address non-independency of markers in association analysis. A Phenome Atlas can be constructed as a collection of diagrammatic representations of chromosome regions that affect trait inheritance (phenome maps) to document the patterns of trait inheritance across the genome. These maps demonstrate the complexity of the genotype-to-phenotype relationships and the context dependency of marker trait association patterns. This methodology is illustrated using the Wheat Phenome Atlas which represents the results of a genome wide association study of 20 economically important traits from the first 25 years of an international wheat breeding program. Three-way principal component analysis can be used to obtain information about which genotypes carry favorable trait-associated marker block combinations, which marker blocks discriminate among genotypes and which marker block combinations are available for any given combination of genotypes and traits. Different patterns of marker trait association profiles are observed when analyzing the same genotypes for different marker blocks and trait combinations and for data obtained from different combinations of environments. These results emphasize the context dependency of association studies.

BAYESIAN ADJUSTMENT FOR MULTIPLICITY

Jim Berger
Department of Statistical Science
Duke University, Durham, NC 27708-0251, USA
Office: 221 Old Chemistry Building
E-mail: berger@stat.duke.edu
Statistical and Applied Mathematical Sciences Institute
P.O. Box 14006, Research Triangle Park, Durham, NC 27709-4006
Office: 262 NISS Building
E-mail: berger@samsi.info

ABSTRACT

Issues of multiplicity in testing are increasingly being encountered in a wide range of disciplines, as the growing complexity of data allows for consideration of a multitude of possible hypotheses (e.g., does gene xyz affect condition abc); failure to properly adjust for multiplicities is possibly to blame for the apparently increasing lack of reproducibility in science.

Bayesian adjustment for multiplicity is powerful, in that it occurs through the prior probabilities assigned to models/hypotheses, and is thus independent of the error structure of the data - the main obstacle to adjustment for multiplicity in classical statistics. Not all assignments of prior probabilities adjust for multiplicity, however, and it is important to understand which do and which don't. These issues will be reviewed through a variety of examples.

If time permits, some surprises will also be discussed, such as the fact that empirical Bayesian approaches to multiplicity adjustment can be flawed.

EXTREME VALUE THEORY - A TUTORIAL

Jef L. Teugels
Katholieke Universiteit Leuven & EURANDOM, Eindhoven

Jan Beirlant
University Center of Statistics, Katholieke Universiteit Leuven
Goedele Dierckx, HUBrussel
E-mail: jef.teugels@wis.kuleuven.be

ABSTRACT

Extreme value theory is still a much underrated part of statistics. In complexity, it compares favorably to classical central limit theory that also depends on the theory or regular variation. For example, the possible limit laws for the maximum of a sample (the *extreme value laws*), are determined by a simple relation that depends on a single parameter, the *extreme value index* γ.

In the case where $\gamma > 0$, the statistical estimation of this index is often done by a Hill-type estimator that uses a number of the largest order statistics. But alternatives are available.

We illustrate the results with a number of examples from environmental sciences, from geology and from insurance, especially from catastrophe modeling.

Many results can be found in a textbook *J. Beirlant, Y.Goegebeur, J. Segers and J.L. Teugels, Statistics of Extremes, Wiley, 2004*.

MASSIVE DATA STREAMS AND CITIZEN SCIENCE

Edward J. Wegman
Center for Computational Statistics
George Mason University
368 Research I, Ffx, MSN: 6A2
E-mail: ewegman@gmu.edu

ABSTRACT

Science has traditionally been built on two major paradigms: Theory and Experiment. These are often believed to be complementary one feeding into the other. More recently, Computation has become a third paradigm, i.e. exploring the possibility space computationally in a way that could never be done without modern computing resources. In some way Computation partially replaces, but also augments Experiment. The Google experience suggests that there is yet another paradigm emerging. This is a data centric perspective, where masses of data almost replace Theory, or, more precisely Massive Data partially replaces, but also augments, Theory with much the same relationship between Computation and Experiment. In this talk, I will give some examples of massive data streams including data coming from the Large Synoptic Survey Telescope. This instrument is expected to generate 100 petabytes of images of 50 billion astronomical objects. Classification of such objects is a task beyond both any single team and beyond the most sophisticated machine learning algorithms. There is no substitute for the human eye-brain classifier. Using the Internet, enthusiastic citizens can be drawn into the process of classifying these objects. Some examples of citizen science will be illustrated.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 4–6.

PANEL DISCUSSIONS

PANEL 1: PUBLIC OPINION POLLING AND GOOD GOVERNANCE

KNOWLEDGE MANAGEMENT TO BUILD TRUST IN GOVERNMENT INSTITUTIONAL AND REGULATORY FRAMEWORK ISSUES

Magued Osman
Chairman, Information and Decision Support Center
The Egyptian Cabinet
E-mail: magued_osman@idsc.net.eg

Moderator:
Dina Al Khawaja, Regional manager of research unit, Ford Foundation

Discussants:
Hafez Al Mirazi, Director of the Kamal Adham Center for Journalism Training and Research, The American University in Cairo, P.O.Box 74, New Cairo 11835, Egypt
Jennifer Bremer, Chair, Public Policy and Administration Department, School of Public Affairs, The American University in Cairo, P.O.Box 74, New Cairo 11835, Egypt

ABSTRACT

Since the late eighties and early nineties, the world has been witnessing a change in paradigm. A new values system based on empowerment, participation, accountability and transparency are replacing traditional institutional and regulatory frameworks of rule and knowledge management. The early nineties witnessed a revival in demand to "reinvent government" in a way to make it more effective, more democratic, and more transparent. The concept of "governance" hence emerged.

The United Nations Development Program defined governance as the exercise of economic, political and administrative authority in managing a country's affairs at all levels. It embraces the mechanisms, processes and institutions through which citizens and groups articulate their interests, exercise their legal rights, meet their obligations and mediate their differences. This definition implies that public policies are a result of formal and informal interactions between the different actors in society. Therefore, governance is not only about how government conducts business in its own sphere, but also how it interacts with civil society. It illustrates how well government encourages and facilitates people's participation, not only in the delivery of services but also in the evaluation and monitoring of government performance. The achievement of good governance – according to the UNDP definition – requires equitable participation of all stakeholders, so that they influence policy making, setting policy agenda, and allocation of resources. It also requires the free flow of information so that transparency and accountability

are promoted. Moreover, good governance is effective and equitable, and it promotes the rule of law. It ensures that political, social and economic priorities are based on broad consensus in society and that the voices of the poorest and the most vulnerable are heard in decision- making over the allocation of development resources.

During the past few years, the Egyptian government has increasingly realized the importance of good governance as in making effective and efficient decisions. Various efforts were taken towards providing an institutional and regulatory framework that facilitates and promotes public participation in policy making. This paper examines the efforts of, The Egyptian Cabinet's Information and Decision Support Center (IDSC) to instill mechanisms and tools aimed at achieving accountability and transparency. More specifically, the paper highlights the efforts undertaken by the IDSC to enhance public polling mechanisms as an indispensable strategy towards achieving such goals. In July 2003, IDSC established the Public Opinion Poll Center (POPC) with the aim of increasing citizens' participation, strengthening the principle of citizenship rights and supporting the democratization process.

PANEL 2: MEASURING THE UNMEASURABLE

CORRUPTION PERCEPTION

Anis Y. Yusoff
Principal Research Fellow, Institute of Ethnic Studies (KITA),
National University of Malaysia
Member of the Advisory Panel of the Corruption Prevention and Advocacy of the Malaysian
Anti Corruption Commission (MACC)
E-mail: anis.yusoff@gmail.com

Moderator:
Mostafa Kamel El Sayed, Professor of Political Science, Cairo University

Discussants:
Nadia Makary, Professor of Statistics, Cairo University
Andrew Stone, World Bank

ABSTRACT

Corruption is a multifaceted social, political and economic phenomenon that is prevalent in all countries in varying degrees. There is no worldwide consensus on the meaning of corruption. In literature, corruption is usually defined as the exploitation of public power for private gain. Although this definition has been broadly adopted, several critics have observed that such a definition is culturally biased and excessively narrow (UNDP 2008). The fundamental question is: is it possible to measure corruption, and if so, how? Corruption is a variable that cannot be measured directly. However, the number of indices focused on corruption measurement has grown exponentially over the past decade. They range from some of the more established and widely used indicators like Transparency International's (TI) Corruption Perceptions Index (CPI)

and the World Bank's Worldwide Governance Indicators (WGI), to newer generation of measurement and assessment tools like TI's Global Corruption Barometer and Global Integrity's Global Integrity Index. This paper seeks to discuss some of these issues, in particular to highlight on TI-CPI and where Muslim countries are ranked in that index. This paper provides an overview of selected international corruption indices as well as an alternative measurement system which will give a more positive approach to measuring corruption and in creating an early warning system for any potential gaps for corruption to take place. This paper also alludes to explain areas that are not discussed when one argues on anti-corruption issues which are more difficult to measure and help to provide some solutions.

PANEL 3: INDICATORS AND POLITICS

STATISTICS FOR GOOD GOVERNANCE: THE IBRAHIM INDEX FOR AFRICAN GOVERNANCE (STATISTICAL CHALLENGES AND LIMITATIONS)

Speaker: Ali S. Hadi
Department of Mathematics and Actuarial Science,
The American University in Cairo, P.O.Box 74,
New Cairo 11835, Egypt
E-mail: ahadi@aucegypt.edu

Moderator:
Lisa Anderson, Provost, The American University in Cairo, P.O.Box 74, New Cairo 11835, Egypt

Discussants:
Lisa Anderson, Provost, The American University in Cairo, P.O.Box 74, New Cairo 11835, Egypt
Stephen Everhart, Associate Dean, School of Business, The American University in Cairo, P.O.Box 74, New Cairo 11835, Egypt
Nabil Fahmy, Ambassador and Dean, School of Public Affairs, The American University in Cairo, P.O.Box 74, New Cairo 11835, Egypt

ABSTRACT

The Ibrahim Index of African Governance (IIAG) is a comprehensive ranking of the 53 African nations according to governance quality. It assesses national governance using 84 criteria (variables). The criteria capture the quality of services provided to citizens by governments. The IIAG aims to (a) provide a tool for civil society and citizens to hold governments to account, (b) stimulate debate on governance, in particular by providing information about leadership performance, and (c) provide a diagnostic framework to assess governance in Africa. Details can be seen at the Ibrahim Foundation Website: www.moibrahimfoundation.org. This talk describes how the IIAG is constructed and discusses some of its statistical challenges and limitations. Recommendations for improving the index are also offered.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 7–16.

INVITED SESSIONS

SESSION 1: DEMOGRAPHY & POPULATION AGEING
Chair: Hafiz T.A. Khan, Middlesex University Business School, London & Associate Research Fellow, Institute of Ageing, University of Oxford, UK

RURAL-URBAN DIFFERENTIALS IN THE PROBLEMS FACED BY THE ELDERLY IN THE ERA OF HIV/AIDS IN MAFIKENG LOCAL MUNICIPALITY, NORTH WEST PROVINCE, SOUTH AFRICA

Paul Bigala
Ph.D Student, Population Studies and Demography
North West University (Mafikeng Campus)
South Africa
E-mail: paulgigs@yahoo.com

Ishmael Kalule-Sabiti
Professor of Population Studies and Demography
North West University (Mafikeng Campus)
South Africa
E-mail: Ishmael.KaluleSabiti@nwu.ac.za

Natal Ayiga
Department of Population Studies
University of Botswana
E-mail: Natal.Ayiga@mopipi.ub.bw

ABSTRACT

Ageing is increasingly becoming a global phenomenon. In sub-Saharan Africa, South Africa has one of the fastest growing elderly populations. These people have been faced with numerous social and economic challenges resulting from the disintegration of the traditional family systems in the era of the HIV/AIDS pandemic. This study investigated the problems faced by the elderly in South Africa in the era of HIV/AIDS. This was a cross-sectional study using both quantitative and qualitative methods. Quantitative data was collected from a sample of 506 households randomly selected from both rural and urban areas. Data analysis was done using chi-square and the Master sheet analysis framework. The study found that the elderly experienced several problems which were aggregated by HIV/AIDS. These problems were different between rural and urban areas. The main problems identified include income poverty, poor health, inadequate food resources and low self-esteem. HIV/AIDS related problems experienced by the elderly include care of both their own sick children and their grandchildren. The most mentioned care needs were identified as healthcare, clothing, education, food and emotional support. Older women were found to provide the bulk of the care for their adult children and their grandchildren. In the light of these findings, the study recommends improvements in the welfare scheme for the elderly, free health care and

education for affected families and a nutrition support programme for people infected and affected by HIV/AIDS.

1. INTRODUCTION

The global rapid growth of the elderly population has created an unprecedented demographic revolution, which started in the most developed countries at the beginning of the 19th century (United Nations, 2002). This was mainly as a result of the demographic transition characterized by significant reductions in fertility and mortality especially among infants and children. The reduction in mortality was occasioned by the epidemiologic transition as well as improved nutrition, improvement in health care, education and income (Ramashala, 2002; Mba, 2007). This phenomenon is now being experienced in much of the developing countries including those in sub-Saharan Africa.

The number of older people in Africa alone will be between 204 million and 210 million by the year 2050 (Help Age International, 2002). The rise in the overall proportion of the elderly populations in this poor world region is likely to create new challenges for the future. The lack of social security and the devastating effects of emerging and re-emerging diseases is likely to make the situation of the elderly worse in a region which has greatly relied on the traditional family structure as a safety net for the care of the vulnerable including the elderly. South Africa is one of the few sub –Saharan African countries with the highest proportions of the older populations with nearly 7 percent aged 60 years and above and is expected to rise to nearly 10 percent by 2025 (Kinsella and Ferreira, 1997; Ferreira and Van Dongen, 2004; Velkoff and Kowal, 2007:10). Most of the elderly in these countries live in rural areas where services and resources to reduce address vulnerability to aging remain low.

In the era of HIV/AIDS the situation of the elderly population has further deteriorated. In the past two decades or so several households have lost their breadwinners to the pandemic, with the elderly having to stay at home to look after relatives that are sick. The health sector has also been put under enormous strain with the demand for care for those living with HIV rising (Kanab us et al., 2007:2). The rising HIV prevalence rates among pregnant women and in the general population in five of the country's nine provinces namely KwaZulu Natal, Gauteng, Free State, Mpumulanga and North West now estimated at over 25% is likely to make the situation worse. It is estimated that 5.3 million South Africans were living with HIV at the end of 2002, which indicates that AIDS deaths are expected to continue rising.

The most devastating consequences of the HIV/AIDS pandemic arise not simply because many people die but because the deaths occur mainly among adults between the ages of 25 and 45 years, who are expected to support families including the elderly, thereby leading to poverty and sometimes destitution not only among the elderly, but also among children. Poverty is one of the biggest concerns that older people face and therefore a threat to their overall well being (UNFPA, 2002: 29). Only a few sub-Saharan countries like South Africa, Namibia and Botswana offer a pension to their elderly citizens which is not adequate because it is shared among extended family members, most of whom are unemployed (Economic Commission for Africa, 2001:2; Moller and Ferreira, 2003:25)) and some of them are infected with HIV/AIDS.

Accordingly the elderly people have now assumed the role of caregivers, having to take care of their grandchildren and at the same time, look after their sick children. A number of studies indicate that in sub-Saharan Africa, older people, especially older women, irrespective of the situation, care for the majority of orphans. In South Africa 40 percent of orphaned children are living with their grandparents (Nhongo, 2004:4) depending on the pensions of the elderly. Despite the inadequacy of pensions, it is an important source of livelihoods to the elderly and their children and grandchildren (Charlton, 2000:3).

One of the biggest impacts of the pandemic is the financial implication on the elderly who have to assume the role of providing for the orphaned children. Because household income falls between 48% and 78% when a Household member dies from HIV/AIDS, which excludes the cost of funerals (Drimie, 2002:11), social pensions become a lifeline to elderly headed households and their grandchildren (Help Age International, 2004; Booysen et al, 2004:5). Without the pensions the situation of these children would be that of abject poverty as the money is required to provide clothes, bedding and education are some of the needs that the elderly were unable to adequately provide due to lack of funds (Help Age International, 2002; 2007). Furthermore, the needs of People Living with AIDS (PWA) tend to be difficult to meet because they require nutritious food, drugs and money for regular hospital visits, which are very expensive. Other demands include clean water, which in some countries may not be easily accessible, especially in rural areas.

The psychological impact of the pandemic on the elderly has also compounded their socio-economic problems. Providing care to people with AIDS (PWA) can be very stressful because of the unpredictability of the infection, uncontrollability of the HIV/AIDS symptoms and the debilitating effects of the disease. (WHO 2002:15). This activity can be a 24-hour shift, requiring the caregiver to fulfil roles of confidante, chauffeur, and housekeeper and in the case of the elderly, a parent (Nova Scotia, 2004; UNFPA: 2002:43). The loss of a child as the main source of support and the inability to care for their grandchildren can lead to increased stress levels, depression and sadness (Help Age International, 2004:14; 2007:8). This has affected the health of the elderly (Essop; 2006:1). The fear of the future of grandchildren also contributes to stress suffered by the elderly due to the stigma associated with the disease (Help Age International 2004:15; WHO, 2002:14). This paper therefore explores the rural-urban differences in the problems faced by the elderly in Mafikeng Local Municipality of South Africa.

2. METHODOLOGY

2.1 Study Design and Data Collection

Primary sources of data collected using a structured questionnaire through interviews with 506 elderly households was used. The elderly were randomly selected from rural and urban areas in the Mafikeng Local Municipality. A small pilot study was conducted in order to identify any potential problems with the questionnaire design and sampling and all the necessary adjustments were done to ensure that the data collected were robust.

The questionnaire had both a household and an individual section. The individual section focused on socio-economic, health, HIV/AIDS and care giving. At the end of the fieldwork all questionnaires were evaluated to check for consistency and completeness. Furthermore, focus group discussions were also conducted in both urban and rural areas.

2.2 Ethical Considerations

Permission was sought from the North West Provincial Department of Social Development to conduct this study. Local councillors and chiefs were also consulted and their support was sought in conducting the study. Informed consent was also obtained from the elderly respondents, before administering interviews and the confidentiality of data collected was ensured by not using the identity of the source of data and using the data for purposes for which it was collected.

2.3 Data Analysis

The unit of analysis was the elderly person. Elderly persons were defined as those men and women age 60 years and older. Two separate analyses were performed. The first used statistical approaches to examine the characteristics of the elderly and assess the level of association between identified indicators of social problems they experienced by place of residence using the chi-square statistic. The Chi-square statistic was chosen because of the categorical nature of the data collected. The second analysis was a qualitative FGD analysis which used the master sheet framework to identify the main problems faced by the elderly. The qualitative data was used to corroborate and explain the quantitative results.

3. RESULTS

3.1 Characteristics of the Elderly

The results of the study are presented in two parts. The first part presented the characteristics of the elderly from whom data was collected and this is presented in Table 1. The second part presents data in Tables 2 and 3, on the economic, health and psychological problems associated with HIV/AIDS reported by the elderly.

Table 1 presents the socio-economic and demographic characteristics of the elderly by gender. The table shows that most of the elderly were aged between 60 and 69 years and as expected the proportion of the elderly decreased with increasing age for both sexes. There are more elderly females in the oldest ages of 80 or older. As expected, the majority of the elderly had 4-6 children and 66.8%, representing the majority, reported that they have experienced mortality among their children. Table 1 also shows that the majority of the elderly live in rural (74%) compared to 27% in urban areas. Regarding marriage, 44% of the elderly were widowed, and most of those widowed were women. About 37% were married and living together and only about 6% were divorced.

3.2 Economic and Social Conditions of the Elderly

Table 2 presents the economic conditions of the elderly by place of residence. Economic conditions were reported by the elderly as some of the main problems they faced in the era of HIV/AIDS. A significant percentage of the elderly (58.3%) reported they were responsible for the provision of food and other essentials in their household. More (60%) rural elderly compared to urban (52.2%) elderly reported they provided their needs and that of their households. Regarding the main source of income, 79.6% of the elderly, of which the majority (85.8%) in rural areas and 62.7% in urban areas depended on government pensions. Most of these reported they did not have alternative sources of support.

However, the majority of the elderly, 67.4% reported their standard of life was good and only 16% reported that their standard of living was poor. As expected, most elderly people who reported their standard of living as good were in urban areas. Conversely most of those who reported their standard of living as poor were in rural areas, where the impact of HIV/AIDS is expected to be greatest. The majority of the elderly (75.9%) reported they cannot save any of their pension money or other incomes. Most of the elderly, nearly 48%, of which nearly 50% and 36% live in rural and urban areas respectively reported that having large households is one of the main reasons for their inability to save. Most of the people living on the pensions of the elderly were young children.

Table 1 Socio economic and Demographic characteristics of elderly by Gender

Characteristics of respondents		Gender: Male	Female	Total: Col %	Number
Age	60-69	50.3	52.2	51.6	261
	70-79	33.9	26.9	29.2	148
	80	15.8	20.9	19.2	97
Total		**100.0**	**100.0**	**100.0**	**506**
Number of Children ever had	0	4.1	5.4	4.9	25
	1-3	32.7	32.2	32.4	164
	4-6	39.2	37.6	38.1	193
	7	24.0	24.8	24.5	124
Total		**100.0**	**100.0**	**100.0**	**506**
Children dead	Yes	56.7	71.9	66.8	338
	No	43.3	28.1	33.2	168
Total		**100.0**	**100.0**	**100.0**	**506**
Place of residence	Urban	23.4	28.1	26.5	134
	Rural	76.6	71.9	73.5	372
Total		**100.0**	**100.0**	**100.0**	**506**
Marital status	Married/together	62.0	23.9	36.8	186
	Divorced	7.0	5.4	5.9	30
	Widow/widower	22.2	55.8	44.5	225
	Never married	8.8	14.9	12.8	65
Total		**100.0**	**100.0**	**100.0**	**506**

Table 2 Economic conditions of the elderly by place rural-urban residence

Economic characteristics		Place of residence: Urban	Rural	Total: Col %	Number
Provider of food and household needs	Self	52.2	60.5	58.3	295
	Spouse/partner	31.3	12.1	17.2	87
	More than one person	16.4	27.4	24.5	124
Total		**100.0**	**100.0**	**100.0**	**506**
Main source of income	Salaries and wages	9.0	6.7	7.3	37
	Gov't pension	62.7	85.8	79.6	403
	Private pension *	20.9	3.5	8.1	41
	Other	7.5	4.0	4.9	25
Total		**100.0**	**100.0**	**100.0**	**506**
Standard of living	Good	77.6	63.7	67.4	341
	Better	17.2	16.4	16.6	84
	Poor	5.2	19.9	16.0	81
Total		**100.0**	**100.0**	**100.0**	**506**
Save money	Yes	53.9	14.3	24.1	112
	No	46.1	85.7	75.9	353
Total		**100.0**	**100.0**	**100.0**	**465**
Reasons for not saving	Inadequate pension	45.3	22.0	25.5	90
	Debt payment	18.9	28.3	26.9	95
	Large households	35.8	49.7	47.6	168
Total		**100.0**	**100.0**	**100.0**	**353**

Table 3 Health Status of elderly by Place of Residence

		Place of residence			Total
Health indicators		**Urban**	**Rural**	**Col %**	**Number**
Health status	Good	22.4	18.3	19.4	98
	Moderate	62.7	50.0	53.4	270
	Bad	14.9	31.7	27.3	138
Total		**100.0**	**100.0**	**100.0**	**506**
Difficulty carrying out household activities	None	61.2	35.5	42.3	214
	Moderate	29.1	52.7	46.4	235
	Severe	9.7	11.8	11.3	57
Total		**100.0**	**100.0**	**100.0**	**506**
Starving elderly	Yes, often	10.4	27.2	22.7	115
	No	89.6	72.8	77.3	391
Total		**100.0**	**100.0**	**100.0**	**506**

3.3 Health Conditions of the Elderly

In Table 3 data on the health status of the elderly is presented. As expected a significant percentage of the rural elderly reported that their health was bad (32%) compared to only 15% in urban areas. Overall urban elderly have a better health status than rural elderly. The majority of the elderly reported that they experienced moderate problems in conducting household chores. Only 11.3% reported that they experience severe difficulties in conducting their household chores. Regarding the food status of the elderly, only 22.7%, of which 27.2% were in rural areas reported that they experienced problems of food leading to frequent starvation. This is compounded by the presence of sick and young children in the household.

3.4 HIV/AIDS Care and Support

Data on HIV/AIDS is presented in Table 4 and it shows that 25% of the elderly reported that they have ever lost a household member to HIV/AIDS. Slightly more rural elderly reported they lost a household member to HIV/AIDS than the urban elderly. Furthermore, 19% of the elderly reported that they were the main care provider in their households. More elderly people in rural areas (21.2%) reported they were the main care providers in their households than in urban areas (13.4%).

The elderly also provided other needs to those sick of HIV/AIDS including moral support (24.7%), Comfort for the sick (12.4%) and financial support (25.8%). The most serious problem faced by the elderly is the provision of care for their grand children. Many of the elderly reported that the child support grants offered by the government did help to cover the needs of children. One of these is the rising cost of education. One elderly woman commented said: "*This child grant we receive is barely enough to send my grandchildren to school. Soon they will have to stop going altogether because of the rising cost of education*".

Most urban elderly caregivers complained that the purchase of basic necessities was expensive and with little or no support at all, it was becoming extremely difficult to provide for themselves and their households.

"*I have to pay rent for this house amongst so many other expenses I have. The pension I get cannot cover everything*." said an elderly woman.

Table 4 Impacts of HIV/AIDS on elderly by place of Residence

		Place of residence			Total
		Urban	Rural	Col %	Number
Any household members lost to HIV/AIDS	Yes	19.8	27.1	25.1	118
	No	80.2	72.9	74.9	353
Total		**100.0**	**100.0**	**100.0**	**471**
Elderly main care provider	Yes	13.4	21.2	19.2	97
	No	86.6	78.8	80.8	409
Caring orphans	Education	23.5	8.3	11.2	10
	Health	23.5	15.3	16.9	15
	Basic necessities	41.2	54.2	51.7	46
	All the above	11.8	22.2	20.2	18
Health needs of terminally ill	Financial support	11.8	29.2	25.8	23
	comfort, motivation	29.4	8.3	12.4	11
	Moral support	41.2	20.8	24.7	22
	Free medication	17.6	41.7	37.1	33
Total		**100.0**	**100.0**	**100.0**	**506**

Nearly all caregivers have indicated that they have been financially drained by the unforeseen circumstances of providing for their grandchildren. Their pensions and child grants are simply inadequate to provide all the needs required. "*I wish the government can at least subsidize school fees for our children. At least we may be able to save. Our situation would be much better like before*" Complained an elderly woman.

In desperate situations some elderly reported they have to go to their colleagues for assistance with the basic necessities like food to feed the young ones. This has prevented most caregivers from being destitute. Most elderly caregivers would prefer if they went hungry. A rural 63 year old elderly caregiver put her situation thus:

"I don't know how I can adequately provide for my grand children. They won't be able to go to school because I have no money".

Psychological problems were also reported by the elderly as some of the main problems they experienced in caring for children which often left them stressed and frustrated. A 65 year old elderly urban woman indicated that looking after children, especially those affected by HIV is very stressful. She reported that her experiences with her grand children are quite difficult saying: *"I have two young grand children, one three year old girl and the other five year old boy. The girl cries a lot, which irritates me a lot but I have to be patient and take time to keep her quiet"*

However, none of the elderly care givers reported suffering discrimination due to their care-giving activities. This may be due to the fact that people are increasingly beginning to better understand how HIV/AIDS was spread, thus minimizing the stigma surrounding the disease. Also they get community support for their efforts in taking care of their sick adult children and orphaned children. It is a known fact that grand children are usually left in the hands of their grandparents particularly those staying in rural areas.

4. DISCUSSION AND CONCLUSIONS

This paper examined the problems faced by the elderly in the era of HIV/AIDS. The results have shown that most elderly live in poverty and are psychologically stressed leading to a poor quality of health. Irrespective of place of residence, most elderly respondents were not in a position to save since the majority relied on government pensions which was the main source of sustaining their livelihoods. Most of the elderly receive a government pension of less than R1000 as their monthly pension. This pension is used to support their extended families that includes adult children and their grandchildren especially those who are affected by HIV/AIDS.

Providing care for the sick and the orphaned children requires a lot of socio-economic and moral support, yet the results in this study show that the elderly, the primary caregivers receive no or very little support to care for affected relatives at a time when they also need care and support. The support the elderly provide include physical and emotional care of the sick, provision of health care, feeding and provision of education to the orphaned children. These forms of care require significant financial and emotional resources which most elderly caregivers do not have. In addition, basic necessities like food and clothing were costly to provide for themselves as the meagre pension received by the caregivers were stretched to the limit. The recent rise in basic food prices and other commodities may have worsened the elderly caregivers' situation such that they now got less with the same amount of money. This situation may lead to elderly households becoming destitute; orphaned children dropping out of school thereby reducing their chances in life and increasing vulnerability to a new cycle of HIV/AIDS.

5. RECOMMENDATIONS

These findings have some important policy implications. Support to the elderly and orphans require urgent intervention in safeguarding the rights of the elderly people in their communities. The South African government has gone some way in creating an older persons bill. However this may require additional resources made available to the elderly. It will also require a more robust approach in dealing with the problems of orphan care giving and upbringing in the context of HIV/AIDS to reduce greater vulnerability of orphans to the fate that befell their parents. Community and neighbourhood solidarity with the elderly in the situation of HIV/AIDS needs to be seriously considered as a means of reducing the stress in care of themselves, the sick and orphans. Efforts to intensify efforts in providing basic services particularly health care in the rural areas where most elderly and orphans live should be addressed. The services provided must adequately cater for health needs of the elderly who mainly suffer from degenerative diseases with more emphasis place on how to maintain a healthy lifestyle.

REFERENCES

Age ways 2002, '*Violence and abuse: Practical issues and ageing and development*' Help Age International London N1 9ZN, United Kingdom, issue 59.

ANC today 2001, '*Rights of the elderly'* Available at: http://www.anc.org.za/ancdocs/anctoday/2001/at10.htm Accessed 20/08/2005.

Ayieko, S. 1998, "From Single parents to Child-headed Households: the case of Children orphaned by AIDS in Kisumu and Siaya Districts in Kenya, available at:

http://QBE_QUERY&XC=/dbtw_wpd/exec/dbtwpcgi.exe&BU=http%3a//www.eldis.org/search.htm

Booysen, F., Bachmann, M., Matebesi, Z., and Meyer, J. 2004, '*The Socio-economic impact of HIV/AIDS on Households in South Africa*: Pilot study in Welkom and Qwaqwa, Free State province'. University of the Free State and Centre for Health Systems Research & Development. Available at: http://www.jeapp.org.za/attachment.php?attachmentid=2

Charlton, K E.2000, 'Nutrition, health and old age- The case of South African Urban elderly', *South African Journal of Clinical Nutrition*, Vol. 13, no. 1, pp 13. Available at: http://www.saspen.com/supplement2000/oldage.htm. Accessed 31/07/2005.

Charlton, K. E. and Rose, D. 2001, 'Nutrition among Older Adults in Africa: the situation at the Beginning of the Millennium', *Journal of Nutrition*. 131:2424s-2428s. Available at: http://www.nutrition.org/cgi/content/full/131/9/2424s. Accessed 13/03/2005

Drimie, S. 2002, The Impact of HIV/AIDS on Rural Households and Land issues in Southern and Eastern Africa, a Background Paper prepared for the Food and Agricultural Organization, Sub-Regional Office for Southern and Eastern Africa. HSRC, Pretoria.

Economic and Social Commission for Asia and The Pacific: Expert Group meeting on Development Issues and Policies. December 2004. Available at: http://www.unescap.org/pdd/calender/EGM_dev2004/ageing.pdf. Accessed 21/10/2005.

Essop, P. 2006. Emotional stress of HIV/AIDS hits Elderly. Available at: http://www.globalaging.org/elderrights/world/2006/bill.htm. Accessed 17

Ferreira, M and Charlton, K (2004) "Ageing inequitably in South Africa." Available at: http://www.islamset.com/healnews/aged/monica.html.Accessed 06/07/2004.

Ferreira, M and Charlton, K (2004) "Ageing inequitably in South Africa." Available at:http://www.islamset.com/healnews/aged/monica.html.Accessed 06/07/2004

Fouad, D. 2005, *Role of elderly people in the era of HIV/AIDS in Africa*, Available at: http://www.globalaging.org/health/world/2005/nigeria.pdf.Accessed 03/05/2005.

Global Action on Aging. 2002. *Chronic poverty among the aged; United Nations integrated regional information networks*. Available at: http://www.globalaging.org/rural_aging/world/povety/htm. Accessed 15/03/2005.

Help Age International, 2002. "HIV/AIDS and Older people: The African Situation". African Regional Development Centre.

Help Age International, 2002. State of the World's older people, London United Kingdom.

Help Age International, 2004. *The cost of Love: Older people in the fight against AIDS in Tanzania*, Help Age International. Tanzania.

Help Age International 2006. "Older people and HIV/AIDS: facts and figures". Available at: http://www.helpage.org/reseaerchandpolicy/HIVAIDS/factsandfugures Accessed 23/06/2006.

Help Age International 2007, Supporting Older carers affected by HIV/AIDS in Africa, Available at: http://www.helpage.org/worldwide/Africa/keyprojects/supportingoldercarersaffected by HIV/AIDS. Assessed 22/03/2008.

Joubert, J. and Bradshaw, D. 2003, "Health of older persons". South African medical Research Council Available at: http://www.hst.org.za/uploads/files/chap11_03.pdf

Accessed 13/08/2006.

Kanabus, A., Fredriksson-bass, J. and Pembrey, P. 2007, HIV *& AIDS in Africa*, Available at: http://www.avert.org/aafrica.htm accessed on 12/02/2008

Kazeze, Z.W. 2005, "Poverty and Ageing in Africa." Paper presented at the Help Age International training Course on Ageing in Africa. Available at: http://www.sarpn.org.za/documents/d0001147/p1264-poverty_africa_kazeze_febr2005.pdf accessed on 12/03/2005

May, J. 2003, "*Chronic poverty and older people in South Africa*." Chronic Poverty Research centre working paper http://www.chronic poverty.org/ Accessed on 22/04/2006.

Mba, C.J 2007, Gender Disparities in living arrangements of older people in Ghana: Evidence from the Ghana Demographic and health Survey. Journal of International women's studies Vol 9 no 1

Nhongo, T. M. 2004, *Impact of HIV/AIDS on generational role and intergenerational relationships*, Paper presented at the workshop on HIV/AIDS and family well-being. Namibia, 28-30. Available at: http://www.un.org/esa/socdev/unyin/workshops/windhoekbackgroundpaper2.pdf

Pick'n Pay- Talk: The Elderly. Available at: http://www.pnp.co.za/pnp/view/pnp/en/page9689 Accessed on 25/02/2007

Ramashala, M. 2004, *Living arrangements, poverty and the health of older people in Africa.* Available at: http://www.un.org/esa/population/publication/publications/bulletin42_43/ramashala.pdf. Accessed on 29/06/2004.

Tewdoros, A (2004). "Older people in conflict situations: experience from Africa''. http://www.globalaging.org/armedconflict/popsideevent/tewdoros.htm Accessed 04/05/2005

The Daily Star (2005) Ageing: A creeping Problem for future Society? Available at: http://www.globalaging.org/elderrights/world/2005/creeping.htm Accessed on 25/07/2005.

UNAIDS 2003, *Sub-Saharan Africa: AIDS epidemic update*. Available at: http://www.unaids.org/html/pub/publications/irc-pub06/ Assessed on 02/07/2004.

United Nations Population Fund 2002, Situation and voices: the older poor and excluded in South Africa and India. Available at: http://www.eldis.org/static/doc10921.htm. Accessed on 25/01/2005

United Nations 2002, *HIV/AIDS and Older people*, World assembly on global ageing. Available at: http://www.globalaging/org/waa2/articles/hivolder.htm Accessed on 24/05/2005.

Velkoff. A. and Lawson. V.A., 1998. "Gender and Aging: International brief." United States Department of commerce economics and statistics administration. http://www.census.gov/ipc/prod/ib-9803.pdf. Accessed on 21/06/2008

Velkoff, V A. and Kowal, P.R. 2007, *Population Aging in Sub-Saharan Africa: Demographic Dimensions 2006*, U.S. Government printing Office, Washington DC.

World Health Organization.2002, "*Impact of AIDS on older people in Africa: case study Zimbabwe*". Switzerland.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 17–24.

EXPERIENCE OF ABUSE IN OLD AGE: AN EMERGING CONCERN

Tengku-Aizan Hamid[1], Siti Farra ZA[1] & Nurizan Y[2]
[1] Institute of Gerontology, Universiti Putra Malaysia
[2] Faculty of Human Ecology, Universiti Putra Malaysia

ABSTRACT

This paper discusses findings from the pioneer survey on elder abuse in Malaysia under the sponsorship of the Ministry of Science, Technology and Innovation (MOSTI), Malaysia. The findings reveal that majority of them live on the edge financially, and dependent on family members for support. Furthermore, single living and institutional placement of older persons, which have been practice recently, will also lead for societal hidden problem such as elder abuse to occur.

1. INTRODUCTION

The full impact of an ageing population will hit Malaysia in 2030, where 15% of her total population is projected as older persons by the United Nations (2006). As recorded in Population and Housing Census of Malaysia 2000, the number of older persons stands at 1.45 million, making up about 6.2% of the total population. Pala (2005) noted that the increasing size and proportion of the older population is a global trend which will persist well into the future. Already, the proportion of elderly population in the country is estimated as 6.9% during 2005 and is expected to double 20 years later. This is because of increasing life expectancy from 70.8 years to 76.9 during that time other than advancement in medical technology.

With a rapidly growing elderly population, the incidence of elder abuse can be expected to increase. Unfortunately, the exact incidence of elder abuse remains a hidden phenomenon in the country. As it is most often not reported (Swargerty, Takahashi & Evans, 1999), a true nationwide picture of elder abuse cannot be accurately drawn. Despite no organize documentation, media reports, journalistic articles and social welfare data; manifest sound evidence of elder abuse in the country. However it remains scanty and untold due to lack of awareness in its culture.

Due to limited research and reporting on elder abuse, Institute of Gerontology, UPM (the only research entity on aged and ageing in Malaysia) has recognize the need to conduct research on the subject entitled ‘Perception, Awareness and Risk Factors of Elder Abuse of Elder Abuse.’ This two-year research commences in 2007. In the absence of any local data on elder abuse (Esther, Shahrul & Low, 2006), primary data from this research should be at least able to inspire and facilitate commitment to curb this problem. No doubt there is a lot more to be done, this paper can provide an overview of elder abuse in the country. It addresses socio-demographic profiles of the elderly and types of abuse experienced by them.

2. WHAT IS ELDER ABUSE?

Report of the United Nations Secretary-General, which was presented at the Second Worlds Assembly on Ageing defined elder abuse as ‘a single or reported act, or lack of appropriate action occurring within any relationship where there is an expectation of trust, which causes harm or distress to an older person'. However, the definition of elder abuse varies within countries due to norms, values and cultures (American Psychological Association, 2006). Similarly, it is most often defined as an action by a person in a position of trust, which causes harm to an elder. Across the world, elder abuse has been commonly divided into five main categories (Table 1.0).

Table 1 Categories and definition of elder abuse

Categories of elder abuse	Definition
Physical abuse	The infliction of pain or injury, physical coercion, physical/chemical restraint.
Psychological or emotional abuse	The infliction of mental anguish.
Financial or material abuse	The illegal or improper exploitation and/or use of funds or resources.
Sexual abuse	Non-consensual contact of any kind with an older person.
Neglect	Intentional or unintentional refusal or failure to fulfil a care-taking obligation.

Source: WHO/ INPEA (2000)

3. EXPERIENCE OF ELDER ABUSE

Being no adequate awareness on the problem in most developing countries, the available data on prevalence of elder abuse globally has relied on five community surveys conducted in developed countries as stated in Table 2.0. The accepted prevalence rates of abuse among older persons which ranges from four to six percent were drawn from this survey (The World Report on Violence and Health, 2002).

Table 2 Prevalence of elder abuse in five developed countries

Country	Prevalence (%)
Netherlands	5.6
Finland	5.4
United Kingdom	5.0
Canada	4.0
United State of America	3.2

At the national level, there is no proper record on the incidence of elder abuse (National Report on Violence and Health Malaysia, 2006). Official correspondence from the Royal Malaysian Police also signified non-existence of record on the subject. Available data and information on elder abuse is only available from the Department of Social Welfare. However, it is limited to the number of older residence living in institution which recorded a significant increase throughout the year.

4. CHARACTERISTICS OF THE ABUSED

Phenomena of elder abuse involve all social economic background. Most studies highlight elder at risk are often female (Penhale, 1998), widowed, frail, cognitively impaired and chronically ill (Rounds, 1999). The National Elder Abuse Incidence Study (1998) concludes that elderly persons who are unable to care for themselves and/ or are mentally confused and depressed are especially vulnerable to abuse.

5. METHODOLOGY

Study sample derived from the general population, not among identified victim of elder abuse. Sampling frame of this survey was obtained from the Department of Statistics, Malaysia based on 90% confidence interval, 10% margin error, design effect value of two, 80% response rate and 168 enumeration blocks. A total of 1 344 respondents were selected with 366 at each zone. The respondents were than divided into two categories (i) adult population aged from 20 to 59 years and (ii) older persons aged 60 years and over. As a result, 1 079 respondents were successfully interviewed with 599 adult respondents and 480 older persons. With aim to present data on experience of elder abuse, this paper utilized the sub-sample of 480 older persons aged 60 years or over living in the community.

This cross-sectional survey was conducted in the state of Perak, Malacca, Kelantan and Selangor, which represent four geographical zones in Peninsular Malaysia. Data collection was conducted in January to May 2008 through a face-to-face interview method using enumerator-administered questionnaires. Respondents were asked whether they have ever experienced any abuse from the age of 50 years and onwards. The experience of elder abuse was measured by 19 items with pre-determined response categories of (1) Yes, (2) No and (3) Not applicable. This instrument measures types of financial abuse (4 items), emotional abuse (5 items), physical abuse (4 items) and sexual abuse (6 items). The items were developed by the research team based on recent literature.

6. RESULTS

Finding shows that 26% respondents reported having experienced at least one incidence of abuse since the age 50 years, with emotional abuse being the most prevalent type, followed by financial/ material, physical and sexual abuse (Figure 1.0). The number is relatively high due to courteous norm in the society and value of 'filial piety' embedded in traditional families. It might also be contributed by the religious line of Malaysian with majority of its population being Moslem. A Moslem is required, as far as Islam is concerned, to be dutiful to his parents and to pay them the gratitude they deserve. It is cited in the Holy Al-Quran Al-Karim, "Your Lord has decreed, that you worship none save Him, and (that you show) kindness to parents. If one of them or both of them attain old age with you, say not 'fie' unto them nor repulse them, but speak unto them a gracious word" (Surah Al-Isra' (17):23).

The respondent's response on the item is in Figure 2.0. Item that received the highest response is 'scolded', while no answer is recorded in most item of sexual abuse.

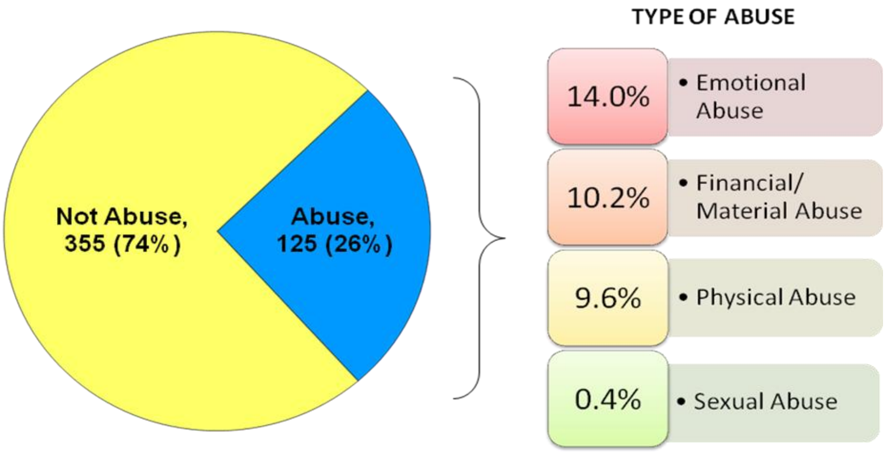


Figure 1 Experience of abuse among the respondents (N = 480)

Experience of elder abuse was measured by the following 19 items develop by the research team based on recent literature:

1. Money, property or assets used, taken, sold or transferred without consent
2. Signature forged on cheque or other financial document
3. All pension given to children
4. Expenditure controlled by carer
5. Stalked or followed around
6. Threatened with punishment
7. Yelled
8. Scolded
9. Called with insulting name
10. Force to eat
11. Tied or locked in a room
12. Hit, kicked or slapped
13. Threatened with knife
14. Coerced nudity
15. Shown pornography movie
16. Forced to photograph in sexual explicit pose
17. Exposed to private part
18. Molested
19. Raped

Elder at risk of abuse is associated with several socio-demographic characteristics. Table 3.0 presents background characteristics of respondents who ever abused since the age of 50 years. The distribution of the respondents according to age structure shows majority of them is of 60 to 74 years old. Most of the respondents who experienced abuse are female, similar as noted in the literature. The ethnicity of respondents is biased toward the Malays because the larger proportion (76.5%) of the sample is Malays. More than half are married with an average of five children and household size of four. Great variation is also noted in education attainment, employment status and house ownership among the respondents.

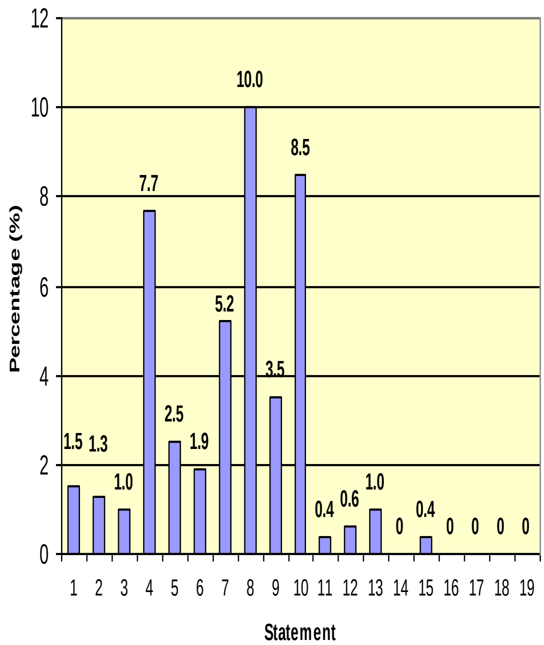


Figure 2 Respondent's response (%) on statement of elder abuse (N = 480)

Table 3 Socio-demographic characteristics of the respondents ever experienced abuse (n = 125)

Variables	***N***	**(%)**
Age (years) [M = 69.42, SD = 8.311]		
Young - old (60 - 74)	94	75.2
[M = 65.51, SD = 4.517]		
Old - old (75 - 84)	23	18.4
[M = 78.61, SD = 2.824]		
Oldest - old (>85)	8	6.4
[M = 88.88, SD = 3.482]		
Sex: Female	70	56.0
Male	55	44.0
Ethnicity: Malays	89	71.2
Non-malays	36	28.8
Marital status: Married	70	56.0
Not married	55	44.0
Number of child(ren) [M = 5.92, SD = 2.778]		
Household size [M = 4.32, SD = 2.555]		
Education level		
Formal education	79	63.2
No formal education	46	36.8
Employment status: Unemployed	111	88.8
Employed	14	11.2
House ownership		
House ownership: Owner	85	68.0
No ownership	40	32.0

Another factor that has been associated with elder abuse is the choice of living arrangement. This factor may contribute to the occurrence of abuse from the interaction with the members living together which may result conflict within them. As presented in Table 4.0, result of the study shows that 48.6% female were staying with their child(ren)/ son in-laws/ daughter in-laws/ grandchild(ren), while 36.4% male were staying with spouse and children. This is due to longer life expectancy among Malaysian female (76.2 years) as compared to their male (71.5 years) counterpart (Department of Statistics, 2006).

Table 4 Living arrangement by sex of the respondents ever experienced abuse (*n* = 125)

Living arrangement	**Sex (%)**	
	Female (n = 70)	**Male (n = 55)**
Alone	12.9	1.8
Spouse only	8.6	21.8
Child(ren) only	11.4	5.5
Spouse and child(ren) only	11.4	36.4
Child(ren)/ Son in-laws/ daughter in-laws/ Grandchild(ren)	48.6	32.7
Relatives	1.4	1.8
Neighbors/friends	5.7	0
Total (%)	**100**	**100**

Income data indicate that almost three quarter (Figure 3) of the respondents are in the poor category based on the national Poverty Line Income (PLI). According to the Economic Planning Unit, 2007 the PLI for Peninsular Malaysia is MYR 720 (1 USD = 3.40 MYR).

The measurement of income is most often under-reported especially among the women's that perform home duties, which is always referred as 'informal work'. However, income is an important variable in predicting the decision-making status and control of earning.

Chi-square (χ^2) test was utilized to measure the association of socio-demographic variables and the type of abuse. The analysis (Table 5.0) showed that financial abuse was significantly related to the income category (χ^2 = 4.225, p = 0.040), house ownership (χ^2 = 7.456, p = 0.006), employment category (χ^2 = 5.298, p = 0.021) and sex (χ^2 = 4.573, p = 0.032) of the respondents. There was also significant relationship between employment category (χ^2 = 4.616, p = 0.032), marital status (χ^2 = 8.758, p = 0.003) and sex (χ^2 = 4.212, p = 0.040) with physical abuse. Overall, the incidence of abuse (all types) was linked to marital (χ^2 = 9.791, p = 0.002) and employment (χ^2 = 5.462, p = 0.019) status of the elderly. The results showed that experience of financial abuse is more common for older persons without their own house and married elderly are more at risk for physical abuse.

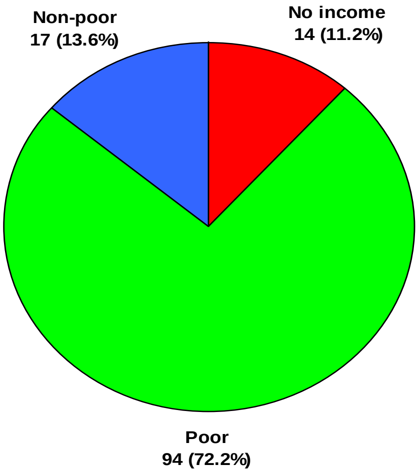


Figure 3 Monthly income (MYR) of respondents ever experienced abuse based on Poverty Line Income 2007, Malaysia (n = 125)

Table 5 Chi-square (χ^2) analysis

Dependent Variable	Independent Variables	χ^2	p
Financial Abuse	Monthly income	4.225	0.040*
	House ownership	7.456	0.006**
	Employment status	5.298	0.021*
	Sex	4.573	0.032*
Physical Abuse	Employment status	4.616	0.032*
	Marital status	8.758	0.003**
	Sex	4.212	0.040*
All type of Abuse	Marital status	9.791	0.002**
	Employment status	5.462	0.019*

7. CONCLUSION AND RECOMMENDATIONS

In conclusion, elder abuse cuts across all socio-economic background. However, married and unemployed older persons are more likely to experience abuse in later life. There is also no significant different between various type of abuse with age, ethnicity and education level. Due to limitations of the study, it was not possible to make out the sequence of events before and after a reported abuse in this analysis. Besides, the relationship of the perpetrator to the older persons was not identified and the study did not determine a specific time frame for the occurrence of the incidence. Although this study has limitations, the research was able to elucidate the phenomenon of elder abuse.

Research is important to better understand the problem in diverse cultural perspective. The available data is then useful to make recommendations that may be valuable in promoting and sharing information on elder abuse as well as its prevention in the country. Important need in preventing this incidence to occur is reliable information on its socio-demographic of the victims and perpetrators. Thus, further study is required to identify the characteristics of perpetrators on the issue in the country.

REFERENCES

Al- Quran. Surah Al-Isra' (17):23.

American Psychological Association (2006). Public Interset Directorate (http://www.apa/pi/aging/eldabuse.html). Retrieved on 19 Nov 2009.

Department of Statistics, Malaysia (2006). Abridged Life Tables, Malaysia, 2002 - 2005. Putrajaya: DOSM.

Department of Statistics, Malaysia (2001). Population Distribution and Basic Demographic Characteristics. Population and Housing Census of Malaysia, 2000. Putrajaya: DOSM.

Economic Planning Unit (2007) Available from Ministry of Housing and Local Government Website, http://www.kpkt.gov.my/kpkt/fileupload/PENDAPATAN_GARIS_KEMISKINAN2007.pdf

Esther, E., Shahrul, K. & Low W.Y. (2006). Elder abuse: a silent cry. Malaysian Journal of Psychiatry. March, vol. 14, no. 1.

Krug, E.G., Dahlberg, L.L., Mercy, J.A., Zwi, A.B. & Lazano, R. (eds.) (2002). World Report on Violence and Health. Geneva, World Health Organization.

National Center on Elder Abuse (1998). The National Elder Abuse Incidence Study. American Public Human Services Association. Washington, DC: United States.

Pala, J. (2005). Population ageing trends in Malaysia. Monograph Series, Population Census 2000. Putrajaya: Deparment of Statistics, Malaysia.

Penhale, B. (1998). Bruises on the soul: older women, domestic violence and elder abuse. Bold. 8(2): 16-30.

Rounds, L. (1999). Elder abuse and neglect: a relationship to health characteristics. American Journal for Nurse Practitioners. Apr - Jun, 47-52.

Swagerty, D.L., Takahashi, P.Y. & Evans, J.M. (1999). Elder Mistreatment. American Family Physician, vol. 59, no. 10, pg. 2804-2808

United Nations (2006). World Population Prospects: The 2006 Revision. Available from Population Division, Department of Economic and Social Affairs Website, http://esa.un.org/unpp/

WHO (2006). National report on violence and health Malaysia. Kobe, World Health Organization.

WHO/ INPEA (2002). Missing voices: views of older persons on elder abuse. Geneva, World Health Organization.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 25–44.

THE ROLE OF WOMEN IN LONG-TERM CARE PROVISION: PERSPECTIVES ON AGING IN THE ARAB AND ISLAMIC WORLD

Shereen Hussein
Senior Research Fellow
Social Care Workforce Research Unit
King's College London
Strand, London, UK, WC2R 2LS
E-mail: shereen.hussein@kcl.ac.uk

ABSTRACT

Populations are aging rapidly in the majority of the developed world and in many developing countries. However, longevity is often accompanied by many years of ill health and disability and in several countries there are declining numbers of people in working age groups. This paper focuses on the multiple roles of women in the labor force; as mothers, widows, single women and wives/partners, and as the main providers of long-term care for parents and other family members. Women are rendered particularly vulnerable in later life by their life expectancy (generally longer than men's) and the high possibility that many of these later years will be characterized by ill health and disability. This paper is based on a review of trends in aging and female labor force participation among a set of developed and less economically developed countries, with a focus on the Arab and Islamic countries. It also provides illustrations of some existing strategies used by women and governments to balance their increasingly competing multiple roles.

1. INTRODUCTION

Populations are aging rapidly in the majority of the developed world; moreover, it is predicted that by 2010 overall populations will start decreasing in countries such as France, Italy, and Japan. In 2004, more than half of the world's older people lived in just six countries: China, the United States, India, Japan, Germany and Russia (Kinsella and Phillips 2005). At the same time, many less economically developed countries suffer from the conjunction of rapidly aging and relatively poor populations. Many Islamic countries will experience demographic transition as part of global demographic changes toward longer life expectancy and lower fertility, but may be less equipped economically to address the multiple implications of such changes. However, very little research is conducted in this area; and relevant data on longevity, ill health at older age, and the use of formal care in the Arab and Islamic countries, are surprisingly sparse (Yount and Sibai, 2009).

Women in the Arab and Islamic world, as well as in other developed countries, live longer than men. In the most Arab and Islamic countries, this longevity gap ranges between 0 and 5 years, as illustrated in Table 1. Nonetheless, although Arab and Islamic women may outlive men, they are generally in worse health than other women in the developed world. The little data available indicate that a considerable part of the increased life expectancy in this region is spent in ill health. For example, Margolis and Reed (2001) found that higher rates of older people using formal care services in the United Arab Emirates suffered from neurological diseases and dementia than in other western countries such as the United States. Similar findings from Iran were illustrated by Tajvar and colleagues (2008), where their sample of older people highlighted the poor health conditions among elderly women and those with lower educational attainment.

Table 1 Life expectancy by sex in selected Arab and Islamic countries

	Life expectancy		
Country	Men	Women	Gender-gap
Turkey+	69.0	74.0	5.0
Malaysia+	72.0	77.0	5.0
Qatar*	71.4	76.2	4.8
Libya*	71.8	76.4	4.6
Egypt*	68.0	72.4	4.4
Morocco*	67.8	72.2	4.4
UAE*	76.7	81.1	4.4
Lebanon*	70.1	74.4	4.3
Kuwait*	75.4	79.7	4.3
Tunisia*	71.4	75.6	4.2
Indonesia+	69.0	73.0	4.0
Saudi Arabia*	70.3	74.2	3.9
Syria*	71.8	75.4	3.6
Mauritania*	51.5	54.7	3.2
Palestine*	71.1	74.2	3.1
Iraq^	57.4	60.5	3.1
Jordan*	70.2	73.2	3.0
Oman*	73.1	76.0	2.9
Sudan*	55.1	58.0	2.9
Bahrain*	73.2	76.0	2.8
Yemen*	59.7	62.4	2.7
Algeria*	70.1	72.7	2.6
Djibouti*	51.8	54.1	2.3
Somalia^	45.4	47.6	2.2
Bangladesh+	64.0	66.0	2.0
Pakistan+	66.0	67.0	1.0
Afghanistan+	44.0	44.0	0.0

* Source: United Nations Development Program (UNDP) Year 2004; ^ Year 2003
http://gender.pogar.org/countries/stats.asp?cid=14&gid=10&ind=120
+ Year 2009 estimates, Sources: PRB 2009 World Population Data Sheet
http://www.prb.org/Datafinder/Topic/Bar.aspx?sort=v&order=d&variable=122

Informal care providers, such as unpaid family members, as well as formal care providers, such as nursing aides, home care assistants, and other paid care workers, constitute two often parallel systems of support providing long-term care in the developed world. At the same time, both informal caregivers and formal care workers, such as home care workers and staff in residential and nursing homes, are in increasingly short supply in most of the developed world.

It is well documented that most care delivered to older people or people with disabilities, particularly in the Arabic and Islamic regions, is provided by families, mainly women, or by other informal caregivers (Rugh 1984, 1997; Yount and Rashad 2008). However, the continued availability of these (predominantly female) family members is uncertain, due to a number of interacting demographic and socioeconomic trends. Changes in family structure, migration and increases in the participation of women in the labor force, as well as other factors, are negatively affecting the availability and willingness of informal carers (WHO 2003; Hussein and Manthorpe 2005).

It is important to highlight that both population changes and women's position in society vary dramatically between the Arab and Islamic countries in the region. For example, Yount and Rashad (2008) highlight the striking variations in family structure and women's political and economic participation in Egypt, Tunisia and Iran. The current picture of family formation, encompassing marriage and fertility, is far from homogenous in these three countries and indeed in other countries in the region (Yount and Rashad, 2008; Hussein 2002; Hussein and Manthorpe 2007).

Nevertheless, it is fair to say that women within this context face great challenges both as receivers and providers of care. There are a number of competing demands on women's time and energy, ranging from carrying out maternal duties, being exemplary wives, and providing care for elderly parents or other relatives, to a growing demand that they participate in the labor force for both social and economic reasons. Such competing demands are increasing, due to many external factors, such as population changes, migration trends, and changes in family structure patterns, as well as the increased importance of female labor participation both at the micro and macro levels.

In this paper, data related to trends in long-term care needs as well as female labor force participation in the Arabic and Islamic regions are examined to highlight the competing demands made on women. Following this, a review of current policies adopted by governments, employers and families in more economically developed countries is conducted to provide some illustrative examples of possible strategies available.

2. POPULATION CHANGES

In the Middle East and most Islamic countries, despite the observed 'Youth Bulge' (Roudi-Fahimi and Kent, 2007), population aging is also occurring. Indicated by increased life expectancy that takes place in a context of socioeconomic change, paralleled by significant changes in family structure, as will be discussed further in the next section. However, policy and governmental support related to formal provision of long-term care are still in their infancy in most of these countries, as in many less economically developed countries, with an undeclared assumption that relies heavily on traditional informal care support. Shakoori (2008), in a statement to the United Nation's Commission on Population and Development, indicates that although the aging process in the Arab region is in its early stage, given its rapidity, the requirements to meet challenges with regard to number of older people have not been fully addressed and are likely to be underestimated.

Recent research in Egypt (Sinunu *et al.* 2008) indicates that many factors lead family caregivers to use formal care for their older relatives in Cairo, despite the longstanding norms of family and informal care provision. As in these accounts provided by Sinunu and colleagues from Egypt, Sibai and colleagues (2004) observe a ‘compressed’ demographic transition in Lebanon. Here, the effect of declining fertility and mortality rates effect has made itself felt in increased life expectancy. However, such changes, and further projected changes, are infrequently addressed by policy makers in terms of pension plans or formal long-term care provision.

A report by the United Nations (2007) highlights the importance of a ‘call for change’ arising from the recent demographic changes in most Arab countries, particularly in terms of changes in their age-structure, including a gradual increase of older people (65 years or more). Figure 1 shows the actual and predicted life expectancy at birth from 1995 to 2025 in some Islamic and Arab countries, compared to those in Japan and the United Kingdom.[1]

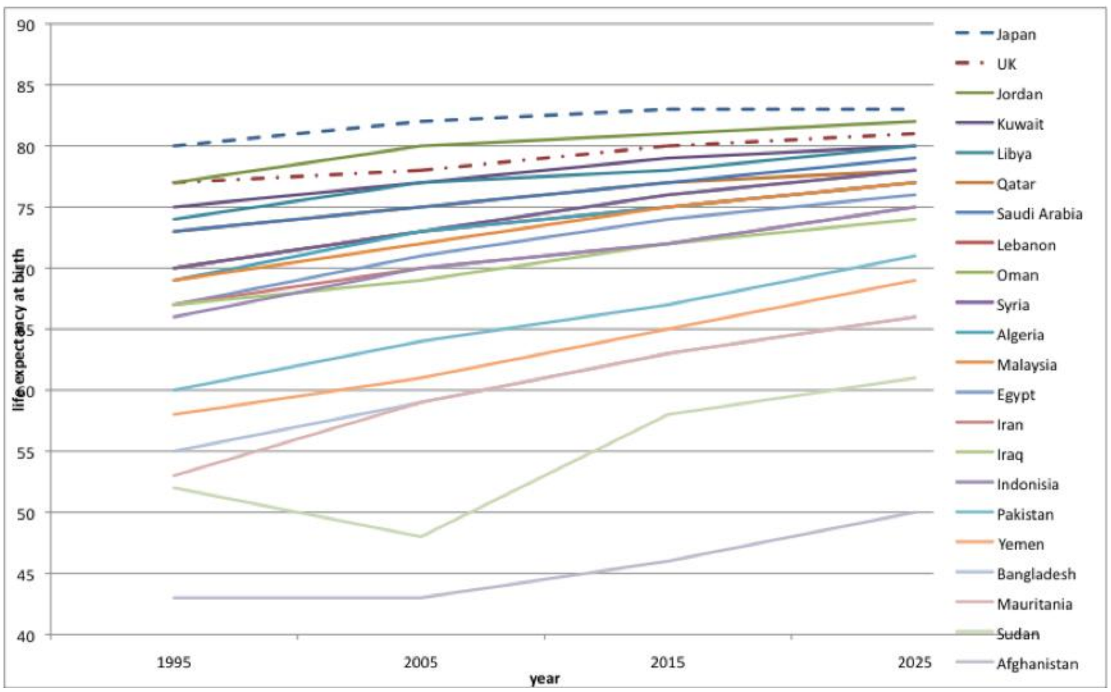


Figure 1 Actual and predicted life expectancy at birth from 1995 to 2025 in the Arab and Islamic world compared to those in Japan and United Kingdom

The data show that different Arabic and Islamic countries are at different levels of life expectancy, with the majority clustered above the 70 years mark from 2005 onward. At the top is Jordan, where life expectancy from 1995 onward equals or exceeds that of the United Kingdom. Two countries, Afghanistan and Sudan, have considerably lower life expectancy, mainly due to local military and political unrest. However, life expectancy is predicted to increase in all countries and will reach at least the 75 years mark in the majority of countries by 2025.

As discussed earlier in this paper, longevity is often accompanied by ill health and increased need for care provision. This is particularly so in the developing world; however, it is also

1 Source U.S. Census Bureau, International Database (accessed October 2009)

observed in many developed countries (Sinunu et al. 2008; Boggatz and Dassen, 2005; Sibai et al. 2004; Hussein *et al.* 2009). As is well documented, informal provision of care for older relatives is the norm, but extended hospitalizations and the admission to care facilities of frail, older adults have been occurring in many countries for at least the past two to three decades (Rugh 1984, Sinunu *et al.* 2008). However, there are no available statistics to establish the volume or trends of the use of formal care, mainly due to the fact that provision of such care is often poorly documented (Margolis and Reed 2001). Boggatz and Dassen (2005), in a review of the literature on formal care provision in Egypt, concluded that despite the scarcity of available data, long-term care provision is an emerging problem in Egypt and needs attention from policy makers. Similar recommendations are made by Sibai and colleagues (2004) in relation to Lebanon.

The fact that the majority of norms governing the region places the duty of informal care, by far, on women, whether unpaid or paid in relation to domestic services, needs further attention because women are faced with a number of competing demands including participation in the labor force. In the next sections the role of interacting factors such as migration, family structure and female labor force participations are examined in relation to the incrementing multiple demands on women in the Arab and Islamic regions.

3. MIGRATION

Migration within and outside the region is a significant factor in the demography of the Arab world. Since the second half of the 20th century migration within, between and outside the Arab world is very evident. Economic factors largely explain major movements within the region, for example, from Egypt and Yemen to Saudi Arabia and other oil-producing countries of the Gulf states (Farques 2006). Some countries, such as Yemen, are labeled ‘countries of emigration’, as large proportions of their population migrate to other countries.

Another important group of migrants in the region are refugees, from Palestinians and Iraqis in Syria to Afghanis in Iran. The region is considered by some to contain the largest refugee and asylum-seeking communities in the world (US Committee for Refugees and Immigrants 2006).

Moreover, large communities from the region have migrated outside the region, to Europe, US and Canada. Historical and cultural links with certain developed countries play an important part. For example, previous colonial ties to some European countries, such as Algeria and France and to a lesser extent Egypt and Britain, are important in understanding population flows. Economic ties also play a significant role, for example the recruitment of Turkish workers to Germany in the 1960s. More recently, many young people from the region seek migration to other developed countries not only for economic reasons, but also for better educational opportunities and to escape political unrest.

Such movements, particularly those that occur within the region, usually involve husbands leaving their wives and families in their home countries to maximize possible remittances to be sent home. Such activity places a considerable burden of care on women, who are expected to look after the family, including children and the elderly, while (in many cases) participating in full time employment. The effects of such phenomena are illustrated by the over-representation of working-age men in oil-rich countries, such as Qatar and Kuwait, while there is a deficit of working-age men in sending countries, such as Yemen (see Figures 1 and 2).

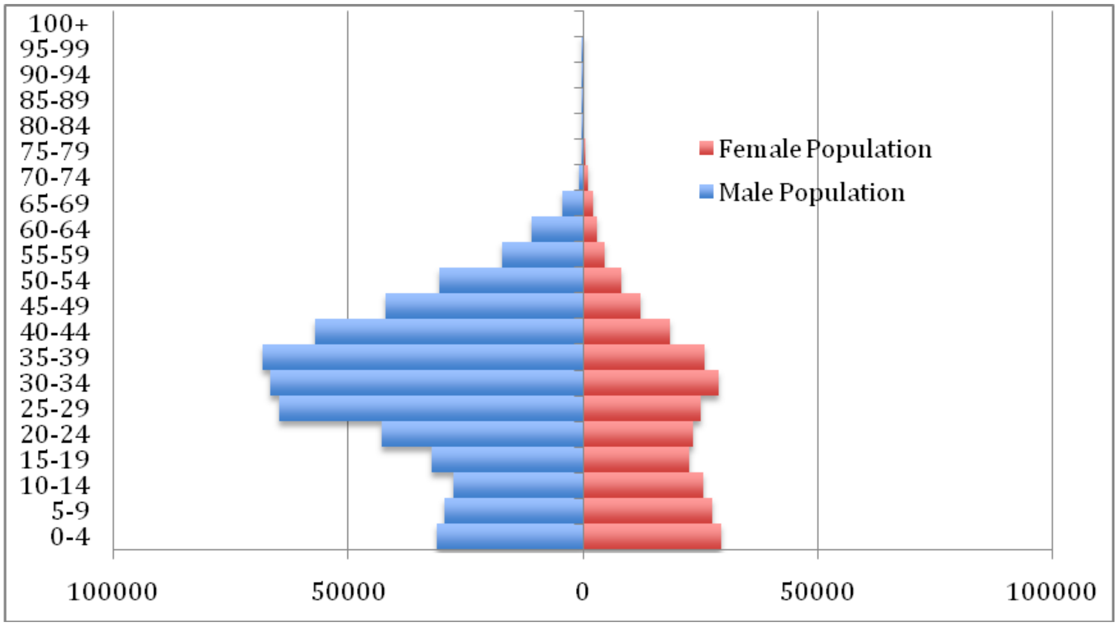


Figure 2 Age and sex structure, population Pyramid, Qatar 2005

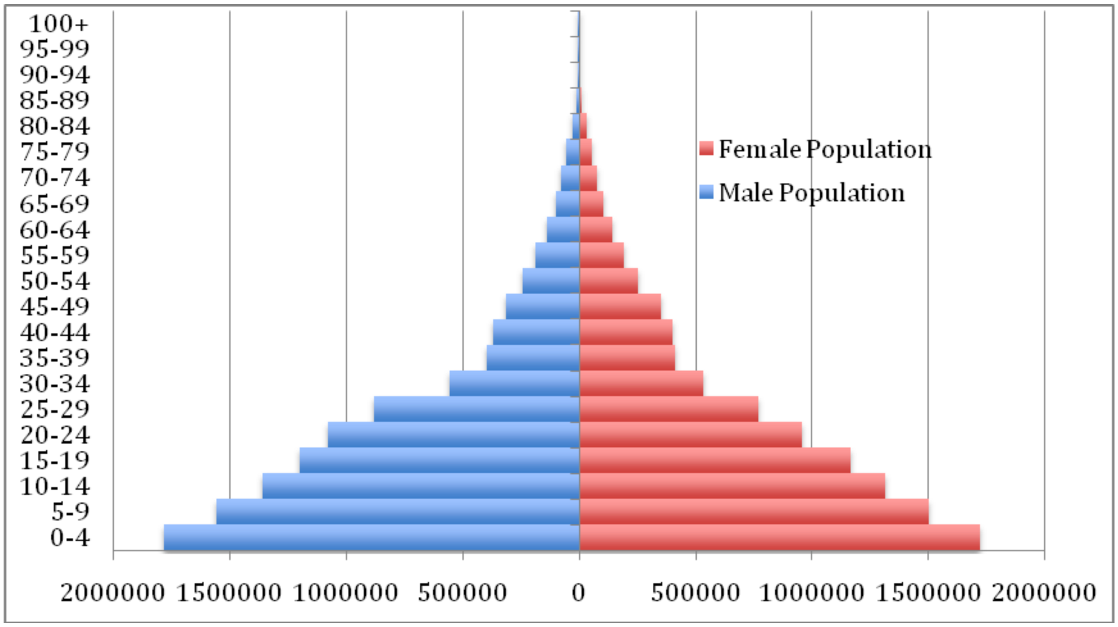


Figure 3 Age and sex structure, population pyramid, Yemen 2005
Source: US Census Bureau, International Data Base.

4. FAMILY STRUCTURE

Another important factor affecting women in Arab and Islamic countries is the shift in family formation and structure. Trends toward higher median age at first marriage, urbanization and a decline in co-residence with in-laws and extended families, as well as recent trends towards higher divorce rates, are observable in some countries (Hussein 2002, Rashad *et al.* 2005). These may lead to a higher probability of women spending their later life on their own, without a close relative who can provide care for them during this phase of life. This state of affairs may be rendered more probable as a consequence of increased migration within and outside the region, as explained above.

Given that marriage is the only socially and religiously approved context for sexuality and parenting in the Arab and Islamic world, family formation is of central importance to men and women in the region alike. Arab kinship structure is mostly described as being one of 'patrilineal endogamy' (marrying within a particular group related to the male line) and the existence of parallel cousin marriage is widespread. It is generally through marriage and having children that adulthood and self-satisfaction are achieved, for both men and women (Rugh, 1984 and 1997, Hoodfar, 1997, Hussein 2002, Hussein and Manthorpe 2007).

In the Arab region marriage is usually not viewed as a partnership between individuals but rather as an association between two families. In such a context, the families of potential partners are likely to make marriage decisions and offer choices of suitable partners. However, the extent of such practices varies widely between and within countries in the region. For example, women from regions in North Africa, which have a large proportion of Barbers, such as some regions of Algeria and Morocco, have different marriage patterns from other women living in different regions within the same country where the Barber population is smaller (Hussein, 2002). The local context usually influences decisions related to family structure, including age at marriage and co-residence after marriage (McNicoll, 1980; Ryder, 1983; Cain, 1985, Bener et al 1996, Joseph 1994). It is worth noting that due to the laws that govern marriage in the region, which give men and women different rights, kin and in-group marriages have been regarded as common strategies for protecting women in case their husbands abuse their rights in terms of divorce or polygyny (Rugh, 1984, 1997).

Currently most Arab countries are going through some sort of 'nuptiality transition' from one pattern of marriage to another, and different countries are at different stages of this transition. Some observe that the universality of marriage, which characterized the region for many decades, is starting to decline (Rashad et al, 2005, Hussein, 2002). Other forms of non-conventional marriages in the Arab world are also emerging. One form is *muta'a* (temporary marriage), which is practiced in some areas in the region, for example by Shi'ites in southern Lebanon and other areas (Hoveyda, 2005), where couples specify in their marriage contract the date upon which the marriage ends. In another emerging form of marriage *'Zawag Urfi'*, or undocumented marriage, the couple usually makes a written declaration stating that the two are married and two witnesses sign it. In most cases, this type of marriage is not declared and is kept secret between the partners; thus, in case of divorce, 'wives' have no right to support and are sometimes unable to prove that the 'marriage' existed. Some anecdotal evidence reports that this form is becoming popular among some groups of educated Egyptian youth (Rashad *et al.* 2005, Shahine 2005, Anonymous, 2005). It is suggested that this is due to the increasing cost of marriage and also as a way for young couples to achieve more autonomy in their marriage choices, albeit without declaring these choices and actions to their families. All these changes in the traditional family structure again risk leaving older women to spend their later lives alone, usually in ill health. This calls for the consideration of alternative long-term care approaches, including investing in formal care provision and considering better pension systems.

A number of recent changes in relation to family formation are observed during the last two decades. Recent trends indicate the women are waiting longer to marry and their spousal age-gap is declining. Although that some anecdotes from the media indicates a recent trends towards higher divorce rates in some countries such as Egypt, no published data can support such observations. Contrary to this, data up to year 2000 indicate that most Arab countries experience decline, albeit slight, in divorce rates (El Saadani, undated). Similarly, trends towards nuclear families and urbanisation are not observed everywhere, particularly not in some of the Gulf countries (El Haddad 2006).

5. LABOR FORCE PARTICIPATION RATES AMONG WOMEN

In any society, women's labor force participation is restricted by many obstacles: lack of education, discrimination in wage rates and employment practices, negative cultural attitudes and obligations to provide informal care for both children and older family members. In the Arab and Islamic world women's experiences in terms of empowerment, freedom of movement and labor force participation range on a wide spectrum from 'being strictly closeted, isolated and voiceless' to 'enjoy the right to work and participate in public affairs' (WLUML 1986; An-Na'im 1995; Syed 2008). Table 2 presents the proportion of adult females who participate in the labor force in different Arab and Islamic countries.[2] However, it is worth noting that women participate in many unrecorded work-activities inside and outside the household, including domestic, trade and agriculture work.

The Arab Human Development Report (UNDP 2002) identifies three deficit areas unique to the Arab region: deficits in freedom, in women's empowerment, and in knowledge. Female illiteracy is high, and female labor force participation, in some countries, is by far the lowest in the world. Saudi Arabia is an example where women have limited opportunity to hold employment or other public offices. However, some recent changes are being actively promoted to improve women's participation in recent years. For example, in 2005, women in Kuwait received the same political rights as men, which enabled them to vote and politically participate if they wished. However, such recent 'gains' need to be addressed within the wider political context in many Arab and Islamic countries and the lack of 'democratic' institutions.

In many countries in the Middle East, the issues of women's right to education, work, and political participation are still debated and are yet to be resolved (UNDP 2006). On the other hand, in a number of countries, such as Tunisia, gender equality in access to education has not been a significant issue. In some situations, women have equaled and even surpassed men in average educational attainment, with women outnumbering men studying in universities in Kuwait, Qatar and the UAE (Rubin 2007). However, high numbers of female university students do not usually equate to higher participation rates in the workforce.

Data presented in Table 2 show that documented labor force participation among adult females in Arab and Islamic countries in 2007 ranged from only 14 percent to 57 percent. It is likely that a much higher proportion of women in the region participated in other unpaid forms of work, such as agriculture, trade and domestic services, and these figures are considered underestimates of female labor force participations.

6. THE ROLE OF WOMEN IN LONG-TERM CARE PROVISION

The provision of long-term care is one of the most important issues facing most developed countries and many less developed ones as well. As shown in the previous sections, changing demographic structures, labor market dynamics, migration policies and systems of financing long-term care all influence care provision. Informal care continues to play a central role in long-term care, and women remain predominant in the provision of both formal and informal care.

---

[2] Source: UN Data: http://data.un.org/CountryProfile.aspx

Table 2 Documented labor force participation among adult females in different Arabic and Islamic countries

Country	Year 2000	Year 2005	Year 2007
Bangladesh	54.8	56.7	57.2
Indonesia	50.2	49.8	49.6
Malaysia	44.3	44.2	44.7
Kuwait	42.8	42.8	43.1
UAE	34.6	39.1	40.0
Algeria	31.2	35.0	36.9
Bahrain	34.0	33.5	33.6
Iran	28.3	30.7	31.8
Sudan	29.6	30.4	31.1
Libya	22.9	24.6	25.9
Oman	23.6	24.9	25.8
Tunisia	24.2	25.6	25.7
Lebanon	23.9	24.5	24.8
Morocco	26.5	25.3	24.7
Turkey	26.6	24.8	24.4
Egypt	21.1	23.3	23.8
Yemen	18.2	21.4	21.6
Syria	18.4	20.1	20.9
Pakistan	16.1	19.3	20.8
Saudi Arabia	16.6	18.4	19.1
Jordan	13.9	15.3	15.5
Iraq	13.0	13.9	14.2

During the past 15 years, some Arab and Islamic countries have introduced a limited system of state and private formal care provision. In most of the Islamic world the available data indicate that the use of formal care by older people is not substantial, but that it may be growing in extent. As in other countries, women are over-represented as employees providing formal care provision. The literature also stresses the sizeable role played by women in relation to informal long-term care, which is the most common form of care provision. Most long-term care is provided informally (for a minority of people with severe health conditions, in hospital settings), and women provide the majority of such informal care.

In most Arab and Islamic countries women work outside the home, whether paid or unpaid, while they simultaneously carry out most household, childcare and care for older relatives' tasks. These multiple roles persist, and during the past 15 years more emphasis has been placed on women's roles and duties as care providers as well as labor force participants. In the region, virtually all care for older people, when sick or disabled, takes place within families and is undertaken mostly by women. Even when hospitalization is necessary, bedside or personal care, such as help with eating and washing, is often provided by relatives rather than nursing staff; with the vast majority of hands-on care being provided by daughters and daughters-in-law. In a study by Ahmed and Abbas (1993) Egyptian women were found to be the main care providers for older people, whether parents or in-laws.

However, demographic, economic, and social factors have obvious and large effects on informal caregivers' availability (Miller and Weissert 2000, Hussein and Manthorpe 2005). For instance, if female longevity increases more rapidly than male longevity, as observed in some countries, women can expect to spend more years in widowhood, without the support of a spouse, leading to a greater risk of needing formal long-term care (Lakdawalla and Philipson 1999). Women's participation in the labor workforce also reduces their availability to provide informal care for parents, parents-in-law and grandparents; therefore the limited availability of potential providers of informal care is directly associated with increased demand for formal long-term care (Coughlin *et al.* 1990, Miller and Weissert 2000, Yoo et al 2004). As the case in the majority of the Arab and Islamic world, such formal care, is rarely available, and may be socially unacceptable, except in cases where hospitalization is necessary.

In most of the developed world, women provide the bulk of informal care as well, although with marked differences across countries (Hussein *et al.* 2009). In some countries, men are more likely to take over the role of carer or caregiver for their spouses than men in other family roles. However, this is unlikely to be the case in Arab and Islamic regions where the agreed 'norms' are that caring responsibilities are the duty of women. However, even in developed countries where men provide some of caring, women predominate among informal carers with the heaviest commitments. They are more likely to be the main carer rather than an additional carer. The greater the need for personal care services, the more likely it is that women provide them (OECD 2005).

In relation to the age of informal carers, across different developed countries there seems to be a peak in care giving amongst those aged 45-65 (Table 3). Due to the informal nature of care provision in the Arab and Islamic world and very limited data, the age of carers is not documented. However, given the low median age at marriage and childbearing of current older people, it is expected that daughters and daughters-in-law of older people who are 65 years or older will be in the age range 40-55. This is the age group that most frequently juggles multiple care responsibilities for children, parents or for a spouse or partner with age-related health problems. It is important for governments to consider how caring responsibilities can be combined with employment in this age group.

Table 3 Age distribution of informal carers in different OECD countries

Year	Year	44 and less	45-64	65 and over
Australia	1998	47	36	17
Austria	2002	27	48	25
Canada[1]	1995	35	42	23
Germany[2]	1998	15	53	33
Ireland	2002	46	43	11
Japan[3]	2001	4	42	54
South Korea	2001	30	39	31
UK[4]	2000	35	45	20
US[5]	1994	12	37	51

1. British Columbia only.
2. Germany: main carer only, age groups refer to –39, 40-64, and 65+.
3. Japan: age groups refer to –30, 40-59, and 60+.
4. United Kingdom: age groups refer to 16-44, 45-64, and 65+.
5. Primary active caregivers only.
Source: OECD 2005.

7. HOW GOVERNMENTS SUPPORT WOMEN PROVIDING INFORMAL LONG-TERM CARE

Several countries in the Middle East, such as Egypt, Bahrain, Lebanon, Tunisia, Morocco, and Kuwait, have tried to make conditions easier for working women with the introduction of paid maternity leave. However, as in some developed countries, support for women who provide long-term care does not exist. In Kuwait, for example, women have been entitled to up to two months at their full salary, and an extra four months at half salary if they could show that they were sick due to pregnancy. Other countries have also passed laws prohibiting gender discrimination in the workplace. For example in 2002, Lebanon's laws were changed to make it illegal for employers to discriminate based on gender in the nature of work, salary, or promotion. However, applications of such laws are very rare in practice (Rubin 2007).

A review of literature, on some examples from the developed world of a number of policies adopted to support those who provide care for older or disabled relatives, can be useful in terms of considering existing strategies. These are extracted from Hussein and colleagues (2009) and presented in Box 1.

Arab and Islamic governments face growing demands to establish formal long-term care provision as well as to facilitate and support those who provide informal care for the elderly and disabled. This goes hand in hand with an increasing need to maximize participation in the labor market, including that of the women who have been the traditional providers of informal care. This gives rise to expectations that governments should adopt sensible and cost-effective polices that aim to maximize the opportunities for women to balance their formal labor participation as well as to provide, in partnership with men, much-needed support and care to supplement that provided by the government.

Within different developed countries, there are several government approaches that address, and seek to improve, work-life balance in general, but rarely with a clear focus on providing care to older people. As mentioned earlier, the majority of policy focus relates to the care of young children, with very little attention to employees who care for older or disabled people. In terms of general work-life strategies, three major types of approaches emerge from the literature. First, some countries such as the United Kingdom, New Zealand and Australia have developed *campaigns to promote work-life balance* in the workplace by targeting employers. The second can be found in the Netherlands, Sweden and Denmark, where efforts to support workers to balance work and other responsibilities, include caring for older people, have been developed through *a broad range of measures*. The third group, such as France, Belgium, United States and Ireland, focus on *developing legislation and initiatives* supporting work-life balance (Hussein *et al.* 2009).

In the first group, a business case is usually presented to promote the value of work-life balance. Employers are encouraged to recognize that available coping mechanisms will increase workers' productivity. Initiatives are usually based on providing information on websites or newsletters on the importance of improving work-life balance. These also provide workers with links to organizations that provide information and support to carers, people with disabilities and others. However, compliance with these guidelines is voluntary with the exception, in the UK, that parents of children under six years or disabled children under 18 years have had the right to request flexible working arrangements (from April 2003). This is now being extended to other age groups. In New Zealand, a Work-Life Balance Project was established in 2002 to provide information aimed mainly at parents; similar resources are available in Australia to inform carers of relevant legislation.

Box 1 Examples of policies adopted in some developed countries to support informal care providers

The ***United Kingdom*** offers a cash benefit known as Carer's Allowance to provide support to carers. To be eligible, carers must have limited income or savings and be providing a minimum of 35 hours of care a week to a person who is themselves in receipt of a benefit awarded to pay for care and attention (Attendance Allowance or Disability Living Allowance). Until 2002, Carer's Allowance was available only to carers aged below 65, but eligibility was then extended to those over this age, subject to a means test.

In ***Germany***, informal care continues to play its traditional strong role, and this is reflected in the benefit system, allowing a recipient to draw a cash allowance that can be used to pay informal carers or careworkers. Non-profit organizations are the major providers of long-term care services at home. The introduction of long-term care insurance has resulted in the rapid growth of providers of home-care services, which by law have to be mainly private providers (either not-for-profit or for-profit).

In ***Ireland***, the Carer's Allowance is a payment for carers with low income who live with and look after people who need full-time care and attention. Carer's Benefit is a payment made to insured persons who leave the workforce temporarily to care for a person in need of full-time care and attention. In order to receive these benefits, the care recipient must be so disabled as to require full-time care and attention, but must not normally live in a hospital, care home or other institution.

In ***Norway***, the government plays the dominant role in long-term care, as the public sector provides most services and these are largely financed by direct taxation. However, provision of long-term care services is largely decentralized and integrated at the level of the municipality. In 1988, Norway introduced payments for informal care under the Municipal Health Services Act of 1986. People caring for older relatives or disabled children on a regular basis may receive a cash benefit from the municipality called 'caregiver pay'.

In ***Sweden***, there are primarily three types of support: respite and relief services, support and educational groups for carers and economic support for caring. Informal carers can be supported through a Home Care Allowance, respite care for the older person in day-care centers or short-term stays for the older person in care homes. A number of cash benefits are available for informal carers, and a carer can be directly employed by the municipality to care for an older person.

Luxembourg has a social insurance system covering old age and acute health care, and in 1998 introduced a new arm of social insurance to cover long-term care. Since 2001, the share of disabled older people who are cared for at home has been steadily increasing, from 53 percent of all long-term care beneficiaries in 2001 to 60.4 percent in 2004, and the size of home-care workforce has consequently increased by 21 percent from 1999 to 2002. There is no evidence of programs aimed at supporting informal carers, including women, in Luxembourg.

Poland provides tax relief on expenses involved in the care of a dependent relative. Polish workers can also take time off work with compensation, up to 14 days per year.

8. HOW DO WOMEN BALANCE THEIR DUAL ROLE AS PARTICIPANTS IN THE LABOR MARKET AND ALSO AS INFORMAL CARERS?

Some studies of employees who provide long-term care to parents or other relatives have shown that the demands of caring responsibilities and of employment often interfere with each other (Gignac et al 1996; Scharlach, 1994, Stephens *et al.* 2001). Bi-directional stress and interference are usually observed, where not only employment interferes with caring responsibilities but *vice versa* (Stephens and Townsend 1997). On the other hand, 'expansion theory' argues that multiple roles can enhance an individual's energy and health. This assumption is based on the principle that different roles provide individuals with different resources (Marks, 1977). In general, there is wide research suggesting that employed women, regardless of their caring responsibilities, tend to have better health indicators, such as less psychological stress, better physical health and higher self-esteem (Aneshensel *et al.* 1995; Bromberger and Matthews, 1994, Ross and Mirowsky 1995). Moreover, research on providing care in later life has shown the direct benefits of employment for informal carers (Brody et al 1987, Skaff and Pearlin 1992, Stephens et al 2001). Time spent at work has the potential to provide women with some form of respite or distraction from their usual caring responsibilities. Furthermore, employment increases women's income as well as their social and psychological resources. However, most of the literature focuses on informal carers who are employed in professional roles, and there is less known about the effect of multiple roles on women who are employed in less skilled occupations. To sum up, the effect of multiple roles on women's health and wellbeing is governed by two theoretical perspectives: 'role strain' and 'role enhancement' theories. Traditionally these have been regarded as opposing perspectives but more recent research is showing that they are more likely to be complementary theories (Stephens and Franks 1999; Martire and Stephens 2003).

Hussein and colleagues (2009) cite recent research suggesting that it is beneficial for employers to support their staff with their caring responsibilities (Yeandle et al 2006) and show that different support models are used by employers in different countries. Flexible and supportive work environments are essential in enabling people who provide informal care to participate in the labor market (Phillips et al 2002; Seddon et al 2004). It is also evident that competing demands may cause carers to cease work completely, in favor of their caring responsibilities (Arber and Ginn 1995), but the literature suggests that the majority of carers manage (or are forced to manage) a combination of the two activities (Glendinning 1992, Joshi 1995).

Data from the 2001 United Kingdom census shows that the majority of working carers in England and Wales are less likely to hold university degrees and are concentrated in lower-level jobs (Buckner and Yeandle 2006). It is not clear whether people with lower qualifications and jobs were more available to provide informal care, or whether their caring responsibilities caused them to lose or give up education and employment opportunities (Lundsgaard 2005; Himmelweit and Land 2008). In the UK, Arksey and Glendinning (2008) explored carers' decision-making process around work and care in interviews with 80 working-age carers. They again concluded that decision-making processes, relating to labor force participation and managing strategies, depend on different characteristics such as age, gender, ethnicity and levels of need among people being cared for. In many cases, carers were forced to cease employment due to pragmatic or practical reasons and against their wishes. Arksey and Glendinning (2008) observed the added challenges of different geographical locations: for example, it takes less time to travel to work in urban than rural locations, which adds another barrier to rural residents if they wish to participate in the labor market. Moreover, fewer job opportunities are generally available in rural than urban locations. One of their main conclusions was the importance of the availability of flexible working hours, which were critical in enabling carers to combine work and caregiving. Self-employment was perceived to be a suitable option, in theory; however, it was unrealistic if it required traveling or being available at certain times during the day.

In the United States, Pavalko and Henderson (2006) found that women who provide care are more likely to cease work if their employment conditions are not supportive. On the other hand, women employed in jobs that have access to flexible hours, paid and unpaid family leave, and paid sick leave are more likely to remain employed and maintain most of their working hours. Using data from the National Longitudinal Survey of Young Women, 1995 to 2001, Pavalko and Henderson found that on average, in a six year period, around 13 percent of employed women started providing unpaid care, as a proportion of those who were not providing any care work previously. Examining the effects of different workplace policies on the implications of care work for women, they revealed that, in general, women are at greater risk of leaving the labor workforce if they take on unpaid care work. However, they observed that women who had access to flexible working policies in their workplace were more likely to stay in employment. An interesting finding was that such associations were observed among all employed women who had access to such policies, regardless of their caring responsibilities. This study highlights the importance of supportive workplace policies for women in order to perpetuate their labor force participation. Also in the United States, research on employed women who provide care for older people found that women who received instrumental support, such as physical assistance, from their partners were better able to balance their roles (Franks and Stephens, 1996).

9. EXAMPLES OF COPING MECHANISMS AND HOW WOMEN BALANCE WORK AND LTC PROVISION

Most existing work-life balance or ‘family friendly’ policies and services, in more economically developed countries, are primarily designed for parents of young children, and rarely address the needs of employees who care for older or disabled people. The issue of work-life balance is receiving different degrees of attention in different countries. Some countries, such as the United Kingdom, New Zealand and Australia, support work-life balance as an explicit policy goal. Various campaigns focus on encouraging the voluntary agreement of employers to develop and implement policies and practices supporting a work-life balance in their organizations. Some countries, such as the Netherlands, Denmark and Sweden, focus on developing legislative and social policy goals to support workers who provide unpaid care including those providing care for older people. Netherlands’ ‘leave savings’ and Sweden’s sabbatical leave, allow workers more time to devote to caring responsibilities (Hussein *et al.* 2009).

The resulting situation, in many countries, is that women or other carers who are faced with caring responsibilities for older people either quit working or opt for part time work which may underuse their skills and education (Swanberg 2007, Marler and Moen 2005). Many women carers find one of the only options available is to work fewer hours; however, most highly paid workplaces lack flexible work arrangements (Todd 2004). In Australia, one-third of employees providing informal care indicated that due to their caring responsibilities they had to work fewer hours, take periods of unpaid leave or take jobs with less responsibilities, which required fewer skills than they possessed; 13 percent had refused promotions in fear of increased responsibilities (Morehead *et al.* 1997).

Research also shows that a large percentage of those caring for younger children are also caring for older relatives (Families and Work Institute 1998, Krouse and Afifi 2007). Women who have multiple caring and work roles usually have bi-directional stress spillover between work and family (Allen *et al.* 2000, Edwards and Rothbard 2000). Stress is particularly evident when women combine their informal caring responsibility with a carer of formal care provision due to high emotional involvement at work and home (Brody 1985, Pearlin *et al.* 1990, Hochschild 1983, Krouse and Afifi 2007). Moreover, societies usually devalue formal care provided by women due to the proximity of caring work with female roles as mothers or wives (England and Folbre 1999).

10. CONCLUSION

The existing evidence suggests a growing proportion of older people in the Arab and Islamic countries, who will regularly require a considerable care provision, due to ill health in later life. The norms, religion and traditions of these countries place the duty of care on women; such care, in the vast majority of cases, is unpaid and informally provided. A number of interacting factors, such as migration, changing family structure and labor force participation, heighten the pressure on women's time and energy through a number of competing demands.

How women balance their multiple roles depends on a number of interacting factors. Some operate at the individual level and others relate to the surrounding community and underlying support from government policies. The availability of formal care, whether on a full-time or flexible basis, needs to be considered by almost all countries in the Arab and Islamic region. A variety of support measures should be introduced, including flexible care arrangements between the formal and informal sector, such as respite care provision. Moreover, flexible work environments and the availability of supportive workplace policies are crucial to enhancing women's participation in the labor force. The relatively low prevalence of women who combine family, care and employment responsibilities in the developed world, such as United Kingdom and North America (Rosenthal 1997; Spitze and Logan 1990; Evandrou and Glaser 2004), indicates the extent of the barriers faced by women who are seeking to balance their multiple roles.

Suitable policies need to be tailored and adopted to the cultural context of the region. Policies enabling more flexible working hours and provision of paid leave are reported to be particularly important. The availability of informal support to women to 'care share' through partners, older children, friends or other members of the community is also fundamental in facilitating women's participation in the labor force.

The impact of informal care and labor force participation is becoming increasingly important, owing to several factors, including the reality of increased demand for both long-term care and labor participation. A large volume of empirical research indicates the difficulties associated with combining care and work responsibilities, particularly caring for older people. Unlike the care of children, which follows a fairly predictable time schedule, care for older or disabled people is unpredictable in duration and intensity; it may also increase in intensity over the course of the care experience. Moreover, there are very limited formal substitutes for informal care provided by the family, most often women. This may explain the continuance of women's multiple roles.

REFERENCES

Ahmed, N. I. and S. Abbas (1993). Family care of elderly problems. *Journal of the Egyptian Public Health Association* 68: 161-177.

Allen, T., D. Herst, C. Bruck and M. Sutton (2000). Consequences associated with work-to-family conflict: A review and agenda for future research. *Journal of Occupational Health Psychology* 5: 278-308.

Aneshensel, C., P. Leonard, J. Mullan, S. Zaeit and C. Whitlatch (1995). *Profiles in Caregiving: The Unexpected Carer*. San Diego, CA, Academic Press.

An-Na'im, A. (1995). The dichotomy between religious and secular discourse in Islamic societies. *Faith and freedom : women's human rights in the Muslim world*. M. Afkhami. London, Tauris**:** 51-60.

Anonymous (2005). Urfi Marriage in Egypt: The Issues. *Sexuality in Africa Magazine* 2 (2): 6-7.

Arber, S. and J. Ginn (1995). Paid employment and Informal Care. *Work, Employment and Society* 9 (3): 445-471.

Arksey, H. and C. Glendinning (2008). Combining work and care: carers' decision-making in the context of competing policy pressures. *Social Policy and Administration* 42 (1): 1-18.

Bener, A. and K. Alali (2006). Consanguineous marriage in a new developed country: the Qatari population. *Journal of Biosocial Science* 38 (2): 239-246.

Boggatz, T. and T. Dassen (2005). Ageing, care dependency, and care for older people in Egypt: A review of the literature. *International Journal of Older People Nursing* 14: 56-63.

Brodsky, J., J. Habib and M. Hirschfeld (2003). *Key policy issues in long-term care*. Geneva, World Health Organization.

Brody, E., M. Kleban, P. Johnsen, C. Hoffman and C. Schoonover (1987). Work Status and Parent Care: A Comparison of Four Groups of Women. *The Gerontologist* 27 (2): 201-208.

Bromberger, J. and K. Matthews (1994). Employment status and depressive symptoms in middle-aged women: A longitudinal investigation. *American Journal of Public Health* 84 (2): 202-206.

Buckner, L. and S. Yeandle (2006). *Statistical Analysis - Working Carers: Evidence from the 2001 Census*. London, Carers UK.

Cain, M. (1985). On the relationship between landholding and fertility. *Population Studies* 39: 5-15.

Coughlin, T., T. McBride, M. Perozek and K. Liu (1990). *Home care for the disabled elderly: Predictors and expected costs*. Washington, DC, Urban Institute.

Edwards, J. and N. Rothbard (2000). Mechanisms linking work and family: Clarifying the relationship betwen work and family constructs. *The Academy of Management Review* 25: 178-99.

El Haddad, Y. (2006). Major Trends Affecting Families in the Gulf Countries. *Major Trends Affecting Families: A Background Document*. United Nations. Geneva, United Nations Department of Economic and Social Affairs.

El-Saadani, S. M. (2006). *Divorce in the Arab Region: Current Levels, Trends and Features*. The European Population Conference 2006, Liverpool. 21-24 June.

England, P. and N. Folbre (1999). The cost of caring. *The Annals of the American Academy* 561: 39-51.

Evandrou, M. and K. Glaser (2004). Family, work and quality of life: changing economic and social roles through the lifecourse. *Ageing and Society* 24: 771-791.

Families and Work Institute (1998). *The 1997 national study of the changing workforce: executive summary*. New York, NY, Families and Work Institute. http://familiesandwork.org/site/research/summary/1997nscwsumm.pdf (retrieved 30 October 2009).

Farques, P. (2006). *International Migration in the Arab Region*. UN Expert Group Meeting on International Migration and Development in the Arab Region: Challenges and Opportunities, Beirut. 10 May.

Franks, M. and M. Stephens (1996). Social support in the context of caregiving: Husbands' provision of support to wives involved in parent care. *Journal of Gerontology: Psychological Sciences* 51 (B): 43-52.

Gignac, M., E. Kelloway and B. Gottlieb (1996). The impact of caregiving on employment: A mediational model of work-family conflict. *Canadian Journal on Aging* 15: 525-542.

Glendinning, C. (1992). Employment and community care: policies for the 1990s. *Work, Employment and Society* 6 (1): 103-111.

Himmelweit, S. and H. Land (2008). *Reducing Gender Inequalities to Create a Sustainable Care System*. York, Joseph Rowntree Foundation.

Hochschild, A. R. (1983). *The managed heart : commercialization of human feeling*. Berkeley, University of California Press.

Hoodfar, H. and I. NetLibrary (1997). *Between marriage and the market : intimate politics and survival in Cairo*. Berkeley, University of California Press.

Hoveyda, F. (2005). Arab women and the future of the Middle East. *American Foreign Policy Interests* 27 (5): 419-438.

Hussein, S. (2002). An exploration of nuptiality patterns in Egypt and the Maghreb countries. London, London School of Economics, University of London. PhD Theses.

Hussein, S. and J. Manthorpe (2005). An International Review of Long Term Care Workforce: Policies and Shortages. *Journal of Aging and Social Policy* 17 (4): 75-94.

Hussein, S. and J. Manthorpe (2007). Women from the Middle East and North Africa in Europe: understanding their marriage and family dynamics. *European Journal of Social Work* 10(4) 465-480.

Hussein, S., J. Manthorpe and A. Bakalana (2009). *The Competing Demands for Women's Labor: the Role of Women in Long-Term Care Provision in the Russian Federation*. Working

Paper prepared for the Russian Demographic Policy Note Team. Washington, DC, World Bank.

Joseph, S. (1994). Brother/Sister Relationships - Connectivity, Love and Power in the Reproduction of Patriarchy in Lebanon. *American Ethnologist* 21 (1): 50-73.

Joshi, H. (1995). The labour market and unpaid caring: conflict and compromise. *The future of family care for older people*. I. Allen and E. Perkins. London, HMSO**:** ix, 234p.

Kinsella, K. and D. R. Phillips (2005). Global aging: The challenge of success. *Population Bulletin* 60 (1): 3-40.

Krouse, S. and T. Afifi (2007). Family-to-work spillover stress: Coping communicatively in the workplace. *Journal of Family Communication* 7 (2): 85-122.

Lakdawalla, D., T. J. Philipson and National Bureau of Economic Research (1999). *Aging and the growth of long-term care*. Cambridge, MA, National Bureau of Economic Research.

Lundsgaard, J. (2005). *Consumer Direction and Choice in Long-term Care for Older Persons, Including Payments for Informal Care*. Health Working Paper No. 20. Paris, OCED.

Margolis, S. A. and R. L. Reed (2001). Institutionalizing older adults in a health district in the United Arab Emirates: Health status and utilization rate. *Gerontology* 47: 161-167.

Marks, S. R. (1977). Multiple Roles and Role Strain - Some notes on human energy, time and commitment. *American Sociological Review* 42 (6): 921-936.

Marler, J. H. and P. Moen (2005). Alternative employment arrangements: A gender perspective. *Sex Roles* 52 (5-6): 337-349.

Martire, L. M. and M. A. P. Stephens (2003). Juggling parent care and employment responsibilities: The dilemmas of adult daughter caregivers in the workforce. *Sex Roles* 48 (3-4): 167-173.

McNicoll, G. (1980). Institutional determinants of fertility change. *Population and Development Review* 6 (3): 441-462.

Miller, E. A. and W. G. Weissert (2000). Predicting elderly people's risk for nursing home placement, hospitalization, functional impairment, and mortality: A synthesis. *Medical Care Research and Review* 57 (3): 259-297.

Morehead, A., M. Steele, M. Alexander, K. Stephen and L. Duffin (1997). *Changes at work : the 1995 Australian workplace industrial relations survey*. South Melbourne, Longman.

Organization for Economic Cooperation and Development (OECD) (2005). *Long-term Care for Older People*. The OECD Health Project. Paris, OECD.

Pavalko, E. K. and K. A. Henderson (2006). Combining care work and paid work - Do workplace policies make a difference? *Research on Aging* 28 (3): 359-374.

Pearlin, L. I., J. T. Mullan, S. J. Semple and M. M. Skaff (1990). Caregiving and the Stress Process: an Overview of Concepts and their Measures. *Gerontologist* 30 (5): 583-594.

Phillips, J., M. Bernard, M. Chittenden and F. Joseph Rowntree (2002). *Juggling work and care : the experiences of working carers of older adults*. Bristol, Published for the Joseph Rowntree Foundation by The Policy Press.

Rashad, H., M. Osman and F. Foudi-Fahimi (2005). *Marriage in the Arab World*. Washington, DC, Population Reference Bureau.

Rosenthal, C. (1997). The changing contexts of family care in Canada. *Ageing International* 24 (1): 13-31.

Ross, C. E. and J. Mirowsky (1995). Does Employment Affect Health? *Journal of Health and Social Behavior* 36 (3): 230-243.

Roudi-Fahimi, F. and M. M. Kent (2007). Challenges and opportunities - The population of the Middle East and North Africa. *Population Bulletin* 62 (2): 3-19.

Rubin, J. C. (2007). Lagging Far Behind: Women in the Middle East. *Middle East Review of International Affairs* 11, http://meria.idc.ac.il/journal/2007/issue2/jv11no2a2.html (retrieved 29 October 2009).

Rugh, A. B. (1981). Orphanages and homes for the aged in Egypt: Contradiction or affirmation in a family oriented society? *International Journal of Sociology of the Family* 11: 203-233.

Rugh, A. B. (1984). *Family in contemporary Egypt*. Syracuse, NY, Syracuse University Press.

Rugh, A. B. (1997). *Within the circle : parents and children in an Arab village*. New York, Columbia University Press.

Ryder, N. (1983). Fertility and Family Structure. *Population Bulletin of the United Nations* 15: 15-34.

Scharlach, A. E. (1994). Caregiving and Employment: Competing or Complementary Roles? *Gerontologist* 34 (3): 378-385.

Seddon, D., C. Robinson and K. Jones (2004). *Carers in Employment: Towards an Integrated System of Support*. Final Report to the Wales Office of Research and Development for Health and Social Care. Bangor, Centre for Social Policy Research and Development, University of Wales.

Shahine, G. (2005). The Double Bind. *Al Ahram Newspaper Weekly*, 1-7 October 2005. http://weekly.ahram.org.eg/1998/397/fe1.htm (retrieved 30 October 2009).

Shakoori, B. (2007). ESCWA Population Activities 2006-2007. *Statement to the Commission on Population and Development* http://www.escwa.un.org/popin/docs/ESCWA_CPD41.pdf (retrieved 28 October 2009).

Sibai, A., K. Sen, M. Baydoun and P. Saxena (2004). Population aging in Lebanon: Current status, future prospects and implications for policy. *Bulletin of the World Health Organisation* 82: 219-225.

Sinunu, M., K. M. Yount and N. A. W. El Afify (2008). Informal and formal long-term care for frail older adults in Cairo, Egypt: family caregiving decisions in a context of social change. *Journal of Cross Cultural Gerontology* 24 (1): 63-76.

Skaff, M. M. and L. I. Pearlin (1992). Caregiving - Role Engulfment and the Loss of Self. *Gerontologist* 32 (5): 656-664.

Spitze, G. and J. Logan (1990). More Evidence on Women (and Men) in the Middle. *Research on Aging* 12 (2): 182-198.

Stephens, M. A. P. and A. L. Townsend (1997). Stress of parent care: Positive and negative effects of women's other roles. *Psychology and Aging* 12 (2): 376-386.

Stephens, M. A. P. and M. M. Franks (1999). Parent care in the context of women's multiple roles. *Current Directions in Psychological Science* 8 (5): 149-152.

Stephens, M. A. P., A. L. Townsend, L. M. Martire and J. A. Druley (2001). Balancing parent care with other roles: Interrole conflict of adult daughter caregivers. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences* 56 (1): P24-P34.

Swanberg, J. E. (2007). Making it work: Informal caregiving, cancer, and employment (vol 24, pg 1, 2006). *Journal of Psychosocial Oncology* 25 (1): 138-138.

Syed, J. (2008). A context-specific perspective of equal employment opportunity in Islamic societies. *Asia Pacific Journal of Management* 25 (1): 135-151.

Tajvar, M., M. Arab and A. Montazeri (2008). Determinants of health-related quality of life in elderly in Tehran, Iran. *BMC Public Health* 8, doi:10.1186/1471-2458-8-323 (retrieved 29 October 2009).

Todd, S. (2004). *Improving work-life balance - what are other countries doing?* Ottowa, Labour Program, Human Resources and Skills Development Canada.

United Nations Development Programme Regional Bureau for Arab States and Arab Fund for Economic and Social Development (2002). *The Arab Human Development Report : creating opportunities for future generations*. New York, NY, United Nations.

United Nations Development Programme Regional Bureau for Arab States and Arab Fund for Economic and Social Development (2006). *The Arab human development report 2005 [electronic resource] : towards the rise of women in the Arab world*. New York, NY, United Nations Development Programme.

US Committee for Refugees and Immigrants (2006). World Refugee Regions 2006. http://www.refugees.org (retrieved 12 October 2009).

WLUML (Women Living Under Muslim Laws) (1986). *Plan of Action*. Aramon.

Yeandle, S., C. Bennett, L. Buckner, L. Shipton and A. Suokas (2006). *Who Cares Wins: The Social and Business Benefits of Supporting Working Carers*. Sheffield, Centre for Social Inclusion, Sheffield Hallam University.

Yoo, B. K., J. Bhattacharya, K. M. McDonald and A. M. Garber (2004). Impacts of informal caregiver availability on long-term care expenditures in OECD countries. *Health Services Research* 39 (6): 1971-1995.

Yount, K. M. and H. Rashad (2008). *Family in the Middle East : ideational change in Egypt, Iran and Tunisia*. London, Routledge.

Yount, K. M. and A. M. Sibai (2009). Demography of Aging in Arab Countries. *International Handbook of Population Aging [electronic resource]*. D. L. Poston and P. Uhlenberg. Dordrecht, Springer Netherlands**:** 277-315.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 45–69.

THE EFFECT OF SELF-RATED-HEALTH ON THE QUALITY OF LIFE OF OLDER ADULTS ACROSS THE WORLD – EVIDENCE FROM A GLOBAL AGEING SURVEY "THE FUTURE OF RETIREMENT"

Hafiz T.A. Khan*
Senior Lecturer in Applied Statistics
Middlesex University Business School
Middlesex University
London NW4 4BT, UK
E-mail: h.khan@mdx.ac.uk

George W. Leeson
Deputy Director
Oxford Institute of Ageing
University of Oxford
Oxford OX2 6PR, UK
E-mail: george.leeson@ageing.ox.ac.uk

ABSTRACT

In social science and public health earlier research has persistently reported significant socio-economic inequalities in health, inequalities in the use of health care, and self-rated-health (SRH) among older adults. However, relatively little attention is paid to the link between SRH and the overall quality of life (QoL) of older adults. Using the data collected in the Future of Retirement Survey (FoR) the study explores the linkages between the self-rated-health and quality of life among older adults in 21 countries and territories in five major regions of the world. The QoL was assessed by two survey instruments designed to capture subjective as well as objective appraisals of individual quality of life. Both bivariate and multiple analyses were performed to examine the impact of SRH on the QoL. The analyses reveal that there are health inequalities across different age cohorts and this is consistent for all selected countries and territories. As expected the proportion reporting poor health increases with age in most countries. The net effect of health on QoL has also been analysed using ordered logistic regression analysis adjusted for age and gender. Age plays an important role alongside with health on the overall quality of life. The study also reveals that women are found to be more likely to have been depressed compared to their male counterparts.

Keywords: Population ageing, self reported health, quality of life.

* ***Also an Associate Research Fellow, Oxford Institute of Ageing, University of Oxford, Oxford OX2 6PR, UK***
Email: hafiz.khan@ageing.ox.ac.uk

1. INTRODUCTION

Health has always been a concern, particularly among older persons and has become a subject of interest among researchers in many areas such as public health and social science disciplines [1-8]. Health is one of the most crucial factors of human life and living. A good health is a legitimate expectation of every human being and is globally universal no matter what age peoples are. The better the individuals' health status the more they can develop themselves, and most importantly contribute to every sphere of life. It is well believed that a healthy society can develop optimally and only a healthy nation is able to enjoy the benefits of development[9].

Earlier research has persistently reported significant socio-economic inequalities in health, inequalities in the use of health care, and self-rated-health (SRH) among older adults. However, relatively little attention is paid to the link between SRH and the overall quality of life (QoL) of older adults. Moreover, evidence from cross-country analysis is rarely seen in the existing literature and yet such analyses are needed to progress our understanding of some of the many trends in global ageing. Public policy and services in the UK and elsewhere have increasingly recognised the importance of quality of life for their own people.

The principle aim of this study has been to explore the effect of self-reported-health on the quality of life of older adults cross-nationally. To address such an aim, we set up three specific objectives:

- Firstly, to examine health inequalities across the selected countries and territories;
- Secondly, to identify differential effects of health across various components of QoL; and
- Finally, to examine the net effect of SRH on the overall QoL as constructed in this study.

2. DATA METHODOLOGY

2.1 The Data

This study uses data collected in the HSBC Global Ageing Survey (GLAS) – the Future of Retirement Survey (FoR) in which information was collected from 21,233 individuals aged 40-79 years across 21 countries and territories in five major geographical regions of the world. The study population comprised of 9,843 men and 11,390 women in four age cohorts: two pre-retirement aged 40-49 and 50-59 years, and two post-retirement aged 60–69 and 70–79 years. The FoR is the largest global ageing survey of its kind that investigates attitudes towards later life, ageing and retirement. Each individual was asked a battery of structured questions regarding their socio-economic status, health conditions, social networks, saving and investments, and preparedness for retirement. The survey covers Denmark, France, Germany, Poland, Sweden, UK, Canada, USA, Brazil, Mexico, Russia, Turkey, Saudi Arabia, South Africa, China, Hong Kong, Taiwan, India, Indonesia, Japan, Malaysia, Philippines, Singapore, and South Korea, which can further be broadly classified into two mutually exclusive groups of mature and transitional economies respectively. Mature economies are those that industrialized early, have large service sectors, affluent populations, long-established pensions infrastructure and legislation and provide a comprehensive welfare 'safety net' for their citizens. On the other hand, transitional economies do not yet meet the definition of a mature economy. In the transitional economies, the survey interviewed so-called 'trendsetters' - people who live mainly in urban settings, and who work in the service sector or other modern areas of the economy. These trendsetters will arguably pick up on the behaviours and attitudes of mature economy populations at an earlier stage than rural populations in the transitional economies. The interviews were conducted by telephone or where this was impractical by face-to-face. Individuals were selected at random and cohort samples are representative of the cohort (with due note of the trendsetter phenomenon). The Oxford Institute of Ageing at the University of Oxford is responsible for the research design and tools. Fieldwork and data-entry were carried out by Harris Interactive. Details of survey methodology and research reports can be obtained on the website http://www.hsbc.com/hsbc/retirement_future/research-summary[10-11].

2.2 Variables Used For Statistical Analysis

The description of the selected variables is discussed in this section. The self-reported health (SRH) of respondent is used as an independent variable in this study. During the interview, respondents were asked about their perception of health status and a question was "*How is your health in general*?" In order to record and compare the health situation of older adults across selected countries, responses are categorised in a five point ordered categorical scale as very good, good, fair, poor, and very poor. The higher the numerical value, the poorer is the respondent's overall health situation. This variable is recorded further into two groups where the SRH was used as an indicator of health which is categorized as "good" (corresponding to good and very good responses) and "poor" (corresponding to fair, poor and very poor responses).

The QoL was assessed by two survey instruments designed to capture subjective and objective appraisals of individual QoL. The feeling of an individual is measured by asking a subjective question with seven dimensions: *How often do you think each statement applies to you?* (Responses were from an ordered range i.e., often, sometimes, rarely and never). Following are various categories of outcome variables:

- i) My age prevents me from doing the things I would like to do (variable1, say V1_1),
- ii) I feel that what happens to me is out of my control (similarly, V1_2),
- iii) I am able to do the things I want to do (V1_3),
- iv) Lack of money prevents me from doing the things I want to do (V1_4),
- v) Family responsibilities prevent me from doing what I want to do (V1_5),
- vi) I feel life is full of responsibilities (V1_6), and
- vii) I feel the future looks good for me (v1_7).

On the other hand, an objective question captures the difficulties they face in real life situations: "*How often have you felt the following over the last week*?" (Similarly, responses were chosen from an ordered range i.e., almost all of the time, most of the time, some of the time, almost none of the time and never). The following are various categories of outcome variables:

- i) I felt depressed (variable2, say V2_1),
- ii) I was happy (similarly, V2_2),
- iii) I felt lonely (V2_3),
- iv) I felt sad (V2_4),
- v) I felt everything was too much effort (V2_5),
- vi) I enjoyed life (V2_6), and
- vii) I was looking forward to the future (V2_7).

2.3 Statistical Tools Used For Data Analysis

In the study, a bivariate cross tabulation analysis was carried out to examine any significant difference between two groups of individuals who reported good and poor health. Then a multiple ordered logistic regression analyses were performed to examine the impact of SRH on QoL across various countries.

3. RESULTS

3.1 Cross Tabulation Analysis

The prevalence of self-reported-health by age and gender is displayed in Table 1. Our analyses show that there are health inequalities across different age cohorts and this is consistent for all selected counties and territories. The proportion reporting poor health increases with age and oppositely good health decreases with age. This result indicates a similar pattern across the world where reporting good heath is inversely related with age and this is now a globe al phenomenon. Moreover, whilst comparing between cohorts we see clear distinctions of prevalence (response) rates. For example, in case of youngest age cohort 40-49, more people reported good health compared with poor, whereas in the oldest cohort 70-79, it is completely opposite. The Chi-squared analysis was applied to isolate the significant difference between groups of individuals and Table 1 shows that there is a significant difference in reporting good and poor health status across selected age cohorts.

Gender is obviously an important variable in human science research. Significant differences in health status are observed to be for gender in some countries such as in Brazil, Hong Kong, Russia, Saudi Arabia, and in Turkey (statistical significance is considered at 1% level). In these countries a higher proportion of females are reported poor health compared to their male counterparts. It indicates that health inequality mainly persists in the developing countries. Evidence from this analysis also reveals that globally no universal health pattern follows for gender and for better understanding this requires further investigation.

3.2 Health and Subjective Measure of QoL

A cross tabulation between health and subjective QoL for various countries is presented in Table 2-8. With few exceptions, the analysis reveals that globally an individual's feeling towards QoL varies significantly as a result of health inequality. As can be seen from Table 2, the subjective QoL is measured by an instrument "My age prevents me from doing the things I would like to do" and the outcome measures are often, sometimes, rarely and never. The response score varies across different segments of feeling due to biological age. Most people in developed countries who responded a good health feel that their age never prevents them from doing the things that they want to do. It has been observed that the biological age rarely prevents enjoying the life fully because of possessing good health across the globe. It is also found that respondents who possessing poor health a higher proportion in developed countries reported that poor health is never a factor, on the contrary, poor health is sometimes prevents enjoy life in developing countries.

Table 3 shows that distribution of responds reported subjective QoL by health status. Irrespective of all geographical regions a vast majority reported that because of their good health they never feel it is not yet out of control. Good health provides the opportunity to do things often what they want across selected countries (Table 4) and there exists significant variations among various levels QoL. Similarly, there is no consistent pattern in answering the QoL for “lack of money prevents from doing the things they want” although there exists significant variation between respondents possessing poor and good health (Table 5). To understand a clear relationship we need to control for other variables such as age and income.

It has been observed from Table 6 that a higher proportion of respondents reported that family responsibilities never prevent them from doing what they like to do irrespective of their health situation and the response rates are higher for developed countries. Respondents also reported that as they feel life is often full of responsibilities and significant variations are also seen for all categories of QoL (Table 7). It has been found that respondents are positive about the future as they often feel the future looks good for them and there is statistically significant across various response of QoL (Table 8).

3.3 Health and Objective Measures of QoL

In contrast, with regard to questions about how an individual felt over the last week, there is significant variation in relation to *I felt depressed* (Table 9). For example, those with good health are less likely to have been depressed – in mature economies 84.3% in the UK, 88.9% in Denmark, and 76.6% in France have almost none of the time or never felt depressed. On the other hand, in transitional economies, these proportions are significantly lower: about 41.0% in India, 35.9% in Philippines, and 33.7% in Turkey. Individuals who possess good health are likely to have been reported to be happy most of the time (Table 10). Loneliness differs significantly with SRH in all countries and territories except for China, Philippines, and Russia where SRH has no significant impact at all (Table 11). When considering *I felt sad* as a QoL indicator there is significant variation as a result of SRH except in China and Japan (Table 12). Similarly, Tables 13-15 shows variations across various categories of QoL among selected individuals in the study. Thus, it can be concluded that health is a major factor for enjoying life and the study confirms this for all areas under investigation.

3.4 Ordered Logistic Regression

The net effect of health on QoL has also been analysed using ordered logistic regression analysis adjusted for age and gender. We considered three independent variables in the model in order to examine their effects on the subjective QoL (Table16). It has been found that age and health are the two predominant factors affecting our QoL. On the other hand, gender is sometimes appeared to be statistically significant in some models. While considering all three variables for predicting Objective QoL we see that all three variables are appeared to significant. What it means that age, gender and health are interlinked and they play important role in our QoL, particularly among the selected respondents aged 40-79 in our study.

Table 1 Percent distribution of self-reported health among adult respondents by age and sex

Country/ Territory	Health	Age				χ^2 test	Gender		χ^2 test	Total
		40-49	50-59	60-69	70-79	*p-value*	Male	Female	*p-value*	cases
Brazil	*Good*	29.1	25.0	23.8	22.0	*0.000*	50.4	49.6	*0.004*	635
	Poor	17.8	24.9	27.0	30.3		41.0	59.0		366
Canada	*Good*	27.8	25.7	25.2	21.3	*0.000*	42.0	58.0	*0.540*	902
	Poor	14.5	23.5	26.5	35.5		44.6	55.4		166
China	*Good*	34.3	28.1	20.6	17.0	*0.000*	52.2	47.8	*0.236*	335
	Poor	20.3	23.5	27.2	29.0		48.3	51.7		665
France	*Good*	29.4	26.8	24.6	19.2	*0.000*	40.7	59.3	*0.114*	765
	Poor	10.6	19.1	26.4	43.8		34.9	65.1		235
Germany	*Good*	29.5	26.3	25.2	19.0	*0.000*	52.2	47.8	*0.384*	691
	Poor	15.5	22.4	26.1	36.1		55.2	44.8		330
Hong Kong	*Good*	30.6	25.3	23.7	20.4	*0.000*	49.7	50.3	*0.006*	553
	Poor	18.5	24.5	26.5	30.5		41.0	59.0		449
India	*Good*	32.3	24.2	23.7	19.7	*0.000*	75.2	24.8	*0.466*	532
	Poor	16.6	26.4	26.8	30.1		73.2	26.8		481
Japan	*Good*	26.6	25.3	26.0	22.1	*0.022*	48.7	51.3	*0.255*	665
	Poor	21.8	24.5	23.0	30.7		52.5	47.5		335
Malaysia	*Good*	35.3	31.8	22.4	10.4	*0.000*	50.8	49.2	*0.233*	606
	Poor	9.3	14.6	28.8	47.2		47.0	53.0		396
Mexico	*Good*	31.7	25.4	22.2	20.7	*0.000*	53.2	46.8	*0.015*	571
	Poor	16.0	24.8	28.5	30.6		45.5	54.5		431
Philippines	*Good*	28.1	29.2	23.7	18.9	*0.000*	49.7	50.3	*0.923*	636
	Poor	19.0	19.8	26.5	34.8		50.0	50.0		374
Russia	*Good*	41.0	32.6	15.7	10.7	*0.000*	44.9	55.1	*0.000*	178
	Poor	21.8	23.7	27.5	27.0		19.4	80.6		854
Saudi Arab.	*Good*	34.1	30.7	23.9	11.3	*0.000*	53.3	46.7	*0.004*	683
	Poor	5.3	14.0	27.0	53.7		43.5	56.5		322
Singapore	*Good*	31.9	27.1	22.9	18.1	*0.000*	47.3	52.7	*0.373*	698
	Poor	9.8	19.9	29.7	40.5		50.3	49.7		306
South Afr.	*Good*	35.6	28.3	20.9	15.2	*0.000*	45.8	54.2	*0.622*	554
	Poor	13.4	20.4	29.8	36.4		44.3	55.7		456
South Korea	*Good*	32.0	30.3	21.5	16.2	*0.000*	51.4	48.6	*0.043*	488
	Poor	18.3	19.9	28.5	33.3		45.0	55.0		513
Denmark	*Good*	27.3	25.1	25.6	22.0	*0.000*	43.0	57.0	*0.018*	758
	Poor	16.9	23.1	26.2	33.8		34.6	65.4		260
Taiwan	*Good*	33.5	28.0	21.8	16.7	*0.000*	45.8	54.2	*0.024*	528
	Poor	15.5	21.6	28.6	34.3		38.8	61.2		472
Turkey	*Good*	33.1	25.2	20.5	21.1	*0.000*	47.8	52.2	*0.000*	341
	Poor	21.3	25.4	26.9	26.3		32.0	68.0		676
UK	*Good*	26.8	26.6	24.0	22.6	*0.000*	45.8	54.2	*0.169*	837
	Poor	15.5	18.2	28.2	38.1		51.4	48.6		181
USA	*Good*	27.8	25.1	24.7	22.4	*0.000*	40.3	59.7	*0.521*	813
	Poor	12.2	26.0	26.0	35.7		42.9	57.1		196

Table 2 Distribution of respondents reporting V1_1 by health status

Country/ Territory	Health	My age prevents me from doing the things I would like to do				χ^2 test	Total
		Often	Sometimes	Rarely	Never	*p-value*	cases
Brazil	*Good*	4.4	26.1	25.0	44.4	*0.000*	635
	Poor	23.5	34.2	21.3	21.0		366
Canada	*Good*	4.7	19.5	28.5	47.3	*0.000*	902
	Poor	19.3	30.7	24.7	25.3		166
China	*Good*	11.6	21.8	13.4	53.1	*0.000*	335
	Poor	16.5	29.2	15.5	38.8		665
France	*Good*	5.2	21.0	17.1	56.6	*0.000*	765
	Poor	23.8	36.2	8.9	31.1		235
Germany	*Good*	3.8	11.1	19.4	65.7	*0.000*	691
	Poor	15.5	22.4	20.9	41.2		330
Hong Kong	*Good*	7.8	26.9	28.0	37.3	*0.000*	553
	Poor	17.8	41.9	14.3	26.1		449
India	*Good*	20.5	50.0	16.5	13.0	*0.014*	532
	Poor	18.1	58.6	15.4	7.9		481
Japan	*Good*	2.9	31.6	46.0	19.5	*0.000*	665
	Poor	7.2	36.4	46.0	10.4		335
Malaysia	*Good*	4.3	23.6	32.8	39.3	*0.000*	606
	Poor	24.2	41.2	20.2	14.4		396
Mexico	*Good*	6.7	23.5	29.2	40.6	*0.000*	571
	Poor	19.0	33.9	17.9	29.2		431
Philippines	*Good*	7.5	28.9	14.0	49.5	*0.000*	636
	Poor	16.0	36.9	13.4	33.7		374
Russia	*Good*	10.7	8.4	30.9	50.0	*0.000*	178
	Poor	23.2	20.6	27.4	28.8		854
Saudi Arab.	*Good*	11.7	36.9	36.3	15.1	*0.000*	683
	Poor	35.7	42.5	18.0	3.7		322
Singapore	*Good*	13.9	31.8	19.5	34.8	*0.000*	698
	Poor	22.9	43.8	19.6	13.7		306
South Afr.	*Good*	9.4	27.6	21.7	41.3	*0.000*	554
	Poor	34.0	37.5	12.3	16.2		456
South Korea	*Good*	9.0	25.0	15.4	50.6	*0.000*	488
	Poor	23.0	25.3	13.8	37.8		513
Denmark	*Good*	3.2	10.3	21.6	64.9	*0.000*	758
	Poor	15.4	23.5	20.8	40.4		260
Taiwan	*Good*	11.2	27.1	23.9	37.9	*0.007*	528
	Poor	17.4	30.7	20.8	31.1		472
Turkey	*Good*	15.5	15.2	20.2	49.0	*0.000*	341
	Poor	34.0	16.0	18.9	31.1		676
UK	*Good*	5.6	18.3	17.9	58.2	*0.000*	837
	Poor	21.5	30.9	12.2	35.4		181
USA	*Good*	3.7	18.7	28.5	49.1	*0.000*	813
	Poor	21.4	30.6	19.9	28.1		196

Table 3 Distribution of respondents reporting V1_2 by health status

Country/ Territory	Health	I feel that what happens to me is out of my control				χ^2 test	Total
		Often	Sometimes	Rarely	Never	*p-value*	cases
Brazil	*Good*	3.1	29.3	34.0	33.5	*0.000*	635
	Poor	10.9	41.0	27.0	21.0		366
Canada	*Good*	9.9	20.1	32.6	37.5	*0.000*	902
	Poor	28.3	25.3	22.3	24.1		166
China	*Good*	5.7	10.1	14.0	70.1	*0.024*	335
	Poor	8.7	14.6	15.8	60.9		665
France	*Good*	11.5	39.0	21.2	28.4	*0.000*	765
	Poor	23.8	40.0	12.8	23.4		235
Germany	*Good*	4.8	10.9	22.6	61.8	*0.000*	691
	Poor	11.5	16.4	20.6	51.5		330
Hong Kong	*Good*	8.7	31.3	30.6	29.5	*0.000*	553
	Poor	13.6	42.5	21.2	22.7		449
India	*Good*	15.0	45.5	20.5	19.0	*0.003*	532
	Poor	14.6	41.6	29.9	13.9		481
Japan	*Good*	14.6	20.9	47.7	16.8	*0.001*	665
	Poor	6.9	26.9	45.7	20.6		335
Malaysia	*Good*	4.1	24.6	29.5	41.7	*0.000*	606
	Poor	16.4	38.1	25.8	19.7		396
Mexico	*Good*	5.1	32.7	27.1	35.0	*0.000*	571
	Poor	11.4	35.5	28.1	25.1		431
Philippines	*Good*	7.7	28.8	25.6	37.9	*0.002*	636
	Poor	5.6	40.4	21.9	32.1		374
Russia	*Good*	7.9	4.5	19.1	68.5	*0.000*	178
	Poor	8.8	11.6	28.8	50.8		854
Saudi Arab.	*Good*	7.3	38.2	43.3	11.1	*0.000*	683
	Poor	21.1	41.9	32.0	5.0		322
Singapore	*Good*	9.7	39.7	21.6	28.9	*0.000*	698
	Poor	15.7	42.8	25.2	16.3		306
South Afr.	*Good*	6.5	28.2	29.4	35.9	*0.000*	554
	Poor	22.8	42.1	20.6	14.5		456
South Korea	*Good*	4.5	12.7	14.3	68.4	*0.004*	488
	Poor	8.8	13.5	18.7	59.1		513
Denmark	*Good*	5.4	6.5	21.6	66.5	*0.000*	758
	Poor	13.1	16.9	24.2	45.8		260
Taiwan	*Good*	8.0	22.3	35.8	33.9	*0.001*	528
	Poor	10.2	31.6	27.3	30.9		472
Turkey	*Good*	15.8	17.0	26.7	40.5	*0.302*	341
	Poor	20.6	16.1	23.7	39.6		676
UK	*Good*	7.3	26.2	20.2	46.4	*0.000*	837
	Poor	18.8	33.1	18.8	29.3		181
USA	*Good*	9.1	23.1	32.3	35.4	*0.000*	813
	Poor	23.0	33.2	21.9	21.9		196

Table 4 Distribution of respondents reporting V1_3 by health status

Country/ Territory	Health	I am able to do the things I want to do				χ^2 test *p-value*	Total cases
		Often	Sometimes	Rarely	Never		
Brazil	*Good*	41.9	48.7	8.8	0.6	*0.000*	635
	Poor	23.0	52.5	20.5	4.1		366
Canada	*Good*	74.4	19.4	4.1	2.1	*0.000*	902
	Poor	53.0	31.9	10.8	4.2		166
China	*Good*	52.8	27.8	10.7	8.7	*0.072*	335
	Poor	44.8	29.5	14.1	11.6		665
France	*Good*	66.3	26.1	4.7	2.9	*0.000*	765
	Poor	47.2	36.2	10.2	6.4		235
Germany	*Good*	69.5	21.0	7.1	2.5	*0.001*	691
	Poor	57.9	24.8	12.7	4.5		330
Hong Kong	*Good*	41.8	38.7	13.2	6.3	*0.000*	553
	Poor	29.0	37.2	23.2	10.7		449
India	*Good*	25.0	43.4	23.5	8.1	*0.000*	532
	Poor	13.1	52.6	22.7	11.6		481
Japan	*Good*	27.2	46.6	21.8	4.4	*0.000*	665
	Poor	13.1	46.3	37.3	3.3		335
Malaysia	*Good*	64.5	27.1	7.3	1.2	*0.000*	606
	Poor	35.4	37.6	25.5	1.5		396
Mexico	*Good*	53.9	37.1	7.7	1.2	*0.000*	571
	Poor	35.5	49.2	13.2	2.1		431
Philippines	*Good*	56.0	34.0	6.8	3.3	*0.000*	636
	Poor	37.4	45.5	13.4	3.7		374
Russia	*Good*	32.6	18.5	25.3	23.6	*0.230*	178
	Poor	25.4	21.3	29.5	23.8		854
Saudi Arab.	*Good*	30.2	52.0	17.3	0.6	*0.000*	683
	Poor	9.6	47.2	36.0	7.1		322
Singapore	*Good*	53.4	34.8	9.3	2.4	*0.000*	698
	Poor	36.9	45.4	15.4	2.3		306
South Afr.	*Good*	54.2	29.1	10.1	6.7	*0.000*	554
	Poor	31.1	40.8	19.7	8.3		456
South Korea	*Good*	34.2	39.1	15.8	10.9	*0.000*	488
	Poor	20.5	27.5	25.3	26.7		513
Denmark	*Good*	77.4	17.7	4.0	0.9	*0.000*	758
	Poor	45.4	28.5	16.5	9.6		260
Taiwan	*Good*	48.7	35.4	11.9	4.0	*0.002*	528
	Poor	43.2	30.9	19.7	6.1		472
Turkey	*Good*	35.2	27.0	25.8	12.0	*0.000*	341
	Poor	23.2	23.1	29.4	24.3		676
UK	*Good*	69.7	25.0	3.5	1.9	*0.000*	837
	Poor	46.4	37.6	7.7	8.3		181
USA	*Good*	76.5	18.1	3.9	1.5	*0.000*	813
	Poor	39.8	44.4	10.7	5.1		196

Table 5 Distribution of respondents reporting V1_4 by health status

Country/ Territory	Health	Lack of money prevents me from doing the things I want to do				χ^2 test	Total
		Often	Sometimes	Rarely	Never	*p-value*	cases
Brazil	*Good*	37.0	39.2	13.4	10.4	*0.000*	635
	Poor	51.4	31.4	10.4	6.8		366
Canada	*Good*	15.9	33.3	22.4	28.5	*0.000*	902
	Poor	32.5	25.3	19.3	22.9		166
China	*Good*	10.7	27.2	21.2	40.9	*0.276*	335
	Poor	14.4	24.2	18.8	42.6		665
France	*Good*	16.6	31.2	18.7	33.5	*0.031*	765
	Poor	23.4	31.9	12.3	32.3		235
Germany	*Good*	16.8	24.0	20.5	38.6	*0.064*	691
	Poor	23.3	24.8	17.9	33.9		330
Hong Kong	*Good*	7.1	25.5	34.9	32.5	*0.000*	553
	Poor	15.4	33.4	25.8	25.4		449
India	*Good*	18.0	43.2	30.6	8.1	*0.021*	532
	Poor	20.6	49.9	22.7	6.9		481
Japan	*Good*	4.5	33.2	41.8	20.5	*0.000*	665
	Poor	10.4	39.4	35.8	14.3		335
Malaysia	*Good*	8.7	42.2	30.2	18.8	*0.024*	606
	Poor	11.4	47.5	28.8	12.4		396
Mexico	*Good*	16.6	44.3	23.1	15.9	*0.000*	571
	Poor	25.3	52.2	12.3	10.2		431
Philippines	*Good*	28.1	40.1	14.8	17.0	*0.137*	636
	Poor	35.0	35.8	14.4	14.7		374
Russia	*Good*	21.3	15.2	37.6	25.8	*0.000*	178
	Poor	34.2	25.5	23.1	17.2		854
Saudi Arab.	*Good*	9.1	44.5	30.7	15.7	*0.000*	683
	Poor	15.8	55.3	20.2	8.7		322
Singapore	*Good*	12.5	34.4	22.9	30.2	*0.000*	698
	Poor	12.4	34.6	34.0	19.0		306
South Afr.	*Good*	25.6	42.4	17.9	14.1	*0.000*	554
	Poor	48.2	31.1	14.3	6.4		456
South Korea	*Good*	11.7	27.5	18.9	42.0	*0.000*	488
	Poor	23.0	29.6	15.4	32.0		513
Denmark	*Good*	6.2	14.9	27.6	51.3	*0.000*	758
	Poor	13.5	14.2	18.8	53.5		260
Taiwan	*Good*	4.0	22.3	25.0	48.7	*0.000*	528
	Poor	11.7	28.2	26.7	33.5		472
Turkey	*Good*	35.2	17.9	24.6	22.3	*0.000*	341
	Poor	51.8	14.5	15.5	18.2		676
UK	*Good*	14.8	32.5	15.4	37.3	*0.002*	837
	Poor	26.5	28.7	13.3	31.5		181
USA	*Good*	14.1	33.9	26.9	25.0	*0.000*	813
	Poor	38.3	27.6	16.3	17.9		196

Table 6 Distribution of respondents reporting V1_5 by health status

Country/ Territory	Health	Family responsibilities prevent me from doing what I want to do				χ^2 test *p-value*	Total cases
		Often	Sometimes	Rarely	Never		
Brazil	*Good*	9.0	38.9	24.7	27.4	*0.001*	635
	Poor	16.1	29.5	28.4	26.0		366
Canada	*Good*	8.0	22.1	30.4	39.6	*0.015*	902
	Poor	13.3	18.7	21.7	46.4		166
China	*Good*	10.1	24.8	20.9	44.2	*0.015*	335
	Poor	15.2	27.8	14.6	42.4		665
France	*Good*	9.3	22.6	20.0	48.1	*0.218*	765
	Poor	7.2	20.4	16.6	55.7		235
Germany	*Good*	9.1	22.0	22.9	46.0	*0.730*	691
	Poor	9.4	19.7	21.5	49.4		330
Hong Kong	*Good*	5.4	29.5	30.6	34.5	*0.000*	553
	Poor	14.0	35.0	21.8	29.2		449
India	*Good*	22.0	46.2	23.3	8.5	*0.053*	532
	Poor	15.6	52.4	22.7	9.4		481
Japan	*Good*	3.0	30.1	45.4	21.5	*0.016*	665
	Poor	4.5	37.0	43.9	14.6		335
Malaysia	*Good*	8.7	28.5	32.0	30.7	*0.004*	606
	Poor	6.8	28.5	41.9	22.7		396
Mexico	*Good*	6.8	30.8	24.7	37.7	*0.856*	571
	Poor	7.9	30.4	23.0	38.7		431
Philippines	*Good*	7.1	28.6	17.0	47.3	*0.929*	636
	Poor	7.2	27.3	18.4	47.1		374
Russia	*Good*	14.6	9.0	23.0	53.4	*0.556*	178
	Poor	13.6	11.0	26.7	48.7		854
Saudi Arab.	*Good*	11.4	47.6	33.1	7.9	*0.001*	683
	Poor	20.5	46.6	28.3	4.7		322
Singapore	*Good*	12.3	33.8	24.9	28.9	*0.006*	698
	Poor	13.7	31.7	34.0	20.6		306
South Afr.	*Good*	8.3	30.3	29.8	31.6	*0.000*	554
	Poor	14.3	28.9	34.9	21.9		456
South Korea	*Good*	8.2	25.8	13.9	52.0	*0.019*	488
	Poor	14.4	23.0	14.0	48.5		513
Denmark	*Good*	3.8	10.8	19.8	65.6	*0.442*	758
	Poor	4.2	9.6	15.8	70.4		260
Taiwan	*Good*	13.8	28.6	21.8	35.8	*0.064*	528
	Poor	12.5	33.7	25.0	28.8		472
Turkey	*Good*	12.9	10.9	20.8	55.4	*0.001*	341
	Poor	23.1	10.9	17.0	49.0		676
UK	*Good*	4.9	26.6	20.1	48.4	*0.014*	837
	Poor	11.0	22.7	18.2	48.1		181
USA	*Good*	5.2	25.0	30.0	39.9	*0.000*	813
	Poor	13.3	23.5	23.5	39.8		196

Table 7 Distribution of respondents reporting V1_6 by health status

Country/ Territory	Health	I feel life is full of responsibilities				χ^2 test	Total
		Often	Sometimes	Rarely	Never	*p-value*	cases
Brazil	*Good*	30.7	52.1	15.1	2.0	*0.000*	635
	Poor	20.2	45.4	25.4	9.0		366
Canada	*Good*	73.3	21.2	3.4	2.1	*0.000*	902
	Poor	50.6	31.3	12.0	6.0		166
China	*Good*	24.5	31.3	21.2	23.0	*0.002*	335
	Poor	16.5	27.4	25.4	30.7		665
France	*Good*	60.3	30.6	6.9	2.2	*0.000*	765
	Poor	48.1	35.3	9.4	7.2		235
Germany	*Good*	72.9	19.7	5.5	1.9	*0.000*	691
	Poor	55.8	27.3	11.5	5.5		330
Hong Kong	*Good*	21.7	40.0	21.5	16.8	*0.000*	553
	Poor	12.5	30.5	34.5	22.5		449
India	*Good*	33.6	44.5	16.9	4.9	*0.000*	532
	Poor	19.3	49.5	25.4	5.8		481
Japan	*Good*	29.6	38.6	29.0	2.7	*0.000*	665
	Poor	16.7	40.6	39.4	3.3		335
Malaysia	*Good*	36.1	50.3	12.7	0.8	*0.000*	606
	Poor	17.4	52.0	29.0	1.5		396
Mexico	*Good*	52.4	39.8	6.5	1.4	*0.000*	571
	Poor	36.0	47.8	12.8	3.5		431
Philippines	*Good*	30.0	36.6	10.5	22.8	*0.002*	636
	Poor	27.0	44.9	13.6	14.4		374
Russia	*Good*	52.8	15.2	19.7	12.4	*0.000*	178
	Poor	31.0	19.2	26.5	23.3		854
Saudi Arab.	*Good*	39.5	52.0	8.3	0.1	*0.000*	683
	Poor	19.9	59.9	17.4	2.8		322
Singapore	*Good*	52.1	32.7	11.2	4.0	*0.000*	698
	Poor	33.0	43.1	20.6	3.3		306
South Afr.	*Good*	41.0	40.4	12.6	6.0	*0.000*	554
	Poor	20.4	38.8	31.4	9.4		456
South Korea	*Good*	24.4	33.2	22.7	19.7	*0.000*	488
	Poor	14.6	22.8	22.8	39.8		513
Denmark	*Good*	80.5	13.6	5.3	0.7	*0.000*	758
	Poor	53.1	24.6	15.4	6.9		260
Taiwan	*Good*	33.7	30.1	21.4	14.8	*0.000*	528
	Poor	19.9	21.0	35.2	23.9		472
Turkey	*Good*	27.9	24.0	29.9	18.2	*0.000*	341
	Poor	29.9	17.2	20.6	32.4		676
UK	*Good*	58.5	31.9	7.0	2.5	*0.000*	837
	Poor	40.3	34.8	14.4	10.5		181
USA	*Good*	77.7	17.0	3.7	1.6	*0.000*	813
	Poor	54.1	35.2	7.7	3.1		196

Table 8 Distribution of respondents reporting V1_7 by health status

Country/ Territory	Health	I feel the future looks good for me				χ^2 test	Total
		Often	Sometimes	Rarely	Never	*p-value*	cases
Brazil	*Good*	31.3	55.6	10.9	2.2	*0.000*	635
	Poor	18.6	52.7	21.9	6.8		366
Canada	*Good*	75.5	18.4	4.1	2.0	*0.000*	902
	Poor	45.8	31.9	13.3	9.0		166
China	*Good*	43.6	26.0	13.7	16.7	*0.000*	335
	Poor	30.5	25.7	21.8	22.0		665
France	*Good*	53.3	29.2	12.0	5.5	*0.000*	765
	Poor	36.2	31.9	17.4	14.5		235
Germany	*Good*	52.7	30.5	10.3	6.5	*0.000*	691
	Poor	30.3	31.5	19.4	18.8		330
Hong Kong	*Good*	24.8	34.9	23.9	16.5	*0.000*	553
	Poor	12.9	27.4	35.9	23.8		449
India	*Good*	27.8	49.1	19.7	3.4	*0.000*	532
	Poor	16.4	55.5	21.8	6.2		481
Japan	*Good*	20.2	39.2	37.6	3.0	*0.000*	665
	Poor	6.9	30.7	56.1	6.3		335
Malaysia	*Good*	28.5	55.0	15.7	0.8	*0.000*	606
	Poor	13.4	50.3	33.1	3.3		396
Mexico	*Good*	44.7	46.4	7.5	1.4	*0.000*	571
	Poor	19.7	50.3	24.8	5.1		431
Philippines	*Good*	52.8	35.2	6.3	5.7	*0.002*	636
	Poor	40.4	44.7	7.2	7.8		374
Russia	*Good*	16.3	14.0	29.8	39.9	*0.017*	178
	Poor	10.1	10.3	29.5	50.1		854
Saudi Arab.	*Good*	22.8	61.1	15.2	0.9	*0.000*	683
	Poor	9.6	50.0	33.5	6.8		322
Singapore	*Good*	49.4	33.5	12.2	4.9	*0.000*	698
	Poor	30.7	40.8	21.6	6.9		306
South Afr.	*Good*	37.0	42.6	15.7	4.7	*0.000*	554
	Poor	18.6	33.3	35.1	12.9		456
South Korea	*Good*	26.4	36.1	20.5	17.0	*0.000*	488
	Poor	16.8	20.3	26.3	36.6		513
Denmark	*Good*	78.8	15.6	5.1	0.5	*0.000*	758
	Poor	41.9	39.6	12.7	5.8		260
Taiwan	*Good*	37.3	28.6	18.4	15.7	*0.000*	528
	Poor	20.8	23.5	31.8	23.9		472
Turkey	*Good*	29.0	17.6	24.9	28.4	*0.002*	341
	Poor	23.1	13.6	22.6	40.7		676
UK	*Good*	62.6	31.4	4.2	1.8	*0.000*	837
	Poor	32.6	44.8	12.7	9.9		181
USA	*Good*	75.2	20.7	3.0	1.2	*0.000*	813
	Poor	41.8	40.8	12.2	5.1		196

Table 9 Distribution of respondents answering question "how often have you felt the following over the last week?" i.e., V2_1 by health

Country/ Territory	Health	I felt depressed					χ^2 test *p-value*
		Almost all of the time	Most of the time	Some of the time	Almost none of the time	Never	
Brazil	*Good*	1.7	6.8	23.0	32.0	36.5	*0.000*
	Poor	9.8	15.3	33.1	24.3	17.5	
Canada	*Good*	1.4	2.1	14.7	45.7	36.0	*0.000*
	Poor	3.0	5.4	37.3	33.1	21.1	
China	*Good*	0.9	3.6	26.0	41.8	27.8	*0.764*
	Poor	0.9	4.8	28.0	41.5	24.8	
France	*Good*	1.6	2.6	19.2	24.6	52.0	*0.000*
	Poor	5.5	8.1	34.5	14.0	37.9	
Germany	*Good*	1.3	2.0	15.3	33.4	47.9	*0.000*
	Poor	4.5	3.6	27.3	28.2	36.4	
Hong Kong	*Good*	0.7	2.2	16.8	59.9	20.4	*0.000*
	Poor	2.2	5.8	25.6	52.3	14.0	
India	*Good*	5.3	26.5	37.2	9.2	21.8	*0.021*
	Poor	7.3	29.7	41.2	6.7	15.2	
Japan	*Good*	0.5	6.3	32.6	41.1	19.5	*0.011*
	Poor	1.8	10.1	31.0	43.0	14.0	
Malaysia	*Good*	0.2	1.5	21.0	58.6	18.8	*0.000*
	Poor	0.8	6.3	38.6	46.2	8.1	
Mexico	*Good*	2.1	4.2	20.3	48.7	24.7	*0.000*
	Poor	5.3	12.5	35.5	29.5	17.2	
Philippines	*Good*	1.3	7.5	52.5	23.0	15.7	*0.000*
	Poor	2.1	11.8	60.2	12.0	13.9	
Russia	*Good*	6.7	2.2	25.3	24.2	41.6	*0.000*
	Poor	13.7	6.9	36.1	13.0	30.3	
Saudi Arab.	*Good*	0.6	5.6	16.7	43.3	33.8	*0.000*
	Poor	2.2	11.5	30.7	37.3	18.3	
Singapore	*Good*	1.4	2.9	32.1	32.8	30.8	*0.000*
	Poor	1.3	8.8	35.3	34.3	20.3	
South Afr.	*Good*	3.1	7.6	37.9	27.8	23.6	*0.000*
	Poor	12.7	21.1	42.3	14.7	9.2	
South Korea	*Good*	2.0	6.1	24.4	40.8	26.6	*0.000*
	Poor	7.2	12.7	28.5	35.5	16.2	
Denmark	*Good*	0.8	1.2	9.1	25.6	63.3	*0.000*
	Poor	3.1	4.6	16.9	30.8	44.6	
Taiwan	*Good*	0.6	0.8	12.3	59.7	26.7	*0.000*
	Poor	1.1	3.2	25.8	42.6	27.3	
Turkey	*Good*	10.0	7.6	29.3	40.2	12.9	*0.000*
	Poor	19.8	11.7	34.8	23.8	9.9	
UK	*Good*	0.2	0.7	14.7	31.4	52.9	*0.000*
	Poor	1.7	5.0	29.8	28.2	35.4	
USA	*Good*	0.9	0.6	16.9	47.7	33.9	*0.000*
	Poor	5.6	4.1	41.3	27.6	21.4	

Table 10 Distribution of respondents answering question "how often have you felt the following over the last week?" i.e., V2_2 by health

Country/ Territory	Health	I was happy					χ^2 test *p-value*
		Almost all of the time	Most of the time	Some of the time	Almost none of the time	Never	
Brazil	*Good*	30.4	41.9	23.6	3.8	0.3	*0.000*
	Poor	15.8	34.4	36.3	11.5	1.9	
Canada	*Good*	47.5	41.6	9.2	1.4	0.3	*0.000*
	Poor	31.3	38.6	26.5	1.8	1.8	
China	*Good*	24.5	44.8	22.4	7.2	1.2	*0.005*
	Poor	17.7	41.2	26.0	12.8	2.3	
France	*Good*	39.7	39.5	16.7	2.7	1.3	*0.000*
	Poor	31.1	33.2	28.5	4.7	2.6	
Germany	*Good*	43.4	34.9	17.8	2.9	1.0	*0.000*
	Poor	31.2	33.9	24.2	5.8	4.8	
Hong Kong	*Good*	14.8	38.3	34.4	10.8	1.6	*0.000*
	Poor	11.8	23.4	43.9	17.8	3.1	
India	*Good*	37.2	38.7	19.7	3.4	0.9	*0.000*
	Poor	20.8	41.8	29.3	5.2	2.9	
Japan	*Good*	33.7	37.6	21.7	6.5	0.6	*0.000*
	Poor	18.2	37.3	31.0	12.8	0.6	
Malaysia	*Good*	33.0	47.4	18.8	0.8	-	*0.000*
	Poor	17.4	47.0	34.1	1.5	-	
Mexico	*Good*	32.2	53.1	13.1	1.4	0.2	*0.000*
	Poor	27.4	45.2	22.3	4.6	0.5	
Philippines	*Good*	37.3	48.1	13.7	0.9	0	*0.002*
	Poor	27.5	50.0	21.1	1.1	0.3	
Russia	*Good*	29.2	24.2	28.1	6.2	12.4	*0.000*
	Poor	19.7	14.2	33.4	12.5	20.3	
Saudi Arab.	*Good*	21.5	50.8	24.7	2.8	0.1	*0.000*
	Poor	7.5	36.0	46.3	9.6	0.6	
Singapore	*Good*	37.4	37.2	19.6	5.3	0.4	*0.000*
	Poor	18.0	43.5	33.0	5.2	0.3	
South Afr.	*Good*	33.8	43.7	20.6	1.4	0.5	*0.000*
	Poor	16.2	40.1	36.0	6.1	1.5	
South Korea	*Good*	33.0	31.6	22.1	9.0	4.3	*0.000*
	Poor	25.1	23.2	27.5	17.0	7.2	
Denmark	*Good*	51.2	35.6	11.2	1.6	0.4	*0.000*
	Poor	37.7	38.5	16.9	6.2	0.8	
Taiwan	*Good*	18.8	44.1	18.9	18.0	0.2	*0.001*
	Poor	14.2	49.2	25.2	11.0	0.4	
Turkey	*Good*	38.7	21.7	26.1	9.1	4.4	*0.000*
	Poor	22.2	18.6	34.9	19.2	5.0	
UK	*Good*	38.9	45.0	13.9	1.2	1.0	*0.000*
	Poor	26.5	42.0	27.6	2.8	1.1	
USA	*Good*	45.0	42.1	11.4	1.2	0.2	*0.000*
	Poor	26.5	37.8	30.6	4.6	0.5	

Table 11 Distribution of respondents answering question "how often have you felt the following over the last week?" i.e., V2_3 by health

Country/ Territory	Health	I felt lonely					χ^2 test *p-value*
		Almost all of the time	Most of the time	Some of the time	Almost none of the time	Never	
Brazil	*Good*	1.9	5.5	21.7	34.6	36.2	*0.000*
	Poor	10.4	12.6	27.3	28.7	21.0	
Canada	*Good*	1.1	2.4	19.2	42.4	34.9	*0.000*
	Poor	3.6	7.2	31.9	35.5	21.7	
China	*Good*	1.2	4.2	21.8	43.9	29.0	*0.606*
	Poor	2.0	4.8	24.5	43.3	25.4	
France	*Good*	3.3	3.5	17.9	24.2	51.1	*0.000*
	Poor	10.2	6.4	24.3	20.4	38.7	
Germany	*Good*	0.7	2.2	10.9	32.1	54.1	*0.003*
	Poor	2.7	3.3	16.4	31.2	46.4	
Hong Kong	*Good*	0.7	1.6	27.7	55.0	15.0	*0.000*
	Poor	2.9	9.1	27.4	45.9	14.7	
India	*Good*	6.8	21.4	31.2	11.8	28.8	*0.002*
	Poor	7.3	24.5	38.7	11.2	18.3	
Japan	*Good*	0.2	5.3	32.8	43.8	18.0	*0.006*
	Poor	2.4	6.3	31.0	45.4	14.9	
Malaysia	*Good*	0.2	3.0	27.7	52.0	17.2	*0.000*
	Poor	0.8	9.1	44.4	38.4	7.3	
Mexico	*Good*	1.6	5.1	17.5	46.6	29.2	*0.000*
	Poor	3.5	11.1	27.8	33.4	24.1	
Philippines	*Good*	2.4	7.9	51.6	23.6	14.6	*0.246*
	Poor	2.1	11.0	54.0	18.7	14.2	
Russia	*Good*	3.9	1.1	16.3	12.4	66.3	*0.071*
	Poor	8.4	3.6	13.7	13.9	60.3	
Saudi Arab.	*Good*	1.0	3.8	20.2	41.3	33.7	*0.000*
	Poor	2.5	9.9	39.1	32.0	16.5	
Singapore	*Good*	1.4	4.7	26.2	36.1	31.5	*0.000*
	Poor	2.9	10.1	36.9	31.0	19.0	
South Afr.	*Good*	2.9	8.8	31.4	30.5	26.4	*0.000*
	Poor	10.5	18.2	32.9	20.8	17.5	
South Korea	*Good*	2.0	6.6	23.8	40.6	27.0	*0.000*
	Poor	8.8	13.3	28.3	33.9	15.8	
Denmark	*Good*	0.1	1.7	5.5	27.6	65.0	*0.000*
	Poor	5.4	2.7	9.6	25.8	56.5	
Taiwan	*Good*	0.4	1.9	15.0	57.2	25.6	*0.002*
	Poor	0.8	3.2	19.5	44.5	32.0	
Turkey	*Good*	9.4	7.0	23.8	48.1	11.7	*0.000*
	Poor	22.0	10.1	24.3	31.2	12.4	
UK	*Good*	1.0	1.4	13.3	30.2	54.1	*0.000*
	Poor	3.9	6.1	30.9	24.3	34.8	
USA	*Good*	1.6	2.1	20.0	46.6	29.6	*0.000*
	Poor	4.1	4.6	39.3	30.6	21.4	

Table 12 Distribution of respondents answering question "how often have you felt the following over the last week?" i.e., V2_4 by health

Country/ Territory	Health	I felt sad					χ^2 test *p-value*
		Almost all of the time	Most of the time	Some of the time	Almost none of the time	Never	
Brazil	*Good*	2.4	6.9	32.9	32.0	25.8	*0.000*
	Poor	11.7	13.4	36.1	23.5	15.3	
Canada	*Good*	1.8	1.8	30.3	43.0	23.2	*0.000*
	Poor	3.0	7.8	41.0	30.7	17.5	
China	*Good*	1.2	3.0	20.0	47.2	28.7	*0.390*
	Poor	0.9	4.7	23.9	44.4	26.2	
France	*Good*	2.5	2.5	32.2	24.6	38.3	*0.000*
	Poor	9.4	6.0	39.1	20.9	24.7	
Germany	*Good*	2.5	3.5	26.6	36.9	30.5	*0.001*
	Poor	3.9	7.6	33.0	29.7	25.8	
Hong Kong	*Good*	0.4	1.3	15.7	62.7	19.9	*0.000*
	Poor	2.4	3.8	26.7	52.6	14.5	
India	*Good*	2.8	18.0	35.2	17.1	26.9	*0.000*
	Poor	7.9	25.6	38.3	11.6	16.6	
Japan	*Good*	0.5	5.3	33.2	43.0	18.0	*0.163*
	Poor	0.9	8.1	33.1	44.5	13.4	
Malaysia	*Good*	0	1.8	28.2	53.1	16.8	*0.000*
	Poor	0.5	6.6	45.2	40.7	7.1	
Mexico	*Good*	1.8	5.4	24.2	48.5	20.1	*0.000*
	Poor	3.7	15.3	38.3	29.5	13.2	
Philippines	*Good*	1.1	6.1	67.0	16.0	9.7	*0.024*
	Poor	2.9	10.2	63.1	13.4	10.4	
Russia	*Good*	6.2	4.5	39.9	17.4	32.0	*0.000*
	Poor	14.2	9.6	40.7	14.6	20.8	
Saudi Arab.	*Good*	1.0	3.1	26.2	42.8	26.9	*0.000*
	Poor	1.9	6.8	40.7	34.8	15.8	
Singapore	*Good*	1.4	3.2	35.2	34.5	25.6	*0.000*
	Poor	3.9	11.8	38.9	29.1	16.3	
South Afr.	*Good*	1.6	7.2	41.2	33.2	16.8	*0.000*
	Poor	8.8	15.1	45.8	19.5	10.7	
South Korea	*Good*	1.6	5.7	20.3	44.7	27.7	*0.000*
	Poor	6.2	11.3	25.7	40.2	16.6	
Denmark	*Good*	1.1	2.6	17.2	36.4	42.7	*0.000*
	Poor	5.8	6.5	26.5	30.0	31.2	
Taiwan	*Good*	0.6	0.9	11.7	61.4	25.4	*0.000*
	Poor	0.8	3.4	23.5	42.2	30.1	
Turkey	*Good*	13.2	10.0	42.5	28.2	6.2	*0.000*
	Poor	27.4	13.8	38.5	14.1	6.4	
UK	*Good*	0.7	1.9	28.3	32.3	36.8	*0.000*
	Poor	3.3	4.4	42.5	24.9	24.9	
USA	*Good*	1.4	1.5	28.8	50.6	17.8	*0.000*
	Poor	2.6	5.1	48.0	31.6	12.8	

Table 13 Distribution of respondents answering question "how often have you felt the following over the last week?" i.e., V2_5 by health

Country/ Territory	Health	I felt everything was too much effort					χ^2 test *p-value*
		Almost all of the time	Most of the time	Some of the time	Almost none of the time	Never	
Brazil	*Good*	2.2	11.3	33.9	29.3	23.3	*0.000*
	Poor	7.9	16.1	40.2	22.4	13.4	
Canada	*Good*	1.8	1.9	18.3	44.7	33.4	*0.000*
	Poor	3.0	5.4	41.6	31.3	18.7	
China	*Good*	2.1	3.9	29.0	40.3	24.8	*0.002*
	Poor	3.3	9.5	31.6	37.9	17.7	
France	*Good*	1.0	1.2	18.3	25.5	54.0	*0.000*
	Poor	6.0	5.5	28.1	23.4	37.0	
Germany	*Good*	1.9	3.8	23.3	34.4	36.6	*0.000*
	Poor	9.1	7.3	31.5	25.5	26.7	
Hong Kong	*Good*	1.3	6.5	30.4	48.6	13.2	*0.000*
	Poor	4.5	8.9	36.7	39.2	10.7	
India	*Good*	9.8	30.8	32.3	10.9	16.2	*0.000*
	Poor	10.2	32.6	40.5	10.0	6.7	
Japan	*Good*	0.9	7.8	25.4	48.6	17.3	*0.965*
	Poor	1.2	8.1	23.6	49.9	17.3	
Malaysia	*Good*	1.2	4.6	34.2	43.6	16.5	*0.000*
	Poor	2.0	13.1	42.4	33.8	8.6	
Mexico	*Good*	4.2	11.0	23.6	38.4	22.8	*0.000*
	Poor	11.1	16.7	33.6	23.4	15.1	
Philippines	*Good*	7.1	17.5	41.0	19.7	14.8	*0.715*
	Poor	6.4	17.9	37.7	20.3	17.6	
Russia	*Good*	7.3	2.8	19.1	23.0	47.8	*0.000*
	Poor	15.0	13.0	27.0	14.4	30.6	
Saudi Arab.	*Good*	6.7	22.4	45.7	21.4	3.8	*0.000*
	Poor	15.5	32.0	38.8	13.0	0.6	
Singapore	*Good*	3.0	9.2	33.1	29.5	25.2	*0.002*
	Poor	4.6	10.8	41.8	28.1	14.7	
South Afr.	*Good*	2.0	9.9	37.0	31.0	20.0	*0.000*
	Poor	10.1	20.6	39.9	20.6	8.8	
South Korea	*Good*	6.4	8.2	13.5	35.5	36.5	*0.000*
	Poor	9.2	13.6	17.3	38.2	21.6	
Denmark	*Good*	1.1	2.6	14.1	36.1	46.0	*0.000*
	Poor	9.2	10.4	26.9	23.8	29.6	
Taiwan	*Good*	0.9	1.7	25.4	48.7	23.3	*0.000*
	Poor	1.5	7.8	34.3	32.8	23.5	
Turkey	*Good*	24.3	19.6	32.6	16.1	7.3	*0.761*
	Poor	27.5	19.7	29.6	15.1	8.1	
UK	*Good*	0.8	1.9	20.0	32.1	45.2	*0.000*
	Poor	3.3	8.3	35.9	29.3	23.2	
USA	*Good*	1.4	1.6	20.4	49.0	27.7	*0.000*
	Poor	4.1	6.1	39.8	34.7	15.3	

Table 14 Distribution of respondents answering question "how often have you felt the following over the last week?" i.e., V2_6 by health

Country/ Territory	Health	I enjoyed life: Almost all of the time	Most of the time	Some of the time	Almost none of the time	Never	χ^2 test *p-value*
Brazil	*Good*	25.4	39.8	29.1	4.9	0.8	*0.000*
	Poor	15.0	28.1	38.5	15.8	2.5	
Canada	*Good*	55.9	35.7	6.7	1.2	0.6	*0.000*
	Poor	36.7	35.5	25.3	1.8	0.6	
China	*Good*	24.2	42.1	21.2	10.4	2.1	*0.002*
	Poor	18.5	36.2	23.2	17.3	4.8	
France	*Good*	48.5	37.9	10.3	2.6	0.7	*0.000*
	Poor	38.7	35.7	18.7	3.8	3.0	
Germany	*Good*	46.5	28.7	17.7	4.8	2.5	*0.000*
	Poor	33.6	30.9	23.0	7.0	5.5	
Hong Kong	*Good*	16.5	34.7	36.2	10.8	1.8	*0.000*
	Poor	11.4	24.5	36.3	24.1	3.8	
India	*Good*	30.6	38.2	22.7	6.8	1.7	*0.000*
	Poor	17.9	40.3	34.7	3.5	3.5	
Japan	*Good*	29.8	30.1	28.3	11.3	0.6	*0.000*
	Poor	14.9	33.4	35.2	15.8	0.6	
Malaysia	*Good*	37.5	45.4	15.5	1.7	0	*0.000*
	Poor	18.9	49.5	27.3	4.0	0.3	
Mexico	*Good*	41.0	50.1	6.1	2.8	0	*0.000*
	Poor	36.7	42.7	16.2	3.7	0.7	
Philippines	*Good*	43.6	43.6	12.3	0.6	0	*0.000*
	Poor	31.3	47.6	18.7	2.1	0.3	
Russia	*Good*	27.0	21.9	28.7	5.6	16.9	*0.000*
	Poor	16.4	10.8	31.4	15.2	26.2	
Saudi Arab.	*Good*	24.6	49.5	22.3	2.5	1.2	*0.000*
	Poor	8.7	45.3	34.8	9.6	1.6	
Singapore	*Good*	43.1	34.4	15.9	5.7	0.9	*0.000*
	Poor	17.0	39.2	31.4	11.1	1.3	
South Afr.	*Good*	35.4	39.9	19.7	4.3	0.7	*0.000*
	Poor	21.1	27.9	36.4	11.4	3.3	
South Korea	*Good*	29.1	30.5	24.4	11.5	4.5	*0.000*
	Poor	21.8	19.9	30.0	20.5	7.8	
Denmark	*Good*	55.9	28.4	12.0	3.2	0.5	*0.000*
	Poor	36.2	33.8	17.7	8.8	3.5	
Taiwan	*Good*	17.2	39.2	23.9	19.5	0.2	*0.043*
	Poor	13.8	36.4	32.2	17.2	0.4	
Turkey	*Good*	25.5	18.5	34.0	16.4	5.6	*0.000*
	Poor	13.6	12.6	35.5	30.0	8.3	
UK	*Good*	51.3	37.2	10.3	0.7	0.6	*0.000*
	Poor	33.7	37.6	21.5	5.5	1.7	
USA	*Good*	58.3	32.6	8.4	0.4	0.4	*0.000*
	Poor	38.3	36.2	21.9	3.1	0.5	

Table 15 Distribution of respondents answering question "how often have you felt the following over the last week?" i.e., V2_7 by health

Country/ Territory	Health	I was looking forward to the future					χ^2 test *p-value*
		Almost all of the time	Most of the time	Some of the time	Almost none of the time	Never	
Brazil	*Good*	16.9	33.4	35.3	11.5	3.0	*0.000*
	Poor	10.9	19.9	39.3	25.4	4.4	
Canada	*Good*	46.6	31.0	15.5	5.0	1.9	*0.000*
	Poor	30.1	30.7	26.5	7.2	5.4	
China	*Good*	15.8	34.0	26.0	17.3	6.9	*0.001*
	Poor	11.3	27.1	24.8	26.3	10.5	
France	*Good*	30.6	33.1	21.3	7.1	8.0	*0.000*
	Poor	23.8	22.1	23.4	12.3	18.3	
Germany	*Good*	38.6	30.0	19.8	7.7	3.9	*0.000*
	Poor	22.7	26.7	31.5	11.2	7.9	
Hong Kong	*Good*	8.9	21.5	38.2	26.9	4.5	*0.000*
	Poor	5.6	14.0	31.6	41.2	7.6	
India	*Good*	19.5	37.6	33.6	4.7	4.5	*0.003*
	Poor	12.3	33.7	42.0	6.9	5.2	
Japan	*Good*	23.8	31.0	23.2	21.1	1.1	*0.000*
	Poor	10.7	23.9	31.9	30.7	2.7	
Malaysia	*Good*	25.9	43.1	27.4	3.6	0	*0.000*
	Poor	13.6	37.1	39.9	9.1	0.3	
Mexico	*Good*	17.5	41.7	30.1	8.1	2.6	*0.000*
	Poor	15.3	29.9	32.5	16.0	6.3	
Philippines	*Good*	39.6	41.7	14.6	2.5	1.6	*0.000*
	Poor	28.3	40.6	25.1	3.2	2.7	
Russia	*Good*	25.3	19.1	24.7	10.1	20.8	*0.000*
	Poor	14.3	13.2	25.2	14.6	32.7	
Saudi Arab.	*Good*	13.2	37.9	35.7	11.7	1.5	*0.000*
	Poor	6.2	28.3	43.2	16.8	5.6	
Singapore	*Good*	36.7	30.2	20.6	9.5	3.0	*0.000*
	Poor	19.0	35.9	22.2	17.6	5.2	
South Afr.	*Good*	36.1	36.3	20.4	5.4	1.8	*0.000*
	Poor	21.9	24.1	32.5	17.1	4.4	
South Korea	*Good*	20.3	24.4	21.7	21.7	11.9	*0.000*
	Poor	14.2	14.4	25.3	31.6	14.4	
Denmark	*Good*	43.7	27.4	16.5	9.1	3.3	*0.000*
	Poor	31.2	26.2	18.1	15.4	9.2	
Taiwan	*Good*	16.9	34.7	22.5	24.8	1.1	*0.000*
	Poor	11.7	28.2	28.2	28.0	4.0	
Turkey	*Good*	15.2	6.2	31.7	33.1	13.8	*0.001*
	Poor	20.6	11.5	22.5	32.2	13.2	
UK	*Good*	40.3	33.2	21.1	3.8	1.6	*0.000*
	Poor	26.0	25.4	31.5	12.2	5.0	
USA	*Good*	49.9	29.6	15.4	4.3	0.7	*0.000*
	Poor	29.6	33.7	27.6	5.6	3.6	

Table 16 Estimates of ordered logistic regression analysis for determining how people feel about their lives and feelings

Country / Territory	Independent variables	Dependent variables						
		V1_1	V1_2	V1_3	V1_4	V1_5	V1_6	V1_7
Brazil	*Age*	-0.040**	0.001	0.015**	0.029**	0.005	0.008	0.002
	Gender	-0.112	-0.223	0.122	-0.021	-0.139	0.202	0.312*
	Health	-1.121**	-0.783**	0.922**	-0.655**	-0.099	0.735**	0.807**
Canada	*Age*	-0.044**	-0.006	-0.001	0.032**	0.043**	0.028**	0.008
	Gender	0.068	0.009	-0.014	-0.151	-0.042	-0.035	-0.014
	Health	-1.065**	-0.942**	0.952**	-0.684**	-0.071	0.964**	1.316**
China	*Age*	-0.009	0.008	0.007	0.026**	0.019**	0.033**	0.000
	Gender	-0.090	-0.046	0.094	0.228*	-0.089	0.248*	-0.078
	Health	-0.496**	-0.462**	0.296*	-0.157	-0.304**	0.301*	0.534**
France	*Age*	-0.050**	0.015**	-0.011	0.026**	0.040**	-0.005	0.018**
	Gender	0.477**	-0.099	0.156	-0.249*	-0.108	0.056	0.474**
	Health	-1.107**	-0.700**	0.880**	-0.437**	-0.008	0.574**	0.606**
Germany	*Age*	-0.060**	0.002	-0.015*	0.031**	0.036**	0.001	0.010*
	Gender	0.051	-0.031	0.255	-0.024	-0.265*	0.123	0.282*
	Health	-0.926**	-0.536**	0.627**	-0.455**	-0.084	0.795**	0.991**
Hong Kong	*Age*	-0.022**	0.001	0.019**	0.016**	0.021**	0.046**	0.035**
	Gender	-0.076	-0.032	0.284*	0.238*	0.088	0292*	0.120
	Health	-0.734**	-0.535**	0.550**	-0.659**	-0.598**	0.454**	0.561**
India	*Age*	-0.013*	0.006	-0.001	0.005	0.009	0.003	0.006
	Gender	-0.306*	-0.449**	-0.064	0.136	0.013	0.188	0.256
	Health	-0.079	-0.031	0.390**	-0.327**	0.131	0.591**	0.433**
Japan	*Age*	-0.026**	-0.027**	-0.004	0.015**	0.018**	-0.006	0.009
	Gender	0.143	-0.375**	-0.144	0.536**	0.145	-0.463**	-0.168
	Health	-0.434**	0.309*	0.700**	-0.548**	-0.442**	0.569**	0.905**
Malaysia	*Age*	-0.077**	-0.017**	0.029**	0.020**	0.023**	0.018**	0.024**
	Gender	-0.088	-0.035	0.332**	-0.039	-0.408**	0.412**	0.358**
	Health	-0.909**	-0.958**	0.951**	-0.566**	-0.330*	0.793**	0.762**
Mexico	*Age*	-0.048**	-0.015**	0.039**	-0.002	-0.005	0.030**	0.037**
	Gender	-0.098	-0.245*	0.052	-0.230	-0.307**	-0.099	0.244**
	Health	-0.648**	-0.385**	0.585**	-0.649**	0.043	0.599**	1.167**
Philippines	*Age*	-0.038**	-0.021**	0.036**	0.011*	0.002	0.011**	0.011*
	Gender	-0.409**	-0.241*	-0.168	-0.247*	-0.294*	-0.042	-0.050
	Health	-0.541**	-0.171	0.555**	-0.282*	0.001	-0.164	0.397**
Russia	*Age*	-0.047**	-0.016**	0.011*	0.001	-0.003	0.031**	0.000
	Gender	-0.116	-0.268	-0.232	-0.230	-0.354*	0.065	-0.184
	Health	-0.671**	-0.536**	0.172	-0.675**	-0..025	0.623**	0.514**

Table 16 (Cont.) Estimates of ordered logistic regression analysis for determining how people feel about their lives and feelings

Country / Territory	Independent variables	Dependent variables						
		V1_1	V1_2	V1_3	V1_4	V1_5	V1_6	V1_7
Saudi Arab.	*Age*	-0.034**	-0.002	0.029**	0.016**	0.019**	0.014*	0.038**
	Gender	0.264*	0.323**	0.159	0.178	0.318**	0.275*	0.300*
	Health	-1.077**	-0.856**	1.011**	-0.910**	-0.725**	0.810**	0.790**
Singapore	*Age*	-0.029**	-0.013*	0.024**	0.012*	0.010*	0.040**	0.044**
	Gender	-0.387**	-0.044	-0.013	0.059	-0.040	-0.066	-0.124
	Health	-0.639**	-0.392**	0.428**	-0.296*	-0.234	0.385**	0.421**
South Afr.	*Age*	-0.058**	-0.028**	0.007	-0.005	0.003	0.002	0.011*
	Gender	-0.098	-0.103	-0.038	-0.200	0.039	0.120	0.031
	Health	-1.117**	-1.081**	0.784**	-0.812**	-0.382**	0.996**	1.048**
South Korea	*Age*	-0.027**	-0.001	0.031**	0.008	0.037**	0.059**	0.047**
	Gender	-0.582**	-0.256*	0.131	-0.092	-0.378**	0.211	-0.388**
	Health	-0.475**	-0.397**	0.789**	-0.592**	-0.400**	0.539**	0.694**
Denmark	*Age*	-0.031**	0.021**	-0.025**	0.062**	0.056**	-0.011	-.016*
	Gender	0.200	-0.108	0.186	-0.343*	0.033	-0.131	0.141
	Health	-1.115**	-1.031**	1.654**	-0.356*	-0.034	1.410**	1.645**
Taiwan	*Age*	-0.041**	-0.011*	0.019**	-0.017**	-0.009	0.069**	0.067**
	Gender	-0.342**	-0.214	0.069	-0.040	-0.371**	0.302*	0.239*
	Health	-0.094	-0.221	0.203	-0.551**	-0.081	0.351**	0.348**
Turkey	*Age*	-0.020**	0.022**	0.006	0.017**	0.028**	0.019**	0.019**
	Gender	-0.548**	-0.032	0.009	-0.390**	-0.444**	0.007	-0.030
	Health	-0.695**	-0.188	0.640**	-0533**	-0.395**	0.209	0.380**
UK	*Age*	-0.050**	0.009	-0.009	0.016**	0.036**	0.014*	0.016**
	Gender	0.438**	0.128	0.120	-0.167	-0.134	0.043	-0.100
	Health	-0.991**	-0.847**	1.081**	-0.515**	-0.299*	0.821**	1.263**
USA	*Age*	-0.036**	0.010	-0.015*	0.033**	0.037**	0.016**	0.009
	Gender	0.233	-0.070	0.095	-0.175	-0.174	-0.221	-0.124
	Health	-1.117**	-0.972**	1.587**	-1.120**	-0.420	0.973**	1.420**

Table 17 Estimates of ordered logistic regression analysis for determining how often have you felt the following over the last week?

Country / Territory	Independent variables	Dependent variables						
		V2_1	V2_2	V2_3	V2_4	V2_5	V2_6	V2_7
Brazil	*Age*	-0.006	0.003	-0.005	0.002	0.011*	0.012*	0.013*
	Gender	-0.727**	0.120	-0.515**	-0.609**	-0.177	0.455**	0.038
	Health	-1.064**	0.925**	-0.879**	-0.837**	-0.737**	0.836**	0.741**
Canada	*Age*	0.005	-0.006	-0.005	0.001	0.000	-0.009	0.009
	Gender	-0.458**	0.213	-0.306**	-0.513**	-0.154	0.252*	-0.077
	Health	-1.133**	0.945**	-0.852**	-0.718**	-1.091**	1.032**	0.713**
China	*Age*	0.011*	-0.017**	0.011*	0.009	-0.006	-0.013*	0.022**
	Gender	-0.153	-0.138	-0.081	-0.118	-0.107	-0.164	0.133
	Health	-0.212	0.546**	-0.246	-0.239	-0.412**	0.534**	0.396**
France	*Age*	0.013*	0.007	-0.012*	-0.007	0.016**	-0.000	0.008
	Gender	-0.686**	0.441**	-0.637**	-0.826**	-0.546**	0.408**	0.390**
	Health	-0.948**	0.483**	-0.564**	-0.700**	-0.960**	0.516**	0.588**
Germany	*Age*	-0.007	-0.002	-0.010	-0.003	0.005	-0.001	0.002
	Gender	-0.366**	0.314**	-0.123	-0.509**	-0.725**	0.352**	0.324**
	Health	-0.630**	0.639**	-0.355**	-0.433**	-0.804**	0.558**	0.755**
Hong Kong	*Age*	-0.004	0.002	-0.007	-0.005	-0.014**	0.004	0.031**
	Gender	-0.005	0.112	0.199	0.106	0.150	0.061	0.264*
	Health	-0.625**	0.614**	-0.380**	-0.673**	-0.426**	0.678**	0.516**
India	*Age*	0.011*	0.002	0.013*	0.016**	0.019**	0.002	0.005
	Gender	-0.425**	0.135	-0.313*	-0.443**	-0.200	0.132	0.135
	Health	-0.373**	0.713**	-0.400**	-0.732**	-0.367**	0.491**	0.425**
Japan	*Age*	0.008	-0.008	0.009	0.005	-0.002	-0.003	0.010*
	Gender	0.312**	-0.626**	0.418**	0.205	0.319*	-0.447**	-0.002
	Health	-0.278*	0.744**	-0.156	-0.226	0.045	0.557**	0.731**
Malaysia	*Age*	0.004	-0.010	-0.012*	0.007	-0.006	-0.010	0.018**
	Gender	-0.065	0.034	-0.322**	-0.129	-0.129	0.220	0.206
	Health	-1.100**	0.940**	-0.867**	-1.039**	-0.705**	0.977**	0.604**
Mexico	*Age*	-0.005	0.011*	-0.015**	0.004	-0.015**	0.009	0.040**
	Gender	-0.601**	0.250*	-0.389**	-0.659**	-0.117	0.175	0.022
	Health	-0.875**	0.410**	-0.512**	-0.912**	-0.739**	0.344**	0.394**
Philippines	*Age*	0.003	-0.004	-0.007	-0.001	0.012*	0.002	0.020**
	Gender	-0.436**	0.072	-0.121	-0.320*	0.040	0.234	-0.213
	Health	-0.523**	0.494**	-0.178	-0.261	0.054	0.548**	0.479**
Russia	*Age*	-0.015**	0.018**	-0.013**	-0.012*	-0.021**	0.013*	0.020**
	Gender	-0.534**	0.002	0.115	-0.415**	-0.242	0.027	0.086
	Health	-0.489**	0.572**	-0.234	-0.485**	-0.699**	0.713**	0.553**

Table 17(Cont.) Estimates of ordered logistic regression analysis for determining how often have you felt the following over the last week?

Country / Territory	Independent variables	Dependent variables						
		V1_1	V1_2	V1_3	V1_4	V1_5	V1_6	V1_7
Saudi Arabia	*Age*	-0.001	0.001	-0.010	-0.003	0.006	0.008	0.027**
	Gender	-0.323**	0.425**	-0.472**	-0.699**	0.294*	0.669**	0.327**
	Health	-0.891**	1.193**	-0.928**	-0.676**	-0.897**	0.856**	0.388**
Singapore	*Age*	0.004	0.013*	-0.012*	0.006	0.003	0.018**	0.032**
	Gender	-0.305**	-0.054	-0.188	-0.298*	-0.259*	-0.042	-0.026
	Health	-0.540**	0.620**	-0.633**	-0.768**	-0.510**	0.937**	0.446**
South Africa	*Age*	0.008	-0.016**	-0.014**	0.000	0.001	-0.010	0.010*
	Gender	-0.035	-0.005	-0.060	-0.067	0.084	0.073	0.012
	Health	-1.347**	1.113**	-0.695**	-0.892**	-1.039**	1.061**	0.898**
South Korea	*Age*	0.003	-0.007	-0.001	-0.002	0.011*	-0.008	0.037**
	Gender	-0.540**	-0.255*	-0.499**	-0.515**	-0.261*	-0.180	-0.211
	Health	-0.699**	0.630**	-0.775**	-0.681**	-0.666**	0.675**	0.336**
Denmark	*Age*	0.016**	-0.019**	0.013*	0.030**	0.032**	-0.040**	-0.004
	Gender	0.072	0.085	-0.009	-0.303*	-0.197	0.205	0.080
	Health	-0.920**	0.698**	-0.544**	-0.847**	-1.287*	1.043**	0.642**
Taiwan	*Age*	-0.008	-0.004	-0.018**	-0.012*	-0.031**	0.004	0.031**
	Gender	-0.238	0.130	-0.059	-0.232	-0.226	0.133	-0.003
	Health	-0.362**	-0.019	0.146	-0.156	-0.234	0.125	0.214
Turkey	*Age*	0.011	-0.010	-0.005	-0.003	0.022**	-0.002	0.008
	Gender	-0.576**	0.604**	-0.622**	-0.640**	-0.270*	0.272*	-0.111
	Health	-0.656**	0.718**	-0.473**	-0.633**	-0.106	0.742**	-0.206
UK	*Age*	-0.001	-0.011	-0.005	0.002	0.001	-0.019**	0.012*
	Gender	-0.154	-0.017	-0.010	-0.174	-0.031	0.079	0.098
	Health	-0.938**	0.765**	-1.038**	-0.776**	-1.126**	1.008**	0.850**
USA	*Age*	0.020**	-0.022**	0.009	0.010	0.012	-0.022**	0.001
	Gender	-0.388**	0.060	-0.290*	-0.328**	-0.245	0.086	-0.159
	Health	-1.378**	1.173**	-0.910**	-0.945**	-1.150**	1.060**	0.833**

4. COMMENTS AND CONCLUSIONS

It is hoped that the study will help us to enhance our understanding of wellbeing in old age. Our analyses show that there are health inequalities across different age cohorts and this is consistent for all selected counties and territories. The proportion reporting poor health increases with the biological age. Quality of life is examined for subjective as well as objective measures and it has been found that age, gender and health status of respondents are important determinants of QoL. From our analyses it may be concluded that the country level findings are helpful in understanding the role of health on the quality of life and they also enable us to make general conclusions at the global level. Finally, the results provide us a unique opportunity to learn findings from various countries and to make reasonable comparison. The findings are obviously helpful for policy-makers in order invest more in older adults and to enhance the wellbeing of people in general.

ACKNOWLEDGEMENTS

This research arises as a part of the strategic alliance between the Oxford Institute of Ageing (OIA) and the HSBC Bank Plc for promoting the understanding of ageing issues across the globe. Authors gratefully acknowledge the financial support of HSBC Bank to carry out research at the Oxford Institute of Ageing, University of Oxford in the UK. Paper presented at the 10th Islamic Countries Conference on Statistical Sciences (ICCS-X) held at the American University in Cairo during December 20-23, 2009.

REFERENCES

[1] Smith, J., Borchelt, M., Maier, H. and Jopp, D. (2002). "Health and well-being in the young old and oldest old", *Journal of Social Issues* 58(4): 715-732.

[2] Khan, H.T.A., and Leeson, G. (2009). Determinants of self appraisal of health status among older adults in Malaysia and Singapore: Findings from the Future of Retirement Survey 2007'. Paper presented at the 25th ASEASUK Conference 11-13 September, Swansea University, UK.

[3] Munir, F., Khan, H.T.A., Yarker, J., Haslam, C., Long, H., Bains, M. and Kalawsky, K. (2009). Self management of health-behaviors among older and younger workers with chronic illness, *Patient Education and Counseling* 77(1):109-115.

[4] Rob, U. and Talukder, M.N. (2007). Health Sector Reform: Lessons from Developing Countries, Dhaka: Utso Prokashan.

[5] Hughes, D. and Leethongdee, S. (2007). Universal Coverage in the Land of Smiles: Lessons from Thailand's 30 Baht Health Reforms. Health Affairs 26(4):999-1008.

[6] Kalam, I.M.S. and Khan, H.T.A. (2007). Determinants of Health Status of Elderly People in Bangladesh, *Asian Profile* **35**(1):71-82.

[7] Khan, H.T.A. and Leeson, G. (2008). Attitudes towards bearing the cost of care in later life: Evidence from a global survey. Paper presented at the Radical Statistics Annual Conference held in Edinburgh, 1 March.

[8] World Health Organization (2006). The World Health Report 2006-Working for Health. Geneva: WHO.

[9] Feinstien, A.R. (1992). Benefits and obstacles for development of health status assessment measures in clinical settings. *Medical Care* 30(5): MS50-MS56.

[10] HSBC (2007). *The Future of Retirement: The new old age.* London: HSBC Insurance.

[11] HSBC (2008). *The Future of Retirement: Investing in later life*. London: HSBC Insurance.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 70–79.

THE DEMOGRAPHIC TREND IN AGEING POPULATION AND THE ROLE OF AGE FRIENDLY PRIMARY HEALTH CENTRES (PHCS) TO SUPPORT AGEING IN PLACE IN INDONESIA

Tri Budi W. Rahardjo[1,2,3], Moertiningsih Adioetomo[1,4], Subarkah[2], Vita Priantina Dewi[1], Yudarini[2], Toni Hartono[3] and Eef Hogervorst[5]

[1] Center for Ageing Studies Universitas Indonesia
[2] Center for Health Research Universitas Indonesia
[3] National Commission for Older Persons Indonesia
[4] Demographic Institute Universitas Indonesia
[5] Department of Human Sciences Loughborough University, UK

ABSTRACT

Indonesia is the fourth most populous country in the world, and tenth largest elderly population.In 2020, the number of older people will steadily increase to 28.8 million (11% of total population) while the under five population will gradually decrease in number. Hence, the increase in life expectancy was predicted from 45.7 years in 1970, to 65.4 in 2000, 69.9 in 2015, and 76.9 in 2050, while the increase in number of older people 60+, was predicted from 4.9 million in 1950, to 21.4 m in 2010 and to 79.8 in 2050. This demographic trend will have a significant impact on the elderly health status, in which the role of Age Friendly Primary Health Centres will be very important. The Age – friendly primary health centre principles are designed by WHO in 2002 to serve as a guide for community – based Primary Health Care Centres (PHC), to modify management and clinical services, staff training and environments to better fit the needs of older persons. Based on this policy, Age Friendly Primary Health Care Centres' implementation was declared by Ministry of Health (MOH), the Republic of Indonesia in 2003. To evaluate the role of Age Friendly PHC in Indonesia, The National Commission for Older Persons (2007) has conducted an assessment study at 4 provinces. The purpose of the study was to evaluate the principles address three major areas of Age Friendly PHC that were information and education; health care management; and physical environment. This study was conducted using a cross – sectional design in Jakarta, Yogyakarta, West Sumatra and West Java. The quantitative data was collected using a check list instrument, then was analyzed descriptively, and a qualitative approach was conducted using Focus Group Discussion among Formal and Informal Leaders and in-depth interview among Elderly Patients. The results showed as follows: (1) Among 20 PHCs, 13 of them were Age Friendly PHC. Both type of PHC have implemented Age Friendly Health Care Policy in which Integrated Community Posts have been used to provide older persons health services in the community. The PHCs had already implemented such activities as follows: information and education (83%), early detection (85%), physical exercise (100%), recreation (95%), medication (95%) and rehabilitation (30%). However, only 50% of health providers have been trained, and only 85% of PHCs have been comfortable (2). It was shown that most of the leaders still need support of the government to improve PHCs quality regarding age friendly services, particularly in conducting training for the doctors and nurses in geriatrics and caring for dementia patients. (3) The patients were satisfied with the services and activities provided by the PHCs, however the improvement of home care is still needed. They prefer to stay in their own houses with familiar environment. It was concluded that the implementation of age friendly PHC policy has been implemented but has not been in optimal goal. To cope with ageing in place programme to improve the wellness in ageing, the quality of age friendly PHC should be improved, in which Integrated Community Posts have important role.

Keywords: Demographic Trend, Ageing Population, Age Friendly PHCs, Ageing in Place

1. BACKGROUND

Population ageing is one of humanity's greatest triumphs. It takes on special meaning in the whole population. It will put increased demands on all aspects including demand on health in all countries. Many countries in the world will experience increasingly ageing populations in the 21st century. It was predicted that the increase in number of old people 60+, from 4.9 million in 1950, to 21.4 m in 2010, to 79.8 in 2050. In this case, the number of oldest old in 2050 will be 11.8 million (World Population Projection, 2006 revision calculated and reported by Adioetomo, 2009).

One of the countries that will have dramatic increase in the number of aged people in its populations is Indonesia. In Indonesia, by 2020 the number of older person will steadily increase to 28.8 million (11% of the total population), while under five population will gradually decrease in number. The increasing proportion of older persons is expected to results in an extraordinary increase in life expectation in this country. Increase in life expectancy from 45.7 years in 1970, to 65.4 in 2000, 69.9 in 2015, and 76.9 in 2050. It was declared by WHO in 2002, that there were aspects that have a particularly significant effect on the increasing of life expectancy. One of these aspects is health.

Knowledge and understanding of health is growing rapidly. Health will also influence the quality of life among older persons. Many studies have shown the relationship between health and the quality of life of older persons (WHO, 2002 and Handajani, 2006). If quality of life is to be maximized and health and welfare needs of an ageing society are to be adequately met, then the practices with regard to older people will be improved.

Ageing and health are trends that must not to be ignored. The fact that many older persons present with illnesses needs a development in the health care services. In relation to this, geriatric syndrome can also exist due to the increasing number of oldest old such as nutrition problem, osteoporosis, depression, dementia, Incontinence, DM, cardiovascular, cancer, hearing impairment, visual impairment, oral health problem and medically compromised, delirium, and immobility and falls (Rahardjo et.al, 2007). In this regard, Age – friendly principles that has been designed by WHO in 2002, has adopted by Indonesia government will solve the geriatric problem in the community care. These principles are used to serve as a guide for community – based Primary Health Care Centres (PHCs), to better fit the needs of older persons including oldest old. Based on this policy, Age Friendly Primary Health Care Centres implementation was declared by Ministry of Health (MOH) the Republic of Indonesia in 2003.

2. RATIONALE OF ASSESSING AGE FRIENDLY PRIMARY HEALTH CARE CENTRES (PHCS)

Based on the above mentioned description in relation to population ageing in Indonesia, we realize that studying the age friendly primary health care and integrated community post to support the improvement of older persons' wellbeing is essential. It is also relevant to evaluate principles addressing three major areas of Age Friendly PHC which are information and education; health care management for the elderly; and physical environment and the role of care givers in Integrated Community Post to support the elderly daily life. It will be crucial too to study this because Indonesia is a very large country with about 225 million people consisting of various ethnic groups as well as a large range of social and economic condition due to geographical, natural and human resources. The study purposes were also to investigate whether the role of PHC benefitted older populations and whether PHC also enable people with temporary or permanent functional limitations to access needed care and to maintain health and independence. Accordingly, living arrangement in which family has very important to take care of older persons should be taken into account, particularly in relation to the partnership with PHCs to support ageing in place programme.

3. THE DEMOGRAPHIC TREND ON AGEING IN INDONESIA

The world is ageing. Today, worldwide, there are approx. 600 millions persons aged 60+. The vast majority of older persons will be living in developing countries such as Indonesia, etc. Especially in Indonesia, there is demographic trends on ageing, as follows: increase in life expectancy from 45.7 years in 1970, to 65.4 in 2000, 69.9 in 2015, 76.9 in 2050; increase in number of old people 60+, from 4.9 million in 1950, to 21.4 m in 2010 and to 79.8 in 2050; and the consequences of economic and health as an effect on different needs for caring and health services (Adioetomo, 2009). This can be seen in chart below.

Older people in Indonesia constitute the fastest growing age group in the population. In this context, population above sixty years is considered aged. The aged population can be decomposed into groups of young old, middle old and grand old. In relation to this, there are literacy levels in the general population and the aged, differentials according to sex and rural/urban habitat. A study found that the literacy rate among elderly persons is relatively high in Indonesia. In this case, the rural elderly, especially women, seem to be more at a disadvantage, owing to poor quantity and quality of educational facilities in rural areas. (See diagram above and below)

With people living longer and fewer children being born, as mentioned earlier, the absolute number of older persons is increasing in Indonesia. The estimation of number of older persons in Indonesia can be seen in the diagram below.

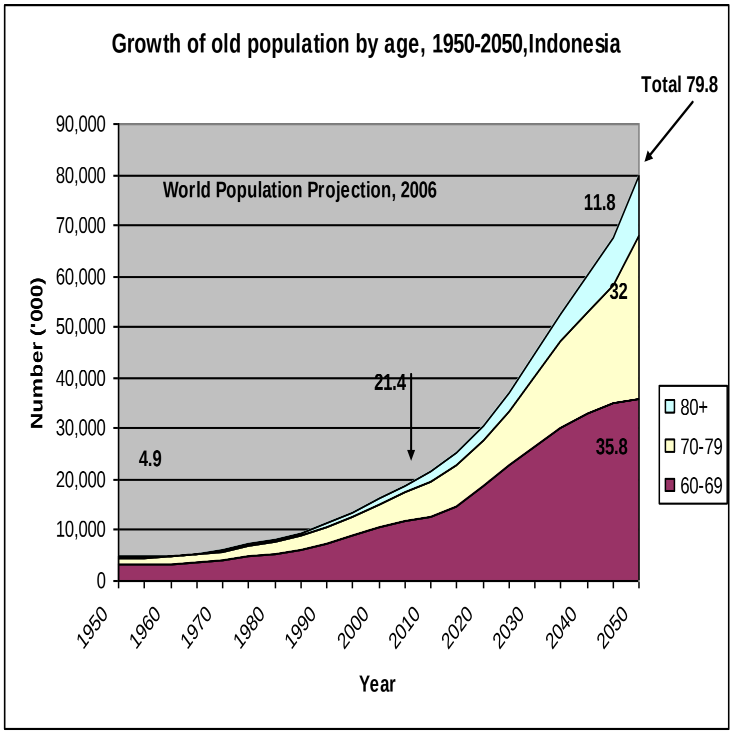


Figure 1. Growth of Indonesian Old Population by age, 1950 – 2050 Calculated by Adioetomo 2009. Source World Population Projection, 2006 revision

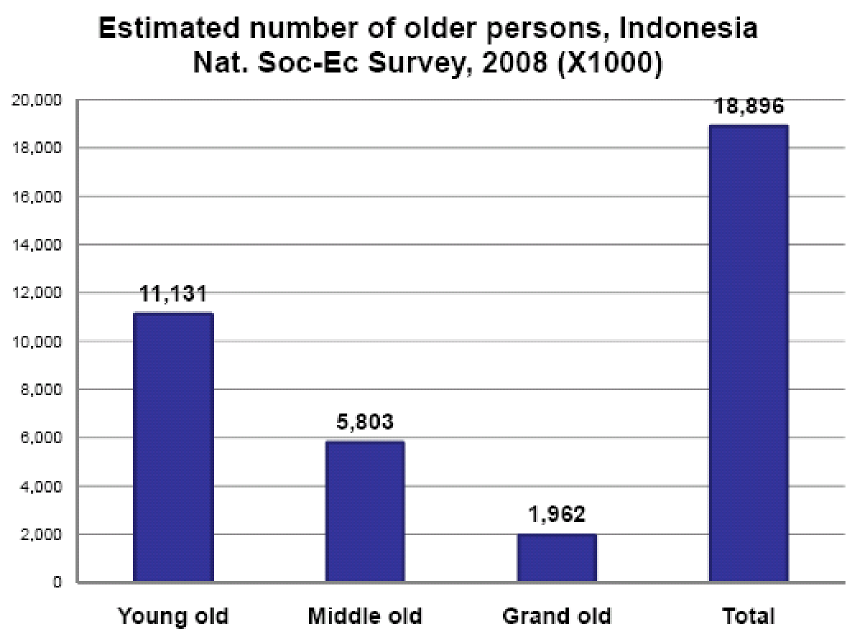


Figure 2. Estimated number of older persons in 2008. Source: National. Socio-Economic Survey 2008. Calculated by Adioetomo, 2009

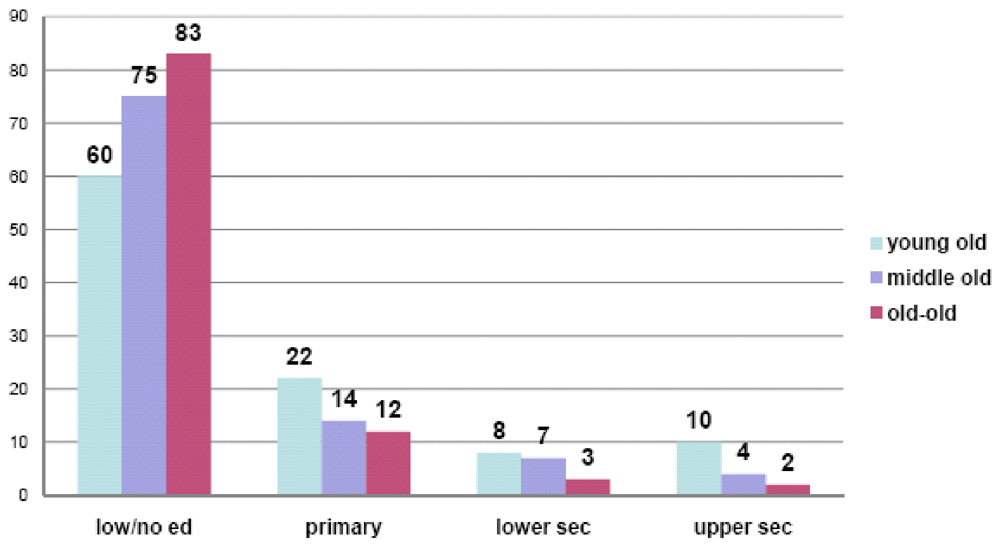


Figure 3. Older persons by education
Source: National. Socio-Economic Survey 2008. Calculated by Adioetomo, 2009

4. AGE FRIENDLY PRIMARY HEALTH CENTRES

The study of role of age friendly primary health centres and integrated community posts to support oldest old care giving in Indonesia is a part of a study to evaluate the role of Age Friendly PHC in Indonesia that was conducted in 2007 and also a part of study on Integrated Community Posts System that was conducted in 2008. Both studies conducted by Indonesia National Commission for Older Persons at 4 provinces, using a cross – sectional design in 2007 to evaluate the role of Age Friendly PHCs in Jakarta and then continued to evaluate the role of Integrated Community Posts Services in 2008. Quantitative data was collected using a check list instrument then was analyzed descriptively while qualitative method was employed at 4 provinces. Qualitative approach was also conducted by using Focus Group Discussion among Formal and Informal Leaders and in-depth interview among Elderly Patients. Socio-economic and health of older persons also studied by employing the cross sectional survey in 33 provinces in Indonesia.

Results of this study showed that both type of PHC have implemented Age Friendly Health Care Policy in which Integrated Community Posts have been used to provide older persons health services in the community. The PHCs had already implemented activities such as information and education (83%), early detection (85%); physical exercise (100%); recreation (95%); medication (95%) and rehabilitation (30%) as shown in figure 5. However, only 50% of health providers and care givers have been trained. It was shown that most of the leaders still need support of the government to improve PHCs quality regarding age friendly services, particularly in conducting training for the doctors and nurses in geriatrics and caring for dementia patients. The patients were satisfied with the services and activities provided by the PHCs, however the improvement of home care is still needed. Furthermore, for living arrangement, most people prefer to stay in their own houses with familiar environment. It means that the living arrangement has followed government policy i.e. the Elderly Act No. 13, 1998. Therefore, the three generation under one roof that was commonly found, should be relevant.

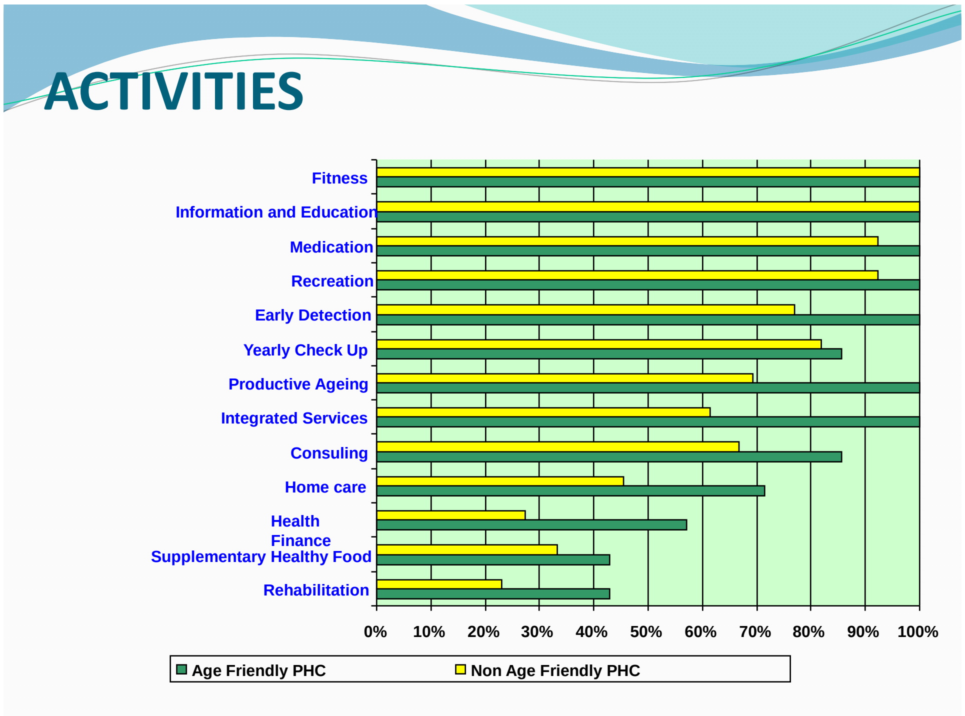


Figure 4. PHCs services. Source: Rahardjo et al, 2009

It was also found in this study that age friendly services, particularly in conducting training for the doctors, nurses and care givers in geriatrics and caring for dementia patients were still limited; supports of local governments were not in routine programme; older persons were satisfied with the services and activities provided by the PHCs and Integrated Community Posts, however, as mentioned earlier, the improvement of home care is still needed; and most of the older persons prefered to stay in their own houses with familiar environment . This is one of the traditional ways of living that support ageing in place. In this regard, home care will be very important, and the capability of family to care the older persons should be improved though collaboration with health providers from PHCs.

In term of health care management, there was a high level of awareness of health provider followed by planning and monitoring, inter-sectoral cooperation, base line data availability, cost subsidy, and elderly empowerment (figure 5).

Among 31 integrated community posts in 8 provinces, there were 90 people participants, consisted of head of integrated post, members and trainees. This study found that the average number of elderly people who actively involved in an integrated community post was approximately between 20 – 200 people. Results also showed that 65% of integrated community posts members were from the integrated community posts' own members while 35% was from the outside. It is also important to know that 75% of integrated community posts were formed based on a community agreement (*musyawarah warga*). One reason for this is because people in the community had has an awareness of the importance of the posts. In this regards, 70% of integrated community posts have already had a complete structure of its board which consisted of chairman, secretary, treasurer, trainee and sections. Again, from community development point of few, Indonesia has significant advantages through integrated posts regarding ageing in place support system.

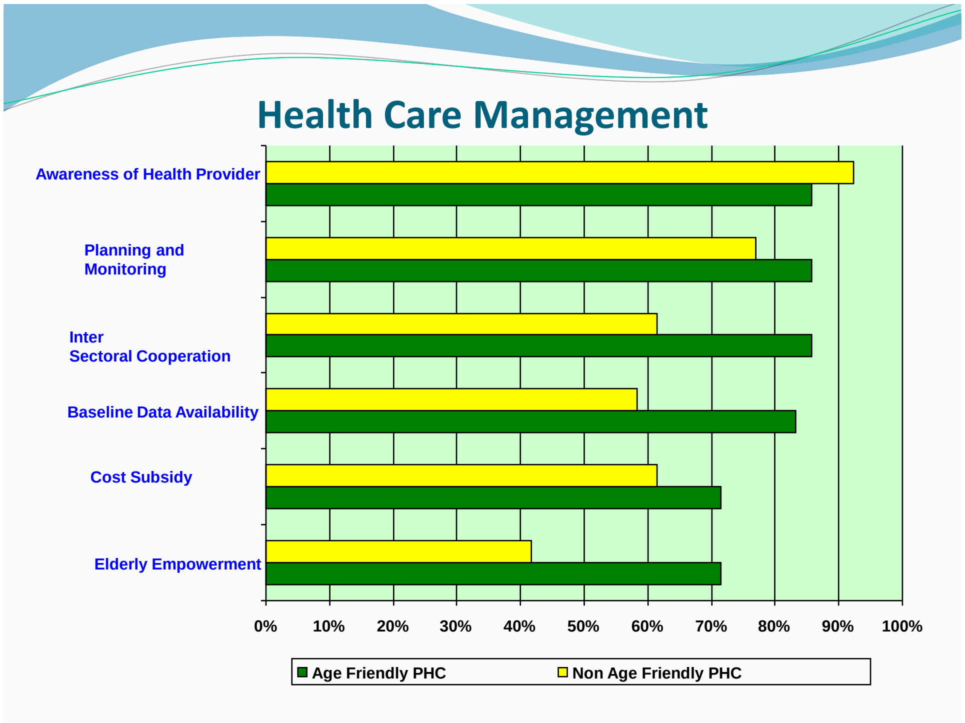


Figure 5. Health Care Management at PHCs. Source: Rahardjo et al, 2009

In relation to older persons activities, some respondents were still active in working, many of them were still doing their activities in or outside their houses. A lot of respondents still got involved in physical exercise (20%), religious activity (24%), health education (10%), medical check up (17%), and social gathering (8%). This figures are similar with national data, reported by Adioetomo, 2009.

What appears to be special in this study is that the role of government, especially local government, in supporting the integrated community posts. Local government played an important role to support actively the integrated community posts. Although this support from local government was not routine, local government was relatively supportive in term of providing funding for the integrated community posts, motivation, visiting, supporting the program and providing place for the activities.

5. THE IMPACT OF AGEING POPULATION IN AGEING PLACE THROUGH THE ROLE OF PHCS

Regarding ageing in place, beside of the numbers and proportions of elders, it is also pertinent to understand another aspect that related to the elders, such as household and living arrangement. The characteristic of household and living arrangement are as follows:

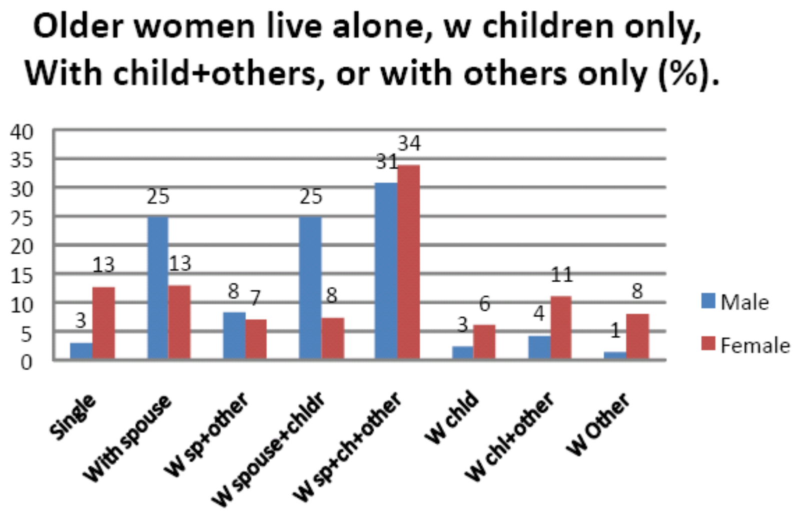


Figure 6. Living arrangement
Source: National. Socio-Economic Survey 2008. Calculated by Adioetomo, 2009

It seems that three generations under one roof was common (30%), balanced transfer but difference in lifestyle is potentials for conflict, however single households among the older persons are apparent in which the oldest, women and rural older persons live alone, and more older women than men live without spouse, instead with children or other persons. This is because changing family system, smaller family size, rural-urban-international migration, female labor participation, less and less caregiver's opportunity to develop home-care system, scarcity of land and housing, and declining job opportunity for the children. In this regard, family ideally should be a part of older person's health services as a team work with health providers of the PHCs (Rahardjo et.al, 2009). However, the awareness and capability of family to take care the older persons are relatively low. Therefore, training for PHCs' staff and family is highly needed, as suggested by Madrid International Plan of Action on Ageing (2002).

Concerning on above mentioned situations, the role of PHCs particularly in rural areas are very important to maintain older persons' health status. According to WHO (2002), most preventive health care and screening for early detection and management takes place in the primary health care setting at the community level. These primary health care (PHC) centres, to which people can self refer, also provide the bulk of ongoing management and care. It is estimated that 80% of front-line health care is provided at the community level where PHC centers form the backbone of the health care system. Similar situation was also found in Indonesia as our study revealed that elderly people has benefited from PHC centres especially for gathering information and education (83%), early detection (85%), physical exercise (100%), medication (95%) and rehabilitation (30%). In regards of PHCs as the backbone of the health care system, our study found that it was only 50% of health providers have been trained.

Older people already account for a sizeable proportion of PHC centre patients and as populations age and chronic disease rates climb, that proportion will increase. PHC centres are on the frontline of health care and are thus familiar to older people and their families. They are ideally positioned to provide the regular and extended contacts and ongoing care that older persons need to prevent or delay disability resulting from chronic health conditions (WHO, 2004). Therefore, it is important to make PHCs more comfortable. This condition has already existed gradually in Indonesia as Indonesian Ministry of Health targeted in 2010 that 50% of

PHCs should have age-friendly health services. This is also similar with a study that revealed 13 of 20 PHCs have already been age- friendly.

Despite the critical role that PHC centres play in older persons' health and well-being, older persons encounter many barriers to care (WHO, 2004). For example, our research found that it was relatively difficult to find transport to the centre, or the transportation was unavailable or too expensive. At PHC centres, the elderly got only a few minutes with a health care provider who did not have time to listen to all their concerns, missed critical warning signs, and did not have the geriatric related training to make the right diagnosis or prescribe the right treatment. Some elderly patients did not able to afford the medicines prescribed, did not understand why to take the prescribed medicine, and what side effects to report. This limitation has predicted by WHO (2002) that older patients may become discouraged from seeking or continuing treatment with potentially serious health consequences.

The World Health Organization (2004) has recognized the critical role PHC centres play in the health of older people worldwide and the need for these centres to be accessible and adapted to the needs of older populations. Based on this, by working with a series of discussion groups, we have gone to the source, asking older people and their providers to describe their benefits and barriers to care and their suggestions for change. These focus groups results that PHC did not only favor older people but also benefits all patients. This in line with the slogan of the UN International Year of Older Persons ("Towards a Society for All Ages". Furthermore, in Indonesia, our study revealed that the implementation of age friendly PHC policy has not been in optimal goal. To cope with oldest old care giving to improve their quality of life, the quality of age friendly PHC should be improved, in which the role of care givers will be highly needed, and Integrated Community Posts have important role. Although there are some barriers that revealed and needed to be improved, most PHCs' patients were satisfied with PHC services.

Finally, Biwas et al. (2009) stated that there is an important opportunity in the over all positioning of health care services to the older persons, due to the availability of health providers' commitment supported by the community through Integrated Community Posts. It also needs mutual coordination between family, community and PHCs. This should be supervised and monitored by local office Ministry of Health.

6. CONCLUSION

Indonesia population is ageing rapidly. Most of elderly in Indonesia, in this case, the rural elderly, especially women, seem to be more at a disadvantage, owing to poor quantity and quality of educational facilities in rural areas. This brings an impact on the health aspect so that the health care services at PHC are important. The implementation of age friendly PHC policy has been implemented, but has not been in optimal goal; the performance of Age Friendly PHs was better than Non Age Friendly PHCs; the facilities were still limited, the funding mostly from Local Goverment Sources; the role of Community Integrated Posts were significant; to cope with oldest old care giving programme to improve their quality of life , the quality of age friendly PHC should be improved, in which Integrated Community Posts have important role; training for Health Providers and Care Givers have to be improved. Furthermore, mutual coordination between family, community and PHCs will support the achievement of ageing in place programme.

ACKNOWLEDGEMENT

This study was supported by the National Commission for Older Persons, the Republic of Indonesia

REFERENCES

Adioetomo, Murtiningsih. The Demographic Outlook of Indonesia's Ageing Population and Social and Economic Profile. *Workshop on Ageing Readiness ESCAP and Demographic Institute Faculty of Economics University of Indonesia.* Jakarta, Indonesia, 7-8 October 2009

Biswas, P; Lloyd-Sherlock, P, Zaman, S. Understanding Elderly Client Satisfaction With Primary Health Care in Rural Bangladesh. *Asian J Gerontol Geriatr 2006;* 1: 133-140.

Handayani, Yvonne Suzy. "The Index of Disability and the Quality of Live of Older Persons in Jakarta". Dissertation of Doctorate Program, University of Indonesia, 2006.

Ministry of Social Affair the Republic of Indonesia. Elderly Act No. 13, 1998. Jakarta, 1999

Rahardo, Tri Budi W., Nugroho Abikusno, Yvonne Suzy Handayani, Isyulianto, Bambang Pujiantoro and Tanti Pratiwi. Older Persons Social Economic and Health Condition in Indonesia: Analysis of 33 Provinces in 2006 in Indonesia. *Indonesian Epidemiology Network Congress.* Semarang, Indonesia, 21 July 2007

Rahardjo, Tri Budi W., Toni Hartono, Vita Priantina Dewi, Eef Hogervorst and Evi Nurvidya Arifi. *Facing the Geriatric Wave in Indonesia: Financial Condition and Social Support* in Older Persons in Southeast Asia: An Emerging Asset. Evi Nurvidya Arifin and Aris Ananta (eds.). Institute of Southeast Asian Studies, Singapore. p. 270-298, 2009

Rahardjo, Tri Budi W., Asviretty, Toni Hartono, Inten Soeweno. The Role of Age Friendly Primary Health Centres (PHCs) and Integrated Community Posts to Support Oldest Old Care Giving inj Indonesia. *International Colloquium on the Oldest Old.* Centre for Gerontological Studies. Thiruvananthapuram, India, 9-11 February 2009

WHO (2002). *Active Ageing: A Policy Framework.* Geneva, World Health Organization (WHO/NMH/NPH/02.8;http//whqlibdoc.who.int/hq/2002/WHO_NMH_NPH_02.8.pdf, accessed 22 June 2004)

WHO (2002). *The World Health Report 2002: Reducing the Risks, Promoting healthy life.* Geneva, World Health Organization (http://whqlibdoc.who.int?whr?2002/WHR_2002.pdf, accessed 22 June 2004)

WHO (2003). *Adolescent Friendly Health Services: An Agenda for Change.* Geneva, WHO (WHO/FCH/CAH/02.14;http://whqlibdoc.who.int/hq/2003/WHO_FCH_CAH_02.14.pdf, accessed 22 June 2004)

WHO (2003). *Primary Health Care: A Framework for Future Strategic Directions.* Global Report.Geneva,WHO(WHO/MNC/OSD/03.01;http://www.who.int/chronic_conditions/primary_health_care/en/phc_report_oct03.pdf, accessed 22 June 2004)

WHO (2004). *Towards Age-Friendly Primary Health Care*. Geneva, WHO.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 80–80.

SESSION 2: FRONTIERS IN STATISTICS

Chair: Qi-Man Shao
Department of Mathematics
Hong Kong University of Science and technology
Clear Water Bay, Kowloon
Hong Kong, China
E-mail: maqmshao@ust.hk

NONPARAMETRIC TESTS FOR THE COEFFICIENT OF VARIATION

Emad-Eldin A. A. Aly
Department of Statistics & Operations Research, Faculty of Science
Kuwait University, P.O. Box 5969, Safat 13060, Kuwait
E-mail: emadeldinaly@yahoo.com

ABSTRACT

We consider two important problems concerning the coefficient of variation. The first problem is concerned with testing the null hypothesis of no change against the alternative of exactly one change point when the change is expressed in terms of the value of the coefficient of variation. In the second problem we consider testing the null hypothesis that $(k+1)$ populations $(k \geq 1)$ have the same unknown distribution against each of the alternatives: a) some populations have different coefficients of variation and b) the coefficients of variation are ordered. For both problems nonparametric test statistics are presented and their limiting distributions are given. The results of Monte Carlo studies are discussed.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 80–81.

MULTISTRATUM FRACTIONAL FACTORIAL DESIGNS

Ching-Shui Cheng
University of California at Berkeley
E-mail: cheng@stat.berkeley.edu

ABSTRACT

Multistratum experiments refer to those with multiple sources of errors. The error structure of an experiment is determined by the structure of the experimental units. Typically multiple strata

arise when some factors require larger experimental units than others, e.g., when their levels are more difficult to change. In this talk, I will present a general and unified approach to the selection and construction of such designs. Applications of the general theory will be illustrated on designs for blocked split-plots, blocked strip-plots and experiments with multiple processing stages.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 81–81.

ON THE JUMP ACTIVITY INDEX FOR SEMIMARTINGALES

Bing-Yi Jing, Xinbing Kong and Zhi Liu
Department of Math, HKUST, Clear Water Bay, Kowloon, Hong Kong
E-mail: majing@ust.hk

ABSTRACT

We present a new estimator of the jump activity index defined in Ait-Sahalia and Jacod (2009) for a discretely sampled semimartingale. The basic idea is to make use of the count of ``large" increments and the power variation of ``small" increments. Compared with that of Ait-Sahalia and Jacod (2009), our estimator has the same asymptotic properties yet has smaller variance since the new estimator effectively increases the number of the usable data. Simulations justify the improvement.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 82–96.

BIB DESIGNS WITH REPEATED BLOCKS: REVIEW AND PERSPECTIVES

Teresa Azinheira Oliveira
CEAUL and DCeT, Universidade Aberta, Rua Fernão Lopes nº9, 2ºdto, 1000-132 Lisboa, Portugal
E-mail: toliveir@univ-ab.pt

ABSTRACT

Experimental Design plays an important role on establishing an interface between Applied Mathematics and statistical applications in several fields, like Agriculture, Industry, Genetics, Biology and Education Sciences. The goal of any Experimental Design is to obtain the maximum amount of information for a given experimental effort, to allow comparisons between varieties and to control for sources of random variability. Randomized block designs are used to control for these sources. A Balanced Incomplete Block Design (BIB Design) is a randomized block design with number of varieties greater than the block size and with all pairs of varieties occurring equally often along the blocks. The Fisher related information of a balanced block design will remain invariant whether or not the design has repeated blocks. This fact can be used theoretically to build a large number of non-isomorphic designs for the same set of design parameters, which could be used for many different purposes both in experimentations and surveys from finite populations. The original and most important method on the construction of BIB Designs with repeated blocks (BIBDR) is due to Hedayat and Li (1979): the trade-off method. Since then, many authors and researchers have been paying particular attention to the construction of BIBDR, but still some unsolved problems remain. This issue will be briefly reviewed and new results on the existence and construction of BIBDR, as well as several unsolved problems for further research will be presented.

Keywords: Experimental Design, BIBD, BIBD with Repeated Blocks.

1. INTRODUCTION AND REMARKS CONNECTED WITH LITERATURE

An experimental design is a set of rules by which the varieties (treatments) to be used in the experiment are assigned to experimental units, so to produce valid results as efficiently as possible. The importance of incomplete block designs is very well known in experiments involving a big number of varieties and a large class of incomplete block designs consists of the so-called balanced incomplete block designs (BIB Designs).

Yates (1936a) formally introduced BIB Designs in agricultural experiments and since that time several challenging problems concerning the construction, non-existence and combinatorial properties of BIB Designs have been posed. A BIB Design is a binary incomplete block design for v varieties in b blocks of size k, so that each variety occur exactly r times along the blocks and every pair of varieties concur in exactly λ blocks. The five integers v,b,r,k,λ are the parameters of the BIBD, and they are not independent.

Applications of BIB Designs are known for long, not only in Agriculture but also for example in Genetics, Industry and Education Sciences. See Raghavarao et al. (1986), Oliveira and Sousa (2002), Gosh and Shrivastava (2001) and Yang (1985).

More recently, applications of BIB Designs are present in fields such as: biotechnology, for example on microarray experiments, Groβmann and Schwabe (2007); Feeding Consume Sciences, Wakeling and Buck (2001); Electronic Engineering, Wireless and Communications problems, Camarda, and Fiume, (2007), Boggia et al. (2009); Computer Science and Code Theory, along with connections with Cryptography, Chakrabarti (2006).

Raghavarao and Padgett (2005) presented an entire chapter devoted to BIB Designs applications. The authors explain the importance and connections of BIB Designs to the following statistical areas: Finite Sample Support and Controlled Sampling, Randomized Response Procedure, Balanced Incomplete Cross Validation, Group Testing, Fractional Plans, Box-Behnken Designs, Intercropping Experiments, Validation Studies, Tournament and Lotto Designs, and Balanced Half-Samples.

BIB Designs are optimal for a number of criteria under the usual homocedastic linear additive model and this optimality is not affected if the design admits block repetition. In fact, several reasons may lead to the consideration of block repetition in the design: the costs of the design implementation may differ using or not block repetition, the experimenter may consider that some treatment combinations are preferable to others or that some treatment combinations must be avoid, see Foody and Hedayat (1977) and Hedayat and Hwang (1984). Also many experiments are affected by the loss of observations, and even in extreme cases all the information contained in one or more blocks is lost. If there are no repeated blocks in the design some of the elementary treatment contrasts are no longer estimable. To avoid these situations, repetitions of blocks in the design are recommended, since it allow rely on the information contained in another block, with the same composition of the lost block and overcoming the missing data problem.

Another reason for using BIBDR in a big number of applications is that, besides the variance expression for comparisons for each design being the same, the number of comparisons of block effects with the same variance is different for BIB Designs considering or not block repetition. BIB designs with repeated blocks have some block contrasts with minimum variance, Ghosh and Shrivastava (2001), Oliveira et al. (2006) and Mandal et al. (2008).

The important role of BIB Designs with repeated blocks in Experimental Design and in Controlled Sampling contexts is explained in Foody and Hedayat (1977) and Wynn (1977).

Important details, namely from the point of view of existence conditions, construction, particular characteristics and applications of BIBDR, as well as some topics for further research can be found for example in Van Lint (1973), Foody and Hedayat (1977), Hedayat and Li (1979), Hedayat and Hwang (1984), Raghavarao et al. (1986), Hedayat et al. (1989), Gosh and Shrivastava (2001). Raghavarao et al. (1986) considered a linear model that will distinguish designs with different support sizes, and they observed the relevance of this model in intercropping experiments and market research.

Besides many authors pay special attention to BIB Designs with repeated blocks, this is a very challenging area since there are yet many research questions with no answer. A review on important known results, some examples and unsolved problems will be presented.

2. IMPORTANT CONCEPTS ON BIB DESIGNS WITH REPEATED BLOCKS

2.1 BIB Designs and the Existence Problem

A BIB Design with parameters v,b,r,k,λ is denoted by BIBD(v,b,r,k,λ), and the following are the necessary but not sufficient conditions for the existence of a BIBD:

$$vr = bk\ ;\ r(k-1) = \lambda(v-1)\ ;\ b \geq v\ .$$

The condition $b \geq v$ is known as Fisher inequality.

For cases k=3 and k=4 to all λ and for k=5 and λ=1,4,20 Hanani (1961) proved that conditions $\lambda(v-1) \equiv 0(mod(k-1))$ and $\lambda v(v-1) \equiv 0(mod\, k(k-1))$ are sufficient for the existence of a BIB Design.

In some particular cases, even when a set (v,k,λ) satisfies the necessary conditions for the existence of a BIB Design, the actual design does not exist or its existence may be unknown. For these situations Hedayat et al. (1995) introduced two classes of designs: Contingently Balanced Incomplete Block (C-BIB) Designs and Virtually Balanced Incomplete Block (V-BIB) Designs. The authors show that both of these classes can be constructed by a sequential search algorithm.

2.2 The Existence of BIB Designs with Repeated Blocks

Consider a particular BIB design, and let B be a specific block randomly selected in this design. Let x_i, i = 0, 1, ..., k be the number of blocks apart from B itself, which have exactly i varieties in common with B.

It is known that for BIB designs with repeated blocks the following conditions hold:

$$\sum_{i=0}^{k}\binom{i}{0}x_i = b-1\ ; \quad \sum_{i=1}^{k}\binom{i}{1}x_i = k(r-1)\ ; \quad \sum_{i=2}^{k}\binom{i}{2}x_i = \binom{k}{2}(\lambda-1)$$

In Sousa and Oliveira (2004) these conditions are developed to obtain a bound to the number of blocks, so that the design admits block repetition. Subtracting the third expression to the second and after some algebraic operations we have:

$$x_0 + \sum_{i=2}^{k}\frac{ix_i(i-3)}{2} + k(r-1) - \frac{k(k-1)(\lambda-1)}{2} + \sum_{i=2}^{k}x_i = b-1$$

Solving in order to b it becomes:

$$\text{(i)} \quad b = x_0 + \sum_{i=3}^{k}x_i\left[\ \frac{(i-1)(i-2)}{2}\ \right] + 1 + k(r-1) - \frac{k(k-1)(\lambda-1)}{2}$$

and since $x_0+\sum_{i=3}^{k}x_i[\ \frac{(i-1)(i-2)}{2}\ ]\geq 0$, then: $b\geq 1+k(r-1)-k(k-1)(\lambda-1)/2$

Developing (i):

$$x_0+\sum_{i=3}^{k-1}x_i[\ \frac{(i-1)(i-2)}{2}\ ]=b-1-k(r-1)+\frac{k(k-1)(\lambda-1)}{2}-\frac{(k-1)(k-2)}{2}x_k$$

then: $x_k\leq\frac{2b-2-2k(r-1)+k(k-1)(\lambda-1)}{(k-1)(k-2)}$

Since x_k denotes the number of blocks identical to block B in the design, then the design just admits repeated blocks if there is at least one block in the design identical to block B, which means, if $x_k\geq 1$. So, to $k\geq 3$:

$$1\leq x_k\leq\frac{2b-2-2k(r-1)+k(k-1)(\lambda-1)}{(k-1)(k-2)}$$

And, solving last inequality in order to b:

$$\text{(ii)}\qquad b\geq\frac{(k-1)(k-2)}{2}+1+k(r-1)-\frac{k(k-1)(\lambda-1)}{2}$$

If there exists a BIBD obeying the parameters (v,b,r,k,λ) then the number of blocks, b, should obey inequality (ii) so that the design admits block repetition.

2.3 Classification into Families

Hedayat and Hwang (1984) divided the collection of all BIB designs with $\lambda\geq 2$ into three mutually exclusive and exhaustive families:

Family 1 consists of all BIBD(v,b,r,k,λ) whose parameters (b,r,λ) have a common integer divisor, t>1, and there exists one or more BIB($v,b/t,r/t,k,\lambda/t$) designs.

Family 2 consists of all BIBD(v,b,r,k,λ) whose parameters (b,r,λ) have one or more common integer divisors greater than one but there is no BIB($v,b/t,r/t,k,\lambda/t$) design if t>1 is one of the divisors of b, r, and λ. No member of this family can be obtained by collecting or taking copies of smaller size BIB designs of type BIBD($v,b/t,r/t,k,\lambda/t$).

Family 3 consists of all BIBD(v,b,r,k,λ) whose parameters (b,r,λ) are relatively prime, thus the parameters v,b,r,k,λ for the members of this family are such that the great integer divisor of (b,r,λ) is one. As in family two, in family three no member can be obtained by collecting or taking copies of smaller size BIB designs of type BIBD($v,b/t,r/t,k,\lambda/t$).

2.4 Results on Possible Parameters for BIBDR Existence (K=3,4,5,6)

Considering the particular cases to k=3,4,5,6 and inequality (ii), we have, respectively the inequalities:

$$\lambda\left[(\nu-5)^2+2\right] \geq 2$$

$$\lambda\left[(\nu-8)^2-\nu+24\right] \geq 72$$

$$\lambda\left[(\nu-13)^2+56\right] \geq 240$$

$$\lambda\left[(\nu-18)^2-\nu+162\right] \geq 600$$

Parameters to possible BIBDR to particular cases k=3, 4, 5, 6, considering $5\leq\nu\leq30$, $\lambda\leq10$ and $b\leq200$, were obtained, by implementing an algorithm to obey known BIBDR existence conditions.

Considering the integer t, $t \geq 2$, if the BIBD with parameters presented in Tables 1,2,3,4 exits, then it allows block repetition. These Tables also present the design classification into families.

2.5 Blocks Multiplicity and Bounds

If in the BIB Design there are less than b distinct blocks then the design has repeated blocks. The set of all distinct blocks in a BIB Design is called the support of the design, and the design cardinality is represented by b^*. The notation BIBD($\nu,b,r,k,\lambda|b^*$) is used to denote a BIBD(ν,b,r,k,λ) with precisely b^* distinct blocks.

Considering BIBDR it seems important to find the answer to some fundamental questions, like for example which blocks and how many times should it be repeated in the design, and what is the minimum admissible value to b* so that the design admits block repetition. Some more concepts are then crucial.

Table 1 Parameters to possible BIBDR, block size 3, and classification into families

v	*b*	*r*	*k*	*λ*	Family		*v*	*b*	*r*	*k*	*λ*	Family
5	10t	6t	3	3t	1		15	35t	7t	3	t	1
6	10t	5t	3	2t	1		16	80	15	3	2	3
7	7t	3t	3	t	1		16	80t	15t	3	2t	1
8	56	21	3	6	3		17	136	24	3	3	3
8	56t	21t	3	6t	1		17	136t	24t	3	3t	1
9	**12t**	**4t**	**3**	**t**	**1***		18	102	17	3	2	3
10	30	9	3	2	3		18	102t	17t	3	2t	1
10	30t	9t	3	2t	1		19	57t	9t	3	t	1
11	55	15	3	3	3		21	70t	10t	3	t	1
11	55t	15t	3	3t	1		22	154	21	3	2	3
12	44	11	3	2	3		22	154t	21t	3	2t	1
12	44t	11t	3	2t	1		24	184	23	3	2	3
13	26t	6t	3	t	1		24	184t	23t	3	2t	1
14	182	39	3	6	3		25	100t	12t	3	t	1
14	182t	39t	3	6t	1							

Table 2 Parameters to possible BIBDR, block size 4, and classification into families

v	*b*	*r*	*k*	*λ*	Family		*v*	*b*	*r*	*k*	*λ*	Family
5	5t	4t	4	3t	1		16	20t	5t	4	t	1
6	15	10	4	6	3		17	68	16	4	3	3
6	15t	10t	4	6t	1		17	68t	16t	4	3t	1
7	7t	4t	4	2t	1		18	153	34	4	6	3
8	14t	7t	4	3t	1		18	153t	34t	4	6t	1
9	18t	8t	4	3t	1		19	57	12	4	2	3
10	15t	6t	4	2t	1		19	57t	12t	4	2t	1
11	55	20	4	6	3		20	95	19	4	3	3
11	55t	20t	4	6t	1		20	95t	19t	4	3t	1
12	33	11	4	3	3		21	105	20	4	3	3
12	33t	11t	4	3t	1		21	105t	20t	4	3t	1
13	**13t**	**4t**	**4**	**t**	**1***		22	77	14	4	2	3
14	91	26	4	6	3		22	77t	14t	4	2t	1
14	91t	26t	4	6t	1		24	138	23	4	3	3
15	105	28	4	6	3		25	50t	8t	4	t	1
15	105t	28t	4	6t	1		28	63t	9t	4	t	1

Table 3 Parameters to possible BIBDR, block size 5, and classification into families

v	b	r	k	λ	Family		v	b	r	k	λ	Family
6	6t	5t	5	4t	1		17	68	20	5	5	3
7	21	15	5	10	3		17	68t	20t	5	5t	1
7	21t	15t	5	10t	1		19	171	45	5	10	3
9	18	10	5	5	3		19	171t	45t	5	10t	1
9	18t	10t	5	5t	1		20	76	19	5	4	3
10	18	9	5	4	3		20	76t	19t	5	4t	1
10	18t	9t	5	4t	1		21	21t	5t	5	t	1
11	11t	5t	5	2t	1		25	30t	6t	5	t	1
13	39	15	5	5	3		26	130	25	5	4	3
13	39t	15t	5	5t	1		26	130t	25t	5	4t	1
15	21t	7t	5	2t	2*		30	174	29	5	4	3
16	48	15	5	4	3		30	174t	29t	5	4t	1
16	48t	15t	5	4t	1							

* There is no BIBD(15,21,7,5,2), so BIBDR(15,42,14,5,4) belongs to family 2, according to Hedayat and Hwang (1984).

Table 4 Parameters to possible BIBDR, block size 6, and classification into families

v	b	r	k	λ	Family		v	b	r	k	λ	Family
7	7t	6t	6	5t	1		21	14p	4p	6	p	1
9	**12t**	**8t**	**6**	**5t**	**1***		22	77	21	6	5	3
10	15t	9t	6	5t	1		22	77t	21t	6	5t	1
11	11t	6t	6	3t	1		24	92	23	6	5	3
12	22	11	6	5	3		24	92t	23t	6	5t	1
12	22t	11t	6	5t	1		25	100	24	6	5	3
13	26	12	6	5	3		25	100t	24t	6	5t	1
13	26t	12t	6	5t	1		26	65	15	6	3	3
15	35	14	6	5	3		26	65t	15t	6	3t	1
15	35t	14t	6	5t	1		27	117	26	6	5	3
16	16t	6t	6	2t	1		27	117t	26t	6	5t	1
18	51	17	6	5	3		28	126	27	6	5	3
18	51t	17t	6	5t	1		28	126t	27t	6	5t	1
19	57	18	6	5	3		30	145	29	6	5	3
19	57t	18t	6	5t	1		30	145t	29t	6	5t	1

The multiplicity of a block is the number of times the block occurs in the design. If in a BIB Design there are exactly α blocks with multiplicity i, exactly β blocks with multiplicity j,…, and all other blocks with multiplicity 1, then design has multiplicity pattern $\alpha^j \beta^j$.

If in a BIB Design with parameters (v,b,r,k,λ) each block occurs exactly m times ($m>1$), then the design is called a multiple block design of multiplicity m and is denoted by M-BIBD (v,b,r,k,λ).

Let s represents the number of identical blocks in a design. The conditions $\lambda \geq 2$ and $s \leq \lambda$ are necessary for the existence of a design with $b^*<b$. For a BIBD (v,b,r,k,λ) with s identical blocks and $r>\lambda$, Mann (1969) proved that $\frac{r}{k}=\frac{b}{v} \geq s$. This inequality covers the Fisher's inequality (case s=1). Van Lint and Ryser (1972) proved that the support size b* of a BIB Design satisfies $b^*=v$ or $b^*>v+1$. Foody and Hedayat (1977) established new limits to b^*, namely they reach the expression $b^*_{min} \geq \frac{b}{\lambda}$ where b^*_{min} represents the minimal support size to a BIBD based on v and k. Constantine and Hedayat (1983) present the construction of complete designs with blocks of maximal multiplicity and set its relevance in simple random sampling context.

For particular cases with $b<\lambda v$ and using the inequality of Mann (1969), Hedayat et al. (1989) obtained a more restrictive bound to b^*_{min} by considering the information about those blocks in the support that are repeated λ times in the design:

$$b^*_{min} \geq 2\frac{v(v-1)}{k(k-1)}$$

2.6 The Variance for Block Effect Contrasts

Consider a BIBDR$(v,\ b,\ r,\ k,\ \lambda|b^*)$ with the incidence matrix $N=(n_{ij})$, where n_{ij} is the number of times that the ith variety occur in the jth block, $i=1,2,\ldots,v,\ j=1,2,\ldots,b$. For this design the coefficient matrix D for estimating a vector of block effects $\beta=(\beta_1,\beta_2,\ldots,\beta_b)'$ has the form $D=kI_b-\frac{1}{r}N'N$, where I_b denote $b \times b$ identity matrix.

Consider any two blocks of the design, B_j and $B_{j'}$, which have h varieties in common. Then the (j,j')th element of the matrix $N'N$ is equal to h and the variance of the difference of block effects is given by:

$$Var(\hat{\beta}_j-\hat{\beta}_{j'})=2\sigma^2(v\lambda+k-h)/vk\lambda .$$

Since h can take at most $k+1$ values, namely $0,1,\ldots,k$, so there will be at most $k+1$ possible variances for the estimated elementary block effect contrasts.

Note that the variance of estimated contrasts of block effects tends to the minimum value as the number of common treatments between the two blocks increases. For BIBDR we can have some block contrasts with minimum variance and this is one of the reasons for using it in a big number of applications, Oliveira et al. (2006).

3. BRIEF REVISION ON THE CONSTRUCTION METHODS FOR BIB DESIGNS AND BIBDR

Bose (1939) introduced the method of cyclical development of initial blocks, which allows the construction of most of the existing BIB Designs. Since then, several methods have been developed for the construction of BIB Designs. The most important ones are presented in Hinkelmann and Kempthorne (2005). These authors explain some Difference Methods, such as Cyclic Development of Difference Sets, the Method of Symmetrically Repeated Differences and the Formulation in Terms of Galois Field Theory and present some other methods using particular cases like irreducible BIB Designs, complement of BIB Designs, residual BIB Designs and Orthogonal Series. Also in Raghavarao and Padgett (2005) the construction of BIB Designs from Finite Geometries and by the Method of Differences is presented.

Raghavarao (1971) presents the complete list of existing BIB Designs obeying the conditions v, $b\leq 100$ and r, $k \leq 15$. Hinkelmann and Kempthorne (2005) present a list of BIB Designs with $v \leq 25$, $k\leq 11$. Julian et al. (2006), Chapter 3, present the existence results for BIB Designs with small block sizes, $k\leq 9$, and point some values of v for which the existence of a BIB Design remains undecided.

Some construction methods have been also developed for the particular BIB Designs with repeated blocks. Van Lint and Ryser (1972) studied this problem and their basic interest was in constructing BIBD (v, b, r, k, λ) with repeated blocks such that parameters b, r, λ were relatively prime. Foody and Hedayat (1977) showed that the combinatorial problem of searching BIBDR was equivalent to the algebraic problem of finding solutions to a set of homogeneous linear equations and presented the construction of BIB designs with $v=8$ and $k=3$ with support sizes 22 to 55.

Hedayat and Li (1979) presented one of the still most important methods for the construction of BIB Designs with variable support sizes: the trade-off method. Using this method Hedayat and Hwang (1984) presented the construction of BIBDR(8,56,21,3,6) and BIBDR(10,30,9,3,2). The basic idea is to trade some blocks with some other blocks, without losing the BIB general characteristics of the design. The resulting design may differ from the original in what concerns the support size.

Hedayat and Hwang (1984) call the following set of blocks a (v,3) trade of volume 4 for any $v\geq 6$, where 6 varieties (x, y, z, u, v, w) appear in each set of four blocks:

I	II
xyz	*uvw*
xuv	*yzw*
yuw	*xzv*
zvw	*xyu*

If in a design we replace blocks in set I by blocks in set II we do not loose BIB property. We will use this method in our examples.

Trades properties are very important. In Handbook of Combinatorial Designs (2006), Chapter 60, by Hedayat and Khosrovshah, is entirely devoted to Trades and related references.

Several authors have been paying particular attention to the construction and properties of *M*-BIB(7,7,3,3,1) Designs. Raghavarao et al. (1986) presented characteristics for distinguishing among balanced incomplete block designs with repeated blocks and obtained a class of BIB Designs β(7,21,9,3,3) where nine BIB Designs out of ten have repeated blocks. In Gosh and Shrivastava (2001) a method for the BIBDR construction, based on a composition of two semi-regular group divisible designs and respective parameters was presented. The authors present a class of BIB designs with repeated blocks and the example of BIBDR (7,28,12,3,4) construction, as well as a table with the respective multiplicity of occurrence of variance for the estimated elementary contrast of block effects. These authors include the comparison of fifteen BIBD with the same parameters, according to the number of distinct blocks and to the multiplicities of variance of elementary contrasts of the block effect. Fourteen of these designs have repeated blocks. Mandal et al. (2008) present a complete class of BIB Designs β(7,35,15,3,5), with thirty one BIB Designs, where one of them has no block repetition and the remaining thirty are BIB Designs with repeated blocks.

Oliveira and Sousa (2004) presented examples of BIBDR(12,44,11,3,2) with different structures for the same cardinality and Sousa and Oliveira (2004) presented the cardinality analysis for BIBDR (9,24,8,3,2). In the Handbook of Combinatorial Designs (2006), Manthon, R. and Rosa, A. chapter 1, results on BIBD(v,b,r,k,λ) of small order are presented, as well as a list of nonisomorphic such designs for particular values of v. Also a table with admissible parameter sets of nontrivial BIB Designs with $r \leq 41$ and $k \leq v/2$ is presented, as well as references for earlier listings of BIB Designs.

4. BIBDR STRUCTURES: SOME EXAMPLES

We will present some examples of BIBDR with different parameters and also some BIBDR with the same parameters and different support sizes. Finally for BIBDR with the same support size we will present BIB Designs with different structures for repeated blocks.

In example 1 we will consider BIBDR(13,26,8,4,2) from Table II and we present one possible structure, the most trivial one, with all blocks repeated once, b*=13. In example 2 we will consider BIBDR (9,24,8,3,2) from Table I, and we will use the trade-off method to obtain designs with two different cardinalities, b*=18 and b*=20. Finally we will present two possible structures for repeated blocks in BIBDR with the same cardinality .

Example 1

The most trivial way of getting a design from an existing one is simply by repeating the blocks in the original design. According to Calinski and Kageyama (2003) this is so-called juxtaposition method. Consider the varieties 1, …, 9, *x*, *y*, *z*, *w*. The representation for BIBD (13,13,4,4,1) structure can be:

BIBD(13,13,4,4,1)

1234	1567	189*x*	1*yzw*
258*y*	269*w*	27*xw*	
359*w*	36*xy*	378*z*	
45*xz*	468*w*	479*y*	

We can easily obtain the structure for BIBDR(13,26,8,4,2|b*=13), considering block multiplicity equal 2 and duplicating BIB(13,13,4,4,1). When we wish to compare any two blocks of the design we have $\binom{26}{2}=325$ possible comparisons, and to any two blocks which have four common varieties we have contrasts with minimum variance.

Example 2

In this example we present BIBDR (9,24,8,3,2) with different cardinalities, b*=18, b*=20 and b*=21. We illustrate the trade-off method application to obtain a possible structure to BIBDR (9,24,8,3,2|*b*=20*) by using BIBDR (*9,24,8,3,2|b*=18*). We present two different structures for repeated blocks considering the BIBDR (9,24,8,3,2|*b*=21*).

Design with 6 repeated blocks **Design with 4 repeated blocks**

BIBDR(9,24,8,3,2|*b*=18*)			TRADE	BIBDR(9,24,8,3,2|*b*=20*)		
123	247	**359**		**123**	247	358
123	249	**359**		**123**	248	359
147	257	**367**	348→259	147	257	**367**
149	258	**367**	359→248	149	259	**367**
157	268	**456**	249→358	157	268	**456**
158	269	**456**	258→349	158	269	**456**
168	**348**	**789**		168	348	**789**
169	**348**	**789**		169	349	**789**

Example 3

In this example we will consider the definition of complementary design to illustrate the construction of BIBDR (*9,24,16,6,10|b*=21*), and two different structures for repeated blocks with the same cardinality. We will use results of example 2).

Design with 3 repeated blocks BIBDR (*9,24,8,3,2|b*=21*)

Structure 1 **Structure 2**

123	248	357		**123**	**246**	357
123	249	358		**123**	**246**	358
147	257	368		**145**	257	367
148	259	369		**145**	259	369
158	267	**456**		167	278	478
159	268	**456**		168	289	479
167	347	**789**		179	348	568
169	349	**789**		189	349	569

Definition of complementary design is very important, particularly for cases with large k size. The complement of a BIBD(v, b, r, k ,λ) is also a BIB design with b blocks each of size v-k, so the existence of a BIBD with block size k implies the existence of another BIBD with block size v-k.

A simple technique for constructing one BIB Design from another is to replace the varieties in each block by the set of varieties that are not in the block, known as the complementary set of varieties.

To k=6 and using complementary designs of BIBDR (*9,24,8,3,2*) presented in example 2, examples follows to BIBDR (*9,24,16,6,10|b*=21*) structure.
Consider the varieties 1,…,9, and block size equal to six.

BIBDR (9,24,16,6,10|*b*=21*)- Structure 1

456789	135679	124689
456789	135678	124679
235689	134689	124579
235679	134678	124578
234679	134589	**123789**
234678	134579	**123789**
234589	125689	**123456**
234578	125678	**123456**

Structure 1: We have three blocks repeated twice and18 non-repeated blocks. Any of the nine varieties appear twice in repeated blocks;

BIBDR (*9,24,16,6,10|b*=21*)- Structure 2:

456789	**135789**	124689
456789	**135789**	124679
236789	134689	124589
236789	134678	124578
234589	134569	123569
234579	134567	123468
234568	125679	123479
234567	125678	123478

Structure 2: We have three blocks repeated twice and18 non-repeated blocks. Varieties 1,2,4 appear once in repeated blocks, varieties 5,3,6 appear twice in repeated blocks and varieties 7,8,9 appear three times in repeated blocks.

5. CONSIDERATIONS AND PERSPECTIVE RESEARCH

The construction of BIB Designs for some particular parameters combinations still remain unsolved, and so are the questions connected with the respective BIBDR. This is an important issue to develop, as well as the necessary or sufficient conditions for the existence of a BIBDR and given parameters. Bounds for the multiplicity of a block in such a BIBD need also more investigation, since for many BIBDR there are still open questions.

In further research, besides continuing the cardinality analysis for particular BIBDR, it seems also important to investigate the relations between the number of possible structures and the design parameters, considering BIBDR with the same cardinality.

Applications of BIB Designs and BIBDR are known in many diverse fields, but there are yet many others to explore and improve, using this area of knowledge as a new challenge.

REFERENCES

Boggia G., Camarda, P., Cormio, C., and Grieco, L.A. (2009): *Iranian Journal of Electrical & Electronic Engineering, Vol. 5, No. 2*

Bose, R.C. (1939): On the construction of balanced incomplete block designs, Annals of Eugenics. 9 , pp. 358–399.

Calinski, T. and Kageyama, S. (2003): *Block Design: A Randomization Approach,* Volume II: Design, Lecture Notes in Statistics, Springer.

Camarda, P. and Fiume, O. (2007): Collision Free MAC Protocols for Wireless Ad Hoc Networks based on BIBD Architecture. Journal of Communications, Vol. 2, number 7, pp 1-8.

Chakrabarti (2006): A new visual cryptography scheme for STS based Access structures. Journal of Discrete Mathematical Sciences & Cryptography, Vol.9, n 1, pp9-23.

Colbourn and Dinitz (2006): *The CRC Handbook of Combinatorial Designs, CRC Press. Chapman & Hall.*

Constantine, C. M. and Hedayat, A. (1983): *Complete designs with blocks of maximal multiplicity.* Journal of Statistical Planning and Inference, 7, 289-294.

Foody, W, and Hedayat, A. (1977): *On theory and applications of BIB Designs with repeated blocks*. The Annals of Statistics, Vol 5, nº5, 932 - 945.

Großmann, H. and Schwabe, R. (2007): The relationship between optimal designs for microarray and paired comparison experiments. Otto–von–Guericke–Universität Magdeburg, Fakultät für Mathematik, Preprint Nr. 23 / 2007.

Gosh and Shrivastava (2001): *A class of BIB designs with repeated blocks*. Journal of Applied Statistics, Vol.28, nº7, 821-833.

Hanani, H. (1961): *The existence and construction of Balanced Incomplete Block Designs*. The Ann. Math. Statist., Vol 32, 361 - 386.

Hedayat, A. S., and Hwang, H. L. (1984): *BIB(8,56,21,3,6) and BIB(10,30,9,3,2) Designs with Repeated Blocks*. Journal of Combinatorial Theory, Series A 36 , 73 - 91.

Hedayat, A.S., Landgev, I.N. and Tonchev, V.D. (1989): *Results on the support of BIB designs*. Journal of Statistical Planning and Inference, 22, 295-306.

Hedayat, A, and Shuo-Yen R. Li (1979): *The trade off method in the construction of BIB Designs with repeated blocks*. The Annals of Statistics , Vol.7, nº6, 1277- 1287.

Hedayat, A, Stufken, J. and Zhang, W.G. (1995): *Contingently and Virtually Balanced Incomplete Block Designs and their efficiencies under various optimality criteria*. Statistica Sinica 5, 575-591.

Hinkelmann, K. and Kempthorne, O. (2005): *Design and Analysis of Experiments, Advanced Experimental Design*, Vol.2. Wiley Series in Probability and Statistics, Wiley-Interscience.

Mandal, S., Ghosh, D.K., Sharma, R.K. and Bagui, S.C. (2008): *A complete class of balanced incomplete block designs β(7,35,15,3,5) with repeated blocks.* Statistics and Probability Letters , Vol 78, Issue 18, 3338-3343.

Mann, H.B. (1969): *A note on balanced incomplete block designs*. The Annals of Math. Statist., Vol.40, nº2, 679-680.

Oliveira, T.A., Sousa, M.F. (2002): *Some correspondence between Diallel Crosses and BIB Designs.* Proceedings of the 17th IWSM, 515-520.

Oliveira, T.A., Sousa, M.F. (2004): *BIBD with repeated block: some analyses on BIBD(12,44,11,3,2) with different structures.* Colloquium Biometryczne, 34a,131-140.

Oliveira, T.A., Ceranka, B., Graczyk, M. (2006): *The variance of the difference of block effects in the balanced incomplete block designs with repeated blocks*, Colloquium Biometryczne, 36, 2006, 115-124.

Raghavarao, D. (1971): *Constructions and Combinatorial Problems in Design of Experiments*, John Wiley and Sons, New York.

Raghavarao, D., Federer, W.T. and Schwager, S.J. (1986): *Characteristics for distinguish among balanced incomplete block designs with repeated blocks*. Journal of Statistical Planning and Inference 13, 151-163.

Raghavarao, D. and Padgett, L.V. (2005): *Block Designs: Analysis, Combinatorics and Applications*. Series on Applied Mathematics, World Scientific Publishing Co.

Sousa, M.F., Oliveira, T.A. (2004): *BIBDR: Some analysis on BIBD(*9,24,8,3,2*) cardinalities*. Colloquium Biometryczne, 34a,161-170.

Van Lint, J.H. (1973): *Block Designs with repeated blocks and* $(b,r,\lambda)=1$. J. Combinatorial Theory Ser A 15, 288 - 309.

Van Lint, J.H. and Ryser , H.J. (1972): *Block Designs with repeated blocks*. Discrete Math., Vol 3, 381 - 396.

Wynn, H.P. (1977): *Convex sets of finite plans*. Ann. Statist.,Vol5,nº2, 414-418.

Wakeling, I.N. and Buck, D. (2001): *Balanced incomplete block designs useful for consumer experimentation*. Food Quality and Preference, 12, 265-268.

Yang, J. (1985): *An Example of the application of a Balanced Incomplete Block Design.* Assessment & Evaluation in Higher Education, 1469-297X, Volume 10, Issue 3, 250 – 253

Yates, F. (1936a). Incomplete randomized blocks, Ann. Eugen. 7, 121–140.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 97–97.

TESTS FOR INDEPENDENCE IN HIGH DIMENSION

Qi-Man Shao
Department of Mathematics
Hong Kong University of Science and technology
Clear Water Bay, Kowloon
Hong Kong, China
E-mail: maqmshao@ust.hk

ABSTRACT

Let $X_1, X_2, \ldots, X_n$ be a random sample from a *p*-dimensional population distribution. Assume that both *p* and *n* are large. In this talk we introduce several test statistics to test the independence of *p*-variates and study the asymptotic properties.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 97–97.

CARRY-OVER EFFECTS WHEN USING CROSSOVER DESIGNS

John Stufken
Professor and Head, University of Georgia
E-mail: jstufken@uga.edu

ABSTRACT

One of the most controversial topics when using a crossover design is how to deal with possible carryover effects. While there is broad agreement that one should attempt to use sufficiently long washout periods, this may not always be possible and it is not always clear what 'sufficiently long' really means. A growing number of researchers recommend that a crossover design should not be used if there is a possibility of carryover effects, with or without a washout period, that persist into the next active treatment period. Others have proposed more realistic models for modeling possible carryover effects. This talk will provide a closer look into these issues and review some of the proposals for alternative models. The talk will also consider efficient designs for different scenarios.

The talk is based on joint work with various collaborators during the last years, including Samad Hedayat (U of Illinois at Chicago), Joachim Kunert (U of Dortmund, Germany), and Min Yang (U of Missouri).

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 98–105.

SESSION 3: A SYMPOSIUM ON MEDICAL META-ANALYSIS

Chair: Shahjahan Khan
Department of Mathematics and Computing, Australian Centre for
Sustainable Catchments, University of Southern Queensland, Toowoomba, Queensland,
Australia
E-mail: khans@usq.edu.au

PRELIMINARY RESULTS OF META-ANALYSIS OF ENDOSCOPIC RETROGRADE CHOLANGIOPANCREATOGRAPHY (ERCP) VERSUS CONSERVATIVE TREATMENT FOR GALL STONE PANCREATITIS

Matthew John Burstow
Department of Surgery, Ipswich Hospital, Queensland, Australia
E-mail: mjburstow@gmail.com

Rossita Mohamad Yunus
Department of Mathematics and Computing, Australian Centre for
Sustainable Catchments, University of Southern Queensland, Toowoomba, Queensland,
Australia
E-mail: yunus@usq.edu.au

Shahjahan Khan
Department of Mathematics and Computing, Australian Centre for
Sustainable Catchments, University of Southern Queensland, Toowoomba, Queensland,
Australia
E-mail: khans@usq.edu.au

Breda Memon
Department of Surgery, Ipswich Hospital, Queensland, Australia
E-mail: bmemon@yahoo.com

Muhammed Ashraf Memon
Department of Surgery, Ipswich Hospital, Queensland, Australia
Department of Surgery, University of Queensland, Herston, Queensland, Australia
Faculty of Medicine and Health Sciences, Bond University, Gold Coast, Queensland, Australia
Faculty of Health Science, Bolton University, Bolton, Lancashire, UK
E-mail: mmemon@yahoo.com

ABSTRACT

Objectives: The aim was to conduct a meta-analysis of randomised control trials (RCTs) investigating the treatment of gallstone pancreatitis (GSP) by early ERCP versus conservative management and subsequent patient outcomes.

Data Sources and Review Methods: A search of Medline, Embase, Science Citation Index, Current Contents, PubMed and the Cochrane Database of Randomised control trials identified all RCTs comparing early ERCP to conservative treatment in gallstone pancreatitis published in the English Language. The meta-analysis was prepared with reference to the guidelines given in the Quality of Reporting of Meta-analysis (QUOROM) statement. Variables that were considered the most objective to analyse were overall mortality, overall morbidity, severity of pancreatitis (mild or severe), pseudocyst formation, organ failure (renal, respiratory and cardiac), abnormal coagulation, development of pancreatic abscess/phlegmon and biliary sepsis.

Results: Six trials were identified totalling 997 patients. There were significantly fewer complications in the active treatment group OR 1.78 (1.19, 2.67) with two further variables (pseudocyst formation and biliary sepsis) strongly favouring treatment but not reaching statistical significance. The other outcome variables examined showed no strong trend for either treatment regimen.

Conclusions: Early ERCP in the setting of acute GSP significantly decreases the risk of complications and biliary sepsis.

Keywords: Endoscopic retrograde cholangiopancreatography (ERCP); Meta-analysis; Randomized controlled trials (RCT); Gallstone pancreatitis; Biliary pancreatitis; Conservative treatment; Human; English

1. INTRODUCTION

Acute pancreatitis is a condition that is responsible for around 220000 hospital admission in the United States (US) and 12700 in Australia (2007-08), with an estimated cost to the US health system of $2.2 billion dollars annually (Whitcomb, 2006; AIHW Principal Diagnosis**;** Fagenholz et al., 2007). There are many and varying causes that lead to this common pathological endpoint, however the vast majority of cases are caused by either alcohol or gallstones; in Australia this is 35% and 45% respectively (Baker, 2006). The spectrum of disease severity seen in acute pancreatitis also varies, 80% of individuals will have mild pathology with a relatively benign course, 20% will suffer a severe attack, and 5% will die (Whitcomb, 2006; Fagenholz et al. 2007; Pandol et al. 2007). Pancreatic inflammation caused by gallstones is likely related to biliary "hypertension" from obstruction and inappropriate activation of pancreatic enzymes, but the exact cause is not fully understood (Whitcomb, 2006; Pandol et al., 2007; Siva and Pereira, 2006; Wysoki and Carter, 2007). Relieving biliary obstruction in GSP by ERCP has been practiced since 1973, and advocated as an early intervention in an attempt to mitigate the morbidity and mortality of this condition (Classen, 2000; Kawai, 2000; Neoptolemos et al. 1988; Fan et al. 1993; Fölsch et al. 1997; Nowak et al. 1995; Acosta et al. 2006; Oria et al. 2007). Randomised controlled trials (RCT) have been conducted comparing conservative (supportive) treatment of GSP with early (usually within 24-72hrs of presentation) ERCP, and the results have been conflicting (Neoptolemos et al. 1988; Fan et al. 1993; Fölsch et al. 1997; Nowak et al. 1995; Acosta et al. 2006; Oria et al. 2007). Meta-analyses performed on these trials have also delivered conflicting results contributing to the uncertainty surrounding the optimum management of these patients (Sharma et al. 1999; Ayub et al. 2004). Since the publication of the last meta-analysis in 2004, 2 further RCTs examining this question have been published, and it is hoped that with this additional data a clearer picture of the appropriate management of these patients will emerge (Acosta et al., 2006; Oria et al., 2007).

2. METHODS

RCTs that compared early ERCP + ES with conservative (supportive) treatment, and were published in full in peer-reviewed journals in the English language between 1970 and 2007, were included. Unpublished studies and abstracts presented at national and international meetings were also included. Published studies that contained insufficient information were also excluded, but only after an effort had been made to obtain unpublished or missing data from the original authors. Six trials were identified by conducting a comprehensive search of Medline, Embase, Science Citation Index, Current Contents and PubMed databases, using medical subject headings 'Endoscopicretrogradecholangiopancreatography (ERCP)'; 'Meta-analysis'; 'Randomized controlled trials'; 'Gallstone pancreatitis'; 'Biliary pancreatitis'; 'Conservative treatment'; 'Human'; 'English' (Neoptolemos et al. 1988; Fan et al. 1993; Fölsch et al. 1997; Nowak et al. 1995; Acosta et al. 2006; Oria et al. 2007). Manual search of the bibliographies of relevant papers was also carried out to identify trials for possible inclusion. Data extraction and critical appraisal were carried out by two authors (MJB and MAM). Standardised data extraction forms were used by these authors to independently and blindly summarise the randomised controlled trials meeting the inclusion criteria. The authors were not blinded to the source of the document or authorship for the purpose of data extraction. The data was compared and discrepancies were resolved by consensus. The primary author also contacted the original authors of some of the trials for clarification of data and to obtain unpublished, missing or additional information on various outcome measures. Variables that were considered the most objective to analyse were overall mortality, overall morbidity, severity of pancreatitis (mild or severe), pseudocyst formation, organ failure (renal, respiratory and cardiac), abnormal coagulation, development of pancreatic abscess/phlegmon and biliary sepsis. The quality of the randomized clinical trials was assessed using Jadad's scoring system and the meta-analysis prepared in accordance with the Quality of Reporting of Meta-analyses (QUOROM) statement (Jadad et al. 1996; Moher et al., 1999).

3. STATISTICAL ANALYSIS

Meta-analyses were performed using odds ratios (ORs) for binary outcome and weighted mean differences (WMDs) for continuous outcome measures. The slightly amended estimator of OR was used to avoid the computation of reciprocal of zeros among observed values in the calculation of the original OR (Agresti, 1996). Random effects models, developed using the inverse variance weighted method approach, were used to combine the data (Sutton et al. 2000). Heterogeneity among studies was assessed using the Q statistic proposed by Cochran and $I2$ index introduced by Higgins and Thompson (Sutton et al., 2000; Cochran, 1954; Hedges and Olkin, 1985; Higgins and Thompson, 2002; Huedo-Medina et al., 2006). If the observed value of Q is larger than the critical value at a given significant level (α), in this case 0.05, we conclude that there is statistically significant between-studies variation. In order to pool continuous data, mean and standard deviation are required. However, some of the published clinical trials did not report the mean and standard deviation, but rather reported the size of the trial, the median and range. Using these available statistics, estimates of the mean and standard deviation were obtained using formulas proposed by Hozo et al. Funnel plots were synthesized in order to determine the presence of publication bias in the meta-analysis. Both total sample size and precision (1/standard error) were plotted against the treatment effects (OR of outcome variable) for re-operation rate, failure rate and complication rate (Sutton et al., 2000; Egger et al., 1997; Tang et al., 2000). All the resulting funnel plots are asymmetrical, suggesting the existence of publication bias (Egger et al., 1997; Tang et al., 2000). The number of studies included in the funnel plots, indicated by the number of plotted points, is not large enough for the detection of study bias (Egger et al., 1997, Span, 2006). All estimates were obtained using a computer program written in R (R: Version 1, 2008) All plots were obtained using the meta-package (Lumley T. The rmeta Package, Version 2.14).

4. RESULTS

Six prospective RCTs were identified by the authors as meeting the inclusion and exclusion criteria for meta-analysis. The studies include 997 patients, 532 treated supportively and 465 having early ERCP +/- ES. Patient demographics and selection methods were detailed in all of the available studies. The design of each RCT was slightly different, namely in time to ERCP in the treatment group (varying from 24-72hrs), and in the specific aspects of the complications reported; the primary endpoints in each study were morbidity and mortality.

There was a significant decrease in overall complications in the treatment group as compared with supportive management (OR 1.78, 95% confidence interval (CI) 1.19, 2.67; *P* 0.0053). Mortality was not shown to be significantly improved by early biliary decompression, however, there was a trend favouring intervention (OR 1.5, CI 0.59, 3.83; *P* 0.39). Other parameters that favoured the treat group without achieving statistical significance were pseudocyst formation (OR 1.57, CI 0.81, 3.01; *P* 0.17) and biliary sepsis (OR 3.77, CI 0.74, 18.99; *P* 0.10). No strong trend favouring either hypothesis was shown for renal failure (OR 0.86, CI 0.34, 2.18; *P* 0.75), cardiac failure (OR 1.29, CI 0.62, 2.69, *P* 0.49), respiratory failure (OR 1.04, CI 0.41, 2.64; *P* 0.93) or coagulation abnormalities (OR 0.91, CI 0.35, 2.36; *P* 0.85).

5. DISCUSSION

The pancreas (Greek *pan,* all; *kreas,* flesh or meat), so named by Rufus of Ephesus around 100 AD, was first connected to the alcohol induced phenomenon of epigastric pain and vomiting by Reginald Fitz, a Harvard pathologist. In 1878 he noticed pancreatic inflammation was associated with this constellation of symptoms, which could progress to severe suppurative and haemorrhagic complications (Townsend et al., 2007; Beger et al., 2007). Biliary calculi were first associated with pancreatitis in 1901 by Opie, also a pathologist, who worked at Johns Hopkins University. He discovered an impacted stone at the Ampulla of Vata during a post-mortem on a patient who had succumbed to severe pancreatitis (Townsend et al., 2007). Acute pancreatitis is still diagnosed clinically by the characteristic epigastric pain, nausea and vomiting, now augmented with tests for serum pancreatic enzyme levels and imaging studies to determine both the cause of pancreatic inflammation and its severity (Whitcomb, 2006; Pandol et al., 2007; Baron et al., 2007). Clinically, acute pancreatitis is categorised according to severity (mild or severe) and aetiology (Whitcomb, 2006; Pandol et al., 2007). The two scoring systems most commonly used to predict severity are Ranson’s criteria and The Acute Physiology and Chronic Health Evaluation (APACHE II) score (Whitcomb, 2006; Pandol et al., 2007). A Ranson’s score ≥ 3 or an APACHE II score ≥ 8 indicates severe acute pancreatitis, and in these individuals more intensive monitoring is crucial (Whitcomb, 2006; Pandol et al., 2007; Baron et al., 2007; Swroop et al., 2004). The overall mortality in acute pancreatitis is 5%, but of the 20% of people who suffer a severe attack 10-30% will die, and in patients requiring ICU admission mortality increases to 30-50%, with outcomes further worsening as the severity scores increase (Whitcomb, 2006; Swroop et al., 2004; Nathens et al., 2004). In 70-80% of individuals, the primary cause of acute pancreatitis is either ethanol ingestion or cholelithiasis; others include drugs, trauma, systemic illness, congenital causes with around 20% classed as idiopathic (Whitcomb, 2006; Pandol et al., 2007; Swroop et al., 2004). However despite the high prevalence of gallstone acute pancreatitis and ever increasing biochemical and imaging investigations, the exact mechanism responsible for the pathological process is still unknown (Pandol et al., 2007). Although the pathogenesis is almost certainly related to aberrant activation of pancreatic enzymes and biliary obstruction, no current animal model adequately explains this, and detailed human data is not available (Whitcomb, 2006; Pandol et al., 2007).

It is no surprise therefore that early surgical intervention in the patient with acute pancreatitis has long been considered and attempted. Specifically in relation to GSP, a study by Acosta found early biliary decompression by cholecystectomy, duct exploration and transduodenal sphincterotomy performed within 48hrs decreased mortality (16% vs. 2%) (Rocha et al., 2008). This finding however was contrary to the commonly held view at the time, that early surgical intervention in patients with GSP was associated with a high morbidity and mortality (Rocha et al., 2008). ERCP, although now routine, is a relatively young intervention in the history of medical and surgical practice. Reports of the first successful endoscopic retrograde cholangiography and endoscopic retrograde pancreatography date from 1968 and 1970 respectively (Goh, 2000). The first ERCP with endoscopic sphincterotomy (ES), was performed by Demling and Classen at Erlangen on June 6, 1973, closely followed by Kawai on August 10 1973 (Classen, 2000; Kawai, 2000). During this first sphincterotomy Drs Classen and Demling also performed a gallstone extraction using the Dorma basket; Dr Classen related that this first successful ERCP + ES resulted in a "massive response" from his autonomic nervous system (Classen, 2000). From these beginnings ERCP has been honed into a formidable diagnostic and therapeutic tool over the preceding forty years.

The utility of ERCP in relieving biliary obstruction in GSP is undoubted, despite this, there is still much controversy as to which patients benefit from this intervention. Both the AGA (American Gastroenterology Association) and ACG (American College of Gastroenterologists) guidelines agree that patients with GSP plus cholangitis should undergo immediate ERCP (Banks et al., 2006; AGA Institute Medical Position Statement on Acute Pancreatitis Gastroenterology 2007). The ACG adds that patients with severe GSP should also undergo early ERCP, however the AGA guidelines sight this as controversial, sighting the conflicting data (Banks et al., 2006; AGA Institute Medical Position Statement on Acute Pancreatitis Gastroenterology 2007). The BSG (British Society of Gastroenterology) further recommends that patients with jaundice, or a dilated common bile duct should also be considered for ERCP preferably within seventy-two hours, although no mention of the severity of pancreatitis is attached to this statement (UK Guidelines for the Management of Acute Pancreatitis. 2005).

Two previous meta-analyses (and their component studies) on which these guidelines have been based reached different conclusions (Sharma et al., 1999; Ayub et al., 2004). The first analysis performed on four RCTs including 834 patients found early (24-72hrs) ERCP significantly decreased complications in the treatment group (regardless of severity) (Neoptolemos et al., 1988; Fan et al., 1993; Fölsch et al., 1997; Nowak et al., 1995; Sharma et al., 1999). The second study focused on only three RCTs with 554 patients, and showed that the significant decrease in complications was found only in patients with severe pancreatitis, neither meta-analysis found that early ERCP significantly decreased mortality (Neoptolemos et al. 1988; Fan et al., 1993; Fölsch et al., 1997; Ayub et al., 2004). Interestingly, a caveat regard the Cochrane meta-analysis must be noted as the authors included data from one study (Fölsch et al) that was not originally included in the published version which enabled further stratification of patients into those with mild and those with severe acute pancreatitis (Ayub et al., 2004). This particular study was stopped early due to poorer patient outcomes in the experimental group both for mortality (in mild and severe GSP) and morbidity (again in both groups), although this was not statistically significant (Ayub et al., 2004). The findings of this meta-analysis support those of the first, that early ERCP +/- ES significantly decreases morbidity, but not mortality.

Unfortunately there is a paucity of data in the RCTs classifying disease severity, making useful analysis of this variable difficult, and thus affecting the consensus statements drawn from them.

6. CONCLUSIONS

This meta-analysis confirms the findings of a previous review, this being that early ERCP +/- ES decreases overall complications in patients with acute GSP. Mortality, biliary sepsis and pseudo-cyst formation favour intervention but do not reach the level of statistical significance. Examination of the data available indicates that further large prospective RCTs are required to confirm the utility of early ECRP +/- ES in mild and severe GSP.

REFERENCES

Acosta J, et al. Early Ductal Decompression Versus Conservative Management for Gallstone Pancreatitis With Ampullary Obstruction A Prospective Randomized Clinical Trial. Annals of Surgery 2006:243:33-40.

AGA Institute Medical Position Statement on Acute Pancreatitis Gastroenterology 2007;132:2019–2021.

Agresti A. An Introduction to Categorical Data Analysis. Wiley & Sons; New York 1996.

AIHW Principal Diagnosis [Internet]. Australian Institute of Health and Welfare. Available from: http://d01.aihw.gov.au/cognos/cgi-bin/ppdscgi.exe?DC=Q&E=/AHS/pdx0708

Ayub K, Imada R, Slavin J. Endoscopic Retrograde Cholangiopancreatography in Gallstone-Associated Acute Pancreatitis. Cochrane Database Syst Rev 2004:CD003630.

Baker S. Diagnosis and Management of Acute Pancreatitis. Critical Care and Resuscitation 2004;6:17-27.

Baron et al. Acute Necrotising Pancreatitis. NEJM 2007;340:18.

Classen M. Back to the Future: the First Papillotomy at Erlangen. Gastrointestinal Endoscopy. 2000;51:5.

Cochran WG. The Combination of Estimates from Different Experiments. *Biometric* 1954;10:101–129.

Diseases of the Pancreas: Current Surgical Therapy By Hans G. Beger, Seiki Matsuno, John L. Cameron Springer, 2007

Egger M, Smith GD, Schneider M, Minder C. Bias in Meta-Analysis Detected by a Simple, Graphical Test. *Br Med J* 1997;315:629-634

Fagenholz P, Et al. Direct Medical Costs of Acute Pancreatitis Hospitalizations in the United States. Pancreas 2007;35(4):302-7.

Fan S-T, Lai ECS, Mok FPT, et al. Early Treatment of Acute Biliary Pancreatitis by Endoscopic Papillotomy. NEJM 1993;328:228–32.

Fölsch UR, Nitsche R, Lu¨dtek R, et al. Early ERCP and Papillotomy Compared with Conservative Management for Acute Biliary Pancreatitis. NEJM 1997;336:237– 42.

Goh KL. Development and Application of Gastrointestinal Endoscopy. Journal of the University of Malaya Medical Centre 2000;5:107-109.

Hedges LV, Olkin I. Statistical Methods for Meta Analysis: Academic Press; Orlando, Florida 1985.

Higgins JPT, Thompson SG. Quantifying Heterogeneity in a Meta-analysis. *Stat Med* 2002; 21:1539–1558.

Hozo SP, Djulbegovic B, Hozo I. Estimating the Mean and Variance from the Median, Range and Size of a Sample. *BMC Medical Research Methodology* 2005;5:13.

Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing Heterogeneity in Meta Analysis: Q Statistic or I^2 Index? *Am Psychol Assoc* 2006;11:193-206.

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the Quality of Reports of Randomized Clinical Trials: is Blinding Necessary? *Control Clin Trials* 1996;17:1-12.

Kawai K. In the beginning. Gastrointestinal Endoscopy. 2000;51:5.

Lumley T. The rmeta Package, Version 2.14, http://cran.rproject.org/web/packages/rmeta/index.html 2008.

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the Quality of Reports of Meta-Analyses of Randomised Controlled Trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999;354:1896-1900.

Nathens A, et al. Management of the Critically Ill Patient With Severe Acute Pancreatitis. Crit Care Med 2004;32:2525-2536.

Neoptolemos JR, Carr-Locke DL, London NJ, et al. Controlled Trial of Urgent Endoscopic Retrograde Cholangicipancreatography and Endoscopic Sphincterotomy Versus Conservative Treatment for Acute Pancreatitis Due to Gallstones. Lancet 1988;2:979–83.

Nowak A, Nowakowska-Dulawa E, Marek TA, et al. Final Results of the Prospective, Randomized Controlled Study on Endoscopic Sphincterotomy Versus Conventional Management in Acute Biliary Pancreatitis. Gastroenterol 1995;108:A380 (abstract).

Oria A, et al. Early Endoscopic Intervention Versus Early Conservative Management in Patients With Acute GallstonePancreatitis and Biliopancreatic Obstruction A Randomized Clinical Trial. Annals of Surgery 2007;245:10-17.

Pandol et al. Acute Pancreatitis: Bench to the bedside. Gastroenterology 2007;132:1127.

Practice Guidelines in Acute Pancreatitis. Banks P, Freeman, M, and the Practice Parameters Committee of the American College of Gastroenterology. Am J Gastroenterol 2006;101:2379–2400.

R: A Language and Environment for Statistical Computing [Computer Program]. Version 1. Vienna: R Foundation for Statistical Computing; 2008.

Rocha F, et al. A Historic Perspective on the Contributions of Surgeons to the Understanding of Acute Pancreatitis. The American Journal of Surgery 2008;196:442-449.

Sabiston Textbook of Surgery 18th Edition. Townsend Jr. C, et al. Saunders 2007.

Sharma VK, Howden CW. Metaanalysis of Randomized Controlled Trials of Endoscopic Retrograde cholangiography and Endoscopic Sphincterotomy for the Treatment of Acute Biliary Pancreatitis. Am J Gastroenterol 1999;94:3211–4.

Siva A, Pereira S. Acute Pancreatitis. Medicine 2006;35:171-177.

Span J, Carière E, Croockewitt S, Smits P. Publication Bias, Effects on the Assessment of Rosiglitasone. *Br J Clin Pharmacol* 2006;62:732.

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for Meta-analysis in Medical Research. John Wiley; London 2000.

Swroop et al. Severe Acute Pancreatitis. JAMA 2004;291:23.

Tang JL, Liu JLY. Misleading Funnel Plot Detection of Bias in Meta-Analysis. *J Clin Epidermiol* 2000;53:477-484.

UK Guidelines for the Management of Acute Pancreatitis. UK Working Party on Acute Pancreatitis. Gut 2005;54;1-9.

Whitcomb D. Acute Pancreatitis. NEJM 2006;354:20.

Wysoki A, Carter C. Acute Pancreatitis. Surgery 2007;25:49-56.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 106–118.

PRELIMINARY RESULTS OF META-ANALYSIS OF LAPAROSCOPIC AND OPEN INGUINAL HERNIA REPAIR

Matthew John Burstow
Department of Surgery, Ipswich Hospital, Queensland, Australia
E-mail: mjburstow@gmail.com

Rossita Mohamad Yunus
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments, University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: yunus@usq.edu.au

Shahjahan Khan
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments, University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: khans@usq.edu.au

Breda Memon
Department of Surgery, Ipswich Hospital, Queensland, Australia
E-mail: bmemon@yahoo.com

Muhammed Ashraf Memon
Department of Surgery, Ipswich Hospital, Queensland, Australia
Department of Surgery, University of Queensland, Herston, Queensland, Australia
Faculty of Medicine and Health Sciences, Bond University, Gold Coast, Queensland, Australia
Faculty of Health Science, Bolton University, Bolton, Lancashire, UK
E-mail: mmemon@yahoo.com

ABSTRACT

Purpose: The aim was to conduct a meta-analysis of the randomized evidence to determine the relative merits of laparoscopic inguinal herniorrhaphy (LIHR) and open inguinal herniorrhaphy (OIHR).

Data Sources and Review Methods: A search of the Medline, Embase, Science Citation Index, Current Contents and PubMed databases identified all randomized clinical trials (RCTs) that compared LIHR and OIHR published in the English literature between January 1990 and January 2007. The six outcome variables analysed were operating time, hospital stay, return to normal activity, return to work, total complications and recurrence rate. Random effects meta-analyses were performed using odds ratios and weighted mean differences.

Results: Fifty-nine trials were considered suitable for the meta-analysis. A total of 8092 patients underwent LIHR and 8580 had OIHR. For three of the six outcomes the summary point estimates favoured LIHR over OIHR; there was a significant reduction of 34% in the relative odds of postoperative complications, 4.99 days in time to return to normal activity and 6.39 days in time to return to work. However, there was a significant increase of 14.08 min in the mean operating time for LIHR. The relative odds of short term recurrence increased by 20 percent for LIHR compared with OIHR. There was a small trend towards decreased duration of hospital stay for LIHR compared with OIHR, although these results were not statistically significant.

Conclusions: Based on this meta-analysis, LIHR offers patients a number of benefits over OIHR at the expense of longer operating time but comparable recurrence rate and hospital stay.

Keywords: Hernia; Inguinal; Comparative study; Prospective studies; Randomized controlled trials; Random allocation; Clinical trial; Human; English

1. INTRODUCTION

Since Bassini introduced the prototypical inguinal hernia repair in 1887, there has been continuous development and refinement related to this surgical procedure. There has been a move away from sutured repairs to tension-free techniques employing mesh with subsequent improvements in recurrence rates and surgical morbidity (Brooks et al., 2007). Early sutured repairs had a recurrence rate as high as 15%, with newer methods this has been reduced to under 5% (Brooks et al., 2007). With the advent of laparoscopic surgery, new techniques became available to apply to the inguinal hernia repair. The success of the laparoscopic cholecystectomy led proponents to argue that the application of minimally invasive surgical techniques to inguinal hernia repair would decrease recovery time and post operative pain (Felix and Michas, 1993; Filipi et al., 1992; Fitzgibbons et al., 1994; Ger et al., 1993; Mckernan and Laws, 1993). Studies to date have largely supported this premise, however recurrence rates, singled out as the most important outcome by some authors, vary widely (see table 1). Meta-analyses comparing LIHR and OIHR have generally shown that LIHR has advantages over OIHR in providing shorter hospital stay, faster return to work and normal activities and fewer overall complications (Kuhry et al., 2006; McCormack, et al., 2003; Memon, et al., 2003; Schmedt et al., 2004; Voyles et al., 2002). However LIHR takes significantly longer to perform, is more expensive per case and shows a trend towards increased recurrence rates versus the open procedure (Kuhry et al., 2006; McCormack et al., 2003; Memon et al., 2003; Schmedt et al., 2004; Voyles et al., 2002).

This meta-analysis is an update to that already published by Memon et al. (2003) examined RCTs comparing LIHR and OIHR, with a total of 59 trials included. The Quality of Reporting of Meta-analyses (QUOROM) statement was followed in the preparation of this study (Moher et al., 1999).

2. METHODS

RCTs that compared LIHR with any type of OIHR, and were published in full in peer-reviewed journals in the English language between January 1990 and the end of Jan 2008, were included. Unpublished studies and abstracts presented at national and international meetings were excluded. Published studies that reported three or fewer outcome variables or that contained insufficient information were also excluded, but only after an effort had been made to obtain unpublished or missing data from the original authors. Trials were identified by conducting a comprehensive search of Medline, Embase, Science Citation Index, Current Contents and PubMed databases, using medical subject headings ‘hernia’, ‘inguinal’, ‘comparative study’, ‘prospective studies’, ‘randomized controlled trials’, ‘random allocation’ and ‘clinical trial’. Manual search of the bibliographies of relevant papers was also carried out to identify trials for possible inclusion. Data extraction and critical appraisal were carried out by three authors, who also contacted the original authors of some of the trials for clarification of data and to obtain unpublished, missing or additional information on various outcome measures. The response to this was extremely good. Six outcome variables were considered most suitable for analysis: operating time, time to discharge from hospital, return to normal activity and return to work, postoperative complications and hernia recurrence rate. Other outcome measures, such as postoperative pain, analgesia requirements and hospital costs, were excluded owing to variations in reporting methodology and the inability to devise uniform objective analysis of these outcomes. The quality of the randomized clinical trials was assessed using Jadad's scoring system (Jadad et al., 1996).

Table 1. Studies Used For Meta-analysis

Aitola P, Airo I, Matikainen M. Laparoscopic versus open preperitoneal inguinal hernia repair: a prospective randomised trial. Ann Chir Gynaecol 1998;87:22–25.
Andersson B, et al. Laparoscopic extraperitoneal inguinal hernia repair versus open mesh repair: A prospective randomized controlled trial. Surgery 2003;133:464-72.
Arvidsson D, et al. Randomized clinical trial comparing 5-year recurrence rate after laparoscopic versus Shouldice repair of primary inguinal hernia. British Journal of Surgery 2005;92:1085–1091.
Barkun JS,Wexler MJ, Hinchey EJ, Thibeault D, Meakins JL. Laparoscopic versus open inguinal herniorrhaphy: preliminary results of a randomized controlled trial. Surgery 1995;118:703–710.
Beets GL, Dirksen CD, Go PM, Geisler FE, Baeten CG, Kootstra G. Open or laparoscopic preperitoneal mesh repair for recurrent inguinal hernia? A randomized controlled trial. Surg Endosc 1999;13:323–327.
Berndsen F, et al. Discomfort Five years after laparoscopic and Shouldice inguinal hernia repair: a randomised trial with 867 patients. A report from the SMIL study group. Hernia 2007;11:307–313.
Berndsen F, et al. Postoperative convalescence after inguinal hernia surgery: Prospective randomized multicenter study of laparoscopic versus Shouldice inguinal hernia repair in 1042 patients. Hernia 2002;6:56–61.
Bessell JR, Baxter P, Riddell P, Watkin S, Maddern GJ. A randomized controlled trial of laparoscopic extraperitoneal hernia repair as a day surgical procedure. Surg Endosc 1996;10:495–500.
Bringman S, et al. Tension-Free Inguinal Hernia Repair: TEP Versus Mesh-Plug Versus Lichtenstein. Annals of Surgery 2003;237:142–147.
Butters M, et al. Long-term results of a randomized clinical trial of Shouldice, Lichtenstein and transabdominal preperitoneal hernia repairs. British Journal of Surgery 2007;94:562–565.
Champault G, Rizk N, Catheline JM, Barrat C, Turner R, Boutelier P. Totally pre-peritoneal laparoscopic approach versus Stoppa operation. Randomized trial: 100 cases. Hernia 1997;1:31–36.
Champault GG, Rizk N, Catheline JM, Turner R, Boutelier P. Inguinal hernia repair: totally preperitoneal laparoscopic approach versus Stoppa operation: randomized trial of 100 cases. Surg Laparosc Endosc 1997;7:445–450.
Colak T, et al. Randomized Clinical Trial Comparing Laparoscopic Totally Extraperitoneal Approach with Open Mesh Repair in Inguinal Hernia. Surgical Laparoscopy, Endoscopy & Percutaneous Techniques 2003;13:191-195.
D. Mahon, B. Decadt, M. Rhodes Prospective randomized trial of laparoscopic (transabdominal preperitoneal) vs open (mesh) repair for bilateral and recurrent inguinal hernia. Surg Endosc 2003;17:1386–1390.

Dedemadi G, Sgourakis G, Karaliotas C, et al. Comparison of laparoscopic and open tension-free repair of recurrent inguinal hernias: a prospective randomized study. Surgical Endoscopy 2006/7;20:1099-1104.
Dirksen CD, Beets GL, Go PM, Geisler FE, Baeten CG, Kootstra G. Bassini repair compared with laparoscopic repair for primary inguinal hernia: a randomised controlled trial. Eur J Surg 1998;164: 439–447.
Douek M, et al. Prospective randomised controlled trial of laparoscopic versus open inguinal hernia mesh repair: five year follow up. BMJ 2005;326:1012-1013.
Eklund A, et al. Short-term results of a randomized clinical trial comparing Lichtenstein open repair with totally extraperitoneal laparoscopic inguinal hernia repair. British Journal of Surgery 2006;93:1060–1068.
Filipi CJ, Gaston-Johansson F, McBride PJ, Murayama K, Gerhardt J, Cornet DA et al. An assessment of pain and return to normal activity. Laparoscopic herniorrhaphy versus open tension free Lichtenstein repair. Surg Endosc 1996;10:983–986.
Fleming W, et al. Randomised Clinical Trial Comparing totally Extraperitoneal Inguinal Hernia Repair with the Shouldice Technique. British Journal of Surgery 2001;88:1183-1188.
Gokalp A, Inal M, Maralcan G, et al. A prospective randomized study of Lichtenstein open tension-free versus laparoscopic totally extraperitoneal techniques for inguinal hernia repair. Acta Chir Belg 2003;103:502–506.
Günal Ö, et al. Does the approach to the groin make a difference in hernia repair? Hernia 2007; 11:429–434.
H. Pokorny et al. Postoperative pain and quality of life after laparoscopic and open inguinal hernia repair: results of a prospective randomized trial. Hernia 2006;10:331–337.
Heikkinen T, et al. Five-year outcome of laparoscopic and Lichtenstein hernioplasties. Surg Endosc 2004;18:518–522.
Heikkinen T, Haukipuro K, Leppala J, Hulkko A. Total costs of laparoscopic and Lichtenstein inguinal hernia repairs: a randomized prospective study. Surg Laparosc Endosc 1997;7:1–5.
Heikkinen TJ, Haukipuro K, Hulkko A. A cost and outcome comparison between laparoscopic and Lichtenstein hernia operations in a day-case unit. A randomized prospective study. Surg Endosc 1998;12:1199–1203.
Heikkinen TJ, Haukipuro K, Koivukangas P, Hulkko A. A prospective randomized outcome and cost comparison of totally extraperitoneal endoscopic hernioplasty versus Lichtenstein hernia operation among employed patients. Surg Laparosc Endosc Percutan Tech 1998;8:338–344.
Hynes D, et al. Cost Effectiveness of Laparoscopic Versus Open Mesh Hernia Operation: Results of a Department of Veterans Affairs Randomized Clinical Trial. J Am Coll Surg 2006:203;447-457.
Johansson B, Hallerback B, Glise H, Anesten B, Smedberg S, Roman J. Laparoscopic mesh versus open preperitoneal mesh versus conventional technique for inguinal hernia repair: a randomized multicenter trial (SCUR Hernia Repair Study). Ann Surg 1999;230:225–231.
Juul P, Christensen K. Randomized clinical trial of laparoscopic versus open inguinal hernia repair. Br J Surg 1999;86:316–319.
Kald A, Anderberg B, Carlsson P, Park PO, Smedh K. Surgical outcome and cost-minimisation-analyses of laparoscopic and open hernia repair: a randomised prospective trial with one year follow up. Eur J Surg 1997;163:505–510.

Khoury N. A randomized prospective controlled trial of laparoscopic extraperitoneal hernia repair and mesh-plug hernioplasty: a study of 315 cases. J Laparoendosc Adv Surg Tech Part A 1998;8:367–372.
Kingsley D, Vogt DM, Nelson T, et al: Laparoscopic intraperitoneal onlay inguinal herniorrhaphy. Am J Surg 1998;176:548-553
Lal P, et al. Randomized controlled study of laparoscopic total extraperitoneal vs open Lichtenstein inguinal hernia repair. Surg Endosc 2003;17:850–856.
Lau H, Patil NG, andYuen WK. Day-case endoscopic totally extraperitoneal inguinal hernioplasty versus open Lichtenstein hernioplasty for unilateral primary inguinal hernia in males: a randomized trial. Surgical endoscopy 2006;20:76-81.
Lawrence K, McWhinnie D, Goodwin A, Doll H, Gordon A, Gray A et al. Randomised controlled trial of laparoscopic versus open repair of inguinal hernia: early results. BMJ 1995;311:981–985.
Leibl BJ, Daubler P, Schmedt C-G, Kraft K, Bittner R. Long-term results of a randomized clinical trial between laparoscopic hernioplasty and Shouldice repair. Br J Surg 2000;87:780–783.
Leigh Neumayer et al. Open Mesh versus Laparoscopic Mesh Repair of Inguinal Hernia. NEJM 2004;350:1819-1827.
Liem M, et al. Recurrences After Conventional Anterior andLaparoscopic Inguinal Hernia Repair A Randomized Comparison. Annals of Surgery 2003;237:136–141.
Liem MSL, van der Graaf Y, van Steensel CJ, Boelhouwer RU, Clevers GJ, Meijer WS et al. Comparison of conventional anterior surgery and laparoscopic surgery for inguinal hernia repair. N Engl JMed 1997;336:1541–1547.
Lorenz D, Stark E, Oestreich K, Richter A. Laparoscopic hernioplasty versus conventional hernioplasty (Shouldice): results of a prospective randomized trial. World J Surg 2000;24:739–746.
Maddern GJ, Rudkin G, Bessell JR, Devitt P, Ponte L. A comparison of laparoscopic and open hernia repair as a day surgical procedure. Surg Endosc 1994;8:1404–1408.
MRC Laparoscopic Groin Hernia Trial Group. Laparoscopic versus open repair of groin hernia: a randomised comparison. Lancet 1999;354:185–190.
Paganini AM, Lezoche E, Carle F, Carlei F, Favretti F, Feliciotti F et al. A randomized, controlled, clinical study of laparoscopic versus open tension-free inguinal hernia repair. Surg Endosc 1998;12:979–986.
Payne JH Jr, Grininger LM, Izawa MT, Podoll EF, Lindahl PJ, Balfour J. Laparoscopic or open inguinal herniorrhaphy? A randomized prospective trial. Arch Surg 1994;129:973–981.
Picchio M, Lombardi A, Zolovkins A, Mihelsons M, La Torre G. Tension-free laparoscopic and open hernia repair: randomized controlled trial of early results. World J Surg 1999;23:1004–1009.
Sarli L, et al. Simultaneous Repair of Bilateral Inguinal Hernias: A Prospective, Randomized Study of Open, Tension-Free Versus Laparoscopic Approach. Surgical Laparoscopy, Endoscopy & Percutaneous Techniques 2001;11:262-267.
Schrenk P,Woisetschlager R, Rieger R,WayandW. Prospective randomized trial comparing postoperative pain and return to physical activity after transabdominal preperitoneal, total preperitoneal or Shouldice technique for inguinal hernia repair. Br J Surg 1996;83:1563–1566.
Simmermacher R, et al. Preperitoneal mesh in groin hernia surgery. A randomized clinical trial emphasizing the surgical aspects of preperitoneal placement via a laparoscopic (TEP)

or Grid-iron (Ugahary) approach. Hernia 2000;4:296-298.
Stoker DL, Spiegelhalter DJ, Singh R, Wellwood JM. Laparoscopic versus open inguinal hernia repair: randomised prospective trial. Lancet 1994;343:1243–1245.
Suter M, Martinet O, Spertini F. Reduced acute phase response after laparoscopic total extraperitoneal bilateral hernia repair compared to open repair with the Stoppa procedure. Surg Endosc. 2002;16:1214-9.
Tanphiphat C, Tanprayoon T, Sangsubhan C, Chatamra K. Laparoscopic versus open inguinal hernia repair. A randomized, controlled trial. Surg Endosc 1998;12:846–851.
Tschudi J, Wagner M, Klaiber C, Brugger J, Frei E, Krahenbuhl L et al. Controlled multicenter trial of laparoscopic transabdominal preperitoneal hernioplasty versus Shouldice herniorrhaphy. Early results. Surg Endosc 1996;10:845–847.
Vantansev C, et al. The Effects of Different Hernia Repair Methods on Postoperative Pain Medication and CRP Levels. Surgical Laparoscopy, Endoscopy & Percutaneous Techniques. 2002;12:243-246.
Vidovic D, Kirac I, Glavan E, et al. Laparoscopic totally extraperitoneal hernia repair versus open Lichtenstein hernia repair: results and complications. J Laparoendosc Adv Surg Tech A 2007;17:585-90.
Vogt DM, Curet MJ, Pitcher DE, Martin DT, Zucker KA. Preliminary results of a prospective randomized trial of laparoscopic onlay versus conventional inguinal herniorrhaphy. Am J Surg 1995;169:84–90.
Wellwood J, Sculpher MJ, Stoker D, Nicholls GJ, Geddes C, Whitehead A et al. Randomised controlled trial of laparoscopic versus open mesh repair for inguinal hernia: outcome and cost. BMJ 1998;317:103–110.
Wright DM, Kennedy A, Baxter JN, Fullarton GM, Fife LM, Sunderland GT et al. Early outcome after open versus extraperitoneal endoscopic tension-free hernioplasty: a randomized clinical trial. Surgery 1996;119:552–557.
Zieren J, Zieren HU, Jacobi CA,Wenger FA, Muller JM. Prospective randomized study comparing laparoscopic and open tension-free inguinal hernia repair with Shouldice's operation. Am J Surg 1998;175:330–333.

3. STATISTICAL ANALYSIS

Meta-analyses were performed using odds ratios (ORs) for binary and weighted mean differences (WMDs) for continuous outcome measures (Sutton et al., 2000). Random effects models were used to combine the data and statistical heterogeneity was assessed using the $\chi 2$ test. To assess whether heterogeneity was explained by study-level co-variates (year of study, length of follow-up and size of study) a random effects meta-regression model was used (Thompson and Sharp, 1998). Subgroup analyses were performed by comparing the results of the two methods of LIHR (transabdominal preperitoneal (TAPP) and totally extraperitoneal (TEP)) and OIHR (tension free and tension creating) separately (Sutton et al., 2000). A sensitivity analysis was carried out to assess the impact of study quality on the results, by identifying poor-quality studies (Jadad score <1) (Jadad et al., 1996). Funnel plots were synthesized in order to determine the presence of publication bias in the meta-analysis. All estimates were obtained using a computer program written in R, and all plots were obtained using the meta-package (Vienna: R Foundation for Statistical Computing, 2008; Lumley T. The rmeta Package, Version 2.14, 2008).

4. RESULTS

Table 2 Summary of pooled data comparing LIHR and OIHR

Outcome Variables	Pooled OR or WMD	Test for Overall Effect		Test for heterogeneity	
		Z	ρ	χ^2	ρ
Operating times	14.08(8.36, 19.80)†	4.83	<0.0001	8830.741	<-0.0001
Hospital stay	-0.06(-0.19,0.08)†	-0.80	0.4223	4261.20	<0.0001
Return to normal	-4.99(-6.1,3.88)†	-8.81	<0.0001	669.16	<0.0001
Return to work	-6.39(-7.95,4.84)†	-8.06	<0.0001	597.95	<0.0001
Complication	0.66(0.53,0.82)*	-3.78	0.0002	186.56	<0.0001
Recurrence	1.19(0.92,1.56)*	1.32	0.1874	82.90	0.0031

Values in parentheses are 95 percent confidence intervals. * stands for OR odds ratio; † stands for WMD weighted mean difference.

For three of the six outcomes (see table 2) the summary point estimates favoured LIHR over OIHR; there was a significant reduction of 34 percent in the relative odds of postoperative complications OR 0.66, 95% confidence interval (CI) 0.53 to 0.82; P = 0.0002), 4.99 (WMD -4.99, 95% CI -6.10 to -3.88; P < 0.0001) days in time to return to normal activity and 6.39 (WMD -6.39, 95% CI -7.95 to - 4.84; P < 0.0001) days in time to return to work. There was a significant increase of 14.08 (WMD 14.08, 95% CI 8.36 to 19.80; P < 0.0001) min in the mean operating time for LIHR. The relative odds of short term recurrence were increased by 20 percent for LIHR compared with OIHR (OR 1.19, 95% CI 0.92 to 1.56; P = 0.1874) and a reduction of 0.06 days of duration of hospital stay (WMD -0.06, 95% CI -0.19 to 0.08; P = 0.4223) for LIHR compared with OIHR, although these results were not statistically significant.

5. DISCUSSION

Laparoscopic surgery continues to be a rapidly advancing and pioneering field. Many different procedures are now being performed laparoscopically, primarily due to the perceived, and in some cases (i.e. laparoscopic cholecystectomy) confirmed benefits of minimally invasive surgery. LIHR has been controversial since its introduction in the early 1990s (Memon and Fitzgibbons, 1998; Memon et al., 1997). Those supporting LIHR claim reduced post-operative pain, earlier return to work or full physical activity and superior cosmesis when compared to OIHR (Felix and Michas, 1993; Filipi et al., 1992; Fitzgibbons et al., 1994; Ger et al., 1993; Mckernan and Laws, 1993). Critics of LIHR have cited increased operating time, as well as cost and the associated technical difficulties of the procedure (Memon and Fitzgibbons, 1998).

Operating time in the data analysed was shown to be significantly reduced in the OIHR group versus the LIHR procedures by 14 minutes. This is not surprising considering the extra time required in preparing equipment and gaining access to the operative area. Surgeon experience also has a marked effect on the time taken per operation, with later trials displaying laparoscopic operating time very similar to the open procedure. It has been estimated that LIHR increased theatre cost by around $600 (US), moreover the effect of longer operating time on the patient may increase morbidity and mortality (Memon et al., 2003; Voyles et al., 2002).

Duration of hospital stay, a frequently sited benefit of laparoscopic surgery, was not significantly different between the two groups. This has been a variable of considerable dispute in the literature. Previous systematic reviews have come to differing results, either in favour of

LIHR or with no significant difference in hospital stay, as this review found. Importantly, no review has found that LIHR increases hospital stay versus OIHR, and in general the trend is toward decreased stay for LIHR (Kuhry et al., 2006; McCormack et al., 2003; Memon et al., 2003; Schmedt et al., 2004; Voyles et al., 2002). An explanation for this may be in the increasing proportion of inguinal hernia repairs done as day cases, reducing overall length of stay for OIHR patients and thus nullifying the statistically significant findings of early reviews with the addition of more recent RCTs (Aylin et al., 2005).

Return to normal activity clearly favoured LIHR (5 days) as did time to return to work (6.4 days), and of all the variables favouring LIHR, this has been a consistent finding (Kuhry et al., 2006; McCormack et al., 2003; Memon et al., 2003; Schmedt et al., 2004; Voyles et al., 2002). Factors influencing this finding obviously include the lack of a muscle disruption, no groin incision and reduced tissue manipulation. Further explanations of this difference between OIHR and LIHR centre on so-call physician and/or patient bias, but this is very difficult to quantify. Odds ratio of overall complications significantly favoured the LIHR group. This finding is consistent with our previous meta-analysis, and other earlier systematic reviews (McCormack et al., 2003; Memon et al., 2003; Schmedt et al., 2004; Voyles et al., 2002). Further sub-group analysis in this area is pending as it has been suggested that LIHR is associated with rare but severe complications such as major vascular injury and bowel perforation; this was not supported by our earlier study (Kuhry et al., 2006; Memon et al., 2003).

Overall recurrence favours OIHR over LIHR, (OR 1.19) however this is not significant (see table 2). LIHR has been shown to be superior than the sutured/tension varieties of OIHR which are little used in current practice, however when a ‘modern’ non-tension mesh OIHR is performed there is no difference between the two in recurrence rates (Brooks, 2007; Memon et al., 2003). Of the two primary LIHR techniques commonly used, our analysis showed that the RCTs employing the Total Extraperitoneal (TEP) repair had a strong trend towards a decreased recurrence rate, the opposite was true for the RCTs employing the TAPP (Transabdominal Pre-Peritoneal) repair; neither reached statistical significance. This trend was not noted in previous meta-analyses, and the opposite for the TAPP repair has been previously noted (McCormack et al., 2003). Many more studies using the TEP method have been published since 2003, the first meta-analysis by Memon et al. (2003) included 22 RCTs reporting TAPP versus OIHR and only 6 employing a TEP approach. This study included 34 RCTs purely comparing TAPP with OIHR and 20 comparing TEP with OIHR, the increase in trials reporting a TEP approach may somewhat explain this change in trend. Long-term follow-up is crucial to meaningful estimation of recurrence, with most recurrences occurring between 5 to 10 years post herniorrhaphy; studies reporting long-term recurrence rates with LIHR are eagerly awaited (Memon et al., 2003).

6. CONCLUSIONS

LIHR is an exciting development in the continuing evolution of general surgery. This meta-analysis both confirms results from previous reviews, but raises some new questions. Operation time for LIHR exceeds that its open counterpart, while return to work and normal activity are dramatically reduced, furthermore the complication rate associated with LIHR is comparable with that of OIHR. While these variables are consistent with previous analyses published, our data suggests a trend to increased recurrence with the TAPP LIHR as compared to OIHR, with the opposite finding for TEP versus OIHR; further sub-group analysis is pending. Long-term recurrence rates are still required however to make a definitive judgement in this area.

REFERENCES

Aitola P, Airo I, Matikainen M. Laparoscopic versus open preperitoneal inguinal hernia repair: a prospective randomised trial. Ann Chir Gynaecol 1998; 87:22–25.

Andersson B, et al. Laparoscopic extraperitoneal inguinal hernia repair versus open mesh repair: A prospective randomized controlled trial. Surgery 2003; 133: 464-72.

Arvidsson D, et al. Randomized clinical trial comparing 5-year recurrence rate after laparoscopic versus Shouldice repair of primary inguinal hernia. British Journal of Surgery 2005; 92: 1085–1091.

Aylin P, et al. Trends in Day Surgery Rates. BMJ 2005; 331;803.

Barkun JS,Wexler MJ, Hinchey EJ, Thibeault D, Meakins JL. Laparoscopic versus open inguinal herniorrhaphy: preliminary results of a randomized controlled trial. Surgery 1995; 118: 703–710.

Beets GL, Dirksen CD, Go PM, Geisler FE, Baeten CG, Kootstra G. Open or laparoscopic preperitoneal mesh repair for recurrent inguinal hernia? A randomized controlled trial. Surg Endosc 1999; 13: 323–327.

Berndsen F, et al. Discomfort Five years after laparoscopic and Shouldice inguinal hernia repair: a randomised trial with 867 patients. A report from the SMIL study group. Hernia 2007; 11: 307–313.

Berndsen F, et al. Postoperative convalescence after inguinal hernia surgery: Prospective randomized multicenter study of laparoscopic versus Shouldice inguinal hernia repair in 1042 patients. Hernia 2002; 6: 56–61.

Bessell JR, Baxter P, Riddell P, Watkin S, Maddern GJ. A randomized controlled trial of laparoscopic extraperitoneal hernia repair as a day surgical procedure. Surg Endosc 1996; 10: 495–500.

Bringman S, et al. Tension-Free Inguinal Hernia Repair: TEP Versus Mesh-Plug Versus Lichtenstein. Annals of Surgery 2003; 237: 142–147.

Brooks, DC. Treatment of Groin Hernias. In: UpToDate, Rose, BD (Ed), UpToDate, Waltham, MA, 2007

Butters M, et al. Long-term results of a randomized clinical trial of Shouldice, Lichtenstein and transabdominal preperitoneal hernia repairs. British Journal of Surgery 2007; 94:562–565.

Champault G, Rizk N, Catheline JM, Barrat C, Turner R, Boutelier P. Totally pre-peritoneal laparoscopic approach versus Stoppa operation. Randomized trial: 100 cases. Hernia 1997; 1: 31–36.

Champault GG, Rizk N, Catheline JM, Turner R, Boutelier P. Inguinal hernia repair: totally preperitoneal laparoscopic approach versus Stoppa operation: randomized trial of 100 cases. Surg Laparosc Endosc 1997; 7: 445–450.

Colak T, et al. Randomized Clinical Trial Comparing Laparoscopic Totally Extraperitoneal Approach with Open Mesh Repair in Inguinal Hernia. Surgical Laparoscopy, Endoscopy & Percutaneous Techniques 2003; 13: 191-195.

Dedemadi G, Sgourakis G, Karaliotas C, et al. Comparison of laparoscopic and open tension-free repair of recurrent inguinal hernias: a prospective randomized study. Surgical Endoscopy 2006/7; 20: 1099-1104.

Dirksen CD, Beets GL, Go PM, Geisler FE, Baeten CG, Kootstra G. Bassini repair compared with laparoscopic repair for primary inguinal hernia: a randomised controlled trial. Eur J Surg 1998; 164: 439–447.

Douek M, et al. Prospective randomised controlled trial of laparoscopic versus open inguinal hernia mesh repair: five year follow up. BMJ 2005;326:1012-1013.

Eklund A, et al. Short-term results of a randomized clinical trial comparing Lichtenstein open repair with totally extraperitoneal laparoscopic inguinal hernia repair. British Journal of Surgery 2006; 93: 1060–1068.

Felix EL, Michas C. Double-buttress laparoscopic herniorrhaphy. *J Laparoendosc Surg* 1993; 3: 1–8.

Filipi CJ, Fitzgibbons RJ Jr, Salerno GM, Hart RO. Laparoscopic herniorrhaphy. *Surg Clin North Am* 1992; 72: 1109–1124.

Filipi CJ, Gaston-Johansson F, McBride PJ, Murayama K, Gerhardt J, Cornet DA et al. An assessment of pain and return to normal activity. Laparoscopic herniorrhaphy versus open tension free Lichtenstein repair. Surg Endosc 1996; 10: 983–986.

Fitzgibbons RJ Jr, Salerno GM, Filipi CJ, HunterWJ, Watson P. A laparoscopic intraperitoneal onlay mesh technique for the repair of an indirect inguinal hernia. *Ann Surg* 1994; 219: 144–156.

Fleming W, et al. Randomised Clinical Trial Comparing totally Extraperitoneal Inguinal Hernia Repair with the Shouldice Technique. British Journal of Surgery 2001; 88: 1183-1188.

Ger R, Mishrick A, Hurwitz J, Romero C, Oddsen R. Management of groin hernias by laparoscopy. *World J Surg* 1993; 17: 46–50.

Gokalp A, Inal M, Maralcan G, et al. A prospective randomized study of Lichtenstein open tension-free versus laparoscopic totally extraperitoneal techniques for inguinal hernia repair. Acta Chir Belg 2003; 103: 502–506.

Günal Ö, et al. Does the approach to the groin make a difference in hernia repair? Hernia 2007; 11:429–434.

Heikkinen T, et al. Five-year outcome of laparoscopic and Lichtenstein hernioplasties. Surg Endosc 2004; 18: 518–522.

Heikkinen T, Haukipuro K, Leppala J, Hulkko A. Total costs of laparoscopic and Lichtenstein inguinal hernia repairs: a randomized prospective study. Surg Laparosc Endosc 1997; 7: 1–5.

Heikkinen TJ, Haukipuro K, Hulkko A. A cost and outcome comparison between laparoscopic and Lichtenstein hernia operations in a day-case unit. A randomized prospective study. Surg Endosc 1998; 12: 1199–1203.

Heikkinen TJ, Haukipuro K, Koivukangas P, Hulkko A. A prospective randomized outcome and cost comparison of totally extraperitoneal endoscopic hernioplasty versus Lichtenstein hernia operation among employed patients. Surg Laparosc Endosc Percutan Tech 1998; 8: 338–344.

Hynes D, et al. Cost Effectiveness of Laparoscopic Versus Open Mesh Hernia Operation: Results of a Department of Veterans Affairs Randomized Clinical Trial. J Am Coll Surg 2006: 203; 447-457.

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ,Gavaghan DJ et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996: 17; 1–12.

Johansson B, Hallerback B, Glise H, Anesten B, Smedberg S, Roman J. Laparoscopic mesh versus open preperitoneal mesh versus conventional technique for inguinal hernia repair: a randomized multicenter trial (SCUR Hernia Repair Study). Ann Surg 1999; 230: 225–231.

Juul P, Christensen K. Randomized clinical trial of laparoscopic versus open inguinal hernia repair. Br J Surg 1999; 86: 316–319.

Kald A, Anderberg B, Carlsson P, Park PO, Smedh K. Surgical outcome and cost-minimisation-analyses of laparoscopic and open hernia repair: a randomised prospective trial with one year follow up. Eur J Surg 1997; 163: 505–510.

Khoury N. A randomized prospective controlled trial of laparoscopic extraperitoneal hernia repair and mesh-plug hernioplasty: a study of 315 cases. J Laparoendosc Adv Surg Tech Part A 1998; 8: 367–372.

Kingsley D, Vogt DM, Nelson T, et al: Laparoscopic intraperitoneal onlay inguinal herniorrhaphy. Am J Surg 1998; 176: 548-553.

Lal P, et al. Randomized controlled study of laparoscopic total extraperitoneal vs open Lichtenstein inguinal hernia repair. Surg Endosc 2003; 17: 850–856.

Lau H, Patil NG, andYuen WK. Day-case endoscopic totally extraperitoneal inguinal hernioplasty versus open Lichtenstein hernioplasty for unilateral primary inguinal hernia in males: a randomized trial. Surgical endoscopy 2006; 20: 76-81.

Lawrence K, McWhinnie D, Goodwin A, Doll H, Gordon A, Gray A et al. Randomised controlled trial of laparoscopic versus open repair of inguinal hernia: early results. BMJ 1995; 311: 981–985.

Leibl BJ, Daubler P, Schmedt C-G, Kraft K, Bittner R. Long-term results of a randomized clinical trial between laparoscopic hernioplasty and Shouldice repair. Br J Surg 2000; 87: 780–783.

Liem M, et al. Recurrences After Conventional Anterior andLaparoscopic Inguinal Hernia Repair A Randomized Comparison. Annals of Surgery 2003; 237: 136–141.

Liem MSL, van der Graaf Y, van Steensel CJ, Boelhouwer RU, Clevers GJ, Meijer WS et al. Comparison of conventional anterior surgery and laparoscopic surgery for inguinal hernia repair. N Engl JMed 1997; 336: 1541–1547.

Lorenz D, Stark E, Oestreich K, Richter A. Laparoscopic hernioplasty versus conventional hernioplasty (Shouldice): results of a prospective randomized trial. World J Surg 2000; 24: 739–746.

Maddern GJ, Rudkin G, Bessell JR, Devitt P, Ponte L. A comparison of laparoscopic and open hernia repair as a day surgical procedure. Surg Endosc 1994; 8: 1404–1408.

Mahon D, et al. Prospective randomized trial of laparoscopic (transabdominal preperitoneal) vs open (mesh) repair for bilateral and recurrent inguinal hernia. Surg Endosc 2003; 17: 1386–1390.

McKernan JB, Laws HL. Laparoscopic repair of inguinal hernias using a totally extraperitoneal prosthetic approach. *Surg Endosc* 1993; 7: 26–28.

Memon MA, Fitzgibbons RJ Jr. Assessing risks, costs and benefits of laparoscopic hernia repair. Annu Rev Med 1998; 49: 95-109.

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the Quality of Reports of Meta-Analyses of Randomised Controlled Trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999; 354: 1896-1900

MRC Laparoscopic Groin Hernia Trial Group. Laparoscopic versus open repair of groin hernia: a randomised comparison. Lancet 1999; 354: 185–190.

Neumayer L, et al. Open Mesh versus Laparoscopic Mesh Repair of Inguinal Hernia. NEJM 2004; 350: 1819-1827.

Paganini AM, Lezoche E, Carle F, Carlei F, Favretti F, Feliciotti F et al. A randomized, controlled, clinical study of laparoscopic versus open tension-free inguinal hernia repair. Surg Endosc 1998; 12: 979–986.

Payne JH Jr, Grininger LM, Izawa MT, Podoll EF, Lindahl PJ, Balfour J. Laparoscopic or open inguinal herniorrhaphy? A randomized prospective trial. Arch Surg 1994; 129: 973–981.

Picchio M, Lombardi A, Zolovkins A, Mihelsons M, La Torre G. Tension-free laparoscopic and open hernia repair: randomized controlled trial of early results. World J Surg 1999; 23: 1004–1009.

Pokorny H, et al. Postoperative pain and quality of life after laparoscopic and open inguinal hernia repair: results of a prospective randomized trial. Hernia 2006; 10: 331–337.

Sarli L, et al. Simultaneous Repair of Bilateral Inguinal Hernias: A Prospective, Randomized Study of Open, Tension-Free Versus Laparoscopic Approach. Surgical Laparoscopy, Endoscopy & Percutaneous Techniques 2001; 11: 262-267.

Schrenk P,Woisetschlager R, Rieger R,WayandW. Prospective randomized trial comparing postoperative pain and return to physical activity after transabdominal preperitoneal, total preperitoneal or Shouldice technique for inguinal hernia repair. Br J Surg 1996; 83: 1563–1566.

Sharp SJ, Sterne J. Fixed and random effects meta-analysis with graphics. Stata Tech Bull 1998; 7: 106–108.

Sharp SJ, Sterne J. Meta-analysis regression. Stata Tech Bull 1998; 7: 148–155.

Simmermacher R, et al. Preperitoneal mesh in groin hernia surgery. A randomized clinical trial emphasizing the surgical aspects of preperitoneal placement via a laparoscopic (TEP) or Grid-iron (Ugahary) approach. Hernia 2000; 4: 296-298.

Stoker DL, Spiegelhalter DJ, Singh R, Wellwood JM. Laparoscopic versus open inguinal hernia repair: randomised prospective trial. Lancet 1994; 343: 1243–1245.

Suter M, Martinet O, Spertini F. Reduced acute phase response after laparoscopic total extraperitoneal bilateral hernia repair compared to open repair with the Stoppa procedure. Surg Endosc. 2002; 16: 1214-9.

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for Meta-analysis in Medical Research. JohnWiley: Chichester, 2000.

Tanphiphat C, Tanprayoon T, Sangsubhan C, Chatamra K. Laparoscopic versus open inguinal hernia repair. A randomized, controlled trial. Surg Endosc 1998;12:846–851.

Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med 1999; 18: 2693–2708.

Tschudi J, Wagner M, Klaiber C, Brugger J, Frei E, Krahenbuhl L et al. Controlled multicenter trial of laparoscopic transabdominal preperitoneal hernioplasty versus Shouldice herniorrhaphy. Early results. Surg Endosc 1996;10:845–847.

Vantansev C, et al. The Effects of Different Hernia Repair Methods on Postoperative Pain Medication and CRP Levels. Surgical Laparoscopy, Endoscopy & Percutaneous Techniques. 2002;12:243-246.

Vidovic D, Kirac I, Glavan E, et al. Laparoscopic totally extraperitoneal hernia repair versus open Lichtenstein hernia repair: results and complications. J Laparoendosc Adv Surg Tech A 2007;17:585-90.

Vogt DM, Curet MJ, Pitcher DE, Martin DT, Zucker KA. Preliminary results of a prospective randomized trial of laparoscopic onlay versus conventional inguinal herniorrhaphy. Am J Surg 1995;169:84–90.

Wellwood J, Sculpher MJ, Stoker D, Nicholls GJ, Geddes C, Whitehead A et al. Randomised controlled trial of laparoscopic versus open mesh repair for inguinal hernia: outcome and cost. BMJ 1998;317:103–110.

Wright DM, Kennedy A, Baxter JN, Fullarton GM, Fife LM, Sunderland GT et al. Early outcome after open versus extraperitoneal endoscopic tension-free hernioplasty: a randomized clinical trial. Surgery 1996;119:552–557.

Zieren J, Zieren HU, Jacobi CA,Wenger FA, Muller JM. Prospective randomized study comparing laparoscopic and open tension-free inguinal hernia repair with Shouldice's operation. Am J Surg 1998;175:330–333.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 119–119.

PARAMETRIC AND SEMIPARAMETRIC BAYESIAN APPROACH TO META-ANALYSIS WITH APPLICATION TO MEDICAL DATA

Pulak Ghosh
Department of Quantitative sciences, Indian Institute of Management,
Bangalore, India
E-mail: pulakghosh@gmail.com

ABSTRACT

Meta-analysis is a statistical procedure that integrates the results of several independent studies considered to be combinable. This paper develops Bayesian methods for testing treatment effects in a meta-analytical framework. Particular attention will be given to develop methods for meta-analysis but relaxing the standard distributional assumptions. The use of a Bayesian approach in evaluating data from clinical trials with many treatment centers is also explored. In particular, different studies in a metaanalysis may involve different treatment comparison, while centers within the same trial usually consider the same treatment. We show the advantage of our method using simulation and a real data set.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 119–119.

SOME METHODOLOGIES FOR COMBINING TESTS AND CONFIDENCE INTERVALS FROM INDEPENDENT STUDIES

K. Krishnamoorthy
University of Louisiana
Lafayette, LA, USA
E-mail: krishna@louisiana.edu

ABSTRACT

In this talk, we shall address the problem of combining the results of independent investigations for the purpose of a common objective. This problem arises, for example, when two or independent agencies involved in measuring the effect of a new drug or when different laboratories are used to measure the toxic waste in a river. Assuming normal models, we shall present methods for combining independent test results and methods for combining the results to find confidence intervals. The properties of the statistical methods will be discussed, and the methods will be illustrated using a few practical examples. Furthermore, we shall consider this problem in the context of simple linear regression models.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 120–120.

COMPARISON OF META-ANALYSIS USING LITERATURE AND USING INDIVIDUAL PATIENT DATA

Thomas Mathew
Department of Mathematics and Statistics,
University of Maryland Baltimore County,
Baltimore, Maryland 21250, USA
E-mail: mathew@umbc.edu

Kenneth Nordstrom
Department of Mathematical Sciences,
University of Oulu, Finland

ABSTRACT

The problem of combining information from separate studies is a key consideration when performing a meta-analysis, or planning a multicenter trial. Although there is a considerable journal literature on summary versus individual patient data, recent articles in the medical literature indicate that there is still confusion and uncertainty as to the precision of an analysis based on aggregate data. In this paper we address this issue by considering the estimation of a linear function of the mean, based on linear models for individual patient data. The setup, which allows for the presence of random effects and covariates in the model, is quite general, and includes many of the commonly employed models. The one-way fixed-effects model and the two-way model without interaction and fixed or random study effects are all obtained as special cases. For this general model we derive a condition for the estimator based on summary data to coincide with the one obtained from individual patient data, extending considerably earlier work. The implications of this result for the specific models mentioned above are illustrated in detail, both theoretically and in terms of two real data sets, and the role of balance is highlighted.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 121–131.

BENEFITS OF EARLY FEEDING VERSUS TRADITIONAL NIL-BY-MOUTH NUTRITIONAL POSTOPERATIVE MANAGEMENT IN GASTROINTESTINAL RESECTIONAL SURGERY PATIENTS: A META-ANALYSIS

Emma Osland
Dept of Nutrition and Dietetics, Ipswich Hospital, Ipswich, Queensland, Australia
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments, University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: Emma_Osland@health.qld.gov.au

Rossita Mohamad Yunus
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments, University of Southern Queensland, Toowoomba, Queensland, Australia
Institute of Mathematical Sciences, University of Malaya, Kuala Lumpur, Malaysia
E-mail: Rossita.MuhamadYunus@usq.edu.au

Shahjahan Khan
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments, University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: Shahjahan.Khan@usq.edu.au

Muhammed Ashraf Memon
Mayne Medical School, School of Medicine, University of Queensland, Brisbane, Queensland, Australia
Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia
Faculty of Health Sciences, Bolton University, Bolton, Lancashire, UK
E-mail: mmemon@yahoo.com

ABSTRACT

The objective of the current work was to conduct a meta-analysis of randomized controlled trials evaluating the effect on surgical outcomes of providing nutrition within 24-hours following gastrointestinal surgery compared with traditional postoperative management. A literature search was conducted to identify randomized controlled trials published in English language between1966 and 2007 comparing the outcomes of early and traditional postoperative feeding. All trials involving resection of the portions of the gastrointestinal tract followed by patients receiving nutritionally significant oral or enteral intake within 24-hours after surgery were included for analysis. Random effects meta-analyses were performed. Outcome variables analyzed were complications, mortality, anastomotic dehiscence, nasogastric reinsertion, days to passing flatus, days to first bowel motion, and length of stay. Fifteen studies (n=1240 patients) were analyzed. A statistically significant forty-five percent reduction in relative odds of total postoperative complications were seen in patients receiving early postoperative feeding (OR 0.55 CI 0.35, 0.87, p=0.01). No effect of early feeding was seen with relation to anastomotic dehiscence (OR 0.75, CI 0.39, 1.4, p=0.39), mortality (OR 0.71, CI 0.32,1.56, p= 0.39), resumption of bowel function as evidenced by days to passage of flatus (WMD -0.42, CI -1.12, 0.28, p=0.23) and first bowel motion (WMD -0.28, CI -1.20, 0.64, p=0.55), or reduced length of

stay (WMD -1.28, CI -2.94, 0.38, p=0.13). Similarly, nasogastric tube reinsertion was less common in traditional feeding interventions, however this was not statistically significant (OR 1.48, CI 0.93, 2.35, p=0.10). Early provision of nutritionally significant oral or enteral intake is associated with a significant reduction in reported total complications when compared with traditional postoperative feeding practices and does not negatively affect outcomes such as mortality, anastomotic dehiscence, resumption of bowel function or hospital length of stay. For these reasons, surgeons should be confident in adopting early feeding as part of standard practice for elective gastrointestinal surgery.

Keywords: Meta-analysis; gastrointesintal surgery; early feeding; surgical complications; human

1. INTRODUCTION

Traditional nil-by-mouth nutritional management of patients in the days following gastrointestinal resectional surgery has been adopted over the years in the belief that it decreases the risk of nausea, vomiting, aspiration pneumonia and anastomotic dehiscence (Nelson, Edwards and Tse, 2008). However, a growing number of well designed randomized controlled clinical trials suggest that it is safe to commence feeding from within 24-hr following surgery. Moreover three meta-analyses on this topic have been published (Andersen, Lewis and Thomas, 2006; Lewis, Andersen and Thomas, 2008; Lewis et al., 2001) which lend further support to the practice of early postoperative feeding. However aspects of nutritional provision that may impact surgical outcomes, such as the location of delivery and composition of nutritional provision, have been left largely unaddressed in these previous meta-analyses. Therefore this present meta-analysis has been undertaken to address these issues and develop a better understanding of the risks and benefits of early feeding when compared to the traditional approach following gastrointestinal resectional surgery.

2. MATERIALS AND METHODS

Electronic databases (Medline, Pubmed, EMBASE, CINAHL, Cochrane Register of Systematic Reviews, Science Citation Index) were cross-searched using search terms customized to each search engine in an attempt to detect relevant English language papers comparing the outcomes of early postoperative feeding in resectional surgery with traditional postoperative nutritional management. Reference lists of review papers and existing meta-analyses were hand searched for further appropriate citations.

All studies comparing early feeding and traditional (nil-by-mouth) postoperative nutritional management published in the English language were reviewed. Only randomized controlled trials with primary comparisons between early and traditional feeding practices were considered for inclusion. Studies must also have reported on clinically relevant outcomes, and have been conducted in adult (>18 years) elective resectional surgical cases in which early feeding was provided proximal to the anastomosis. Unpublished studies and abstracts presented at national and international meetings were excluded. Similarly duplicate publications were also excluded.

Early feeding was defined as the provision of nutritionally significant oral or enteral nutrition via nasogastric or jejunal feeding tube, provided within 24-hours postoperatively. Examples of nutritionally significant oral nutrition include free fluids or standard hospital diet; clear fluids were not included due to their inability to meet nutritional requirements irrespective of volume consumed (Hancock, Cresci and Martindale, 2002). Traditional postoperative management was defined as withholding nutritional provision until bowel function had resumed, as evidenced by either passage of flatus or bowel motion. Exclusion criteria included use of immune modulating enteral feed products such as Oral Impact® (Nestle Healthcare Nutrition, Minneapolis, USA) as these may independently improve postoperative outcomes in some patient populations (Zheng et al., 2007), early feeding provided distal to the anastomosis, use of parenteral nutrition in either interventional group, patients <18 years of age and non-resectional or emergency surgeries. Data extraction and critical appraisal were carried out by two authors (EO and MAM) for compliance with inclusion criteria and methodological quality. Standardized data extraction forms were used by both authors to independently and blindly summarize all the data available in the randomized controlled trials meeting the inclusion criteria. The authors were not blinded to the source of the document or authorship for the purpose of data extraction. The data were compared and discrepancies were addressed with discussion until consensus was achieved.

Evaluation of methodological quality of identified studies was conducted using the Jadad scoring system which provides a numerical quality score based on reporting of randomization, blinding and reporting of withdrawals (Jadad et al., 1996).

Outcomes assessed were those considered to exert influence over practical aspects of surgical practice and policy decisions within institutions such as rates of postoperative complications and mortality outcomes. All studies with reporting on any number of outcomes of this nature were considered and final analyses were run on outcome parameters where numbers were sufficient to allow statistical analysis. Where required, authors were contacted for clarification of data or additional information.

Meta-analyses were performed using odds ratios (ORs) for binary outcomes and weighted mean differences (WMDs) for continuous outcome measures. A slightly amended estimator of OR was used to avoid the computation of reciprocal of zeros among observed values in the calculation of the original OR (Agresti, 1996). Random effects models, developed by using the inverse variance weighted method approach (Sutton et al., 2000), were used to combine the data. Heterogeneity among studies was assessed using the Q statistic proposed by Cochran (Cochran, 1954; Hedges and Olkin, 1985; Sutton et al., 2000) and I^2 index introduced by Higgins and Thompson (Higgins and Thompson, 2002; Huedo-Medina et al., 2006). If the observed value of Q is equal to or larger than the critical value at a given significant level (α), in this case 0.05, we conclude that the outcome variable is statistically significant. The drawback of the Q statistic is that its statistical power depends on the number of studies. The I^2 statistic describes the proportion of variation across studies that is due to between-studies heterogeneity rather than chance and unlike Q statistic it does not inherently depend upon the number of studies considered (Huedo-Medina et al., 2006).

The issue of heterogeneity was further explored based on year of publication i.e. year of publication: before and after 2000. For the computations of the confidence intervals estimates of mean and standard deviation are required. However, some of the published clinical trials did not report the mean and standard deviation, but rather reported the size of the trial, the median and range. From these available statistics, estimates of the mean and standard deviation were obtained using formulas proposed by Hozo et al. (2005). Funnel plots were synthesized in order to determine the presence of publication bias in the meta-analysis. Standard error was plotted against the treatment effects (Log OR for the dichotomous and WMD for continuous variables respectively) (Egger et al., 1997; Sutton et al., 2000; Tang and Liu, 2000) to allow 95% confidence interval limits to be displayed. All estimates were obtained using computer programs written in R (Hornik, 2008). All plots were obtained using the 'rmeta' package (Lumley, 2008). In the case of tests of hypotheses, the paper reports p-values for different study variables. In general, the effect is considered to be statistically significant if the p-value is small. If one uses a 5% significance level then the effect is significant only if the associated p-value is less than or equal to 5%.

3. RESULTS

Cross searching of the electronic databases yielded 87 unique abstracts of potential relevance which were retrieved for independent review. Figure 1 presents the results of the study selection following the Quality of Reporting of Meta-analyses (QUOROM) recommendations (Moher et al., 1999). Pooled results yielded 1240 patients, with a near even distribution between feeding interventions (*n*=617 traditional postoperative management, *n*=623 early post operative feeding) from 15 studies dating from 1979 to 2007. A summary of the randomised controlled trials included in the final meta-analysis are presented in Table 1.

Publication bias is one of the major criticisms of meta-analysis as its validity is reliant on a thorough representation of eligible studies being located (Higgins and Green, 2006; Ng et al., 2006; Sutton and Higgins, 2008; Tang and Liu, 2000). Funnel plots demonstrate symmetry for all outcomes except 'total complications'. This suggests publication bias occur within this meta-analysis in the total complications outcome, but is absent from the other assessed variables (Hedges and Olkin, 1985; Higgins and Thompson, 2002). However the number of studies included in the funnel plots are inadequate to sensitively detect a study bias (Hedges and Olkin, 1985; Huedo-Medina et al., 2006).

None of the 15 included studies achieved a modified Jadad score of over three (range 1 to 3, median 2). Six studies described the method of randomization (Carr et al., 1996; Han-Geurts et al., 2007; Han-Geurts et al., 2001; Nessim et al., 1999; Stewart et al., 1998), six reported on withdrawals (Carr et al., 1996; Han-Geurts et al., 2007; Ortiz, Armendariz and Yarnoz, 1996; Ryan, Page and Babcock, 1981; Schroeder et al., 1991; Stewart et al., 1998), and one study (Beier-Holgersen and Boesby, 1996) reported on blinding. Jadad scores are reported in Table 1.

Sufficient data were available for the analysis for seven clinically relevant outcomes: total complications (defined as any complication reported within the postoperative period, excluding mortality and nausea/vomiting); anastomotic dehiscence; in-hospital mortality; days to passage of bowel motion; days to passage of flatus; length of hospital stay; and nasogastric reinsertion.

A statistically significant forty-five percent reduction in relative odds of total postoperative complications were observed in patients receiving early postoperative feeding (OR 0.55 CI 0.35, 0.87, p=0.01, Q=29.07, p=0.01, I^2 = 51.8%, CI 13, 73%). Early feeding was not associated with significant effects on anastomotic dehiscence (OR 0.75, CI 0.39, 1.4, p=0.39, Q=3.31, p=0.99 , I^2 = 0%, CI 0, 0%), mortality (OR 0.71, CI 0.32,1.56, p= 0.39, Q=4.24 , p= 0.99, I^2 = 0%, CI 0, 0%), resumption of bowel function as evidenced by days to passage of flatus (WMD -0.42, CI -1.12, 0.28, p=0.23, Q=75.6 , p<0.0001, I^2 = 96%, CI 29, 98%) and first bowel motion (WMD-0.28, CI -1.20, 0.64, p=0.55, Q=79 , p<0.001, I^2 = 96%, CI 93, 98%), and reduced length of stay (WMD -1.28, CI -2.94, 0.38, p=0.13, Q=61 , p<0.001, I^2 = 85%, CI 75, 91%). A non-statistically significant reduction in the odds of requiring nasogastric tube reinsertion was seen for traditional feeding practices (OR 1.48, CI 0.93, 2.35, p=0.10, Q= 3.24, p=0.86, I^2 = 0%, CI 0, 30%).

The intervention effects of early postoperative feeding were more pronounced in pre-2000 studies when compared with those conducted post-2000 for the parameters of postoperative complications, mortality, anastomotic dehiscence, days to passage of flatus and first bowel motion, and length of hospital stay. Only studies pre-2000 reported on incidence of nausea and vomiting, with no significant differences observed between intervention groups (OR 0.93, CI 0.53, 1.65, p=0.8). Sample forest plots are presented in Figures 1 and 2.

Table 1 – Summary of included studies

Study	Patient population	n (trad early)	Jadad Score	Early feeding protocol
Sagar et al. (1979)	Major intestinal surgery – oesophagogastrectomy (n=2), gastrectomy (n=6), colectomy, anterior resection, abdominoperineal resection	15/15	1	½ strength Flexical (elemental feed product) @ 25ml/hr for 24hrs D1 post op, full strength Flexical @ 25ml/hr for 24hrs D2 post op, full strength Flexical @ 50ml/hr for 24hrs D3 post op, full strength Flexical @ 100mL/hr D4 post op via jejunal port of nasogastric/jejunal tube
Ryan et al. (1981)	Partial colectomy	7/7	2	Vivonex HN (elemental feed product) 10% w/v @ 50mL/hr on day of operation, 10% w/v @100mL/hr D1 post op, 10% w/v @125mL/hr D2, 15% w/v @ 125mL/hr D3, 20% w/v @ 125mL/hr D4, 20% w/v @ 125mL/hr D5, 25% w/v @ 125mL/hr D6 & D7
Schroeder et al. (1991)	Small or large bowel resections or reanastomosis – colonic resection, abdominoperineal resection, ileoanal J pouch, small bowel resection	16/16	2	50mL/hr Osmolite day of operation, 80mL/hr Osmolite if tolerated thereafter. Oral intake D3 post op
Binderow et al. (1994)	Laparoscopic assisted Laparotomy with colonic or ileal resection	32/32	1	Regular diet from D1 post op
Beier-Holgersen et al. (1996)	Gastrointestinal disease treated with bowel resection with anastomosis, enterostomy, gastric (n=5) or oesophageal resection (n=3).	30/30	2	Clear fluids orally + increasing volumes of nutridrink via nasojejunal tube from day of surgery

Table 1 (Cont.) – Summary of included studies

Carr et al. (1996)	Unspecified intestinal surgery	14/14	3	Immediate post op nasojenunal feeding – 25ml/hr Fresubin (1kcal/mL) and increased by 25ml/hr q4h until individual caloric goals met
Ortiz et al. (1996)	Laparotomy for colon or rectal surgery	95/93	2	Clear fluids on day of surgery (?pre/post op), Regular diet from D1 post op
Hartsell et al. (1997)	Open colorectal surgery	29/29	1	Full liquid diet D1 post op, regular diet once tolerating >1L in 24hrs
Nessim et al. (1999)	Anorectal reconstructive surgery	27/27	2	Regular diet from D1 post op
Stewart et al. (1998)	Colorectal resection with anastomosis and without stoma formation	40/40	3	Free fluids from 4 hours post op on day of surgery, Regular diet from D1 post op
Han-Geurts et al. (2001)	Abdominal surgery (vascular + colonic)	49/56	2	Regular diet from D1 post op
Delaney et al. (2003)	segmental intestinal or rectal resection by laparotomy, including reoperation or pelvic surgery and those with comorbidities	33/31	2	Fluid diet D1 post op with regular diet in PM of D1 post op
Lucha et al. (2005)	Open colorectal surgery	25/26	1	Regular diet from 8hrs day of surgery
Zhou et al. (2006)	Excision and anastomosis for colorectal tumour	155/161	1	Liquid fibreless diet D1-3 post op
Han-Geurts et al. (2007)	Open colorectal surgery	50/46	3	Regular diet from D1 post op

Values in left panel are observed counts for early and traditional feeding, odds ratio and limits of 95% confidence intervals for odds ratio of the outcome variable. In the graph, squares indicate point estimates of treatment effect (odds ratio for early over traditional groups) with the size of the squares representing the weight attributed to each study. The horizontal lines represent 95% confidence interval for odds ratio of individual studies. The pooled estimate for the complication rate is the pooled odds ratio, obtained by combining all odds ratio of the 15 studies using the inverse variance weighted method. The 95% confidence interval for the pooled estimate is represented by the diamond and the length of the diamond depicts the width of the confidence interval. Values to the left of the vertical line favour early feeding.

Figure 1: Odds ratio for complication (nausea and vomiting excluded)

Study	Early	Traditional	OR	L	U
pre 2000					
Sagar	3 of 15	5 of 15	0.53	0.08	3.78
Ryan	2 of 7	7 of 7	0.03	0	0.94
Schroeder	4 of 16	7 of 16	0.46	0.07	2.91
Binderow	0 of 32	0 of 32	1	0.02	61.41
Beier-Holgersen	8 of 30	19 of 30	0.22	0.05	1.08
Carr	0 of 14	4 of 14	0.08	0	2.06
Ortiz	17 of 93	18 of 95	0.96	0.24	3.77
Hartsell	1 of 29	1 of 29	1	0.07	13.42
Nessim	3 of 27	4 of 27	0.75	0.11	5.01
Stewart	10 of 40	12 of 40	0.78	0.17	3.56
subtotal	**48 of 303**	**77 of 305**	**0.55**	**0.34**	**0.9**
post 2000					
Han-Geurts	12 of 56	13 of 49	0.76	0.18	3.27
Delaney	7 of 31	10 of 33	0.69	0.14	3.38
Lucha	1 of 26	1 of 25	0.96	0.07	12.99
Zhou	23 of 161	70 of 155	0.21	0.06	0.74
Han-Geurts	22 of 46	20 of 50	1.37	0.33	5.61
subtotal	**65 of 320**	**114 of 312**	**0.62**	**0.26**	**1.51**
POOLED	**113 of 623**	**191 of 617**	**0.55**	**0.35**	**0.87**

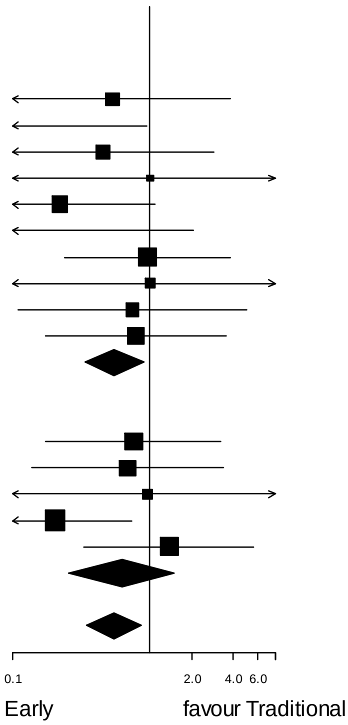
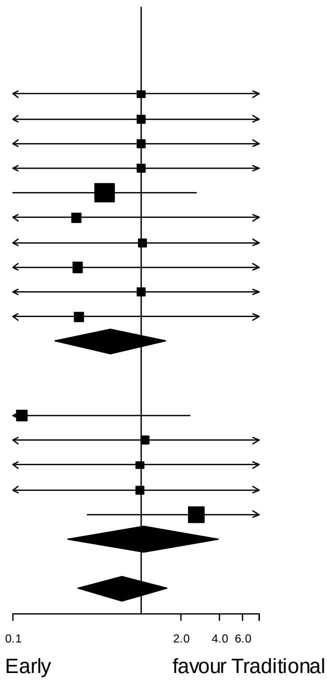


Figure 2: Odds ratio for mortality

Study	Early	Traditional	OR	L	U
pre 2000					
Sagar	0 of 15	0 of 15	1	0.02	53.66
Ryan	0 of 7	0 of 7	1	0.02	57.31
Schroeder	0 of 16	0 of 16	1	0.02	53.46
Binderow	0 of 32	0 of 32	1	0.02	51.94
Beier-Holgersen	2 of 30	4 of 30	0.52	0.1	2.65
Carr	0 of 14	1 of 14	0.31	0.01	8.29
Ortiz	0 of 93	0 of 95	1.02	0.02	52.01
Hartsell	0 of 29	1 of 29	0.32	0.01	8.24
Nessim	0 of 27	0 of 27	1	0.02	52.22
Stewart	0 of 40	1 of 40	0.33	0.01	8.22
subtotal	**2 of 303**	**7 of 305**	**0.58**	**0.22**	**1.54**
post 2000					
Han-Geurts	0 of 56	3 of 49	0.12	0.01	2.33
Delaney	0 of 31	0 of 33	1.06	0.02	55.24
Lucha	0 of 26	0 of 25	0.96	0.02	50.35
Zhou	0 of 161	0 of 155	0.96	0.02	48.83
Han-Geurts	3 of 46	1 of 50	2.66	0.38	18.77
subtotal	**3 of 320**	**4 of 312**	**1.03**	**0.27**	**3.88**
POOLED	**5 of 623**	**11 of 617**	**0.71**	**0.32**	**1.56**



Values in left panel are observed counts for early and traditional feeding, odds ratio and limits of 95% confidence intervals for odds ratio of the outcome variable. In the graph, squares indicate point estimates of treatment effect (odds ratio for early over traditional groups) with the size of the squares representing the weight attributed to each study. The horizontal lines represent 95% confidence interval for odds ratio of individual studies. The pooled estimate for the mortality rate is the pooled odds ratio, obtained by combining all odds ratio of the 15 studies using the inverse variance weighted method. The 95% confidence interval for the pooled estimate is represented by the diamond and the length of the diamond depicts the width of the confidence interval. Values to the left of the vertical line favour early feeding.

4. COMMENTS AND CONCLUSION

This meta-analysis re-enforces previous findings that traditional postoperative feeding practices confer no benefit in terms of outcomes following gastrointestinal resectional surgery (Andersen et al., 2006; Lewis et al., 2008; Lewis et al., 2001). Our pooled findings suggest that a statistically significant reduction in total postoperative complications following surgery is associated with the introduction of nutritionally significant food or fluid within 24-hours postoperatively: this is the first meta-analysis to demonstrate this. In the stratified results most outcomes observed (total complications, mortality, anastomotic dehiscence, days to passage of flatus and bowel motion, length of stay) results were seen to more strongly favour early feeding in the pre-2000 subgroup than in the post-2000 studies. This may be explained by the greater statistical power present in the pre-2000 subgroup due to the larger number of studies (k=10 vs k=5), however this does not explain the effect for all variables, specifically the measure of bowel function return. Therefore numbers alone may not account for these differences. Other possible explanations include the possibility of a greater quantity of nutrition provided to patients in the pre-2000 by virtue of the higher number of studies utilizing tube feeding rather than voluntary oral intake; differences arising from changes to perioperative practice over the 28 years encompassed by the included studies; or unexplained differences in results from the two Dutch studies included in the post-2000 subgroup.

There are a number of limitations associated with this meta-analysis. Firstly in an attempt to standardize the differences in reporting between articles, we contacted several authors for clarification of reported data or additional information within their published data. In cases where no response was returned (Delaney et al., 2003; Hartsell et al., 1997; Stewart et al., 1998; Zhou et al., 2006) assumptions relating to the interpretation of various aspects of their published reports were made, such as the composition of the fluid diets reported (Delaney et al., 2003; Hartsell et al., 1997; Zhou et al., 2006), or discrepancies in the reporting within the paper (Stewart et al., 1998). Therefore while every attempt has been made to ensure analysed studies meet inclusion criteria and that other data are accurate, there may still be inconsistencies between the studies included.

Secondly the studies that met inclusion criteria for this meta-analysis consistently yielded poor scores for methodological quality using the Jadad scoring system (Jadad et al., 1996). Out of a possible score of five, a mean score of 1.9 was achieved, with a maximum score of three. Even with the increasing emphasis on improving the quality of reporting in clinical trials in the medical literature in recent years, no difference was seen in the Jadad score in the average pre- and post-2000 scores (pre 2.0, post 1.8, NS).

Thirdly, there currently exist some limitations to the application of meta-analysis to both surgical and nutrition research at the present time. Examples of areas where more statistical research and modelling are required as highlighted by the current work include: (i) improved methods to detect publication bias, particularly when random effects models of meta-analysis are applied, and where the meta-analysis is comprised of a small number of studies, (ii) development of tests for heterogeneity with improved sensitivity to detect between-study variation in circumstances where small numbers of studies are involved, (iii) empirical investigation into the effect of assuming normal distribution during the application of random effects model of meta-analysis, (iv) guidance on investigation of heterogeneity in the circumstances where a small number of studies make subgroup analysis, meta-regression and other methods of sensitivity analysis difficult or invalid, and (v) further investigation on the effect of methodological quality on the influence of effect size in areas of surgery and nutrition.

In conclusion the results of this meta-analysis fail to demonstrate merit in continuing the traditional postoperative feeding practices of withholding nutrition provided proximal to the anastomosis until bowel function is resumed. This is the first meta-analysis to demonstrate statistically significant reductions in total complications in the postoperative course with early feeding. Furthermore, no negative effect of early feeding was demonstrated with regard to in-hospital mortality, anastomotic dehiscence, length of stay and time to recovery of bowel function. For these reasons, surgeons should be confident in adopting early feeding as part of standard practice for elective gastrointestinal surgery.

ACKNOWLEDGEMENTS

We gratefully acknowledge and thank Dr Han-Guerts and her co-authors for allowing us access to their 2007 study data, and Drs Feo and Ortiz for their prompt replies to our enquiries.

REFERENCES

Agresti, A. (1996). *An introduction to Categorical Data Analysis*: Wiley & Sons.

Andersen, H. K., Lewis, S. J., and Thomas, S. (2006). Early enteral nutrition within 24h of colorectal surgery versus later commencement of feeding for postoperative complications. *Cochrane Database Syst Rev*, CD004080.

Beier-Holgersen, R., and Boesby, S. (1996). Influence of postoperative enteral nutrition on postsurgical infections. *Gut* 39, 833-835.

Binderow, S. R., Cohen, S. M., Wexner, S. D., and Nogueras, J. J. (1994). Must early postoperative oral intake be limited to laparoscopy? *Dis Colon Rectum* 37, 584-589.

Carr, C. S., Ling, K. D., Boulos, P., and Singer, M. (1996). Randomised trial of safety and efficacy of immediate postoperative enteral feeding in patients undergoing gastrointestinal resection. *BMJ* 312, 869-871.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics* 10, 101-129.

Delaney, C. P., Zutshi, M., Senagore, A. J., Remzi, F. H., Hammel, J., and Fazio, V. W. (2003). Prospective, randomized, controlled trial between a pathway of controlled rehabilitation with early ambulation and diet and traditional postoperative care after laparotomy and intestinal resection. *Dis Colon Rectum* 46, 851-859.

Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315, 629-634.

Han-Geurts, I. J., Hop, W. C., Kok, N. F., Lim, A., Brouwer, K. J., and Jeekel, J. (2007). Randomized clinical trial of the impact of early enteral feeding on postoperative ileus and recovery. *Br J Surg* 94, 555-561.

Han-Geurts, I. J., Jeekel, J., Tilanus, H. W., and Brouwer, K. J. (2001). Randomized clinical trial of patient-controlled versus fixed regimen feeding after elective abdominal surgery. *Br J Surg* 88, 1578-1582.

Hancock, S., Cresci, G., and Martindale, R. (2002). The clear liquid diet: when is it appropriate? *Curr Gastroenterol Rep* 4, 324-331.

Hartsell, P. A., Frazee, R. C., Harrison, J. B., and Smith, R. W. (1997). Early postoperative feeding after elective colorectal surgery. *Arch Surg* 132, 518-520; discussion 520-511.

Hedges, L. V., and Olkin, I. (1985). *Statistical methods for meta analysis*. Orlando: Academic Press.

Higgins, J. P., and Green, S. (eds) (2006). *Analysing and Presenting Results. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [updated September 2006]. Section 8. In: The Cochrane Library, Issue 4*Chichester: John Wiley & Sons, Ltd.

Higgins, J. P. T., and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21, 1539-1558.

Hornik, K. (2008). R: A language and environment for statistical computing. Vienna: Foundation for Statistical computing.

Hozo, S. P., Djulbegovic, B., and Hozo, I. (2005). Estimating the mean and variance from the median, range and size of a sample. *BMC Medical Research Methodology* 5, 13.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., and Botella, J. (2006). Assessing heterogeneity in meta analysis: statistic or index? . *The American Psychological Association* 11, 193 - 206.

Jadad, A. R., Moore, A., Carroll, D., *et al.* (1996). Assessing the Quality of Reports of Randomized Clinical Trials: Is Blinding Necessary? *Controlled Clinical Trials* 17, 1-12.

Lewis, S. J., Andersen, H. K., and Thomas, S. (2008). Early Enteral Nutrition Within 24h of Intestinal Surgery Versus Later Commencement of Feeding: A Systematic review and Meta-analysis. *J Gastrointest Surg*.

Lewis, S. J., Egger, M., Sylvester, P. A., and Thomas, S. (2001). Early enteral feeding versus "nil by mouth" after gastrointestinal surgery: systematic review and meta-analysis of controlled trials. *BMJ* 323, 773-776.

Lucha, P. A., Jr., Butler, R., Plichta, J., and Francis, M. (2005). The economic impact of early enteral feeding in gastrointestinal surgery: a prospective survey of 51 consecutive patients. *Am Surg* 71, 187-190.

Lumley, T. (2008). The rmeta Package Version 2.14.

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., and Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomised controled trials: the QUOROM statement. *The Lancet* 354, 1896-1900.

Nelson, R., Edwards, S., and Tse, B. (2008). Prophylactic nasogastric decompression after abdominal surgery *Cochrane Database Syst Rev*.

Nessim, A., Wexner, S. D., Agachan, F., *et al.* (1999). Is bowel confinement necessary after anorectal reconstructive surgery? A prospective, randomized, surgeon-blinded trial. *Dis Colon Rectum* 42, 16-23.

Ng, T. T., McGory, M. L., Ko, C. Y., and Maggard, M. A. (2006). Meta-analysis in surgery: methods and limitations. *Arch Surg* 141, 1125-1130; discussion 1131.

Ortiz, H., Armendariz, P., and Yarnoz, C. (1996). Is early postoperative feeding feasible in elective colon and rectal surgery? *Int J Colorectal Dis* 11, 119-121.

Ryan, J. A., Jr., Page, C. P., and Babcock, L. (1981). Early postoperative jejunal feeding of elemental diet in gastrointestinal surgery. *Am Surg* 47, 393-403.

Sagar, S., Harland, P., and Shields, R. (1979). Early postoperative feeding with elemental diet. *Br Med J* 1, 293-295.

Schroeder, D., Gillanders, L., Mahr, K., and Hill, G. L. (1991). Effects of immediate postoperative enteral nutrition on body composition, muscle function, and wound healing. *JPEN J Parenter Enteral Nutr* 15, 376-383.

Stewart, B. T., Woods, R. J., Collopy, B. T., Fink, R. J., Mackay, J. R., and Keck, J. O. (1998). Early feeding after elective open colorectal resections: a prospective randomized trial. *Aust N Z J Surg* 68, 125-128.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., and Song, F. (2000). *Methods for Meta-analysis in Medical Research*: John Wiley.

Sutton, A. J., and Higgins, J. P. (2008). Recent developments in meta-analysis. *Stat Med* 27, 625-650.

Tang, J. L., and Liu, J. L. (2000). Misleading funnel plot for detection of bias in meta-analysis. *J Clin Epidemiol* 53, 477-484.

Zheng, Y., Li, F., Qi, B., *et al.* (2007). Application of perioperative immunonutrition for gastrointestinal surgery: a meta-analysis of randomized controlled trials. *Asia Pac J Clin Nutr* 16 Suppl 1, 253-257.

Zhou, T., Wu, X. T., Zhou, Y. J., Huang, X., Fan, W., and Li, Y. C. (2006). Early removing gastrointestinal decompression and early oral feeding improve patients' rehabilitation after colorectostomy. *World J Gastroenterol* 12, 2459-2463.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 132–139.

PRELIMINARY RESULTS OF A META-ANALYSIS EVALUATING THE EFFECT OF IMMUNONUTRITION ON OUTCOMES OF ELECTIVE GASTROINTESTINAL SURGERY

Emma Osland
Dept of Nutrition and Dietetics, Ipswich Hospital, Ipswich, Queensland, Australia
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments,
University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: Emma_Osland@health.qld.gov.au

Md Belal Hossain
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments,
University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: hossainm@usq.edu.au

Shahjahan Khan
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments,
University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: Shahjahan.Khan@usq.edu.au

Muhammed Ashraf Memon
Mayne Medical School, School of Medicine, University of Queensland, Brisbane, Queensland, Australia
Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia
Faculty of Health Sciences, Bolton University, Bolton, Lancashire, UK
E-mail: mmemon@yahoo.com

ABSTRACT

Providing tube feeds or liquid oral supplements containing additional arginine and/or omega-3 fatty acids and RNA – often referred to as 'immunonutrition' and more recently 'pharmaconutrition' – has been proposed as a strategy to decrease postoperative complications and duration of hospital length of stay for elective surgical patients. However outcomes of individual studies vary and the purported benefits remain controversial. A meta-analysis on this topic was recently published (Zheng et al., 2007), however further randomised controlled trials on this topic have appeared in the literature since this time. This meta-analysis has been undertaken to update the previously published meta-analysis and to attempt to elucidate the potential benefit of providing immune-enhanced nutrition in surgical patients. A search of electronic databases identified all RCTs comparing the use of pre and/or perioperative immunonutrition with standard nutrition provision in elective adult surgical patients between 1990 and 2008. The meta-analysis was prepared in accordance with the Quality of Reporting of Meta-analyses (QUOROM) statement. The variables analysed included mortality, total complications, infective complications, anastomotic leak (where applicable), and length of hospital stay. 20 distinct studies were identified that met inclusion criteria involving 1966 patients (immunonutrition *n*= 1048; standard nutrition *n*= 918). The provision of immunonutrition was shown to be associated with significant reductions in the incidence of total postoperative complications (OR 0.59, CI 0.41, 0.83, p=0.0023, Q=51.5, p=0.0001, I^2 = 59%, CI 34.8%, 74.5%), postoperative infective complications (OR 0.49, CI 37, 64, p=0.0001, Q=25.2, p=0.1939 , I^2 = 23%, CI 0%, 53.3%), anastomotic breakdown (OR 0.51, CI 0.31, 0.84 , p=0.0085, Q= 6.31, p=0.03 , I^2 = 0%, CI 0%, 7.3%) when compared to standard nutritional provision. No effect of the differences in feed product formulation were seen with relation to mortality (OR 0.94, CI 0.49, 1.8, p=0.861, Q=4.49, p=0.99, I^2 = 0%, CI 0%, 0%) or length of hospital stay (WMD -2.52, CI -3.71, -1.33, p=0.0001, Q= 219, p=0.0001, I^2 = 90.2%, CI 87.5%, 93.4%). This meta-analysis lends strong support to the beneficial effects of immune-enhanced nutrition in the management of elective gastrointestinal surgical patients.

Keywords: Meta-analysis; surgery; immunonutrition; pharmaconutrition; surgical complications; human

1. INTRODUCTION

Nutrition provision is recognized to be an important aspect in the perioperative management of elective gastrointestinal surgery patients, and the timely provision of nutrition has been associated with improved postoperative outcomes (Lewis, Andersen and Thomas, 2009). While the benefits of nutritional provision in surgical patients are traditionally thought to arise from the provision of macronutrients such as calories for energy and protein for wound healing, other nutritional components obtained from food or artificial forms of nutrition support are now though to interact with the immune system and modulate the responses to conditions such a trauma, sepsis or surgery (Jones and Heyland, 2008). In view of this, during the early 1990s new artificial nutrition support formulas emerged in the literature and on the commercial market that contained higher quantities of nutrients such as arginine, glutamine, omega-3 fatty acids, and nucleotides: these are thought to enhance the body's immune response. Since that time, the effects of these formulations have been studied in a variety of patient populations including a variety of surgical specialties and in the critically ill (Dupertuis, Meguid and Pichard, 2009). Many of these studies have demonstrated conflicting results (Dupertuis et al., 2009). Further to these individual studies, five meta-analyses on this topic have been conducted, again with conflicting results depending on the patient population investigated. To date only one meta-analysis has been conducted on the use of immunonutrition in elective surgical patients (Zheng et al., 2007)

The current work has been undertaken in an attempt to further explore the literature on immune-enhancing nutritional formations – more recently termed 'pharmoconutrition' – specifically in the area of elective gastrointestinal surgery.

2. MATERIALS AND METHODS

Electronic databases (Medline, Pubmed, EMBASE, CINAHL, Cochrane Register of Systematic Reviews, Science Citation Index) were cross-searched using search terms customized to each search engine in an attempt to detect relevant papers comparing the outcomes of provision on immune-enhanced nutritional formulas with those of standard composition provided perioperatively in patients receiving elective surgery. Reference lists of review papers and existing meta-analyses were hand searched for further appropriate citations.

All studies comparing the provision of immune-enhanced nutritional formulations (commercial or experimental) with those of standard nutritional composition providing isocaloric and isonitrogenuous nutritional provision, published in the English language were reviewed. Only randomized controlled trials with primary comparisons between the different nutritional formulations were considered for inclusion. Studies must also have reported on clinically relevant outcomes, and have been conducted in adult (>18 years) elective gastrointestinal surgical cases. Additional exclusion criteria included studies that investigated the effect of parenteral provision supplemented with nutrients believed to be immune-enhancing, unpublished studies and abstracts presented at national and international meetings, and duplicate publications.

The meta-analysis was prepared in accordance with the Quality of Reporting of Meta-analyses (QUOROM) statement. Data extraction and critical appraisal of identified studies were carried out by two authors (EO and MAM) for compliance with inclusion criteria and methodological quality. The authors were not blinded to the source of the document or authorship for the purpose of data extraction. The data were compared and discrepancies were addressed with discussion until consensus was achieved.

Evaluation of methodological quality of identified studies was conducted using the Jadad scoring system which provides a numerical quality score based on reporting of randomization, blinding and reporting of withdrawals (Jadad et al., 1996).

Outcomes assessed were those considered to exert influence over practical aspects of surgical practice and policy decisions within institutions such as rates of postoperative complications and mortality outcomes. All studies with reporting on any number of outcomes of this nature were considered and final analyses were run on outcome parameters where numbers were sufficient to allow statistical analysis.

Meta-analyses were performed using odds ratios (ORs) for binary outcomes and weighted mean differences (WMDs) for continuous outcome measures. A slightly amended estimator of OR was used to avoid the computation of reciprocal of zeros among observed values in the calculation of the original OR (Agresti, 1996). Random effects models, developed by using the inverse variance weighted method approach were used to combine the data (Sutton et al., 2000). Heterogeneity among studies was assessed using the Q statistic proposed by Cochran (Cochran, 1954; Hedges and Olkin, 1985; Sutton et al., 2000) and I^2 index introduced by Higgins and Thompson (Higgins and Thompson, 2002). If the observed value of Q is equal to or larger than the critical value at a given significant level (α), in this case 0.05, we conclude that the outcome variable is statistically significant. The drawback of the Q statistic is that its statistical power depends on the number of studies. The I^2 statistic describes the proportion of variation across studies that are due to between-studies heterogeneity rather than chance and unlike Q statistic it does not inherently depend upon the number of studies considered (Huedo-Medina et al., 2006).

Funnel plots were synthesized in order to determine the presence of publication bias in the meta-analysis. Standard error was plotted against the treatment effects (Log OR for the dichotomous and WMD for continuous variables respectively) (Egger et al., 1997; Sutton et al., 2000; Tang and Liu, 2000) to allow 95% confidence interval limits to be displayed. All estimates were obtained using computer programs written in R (Hornik, 2008). All plots were obtained using the 'rmeta' package (Lumley, 2008). In the case of tests of hypotheses, the paper reports p-values for different study variables. In general, the effect is considered to be statistically significant if the p-value is small. If one uses a 5% significance level then the effect is significant only if the associated p-value is less than or equal to 5%.

3. RESULTS

Cross searching of the electronic databases yielded 81 unique abstracts of potential relevance which were retrieved for independent review. Of these, 18 studies met the inclusion criteria. Pooled results yielded 1956 patients, with a near even distribution between feeding interventions (n=1061 immune formulas, n=895 standard composition) from studies dating from 1995 to 2008. The protocol for provision of nutrition differed between studies: 11 studies studied the effects of immunonutrition postoperatively, three studies three operatively and five provided nutrition both pre and postoperatively. All patients receiving postoperative feeding were tube fed distal to the anastomosis within 24hrs post surgery for malignant disease. Five studies report that significant percentages (20-100%) of their study population were malnourished. A summary of the randomised controlled trials included in the final meta-analysis are presented in Table 1.

The included studies collectively demonstrate moderate methodological quality according to the Jadad score with an average score of 2.8 (out of 5), with a range of 1 to 5.

Sufficient data were available for the analysis for five clinically relevant outcomes: total complications; infective complication; anastomotic dehiscence; in-hospital mortality; and length of hospital stay.

Statistically significant reductions in the relative odds of total postoperative complications (OR 0.59 CI 0.41, 0.83, p=0.0023, Q=51.5, p=0.0001, I^2 = 59%, CI 34.8%, 74.5%), infective complications (OR 0.49, CI 37, 64, p=0.0001, Q=25.2, p=0.1939 , I^2=23%, CI 0%, 53.3%), and anastomotic dehiscence (OR 0.51, CI 0.31, 0.84 , p=0.0085, Q= 6.31, p=0.03 , I^2 = 0%, CI 0%, 7.3%) were observed in patients receiving immune- enhancing feeding products. No effect of the differences in feed product formulation were seen with relation to mortality (OR 0.94, CI 0.49, 1.8, p=0.861, Q=4.49, p=0.99,I^2=0%, CI 0%,s 0%) or length of hospital stay (WMD -2.52, CI -3.71, -1.33, p=0.0001, Q= 219, p=0.0001, I^2 = 90.2%, CI 87.5%, 93.4%).

Figure 1 – Forest plot of outcomes for infective complications. Forest plot draws the 95% confidence intervals for treatment effects (odds ratio) as horizontal lines. Confidence intervals show arrows when they exceed specified limits. In the forest plot, squares indicate the estimated treatment effects (odds ratio for immune-nutrition over normal feed groups) with the size of the squares representing the weight attributed to each study. The pooled estimated odds ratio is obtained by combining all the odds ratios of the studies using the inverse weighted method, represented by the diamond and the size of the diamond depicts the 95% confidence interval. Values to the left of the vertical line are at one favour immune-nutrition. Values to right of the vertical line are at one favour Normal feed.

Figure 2 – Forest plot of outcomes for LOS. Forest plot draws the 95% confidence intervals for treatment effects (WMD) as horizontal lines. Confidence intervals show arrows when they exceed specified limits. In the forest plot, squares indicate the estimated treatment effects (WMD for immune-nutrition over normal feed groups) with the size of the squares representing the weight attributed to each study. The pooled estimated odds ratio (WMD) is obtained by combining all the WMD of the studies using the inverse weighted method, represented by the diamond and the size of the diamond depicts the 95% confidence interval. Values to the left of the vertical line are at one favour immune-nutrition. Values to right of the vertical line are at one favour Normal feed.

Funnel plots demonstrate symmetry for mortality, infective complications and anastomotic dehiscence outcomes suggesting the absence of publication bias within the meta-analyses performed for these outcomes. The presence of publication bias is indiciated by asymmetric funnel plots for total complications and length of hospital stay.

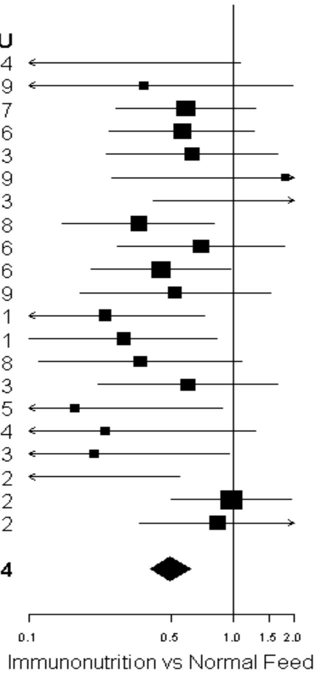
4. COMMENTS AND CONCLUSION

This meta-analysis demonstrates clinically and statistically signficiant reductions in the relative odds of developing postoperative complications (total, infective and that of anastomotic dehiscence) in patients receiving immune-enhancing nutrition formulations following elective gastrointestinal surgery. No effect was seen on postoperative mortality or length of hospital stay.

Sample forest plots are presented in Figures 1 and 2.

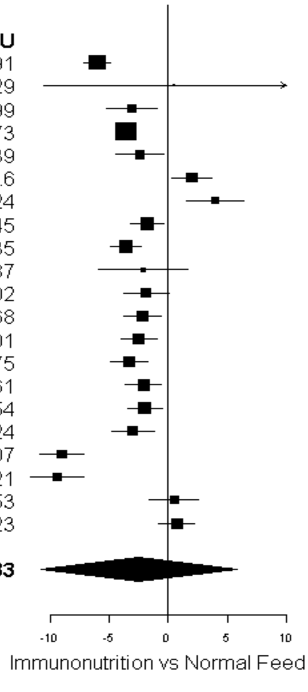
Study	Int n/N	Normal n/N	OR	L	U
Daly et al 95	0/30	7/30	0.055	0.003	1.04
Schilling et al 96	3/14	6/14	0.36	0.07	1.9
Gianotti et al 97	13/87	20/87	0.59	0.27	1.27
Senkal et al 97	12/77	19/77	0.56	0.25	1.26
Braga CCM	9/55	13/55	0.63	0.24	1.63
McCarter Arg	4/14	2/11	1.8	0.26	12.29
McCarter omega 3	5/13	2/11	2.81	0.42	18.73
Braga et al 99	9/85	22/86	0.34	0.14	0.8
Senkal et al 99	9/78	12/76	0.69	0.27	1.76
Snyderman et al 99	19/82	19/47	0.44	0.2	0.96
Gianotti et al 00	6/71	11/73	0.52	0.18	1.49
Braga et al peri 02	5/50	16/50	0.24	0.07	0.71
Braga et al pre 02	6/50	16/50	0.29	0.1	0.81
Braga et al pre/peri 02	5/50	12/50	0.35	0.11	1.08
Braga et al pre/std peri 02	8/50	12/50	0.6	0.22	1.63
Farreras et al 05	2/30	9/30	0.17	0.03	0.85
Xu et al 06	2/30	7/30	0.23	0.04	1.24
Giger et al 07 I-2/7	5/17	10/15	0.21	0.04	0.93
Giger et al 07 I-5/7	2/14	10/15	0.08	0.01	0.52
Klek clin nutr 08	23/92	23/91	0.98	0.5	1.92
Klek ann surg 08	13/52	15/53	0.84	0.35	2
Meta-analysis			**0.49**	**0.37**	**0.64**

Q=25.2 df=19 p=0.1939 I^2=23%



Study	Int mean(sd)	Normal mean(sd)	WMD	L	U
Daly et al 95	16(0.9)	22(2.9)	-6	-7.09	-4.91
Schilling et al 96	14.5(8)	14(19)	0.5	-10.3	11.29
Gianotti et al 97	16.1(6.2)	19.2(7.9)	-3.1	-5.21	-0.99
Senkal et al 97	27(2.3)	30.6(3.1)	-3.6	-4.46	-2.73
Braga CCM	13.7(4.8)	16.1(5.9)	-2.4	-4.41	0.39
McCarter Arg	15(2.4)	13(1.7)	2	0.39	3.6
McCarter omega 3	17(3.7)	13(1.7)	4	1.75	6.24
Braga et al 99	11.1(4.4)	12.9(4.6)	-1.8	-3.14	-0.45
Senkal et al 99	22.2(4.1)	25.8(3.8)	-3.6	-4.84	-2.35
Snyderman et al 99	15.3(9.1)	17.4(11.9)	-2.1	-6.07	1.87
Gianotti et al 00	15.1(5.4)	17.0(6.1)	-1.9	-3.78	-0.02
Braga et al peri 02	9.8(3.1)	12(4.5)	-2.2	-3.71	-0.68
Braga et al pre 02	9.5(2.9)	12(4.5)	-2.5	-3.98	-1.01
Braga et al pre/peri 02	12.0(3.8)	15.3(4.1)	-3.3	-4.84	-1.75
Braga et al pre/std peri 02	13.2(3.5)	15.3(4.1)	-2.1	-3.59	-0.61
Farreras et al 05	13(2.75)	15(3)	-2	-3.45	-0.54
Xu et al 06	9(3.2)	12(3.7)	-3	-4.75	-1.24
Giger et al 07 I-2/7	14.1(1.3)	23.1(3.6)	-9	-10.92	-7.07
Giger et al 07 I-5/7	13.7(2.3)	23.1(3.6)	-9.4	-11.58	-7.21
Klek clin nutr 08	12.9(8)	12.4(5.9)	0.5	-1.53	2.53
Klek ann surg 08	13.1(4.1)	12.4(3.9)	0.7	-0.83	2.23
Meta-analysis			**-2.52**	**-3.71**	**-1.33**

Q=219.57 df=20 p=0.0001 I^2=90.9%



The results of this meta-analysis differ somewhat from the previous meta-analysis on this topic. Zheng et al. (2007) reported signfiicant reductions in infective complications and duration of hospital stay, and no effect on mortality. Investigation of the effect of immunonutrition on total complications or anastomotic dehiscence was not conducted. They also demonstrated an improvement on measures of immune function such as total lymphocytes, CD4 levels, IgG levels and IL6 levels with the provision of immunonutrition. While the latter cluster of immunological outcomes have not been analysed in the current work, it is interesting to note the difference in the length of stay results between this and the Zheng et al. (2007) analysis. The most likely explanation for the difference between outcomes lie in the included studies – the only study included in the 2007 meta-analysis not to report LOS benefits with early feeding was a comparitively by Schilling (1996). Furthermore, this study would have little effect on the pooled outcomes in view of the fixed effects model used and its small study size. In contrast, the present work includes four studies (including two available since the publication of the Zheng et al. (2007) paper) that report increased length of study with the provision of immunonutrition, and utislises a random effects model of meta-analysis that weights smaller studies more heavily than occurs in the fixed effects model.

This is a prelimiary assessment of the data available. To gain a more thorough understanding of the benefits and risks associated with the provision of immunonutrition when compared to standard formulations in this population further analysis of the data is required. For example, as different formulas with different immune-enhancing componant levels have been used (ie Impact®, Stressor®, experimental solutions, etc) stratifying the studies for re-analysis by arginine content may demonstrate differences in outcomes. Similarly, as the presence of malnutrition is known to affect the outcomes of nutritional provision, a sensitivity analysis should be run assessing the impact of this factor on the outcomes seen. Finally, the different protocols used in the 20 studies included in this analysis introduce considerable variablity to the timing and quantity of the immunonutrition provided to the patients involved: these differences should be controlled for to ensure accurate interpretation of the reported outcome data.

In conclusion, this meta-analysis lends support to the beneficial effects of immune-enhanced nutrition in the management of elective gastrointestinal surgical patients. Provision of immunonutrition in preference to standard nutrition appears on preliminary assessment to be associated with statistically and clinically significant reductions in the development of postoperative complications including those of infective origin, anastomotic dehiscence in patients who receive a primary anastomosis, does not show the detrimental mortality outcomes believed to be associated with the use immunonutrition in critically ill populations. It does not, however, concur with previous meta-analysis findings of reduced length of hospital stay associated with the provision of immunonutrition when compared to standard nutritional formulations.

REFERENCES

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*: Wiley & Sons.

Braga, M., Gianotti, L., Nespoli, L., Radaelli, G., and Di Carlo, V. (2002a). Nutritional approach in malnourished surgical patients: a prospective randomized study. *Arch Surg* 137, 174-180.

Braga, M., Gianotti, L., Radaelli, G., *et al.* (1999). Perioperative immunonutrition in patients undergoing cancer surgery: results of a randomized double-blind phase 3 trial. *Arch Surg* 134, 428-433.

Braga, M., Gianotti, L., Vignali, A., and Carlo, V. D. (2002b). Preoperative oral arginine and n-3 fatty acid supplementation improves the immunometabolic host response and outcome after colorectal resection for cancer. *Surgery* 132, 805-814.

Braga, M., Gianotti, L., Vignali, A., Cestari, A., Bisagni, P., and Di Carlo, V. (1998). Artificial nutrition after major abdominal surgery: impact of route of administration and composition of the diet. *Crit Care Med* 26, 24-30.

Chen, D. W., Wei Fei, Z., Zhang, Y. C., Ou, J. M., and Xu, J. (2005). Role of enteral immunonutrition in patients with gastric carcinoma undergoing major surgery. *Asian J Surg* 28, 121-124.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics* 10, 101-129.

Daly, J. M., Weintraub, F. N., Shou, J., Rosato, E. F., and Lucia, M. (1995). Enteral nutrition during multimodality therapy in upper gastrointestinal cancer patients. *Ann Surg* 221, 327-338.

Dupertuis, Y. M., Meguid, M. M., and Pichard, C. (2009). Advancing from immunonutrition to a pharmaconutrition: a gigantic challenge. *Curr Opin Clin Nutr Metab Care* 12, 398-403.

Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315, 629-634.

Farreras, N., Artigas, V., Cardona, D., Rius, X., Trias, M., and Gonzalez, J. A. (2005). Effect of early postoperative enteral immunonutrition on wound healing in patients undergoing surgery for gastric cancer. *Clin Nutr* 24, 55-65.

Gianotti, L., Braga, M., Fortis, C., *et al.* (1999). A prospective, randomized clinical trial on perioperative feeding with an arginine-, omega-3 fatty acid-, and RNA-enriched enteral diet: effect on host response and nutritional status. *JPEN J Parenter Enteral Nutr* 23, 314-320.

Gianotti, L., Braga, M., Vignali, A., *et al.* (1997). Effect of route of delivery and formulation of postoperative nutritional support in patients undergoing major operations for malignant neoplasms. *Arch Surg* 132, 1222-1229; discussion 1229-1230.

Giger, U., Buchler, M., Farhadi, J., *et al.* (2007). Preoperative immunonutrition suppresses perioperative inflammatory response in patients with major abdominal surgery-a randomized controlled pilot study. *Ann Surg Oncol* 14, 2798-2806.

Hedges, L. V., and Olkin, I. (1985). *Statisical methods for meta-analysis*. Orlando: Academic Press.

Higgins, J. P., and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Stat Med* 21, 1539-1558.

Hornik, K. (2008). R: A language and environment for statistical computing. Vienna: Foundation for Statistical Computing.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., and Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods* 11, 193-206.

Jones, N. E., and Heyland, D. K. (2008). Pharmaconutrition: a new emerging paradigm. *Curr Opin Gastroenterol* 24, 215-222.

Klek, S., Kulig, J., Sierzega, M., *et al.* (2008a). Standard and immunomodulating enteral nutrition in patients after extended gastrointestinal surgery--a prospective, randomized, controlled clinical trial. *Clin Nutr* 27, 504-512.

Klek, S., Kulig, J., Sierzega, M., *et al.* (2008b). The impact of immunostimulating nutrition on infectious complications after upper gastrointestinal surgery: a prospective, randomized, clinical trial. *Ann Surg* 248, 212-220.

Lewis, S. J., Andersen, H. K., and Thomas, S. (2009). Early enteral nutrition within 24 h of intestinal surgery versus later commencement of feeding: a systematic review and meta-analysis. *J Gastrointest Surg* 13, 569-575.

Lumley, T. (2008). The rmeta Package version 2.14.

McCarter, M. D., Gentilini, O. D., Gomez, M. E., and Daly, J. M. (1998). Preoperative oral supplement with immunonutrients in cancer patients. *JPEN J Parenter Enteral Nutr* 22, 206-211.

Schilling, J., Vranjes, N., Fierz, W., *et al.* (1996). Clinical outcome and immunology of postoperative arginine, omega-3 fatty acids, and nucleotide-enriched enteral feeding: a randomized prospective comparison with standard enteral and low calorie/low fat i.v. solutions. *Nutrition* 12, 423-429.

Senkal, M., Mumme, A., Eickhoff, U., *et al.* (1997). Early postoperative enteral immunonutrition: clinical outcome and cost-comparison analysis in surgical patients. *Crit Care Med* 25, 1489-1496.

Senkal, M., Zumtobel, V., Bauer, K. H., *et al.* (1999). Outcome and cost-effectiveness of perioperative enteral immunonutrition in patients undergoing elective upper gastrointestinal tract surgery: a prospective randomized study. *Arch Surg* 134, 1309-1316.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., and Song, F. (2000). *Methods for Meta-analysis in Medical Meta-analysis*: John Wiley.

Tang, J. L., and Liu, J. L. (2000). Misleading funnel plot for detection of bias in meta-analysis. *J Clin Epidemiol* 53, 477-484.

Xu, J., Zhong, Y., Jing, D., and Wu, Z. (2006). Preoperative enteral immunonutrition improves postoperative outcome in patients with gastrointestinal cancer. *World J Surg* 30, 1284-1289.

Zheng, Y., Li, F., Qi, B., *et al.* (2007). Application of perioperative immunonutrition for gastrointestinal surgery: a meta-analysis of randomized controlled trials. *Asia Pac J Clin Nutr* 16 Suppl 1, 253-257.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 140–147.

META-ANALYSIS OF D1 VERSUS D2 GASTRECTOMY FOR GASTRIC ADENOCARCINOMA

Manjunath S Subramanya
Department of Surgery, Mount Isa Base Hospital, Mount Isa, Queensland, Australia
E-mail: manjunathbss9@yahoo.com

Md Belal Hossain
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments,
University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: bjoardar2003@yahoo.com

Shahjahan Khan
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments,
University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: khans@usq.edu.au

Breda Memon
Ipswich Hospital, Chelmsford Avenue, Ipswich, Queensland, Australia
E-mail: bmemon@yahoo.com

Muhammed Ashraf Memon
Ipswich Hospital, Chelmsford Avenue, Ipswich, Queensland, Australia
Mayne Medical School, School of Medicine, University of Queensland, Brisbane, Queensland, Australia
Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia
Faculty of Health Science, Bolton University, Bolton, Lancashire, UK
E-mail: mmemon@yahoo.com

ABSTRACT

Objectives: To conduct a meta-analysis of randomized controlled trials evaluating the relative merits of limited (D1) versus extended lymphadenectomy (D2) for proven gastric adenocarcinoma.

Data Sources and Review Methods: A search of Cochrane, Medline, PubMed, Embase, Science Citation Index and Current Contents electronic databases identified randomized controlled trials published in the English language between 1980 and 2008 comparing the outcomes D1 vs D2 gastrectomy for gastric adenocarcinoma. The meta-analysis was prepared in accordance with the Quality of Reporting of Meta-analyses (QUOROM) statement. The six outcome variables analysed included length of hospital stay; overall complication rate; anastomotic leak rate; re- operation rate; 30 day mortality rate and 5 year survival rate. Random effects meta-analyses were performed using odds ratios and weighted mean differences.

Results: Six trials totalling 1876 patients (D1=946, D2=930) were analyzed. In five out of the six outcomes the summary point estimates favoured D1 over D2 group with a statistically significant reduction of (i) 7.12 days reduction in hospital stay (WMD -7.12, CI -12.90, -1.35, *p*=0.0001); (ii) 58% reduction in relative odds of developing postoperative complications (OR 0.42, CI 0.24, 0.71, *p*=0.0014); (iii) 59% reduction in anastomotic breakdown (OR 0.41, CI 0.26, 0.65, *p*=0.0002); 67% reduction in re-operation rate (OR 0.33, CI 0.15, 0.72, *p*=0.006) and 42% reduction in 30 day mortality rate (OR 0.58, CI 0.4, 0.85, *p*=0.0052). Lastly there was no significant difference in the 5 year survival (OR 0.97, CI 0.78, 1.20, *p*=0.76) between D1 and D2 gastrectomy patients.

Conclusions: Based on this meta-analysis, D1 gastrectomy is associated with significant fewer anastomotic leaks, postoperative complication rate, reoperation rate, decreased length of hospital stay and 30 day mortality rate. Lastly the five year survival in D1 gastrectomy patients was similar to the D2 cohort.

Keywords: D1 gastrectomy; D2 gastrectomy; Gastric Cancer; Lymphadenectomy; Meta-analysis; Randomized controlled trials; Patient's outcome; Postoperative complications

1. INTRODUCTION

Gastric adenocarcinoma is a locoregional disease with a high propensity for nodal metastasis. Therefore nodal status remains one of the most critical independent predictor of patient survival following gastrectomy for this disease (Seto et al., 1997, Siewert et al., 1998). Studies have shown that lymph node involvement occurs in 3%-5% of cases when the cancer is limited to the mucosa; 11%-25% of cases for those limited to the submucosa; 50% for T2 cancers and 83% for T3 cancers (Onate-ocana et al., 2000, De Gara et al., 2003). D1 gastrectomy entails removing lymph nodes adjacent to the stomach where as a D2 dissection extends this resection to include the nodes around the branches of the celiac axis. Therefore for T2 and T3 tumors, D1 dissection leads to non-curative intervention in the majority of patients leading to poor patient outcome. Despite these facts, the relative merits of gastrectomy with limited (D1) versus extended lymphadenectomy (D2) as an oncological treatment of gastric adenocarcinoma remains contentious. Surgeons from the west have conventionally preferred the D1 approach because (a) of lower incidence of gastric cancer and therefore fewer opportunities even in large tertiary referral centres to perform more radical forms of gastrectomy; (b) of lack of training in performing D2 resection compared to their Japanese counterpart; (c) it is technically demanding with unproven benefits based on a number of randomized controlled trials (RCTs)published to date; and (d) fear of increased risk of complications (Edwards et al., 2004). Even then, the western surgeons have achieved a 5-year survival rate of 10-30% (Degiuli m et al) with D1 resection. On the other hand surgeons in Japan (and east) have traditionally preferred the D2 approach achieving an impressive 5-year survival of 50-60% with a low morbidity and mortality (Soga et al., 1979, Maruyama et al., 1987). However, the greatest criticism of these reports from Japan demonstrating such an impressive benefit and modest morbidity and mortality from D2 resection has been the retrospective nature of the data. Nonetheless this debate has lead some researchers to address this issue objectively in the form of well designed RCTs (Degiuli et al., 2004, Dent et al., 1988, Robertson et al., 1994, Bonenkamp et al., 1995, Cuschieri et al., 1999, Wu et al., 2006) . The "issue of extended lymphadenectomy" in these RCTs has produced some conflicting results further polarizing the literature on this subject. Therefore this present meta-analysis has been undertaken to develop a better understanding of the risks and benefits of D1 and D2 procedures for the treatment of gastric cancer by pooling data from all of the available RCTs.

2. METHODS

All RCTs of any size that compared D1 gastrectomy with D2 gastrectomy for the treatment of gastric adenocarcinoma and which were published in full peer-reviewed journals in the English language between January 1980 and the end of May 2008, were considered for inclusion (Table 1). Only those studies which have reported on at least one clinically relevant outcome were included. Unpublished RCTs, non-randomized prospective and retrospective comparative trials and abstracts of RCTs presented at national and international meetings were excluded. Furthermore, studies which reported on gastric cancers other than adenocarcinoma such as lymphomas were excluded because of different biological behaviour and treatment options for these tumours.

The six outcome variables analysed included (a) length of hospital stay; (b) overall complication rate; (c) anastomotic leak rate; (d) re- operation rate; (e) 30 day mortality rate and (f) 5 year survival rate. These outcomes were thought to be important because they exert influence over practical aspects of surgical practice and policy decisions within institutions.

Trials were identified by conducting a comprehensive search of Medline, Embase, Science Citation Index, Current Contents and Pub Med databases, using medical subject headings "D1 gastrectomy", "D2 gastrectomy", "gastric cancer", "comparative study", "prospective studies", "randomised or randomized controlled trials", "random allocation" and "clinical trial". Manual search of the bibliographies of relevant papers was also carried out to identify trials for possible inclusion. Data extraction and critical appraisal were carried out by three authors (MSS, BM and MAM) for compliance with inclusion criteria and methodological quality. Standardised data extraction forms (Moher et al., 1999) were used by authors to independently and blindly summarise all the data available in the RCTs meeting the inclusion criteria. The authors were not blinded to the source of the document or authorship for the purpose of data extraction. The data were compared and discrepancies were addressed with discussion until consensus was achieved.

Evaluation of the methodological quality of identified studies was conducted using the Jadad scoring system (Jadad et al., 1996) in which each study was assigned a score of between zero (lowest quality) and 5 (highest quality) based on reporting of randomization, blinding, and withdrawals occurring within the study (Table 2).

3. STATISTICAL ANALYSIS

Meta-analyses were performed using odds ratios (ORs) and relative risk (RR) for binary outcomes and weighted mean differences (WMDs) for continuous outcome measures. A slightly amended estimator of OR was used to avoid the computation of reciprocal of zeros among observed values in the calculation of the original OR (Liu et al., 1996). Random effects models, developed by using the inverse variance weighted method approach were used to combine the data (Sutton et al., 2000). Heterogeneity among studies was assessed using the Q statistic proposed by Cochran (Sutton et al., 2000, Wermuth et al., 1979, Hedges et al., 1985) and I^2 index introduced by Higgins and Thompson (Higgins et al., 2002, Huedo-Medina et al., 2006). If the observed value of Q is larger than the critical value at a given significant level, in this case 0.05, we conclude that the outcome variable is statistically significant. For the computations of the confidence intervals estimates of mean and standard deviation are required. However, some of the published clinical trials did not report the mean and standard deviation, but rather reported the size of the trial, the median and range. From these available statistics, estimates of the mean and standard deviation were obtained using formulas proposed by Hozo (Hozo et al., 2005). Funnel plots were synthesized in order to determine the presence of publication bias in the meta-analysis. Both total sample size and precision (1/standard error) were plotted against the treatment effects (OR for the dichotomous variables: complications, anastomotic leak, reoperation and mortality (Sutton et al., 2000, Egger et al., 1997, Tang et al., 2000). All estimates were obtained using computer programs written in R (R: A language and environment for statistical computing [computer program]. Version 2.8.0. Vienna: Foundation for Statistical computing; 2008). All plots were obtained using the 'rmeta' package (Lumley, The rmeta Package Version 2.14: http://cran.rproject.org/web/packages/rmeta/index.html0.

In the case of tests of hypotheses, the paper reports *p*-values for different study variables. In general, the effect is considered to be statistically significant if the p-value is small. If one uses a 5% significance level then the effect is significant only if the associated p-value is less than or equal to 5%.

4. RESULTS

There was almost a perfect agreement (κ=0.99) between the three authors (MSS, BM, MAM) regarding the inclusion and exclusion of various randomized controlled trials. Based on this agreement, a total of 6 randomized prospective clinical trials (Europe =3, Asia =2, Africa=1) (Degiuli et al., 2004, Dent et al., 1988, Robertson et al., 1994, Bonenkamp et al., 1995, Cuschieri et al., 1999, Wu et al., 2006) that included 1876 gastrectomies (D1=946 and D2=930) were considered suitable for meta-analysis (Table 1).

None of the six trials achieved a modified Jadad score of more than 2 (Table 2).

In five out of the six outcomes the summary point estimates favoured D1 over D2 group with a statistically significant reduction of (i) 7.12 days reduction in hospital stay (WMD -7.12, CI -12.90, -1.35, *p*=0.0001); (ii) 58% reduction in relative odds of developing postoperative complications (OR 0.42, CI 0.24, 0.71, *p*=0.0014); (iii) 59% reduction in anastomotic breakdown (OR 0.41, CI 0.26, 0.65, *p*=0.0002); 67% reduction in re-operation rate (OR 0.33, CI 0.15, 0.72, *p*=0.006) and 42% reduction in 30 day mortality rate (OR 0.58, CI 0.4, 0.85, *p*=0.0052). Lastly there was no significant difference in the 5 year survival (OR 0.97, CI 0.78, 1.20, *p*=0.76) between D1 and D2 gastrectomy patients (Table 3).

5. DISCUSSION

Gastric adenocarcinoma survival is proportional to the level of lymph node metastases in nodal echelons N1-N4 based on the nodal classification by the JRSGC. Surgeons in the east have routinely practiced D2 gastrectomies involving extended lymphadenectomy which provides both diagnostic and therapeutic advantage (Soga et al., 1979, Maruyama et al., 1987). However surgeons in the west have struggled to achieve similar outcome with D2 gastrectomies compared with their eastern counterparts except on occasion (Edwards et al., 2004, Lewis et al., 2002, Diaz et al., 2008, Sue-Ling et al., 1993, Pacelli et al., 1993, Siewert et al., 1993).

A number of RCTs (Degiuli et al., 2004, Dent et al., 1988, Robertson et al., 1994, Bonenkamp et al., 1995, Cuschieri et al., 1999, Wu et al., 2006) have been undertaken to investigate the issues of risks and benefits of limited vs extended lymphadenectomy. Confounding factors including patient population and their selection, operative techniques especially level of lymph node dissection, experience of the operating surgeons especially in D2 resection and outcome descriptors have fuelled the ongoing debate despite reasonable attempts being made by the authors to provide trials of high quality. The authors of this paper have undertaken a meta-analytical review based on the available RCTs data in an attempt to provide some clarification. To date D2 gastrectomy has shown better results than D1 gastrectomy in mainly retrospective studies (Pacelli et al., 1993, Siewert et al., 1993, Mansfield et al., 2004, Volpe et al., 1995). However, this has not been seen in two of the largest RCTs published in

Europe (Bonenkamp et al., 1995, Cuschieri et al., 1999). In the Dutch trial with a median follow-up of 11 year, survival rates were 30% for D1 and 35% for D2 group. The risk of relapse was 70% for D1 and 65% for D2 group. The only group which seemed to benefit was the N2 disease cohort. The 11 year survival was 0% in D1 group whereas it was 20% for D2 group (Hartgrink et al., 2004). However there is a serious concern regarding the Dutch RCT for non-compliance or contamination in the extent of lymphadenectomy performed in the two randomized arms (Bunt et al., 1994).

Five of the 6 studies reported length of hospital stay. Of these 4 had significantly longer stay for D2 group compared to D1 group. The reasons for the longer stay in D2 cohort could be multiple and include (a) prolong and more complex surgical procedures; and (b) more peri-operative complications. A protracted operating time exposes the patient to a longer duration of anaesthesia, and a greater risk of thermic, thromboembolic, cardiac and respiratory complications.

Five of the 6 trials reported data on complications. Complications ranged from simple wound infection, intra-abdominal abscess formation to anastomotic leak. Different studies have enumerated complications on different bases which made summation of the results difficult. Most of the studies have reported complication under 2 major headings, surgical and non surgical. Among the surgical complications, intra-abdominal or subphrenic abscess formation was not only commonly seen but also required reoperation in the majority of cases. The intra-abdominal sepsis was also responsible for other complications such as secondary haemorrhages and death (Robertson et al., 1994). This complication was seen exclusive in D2 group. Amongst the non-surgical complications, the pulmonary infections predominate in both the groups. The overall complication rates were significantly higher in D2 group compared to D1 cohort.

Three of the 6 trials reported an anastomotic leak rate. In all the studies the D2 group had a higher leak rate. The overall pooled data similarly showed that it was statistically significant for the D2 group. The UK (Cuschieri et al., 1999) and the Dutch (Bonenkamp et al., 1995) trials, have not specified the site(s) of the leak or timing of the leaks. In the Hong Kong study (Robertson et al., 1994). 3 anastomotic leaks were recorded, all in the D2 group at the oesophagojejunal junction (i.e following a total gastrectomy) which were treated conservatively with total parenteral nutrition with favourable outcome.

Four of the 6 trials have documented a re-operation rate. The result was statistically significant. The most common cause reported for re-operation was a sub-phrenic abscess, although the causes ranged from haemorrhage to anastomotic leaks and from retro-colic hernia to intra abdominal abscesses.

Five of the 6 trials reported on the 30 day mortality rate. Pooled data shows higher mortality rates in the D2 group, although Taiwanese and Italian studies have reported lower mortality rates in the D2 group (Degiuli et al., 2004, Wu et al., 2006). The overall result was statistically significant for D2 cohort. The factors which may be responsible for higher morbidity and mortality include (a) surgical experience in performing gastric resection (high vs low volume centre); (b) the previous gastric surgery experience; (c) the pancreatic or splenic resection. Pancreatosplenectomy recommended as part of the D2 gastrectomy in the second edition of Japanese classification (Japanese Gastric Cancer Association Japanese Classification of Gastric Carcinoma - 2nd English Edition (1998). *Gastric Cancer*. 1(1):10-24, Aiko and Sasako, 1998) [27,48]was followed by some and not by others (Degiuli et al., 2004, Dent et al., 1988, Robertson et al., 1994, Bonenkamp et al., 1995, Cuschieri et al., 1999, Wu et al., 2006). While surgeons in the Hong Kong and Taiwanese (Robertson et al., 1994, Wu et al., 2006) studies performed pancreatic and splenic resection routinely, the rest of the surgeons except Italians, performed the resection only when the upper 2/3 of the stomach was involved (Dent et al., 1988, Bonenkamp et al., 1995, Cuschieri et al., 1999). The Italians resected the pancreas or the spleen only when the cancer had extended into these organs (Degiuli et al., 2004). The most important outcome that all the surgeons look forward to in any cancer treatment is the 5 year survival rate. Interestingly this meta-analysis does not show any difference in the 5 year survival between the D1 and D2 group which casts doubt on the benefit of the extended gastrectomy.

6. CONCLUSIONS

D1 gastrectomy is associated with significantly fewer post-operative complications including anastomotic leaks, lower re-operation rate , decreased length of hospital stay and decreased 30 day mortality rate. Most importantly there is no difference in 5 year survival rate between the two groups. It is therefore difficult to justify the routine use of D2 gastrectomy as the standard for the management of gastric carcinoma especially when the morbidity and mortality remains high.

APPENDIX

Table 1: Details of all the RCTs

Authors	Year/Country	Type of Study	Number of Patients	
			D1	D2
Dent et al	1988/ South Africa	RCT	22	21
Robertson et al	1994/ Hong Kong	RCT	25	29
Bonenkamp et al	1995/ Netherlands	RCT	513	483
Cuschieri et al	1999/ UK	RCT	200	200
Degiuli et al	2004/ Italy	RCT	76	86
Chew-Wun Wu et al	2006/ Taiwan	RCT	110	111
Total			**946**	**930**

Table 2: Jadad's Score

Authors	Year/Country	Jadad's Score	Randomization	Blinding	Withdrawal
Dent et al	1988/ South Africa	2	2	0	0
Robertson et al	1994/ Hong Kong	2	2	0	0
Bonenkamp et al	1995/ Netherlands	2	2	0	0
Cuschieri et al	1999/ UK	2	2	0	0
Degiuli et al	2004/ Italy	2	2	0	0
Chew-Wun Wu et al	2006/ Taiwan	2	2	0	0

Table 3: Summary of pooled data comparing D1 vs D2 gastrectomy

Outcome Variables	Pooled OR or WMD (95% CI)	Test for overall effect		Test for heterogeneity		
		Z	p-value	Q	p-value	I-squared index
Length of hospital stay	-7.12 (-12.90, -1.35)	-2.4177	0.0156	36.04	0.0001	88.9%
Overall complication rate	0.42 (0.24, 0.71)	-3.2009	0.0014	11.5	0.0215	65.2%
Anastomotic leak rate	0.41 (0.26, 0.65)	-3.7644	0.0002	0.8	0.939	0%
Re-operation rate	0.33 (0.15, 0.72)	-2.7485	0.006	3.52	0.3179	14.8%
30 day mortality rate	0.58 (0.4, 0.85)	-2.7928	0.0052	1.78	0.7753	0%
5 year survival rate	0.97 (0.78, 1.20)	-0.7662	0.7662	1.67	0.797	0%

REFERENCES

Aiko T, Sasako M. (1998). The new Japanese Classification of Gastric Carcinoma: Points to be revised. *Gastric Cancer*. 1(1): 25-30

Bonenkamp JJ, Songun I, Hermans J, Sasako M, Welvaart K, Plukker JT, van Elk P, Obertop H, Gouma DJ, Taat CW. (1995). Randomised comparison of morbidity after D1 and D2 dissection for gastric cancer in 996 Dutch patients. *Lancet*. 25; 345(8952): 745-8

Cuschieri A, Weeden S, Fielding J, Bancewicz J, Craven J, Joypaul V, Sydes M, Fayers P. (1999). Patient survival after D1 and D2 resections for gastric cancer: long-term results of the MRC randomized surgical trial. Surgical Co-operative Group. *Br J Cancer*. 79(9-10): 1522-30.

Degiuli M, Sasako M, Calgaro M, Garino M, Rebecchi F, Mineccia M, Scaglione D, Andreone D, Ponti A, Calvo F; Italian Gastric Cancer Study Group. (2004). Morbidity and mortality after D1 and D2 gastrectomy for cancer: interim analysis of the Italian Gastric Cancer Study Group (IGCSG) randomised surgical trial. *Eur J Surg Oncol*. 30(3): 303-8.

De Gara CJ, Hanson J, Hamilton S. (2003). A population-based study of tumor-node relationship, resection margins, and surgeon volume on gastric cancer survival. *Am J Surg*. 186(1): 23-7.

Dent DM, Madden MV, Price SK. (1998). Randomized comparison of R1 and R2 gastrectomy for gastric carcinoma. *Br J Surg*. 75(2): 110-2.

Díaz de Liaño A, Yarnoz C, Aguilar R, Artieda C, Ortiz H. (2008). Rationale for gastrectomy with D2 lymphadenectomy in the treatment of gastric cancer. *Gastric Cancer*. 11(2): 96-102. Epub 2008 Jul 2.

Edwards P, Blackshaw GR, Lewis WG, Barry JD, Allison MC, Jones DR. (2004). Prospective comparison of D1 vs modified D2 gastrectomy for carcinoma. *Br J Cancer*. 17; 90(10): 1888-92.

Egger M, Davey Smith G, Schneider M, Minder C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ. 13; 315(7109): 629-34.

Hedges LV, Olkin I. (1985). Statistical methods for meta analysis. Orlando: Academic Press.

Higgins JP, Thompson SG. (2002). Quantifying heterogeneity in a meta-analysis. Stat Med. 15; 21(11): 1539-58

Hozo SP, Djulbegovic B, Hozo I. (2005). Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol. 20; 5(1):13

Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods. 11(2): 193-206

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ. (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 17: 1-12.

Japanese Gastric Cancer Association Japanese Classification of Gastric Carcinoma - 2nd English Edition. (1998). *Gastric Cancer*. 1(1): 10-24.

Lewis WG, Edwards P, Barry JD, Khan S, Dhariwal D, Hodzovic I, Allison MC, Shute K. (2002). D2 or not D2? The gastrectomy question. *Gastric Cancer*. 5(1):2 9-34.

Liu IM, Agresti A. (1996). Mantel-Haenszel-type inference for cumulative odds ratios with a stratified ordinal response. Biometrics. 52(4): 1223-34.

Lumley T. The rmeta Package Version 2.14, http://cran.rproject.org/web/packages/rmeta.

Mansfield PF. (2004). Lymphadenectomy for gastric cancer. *J Clin Oncol*. 15;22(14): 2759-61.

Maruyama K, Okabayashi K, Kinoshita T. (1987). Progress in gastric cancer surgery in Japan and its limits of radicality. World J Surg. 11(4): 418-25.

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet.* 354: 1896-1900

Oñate-Ocaña LF, Aiello-Crocifoglio V, Mondragón-Sánchez R, Ruiz-Molina JM. (2000). Survival benefit of D2 lympadenectomy in patients with gastric adenocarcinoma. *Ann Surg Oncol*. 7(3): 210-7.

Pacelli F, Doglietto GB, Bellantone R, Alfieri S, Sgadari A, Crucitti F. !1993). Extensive versus limited lymph node dissection for gastric cancer: a comparative study of 320 patients. Br J Surg. 80(9): 1153-6

R: A language and environment for statistical computing [computer program]. Version 2.8.0. Vienna: Foundation for Statistical computing; 2008

Robertson CS, Chung SC, Woods SD, Griffin SM, Raimes SA, Lau JT, Li AK. (1994). A prospective randomized trial comparing R1 subtotal gastrectomy with R3 total gastrectomy for antral cancer. *Ann Surg*. 220(2): 176-82.

Seto Y, Nagawa H, Muto T. (1997). Impact of lymph node metastasis on survival with early gastric cancer. *World J Surg*. 21(2): 186-89

Siewert JR, Böttcher K, Roder JD, Busch R, Hermanek P, Meyer HJ.(1993). Prognostic relevance of systematic lymph node dissection in gastric carcinoma. German Gastric Carcinoma Study Group. Br J Surg. 80(8): 1015-8

Siewert JR, Böttcher K, Stein HJ, Roder JD. (1998). Relevant prognostic factors in gastric cancer: ten-year results of the German Gastric Cancer Study. *Ann Surg.* 228(4): 449-61.

Soga J, Kobayashi K, Saito J, Fujimaki M, Muto T. (1979). The role of lymphadenectomy in curative surgery for gastric cancer. *World J Surg*. 3(6): 701-8

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. (2000). Methods for Meta-analysis in Medical Research. *John Wiley*: Chichester.

Sue-Ling HM, Johnston D, Martin IG, Dixon MF, Lansdown MR, McMahon MJ, Axon AT. (1993). Gastric cancer: a curable disease in Britain. BMJ. 4 ;307(6904): 591-6.

Tang JL, Liu JL.(2000). Misleading funnel plot for detection of bias in meta-analysis. J Clin Epidemiol. 53(5): 477-84.

Volpe CM, Koo J, Miloro SM, Driscoll DL, Nava HR, Douglass HO Jr. (1995). The effect of extended lymphadenectomy on survival in patients with gastric adenocarcinoma. *J Am Coll Surg.* 181(1): 56-64.

Wermuth N, Cochran WG. (1979). Detecting systematic errors in multi-clinic observational data. Biometrics. 35(3): 683-6.

Wu CW, Hsiung CA, Lo SS, Hsieh MC, Chen JH, Li AF, Lui WY, Whang-Peng J. (2006).Nodal dissection for patients with gastric cancer: a randomised controlled trial. *Lancet Oncol*. 7(4): 309-15.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 148–156.

META-ANALYSIS OF LAPAROSCOPIC POSTERIOR AND ANTERIOR FUNDOPLICATION FOR GASTRO-OESOPHAGEAL REFLUX DISEASE

Manjunath S Subramanya
Department of Surgery, Mount Isa Base Hospital, Mount Isa, Queensland, Australia
E-mail: manjunathbss9@yahoo.com

Md Belal Hossain
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments, University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: bjoardar2003@yahoo.com

Shahjahan Khan
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments, University of Southern Queensland, Toowoomba, Queensland, Australia
E-mail: khans@usq.edu.au

Breda Memon
Ipswich Hospital, Chelmsford Avenue, Ipswich, Queensland, Australia
E-mail: bmemon@yahoo.com

Muhammed Ashraf Memon
Ipswich Hospital, Chelmsford Avenue, Ipswich, Queensland, Australia
Mayne Medical School, School of Medicine, University of Queensland, Brisbane, Queensland, Australia
Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia
Faculty of Health Science, Bolton University, Bolton, Lancashire, UK
E-mail: mmemon@yahoo.com

ABSTRACT

Objectives: Although laparoscopic posterior (Nissen) fundoplication (LPF) has the proven efficacy for controlling gastro-oesophgeal reflux surgically, there remain problems with postoperative dysphagia and the inability to belch or vomit. To decrease some of these postoperative complications, laparoscopic anterior fundoplication (LAF) was introduced. The aim of this study was to conduct a meta-analysis of RCTs to investigate the merits of LPF vs LAF for the treatment of gastro-oesophageal reflux disease (GORD).

Data Sources and Review Methods: A search of Medline, Embase, Science Citation Index, Current Contents, PubMed and the Cochrane Database identified all RCTs comparing different types of laparoscopic posterior and anterior fundoplications published in the English Language between 1990 and 2008. The eight variables analysed included operative time, overall complications, rate of conversion to open, re-do operative rate, dysphagia score, heartburn rate, visick grading of satisfaction and overall satisfaction.

Results: Five trials totalling 556 patients (Posterior=277, Anterior=279) were analysed. The analysis showed trends favouring LPF in terms of overall complication rate, conversion rate, incidence of postoperative heartburn and re-do operative rate. There was significant improvement in the postoperative satisfaction score in terms of reflux symptoms favouring LPF while there was significant reduction in the dysphagia score favouring LAF. No difference was noted in operating time and Visick's grading of satisfaction between the two groups.

Conclusions: Based on this meta analysis, LPF is associated with fewer complications, decreased rate of conversion, heartburn rate, re-operation rate and significantly higher overall satisfaction rate for controlling GORD symptoms. However the LAF was associated with a significantly lower incidence of dysphagia compared to its posterior counterpart. We therefore conclude that LPF is a better alternative to AFP at the expense of higher dysphagia rate.

Keywords: Fundoplication; Anti-reflux; Posterior fundoplication; Nissen fundoplication; Anterior antireflux surgery; Gastro-oesophagael reflux disease; Randomised controlled trials

1. INTRODUCTION

The "Open Nissen Fundoplication", and its modifications have been employed to treat moderate to severe gastro-oesophageal reflux disease (GORD) for almost six decade now. The procedure involves 360^0 posterior wrapping of the fundus of the stomach around the distal oesophagus. Laparoscopic fundoplication, the minimally access modification of the original Nissen fundoplication was first described by Dallemagne in 1991 (Dallemagne et al., 1991) following the success of laparoscopic cholecystectomy. Although this type of fundoplication achieves good control of GORD, some patient may experience troublesome postoperative side effects such as dysphagia, an inability to vomit or belch (gas bloat syndrome) and excessive flatulence. To alleviate this problem partial posterior and anterior fundoplications were introduced where the wrap extended from 90^0 to 270^0 either posteriorly or anteriorly. The question of which technique offers the best results in terms of control of GORD with minimal side effects remains controversial. Many uncontrolled single centre series claim good results with partial posterior (Bell, 1996, O'Reilly, 1996) or anterior fundoplication (Watson et al., 1991, 1995, Rice et al., 2006). However, there have been few RCTs undertaken to address this controversial issue in an evidence based manner (Hagedorn et al., 2003, Watson et al., 1999, 2004, Baigrie et al., 2005, Spence et al., 2006). The problem is further compounded by the fact that only a few of these RCTs have long term follow-up.

The aim of this meta-analysis was to investigate the benefits and risks of LAP versus LPF for the treatment of GORD. Five randomised controlled trials (Hagedorn et al., 2003, Watson et al., 1999, 2004, Baigrie et al., 2005, Spence et al., 2006) comparing these procedures have been published over the last seven years. This meta-analysis considers pooled data from all of the available randomised clinical trials that compared LAF and LPF methods of anti-reflux surgery, and was prepared in accordance with the Quality of Reporting of Meta-analyses (QUOROM) statement (Moher et al., 1999).

2. MATERIAL AND METHODS

All randomised clinical trials of any size that compared any type of LAF with LPF for the treatment of GORD, and were published in full in peer-reviewed journals in the English language between January 1990 and the end of May 2008 were included (Table 1). Studies must also have reported on at least one clinically relevant outcome. Unpublished studies and abstracts presented at national and international meetings were excluded. Eight outcome variables which were considered most suitable for analysis included:

1. Operative time
2. Overall complications
3. Rate of conversion to open
4. Re-do operative rate
5. Dysphagia score
6. Heartburn rate
7. Visick grading of satisfaction
8. Overall satisfaction

Trials were identified by conducting a comprehensive search of Medline, Embase, Science Citation Index, Current Contents and Pubmed databases, using medical subject headings “fundoplication”, “anti-reflux”, “posterior fundoplication”, “Nissen fundoplication”, “anterior antireflux surgery”, “gastro-oesophagael reflux disease”, “comparative studies”, “prospective studies”, “randomised controlled trials”, “random allocation” and “clinical trial”. Manual search of the bibliographies of relevant papers was also carried out to identify trials for possible inclusion.

Data extraction and critical appraisal were carried out by three authors (MSS, BM and MAM) for compliance with inclusion criteria and methodological quality. Standardised data extraction forms (Moher et al., 1999) were used by authors to independently and blindly summarise all the data available in the RCTs meeting the inclusion criteria. The authors were not blinded to the source of the document or authorship for the purpose of data extraction. The data were compared and discrepancies were addressed with discussion until consensus was achieved.

Evaluation of the methodological quality of identified studies was conducted using the Jadad scoring system (Jadad et al., 1996) in which each study was assigned a score of between zero (lowest quality) and 5 (highest quality) based on reporting of randomization, blinding, and withdrawals occurring within the study (Table 2).

3. STATISTICAL ANALYSIS

Meta-analyses were performed using odds ratios (ORs) and relative risk (RR) for binary outcomes and weighted mean differences (WMDs) for continuous outcome measures. A slightly amended estimator of OR was used to avoid the computation of reciprocal of zeros among observed values in the calculation of the original OR (Liu et al., 1996). Random effects models, developed by using the inverse variance weighted method approach were used to combine the data (Sutton et al., 2000). Heterogeneity among studies was assessed using the Q statistic proposed by Cochran (Sutton et al., 2000, Hedges et al., 1985) and I^2 index introduced by Higgins and Thompson (Higgins et al., 2002, Huedo-Medina et al., 2006). If the observed value of Q is larger than the critical value at a given significant level, in this case 0.05, we conclude that the outcome variable is statistically significant. For the computations of the confidence intervals estimates of mean and standard deviation are required. However, some of the published clinical trials did not report the mean and standard deviation, but rather reported the size of the trial, the median and range. From these available statistics, estimates of the mean and standard deviation were obtained using formulas proposed by Hozo (Hozo et al., 2005). Funnel plots were synthesized in order to determine the presence of publication bias in the meta-analysis. Both total sample size and precision (1/standard error) were plotted against the treatment effects (OR for the dichotomous variables: complications, conversion rate, dysphagia, heartburn and reoperation (Sutton et al., 2000, Egger et al., 1997,). All estimates were obtained using computer programs

written in R (R: A language and environment for statistical computing [computer program]. Version 2.8.0. Vienna: Foundation for Statistical computing; 2008). All plots were obtained using the ‘rmeta’ package (Lumley T. The rmeta Package Version 2.14.
In the case of tests of hypotheses, the paper reports *p*-values for different study variables. In general, the effect is considered to be statistically significant if the p-value is small. If one uses a 5% significance level then the effect is significant only if the associated p-value is less than or equal to 5%.

4. RESULTS

There was almost a perfect agreement (κ=0.99) between the three authors (MSS, BM, MAM) regarding the inclusion and exclusion of various randomized controlled trials. Based on this agreement, a total of five randomized prospective clinical trials (Hagedorn et al., 2003, Watson et al., 1999, 2004, Baigrie et al., 2005, Spence et al., 2006) involving a total of 556 patients (LAF=279, LPF=277) were considered suitable for meta-analysis. None of the five trials achieved a modified Jadad score of more than 2 (Table 1).

A statistically significant reduction of 4.9 in WMD for the dysphagia score was noted favouring LAF (WMD 4.8982, CI 2.4753, 7.3211, p=<0.0001). On the other hand, there was a significant 59% reduction in the relative odds of heartburn seen in patients receiving LPF (OR 0.4145, CI 0.1785, 0.9625, p=0.0405). Comparable effects were seen for LPF and LAP for other variables which include operating time (WMD 1.5581, CI -4.8591; 7.9753, p=0.6342), overall complications (OR 0.4547, CI 0.1874, 1.1035, p=0.0815), conversion rate (OR 0.899, CI 0.2279, 3.5465, p=0.8791), Visick grading (OR 1.41, CI 0.67, 2.95, p=0.3631), patient’s satisfaction (WMD -0.1848, CI -0.6002; 0.2305, p=0.383) and redo surgery (OR 0.6538, CI 0.2889, 1.4799, p=0.308) (Table 2).

5. DISCUSSION

It is well documented that while LPF is associated with higher rates of dysphagia, the LAF is accompanied by worse reflux control (Hagedorn et al., 2003). This is supported objectively by the findings of Anderson (Anderson et al., 1998) in his RCT where Nissen (LPF) fundoplication was found to have a greater elevation of resting (33.5 vs. 23 mm Hg) and residual lower esophageal sphincter pressures (17 vs. 6.5 mm Hg) and lower esophageal ramp pressure (26 vs. 20.5 mm Hg) than the LAF. A smaller radiologically measured sphincter opening diameter was seen following Nissen fundoplication (9 mm) compared with anterior fundoplication (12 mm).

Various modifications have been tried to better the outcome in relation to reflux control and dysphagia, such as varying the degree of wrap and division of short gastric vessels. Division of short gastric vessels have failed to show any improvement in the overall outcome with regards to postoperative symptoms such as dysphagia in either laparoscopic or open fundop-lication (Watson et., 1997, 1999, Blomqvist et al., 2000, Luostarinen and Isolauri, 1999). This fact is further proven by Engstrom (Engstrom et al., 2004) in her RCT which involves 24 patients undergoing Laparoscopic Nissen Fundoplication divided into 2 groups, with and without short gastric divisions, where there was no difference in the rate of dysphagia or reflux control achieved.

There remains controversy regarding the correlation between division of short gastric vessels and wind related problems mainly belching and bloating. While Luostarinen (Luostarinen and Isolauri, 1999) showed no change in wind related problems with division of the short gastrics, Boyle (O’Boyle et al., 2002) and Chrysos (Chrysos et al., 2001) showed increased bloating and flatus with division of the short gastrics. Luostarinen’s RCT involved 50 patients who underwent

Nissen Fundoplication, with and without short gastrics division and at 5 years showed no difference in wind related problems with either group. Boyle's RCT involved 102 patients undergoing Nissen fundoplication and at 3 years showed, in the group having division of the short gastrics, an increased incidence of epigastric bloating (71% vs. 48%) and increased flatus (88% vs. 70%) compared to the group without the division of short gastric. Chrysos in his RCT involving 56 patients undergoing Nissen fundoplication, at 1 year showed an increased incidence of bloating in the group who had division of the short gastric (38%) compared to the group without short gastric division (19%).

Trials involving partial posterior fundoplication which were conducted in an attempt to eliminate certain postoperative problems such as dysphagia and bloating associated with total wrap posterior fundoplication managed to eliminate bloating but failed to decrease the rate of dysphagia. A ten year follow up result of a RCT involving 137 patients comparing laparoscopic Toupet (180 degree posterior) and Nissen fundoplication reported by Hagedorn (Hagedorn et al., 2002), showed no difference in the rate of dysphagia (41% vs. 38%) but significantly less bloating (54% vs. 66%, p <0.03) and flatulence (72% vs. 90%, p <0.001). Walker et al. (1992) in his RCT involving 52 patients compared Lind (300 degree posterior) and Nissen fundoplication where he found no difference in the rate of postoperative dysphagia but a significant improvement in bloating in patients undergoing Lind fundoplication (p <0.05)

Partial posterior fundoplication on the other hand has better reflux control than the partial anterior fundoplication as shown by the long term results of the Swedish triall (Hagedorn et al., 2003), the only RCT to compare 2 partial fundoplications (Posterior 180 vs. anterior 120). The 5 year follow up result of this trial reported by Engstrom (Engstrom et al., 2007) showed that 82.2% has no reflux symptoms in the LPF group in comparison to only 34.9% in the LAF group. It had similar figures with respect to the patient being able to belch (86.7% in LPF, 88.4% in LAF) but had much worse results with ability to vomit (27.3% in LPF, 63.4% in LAF).

Ludemann (Ludemann et al., 2005), in his 5 year follow up report of an RCT involving 105 patients, initially conducted by Watson (Watson et al., 1999) concluded that anterior fundoplication (180 degree) has a reflux control as good as Nissen fundoplication. Both at the end of 5 years had 90% of patients with good reflux control and not on any PPIs.

Hagedorn (Hagedorn et al., 2003) on the other hand in his RCT involving 96 patients comparing 120 degree anterior and Nissen fundoplication reports a worse reflux control with the former procedure (34% vs. 73%). Similarly Watson (Watson et al., 2004) in his RCT involving 112 patients comparing 90 degree anterior with Nissen fundoplication reports a worse reflux control with the anterior fundoplication (79% vs. 95%). A single centre prospective trial by Rice (Rice et al., 2006) involving 117 patients who underwent 180 degree anterior fundoplication revealed that at 5 years (99% of patients were available for follow up) 80% of patients have a full reflux control and not needing any PPIs. Furthermore 95% of the patients were highly satisfied overall. These results were based on patient questionnaires and not on any manometric or endoscopic follow up.

Very few studies have published long term outcomes for laparoscopic fundoplication. A number of authors have shown previously better outcomes with short term follow up which are not matched by long term follow up results (Ludemann et al., 2003). Kneist et al. (2003) showed in a retrospective study a significant improvement of heartburn with anterior fundoplication after 44 months of follow up, where as Engstrom (Engstrom et al., 2007) in a RCT comparing anterior and posterior fundoplication found that posterior fundoplication yielded better control of reflux after 65 months of follow up. In the RCT by Watson (Watson et al., 1999), at 5 years there was a higher number of patients with heartburn in the anterior fundoplication group, easily controlled with PPIs, however the posterior fundoplication group had a higher number of patients with dysphagia, although patients in both groups were more satisfied overall postoperatively

(Ludemann et al., 2005). At 10 years both groups of patients had a good results in terms of heartburn, but anterior fundoplication group had decreased incidence of dysphagia (Cai et al., 2008). A 10 year retrospective study from Fein (2008) involving 120 patients also showed that regurgitation persisted in only 15% of the patients in the Nissen group compared to 44% in the anterior fundoplication group.

Finally, surgeons experience has to be taken into account when determining the long term outcome of reflux control. A 5 year follow up of a prospective trial comparing open and laparoscopic Nissen with anterior fundoplication showed no difference in the control of reflux symptoms (Stewart et al., 2004). A 5 year follow up of a prospective study of laparoscopic fundoplication involving 2 groups, during the learning curve and thereafter, shows a significant improvement in the re-operation rate for persisting reflux from 15% to only 6% (Jamieson, 2005). Therefore surgeons experience has a significant role to play especially if the trial is conducted during the learning curve of the surgeons involved.

6. CONCLUSIONS

Based on this meta-analysis, LPF is associated with fewer complications, decreased rate of conversion, heartburn and re-operation, and significantly higher overall satisfaction among patients. However the LAF was associated with a significantly lower incidence of dysphagia. We therefore conclude that LPF is a better alternative to LAF at the expense of a higher dysphagia rate.

APPENDIX

Table 1: Details of all the RCTs

Authors	**Year/Country**	**No of Patients**		**Types of Fundoplications**	**Division of Short Gastrics**	**Jadad Score**
		LAF	LPF			
Watson et al	1999/Australia	54	53	Anterior 180 vs Nissen 360	No	2
Hagedorn et al	2003/Sweden	47	48	Anterior 120 vs Posterior 180	Yes	2
Watson et al	2004/Australia	60	52	Anterior 90 vs Nissen 360	Yes	2
Baigrie et al	2005/South Africa	79	84	Anterior 180 vs Nissen 360	No	2
Spence et al	2006/Australia	39	40	Anterior 90 vs Nissen 360	No	2

Table2: Summary of pooled data comparing posterior and anterior fundoplication

Outcome Variables	Patients	Pooled OR WMD (95% CI)	Test for overall effect		Test for heterogeneity		
	n		Z	p value	*Q*	*p* value	I^2 index
Operation time	512	1.5581 (-4.8591; 7.9753)	0.4759	0.6342	33.83	0.0001	91.1%
Overall complications	298	0.4547 (0.1874; 1.1035)	-1.7422	0.0815	1.78	0.411	0%
Conversion rate	461	0.899 (0.2279; 3.5465)	-0.1521	0.8791	2.31	0.5099	0%
Dysphagia	393	4.8982 (2.4753; 7.3211)	3.9623	< 0.0001	223.94	0.0001	98.7%
Heartburn	393	0.4145 (0.1785; 0.9625)	-2.0488	0.0405	5.71	0.1266	47.5%
Visick grading	462	1.41 (0.67; 2.95)	0.9095	0.3631	0.23	0.9719	0%
Satisfaction	461	-0.1848 (-0.6002;0.2305)	-0.8723	0.383	1.65.47	0.0001	98.2%
Redo surgery	461	0.6538 (0.2889; 1.4799)	-1.0194	0.308	0.36	0.948	0%

REFERENCES

Anderson JA, Myers JC, Watson DI, Gabb M, Mathew G, Jamieson GG. (1998). Concurrent fluoroscopy and manometry reveal differences in laparoscopic Nissen and anterior fundoplication. *Dig Dis Sci.* 43(4):847-53

Baigrie RJ, Cullis SN, Ndhluni AJ, Cariem A. (2005). Randomized double-blind trial of laparoscopic Nissen fundoplication versus anterior partial fundoplication. Br J Surg. 92(7):819-23

Bell RC, Hanna P, Powers B, Sabel J, Hruza D. (1996). Clinical and manometric results of laparoscopic partial (Toupet) and complete (Rosetti-Nissen) fundoplication. *Surg Endosc.* 10(7):724-8

Blomqvist A, Dalenbäck J, Hagedorn C, Lönroth H, Hyltander A, Lundell L. (2005). Impact of complete gastric fundus mobilization on outcome after laparoscopic total fundoplication. *J Gastrointest Surg.* 4(5):493-500

Cai W, Watson DI, Lally CJ, Devitt PG, Game PA, Jamieson GG. (2008). Ten-year clinical outcome of a prospective randomized clinical trial of laparoscopic Nissen versus anterior 180 (degrees) partial fundoplication. *Br J Surg*. 95(12):1501-5

Chrysos E, Tzortzinis A, Tsiaoussis J, Athanasakis H, Vasssilakis J, Xynos E. (2001). Prospective randomized trial comparing Nissen to Nissen-Rossetti technique for laparoscopic fundoplication. *Am J Surg*. 182(3):215-21.

Dallemagne B, Weerts JM, Jehaes C, Markiewicz S, Lombard R. (1991). Laparoscopic Nissen fundoplication: preliminary report. *Surg Laparosc Endosc* 1(3):138-43

Egger M, Davey Smith G, Schneider M, Minder C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ. 13; 315(7109):629-34.

Engström C, Blomqvist A, Dalenbäck J, Lönroth H, Ruth M, Lundell L. (2004). Mechanical consequences of short gastric vessel division at the time of laparoscopic total fundoplication. *J Gastrointest Surg*. 8(4):442-7

Engström C, Lönroth H, Mardani J, Lundell L. (2007). An anterior or posterior approach to partial fundoplication? Long-term results of a randomized trial. *World J Surg.* 31(6):1221-5; discussion 1226-7.

Fein M, Bueter M, Thalheimer A, Pachmayr V, Heimbucher J, Freys SM, Fuchs KH. (2008). Less reflux recurrence following Nissen fundoplication : results of laparoscopic antireflux surgery after 10 years. *Chirurg*. 79(8):759-64.

Hagedorn C, Jönson C, Lönroth H, Ruth M, Thune A, Lundell L. (2003). Efficacy of an anterior as compared with a posterior laparoscopic partial fundoplication: results of a randomized, controlled clinical trial. *Ann Surg.* 238(2):189-96

Hagedorn C, Lönroth H, Rydberg L, Ruth M, Lundell L. (2002). Long-term efficacy of total (Nissen-Rossetti) and posterior partial (Toupet) fundoplication: results of a randomized clinical trial. *J Gastrointest Surg*. 6(4):540-5.

Hedges LV, Olkin I. (1985). Statistical methods for meta analysis. Orlando: Academic Press.

Higgins JP, Thompson SG. (2002). Quantifying heterogeneity in a meta-analysis. Stat Med. 15; 21(11):1539-58.

Hozo SP, Djulbegovic B, Hozo I. (2005). Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol. 20; 5(1):13.

Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods. 11(2):193-20

Jamieson GG. (2005). Results of laparoscopic antireflux surgery at five years and beyond]. *Bull Acad Natl Med.* 189(7):1519-25.

Kneist W, Heintz A, Trinh TT, Junginger T. (2003). Anterior partial fundoplication for gastroesophageal reflux disease. *Langenbecks Arch Surg*. 388(3):174-80.

Liu IM, Agresti A. (1996). Mantel-Haenszel-type inference for cumulative odds ratios with a stratified ordinal response. Biometrics. 52(4):1223-34.

Ludemann R, Watson DI, Jamieson GG. (2003). Influence of follow-up methodology and completeness on apparent clinical outcome of fundoplication. *Am J Surg*. 186(2):143-7.

Ludemann R, Watson DI, Jamieson GG, Game PA, Devitt PG. (2005). Five-year follow-up of a randomized clinical trial of laparoscopic total versus anterior 180 degrees fundoplication. *Br J Surg*. 92(2):240-3.

Lumley T. The rmeta Package Version 2.14. http://cran.rproject.org/web/packages/rmeta/

Luostarinen ME, Isolauri JO. (1999). Randomized trial to study the effect of fundic mobilization on long-term results of Nissen fundoplication. *Br J Surg.* 86(5):614-8.

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 354: 1896-1900.

O'Boyle CJ, Watson DI, Jamieson GG, Myers JC, Game PA, Devitt PG. (2002). Division of short gastric vessels at laparoscopic nissen fundoplication: a prospective double-blind randomized trial with 5-year follow-up. *Ann Surg*. 235(2):165-70.

O'Reilly MJ, Mullins SG, Saye WB, Pinto SE, Falkner PT. (1996). Laparoscopic posterior partial fundoplication: analysis of 100 consecutive cases. *Laparoendosc Surg*. 6(3):141-50

R: A language and environment for statistical computing [computer program]. Version 2.8.0. Vienna: Foundation for Statistical computing; 2008.

Rice S, Watson DI, Lally CJ, Devitt PG, Game PA, Jamieson GG. (2006). Laparoscopic anterior 180 degrees partial fundoplication: five-year results and beyond. *Arch Surg*. 141(3):271-5.

Spence GM, Watson DI, Jamiesion GG, Lally CJ, Devitt PG. (2006). Single center prospective randomized trial of laparoscopic Nissen versus anterior 90 degrees fundoplication. J Gastrointest Surg. 10(5):698-705.

Stewart GD, Watson AJ, Lamb PJ, Lee AJ, Krukowski ZH, Griffin SM, Paterson-Brown S. (2004). Comparison of three different procedures for antireflux surgery. *Br J Surg*. 91(6):724-9.

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. (2000). Methods for Meta-analysis in Medical Research. *John Wiley*: Chichester.

Walker SJ, Holt S, Sanderson CJ, Stoddard CJ. (1992). Comparison of Nissen total and Lind partial transabdominal fundoplication in the treatment of gastro-oesophageal reflux. *Br J Surg*. 79(5):410-4

Watson A, Jenkinson LR, Ball CS, Barlow AP, Norris TL. (19910. A more physiological alternative to total fundoplication for the surgical correction of resistant gastro-oesophageal reflux. *Br J Surg.* 78(9):1088-94.

Watson A, Spychal RT, Brown MG, Peck N, Callander N. (1995). Laparoscopic 'physiological' antireflux procedure: preliminary results of a prospective symptomatic and objective study. *Br J Surg*. 82(5):651-6

Watson DI, Jamieson GG, Pike GK, Davies N, Richardson M, Devitt PG. (1999). Prospective randomized double-blind trial between laparoscopic Nissen fundoplication and anterior partial fundoplication. *Br J Surg.* 86(1):123-30

Watson DI, Jamieson GG, Lally C, Archer S, Bessell JR, Booth M, Cade R, Cullingford G, Devitt PG, Fletcher DR, Hurley J, Kiroff G, Martin CJ, Martin IJ, Nathanson LK, Windsor JA; International Society for Diseases of the Esophagus--Australasian Section. (2004). Multicenter, prospective, double-blind, randomized trial of laparoscopic nissen vs anterior 90 degrees partial fundoplication. *Arch Surg*. 139(11):1160-7.

Watson DI, Pike GK, Baigrie RJ, Mathew G, Devitt PG, Britten-Jones R, Jamieson GG. (1997). Prospective double-blind randomised trail of laparoscopic Nissen fundoplication with division and without division of short gastric vessels. *Ann Surg*. 226(5):642-5.

Zornig C, Strate U, Fibbe C, Emmermann A, Layer P. (2002). Nissen vs Toupet laparoscopic fundo-plication. *Surg Endosc*. 16(5):758-66.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 157–157.

SESSION 4: SMALL AREA ESTIMATION

Chair: Malay Ghosh
University of Georgia, University of Florida and US Bureau of the Census
E-mail: ghoshm2000@yahoo.com

EMPIRICAL LIKELIHOOD FOR SMALL AREA ESTIMATION

Sanjay Chaudhuri
Department of Statistics and Applied probability, National University of Singapore, Singapore 117546
E-mail: stasc@nus.edu.sg

Malay Ghosh
Department of Statistics, University of Florida, Gainesville, FL 32611
E-mail: ghoshm2000@yahoo.com

ABSTRACT

Current methodologies in small area estimation are mostly either parametric or are heavily dependent on the assumed linearity of the estimators of the small area means. We discuss an alternative empirical likelihood based Bayesian approach. This approach neither requires a parametric likelihood nor needs to assume any linearity of the estimators. Moreover, the proposed method can handle both discrete and continuous data in a unified manner. Empirical likelihood for both area and unit level models are introduced. Performance of our method is illustrated through real datasets.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 158–158.

GENERALIZED MAXIMUM LIKELIHOOD METHOD IN LINEAR MIXED MODELS WITH AN APPLICATION IN SMALL-AREA ESTIMATION

Parthasarathi Lahiri
JPSM, 1218 Lefrak Hall, University of Maryland, College Park, MD 20742, USA
E-mail: plahiri@survey.umd.edu

Huilin Li
Division of Cancer Epidemiology and Genetics
National Cancer Institute, USA
E-mail: lih5@mail.nih.gov

ABSTRACT

Standard methods frequently produce zero estimates of dispersion parameters in the underlying linear mixed model. As a consequence, the EBLUP estimate of a small area mean reduces to a simple regression estimate. In this paper, we consider a class of generalized maximum residual likelihood estimators that covers the well-known profile maximum likelihood and the residual maximum likelihood estimators. The general class of estimators has a number of different estimators for the dispersion parameters that are strictly positive and enjoy good asymptotic properties. In addition, the mean squared error of the corresponding EBLUP estimator is asymptotically equivalent to those of the profile maximum likelihood and residual maximum likelihood estimators in the higher order asymptotic sense. However, the strictly positive generalized maximum likelihood estimators have an advantage over the standard methods in estimating the shrinkage parameters and in constructing the parametric bootstrap prediction intervals of the small area means. We shall illustrate our methodology using a real survey data.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 159–185.

Bayesian Benchmarking with Applications to Small Area Estimation

G.S. Datta, M. Ghosh, R. Steorts and J. Maples

University of Georgia, University of Florida and US Bureau of the Census

Abstract

It is well-known that small area estimation needs explicit or at least implicit use of models (cf. Rao, 2003). These model-based estimates can differ widely from the direct estimates, especially for areas with very low sample sizes. While model-based small area estimates are very useful, one potential difficulty with such estimates is that when aggregated, the overall estimate for a larger geographical area may be quite different from the corresponding direct estimate, the latter being usually believed to be quite reliable. This is because the original survey was designed to achieve specified inferential accuracy at this higher level of aggregation. The problem can be more severe in the event of model failure as often there is no real check for validity of the assumed model. Moreover, an overall agreement with the direct estimates at an aggregate level may sometimes be politically necessary to convince the legislators of the utility of small area estimates.

One way to avoid this problem is the so-called "benchmarking approach" which amounts to modifying these model-based estimates so that one gets the same aggregate estimate for the larger geographical area. Currently, the most popular approach is the so-called "raking" or ratio adjustment method which involves multiplying all the small area estimates by a constant data-dependent factor so that the weighted total agrees with the direct estimate. There are alternate proposals, mostly from frequentist considerations, which meet also the aforementioned benchmarking criterion.

We propose in this paper a general class of constrained Bayes estimators which achieve as well the necessary benchmarking. Many of the frequentist estimators, including some of the raked estimators, follow as special cases of our general result. Explicit Bayes estimators are derived which benchmark the weighted mean or both the weighted mean and weighted variability. We illustrate our methodology by developing poverty rates in school-aged children at the state level, and then benchmarking these estimates to match at the national level. Unlike the existing frequentist benchmarking literature which is primarily based on linear models, the proposed Bayesian approach can accomodate any arbitrary model, and the benchmarked Bayes estimators are based only on the posterior mean and the posterior variance-covariance matrix.

Keywords: Area-level, penalty parameter, two-stage, weighted mean, weighted variability.

1 Introduction

Empirical Bayesian (EB) and hierarchical Bayesian (HB) methods are now widely used for simultaneous inference. The biggest advantage of these methods is their ability to enhance the precision of individual estimators by "borrowing strength" from similar other estimators. These methods have been used very successfully in a wide array of disciplines including sociology, epidemiology, business, economics, political science and insurance.

One important application of EB and HB methods is in small area estimation. The topic has become a prime area of research globally in many government statistical agencies. As an example, in the United States Bureau of the Census, the Small Area Income and Poverty Estimates (SAIPE) and Small Area Health Insurance Estimates (SAHIE) research groups are actively engaged in various small area projects. A typical small area estimation problem involves simultaneous estimation of quantities of interest for several small geographical areas (for example counties) or several small domains cross-classified by age, sex, race and other demographic/geographic characteristics. The need for borrowing strength arises in these problems because the original survey was designed to achieve a specific accuracy at a higher level of aggregation than that of small areas or domains. Due to limited resources, the same data needs to be used at lower levels of geography, but individual direct estimates are usually accompanied with large standard errors and coefficients of variation.

It is well-known that small area estimation needs explicit or at least implicit use of models (cf. Rao, 2003). These model-based estimates can differ widely from the direct estimates, especially for areas with very low sample sizes. One potential difficulty with model-based estimates is that when aggregated, the overall estimate for a larger geographical area may be quite different from the corresponding direct estimate, the latter being often believed to be quite reliable. This is because the original survey was designed to achieve specified inferential accuracy at this higher level of aggregation. As an example, one may mention the SAIPE county estimates of the United States Bureau of the Census based on the American Community Survey (ACS) data which are controlled so that the overall weighted estimates agree with the corresponding state estimates which though model-based, are quite close to the direct estimates. The problem can be more severe in the event of model failure as often there is no real check for validity of the assumed model. Pfeffermann and Tiller (2006), in the context of time series models for small area estimation, noted that benchmarked estimates reflect a sudden change in the direct estimates due to some external shock not accounted for in the model much faster than the model-based estimates. Moreover, an overall agreement with the direct estimates at some higher level may sometimes be politically necessary to convince the legislators of the utility of small area estimates (Fay and Herriot, 1979).

One way to avoid this problem is the so-called "benchmarking approach" which amounts to modifying these model-based estimates so that one gets the same aggregate estimate for the larger geographical area. A simple illustration is to modify the model-based state level estimates so that one matches the national estimates. Currently the most popular approach is the so-called "raking" or ratio adjustment method which involves multiplying all the small area estimates by a constant factor so that the weighted total agrees with the direct estimate. The raking approach is ad-hoc, although, later in this paper, we have given it a constrained Bayes interpretation.

We now discuss some of the existing benchmarking literature (mostly frequentist) for small area estimation. We begin with a simple stratified sampling model with m strata having population sizes $N_1, \cdots, N_m$. You and Rao (2002) required estimates $\hat{\theta}_i$ of the stratum means θ_i such that $\sum_{i=1}^m N_i\hat{\theta}_i$ equals the direct survey regression estimator of the overall total. You and Rao (2002, 2003) considered unit-level small area models with known survey weights attached to the different units. More recently, Fuller (2007) has considered a procedure which allows sampling of small areas from a larger pool of small areas and requires a weighted sum of small area predictors equal a design consistent estimator of the population total.

Pfeffermann and Tiller (2006) considered benchmarking in small area estimation based on time series data. They took a frequentist approach and used the Kalman filter for time series to obtain first the model-based estimators of the small area means. Then they obtained the benchmarked estimators satisfying certain agreement of these estimators with some direct estimators at a higher level of aggregation. They did not use any cross-sectional model to borrow strength from other areas. This is an example of internal benchmarking. Earlier examples of internal benchmarking include those of Pfeffermann and Barnard (1991), Isaki, Tsay and Fuller (2000) and Wang, Fuller and Qu (2008).

You *et al.* (2004) considered benchmarking of HB estimates through ratio adjustment for area-level models. Nandram et al. (2007) suggested a different benchmarked HB estimation of small area means based on unit level models. In this exact benchmarking, they proceeded with the conditional distribution of the unobserved units within a small area given the benchmark constraint on the total of all the units in that area. A disadvantage to such an approach is that results can differ depending on which unit is dropped.

The objective of this paper is to develop a general class of Bayes estimators which achieves the necessary benchmarking. For definiteness, we will concentrate only on area-level models. As we will see later, many of the currently proposed benchmarked estimators including the raked ones belong to the proposed class of Bayes estimators. In particular, some of the estimators proposed in Pfeffermann and Barnard (1991), Isaki, Tsay and Fuller (2000), Wang, Fuller and Qu (2008) and You, Rao and Dick (2004) are members of this class.

The proposed Bayesian approach has been motivated from a decision-theoretic framework, and is similar in spirit to one in Louis (1984) and Ghosh (1992) who considered constrained Bayes and empirical Bayes estimators with a slightly different objective. It was pointed out in these papers that the empirical histogram of the posterior means of a set of parameters of interest is underdispersed as compared to the posterior histogram of the same set of parameters. Thus, adjustment of Bayes estimators is needed in order to meet the twin objective of accuracy and closeness of the histogram of the estimates with the posterior estimate of the parameter histogram. In contrast, the present method achieves matching with some aggregate measure such as a national total. In addition, if necessary, we can also match the empirical variability of the estimates of the parameters of interest with the posterior variability of these parameters, or even some preassigned number.

The organization of the remaining sections is as follows. In Section 2, we develop the constrained Bayes estimators requiring only the matching of a weighted average of small area means with some prespecified estimators. These prespecified estimators can be a weighted average of the direct small area estimators, a situation which will be referred to as *internal benchmarking*. On the other hand, if the prespecified estimator is obtained from some other source, for example, a different survey, census or other administrative records, then it becomes an instance of *external benchmarking*. We will also point out how the proposed benchmarked estimators arise as the limit of a general class of Bayes estimators where one needs only partial benchmarking. The general result is illustrated with several constrained benchmarked estimators of area-level means based on the usual random effects or the Fay-Herriot (1979) (also Pfeffermann and Nathan, 1981) model.

In Section 3, we develop benchmarked Bayes estimators which meet the dual objective of overall matching with the prespecified benchmarks as well as the variability agreement as mentioned earlier. We have also pointed out in this section how some of the benchmarked estimators proposed in a frequentist framework by earlier authors are indeed special cases of the proposed Bayesian estimators. Multiparameter extensions of these results will also be given in this section. Section 4 contains an application of the proposed method in a real small area problem. Section 5 contains a summary of the results developed in this paper along with a few suggestions for future research.

2 Benchmarked Bayes Estimators

2.1 Development of the Estimators

Let $\hat{\theta}_1, \cdots, \hat{\theta}_m$ denote the direct estimators of the m small area means $\theta_1, \ldots, \theta_m$. We write $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_m)^T$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$. Initially, we seek the benchmarked

Bayes estimator $\hat{\boldsymbol{\theta}}^{BM1} = (\hat{\theta}_1^{BM1}, \cdots, \hat{\theta}_m^{BM1})^T$ of $\boldsymbol{\theta}$ such that $\sum_{i=1}^m w_i \hat{\theta}_i^{BM1} = t$, where either t is prespecified from some other source or $t = \sum_{i=1}^m w_i \hat{\theta}_i$. The w_i are given weights attached to the direct estimators $\hat{\theta}_i$'s, and without any loss of generality, $\sum_{i=1}^m w_i = 1$. These weights may depend on $\hat{\boldsymbol{\theta}}$ (which is most often not the case), but do not depend on $\boldsymbol{\theta}$. For example, one may take $w_i = N_i / \sum_{j=1}^m N_j$, where the N_i are the population sizes for the m small areas.

A Bayesian approach to this end is to minimize the posterior expectation of the weighted squared error loss $\sum_{i=1}^m \phi_i E[(\theta_i - e_i)^2 | \hat{\boldsymbol{\theta}}]$ with respect to the e_i's satisfying $\bar{e}_w = \sum_{i=1}^m w_i e_i = t$. These ϕ_i may be the same as the w_i, but that need not always be the case. Also, like w_i, ϕ_i may depend on $\hat{\boldsymbol{\theta}}$, but not on $\boldsymbol{\theta}$. Wang *et al.* (2008) considered the same loss, but minimized instead the MSE (which amounts to conditioning on θ), and came up with a solution different from ours. Moreover, they restricted themselves to linear estimators. One of the advantages of the proposed Bayesian approach is that adjustment is possible for any general Bayes estimator, linear or non-linear.

The ϕ_i can be regarded as weights for a multiple-objective decision process. That is, each specific weight is relevant only to the decision-maker for the corresponding small area, who may not be concerned with the weights related to decision-makers in other small areas. Combining losses in such situations in a linear fashion is discussed for example in Berger (1985, p. 279).

We now prove a theorem which provides a solution to our problem. A few notations are needed before stating the theorem. Let $\hat{\theta}_i^B$ denote the posterior mean of θ_i, $i = 1, \cdots, m$ under a certain prior. The vector of posterior means and the the corresponding weighted average are denoted respectively by $\hat{\boldsymbol{\theta}}^B = (\hat{\theta}_1^B, \cdots, \hat{\theta}_m^B)^T$ and $\bar{\hat{\theta}}_w^B = \sum_{i=1}^m w_i \hat{\theta}_i^B$. Also, let $\boldsymbol{r} = (r_1, \cdots, r_m)^T$, where $r_i = w_i/\phi_i$, $i = 1, \cdots, m$, and $s = \sum_{i=1}^m w_i^2/\phi_i$. Then we have the following theorem.

Theorem 1. The minimizer $\hat{\boldsymbol{\theta}}^{BM1}$ of $\sum_{i=1}^m \phi_i E[(e_i - \theta_i)^2|\hat{\boldsymbol{\theta}}]$ subject to $\bar{e}_w = t$ is given by

$$\hat{\boldsymbol{\theta}}^{BM1} = \hat{\boldsymbol{\theta}}^B + s^{-1}(t - \bar{\hat{\theta}}_w^B)\boldsymbol{r}. \tag{1}$$

Proof. First rewrite $\sum_{i=1}^m \phi_i E[(e_i - \theta_i)^2|\hat{\boldsymbol{\theta}}] = \sum_{i=1}^m \phi_i V(\theta_i|\hat{\boldsymbol{\theta}}) + \sum_{i=1}^m \phi_i (e_i - \hat{\theta}_i^B)^2$. Now the problem reduces to minimization of $\sum_{i=1}^m \phi_i(e_i - \hat{\theta}_i^B)^2$ subject to $\bar{e}_w = t$. A Lagrangian multiplier approach provides the solution. But then one needs to show in addition that the solution provides a minimizer and not a maximizer. Alternately, we can use the identity

$$\sum_{i=1}^m \phi_i(e_i - \hat{\theta}_i^B)^2 = \sum_{i=1}^m \phi_i\{e_i - \hat{\theta}_i^B - s^{-1}(t - \bar{\hat{\theta}}_w^B)r_i\}^2 + s^{-1}(t - \bar{\hat{\theta}}_w^B)^2. \tag{2}$$

The solution is now immediate from (2).

Remark 1. The constrained Bayes benchmarked estimators $\hat{\boldsymbol{\theta}}^{BM1}$ as given in (1) can also be viewed also as limiting Bayes estimators under the loss

$$L(\boldsymbol{\theta}, \boldsymbol{e}) = \sum_{i=1}^m \phi_i(\theta_i - e_i)^2 + \lambda(t - \bar{e}_w)^2, \tag{3}$$

where $\boldsymbol{e} = (e_1, \cdots, e_m)^T$, and $\lambda(> 0)$ is the penalty parameter. Like the ϕ_i, the penalty parameter λ can differ for different policy makers. The Bayes estimator of $\boldsymbol{\theta}$

under the above loss (after some algebra) is given by

$$\hat{\boldsymbol{\theta}}_{\lambda}^{B} = \hat{\boldsymbol{\theta}}^{B} + (s+\lambda^{-1})^{-1}(t - \bar{\hat{\theta}}_{w}^{B})\boldsymbol{r}.$$

Clearly, when $\lambda \to \infty$, i.e., when one invokes the extreme penalty for not having the exact equality $\bar{e}_w = t$, one gets the estimator given in (1). Otherwise, λ serves as a trade-off between t and $\bar{\hat{\theta}}_{w}^{B}$ since

$$\boldsymbol{w}^T\hat{\boldsymbol{\theta}}_{\lambda}^{B} = \frac{s\lambda}{s\lambda+1}t + \frac{1}{s\lambda+1}\bar{\hat{\theta}}_{w}^{B}.$$

Remark 2. The balanced loss of Zellner (1986, 1988, 1994) is not quite the same as the one in Remark 1, and is given by

$$L(\boldsymbol{\theta}, \boldsymbol{e}) = \sum_{i=1}^{m} \phi_i(\theta_i - e_i)^2 + \lambda \sum_{i=1}^{m}(\hat{\theta}_i - e_i)^2.$$

This leads to the Bayes estimator $\hat{\boldsymbol{\theta}}^{B} + \lambda(\lambda\boldsymbol{I} + \boldsymbol{\phi})^{-1}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{B})$, where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{\phi} = \mathrm{Diag}(\phi_1, \cdots, \phi_m)$ which is a compromise between the Bayes estimator $\hat{\boldsymbol{\theta}}^{B}$ and the direct estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, and converges to the direct estimator as $\lambda \to \infty$ and to the Bayes estimator when $\lambda \to 0$.

We now provide a generalization of Theorem 1 where one considers multiple instead of one single constraint. As an example, for the SAIPE county level analysis, one may need to control the county estimates in each state so that their weighted total agrees with the corresponding state estimates. One now considers a more general quadratic loss given by

$$L(\boldsymbol{\theta}, \boldsymbol{e}) = (\boldsymbol{e} - \boldsymbol{\theta})^T\boldsymbol{\Omega}(\boldsymbol{e} - \boldsymbol{\theta}), \tag{4}$$

where $\boldsymbol{\Omega}$ is a positive definite matrix. The following theorem provides a Bayesian solution for the minimization of $E[L(\boldsymbol{\theta}, \boldsymbol{e})|\hat{\boldsymbol{\theta}}]$ subject to the constraint $\boldsymbol{W}^T\boldsymbol{e} = \boldsymbol{t}$, where $\boldsymbol{t}$ is a q-component vector, and $\boldsymbol{W}$ is a $m \times q$ matrix of rank $q(< m)$.

Theorem 2. The constrained Bayesian solution under the loss (4) is given by
$\hat{\boldsymbol{\theta}}^{MBM} = \hat{\boldsymbol{\theta}}^B + \boldsymbol{\Omega}^{-1}\boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{\Omega}^{-1}\boldsymbol{W})^{-1}(\boldsymbol{t} - \bar{\hat{\boldsymbol{\theta}}}_w^B)$, where $\bar{\hat{\boldsymbol{\theta}}}_w^B = \boldsymbol{W}^T\hat{\boldsymbol{\theta}}^B$.

Proof. First write
$E[(\boldsymbol{e}-\boldsymbol{\theta})^T\boldsymbol{\Omega}(\boldsymbol{e}-\boldsymbol{\theta})|\hat{\boldsymbol{\theta}}] = E[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}^B)^T\boldsymbol{\Omega}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}^B)|\hat{\boldsymbol{\theta}}]+(\boldsymbol{e}-\hat{\boldsymbol{\theta}}^B)^T\boldsymbol{\Omega}(\boldsymbol{e}-\hat{\boldsymbol{\theta}}^B)$. Hence, the problem reduces to minimization of $(\boldsymbol{e}-\hat{\boldsymbol{\theta}}^B)^T\boldsymbol{\Omega}(\boldsymbol{e}-\hat{\boldsymbol{\theta}}^B)$ with respect to $\boldsymbol{e}$ subject to $\boldsymbol{W}^T\boldsymbol{e} = \boldsymbol{t}$. The result follows from the identity

$$
\begin{aligned}
(\boldsymbol{e}-\hat{\boldsymbol{\theta}}^B)^T\boldsymbol{\Omega}(\boldsymbol{e}-\hat{\boldsymbol{\theta}}^B) &= [(\boldsymbol{e}-\hat{\boldsymbol{\theta}}^B - \boldsymbol{\Omega}^{-1}\boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{\Omega}^{-1}\boldsymbol{W})^{-1}(\boldsymbol{t}-\hat{\boldsymbol{\theta}}^B)]^T \\
&\times \boldsymbol{\Omega}[(\boldsymbol{e}-\hat{\boldsymbol{\theta}}^B - \boldsymbol{\Omega}^{-1}\boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{\Omega}^{-1}\boldsymbol{W})^{-1}(\boldsymbol{t}-\hat{\boldsymbol{\theta}}^B)] \\
&+ (\boldsymbol{t}-\bar{\hat{\boldsymbol{\theta}}}_w^B)^T(\boldsymbol{W}^T\boldsymbol{\Omega}^{-1}\boldsymbol{W})^{-1}(\boldsymbol{t}-\bar{\hat{\boldsymbol{\theta}}}_w^B).
\end{aligned}
$$

The choice of the weight matrix $\boldsymbol{\Omega}$ usually depends on the experimenter depending on how much penalty she/he is willing to assignfor a misspecified estimator. In the special case of a diagonal $\boldsymbol{\Omega}$, Wang *et al.* (2008) have argued in favor of $\omega_i = [\mathrm{Var}(\hat{\theta}_i)^{-1}$.

2.2 Relationship with some Existing Estimators

We now show how some of the existing benchmarked estimators follow as special cases of the proposed Bayes estimators. Indeed, the proposed class of Bayes estimators in this paper includes the some of the raked benchmarked estimators as well as some of the other benchmarked estimators proposed by several authors.

Example 1. It is easy to see why the raked Bayes estimators, considered for example in You and Rao (2004), belong to the general class of estimators proposed in Theorem 1. If one chooses (possibly quite artificially) $\phi_i = w_i/\hat{\theta}_i^B$, $i = 1, \cdots, m$, ($\hat{\theta}_i^B > 0$ for all $i = 1, \cdots, m$), then $\boldsymbol{r} = \hat{\boldsymbol{\theta}}^B$ and $s = \bar{\hat{\theta}}_w^B$. Consequently, the constrained Bayes estimator proposed in Theorem 1 simplifies to $(t/\bar{\hat{\theta}}_w^B)\hat{\boldsymbol{\theta}}^B$, which is the raked Bayes estimator. In particular, one can take $t = \bar{\hat{\theta}}_w$. We may also note that this choice of the ϕ_i's is different from the one in Wang *et al.* (2008) who considered $\phi_i = w_i/\hat{\theta}_i$.

Example 2. The next example considers the usual random effects model as considered in Fay and Herriot (1979) or Pfeffermann and Nathan (1981). Under this model, $\hat{\theta}_i|\theta_i \overset{ind}{\sim} N(\theta_i, D_i)$ and $\theta_i \overset{ind}{\sim} N(\boldsymbol{x}_i^T\boldsymbol{\beta}, \sigma_u^2)$, the $D_i(> 0)$ being known. For the HB approach, one then uses the the prior $\pi(\boldsymbol{\beta}, \sigma_u^2) = 1$ although other priors are also possible as long as the posteriors are proper. The HB estimators $E(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ cannot be obtained analytically, but it is possible to find them numerically either through Markov chain Monte Carlo (MCMC) or through numerical integration. Denoting the HB estimators by $\hat{\theta}_i^B$, one can obtain the benchmarked Bayes estimators $\hat{\theta}_i^{BM1}(i = 1, \cdots, m)$ by applying Theorem 1.

Example 3. Wang *et al.* (2008) considered a slightly varied form of the Fay-Herriot random effects model, the only change, in our notations, being that the marginal variance of the θ_i are now $z_i^2\sigma_u^2$, where the z_i are known. They did not assume normality, but restricted their attention to the class of linear estimators of $\boldsymbol{\theta}$, and benchmarked the best linear unbiased predictor (BLUP) of $\boldsymbol{\theta}$ when σ_u^2 is known. For this example, the benchmarked estimators given in (6) of Wang *et al.* are also derivable from Theorem 1. First for known σ_u^2, consider the uniform prior for $\boldsymbol{\beta}$. Write $B_i = D_i/(D_i + z_i^2\sigma_u^2)$, $\boldsymbol{B} = \text{Diag}(B_1, \cdots, B_m)$, $\boldsymbol{\Sigma} = \text{Diag}(D_1 + z_1^2\sigma_u^2, \cdots, D_m + z_m^2\sigma_u^2)$,

$\boldsymbol{X}^T = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m)$, and $\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\theta}}$, assuming $\boldsymbol{X}$ to be a full column rank matrix. Then the Bayes estimator of $\boldsymbol{\theta}$ is

$$\tilde{\boldsymbol{\theta}}^B = (\boldsymbol{I}_m - \boldsymbol{B})\hat{\boldsymbol{\theta}} + \boldsymbol{B}\boldsymbol{X}\tilde{\boldsymbol{\beta}},$$

which is the same as the BLUP of $\boldsymbol{\theta}$ as well. Now identify the $r_i's$ in this paper with the a_i of Wang and Fuller (2002) to get (26) in their paper.

Example 3 (continued). As shown in Wang *et al.* (2008), the Pfeffermann-Barnard (1991) estimator belongs to their (and accordingly our) general class of estimators where one chooses $\phi_i = w_i/\mathrm{Cov}(\hat{\theta}_i^B, \bar{\hat{\theta}}^B$, where the covariance is calculated over the joint distribution of $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$, treating $\boldsymbol{\beta}$ as an unknown but fixed parameter. Then $\boldsymbol{r}$ contains the elements of $\mathrm{Cov}(\hat{\theta}_i^B, \bar{\hat{\theta}}^B)$ as its components, while $s = V(\bar{\hat{\theta}}_w^B)$.

Instead of the constrained Bayes estimators as given in (1), it is possible to obtain constrained empirical Bayes (EB) estimators as well when one estimates the prior parameters from the marginal distribution of $\hat{\boldsymbol{\theta}}$ (after integrating out $\boldsymbol{\theta}$). The resulting EB estimators are given by

$$\hat{\boldsymbol{\theta}}^{EBM1} = \hat{\boldsymbol{\theta}}^{EB} + s^{-1}(t - \bar{\hat{\boldsymbol{\theta}}}_w^{EB})\boldsymbol{r}, \tag{5}$$

where $\hat{\boldsymbol{\theta}}^{EB} = (\hat{\theta}_1^{EB}, \cdots, \hat{\theta}_m^{EB})^T$ is an EB estimator of $\boldsymbol{\theta}$ and $\bar{\hat{\boldsymbol{\theta}}}_w^{EB} = \sum_{i=1}^m w_i\hat{\theta}_i^{EB}$.

Remark 3. In the model as considered in Example 3, for unknown σ_u^2, one gets estimators of $\boldsymbol{\beta}$ and σ_u^2 simultaneously from the marginals $\hat{\theta}_i \stackrel{\text{ind}}{\sim} N(\boldsymbol{x}_i^T\boldsymbol{\beta}, D_i + z_i^2\sigma_u^2)$ (Fay and Herriot, 1979; Prasad and Rao, 1990; Datta and Lahiri, 2000; Datta, Rao and Smith, 2005). Denoting the estimator of σ_u^2 by $\hat{\sigma}_u^2$, one estimates $\boldsymbol{\Sigma}$ by $\hat{\boldsymbol{\Sigma}} = \text{Diag}(D_1 + z_1^2\hat{\sigma}_u^2, \cdots, D_m + z_m^2\hat{\sigma}_u^2)$, $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\theta}}$ and $\boldsymbol{B}$ by $\hat{\boldsymbol{B}} = \boldsymbol{D}\hat{\boldsymbol{\Sigma}}^{-1}$, where $\boldsymbol{D} = \text{Diag}(D_1, \cdots, D_m)$. Denoting the resulting EB estimator of $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}^{EB}$, one gets

$$\hat{\boldsymbol{\theta}}^{EB} = (\boldsymbol{I}_m - \hat{\boldsymbol{B}})\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{B}}\boldsymbol{X}\hat{\boldsymbol{\beta}}. \tag{6}$$

The benchmarked EB estimator is now obtained from (5).

Remark 4. The benchmarked EB estimator as given in (5) includes the one given in Isaki, Tsay and Fuller (2000), where one takes ϕ_i as the reciprocal of the ith diagonal element of $\hat{\boldsymbol{\Sigma}}$ for all $i = 1, \cdots, m$. Another option is to take ϕ_i as the reciprocal of an estimator of $\text{V}(\hat{\theta}_i^B)$, the variance being computed once again under the joint distribution of $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$, treating $\boldsymbol{\beta}$ as an unknown but fixed parameter.

3 Benchmarking with both Mean and Variability Constraints

There are situations where in addition to benchmarking the first moment, one demands also benchmarking the variability of the Bayes estimators as well. We will address this issue in the special case when $\phi_i = cw_i$ for some $c(> 0)$, $i = 1, \cdots, m$. In this case, $\hat{\theta}_i^{BM1}$ given in (1) simplifies to $\hat{\theta}_i^B + (t - \bar{\hat{\theta}}_w^B)$ for all $i = 1, \cdots, m$. This itself is not a very desirable estimator since then $\sum_{i=1}^m w_i(\hat{\theta}_i^{BM1} - t)^2 = \sum_{i=1}^m w_i(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)^2$. It can be shown as in Ghosh (1992) that $\sum_{i=1}^m w_i(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)^2 < \sum_{i=1}^m w_i E[(\theta_i - \bar{\theta}_w)^2|\hat{\boldsymbol{\theta}}]$. In other words, the weighted ensemble variability of the estimators $\hat{\theta}_i^{BM1}$ is an underestimate of the posterior expectation of the corresponding weighted ensemble variability of the population parameters. To address this issue, or from other considerations, we will consider estimators $\hat{\theta}_i^{BM2}$, $i = 1, \ldots, m$ which satisfy two constraints, namely, (i) $\sum_{i=1}^m w_i \hat{\theta}_i^{BM2} = t$ and (ii) $\sum_{i=1}^m w_i(\hat{\theta}_i^{BM2} - t)^2 = H$, where H is a preassigned number taken from some other source, for example from census data, or it could be $\sum_{i=1}^m w_i E[(\theta_i - \bar{\theta}_w)^2|\hat{\boldsymbol{\theta}}]$ more in the spirit of Louis (1984) and Ghosh (1992). Subject to these two constraints, one minimizes $\sum_{i=1}^m w_i E[(\theta_i - e_i)^2|\hat{\boldsymbol{\theta}}]$. The following theorem provides the resulting estimator.

Theorem 3. Subject to (i) and (ii), the benchmarked Bayes estimators of θ_i ($i = 1, \cdots, m$) are given by

$$\hat{\theta}_i^{BM2} = t + a_{CB}(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B), \tag{7}$$

where $a_{CB}^2 = H / \sum_{i=1}^m w_i(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)^2$. Note that $a_{CB} \geq 1$ when $H = \sum_{i=1}^m w_i E[(\theta_i - \bar{\theta}_w)^2|\hat{\boldsymbol{\theta}}]$.

Proof. As in Theorem 1, the problem reduces to minimization of $\sum_{i=1}^m w_i(e_i - \hat{\theta}_i^B)^2$. We will write

$$\sum_{i=1}^m w_i(e_i - \hat{\theta}_i^B)^2 = \sum_{i=1}^m w_i[(e_i - \bar{e}_w) - (\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)]^2 + (\bar{e}_w - \bar{\hat{\theta}}_w^B)^2. \tag{8}$$

Now define two discrete random variables Z_1 and Z_2 such that

$$P(Z_1 = e_i - \bar{e}_w, Z_2 = \hat{\theta}_i^B - \bar{\hat{\theta}}_w^B) = w_i,$$

$i = 1, \cdots, m$. Hence,

$$\sum_{i=1}^m w_i[(e_i - \bar{e}_w) - (\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)]^2 = V(Z_1) + V(Z_2) - 2Cov(Z_1, Z_2)$$

which is minimized when the correlation between Z_1 and Z_2 equals 1, i.e.

$$e_i - \bar{e}_w = a(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B) + b, \tag{9}$$

$i = 1, \cdots, m$ with $a > 0$. Multiplying both sides of (9) by w_i and summing over $i = 1, \cdots, m$, one gets $b = 0$. Next squaring both sides of (9), then multiplying both sides by w_i and summing over $i = 1, \cdots, m$, one gets $H = a^2 \sum_{i=1}^m w_i(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)^2$ due to condition (ii). Finally, by condition (i), the result follows from (9).

Remark 5. As in the case of Theorem 1, it is possible to work with arbitrary ϕ_i rather than $\phi_i = w_i$ for all $i = 1, \cdots, m$. But then one does not get a closed form minimizer, although it can be shown that such a minimizer exists. We can also provide an algorithm for finding this minimizer numerically.

The multiparameter extension of the above result proceeds as follows. Suppose now $\hat{\boldsymbol{\theta}}_1, \cdots, \hat{\boldsymbol{\theta}}_m$ are the q-component direct estimators of the small area means $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_m$. We generalize the constraints (i) and (ii) as (iM) $\bar{\boldsymbol{e}}_w = \sum_{i=1}^m w_i \boldsymbol{e}_i = \boldsymbol{t}$ for some specified $\boldsymbol{t}$ and (iiM) $\sum_{i=1}^m w_i(\boldsymbol{e}_i - \bar{\boldsymbol{e}}_w)(\boldsymbol{e}_i - \bar{\boldsymbol{e}}_w)^T = \boldsymbol{H}$, where $\boldsymbol{H}$ is a positive definite (possibly data dependent matrix), and is often taken as $\sum_{i=1}^m w_i E[(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_w)(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_w)^T | \hat{\boldsymbol{\theta}}]$. The second condition is equivalent to $\boldsymbol{c}^T \{\sum_{i=1}^m w_i(\boldsymbol{e}_i - \bar{\boldsymbol{e}}_w)(\boldsymbol{e}_i - \bar{\boldsymbol{e}}_w)^T\} \boldsymbol{c} = \boldsymbol{c}^T \boldsymbol{H} \boldsymbol{c}$ for every $\boldsymbol{c} = (c_1, \cdots, c_q)^T \neq \boldsymbol{0}$ which simplifies to $\sum_{i=1}^m w_i \{\boldsymbol{c}^T(\boldsymbol{e}_i - \bar{\boldsymbol{e}}_w)\}^2 = \boldsymbol{c}^T \boldsymbol{H} \boldsymbol{c}$. An argument similar as before now leads to $\boldsymbol{c}^T \hat{\boldsymbol{\theta}}_i^{BM2} = \boldsymbol{c}^T \bar{\hat{\boldsymbol{\theta}}}_w + a_{CB} \boldsymbol{c}^T (\hat{\boldsymbol{\theta}}_i^B - \bar{\hat{\boldsymbol{\theta}}}_w^B)$ for every $\boldsymbol{c} \neq \boldsymbol{0}$, where $\hat{\boldsymbol{\theta}}_i^B$ is the posterior mean of $\boldsymbol{\theta}_i$, $\bar{\hat{\boldsymbol{\theta}}}_w^B = \sum_{i=1}^m w_i \hat{\boldsymbol{\theta}}_i^B$, and $a_{CB}^2 = \boldsymbol{c}^T \boldsymbol{H} \boldsymbol{c} / \sum_{i=1}^m w_i \{\boldsymbol{c}^T (\hat{\boldsymbol{\theta}}_i^B - \bar{\hat{\boldsymbol{\theta}}}_w^B)\}^2$. The coordinatewise benchmarked Bayes estimators are now obtained by putting $\boldsymbol{c} = (1, 0, \cdots, 0)^T, \cdots, (0, 0, \cdots, 1)^T$ in succession.

The proposed approach can be extended also to a two-stage benchmarking somewhat similar to what is considered by Pfeffermann and Tiller (2006). To cite an example, consider the SAIPE scenario where we want to estimate the number of poor school children in different counties within a state, as well as those numbers within the different school districts in all these counties. Let $\hat{\theta}_i$ denote the Current Population Survey (CPS) estimate of θ_i, the true number of poor school children for the ith county and $\hat{\theta}_i^B$ the corresponding Bayes estimate, namely the posterior mean. Subject to the constraints $\bar{e}_w = \sum_{i=1}^m w_i e_i = \sum_{i=1}^m w_i \hat{\theta}_i = \bar{\hat{\theta}}_w$, and $\sum_{i=1}^m w_i (e_i - \bar{e}_w)^2 = H$, the benchmarked Bayes estimate for θ_i in the ith county is $\hat{\theta}_i^{BM2}$ as given in (7). Next, suppose that $\hat{\xi}_{ij}$ is the CPS estimator of ξ_{ij}, the true number of poor school children for the jth school district in the ith county and η_{ij} is the weight attached to the direct CPS estimator of $\xi_{ij}, j = 1, \ldots, n_i$. We seek estimators e_{ij} of ξ_{ij} such that (i) $\bar{e}_{i\eta} = \sum_{j=1}^{n_i} \eta_{ij} e_{ij} = \bar{\xi}_{i\eta}^{BENCH}$, the benchmarked estimator of $\bar{\xi}_{i\eta} = \sum_{j=1}^{n_i} \eta_{ij} \xi_{ij}$, and (ii) $\sum_{j=1}^{n_i} \eta_{ij} (e_{ij} - \bar{e}_{i\eta})^2 = H_i^*$ for some preassigned H_i^*, where again H_i^* can be taken as $\sum_{j=1}^{n_i} \eta_{ij} E[(\xi_{ij} - \bar{\xi}_{i\eta})^2 | \hat{\boldsymbol{\xi}}_i]$, $\hat{\boldsymbol{\xi}}_i$ being the vector with elements ξ_{ij}. A benchmarked estimator similar to (7) can now be found for the ξ_{ij} as well.

4 An Illustrative Example

The motivation behind this example is primarily to illustrate how the proposed Bayesian approach can be used for real life data. The Small Area Income and Poverty Estimates (SAIPE) program at the U.S. Bureau of the Census produces model-based estimates of the number of poor school-aged children (5-17 years old) at the national, state, county and school district levels. the school district estimates are used by the Department of Education to ased state estimates were benchmarkedllocate funds under the No Child Left Behind Act of 2001. In the SAIPE program, the model-based state estimates were benchmarked using ratio adjustments to the national design-based estimate of the number of poor school aged children from the Annual Social and Economic Supplement (ASEC) of the CPS up through 2004, while ACS estimates are used from 2005 onwards. Additionally, the model-based county estimates are benchmarked to the model-based state estimates in a hierarchical fashion, once again using ratio adjustments. In this section we will consider the implications of different benchmarking methods, using different weights on the state level estimates applying Theorems 1 and 3.

In the SAIPE program, the state model for poverty rates in school-aged children follows the basic Fay-Herriot framework (see e.g. Bell, 1999):

$$\hat{\theta}_i = \theta_i + e_i \tag{10}$$

$$\theta_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + u_i \tag{11}$$

where θ_i is the true state level poverty rate, $\hat{\theta}_i$ is the direct survey estimate (from CPS ASEC), e_i is the sampling error term with assumed known variance D_i, $\boldsymbol{x}_i$ are the predictors, $\boldsymbol{\beta}$ is the vector of regression coefficients and u_i is the model error with constant variance σ_u^2. The explanatory variables in the model are: IRS income tax based pseudo-estimate of the child poverty rate, IRS non-filer rate and the residual term from the regression of the 1990 Census estimated child poverty rate. The parameters $(\boldsymbol{\beta}, \theta_i, \sigma_u^2)$ are estimated using numerical integration for Bayesian inference (Bell, 1999).

The state estimates were benchmarked to the CPS direct estimate of the national school-aged child poverty rate until 2004. The weights, w_i, to calibrate the state's poverty rates to the national poverty rate, are proportional to the population estimates of the number of school-aged children in each state. Three different sets of risk function weights, ϕ_i, will be used to benchmark the estimated state poverty rates based on Theorem 1. The first set of weights will be the weights used in the benchmarking, i.e. $\phi_i = w_i$. The second set of weights creates the ratio adjusted benchmarked estimators, $\phi_i = w_i/\hat{\theta}_i^B$ (Example 1). The third set of weights uses the results from Pfeffermann and Barnard (1991) where $\phi_i = w_i/Cov(\hat{\theta}_i^B, \bar{\hat{\theta}}_w^B)$ (Example 3). Let the set of benchmarked estimates be denoted as $\hat{\boldsymbol{\theta}}^{(1)}$, $\hat{\boldsymbol{\theta}}^{(r)}$ and $\hat{\boldsymbol{\theta}}^{(PB)}$ respectively. Finally, we will benchmark the state poverty estimates using the results from Theorem 3 and denote the estimator as $\hat{\boldsymbol{\theta}}^{(2)}$.

To compute $\sum_i w_i E[(\theta_i - \bar{\theta}_w)^2|\hat{\boldsymbol{\theta}}]$, two methods are given. The first method uses the first two posterior moments of $\boldsymbol{\theta}$.

$$\begin{aligned} H &= \sum_i w_i E[(\theta_i - \bar{\theta}_w)^2|\hat{\boldsymbol{\theta}}] \\ &= E\left[\boldsymbol{\theta}^T(\boldsymbol{W} - \boldsymbol{w}\boldsymbol{w}^T)\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right] \\ &= \text{trace}\left[(\boldsymbol{W} - \boldsymbol{w}\boldsymbol{w}^T)(\text{Var}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) + E(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})E(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})^T)\right], \end{aligned} \tag{12}$$

where $\boldsymbol{W} = \text{diag}(w_1, \ldots, w_m)$ and $\boldsymbol{w} = (w_1, \ldots, w_m)^T$. The second method is to compute the posterior mean of $\sum_i w_i(\theta_i - \bar{\theta}_w)^2$ from the MCMC output of the Gibbs sampler. While both give equivalent values to benchmark the variability, the second method may be more practical as the number of small areas becomes large because it does not require manipulating an $m \times m$ matrix.

Table 1: Benchmarking Statistics for ASEC CPS

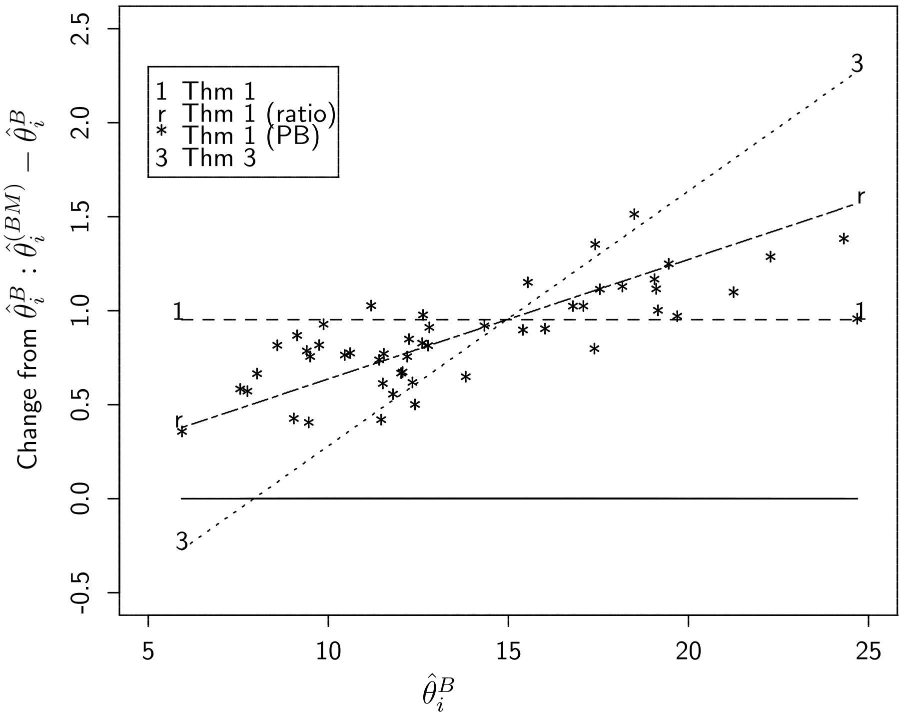
year	t	$\bar{\hat{\theta}}_w^B$	a_{CB}
1995	18.7	17.9	1.09
1997	18.4	17.8	1.07
1998	17.5	16.8	1.14
1999	15.9	14.9	1.14
2000	14.6	15.4	1.11
2001	14.8	15.3	1.10

For benchmarking, as given by Theorems 1 and 3, the key summary quantities are $t = \sum_i w_i \hat{\theta}_i$, $\bar{\hat{\theta}}_w^B = \sum_i w_i \hat{\theta}_i^B$ and a_{CB}. As noted earlier in Theorem 3, with the choice of H as given in (12), $a_{CB} \geq 1$. Six years of historical data from the CPS and the SAIPE program are analyzed and benchmarked using the four criteria mentioned above. Table 1 gives the key quantities for these six years. The hierarchical Bayes estimates for the years 1995-1999 underestimate the benchmarked poverty rate and overestimate the poverty rate in 2000 and 2001. Even if the estimate $\bar{\hat{\theta}}_w^B$ is close to the benchmarked value t, there is still a strong desire to have exact agreement between the quantities when producing official statistics.

Figure 1 shows the differences of the various benchmarked estimates from the hierarchical Bayes estimate, $\hat{\theta}_i^B$, made for the year 1999 when the overall poverty level had to be raised to agree with the national direct estimate. Figure 2 shows the differences for year 2000 when the overall poverty level had to be lowered to obtain agreement. The differences of the benchmarked estimators $\hat{\boldsymbol{\theta}}^{(1)}$, $\hat{\boldsymbol{\theta}}^{(r)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ from the HB estimator all fall on straight lines that pass through the same point $(\bar{\hat{\theta}}_w^B, t)$. In fact, these benchmarked estimators can be written in the form:

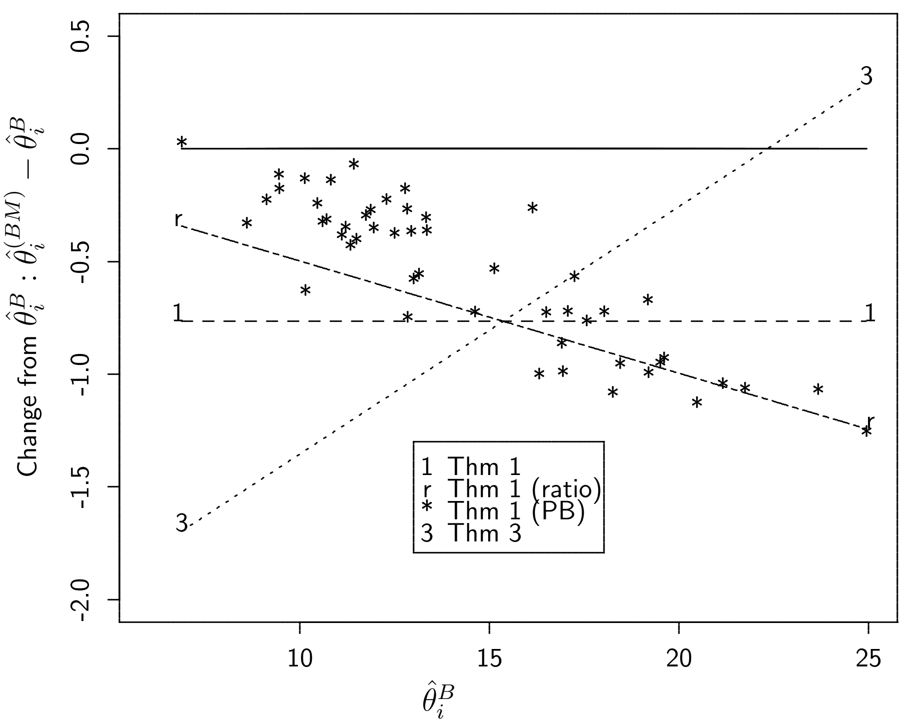
$$\hat{\theta}_i^{BM} = t + \alpha(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)$$

Figure 1: Change due to Benchmarking: 1999



where $\alpha = 1$ for $\hat{\boldsymbol{\theta}}^{(1)}$, $\alpha = t/\bar{\hat{\theta}}_w^B$ for $\hat{\boldsymbol{\theta}}^{(r)}$ and $\alpha = a_{CB}$ for $\hat{\boldsymbol{\theta}}^{(2)}$. The slopes of the lines in Figures 1 and 2 for differences in the benchmarked estimates from $\hat{\theta}_i$ are $\alpha - 1$. The slopes for $\hat{\theta}_i^{(1)}$ and $\hat{\theta}_i^{(r)}$ depend on whether the benchmarked total t is larger or smaller than the model-based estimate $\bar{\hat{\theta}}_w^B$. However, since $a_{CB} \geq 1$, the slope for the difference $\hat{\theta}_i^{(2)} - \hat{\theta}_i^B$ will always be non-negative. The Pfeffermann-Barnard benchmarked estimator does not follow this form. However, it does show a trend in a similar direction as the benchmarked estimator $\hat{\theta}_i^{(r)}$ based on ratio adjustment.

Figure 2: Change due to Benchmarking: 2000



5 Summary and Conclusion

The paper develops some general Bayesian benchmarked estimators with either a single or multiple benchmarking constraints. Much of the previous benchmarking literature is devoted to a single benchmarked constraint and also are restricted to linear models. The proposed Bayesian approach does not require either one of the two, and has a much wider applicability than what is known so far. In addition, unlike the existing benchmarking literature which deals almost exclusively to benchmarking the total or the weighted total, the present work allows benchmarking either with respect to some weighted total or with respect to both weighted total and weighted variability. Moreover, the proposed benchmarked Bayes estimators include as special cases many benchmarked estimators proposed earlier. This includes not only the raked estimators, but also other benchmarked estimators proposed by Pfeffermann and Barnard (1991), Isaki, Tsay and Fuller (2000), and Wang *et al.* (2008) among others. Further work needs evaluation of the benchmarked EB estimators, in particular, development of mean squared errors of EB benchmarked estimators as well as the associated confidence intervals.

The example is primarily intended to illustrate the proposed methodology. However, its analysis demonstrates some features of practical interest. In particular, the figures given in Table 1 show that at the turn of the century, there is a reverse trend in the weighted aggregate of the model-based state level estimates of the number of poor school-aged (5-17 years old) than what was evidenced in the years 1995 and 1997-1999. Specifically, in the years 2000 and 2001, this weighted aggregate exceeded the national total, while opposite was the case prior to the year 2000. While not too much conclusion should be drawn out of this, there is a mild suggestion that this particular feature is possibly due to change in the method of estimation starting from 2000.

Acknowledgments

We thank Don Luery for a comment which led to the general loss as considered in Theorem 1. The paper has also benefitted from the comments of William R. Bell and Donald J. Malec. Most of this work was completed when Ghosh was visiting the United States Bureau of the Census as an ASA/NSF/Census Senior Research Fellow. His research was also partially supported by a NSF Grant SES-0631426 and an NSA Grant MSPF-07G-097. Datta's research was partially supported by an NSA Grant MSPF-07G-082. The views expressed in this paper are due to the authors and do not reflect in any way those of the United States Bureau of the Census.

References

1. Bell, W.R. (1999). Accounting for uncertainty about variances in small area estimation. *Bulletin of the International Statistical Institute*, 52nd Session, Helsinki.

2. Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd Edition. Springer-Verlag, New York.

3. Datta, G.S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, **10**, 613-627.

4. Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, **92**, 183-196.

5. Fay, R.E. and Herriot, R.A. (1979). Estimates of income from small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269-277.

6. Fuller, W.A. (2007). Small area prediction subject to a restriction. Preprint.

7. Ghosh, M. (1992). Constrained Bayes estimation with applications. *Journal of the American Statistical Association*, **87**, 533-540.

8. Isaki, C.T., Tsay, J.H. and Fuller, W.A. (2000). Estimation of census adjustment factors. *Survey Methodology*, **26**, 31-42.

9. Louis, T.A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association*, **79**, 393-398.

10. Pfeffermann, D. and Barnard, C.H. (1991). Some new estimators for small area means with application to the assessment of farmland values. *Journal of Business and Economic Statistics*, **9**, 31-42.

11. Pfeffermann, D. and Nathan, G. (1981). Regression analysis of data from a cluster sample. *Journal of the American Statistical Association*, **76**, 681-689.

12. Pfeffermann, D. and Tiller, R. (2006). Small-area estimation with state-space models subject to benchmark constraints. *Journal of the American Statistical Association*, **101**, 1387-1397.

13. Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163-171.

14. Nandram, B., Toto, Ma. C.S. and Choi, J.W. (2007). A Bayesian benchmarking for small areas. Preprint.

15. Rao, J.N.K. (2003). *Small Area Estimation.* Wiley, New York.

16. Wang, J. and Fuller, W.A. (2002). Small area estimation under a restriction. *Proceedings of the Section on Survey Research methods of the American Statistical Association*, 3627-3630.

17. You, Y. and Rao, J.N.K. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *The Canadian Journal of Statistics*, **30**, 431-439.

18. You, Y. and Rao, J.N.K. (2003). Pseudo hierarchical Bayes small area estimation combining unit level models and survey weights. *Journal of Statistical Planning and Inference*, **111**, 197-208.

19. You, Y., Rao, J.N.K. and Dick, P. (2004). Benchmarking hierarchical Bayes small area estimators in the Canadian census undercoverage estimation. *Statistics in Transition*, **6**, 631-640.

20. Zellner, A. (1986). Further results on Bayesian minimum expected loss (MELO) estimates and posterior distributions for structural coefficients. In *Advances in Econometrics.* Ed. D.J. Slottje. JAI Press Inc., pp, 171-182.

21. Zellner, A. (1988). Bayesian analysis in econometrics. *Journal of Econometrics*, **37**, 27-50.

22. Zellner, A. (1994). Bayesian and non-Bayesian estimation using balanced loss functions. In *Statistical Decision Theory and Related Topics V*. Eds. S.S. Gupta and J.O. Berger. Springer-Verlag, New York, pp. 377-390.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 186–204.

Small Area Estimation, Some New Developments and Applications

Danny Pfeffermann

Southampton Statistical Sciences Research Institute, University of Southampton, S[illegible]1[illegible]1[illegible]

Department of Statistics, Hebrew University of Jerusalem, 91905, Israel

msdanny@soton.ac.il

Abstract

The purpose of this paper is to provide a critical review of some of the new important developments in small area estimation (SAE) methods that I am aware of. Rao (2003) published a very comprehensive book on SAE that covers all the main developments in this topic until that time and so the focus of this review is on new developments in the last [illegible] years. The review covers both design-based and model-dependent methods with emphasis on point prediction of the target area quantities, and mean square error assessments. The style of the paper is similar to the style of my previous review published in 2002, explaining the new problems investigated and describing the proposed solutions, but without dwelling on theoretical details, which can be found in the original articles. For further clarity and to make this paper more self contained, I also repeat shortly some of the 'old' developments. I am hoping that this paper will be useful both to researchers who like to learn more on the research carried out in SAE and to practitioners who might be interested in the application of the new methods.

Key words: Benchmarking, Calibration, Errors in variables, [illegible], M-quantiles, Ordered means, Outliers, P-spline.

1. Preface

In 2002 I published a review paper in the *International Statistical Review* with a similar title. At that year small area estimation (SAE) was flourishing, both in terms of research and applications, but my own feeling in those days was that the topic has been more or less exhausted in terms of research and that it will just turn into a routine application as part of sample survey practice. As the last [illegible] years show, I was completely wrong and not only that research in this area is accelerating, it now involves some of the best known statisticians, who otherwise are not involved in survey sampling theory or applications. The diversity of new problems investigated is overwhelming, and the solutions proposed are not only elegant and employ sophisticated statistical techniques, but are also very practical.

Rao (2003) published a very comprehensive book on SAE that covers all the main developments in this topic until that time. As in other areas of science, new developments in SAE are published in all kinds of journals, and some developments have not been published yet. Since SAE is so broadly applied, I thought that the time is ripe for a new critical review paper that focuses on some of the main developments in the last [illegible] years or so that I am aware of. The style of the paper is similar to the style of my previous review, explaining the new problems investigated and describing the proposed solutions, but without dwelling on theoretical details, which can be found in the original articles. For further clarity and to make this paper self contained, I also repeat shortly some of the "old" developments. I am hoping that this paper will be useful both to researchers who like to learn more on research carried out in SAE and to practitioners who might be interested in the application of the new methods.

2. Some background

The problem of SAE is how to produce reliable estimates of characteristics of interest such as means, counts, quantiles, etc., for small areas or domains for which only small samples or no samples are available. The latter case of no samples clearly requires the use of statistical models that define how to borrow information from neighboring areas or over time. As I shall explain later, no design-based (randomization) theory exists for the prediction of small area quantities in areas with no samples. Although the point estimators or predictors are usually of first priority, a related problem is how to assess the estimation error.

The use of the term "SAE" is a bit confusing, since it is the size of the sample in the area that creates the problem and not the size of the area. Also, the term "areas" does not necessarily refer to geographical zones and may define another grouping, such as socio-demographic groups or types of industry, in which case it is often referred to as small domain estimation. Closely related terms in common use are "poverty mapping" or "disease mapping", which amount to SAE of poverty measures or disease incidence and then presenting the results on a map, with different colors defining different levels of the estimators. What is common to all small area estimation problems is that estimators and estimators of errors such as mean square error are required for every area separately, and not just as an average over all the areas under consideration.

The great importance of SAE stems from the fact that many new policies such as fund allocation for needed areas, new educational or health programmes and environmental planning rely heavily on these estimates. Interest in SAE methods has further enhanced in recent years due to the reliance of modern censuses in many countries on administrative records. SAE is used to test and correct the administrative data for under-coverage or over-coverage, and also to supplement information not contained in these records.

SAE methods can be divided broadly into "design-based" methods and "model-based" methods. The latter methods use either the frequentist approach or a full Bayesian methodology, and in some cases combine the two, which is known in the SAE literature as "empirical Bayes". Design-based methods often use a model for the construction of the estimators (known as "model assisted"), but the bias, variance and other properties of the estimators are evaluated under the randomization (design-based) distribution. The randomization distribution is over all the samples possibly selected from the target population of interest under the sampling design used to select the sample, with the population values considered as fixed values. Model-based methods on the other hand generally condition on the selected sample and the inference is with respect to the underlying model.

The common feature to design-based and model-based SAE is the use of covariate (auxiliary) information as obtained from surveys or administrative records such as censuses or registers. Some estimators only require knowledge of the covariates for the sampled units and the true area means of these covariates. Other estimators require knowledge of the covariates for every unit in the population. Covariates that are only observed for the sample cannot be used for SAE. The use of auxiliary information for SAE is vital because with the small sample sizes often encountered in practice, even the most sophisticated model can be of little use if it does not involve a set of covariates that provide good predictors of the small area quantities of interest.

3. Notation

Consider a population U of size N, divided into M exclusive and exhaustive areas $U_1 \cup ... \cup U_M$ with N_i units in area i, $i = 1,...,M$, $\sum_{i=1}^{M} N_i = N$. We assume the availability of samples for m out of the M areas, $m \leq M$. Let $s = s_1 \cup ... \cup s_m$ define the sample, where s_i of size n_i defines the sample observed for area i, $\sum_{i=1}^{m} n_i = n$. Notice that n_i is random, unless a separate sample of fixed sample size is taken in that area. Let y define the characteristic of interest and denote by y_{ij} the outcome value for unit j belonging to area i, $i = 1,...,M$, $j = 1...N_i$, with sample means $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i$. When information is available for p covariates, we denote by $\mathbf{x}_{ij} = (x_{1ij},...,x_{pij})'$ the covariates associated with unit (i, j) and by $\bar{\mathbf{x}}_i = \sum_{j=1}^{n_i} \mathbf{x}_{ij} / n_i$ the sample means. The corresponding true are means are $\bar{X}_i = \sum_{j=1}^{N_i} \mathbf{x}_{ij} / N_i$. Let θ_i define the quantity of interest. For example, $\theta_i = \bar{Y}_i = \sum_{j=1}^{N_i} y_{ij} / N_i$, the true area mean. The common case of estimating a proportion is a special case where y_{ij} is binary. In other applications θ_i may represent a count or a quantile (or a set of quantiles).

4. Design-based methods

4.1 *Design-based estimators in common use*

For a recent comprehensive review of design-based methods in SAE the reader is referred to Lehtonen and Veijanen (2009). Here we only overview some of the basic ideas. Suppose that the sample is selected by simple random sampling without replacement (SRSWOR) and that the target quantities of interest are the area means, $\bar{Y}_i$. Estimation of a mean contains as special cases the estimation of a proportion, in which case y_{ij} is binary, and the estimation of the distribution $F_i(t) = \sum_{j \in U_i} v_{ij} / N_i$, in which case $v_{ij} = I(y_{ij} \leq t)$. Estimators of the percentiles of the distribution are usually obtained from the estimated distribution.

If no covariates are available, the *direct* design unbiased estimator and its conditional design variance over the *randomization* distribution are given by,

$$\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i \; ; \; V_D(\bar{y}_i | n_i) = (S_i^2 / n_i)[1 - (n_i / N_i)], \tag{4.1}$$

where $S_i^2 = \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_i)^2 / (N_i - 1)$. The term 'direct' is often used to signify an estimator that only uses the data for the relevant area corresponding to the specific time of interest. The variance $V_D(\bar{y}_i | n_i)$ is $O(1/n_i)$ and for small n_i the variance will be large, unless S_i^2 is sufficiently small.

Next suppose that covariates $\mathbf{x}_{ij}$ are also observed with $x_{1ij} \equiv 1$. An estimator in common use that utilizes the covariates is the *synthetic Regression* estimator,

$$\hat{\bar{Y}}_{reg,i}^{syn} = \bar{X}_i' \hat{B} = \frac{1}{N_i} \sum_{j=1}^{N_i} (x_{ij}' \hat{B}), \tag{4.2}$$

where $\hat{B} = [\sum_{i=1}^{m} \sum_{j=1}^{n_i} x_{ij} x_{ij}']^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_i} x_{ij} y_{ij}$ is the ordinary least square (OLS) estimator. In the special case of a single covariate and zero intercepts, $\hat{\bar{Y}}_{reg,i}^{syn}$ is often replaced by the *synthetic Ratio* estimator $\hat{\bar{Y}}_{ratio,i}^{syn} = \bar{X}_i \bar{y} / \bar{x}$, where $\bar{y}, \bar{x}$ are the sample means of y and x. Notice that $\hat{B}$ is approximately design-unbiased for the vector B of regression coefficients computed from all the population values, irrespective of whether a linear relationship between y and x exists in the population. The term "synthetic" refers to the fact that an (approximately) design-unbiased estimator computed from all the areas ($\hat{B}$ in the present case) is used for each area separately, assuming that the areas are 'homogeneous' with respect to the quantity that is estimated (similar vectors of coefficients in the various areas in the present case). Thus, synthetic estimators borrow information from other 'similar areas' and they are therefore *indirect* estimators.

The obvious advantage of the estimator $\hat{\bar{Y}}_{reg,i}^{syn}$ over the simple sample mean or any other direct estimator such as the regression estimator $\hat{\bar{Y}}_{reg,i} = \bar{y}_i + (\bar{X}_i - \bar{x}_i)'\hat{B}_i$, where $\hat{B}_i$ is computed only from the data observed for area i, is that $Var_D(\hat{\bar{Y}}_{reg,i}^{syn}) = O(1/n)$, and $n = \sum_{i=1}^{m} n_i$ is usually very large. The use of the synthetic estimator is motivated (assisted) by a linear regression model of y on x in the population with a common vector of coefficients. However, assuming that $x_{1ij} \equiv 1$, $E_D(\hat{\bar{Y}}_{reg,i}^{syn} - \bar{Y}_i) \cong -\bar{X}_i'(B_i - B)$, where B_i is the OLS computed from all the population values in area i (follows from the fact that the sum of the OLS residuals is zero). Thus, if in fact different regression coefficients B_i operate in different areas, the bias can be large. The bias could be removed by estimating instead a different vector for every area, but then the variance will be $O(1/n_i)$. When the sample is selected with unequal probabilities, the OLS estimator $\hat{B}$ in (2.2) is commonly replaced by the probability weighted estimator $\hat{B}_{pw} = \left[\sum_{i=1}^{m}\sum_{j=1}^{n_i} w_{ij}x_{ij}x_{ij}'\right]^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i} w_{ij}x_{ij}y_{ij}$, where $w_{ij} = 1/\Pr[(i,j)\in s]$ are the (base) sampling weights. Calibration of the sampling weights is considered below.

In order to deal with the possible large bias of the synthetic estimator, it is common to estimate the bias and then subtract it from the synthetic estimator. The resulting generalized regression estimator (GREG) takes the form,

$$\hat{\bar{Y}}_i^{GREG} = \frac{1}{N_i}\sum_{j=1}^{N_i}(x_{ij}'\hat{B}_{pw}) + \frac{1}{N_i}\sum_{j=1}^{n_i} w_{ij}(y_{ij} - x_{ij}'\hat{B}_{pw}) = \hat{\bar{Y}}_{i,H-T} + (\bar{X}_i - \hat{\bar{X}}_{i,H-T})'\hat{B}_{pw}, \tag{2.3}$$

where $(\hat{\bar{Y}}_{i,H-T}, \hat{\bar{X}}_{i,H-T})$ are the Horvitz-Thomson estimators of $(\bar{Y}_i, \bar{X}_i)$. The variance of (2.3) is generally reduced by multiplying the bias correction by $N_i/\hat{N}_i = N_i / \sum_{j=1}^{n_i} w_{ij}$. The GREG is almost design-unbiased and performs well when the covariates have good predictive power, but the variance is back to order $O(1/n_i)$.

A compromise between the possibly large bias of the synthetic estimator and the possibly large variance of the GREG is achieved by use of a linear combination of the two. The resulting combined (composite) estimator is defined as,

$$\hat{\bar{Y}}_i^{COM} = \delta_i \hat{\bar{Y}}_i^{GREG} + (1-\delta_i)\hat{\bar{Y}}_{reg,i}^{syn}, \tag{2.4}$$

where $0 \le \delta_i \le 1$. Ideally, the coefficient δ_i should be determined such that it minimizes the mean square error (MSE) of $\hat{\bar{Y}}_i^{COM}$, but assessing the bias of the synthetic estimator for any given area is formidable. As such, it is common to let δ_i depend on the achieved sample size, n_i, in the area such that the larger n_i, the larger is δ_i. See Drew *et. al* (1982), Särndal and Hidiroglou (1989) and Rao (2003).

2.2 *New developments in design-based small area estimation*

One family of estimators is obtained by calibrating the base sampling weights $w_{ij} = 1/\Pr[(i,j)\in s]$. Suppose that the population can be divided into C calibration groups, $U = U_{(1)} \cup ... \cup U_{(C)}$ with known totals t_{cx} of auxiliary variables x in the groups, such that each area U_i belongs to one of the calibration groups. Let $s = s_{(1)} \cup ... \cup s_{(C)}$ define the corresponding partitioning of the sample. In a special case $C = 1$ and $U_{(1)} = U$. The calibrated estimator of the area mean $\bar{Y}_i$ is defined as,

$$\hat{\bar{Y}}_i^{cal} = \sum_{j=1}^{n_i} w_{ij}^c y_{ij} / N_i \ \text{ s.t. } \sum_{j\in s_{(c)}} w_{ij}^c x_{ij} = t_{cx}. \tag{2.5}$$

The calibration weights $\{w_{ij}^c\}$ are chosen in such a way that they minimize an appropriate distance from the base weights $\{w_{ij}\}$, subject to satisfying the constraints $\sum_{j\in s_{(c)}} w_{ij}^c x_{ij} = t_{cx}$. For example, when using the distance $\chi^2 = \sum_{j\in s_{(c)}} (w_{ij}^c - w_{ij})^2 / w_{ij}$ and assuming $x_{1ij} \equiv 1$, the resulting calibrated estimator has the form,

$$\hat{\bar{Y}}_i^{cal} = \hat{\bar{Y}}_{i,H-T} + (\bar{X}_c - \hat{\bar{X}}_{c,H-T})'\hat{B}_{c,pw}, \tag{2.6}$$

where $\hat{\bar{Y}}_{i,H-T}$ is the H-T estimator in area i, $\bar{X}_c = t_{cx}/N_c$ is the vector of true means of the auxiliary variables in the group $U_{(c)}$, $\hat{\bar{X}}_{c,H-T}$ is the corresponding H-T estimator and $\hat{B}_{c,pw}$ is the probability weighted estimator of the regression coefficients in the group. Comparing with (□□), $\hat{\bar{Y}}_i^{cal}$ is the GREG estimator with the synthetic estimator $\bar{X}_c'\hat{B}_{c,pw}$ computed in the group U_c. When $U_c = U_i$ (the calibration group is the domain), $\hat{\bar{Y}}_i^{cal}$ is the direct GREG in the domain.

The use of calibration weights was first introduced by Deville and Särndal (1992) and it is now in broad use in sample survey practice (not only for SAE). See Kott (2009) for a recent comprehensive review and discussion. The basic idea behind the use of calibrated estimators in SAE is that if y is approximately a linear combination of x in all the domains belonging to U_c, then $\bar{Y}_i \cong \bar{X}_i'B_c$ for domain i in U_c and since $\sum_{j\in s_{(c)}} w_{ij}^c x_{ij} = t_{cx}$, $\hat{\bar{Y}}_i^{cal} = \sum_{j=1}^{n_i} w_{ij}^c y_{ij} / N_i$ will be a good estimator of $\bar{Y}_i$. Indeed, the advantage of the use of (□□) over (□□) is that it is assisted by a model that only assumes common regression coefficients within the groups U_c and not for all the domains, as implicitly assumed by the use of (□□). The estimator (□□) is approximately design-unbiased irrespective of any model, but $Var_D(\hat{\bar{Y}}_i^{cal} | n_i) = O(1/n_i)$, which may still be large. Notice also that in general, $Var(\hat{B}_{c,pw}) < Var(\hat{B}_{pw})$.

Another way of calibration is by use of instrumental variables (Estevao and Särndal, 200□). Denote the vector of instrumental values for unit (i,j) by h_{ij}. The calibration weights are defined as,

$$w_{ij}^{ins} = w_{ij}(1 + g_c'h_{ij});\ g_c' = (t_{cx} - \hat{t}_{cx,H\text{-}T})'\left[\sum_{i,j\in S_{(c)}} w_{ij} h_{ij} x_{ij}'\right]^{-1}, \qquad (□□)$$

where $\hat{t}_{cx,H\text{-}T}$ is the H-T estimator of t_{cx}. Notice that the instrumental values need only be known for the units in $s_{(c)}$. The calibrated estimator of $\bar{Y}_i$ is now $\hat{\bar{Y}}_i^{cal} = \sum_{j=1}^{n_i} w_{ij}^{ins} y_{ij} / N_i$. As easily checked, $\sum_{j=1}^{n_i} w_{ij}^{ins} x_{ij} = t_{cx}$, thus satisfying the same calibration constraints on the auxiliary variables as before. The instruments may include some or all of the variables in x. Estevao and Särndal (200□) study optimal choice of the instruments, which requires knowledge of the joint selection probabilities $\Pr(i,j \in S)$.

The synthetic estimator (□2), the GREG (□□) and the various calibration-based estimators considered above are all assisted by models that assume a linear relationship between y and x. These estimators only require knowledge of the covariates for the sampled units and the area (group) totals of these covariates. Lehtonen *et al.* (200□, 2005) consider the use of generalized linear models (GLM) and even generalized linear mixed models (GLMM) as the assisting models. Suppose that $E_M(y_{ij}) \cong f(x_{ij};\psi)$ for some nonlinear function $f(\cdot)$ with an unknown vector parameter ψ. A simple important example is where $f(x_{ij};\psi)$ is the logistic function. Estimating ψ by the pseudo-likelihood approach, which consists of estimating the likelihood equations by the corresponding H-T estimator, (or alternatively weighting each sample observation by its base sampling weight), and then maximizing the resulting estimated equations yields the estimator $\hat{\psi}_{pl}$ and the predicted values $\{\hat{y}_{ij} = f(x_{ij};\hat{\psi}_{pl})\}$. The synthetic and GREG estimators are now defined as,

$$\hat{\bar{Y}}_{GLM,i}^{syn} = \frac{1}{N_i}\sum_{j=1}^{N_i} f(x_{ij};\tilde{\psi}_{pl});\ \hat{\bar{Y}}_{GLM,i}^{GREG} = \hat{\bar{Y}}_{GLM,i}^{syn} + \frac{1}{N_i}\sum_{j=1}^{n_i} w_{ij}[y_{ij} - f(x_{ij};\hat{\psi}_{pl})]. \qquad (□8)$$

Notice that the use of the estimators in (□8) requires that the covariates are known for every unit $(i,j) \in U$.

A further extension consists of adding random area effects to the assisting model, that is, assuming $E_M(y_{ij} | x_{ij}, u_i) \cong f(x_{ij}, u_i; \delta)$, with $E_M(u_i) = 0$, $Var_M(u_i) = \sigma_u^2$. Estimation of the model parameters δ, σ_u^2 is now under the model, ignoring the sampling weights. The synthetic and GREG estimators are defined similarly to (□8) but with $f(x_{ij};\hat{\psi}_{pl})$ replaced by $f(x_{ij}, \hat{u}_i; \hat{\delta}, \hat{\sigma}_u^2)$. For a sufficiently large sample size within the area the extended GREG is approximately design unbiased for the true area mean but it is not clear how to estimate the design (randomization) variance in this case in a way that accounts for the estimation of the random effects (see below). Torabi and Rao (2008) compare the MSE of model-based predictors and a GREG assisted by a GLMM.

Jiang and Lahiri (200□a) propose the use of model-dependent estimators that are design consistent (under the randomization distribution) as the area sample sizes increase. The basic idea is to model the direct estimators

$\bar{Y}_{iw} = \sum_{j=1}^{n_i} w_{ij} y_{ij} / \sum_{j=1}^{n_i} w_{ij}$ instead of the individual observations y_{ij}, and then obtain the empirical best predictor of the true mean in each area under the model. The authors consider the general two-level model $E(\bar{Y}_{iw} | u_i) = \xi_i = \xi(u_i, \bar{X}_{iw}; \psi)$, where the u_i s are independent random effects with mean zero and variance σ_u^2 and $\xi(\cdot)$ is some known function governed by the unknown vector parameter ψ. The empirical best predictor is the best predictor under the model but with the unknown parameters replaced by model consistent estimators, $\hat{\bar{Y}}_i^{EBP} = E_M(\xi_i | \bar{Y}_{iw}; \hat{\psi})$. The estimator is shown to be design consistent for large n_i even if the model is misspecified thus robustifying the estimation, and also consistent under the model. Note, however, that the problem of SAE is the small sample sizes in some or all of the areas so that design consistency is not a strong property in this case. The authors develop estimators of the mean squared prediction error (MSPE) for bounded sample sizes with bias of desired order $o(1/m)$. The MSPE is taken with respect to the model holding for the individual observations y_{ij} and the randomization distribution.

[illegible] *Pros and cons of design-based estimation*

The apparent advantage of design-based methods is that the estimation is less dependent on an assumed model, although models are used (assisted) for the construction of the estimators. Approximate design unbiasedness and consistency for large sample sizes within the areas are desirable properties that add some protection against possible model failures, but as noted above in practical applications the sample sizes are often very small, such that these properties should be judged with caution.

Against these advantages stand many disadvantages. Direct estimators generally have large variance due to the small sample sizes. Indirect estimators such as the synthetic estimators are generally biased, with very limited possibilities to assess the bias. Composite estimators are likewise biased and it is not obvious how to best choose the weights attached to the synthetic estimator and the GREG (or other direct estimators). Computation of randomization-based confidence intervals generally requires large sample normality assumptions, but the sample sizes in at least some of the small areas may be very small.

Another limitation of design-based inference (not restricted to SAE) is that it does not lend itself to conditional inference, for example, conditioning on the sampled values of the covariates or the sampled clusters in a two-stage sampling design. This again inflates the variance of the estimators. Conditional inference is in the heart of classical statistical inference under both the frequentist and the Bayesian approaches. Last, but in many ways the most important limitation is the fact that there is no founded design-based theory for estimation in areas with no samples. The use of the randomization distribution does not lend itself to prediction problems such as the prediction of the dependent variable for given covariates under a regression model, or the prediction of small area means for areas with no samples. Design-based theory is restricted to estimation of population quantities from a sample taken from the population. As stated in the introduction, it is often the case that samples are available for only a minority of the areas but estimators and MSE estimators are required for each of the areas, whether sampled or not.

5. Model-based methods

5.1 *General formulation*

Model-based methods assume a model for the sample data and use the optimal or approximately optimal predictor of the area characteristic of interest under the model. The MSE of the prediction error is likewise defined and estimated under the model. Note that we now use the term "prediction" rather than estimation because the target characteristics are generally random under the model. The use of models overcomes the problems underlying the use of design-based methods discussed in Section [illegible], but it is important to emphasize again that even the most sophisticated model cannot produce predictors of acceptable precision when the area sample sizes are too small and no covariates with good predictive power are available. The use of models raises also the question of the robustness of the inference to possible model misspecification, and we later discuss various attempts to deal with this problem.

As before, we denote by θ_i the quantity of interest in area i (mean, proportion, quantile,…). Let y_i define the observed response data for area i (when the area is sampled) and x_i define the corresponding values of the covariates (when available). As becomes evident below, y_i is either a scalar, in which case x_i is usually a vector, or a vector, in which case x_i is usually a matrix. A typical small area model consists of two parts: The first part models the distribution (moments) of $y_i | \theta_i; \psi_{(1)}$. The second part models the distribution (moments) of $\theta_i | x_i; \psi_{(2)}$, linking the θ_i's to the known covariates and to each other. This is achieved by including in the model random effects that account for the variability of the θ_i's not explained by the covariates. The hyper-parameters $\psi = (\psi_{(1)}, \psi_{(2)})$ are usually unknown and are estimated either under the frequentist approach or under the full Bayesian approach after setting appropriate prior distributions. In some applications the index i may define time, in which case the model for $\theta_i | x_i; \psi_2$ is a time series model.

5.2 *Models in common use*

In this section we review briefly three models in common use, as most of the recent developments in SAE relate to these models or extensions of them. For more details and many references, see Pfeffermann (2002), Rao (2003) and Datta (2009).

5.2.1 Area level model

This model is in broad use when the covariate information is only at the area level, so that x_i is a vector of known area characteristics. The model, used originally by Fay and Herriot (1979) is defined as,

$$\tilde{y}_i = \theta_i + e_i; \quad \theta_i = x_i'\beta + u_i, \tag{5.1}$$

where $\tilde{y}_i$ denotes the direct sample estimator of θ_i (for example, the sample mean $\bar{y}_i$). The residual e_i represents the sampling error, assumed to have zero mean and known design (randomization) variance $Var_D(e_i) = \sigma^2_{Di}$. The random effects u_i are assumed to be independent with zero mean and common variance σ_u^2. For known σ_u^2, the best linear unbiased predictor (BLUP) of θ_i under this model is,

$$\hat{\theta}_i = \gamma_i \tilde{y}_i + (1-\gamma_i) x_i'\hat{\beta}_{GLS} = x_i'\hat{\beta}_{GLS} + \gamma_i(\tilde{y}_i - x_i'\hat{\beta}_{GLS}), \tag{5.2}$$

where $\hat{\beta}_{GLS}$ is the generalized least square (GLS) estimator of β under the model. The BLUP $\hat{\theta}_i$ is in the form of a composite estimate ([illegible]), but with a tuning coefficient $\gamma_i = \sigma_u^2 / (\sigma_u^2 + \sigma^2_{Di})$ that depends optimally on the ratio $\sigma_u^2 / \sigma^2_{Di}$ of the variances of the prediction errors of $x_i'\beta$ and $\tilde{y}_i$. Under normality of u_i, e_i and a uniform prior for β, $\hat{\theta}_i$ is also the Bayesian predictor (posterior mean) of θ_i. For a nonsampled area k (but known x_k), the BLUP is now well defined to be $x_k'\hat{\beta}_{GLS}$. In practice, the variance σ_u^2 is seldom known and it is replaced in γ_i and $\hat{\beta}_{GLS}$ by a sample estimate, yielding what is known as the empirical BLUP (EBLUP) under the frequentist approach, or the empirical Bayes (EB) predictor. Alternatively, one may assume a prior distribution for σ_u^2 and compute the posterior distribution of θ_i, which may be used for the point predictor and for the construction of credibility intervals.

Remark 1. The synthetic estimator $x_i'\hat{\beta}_{GLS}$ in (5.2), and hence also the BLUP $\hat{\theta}_i$ are biased when conditioning on u_i, similarly to what we had under the randomization distribution. Conditioning on u_i amounts to assuming different fixed intercepts in different areas, and the unbiasedness of $\hat{\theta}_i$ in (5.2) under the model is achieved by viewing the intercepts as random. The same applies for the BLUP (5.4) below.

5.2.2 Nested error unit level model

This model uses the individual observations y_{ij} such that y_i is now a vector and x_i is a matrix. As for design-based methods, the use of this model for SAE requires that the area means, $\bar{X}_i = \sum_{j=1}^{N_i} x_{ij} / N_i$, are known. The model, first proposed for SAE by Battese *et al.* (1988) has the form,

$$y_{ij} = x_{ij}'\beta + u_i + \varepsilon_{ij}, \tag{5.3}$$

where u_i (the random effect) and ε_{ij} (residual terms) are mutually independent with zero means and variances σ_u^2 and σ_ε^2 respectively. Under the model the true small area means are $\bar{Y}_i = \bar{X}_i'\beta + u_i + \bar{\varepsilon}_i$, but since $\bar{\varepsilon}_i = \sum_{j=1}^{N_i} \varepsilon_{ij} / N_i \cong 0$ for large N_i, the target means are often defined as $\theta_i = \bar{X}_i'\beta + u_i = E(\bar{Y}_i | u_i)$. For known variances $(\sigma_u^2, \sigma_\varepsilon^2)$, the BLUP of θ_i is,

$$\hat{\theta}_i = \gamma_i [\bar{y}_i + (\bar{X}_i - \bar{x}_i)'\hat{\beta}_{GLS}] + (1-\gamma_i)\bar{X}_i'\hat{\beta}_{GLS}, \tag{5.4}$$

where $\hat{\beta}_{GLS}$ is the GLS of β computed from all the observed data, $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$ and $\gamma_i = \sigma_u^2/(\sigma_u^2+\sigma_\varepsilon^2/n_i)$. For area k with no sample (but known $\bar{X}_k$), the BLUP is $\hat{\theta}_k = \bar{X}_k'\hat{\beta}_{GLS}$. See Rao (2003) for the BLUP of the actual mean $\bar{Y}_i$ in sampled areas.

As with the area level model, the BLUP (5.4) is also the Bayesian predictor (posterior mean) under normality of the error terms and a uniform prior for β. Replacing the unknown variances σ_u^2 and σ_ε^2 in γ_i and $\hat{\beta}_{GLS}$ by sample estimates yields the corresponding EBLUP or EB predictors. Common methods of estimating the variances under the frequentist approach are Maximum Likelihood (MLE), Restricted MLE (REML) or Analysis of Variance (ANOVA) type estimators. See Rao (2003) for details. Hierarchical Bayes (HB) predictors are obtained by specifying prior distributions for β and the two variances and computing the posterior distribution of θ_i (or $\bar{Y}_i$) given all the observations in all the areas.

5.2. Mixed logistic model

The previous two models assume continuous response values. Suppose now that y_{ij} is binary taking the values 1 and 0, in which case the small area quantities of interest are usually proportions or counts (say, the proportion or total of unemployed persons in the area). The following generalized linear mixed (GLMM) considered originally by MacGibbon and Tomberlin (1989) for SAE is in broad use for this kind of problem.

$$\Pr(y_{ij}=1|p_{ij})=p_{ij};\ logit(p_{ij})=x_{ij}'\beta+u_i;\ u_i\sim N(0,\sigma_u^2). \qquad (5.5)$$

The outcomes y_{ij} are assumed to be conditionally independent given the random effects, u_i, and likewise for the random effects. The purpose is to predict the true area proportions $p_i=\sum_{j=1}^{N_i} y_{ij}/N_i$. Let $\psi=(\beta,\sigma_u^2)$ denote the model parameters. For this model there is no explicit expression for the best predictor (BP) $E(p_i|y_i,x_i;\psi)$, but as shown in Jiang and Lahiri (2006a), for known ψ the BP can be computed (approximated) numerically as the ratio of two one-dimensional integrals. The authors review methods of estimating ψ, yielding the empirical BP (EBP) $\hat{p}_i^{EBP}=E(p_i|y_i,x_i;\hat{\psi})$. Alternatively, the predictors $\hat{p}_i$ can be obtained by application of the EB approach (setting $\psi=\hat{\psi}$), as in MacGibbon and Tomberlin (1989), or by application of the full HB approach, as in Malec *et al.* (1997). Let us describe briefly the application of the HB approach. It consists of the following steps:

1. Specify prior distributions for σ_u^2 and β;

2. Compute the posterior distributions of β and $\{u_i, i=1,...,m\}$ by, say, MCMC simulations and draw a large number of realizations $(\beta^{(r)},\{u_i^{(r)}\})$, and hence realizations, $y_{ik}^{(r)} \sim p_{ik}^{(r)} = \frac{\exp(x_{ik}'\beta^{(r)}+u_i^{(r)})}{1+\exp(x_{ik}'\beta^{(r)}+u_i^{(r)})}$, $r=1,...,R,\ i=1,...,m,\ k\notin s_i$.

3. Predict: $\hat{p}_i=(\sum_{j\in s_i} y_{ij}+\sum_{k\notin s_i}\hat{y}_{ik})/N_i;\ \hat{y}_{ik}=\sum_{r=1}^R y_{ik}^{(r)}/R,\ k\notin s_i$.

In order to compute the posterior variance write $\hat{p}_i=\frac{1}{N_i}\frac{1}{R}\sum_{r=1}^R(\sum_{j\in s_i} y_{ij}+\sum_{k\notin s_i}\hat{y}_{ik}^{(r)}) = \frac{1}{N_i}\frac{1}{R}\sum_{r=1}^R \hat{p}_i^{(r)}$.

Compute, $\hat{V}_{post}(\hat{p}_i)=\frac{1}{N_i^2}\frac{1}{R(R-1)}\sum_{r=1}^R(\hat{p}_i^{(r)}-\hat{p}_i)^2$.

6. New developments in model-based SAE

6.1 *Estimation of prediction MSE*

As stated in the introduction, an important aspect of SAE is the assessment of the prediction errors. This problem is solved "automatically" under the full Bayesian paradigm, which produces the posterior distribution of the target quantities around their posterior means. However, estimation of the prediction MSE (PMSE) under the EBLUP and EB approaches is complicated because of the added variability induced by the estimation of the model hyper-parameters. Prasad and Rao (1990) develop PMSE estimators with bias of order $o(1/m)$, (m is the number of sampled areas), for the linear mixed models 5.1.1 and 5.1.2 for the case where the random errors have a normal distribution and the model variances are estimated by the ANOVA type method of moments estimators. Lahiri and Rao (1995) show that the PMSE estimator in the case of the model 5.1.1 is robust to departures from normality of the random area effects u_i (but not the sampling errors e_i). Datta and Lahiri (2000) extend the estimation of Prasad and Rao to general linear mixed models of the form,

$$y_i = X_i\beta + Z_i u_i + e_i = \xi_i + e_i,\ \ i = 1...m, \qquad (□1)$$

where Z_i is a fixed matrix of order $n_i \times d$ and u_i and e_i are independent vector random effects and residual terms of orders d and n_i respectively. It is assumed that $u_i \sim N_d(0,Q_i)$, $e_i \sim N_{n_i}(0,R_i)$. The authors develop MSE estimators with bias of order $o(1/m)$ for the EBLUP obtained when estimating Q_i and R_i by MLE or REML. More recently, Datta *et al*. (2005) showed that for the area level model 5.1.1, if the variance is estimated by the method proposed by Fay and Herriot (1989) and Pfeffermann and Nathan (1981), then it is required to add an extra term to the PMSE estimator to achieve the desired order of bias of $o(1/m)$. See Datta (2009) for further discussion of the estimation of the PMSE of the EBLUP and EB under linear models.

Estimation of the PMSE under GLMM is more involved and below we review briefly resampling procedures that can be used in such cases. For convenience we consider the mixed logistic model (5.5) but the procedures are applicable to other models belonging to this family.

The first procedure, proposed by Jiang *et. al* (2002) uses the jackknife method. Let $\lambda_i = E(\hat{p}_i^{EBP} - p_i)^2$ denote the PMSE where $p_i = \sum_{j=1}^{N_i} y_{ij}/N_i$ is the true proportion and $\hat{p}_i^{EBP} = E(p_i|y_i, x_i; \hat{\psi})$ is the empirical best predictor. The following decomposition always holds,

$$\lambda_i = E(\hat{p}_i^{(BP)} - p_i)^2 + E(\hat{p}_i^{(EBP)} - \hat{p}_i^{(BP)})^2 = M_{1i} + M_{2i}, \qquad (□2)$$

where M_{1i} is the PMSE of the BP (assumes known parameter values) and M_{2i} is the contribution to the PMSE from having to estimate the model parameters. Denote by $\hat{\lambda}_i^{BP}(\hat{\psi})$ the 'naive' estimator of M_{1i}, obtained by setting $\hat{\psi} = \psi$. Let $\hat{\lambda}_i^{BP}(\hat{\psi}_{-l})$ denote the naive estimator when estimating ψ from all the areas except for area l, and $\hat{p}_i^{EBP}(\hat{\psi}_{-l})$ denote the corresponding EBP. The jackknife estimator is:

$$\lambda_i^{JK} = \hat{M}_{1i} + \hat{M}_{2i};\quad \begin{aligned} \hat{M}_{1i} &= \hat{\lambda}_i^{BP}(\hat{\psi}) - \frac{m-1}{m}\sum_{l=1}^{m}[\hat{\lambda}_i^{BP}(\hat{\psi}_{-l}) - \hat{\lambda}_i^{BP}(\hat{\psi})] \\ \hat{M}_{2i} &= \frac{m-1}{m}\sum_{l=1}^{m}[\hat{p}_i^{EBP}(\hat{\psi}_{-l}) - \hat{p}_i^{EBP}(\hat{\psi})]^2 \end{aligned}. \qquad (□□)$$

Under some regularity conditions, $E(\hat{\lambda}_i^{JK}) - \lambda_i = o(1/m)$ as desired.

The jackknife method estimates the unconditional PMSE over the joint distribution of the random effects and the responses. Lohr and Rao (2009) proposed a modification of the jackknife estimator that is computationally simpler and estimates the conditional PMSE, $E[(\hat{p}_i^{(EBP)} - p_i)^2 | y_i]$, which for nonlinear models is different from the unconditional PMSE. The conditional PMSE is close conceptually to the posterior variance. Denoting $q_i(\psi, y_i) = Var(\theta_i | y_i; \psi)$, the modification consists of replacing $\hat{M}_{1i}$ by $\hat{M}_{1i,c} = q_i(\hat{\psi}, y_i) - \sum_{l \neq i}^{m}[q_i(\hat{\psi}_{-l}, y_i) - q_i(\hat{\psi}, y_i)]$. The modified jackknife estimator, $\hat{\lambda}_{i,c}^{JK} = \hat{M}_{1i,c} + \hat{M}_{2i}$ is shown to have bias of order $o_p(1/m)$ in estimating the conditional PMSE and a bias of order $o(1/m)$ in estimating the unconditional PMSE.

Hall and Maiti (200□) propose a third procedure of estimating the PMSE based on double-bootstrap. For the model (5.5) the procedure consists of the following steps:

1. Generate a new population from the model (5.5) with parameters $\hat{\psi}$ and compute the 'true' area proportions for this population. Compute the EBPs based on the corresponding new sample data and newly estimated parameters. The new population and sample use the same covariates as in the original population and sample. Repeat the same process independently B_1 times, with B_1 sufficiently large. Denote by $p_{i,b_1}(\hat{\psi})$ and $\hat{p}_{i,b_1}^{(EBP)}(\hat{\psi}_{b_1})$ the true proportions and corresponding EBPs for population and sample b_1, $b_1 = 1,...,B_1$. Compute the first-step bootstrap MSE estimator,

$$\hat\lambda_{i,1}^{BS} = \frac{1}{B_1}\sum_{b_1=1}^{B_1}[\hat p_{i,b_1}^{(EBP)}(\hat\psi_{b_1}) - p_{i,b_1}(\hat\psi)]^2 . \tag{6.4}$$

2. For each sample drawn in Step 1, repeat the computations of Step 1 B_2 times with B_2 sufficiently large, yielding new 'true proportions' $p_{i,b_2}(\hat\psi_{b_1})$ and EBPs $\hat p_{i,b_2}^{(EBP)}(\hat\psi_{b_2})$, $b_2 = 1,...,B_2$. Compute the second-step bootstrap MSE estimator,

$$\hat\lambda_{i,2}^{BS} = \frac{1}{B_1}\sum_{b_1}^{B_1}\frac{1}{B_2}\sum_{b_2=1}^{B_2}[\hat p_{i,b_2}^{(EBP)}(\hat\psi_{b_2}) - p_{i,b_2}(\hat\psi_{b_1})]^2 . \tag{6.5}$$

The double-bootstrap MSE estimator is obtained by computation of one of the classical bias corrected estimators. For example,

$$\hat\lambda_i^{D-BS} = \begin{cases} \hat\lambda_{i,1}^{BS} + (\hat\lambda_{i,1}^{BS} - \hat\lambda_{i,2}^{BS}), & \text{if } \hat\lambda_{i,1}^{BS} \ge \hat\lambda_{i,2}^{BS} \\ \hat\lambda_{i,1}^{BS}\exp[(\hat\lambda_{i,1}^{BS} - \hat\lambda_{i,2}^{BS})/\hat\lambda_{i,2}^{BS}], & \text{if } \hat\lambda_{i,1}^{BS} < \hat\lambda_{i,2}^{BS} \end{cases} . \tag{6.6}$$

The estimator has bias of order $o(1/m)$ under some regularity conditions.

Remark 2. All the above procedures although being resampling methods are actually parametric methods that rely on the underlying model.

6.2 *Choice of priors in Bayesian applications*

For the area level model (5.1), Ganesh & Lahiri (2008) develop a class of priors for σ_u^2 satisfying for a given set of weights $\{\omega_i\}$,

$$\sum_{i=1}^{m}\omega_i E[Var(\theta_i|y) - \text{PMSE}[\tilde\theta_i(\hat\sigma_u^2)]] = o(1/m), \tag{6.7}$$

where $Var(\theta_i|y)$ is the posterior variance under the desired prior $p(\sigma_u^2)$, $\tilde\theta_i(\hat\sigma_u^2)$ is the EBLUP of θ_i obtained by substituting $\hat\sigma_u^2$ for σ_u^2 in (5.2) and the expectation and PMSE are computed under the joint distribution of θ and y. The prior $p(\sigma_u^2)$ satisfying (6.7) is shown to be,

$$P(\sigma_u^2) \propto \sum_{i=1}^{m}1/(\sigma_{Di}^2+\sigma_u^2)^2 \Big/ \sum_{i=1}^{m}\omega_i[\sigma_{Di}^2/(\sigma_{Di}^2+\sigma_u^2)]^2 . \tag{6.8}$$

The motivation for using the prior (6.8) is to warrant some 'frequency validity' to the Bayesian measure of uncertainty by guaranteeing that the weighted average of the expected difference between the posterior variance and the PMSE of the EBLUP is sufficiently close. Having satisfied (6.7), the analyst may then take advantage of all the flexibility of Bayesian inference resulting from the ability to draw observations from the posterior distribution. By appropriate choice of the weights $\{\omega_i\}$, the prior (6.8) contains as special cases the flat prior $p(\sigma_u^2) = U(0,\infty)$, the prior developed by Datta *et al.* (2005) for a given area satisfying $E[Var(\theta_i|y)]$ $= PMSE[\tilde\theta_i(\hat\sigma_u^2)] + o(1/m)$, (different prior for different areas), and the average moment matching prior (obtained by setting $\omega_i \equiv 1$).

6.3 *Benchmarking*

Model-based SAE depends on models that could be hard to validate. When the model is misspecified, the resulting predictors may perform poorly. Benchmarking is another way of trying to robustify the inference by forcing the model-based predictors to agree with the design-based estimator for an aggregation of the areas where it can be trusted. Assuming for convenience that this aggregation contains all the areas, the benchmarking equation takes the general form,

$$\sum_{i=1}^{m} b_i\hat\theta_{i,model} = \sum_{i=1}^{m} b_i\hat\theta_{i,design} , \tag{6.9}$$

where the coefficients $\{b_i\}$ are fixed weights, assumed without loss of generality to sum to 1 (e.g., relative area sizes). The modification (6.9) has the further advantage of guaranteeing consistency of publication between the small area predictors and the design-based estimator for the aggregated area, which is often required by statistical agencies. For example, the model-based predictors of total unemployment in, say, counties should add to the design-based estimate of total unemployment in the country, which is deemed accurate.

A benchmarking procedure in common use, often referred to as ratio or pro-rata adjustment is defined as,

$$\hat\theta_i^{bench} = \Big(\sum_{i=1}^{m} b_i\hat\theta_{i,design} \Big/ \sum_{i=1}^{m} b_i\hat\theta_{i,model}\Big)\times\hat\theta_{i,model} . \tag{6.10}$$

The use of this procedure, however, applies the same percentage correction for all the areas, irrespective of the relative precision of the original model-based predictors (before benchmarking). Also, estimation of the PMSE of the prorated predictors is not straightforward. As a result, other procedures have been proposed in the literature.

Wang *et al.* (2008) derive a benchmarked BLUP (BBLUP) under the area level model (5.1) by minimizing $\sum_{i=1}^{m}\varphi_i E(\theta_i - \hat{\theta}_i^{bench})^2$ subject to (□9), where the φ_i s are a chosen set of positive weights. The BBLUP predictor is,

$$\hat{\theta}_{i,BLUP}^{bench} = \hat{\theta}_{i,model}^{BLUP} + \delta_i \sum_{j=1}^{m} b_j (\hat{\theta}_{j,design} - \hat{\theta}_{j,model}^{BLUP}) ;\ \delta_i = (\sum_{j=1}^{m} \varphi_j^{-1} b_j^2)^{-1} \varphi_i^{-1} b_i) . \qquad (□11)$$

When the variance σ_u^2 is unknown, it is replaced by its estimator everywhere in (□11), yielding the empirical BBLUP. The PMSE of the latter predictor can be estimated by a method developed in by Isaki *et al.* (2000), or by one of the resampling procedures described in Section □1.

You & Rao (2002) achieve "automatic benchmarking" for the unit level model (5.□) by changing the estimator of β. Wang *et al.* (2008) consider a similar procedure for the area level model. The approach is further extended by augmenting the covariates x_i to $\tilde{x}_i = [x_i', b_i\sigma_{Di}^2]'$. (The variances σ_{Di}^2 are considered known under the area level model.) The use of the augmented model yields a BLUP that satisfies the benchmark constraint (□9) and is more robust to model misspecification.

Pfeffermann & Tiller (200□) added monthly benchmark constraints of the form (□9) to the measurement (observation) equation of a time series state-space model fitted jointly to the direct estimates in several areas. Adding benchmark constraints to time series models used for the production of model-dependent small area predictors is particularly important since time series models are slow to adapt to abrupt changes. By incorporating the constraints in the model, the use of this approach permits estimating the variance of the benchmarked estimators as part of the model fitting. The variance accounts for the variances of the model error terms and the variances and covariances of the sampling errors of the monthly direct estimators and their mean $\sum_{i=1}^{m} b_i \hat{\theta}_{i,direct}$.

Datta *et al.* (2010) develop Bayesian benchmarking by minimizing,

$$\sum_{i=1}^{m}\varphi_i E[(\theta_i - \hat{\theta}_i^{bench})^2 \,|\, \hat{\theta}_{design}] \ \text{s.t.}\ \sum_{i=1}^{m} b_i \hat{\theta}_i^{bench} = \sum_{i=1}^{m} b_i \hat{\theta}_{i,design} , \qquad (□12)$$

where $\hat{\theta}_{design} = (\hat{\theta}_{1,design},...,\hat{\theta}_{m,design})'$. The solution of this minimization problem is the same as (5.1□), but with $\hat{\theta}_{i,model}^{BLUP}$ replaced everywhere by the posterior mean $\hat{\theta}_{i,Bayes}$. Denote the resulting predictors by $\hat{\theta}_{i,Bayes}^{bench,1}$. The use of these predictors has the drawback of "over shrinkage" in the sense that $\sum_{i=1}^{m} b_i (\hat{\theta}_{i,Bayes}^{bench,1} - \bar{\hat{\theta}}_{b,Bayes}^{bench,1})^2 < \sum_{i=1}^{m} b_i E[(\theta_i - \bar{\theta}_b)^2 \,|\, \hat{\theta}_{design}]$, where $\bar{\hat{\theta}}_{b,Bayes}^{bench,1} = \sum_{i=1}^{m} b_i \hat{\theta}_{i,Bayes}^{bench,1}$ and $\bar{\theta}_b = \sum_{i=1}^{m} b_i \theta_i$. To deal with this problem, Datta *et al.* (2010) propose to consider instead the predictors $\hat{\theta}_{i,Bayes}^{bench,2}$ satisfying,

$$\sum_{i=1}^{m} b_i \hat{\theta}_{i,Bayes}^{bench,2} = \sum_{i=1}^{m} b_i \hat{\theta}_{i,design} ;\ \sum_{i=1}^{m} b_i (\hat{\theta}_{i,Bayes}^{bench,2} - \sum_{i=1}^{m} b_i \hat{\theta}_{i,design})^2 = H , \qquad (□1□)$$

where $H = \sum_{i=1}^{m} b_i E(\theta_i - \bar{\theta}_b)^2 \,|\, \hat{\theta}_{design}]$. The desired predictors have now the form,

$$\hat{\theta}_{i,Bayes}^{bench,2} = \sum_{i=1}^{m} b_i \hat{\theta}_{i,design} + A_{CB}(\hat{\theta}_{i,Bayes} - \bar{\hat{\theta}}_{Bayes}) ;\ A_{CB}^2 = H / \sum_{i=1}^{m} b_i (\hat{\theta}_{i,Bayes} - \bar{\hat{\theta}}_{Bayes})^2 . \qquad (□1□)$$

Notice that the development of the Bayesian benchmarked predictors by Datta *et al.* (2010) is general and not restricted to any particular model.

□.□ Accounting for errors in the covariates

Ybarra and Lohr (2008) consider the case where some or all of the true covariates x_i in the area level model (5.1) are unknown and one uses instead an estimator $\hat{x}_i$ with $MSE(\hat{x}_i) = C_i$ in the expression for the BLUP $\hat{\theta}_i$ defined by (5.2). Estimates for the missing covariates may be obtained from another survey. Denoting the resulting predictor by $\hat{\theta}_i^{Err}$, it follows that

$$MSE(\hat{\theta}_i^{Err}) = MSE(\hat{\theta}_i) + (1-\gamma_i)^2 \beta' C_i \beta . \qquad (□15)$$

Thus, reporting $MSE(\hat{\theta}_i)$ as the PMSE as if the covariates are measured correctly results in under-reporting of the true PMSE. Moreover, if $\beta' C_i \beta > \sigma_u^2 + \sigma_{Di}^2$, then $MSE(\hat{\theta}_i^{Err}) > \sigma_{Di}^2$, the variance of the direct estimator $\tilde{y}_i$. The authors propose using instead the predictor,

$$\hat{\theta}_i^{Me} = \tilde{\gamma}_i \tilde{y}_i + (1-\tilde{\gamma}_i)\hat{x}_i'\beta \ ; \ \tilde{\gamma}_i = (\sigma_u^2 + \beta' C_i \beta)/(\sigma_{Di}^2 + \sigma_u^2 + \beta' C_i \beta) . \qquad (□1□)$$

The predictor $\hat{\theta}_i^{Me}$ minimizes the MSE of linear combinations of $\tilde{y}_i$ and $\hat{x}_i'\beta$. $E(\hat{\theta}_i^{Me} - \theta_i) = (1-\tilde{\gamma}_i)[E(\hat{x}_i) - x_i]'\beta$, such that the bias vanishes when $\hat{x}_i$ is unbiased for x_i and $E(\hat{\theta}_i^{Me} - \theta_i)^2 = \tilde{\gamma}_i \sigma_{Di}^2 \le \sigma_{Di}^2$. The authors develop estimators for the unknown σ_u^2 and β, which are then substituted in (□1□) to yield the corresponding empirical predictor. The PMSE of the empirical predictor is estimated using the jackknife procedure of Jiang *et al*. (2002) described in Section □1.

Torabi *et al*. (2009) consider the unit level model (5.□) with $x_{ij}'\beta = \beta_0 + \beta_1 x_i$ (a single covariate common to all the units in the same area), and assume that instead of measuring x_i, one measures instead $x_{ij} = x_i + \eta_{ij}$, where $x_i \sim N(\mu_x, \sigma_x^2)$. It is assumed also that $(u_i, \varepsilon_{ij}, \eta_{ij})$ are mutually independent normally distributed random errors with zero means and variances $\sigma_u^2, \sigma_\varepsilon^2$ and σ_η^2 respectively. The sample observations consist of $\{y_{ij}, x_{ij};\ i = 1,...,m,\ j = 1,...,n_i\}$. An example giving rise to such a model is where x_i indicates the true level of air pollution in the area and the x_{ij} s represent measures of the pollution at different sites located in the area, with the response values y_{ij} measuring a certain health indicator at the different sites .

The authors show that when all the model parameters are known, the posterior distribution of the unobserved y-values in any given are i is multivariate normal, which yields to following predictor for $\bar{Y}_i$:

$$\hat{\bar{Y}}_{i,Bayes} = (1 - f_i K_i)\bar{y}_i + f_i K_i(\beta_0 + \beta_1 \mu_x) + f_i K_i \gamma_{xi} \beta_1 (\bar{X}_i - \mu_x) , \qquad (□1□)$$

where $f_i = 1-(n_i / N_i)$ is the finite population correction in area i, $\gamma_{xi} = n_i \sigma_x^2 (\sigma_\eta^2 + n_i \sigma_x^2)^{-1}$ and $A_i = [n_i \beta_1^2 \sigma_x^2 \sigma_\eta^2 + (n_i \sigma_u^2 + \sigma_\varepsilon^2) v_i]^{-1} \sigma_\varepsilon^2 v_i$ with $v_i = (\sigma_\eta^2 + n_i \sigma_x^2)$. For large N_i and small (n_i / N_i), the PMSE of $\hat{\bar{Y}}_{i,Bayes}$ is $E[(\hat{\bar{Y}}_{i,Bayes} - \bar{Y}_i)^2 | y_i, x_{ij}] = K_i[\beta_1^2 \sigma_x^2 + \sigma_u^2 - n_i \beta_1^2 \sigma_x^4 v_i^{-1}]$.
The unknown model parameters $\psi = (\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)$ are estimated by a method of moments. Replacing the unknown model parameters by their estimates yields the corresponding EB estimate, which is shown to be asymptotically optimal as the number of sampled areas increases. The PMSE of the EB estimator is estimated by use of a weighted jackknife procedure proposed by Chen and Lahiri (2002).

Remark □. In an earlier study, Ghosh *et al*. (200□) used the same model and inference procedure as Torabi *et al*. (2009), but they used only the observed y-values (and not the observed covariates) for estimating the area means under the model.

□5 *Treatment of outliers*

Bell and Huang (200□) consider the area level model (5.1) from a Bayesian perspective, assuming that the random effect or the sampling error have a Student's $t_{(k)}$ distribution. The t distribution is often used in statistical modeling to account for outliers because of its long tails. One of the models considered is,

$$e_i \sim N(0, \sigma_{Di}^2) ,\ u_i | \delta_i, \sigma_u^2 \sim N(0, \delta_i \sigma_u^2) ; \delta_i^{-1} \sim Gamma[k/2, (k-2)/2] , \qquad (□18)$$

which implies $u_i | \sigma_u^2 \sim t_{(k)}(0, \sigma_u^2 (k-2)/k)$ and $E(\delta_i) = 1, Var(u_i | \sigma_u^2) = \sigma_u^2$. Thus, the coefficient δ_i can be viewed as a multiplicative random effect distributed around 1 that inflates or deflates the variance of $u_i = \theta_i - x_i'\beta$. A large value δ_i signals the existence of an outlying area mean θ_i. The degrees of freedom parameter, k, is assumed known. Setting $k = \infty$ is equivalent to assuming the model (5.1). The authors consider several possible (small) values for k in their application, but the choice of an appropriate value is very much the result of data exploration. Alternatively, the authors assume the model (□18) for the sampling error, e_i, (with σ_u^2 replaced by σ_{Di}^2), in which case $u_i \sim N(0, \sigma_u^2)$. The use of either model is shown empirically to perform well in identifying outlying areas, but the effect of assuming the model for the random effects is to push the small area predictor (the posterior mean) towards the direct estimator, whereas assuming the model for the sampling errors is to push the predictor towards the synthetic part. It is not obvious how to choose between the two models. Huang and Bell (200□) extend the approach to a bivariate area level model where two direct estimates are available for every area with uncorrelated sampling errors but correlated random effects. Such a model applies, for example, when estimates are obtained from two different surveys.

Ghosh *et al*. (2008) likewise consider the model (5.1) from a Bayesian perspective. The starting point in this study is that an outlying direct estimate may arise either from a large sampling error or from an outlying random effect. The authors propose to replace therefore the EB predictor (5.2) by the robust empirical Bayes predictor,

$$\hat{\theta}_i^{Rob} = \tilde{y}_i - (1-\hat{\gamma}_i)\hat{D}_i \Psi_G[(\tilde{y}_i - x_i'\hat{\beta}_{GLS})\hat{D}_i^{-1}];\ \hat{D}_i^2 = \hat{Var}(\tilde{y}_i - x_i'\hat{\beta}_{GLS}), \qquad (□19)$$

where $\hat{\beta}_{GLS}$ is the GLS estimator of β under the model with estimated variance $\hat{\sigma}_u^2$, and Ψ is the Huber influence function $\Psi_G(t) = sign(t)\min(G, |t|)$ for some given $G > 0$. Thus, for large positive standardized residuals $(\tilde{y}_i - x_i'\hat{\beta}_{GLS})\hat{D}_i^{-1}$, the EB $\hat{\theta}_i = \tilde{y}_i - (1-\hat{\gamma}_i)\hat{D}_i(\tilde{y}_i - x_i'\hat{\beta}_{GLS})\hat{D}_i^{-1}$ under the model is replaced by $\hat{\theta}_i^{Rob} = \tilde{y}_i - (1-\hat{\gamma}_i)\hat{D}_i G$, and similarly for large negative standardized residuals, where as in other cases the EB predictor is unchanged. The constant G may actually change from one area to the other and chosen adaptively in such a way that the excess Bayes risk from using the predictor (□19) under the model is bounded by some percentage point. Alternatively, *G* may be set to 1 or 2 as often found in the robustness literature. The authors derive the PMSE of $\hat{\theta}_i^{Rob}$ under the model for the case where σ_u^2 is estimated by MLE with bias of order $o(1/m)$, and develop an estimator for the approximate PMSE, which is correct up to the order of $O(1/m)$.

Under the Ghosh *et al*. (2008) approach the EB small area predictor (5.2) is replaced by the outlier robust predictor (□19), but the estimators of the unknown model parameters remain intact. Sinha and Rao (2009) propose to robustify also the estimation of the model parameters. The authors consider the model considered by Datta and Lahiri (2000) defined by (□1), which when written compactly for all the observations $y = (y_1',...,y_m')'$ has the form,

$$y = X\beta + Zu + e,\ E(u) = 0,\ E(uu') = Q;\ E(e) = 0,\ \ E(ee') = R, \qquad (□20)$$

where $X = [x_1,...,x_m]'$ is the corresponding design matrix of the covariates, u is a vector of random effects and e is the vector of residuals (or sampling errors). The matrices Q and R are block diagonal and depend on some vector parameter $\zeta = (\zeta_1,...,\zeta_L)$ of variance components, such that $V(y) = V = ZQZ' + R = V(\zeta)$. This model contains as special cases the area level model (5.1) and the unit level model (5.□). The target is to predict the linear combination $\mu = l'\beta + h'u$ by $\hat{\mu} = l'\hat{\beta} + h'\hat{u}$.

The MLE of β and ζ under the model (□20), are obtained by solving the normal equations $X'V^{-1}(y - X\beta) = 0;\ (y - X\beta)'V^{-1}\frac{\partial V}{\partial \zeta_l}V^{-1}(y - X\beta) - tr(V^{-1}\frac{\partial V}{\partial \zeta_l}) = 0,\ l = 1,...,L$. In order to deal with possible outliers, the authors propose solving instead,

$$X'V^{-1}U^{1/2}\Psi_G(r) = 0;\ \Psi_G'(r)U^{1/2}V^{-1}\frac{\partial V}{\partial \zeta_l}V^{-1}U^{1/2}\Psi_G(r) - tr(V^{-1}\frac{\partial V}{\partial \zeta_l}cI) = 0,\ l = 1,...,L, \qquad (□21)$$

where $r = U^{-1/2}(y - X\beta)$, $U = Diag[V]$, $\Psi_G(r) = [\Psi_G(r_1), \Psi_G(r_2),...]'$ with $\Psi_G(r_i)$ defining the same influence function as used by Ghosh *et al*. (2008) and $c = E[\Psi_G^2(r_i)]$ ($r_i \sim N(0,1)$). Notice that the normal equations and the robust estimating equations can both be written as sums over the *m* areas. The two sets of equations coincide for $G = \infty$. Denote by $\hat{\beta}_{Rob}, \hat{\zeta}_{Rob}$ the solutions of (□21). The random effects are estimated by solving,

$$Z'R^{-1/2}\Psi[R^{-1/2}(y - X\beta - Zu)] - Q^{-1/2}\Psi(Q^{-1/2}u) = 0. \qquad (□22)$$

Sinha and Rao (2009) estimate the PMSE of the robust predictors by application of the first step of the double-bootstrap procedure of Hall and Maiti (200□) described in Section □1 (Equation □□). All the parameter estimates and random effects predictors needed for the application of this procedure are computed by the robust estimating equations defined by (□21) and (□22).

[illegible] M-quantile estimation

Classical SAE methods under the frequentist approach model the expectations $E(y_i | x_i, u_i)$ and $E(u_i)$. Tzavidis and Chambers (2005) and Chambers and Tzavidis (200[illegible]) propose modelling instead the quantiles of the conditional distribution $f(y_i | x_i)$, where for now y_i is a scalar and x_i is a vector of covariates. Under a linear model for the quantiles, this leads to a family of models indexed by the coefficient $q \in (0,1)$; $\Pr(y_i \le x_i'\beta_q) = q$. In quantile regression the vector β_q is estimated by the vector $\tilde\beta_q$ minimizing,

$$\sum_{i=1}^{n} |y_i - x_i'\tilde\beta_q| \{(1-q)I(y_i - x_i'\tilde\beta_q \le 0) + qI(y_i - x_i'\tilde\beta_q > 0)\}. \quad ([illegible])$$

M-quantile regression uses influence functions for the estimation of β_q by solving the estimating equations,

$$\sum_{i=1}^{n} \Psi_q(r_{iq\Psi}) x_i = 0;\ r_{iq\Psi} = (y_i - x_i'\beta_{q\Psi}),\ \Psi_q(r_{iq\Psi}) = 2\Psi(s^{-1} r_{iq\psi})\{(1-q)I(r_{iq\Psi} \le 0) + qI(r_{iq\Psi} > 0)\}, \quad ([illegible])$$

where s is a robust estimate of scale, and Ψ is an appropriate influence function. The solution of ([illegible]) can be obtained by application of an iterative reweighted least square algorithm (assuming that the influence function is continuous). Denote by $\hat\beta_q$ the (unique) estimate solving ([illegible]). Notice that each sample value (y_i, x_i) lies on one and only one of the M-quantiles $m_q(x_i) = x_i'\hat\beta_q$ (follows from the fact that the quantiles are continuous in q).

How is the M-quantile theory used for SAE? Suppose that the sample consists of unit level observations $\{y_{ij}, x_{ij};\ i = 1,...,m,\ j = 1,...,n_i\}$ (same as under the unit level model (5.[illegible])). Define for unit (i,j) the value q_{ij} such that $x_{ij}'\hat\beta_{q_{ij}} = y_{ij}$. A small area predictor of the mean θ_i is obtained under this approach by averaging the M-quantile coefficients q_{ij} over the sampled units $j \in s_i$, and then computing,

$$\hat\theta_i^M = N_i^{-1}\Big(\sum_{j\in s_i} y_{ij} + \sum_{k\notin s_i} x_{ik}'\hat\beta_{\bar q_i}\Big);\ \bar q_i = \sum_{j=1}^{n_i} q_{ij}/n_i. \quad ([illegible]25)$$

Alternatively, one can average the vector coefficients $\beta_{q_{ij}}$ and replace the vector $\hat\beta_{\bar q_i}$ in ([illegible]25) by the mean $\hat{\bar\beta}_i = \sum_{j=1}^{n_i} \hat\beta_{q_{ij}}/n_i$. The vectors $\hat\beta_{\bar q_i}$ and $\hat{\bar\beta}_i$ account for differences between the areas, playing a similar role to the random effects under the unit level model (5.[illegible]).

The use of this approach is not restricted to the estimation of means although it does assume continuous *y*-values. For example, the distribution function in Area *i* can be estimated as

$$\hat F_i(t) = N_i^{-1}\Big[\sum_{j\in s_i} I(y_{ij} \le t) + \sum_{k\notin s_i} I(x_{ik}'\hat{\bar\beta}_i \le t)\Big]. \quad ([illegible])$$

The M-quantile approach has the apparent advantage of not assuming a parametric model although at the current applications it imposes an implicit assumption that the quantiles are linear in the covariates. Clearly, if the unit level model holds, the use of this model for SAE is more efficient, but the authors illustrate that the use of the M-quantile approach can be more robust to model misspecification. Notice in this regard that the approach is not restricted to any specific definition of the small areas. It accounts also for the possibility of outliers by choosing an appropriate influence function in the estimating equations ([illegible]). On the other hand, there seems to be no obvious way of how to predict the means or other quantities in nonsampled areas. One simple solution would be to set $q = 0.5$ for such areas, or weight the *q*-values of neighboring areas, but then is the question of how to estimate the PMSE without assuming a model. See my comment about deign-based SAE.

[illegible] SAE using Penalized spline regression

Another way of robustifying the general mixed linear regression model defined by ([illegible]1) and ([illegible]20), or other parametric models is by using penalized spline (P-spline) regression. The idea here is not to assume apriori a functional form for the relationship between the outcome variable and the covariates. Suppose that there is a single covariate x. The simple P-spline model assumes simply that $y_i = m_0(x_i) + \varepsilon_i$, $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma_\varepsilon^2$. The mean $m_0(x_i)$ is taken as unknown and is approximated as,

$$m(x;\beta,\gamma) = \beta_0 + \beta_1 x + ... + \beta_p x^p + \sum_{k=1}^{K} \gamma_k (x - K_k)_+^p;\ (x - K_k)_+^p = \max(0, x - K_k)^p, \quad ([illegible])$$

where p is the degree of the spline and $K_1 < ... < K_K$ are fixed knots. For large *K* and a good spread of the knots over the range of *x*, the spline function approximates well most smooth functions. The spline function ([illegible]) uses the basis $\{1, x, \ldots, x^p, (x-K_1)_+^p, \ldots (x-K_k)_+^p\}$ to approximate the function $m_0(x)$ but other bases can be used, particularly when there is more than one covariate.

Opsomer *et al.* (2008) use P-spline regression for SAE by treating the γ-coefficients as an additional set of random effects. Suppose as under the unit level model that the data consist of the n observations $\{y_{ij}, x_{ij}; i=1,\ldots,m,\ j=1,\ldots,n_i\}$. For unit j in area i, the model considered is,

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \ldots + \beta_p x_{ij}^p + \sum_{k=1}^{K} \gamma_k (x_{ij} - K_k)_+^p + u_i + \varepsilon_{ij}\ , \tag{5.28}$$

where the u_i's are the area random effects and the ε_{ij}'s are model residuals. Let $u = (u_1, \ldots, u_m)'$ and $\gamma = (\gamma_1, \ldots, \gamma_K)'$. Defining $d_{ij} = 1(0)$ if unit j is (is not) in area i and denoting $d_j = (d_{1j}, \ldots, d_{mj})'$, $D = [d_1, \ldots, d_n]'$, the full model for the vector y of all the sample observations can be written compactly as,

$$y = X\beta + Z\gamma + Du + \varepsilon;\ \gamma \sim (0, \sigma_\gamma^2 \mathrm{I}_k),\ u \sim (0, \sigma_u^2 \mathrm{I}_m),\ \varepsilon \sim (0, \sigma_\varepsilon^2 \mathrm{I}_n), \tag{5.29}$$

where, $X = [x_1^{(p)}, \ldots, x_n^{(p)}]'$ with $x_l^{(p)} = (1, x_l, \ldots, x_l^p)'$ and $Z = [z_1, \ldots, z_n]'$ with $z_l = [(x_l - K_1)_+^p, \ldots, (x_l - K_K)_+^p]'$. Written that way the model (5.29) looks similar to the mixed linear model (5.20) but notice that under (5.29) the observations y_{ij} are not independent between areas because of the common random effects γ. Nonetheless, the BLUP and EBLUP predictors of (β, u, γ) are obtained using standard results. The small area predictors are then,

$$\hat{\theta}_{i,EBLUP}^{\text{P-}spline} = \hat{\beta}' \bar{X}_i^{(p)} + \hat{\gamma}' \bar{Z}_i + \hat{u}_i;\ \bar{X}_i^{(p)} = \sum_{l \in U_i} x_l^{(p)} / N_i,\ \bar{Z}_i = \sum_{l \in U_i} z_l / N_i\ . \tag{5.30}$$

Opsomer et al. (2008) derive the PMSE of the EBLUP in (5.30) when estimating the unknown variances by restricted ML that is correct to the second order, and an estimator of the PMSE with bias that is also correct to the second order. The authors propose also a nonparametric bootstrap algorithm for estimating the PMSE and for testing the hypotheses $\sigma_u^2 = 0$ and $\sigma_\gamma^2 = 0$. Rao *et al.* (2009) use a similar model to (5.29) with $p=1$, but rather than using the EBLUP predictors under the model the authors develop estimators that are robust to outliers, similarly (but not the same) to the methodology developed by Sinha and Rao (2009) for the mixed linear model described in Section 5.5.

5.8 *Prediction of ordered area means*

Malinovsky and Rinott (2010) consider the following (hard) problem: predict the ordered area means $(\theta_{(1)} \le \theta_{(2)} \le \ldots \le \theta_{(m)})$ under the area level model $\tilde{y}_i = \mu + u_i + e_i = \theta_i + e_i$ (special case of Eq. 5.1) with $u_i \overset{iid}{\sim} H(0, \sigma_u^2)$, $e_i \overset{iid}{\sim} G(0, \sigma_e^2)$. H and G are general distributions with zero means and variances σ_u^2 and σ_e^2. This problem is different from just ranking the small area means, one of the famous triple-goal estimation in SAE put forward by Shen and Louis (1998). The triple-goal estimation consists of 'good' area specific estimation, 'good' estimates of the histogram (distribution) and 'good' estimates of the ranks. See Rao (2003) for further discussion.

In order to illustrate the difference between the prediction of the unordered means and the ordered means, consider the prediction of $\theta_{(m)} = \max(\theta_i)$. If $E(\hat{\theta}_i|\theta_i) = \theta_i$, then $E(\max_i \hat{\theta}_i | \{\theta_j\}) > \theta_{(m)}$ so that taking simply the largest estimator of the area means as the predictor of the true largest mean results in overestimation. On the other hand, using the Bayesian predictor $\theta_i^* = E[\theta_i | \{\hat{\theta}_j\}]$ implies that $\max_i(\theta_i^*) < E(\theta_{(m)})$, an underestimation in expectation.

Wright, Stern and Cressie considered the problem of predicting ordered means from a Bayesian perspective but this approach requires heavy numerical calculations and seems to be sensitive to the choice of priors. Malinovsky and Rinott (2010) compare three simple predictors under the frequentist approach using the loss function, $L(\tilde{\theta}_{()}, \theta_{()}) = \sum_{i=1}^{m} (\tilde{\theta}_{(i)} - \theta_{(i)})^2$ and the Bayesian risk $E[L(\tilde{\theta}_{()}, \theta_{()})]$. Let $\hat{\theta}_i$ define the direct estimator of θ_i ($\tilde{y}_i$ in the notation of Section 5.2). The predictors compared are:

$$\tilde{\theta}_{(i)}^{(1)} = \hat{\theta}_{(i)};\ \tilde{\theta}_{(i)}^{(2)}(\delta) = \delta\hat{\theta}_{(i)} + (1-\delta)\bar{\hat{\theta}},\ \bar{\hat{\theta}} = \sum_{i=1}^{m} \hat{\theta}_i / m;\ \tilde{\theta}_{(i)}^{(3)} = E(\theta_{(i)} | \hat{\theta}),\ \hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)\ , \tag{5.31}$$

where $\hat{\theta}_{(i)}$ is the i-th ordered statistic of the direct estimators. The predictors in (5.31) assume that σ_u^2 and σ_e^2 are known but μ is unknown and estimated by $\bar{\hat{\theta}}$. In practice, the unknown variances are substituted by appropriate sample estimates, yielding the corresponding empirical predictors. The predictor $\tilde{\theta}_{(i)}^{(1)}$ cannot in general be computed explicitly for $m>2$ but the authors show that it can be efficiently replaced by the predictor $\tilde{\theta}_{(i)}^{(2)}$ with an appropriate choice of the shrinkage coefficient δ.

Denote by $\tilde{\theta}^{(k)}$ the predictor of the ordered true means when using the predictors $\theta_{(i)}^{(k)}$, $i=1,...,m$, $k=1,2$, and let $\gamma=\sigma_u^2(\sigma_u^2+\sigma_e^2)^{-1}$ be the optimal shrinkage coefficient under the model when predicting the unordered means (see Equation 5.2). The authors derive several interesting theoretical comparisons and further comparisons based on simulations. For example,

$$\text{If } \gamma \le (m-1)^2/(m+1)^2 \text{ then } E[L(\tilde{\theta}_{()}^{(2)}(\delta),\theta_{()})] \le E[L(\tilde{\theta}_{()}^{(1)},\theta_{()})] \text{ for all } \gamma \le \delta \le 1. \quad (6.2)$$

Notice that $\lim_{m\to\infty}[(m-1)^2/(m+1)^2]=1$ implying that (6.2) holds asymptotically for all γ, and that the inequality $\gamma \le \delta \le 1$ implies less shrinkage of the direct estimators towards the mean. In particular, the optimal choice δ^{opt} of δ for $\tilde{\theta}^{(2)}(\delta)$ satisfies $\lim_{m\to\infty}\delta^{opt}=\sqrt{\gamma}$.

The results so far assume general distributions *H* and *G* of the random effects and residual terms. When these distributions are normal, then for $m=2$ $E[L(\tilde{\theta}_{()}^{(1)},\theta_{()})] \le E[L(\tilde{\theta}_{()}^{(2)}(\delta),\theta_{()})]$. A conjecture supported by simulations is that this relation holds also for $m>2$ and the optimal choice δ^{opt}. The simulations suggest also that for sufficiently large m (e.g., $m \ge 25$), $\tilde{\theta}^{(1)}$ is efficiently replaced by $\tilde{\theta}^{(2)}(\sqrt{\gamma})$.

7. Summary

In this article I tried to review some of the main developments in SAE in recent years. There are a few other developments listed below that I intend to add to the article in the near future.

1. *The use of SAE in conjunction with censuses*: as mentioned in the introduction, SAE is increasingly used in modern censuses to test and correct the administrative data for under-coverage or over-coverage and to supplement information not contained in these records. A good reference is Nirel and Glickman (2009).

2. *Small area estimation with varying boundaries*: Moura *et al*. (2005) investigate the use of hierarchical models for SAE with varying area boundaries. The paper shows how area estimates and corresponding PMSE estimates can be obtained at a variety of nested and intersecting boundary systems by fitting the model at the lowest possible level. See also Section [illegible] on M-quantile estimation.

3. *Accounting for Spatial correlations between the area means*: The models considered in this article assume that the area random effects are independent between the areas. Since the random effects account for the unexplained variation by the existing covariates, it is often reasonable to assume that the random effects in neighbouring areas are correlated. See, e.g., Pratesi and Salvati (200[illegible]) and De Souza *et al*. (2009).

4. *SAE estimation under informative sampling*: All the studies reviewed so far assume implicitly that the selection of the sampled areas is noninformative, and similarly for the sampling within the selected areas, implying that the population model applies to the observed sample data. This, however, may not be the case and ignoring the effects of informative sampling may bias the inference very severely. This problem is considered in Malec *et al*. (1999) and Pfeffermann and Sverchkov (200[illegible]).

5. *The use of a two-part model for SAE*: It is sometimes the case that the outcome value is either zero or an observation from a continuous distribution. A typical example is the assessment of literacy proficiency with the possible outcome being either zero, indicating illiteracy, or a positive score measuring the level of literacy. Another example is the consumption of illicit drugs. Available SAE methods are not suitable for this kind of data because of the mixed distribution of the outcome variable. Pfeffermann *et al*. (2008) considered the estimation of the average outcome, or the proportion of positive outcomes in small areas for this kind of situations using a Bayesian methodology.

6. *Predictive fence mmethod for Small Area Estimation*: Model selection is one of the major problems in SAE because the models usually involve unobserved random effects with limited or no information on their distribution. Jiang *et al*. (2008) propose a class of strategies called fence methods for mixed model selection, which includes linear and generalized linear mixed models. The strategies involve a procedure to isolate a

su□grou□ of correct models (of □hich the o□timal model is a mem□er) and then selecting the o□timal model according to some s□ecified criterion. □iang *et al*. (2010) further de□elo□ the method for selecting non□arametric s□line models for small area estimation.

References

□attese, □.□., □arter, □.M. and □uller, □ .□. (1988). □n error com□onents model for □rediction of county cro□ area using sur□ey and satellite data. *Journal of the American Statistical Association* **83**, 28–□□.

□ell, □ .□. and □uang, □.T. (200□). □sing the *t*-distri□ution to deal □ith outliers in small area estimation. Proceedings of Statistics Canada Sym□osium on Methodological Issues in Measuring Po□ulation □ealth.

Cham□ers, □. and T□a□idis, □. (200□). M-□uantile models for small area estimation. *Biometrika* **93**, 255-2□8.

Chen, S. and Lahiri, P. (2002). □n mean s□uared □rediction error estimation in small area estimation □ro□lems. In Proceedings of the Sur□ey □esearch Methods Section, □merican Statistical □ssociation. □□. □□□-□□□.

□atta, □.S. and Lahiri, P. (2000). □ unified measure of uncertainty of estimated □est linear un□iased □redictors in small area estimation □ro□lems. *Statistica Sinica* **10**, □1□–□2□.

□atta, □.S., □ao, □□.□. and Smith, □.□. (2005). □n measuring the □aria□ility of small area estimators under a □asic area le□el model. *Biometrika* **92**, 18□–19□.

□atta, □.S. (2009). Model-□ased □□roach to Small □rea □stimation. In: □and□ook of Statistics 29□□*Sample Surveys*: *Inference and Analysis*. □ds. □. Pfeffermann and C.□. □ao. □orth □olland. □□. 251-288.

□atta, □.S., □hosh, M., Steorts, □. and Ma□les, □ (2010). □ayesian □enchmarking □ith □□□lications to Small □rea □stimation. *The Canadian Journal of Statistics* **38** (to a□□ear).

□e Sou□a, □.□., Moura, □.□.S. and Migon, □.S. (2009). Small □rea Po□ulation Prediction □ia □ierarchical Models. *Survey Methodology* **35** (forthcoming).

□e□ille, □C. and S□rndal, C.□. (1992). Cali□ration estimators in sur□ey sam□ling. *Journal of the American Statistical Association* **87**, □□□–□82.

□re□, □., Singh, M.P., and Choudhry, □.□. (1982). □□aluation of small area estimation techni□ues for the Canadian La□our □orce Sur□ey. *Survey Methodology* **8**, 1□-□□.

□ste□ao, V.M. and S□rndal, C.□. (200□). □orro□ing strength is not the □est techni□ue □ithin a □ide class of design-consistent domain estimators. *Journal of Official Statistics* **20**, □□5–□□9.

□ay, □.□. and □erriot, □.□. (19□9). □stimates of income for small □laces: an a□□lication of □ames-Stein □rocedures to census data. *Journal of the American Statistical Association* **74**, 2□9–2□□.

□anesh, □. and Lahiri, P. (2008). □ ne□ class of a□erage moment matching □rior. *Biometrika* **95**, 51□-520.

□hosh, M., Sinha, □. and □im, □. (200□). □m□irical and hierarchical □ayesian estimation in finite □o□ulation sam□ling under structural measurement error models. *Scandinavian Journal of Statistics* **33**, 591-□08.

□hosh, M., Maiti, T. and □oy, □. (2008). Influence functions and ro□ust □ayes and em□irical □ayes small area estimation. □iometrika **95**, 5□□-585.

□all, P. and Maiti, T. (200□□). □n □arametric □ootstra□methods for small area □rediction. *Journal of the Royal Statistical Society, Series B* **68**, 221- 2□8.

□uang, □.T. and □ell, □ .□. (200□). □sing the *t*-distri□ution in small area estimation: an a□□lication to S□IP□ state □o□erty models. In Proceedings of the Sur□ey □esearch Methods Section, □merican Statistical □ssociation. □□. □1□2-□1□9.

Isaki, C.T., Tsay, □□. and □uller, □ .□. (2000). □stimation of census ad□ustment factors. Sur□ey Methodology **26**, □1-□2.

□iang, □ Lahiri, P. and □ an, S. M. (2002). □ unified □ackknife theory for em□irical □est □rediction □ith M-estimation. *The Annals of Statistics* **30**, 1□82–1810.

□iang, □ and Lahiri, P.S. (200□a). □stimation of finite □o□ulation domain means: a model-assisted em□irical □est □rediction a□□roach. *Journal of the American Statistical Association* **101**, □01–□11.

□iang, □ and Lahiri, P.S. (200□□). Mi□ed model □rediction and small area estimation. *Test* **15**, 1–9□

□iang, □, □ao, □ S., □u, □. and □guyen, T. (2008). □ence methods for mi□ed model selection. *Annals of Statistics* **36**, 1□□9-1□92.

□iang, □, □guyen, T. and □ao, □ S. (2010). □ence method for non□arametric small a□rea estimation. *Survey Methodology* **36** (□orthcoming).

□ott, P.S. (2009). Cali□ration□ eighting: Com□ining Pro□a□ility Sam□les and Linear Prediction Models. In: □and□ook of Statistics 29□□*Sample Surveys*: *Inference and Analysis*. □ds. □. Pfeffermann and C.□. □ao. □orth □olland. □□ 55-82.

Lahiri, P. and □ao, □□.□. (1995). □o□ust estimation of mean s□uared error of small area estimators. *Journal of the American Statistical Association* **90**, □58–□□□

Lehtonen, □., S□rndal, C.□. and Vei□anen, □. (200□). The effect of model choice in estimation for domains, including small domains. *Survey Methodology* **29**, □□–□□

Lehtonen, □., S□rndal, C.□. and Vei□anen, □. (2005). □oes the model matter□ com□aring model-assisted and model-de□endent estimators of class fre□uencies for domains. *Statistics in Transition* **7**, □9–□□□

Lehtonen, □. and Vei□anen, □. (2009). □esign-□ased Methods of □stimation for □omains and Small □reas. In: □and□ook of Statistics 29□□*Sample Surveys*: *Inference and Analysis*. □ds. □. Pfeffermann and C.□. □ao. □orth □olland. □□ 219-2□9.

Lohr, S.L. and □ao, □□.□. (2009). □ackknife estimation of mean s□uared error of small area □redictors in nonlinear mi□ed models. *Biometrika* **96**, □5□-□□8.

Macgi□□on, □. and Tom□erlin, T. □ (1989). Small area estimates of □ro□ortions □ia em□irical □ayes techni□ues. *Survey Methodology* **15**, 2□□-252.

Malec, □., Sedransk, □, Moriarity, C.L. and LeClere, □.□. (199□). Small area inference for □inary □aria□les in the □ational □ealth Inter□ie□ Sur□ey. *Journal of the American Statistical Association* **92**, 815–82□

Malec, □., □a□is, □ . □ . and Cao, X. (1999). Model-□ased small area estimates of o□er□eight □re□alence using sam□le selection ad□ustment. *Statistics in Medicine* **18**, □189–□200.

Malino□sky, □. and □inott, □. (2010). Prediction of ordered random effects in a sim□le small area model. *Statistica Sinica* (forthcoming).

Moura, f., Clarke, P. and Pfeffermann, □. (2005). Small area estimation under □arying area □oundaries using the synthetic estimator. Statistics in Transition □ □25-□□□

□irel, □. and □lickman, □. (2009). Sam□le sur□eys and censuses. In: □and□ook of Statistics 29□□*Design, Methods and Applications.* □ds. □. Pfeffermann and C.□. □ao. □orth □olland. □□ 5□9-5□5.

□□somer, □□., Claeskens,□., □analli, M.□., □auemann, □. and □reidt, □.□ (2008). □on-□arametric small area estimation using □enali□ed s□line regression. *Journal of the Royal Statistical Society*, series □ **70**, 2□5-28□

Pfeffermann, □. and □athan, □. (1981). □egression □nalysis of □ata from a Cluster Sam□le. *Journal of the American Statistical Association* **76**, □81-□89.

Pfeffermann, □. (2002). Small □rea □stimation- □e□ □e□elo□ments and □irections. I*nternational Statistical Review* **70**, 125-1□□.

Pfeffermann, □. and Tiller, □. (200□). Small □rea □stimation □ith State-S□ace Models Su□□ect to □enchmark Constraints. *Journal of the American Statistical Association* **101**, 1□8□-1□9□.

Pfeffermann, □. and S□erchko□, M. (200□). Small-area estimation under informati□e □ro□a□ility sam□ling of areas and □ithin the selected areas. □ournal of the □merican Statistical □ssociation **102**, 1□2□-1□□9.

Pfeffermann, □., Terryn □. and Moura, □.□.S. (2008). Small area estimation under a t□o-□art random effects model □ith a□□lication to estimation of literacy in de□elo□ing countries. *Survey Methodology* **34**, 2□5-2□9.

Prasad, □.□.□. and □ao, □.□.□. (1990). The estimation of mean s□uared errors of small area estimators. *Journal of the American Statistical Association* **85**, 1□□-1□1.

Pratesi, M. and Sal□ati, □. (200□). Small area estimation: the □□L□P estimator □ased on s□atially correlated random area effects. *Statistical Methods and Application* 1□, 11□-1□1.

□ao, □ □. □. (200□). *Small area estimation*. □ohn □ iley □ Sons, □e□ □ork.

□ao, □.□.□., Sinha, S.□. and □oknossadati M. (2009). □o□ust small area estimation using □enali□ed s□line mi□ed models. In Proceedings of the Sur□ey □esearch Methods Section, □merican Statistical □ssociation. □□ 1□5-15□.

S□rndal, C.□. and □idiroglou, M.□. (1989). Small domain estimation: a conditional analysis. *Journal of the American Statistical Association* **84**, 2□□-2□5.

Shen, □ . and Louis, T.□. (1998). Tri□le-goal estimates in t□o-stage hierarchical models. *Journal of the Royal Statistical Society*, Series □ **60**, □55-□□1.

Sinha, S.□. and □ao, □.□.□ (2009). □o□ust small area estimation. *The Canadian Journal of Statistics* **37**, □81-□99.

T□a□idis, □. and Cham□ers, □. (2005). □ias ad□usted estimation for small areas □ith M-□uantile models. Statistics in Transition **7**, □0□-□1□.

Tora□i, M. and □ao, □.□.□. (2008). Small area estimation under a t□o-le□el model. *Survey Methodology* **34**, 11–1□.

Tora□i, M., □atta, □.S. and □ao, □.□.□. (2009). □m□irical □ayes estimation of small area means under a nested error linear regression model □ith measurement errors in the co□ariates. Scandina□ian □ournal of Statistics **36**, □55-□□8.

□ ang, □., □uller, □ .□. and □u, □. (2008). Small area estimation under a restriction. *Survey Methodology* **34**, 29-□□.

□ right, □.L., Stern, □.S. and Cressie, □. (200□). Loss function for estimation of e□treme □ith an a□□lication to disease ma□□ing. *The Canadian Journal of Statistics* **31**, 251-2□□.

□□arra, L.M.□. and Lohr, S. L. (2008). Small area estimation □hen au□iliary information is measured □ith error. *Biometrika* **95**, 919-9□1.

□ou, □. and □ao, □.□.□. (2002). □ □seudo-em□irical □est linear un□iased □rediction a□□roach to small area estimation using sur□ey □eights. *The Canadian Journal of Statistics* **30**, □□1-□□9.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 205–205.

SESSION 5: STATISTICAL INFERENCE

Chair: Zahirul Hoque
Department of Statistics
College of Business and Economics
United Arab Emirates University
E-mail: zahirul.hoque@uaeu.ac.ae

MULTI-TREATMENT LOCATION-INVARIANT OPTIMAL RESPONSE-ADAPTIVE DESIGNS FOR CONTINUOUS RESPONSES

Atanu Biswas
Applied Statistics Unit, Indian Statistical Institute
203 B. T. Road, Kolkata 700 108, India
E-mail: atanu@isical.ac.in

Saumen Mandal
Department of Statistics, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada

ABSTRACT

Optimal response-adaptive designs in phase III clinical trial involving two or more treatments at hand is of growing interest. Optimal response-adaptive designs were provided by Rosenberger et al. (2001) and Biswas and Mandal (2004) [BM] for binary responses and continuous responses respectively. Zhang and Rosenberger (2006) [ZR] provided another design for normal responses. Biswas, Bhattacharya and Zhang (2007) [BBZ] pointed out some serious drawback of the ZR design. Moreover, all the earlier works of BM, ZR and BBZ suffer seriously if there is any common shift in location to observe the responses. The present paper provides a location invariant design for that purpose, and then extends the present approach for more than two treatments. The proposed methods are illustrated using some real data sets.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 206–206.

ESTIMATION UNDER ASYMMETRIC LOSSES

Zahirul Hoque
Department of Statistics
College of Business and Economics
United Arab Emirates University
E-mail: zahirul.hoque@uaeu.ac.ae

ABSTRACT

Due to the symmetric nature of the squared error loss function it fails to differentiate between over- and under-estimation. This paper considers the estimation of the parameters of the linear models under an asymmetric loss function, namely the linex loss. To illustrate the procedure the risk functions of the preliminary test estimators (PTE) of the parameters of a simple linear regression model are derived. The performance of the PTE is compared with that of the least square estimator. We also consider a class of distribution, called, elliptically contoured distribution (ECD). The derivation of the risk function of the parameters of ECD is also shown in this paper.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 206–206.

INTERMEDIATE MONITORING SAMPLE SIZE REASSESSMENT IN MULTI-TREATMENT OPTIMAL RESPONSE-ADAPTIVE GROUP SEQUENTIAL DESIGNS WITH CONTINUOUS RESPONSES

Pinakpani Pal
Indian Statistical Institute, Calcutta, India
E-mail: pinak@isical.ac.in

ABSTRACT

Optimal response-adaptive designs in phase III clinical trial set up are becoming more and more current interest. In the present paper, an optimal response-adaptive design is introduced for more than two treatments at hand. We consider the situation of continuous responses. We minimize an objective function subject to one or more inequality constraints. We propose an extensive computer search algorithm. The proposed procedure is illustrated with extensive numerical computation and simulations. Some real data set is used to illustrate the proposed methodology. We also extend our approach in the presence of covariates.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 207–235.

WHICH QUANTILE IS THE MOST INFORMATIVE? MAXIMUM LIKELIHOOD, MAXIMUM ENTROPY AND QUANTILE REGRESSION

Anil K. Bera
Department of Economics, University of Illinois,
1407 W. Gregory Drive, Urbana, IL 61801.
E-mail: abera@illinois.edu

Antonio F. Galvao Jr.
Department of Economics, University of Iowa,
W334 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA, 52242.
E-mail: antonio-galvao@uiowa.edu

Gabriel V. Montes-Rojas
Department of Economics, City University of London,
10 Northampton Square, London EC1V 0HB, U.K.
E-mail: Gabriel.Montes-Rojas.1@city.ac.uk

Sung Y. Park
Department of Economics, Chinese University of Hong Kong,
Shatin, NT, Hong Kong.
E-mail: sungpark@cuhk.edu.hk

ABSTRACT

This paper studies the connections among quantile regression, the asymmetric Laplace distribution, maximum likelihood and maximum entropy. We show that the maximum likelihood problem is equivalent to the solution of a maximum entropy problem where we impose moment constraints given by the joint consideration of the mean and median. Using the resulting score functions we propose an estimator based on the joint estimating equations. This approach delivers estimates for the slope parameters together with the associated "most probable" quantile. Similarly, this method can be seen as a penalized quantile regression estimator, where the penalty is given by deviations from the median regression. We derive the asymptotic properties of this estimator by showing consistency and asymptotic normality under certain regularity conditions. Finally, we illustrate the use of the estimator with a simple application to the U.S. wage data to evaluate the effect of training on wages.

Keywords: Asymmetric Laplace Distribution, Quantile Regression, Treatment Effects

1. INTRODUCTION

Different choices of loss functions determine different ways of defining the location of a random variable y. For example, squared, absolute value, and step function lead to mean, median and mode, respectively (see Manski, 1991, for a general discussion). For a given quantile $\tau \in (0,1)$, consider the loss function in a standard quantile estimation problem,

$$L_{1,n}(\mu;\tau) = \sum_{i=1}^{n} \rho_\tau(y_i - \mu) = \sum_{i=1}^{n} (y_i - \mu)\left(\tau - 1(y_i \leq \mu)\right), \tag{1}$$

as proposed by Koenker and Bassett (1978). Minimizing $L_{1,n}$ with respect to the location parameter μ is identical to maximizing the likelihood based on the asymmetric Laplace probability density (ALPD):

$$f(y;\mu,\tau,\sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left(-\frac{\rho_\tau(y-\mu)}{\sigma}\right), \tag{2}$$

for given τ. The well known symmetric Laplace (double exponential) distribution is a special case of (2) when τ=1/2.

Several studies developed the properties of the maximum likelihood (ML) estimators based on ALPD. Hinkley and Revankar (1977) derived the asymptotic properties of the unconditional MLE under ALPD. Kotz, Kozubowski, and Podgórsk (2002b) and Yu and Zhang (2005) consider alternative MLE approaches for ALPD. Moreover, models based on ALPD have been proposed in different contexts. Machado (1993) used the ALPD to derive a Schwartz information crietrion for model selection for quantile regression (QR) models, and Koenker and Machado (1999) introduced a goodness-of-fit measure for QR and related inference processes. Yu and Moyeed (2001) and Geraci and Botai (2007) used a Bayesian QR approach based on the ALPD. Komunjer (2005) constructed a new class of estimators for conditional quantiles in possibly misspecified nonlinear models with time series data. The proposed estimators belong to the family of quasi-maximum likelihood estimators (QMLEs) and are based on a family of 'tick-exponential' densities. Under the asymmetric Laplace density, the corresponding QMLE reduces to the Koenker and Bassett (1978) linear quantile regression estimator. In addition, Komunjer (2007) developed a parametric estimator for the risk of financial time series expected shortfall based on the asymmetric power distribution, derived the asymptotic distribution of the maximum likelihood estimator, and constructed a consistent estimator for its asymptotic covariance matrix.

Interestingly, the parameter μ in functions (1) and (2) is at the same time the location parameter, the τ-th quantile, and the mode of the ALPD. For the simple (unconditional) case, the minimization of (1) returns different order-statistics. For example, if we set $\tau = \{0.1, 0.2, ..., 0.9\}$, the solutions are, respectively, the nine deciles of y. In order to extract important information from the data a good summary statistic would be to choose *one* order statistics accordingly the most likely value. For a symmetric distribution one would choose the median. Using the ALPD, for given τ, maximization of the corresponding likelihood function gives that particular order statistics. Thus, the main idea of this paper is to *jointly* estimate τ and the corresponding order statistic of y which can be taken as a good summary statistic of the data. The above notion can be easily extended to modeling the "conditional location" of y given covariates x, as we do in Section 2.3. In this case, the ALPD model provides a twist to the QR problem, as now τ becomes the *most likely quantile* in a regression set-up.

The aim of this paper is threefold. First, we show that the score functions implied by the ALPD-ML estimation are not restricted to the true data generating process being ALPD, but they arise as the solution to a maximum entropy (ME) problem where we impose moment constraints given by the joint consideration of the mean and median. By so doing, the ALPD-ML estimator combines the information in the mean and the median to capture the asymmetry of the underlying empirical distribution (see e.g. Park and Bera (2009) for a related discussion).

Second, we propose a novel Z-estimator that is based on the estimating equations from the MLE score functions (which also correspond to the ME problem). We refer to this estimator as ZQR. The approximate Z-estimator do not impose that the underline distribution is ALPD. Thus, although the original motivation for using the estimating equations is based on the ALPD, the final estimator is independent of this requirement. We derive the asymptotic properties of the estimator by showing consistency and asymptotic normality under certain regularity conditions. This approach delivers estimates for the slope parameters together with the associated *most probable* quantile. The intuition behind this estimator works as follows. For the symmetric and unimodal case the selected quantile is the median, which coincides with the mean and mode. On the other hand, when the mean is larger than the median, the distribution is right skewed. Thus, taking into consideration the empirical distribution, there is more probability mass to the left of the distribution. As a result it is natural to consider a point estimate in a place with more probability mass. The selected τ-quantile does not necessarily lead to the mode, but to a point estimate that is most probable. This provides a new interpretation of QR and frames it within the ML and ME paradigm.

The proposed estimator has an interesting interpretation from a policy perspective. The QR analysis gives a full range of estimators that account for heterogeneity in the response variable to certain covariates. However, the proposed ZQR estimator answers the question: of all the heterogeneity in the conditional regression model, which one is more likely to be observed? In general, the entire QR process is of interest because we would like to either test global hypotheses about conditional distributions or make comparisons across different quantiles (for a discussion about inference in QR models see Koenker and Xiao, 2002). But selecting a particular quantile provides an estimator as parsimonious as ordinary least squares (OLS) or the median estimators. The proposed estimator is, therefore, a complement to the QR analysis rather than a competing alternative. This set-up also allows for an alternative interpretation of the QR analysis. Consider, for instance, the standard conditional regression set-up, $y = x'\beta + u$, and let β be partitioned into $\beta = (\beta_1, \beta_2)$. For a given value of $\beta_1 = \bar{\beta}_1$, we may be interested in finding the representative quantile of the unobservables distribution that corresponds to this level of β_1. For such a case, instead of assuming a given quantile τ we would like to estimate it. In other words, the QR process provides us with the graph $\beta_1(\tau)$, but the graph $\tau(\beta_1)$ could be of interest too.

Finally, the *third* objective of this work is to illustrate the implementation of the proposed ZQR estimator. We apply the estimator to the estimation of quantile treatment effects of subsidized training on wages under the Job Training Partnership Act (JTPA). We discuss the relationship between OLS, median regression and ZQR estimates of the JTPA treatment effect. We show that each estimator provides different treatment effect estimates. Moreover, we extend our ZQR estimator to Chernozhukov and Hansen (2006, 2008) instrumental variables strategy in QR.

The rest of the paper is organized as follows. Section 2 develops the ML and ME frameworks of the problem. Section 3 derives the asymptotic distribution of the estimators. In Section 4 we report a small Monte Carlo study to assess the finite sample performance of the estimator. Section 5 deals with an empirical illustration to the effect of training on wages. Finally, conclusions are in the last section.

2. MAXIMUM LIKELIHOOD AND MAXIMUM ENTROPY

In this section we describe the MLE problem based on the ALPD and show its connection with the maximum entropy. We show that they are equivalent under some conditions. In the next section we will propose an Z-estimator based on the resulting estimating equations from the MLE problem, which corresponds to ME.

2.1 Maximum Likelihood

Using (2), consider the maximization of the log-likelihood function of an ALPD:

$$L_{2,n}(\mu,\tau,\sigma) = n\ln\left(\frac{1}{\sigma}\tau(1-\tau)\right) - \sum_{i=1}^{n}\frac{1}{\sigma}\rho_\tau(y_i-\mu) = n\ln\left(\frac{1}{\sigma}\tau(1-\tau)\right) - \frac{1}{\sigma}L_{1,n}(\mu;\tau), \quad (3)$$

with respect to μ, τ and σ. The first order conditions from (3) lead to the following estimating equations (EE):

$$\sum_{i=1}^{n}\frac{1}{\sigma}\left(\frac{1}{2}\mathrm{sign}(y_i-\mu)+\tau-\frac{1}{2}\right) = 0, \quad (4)$$

$$\sum_{i=1}^{n}\left(\frac{1-2\tau}{\tau(1-\tau)} - \frac{(y_i-\mu)}{\sigma}\right) = 0, \quad (5)$$

$$\sum_{i=1}^{n}\left(-\frac{1}{\sigma}+\frac{1}{\sigma^2}\rho_\tau(y_i-\mu)\right) = 0. \quad (6)$$

Let $(\hat{\mu}, \hat{\tau}, \hat{\sigma})$ denote the solution to this system of equations. The first equation leads to the most probable order statistic. Once we have $\hat{\tau}$, $(1-2\hat{\tau})$ will provide a measure of asymmetry of the distribution. Equation (6) provides a straightforward measure of dispersion, namely,

$$\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n} \rho_{\hat{\tau}}(y_i - \hat{\mu}).$$

Then, the loss function corresponding to (3) can be rewritten as a two-parameter loss function

$$-\frac{1}{n}L_{2,n}(\mu,\tau) = \ln\left(\frac{1}{n}L_{1,n}(\mu;\tau)\right) - \ln\left(\tau(1-\tau)\right). \tag{7}$$

This determines that $L_{2,n}(\mu,\tau,\sigma)$ can be seen as a penalized quantile optimization function, where we minimize $\ln\left(\frac{1}{n}L_{1,n}(\mu;\tau)\right)$ and penalize it by $-\ln\left(\tau(1-\tau)\right)$. The penalty can be interpreted as the cost of deviating from the median, i.e. for $\tau = 1/2$, $-\ln\left(\tau(1-\tau)\right) = -\ln(1/4)$ is the minimum, while for either $\tau \to 0$ or $\tau \to 1$ the penalty goes to $+\infty$.

It is important to note that the structure of the estimating functions suggests that the solution to the MLE problem can be obtained by first obtaining every quantile of the distribution, and then plugging them (with the corresponding estimator for σ) in (5) until this equation is satisfied (if the solution is unique). In other words, given all the quantiles of y, the problem above selects the most likely quantile as if the distribution of y were ALPD.

2.2 Maximum Entropy

The ALPD can be characterized as a maximum entropy density obtained by maximizing Shannon's entropy measure subject to two moment constraints (see Kotz, Kozubowski, and Podgórsk, 2002a):

$$f_{ME}(y) \equiv \arg\max_{f} - \int f(y)\ln f(y)dy \tag{8}$$

subject to

$$E|y-\mu| = c_1, \tag{9}$$
$$E(y-\mu) = c_2, \tag{10}$$

and the normalization constraint, $\int f(y)dy = 1$, where c_1 and c_2 are known constants. The solution to the above optimization problem using the Lagrangian has the familiar exponential form

$$f_{ME}(y:\mu,\lambda_1,\lambda_2) = \frac{1}{\Omega(\theta)}\exp\left[-\lambda_1|y-\mu| - \lambda_2(y-\mu)\right], \quad -\infty < y < \infty, \tag{11}$$

where λ_1 and λ_2 are the Lagrange multipliers corresponding to the constraints (9) and (10), respectively, $\theta = (\mu,\lambda_1,\lambda_2)'$ and $\Omega(\theta)$ is the normalizing constant. Note that $\lambda_1 \in \mathbb{R}^+$ and $\lambda_2 \in [-\lambda_1,\lambda_1]$ so that $f_{ME}(y)$ is well-defined. Symmetric Laplace density (LD) is a special case of ALPD when λ_2 is equal to zero.

Interestingly, the constraints (9) and (10) capture, respectively, the dispersion and asymmetry of the ALPD. The marginal contribution of (10) is measured by the Lagrangian multiplier λ_2. If λ_2 is close to 0, then (10) does not have useful information for the data, and therefore, the symmetric LD is the most appropriate. In this case, μ is known to be the median of the distribution. On the other hand, when λ_2 is not close to zero, it measures the degree of asymmetry of the ME distribution. Thus the non-zero value of λ_2 makes $f_{ME}(\cdot)$ deviate from the symmetric LD, and therefore, changes the location, μ, of the distribution to adhere the maximum value of the entropy (for general notion of entropy see Soofi and Retzer, 2002).

Let us write (9) and (10), respectively, as

$$\int \phi_1(y,\mu) f_{ME}(y:\mu,\lambda_1,\lambda_2)dy = 0 \quad \text{and} \quad \int \phi_2(y,\mu) f_{ME}(y:\mu,\lambda_1,\lambda_2)dy = 0,$$

where $\phi_1(y,\mu) = |y-\mu| - c_1$ and $\phi_2(y,\mu) = (y-\mu) - c_2$. By substituting the solution $f_{ME}(y:\mu,\lambda_1,\lambda_2)$ into the Lagrangian of the maximization problem in (8), we obtain the profiled objective function

$$h(\lambda_1,\lambda_2,\mu) = \ln \int \exp\left[-\sum_{j=1}^{2} \lambda_j \phi_j(y,\mu)\right] dy. \tag{12}$$

The parameters λ_1, λ_2 and μ can be estimated by solving the following saddle point problem (Kitamura and Stutzer, 1997)

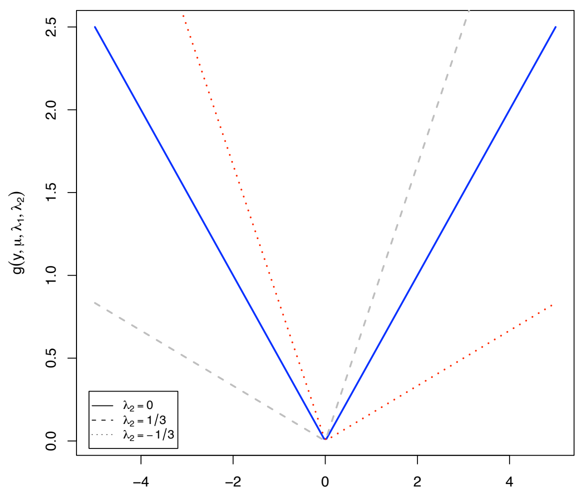
$$\hat{\mu}_{ME} = \arg\max_{\mu} \ln \int \exp\left[-\sum_{j=1}^{2} \hat{\lambda}_{j,ME} \phi_j(y,\mu)\right] dy,$$

where $\hat{\lambda}_{ME} = (\hat{\lambda}_{1,ME}, \hat{\lambda}_{2,ME})$ is given by

$$\hat{\lambda}_{ME}(\mu) = \arg\min_{\lambda} \ln \int \exp\left[-\sum_{j=1}^{2} \lambda_j \phi_j(y,\mu)\right] dy.$$

Solving the above saddle point problem is relatively easy since the profiled objective function has the exponential form. However, generally, c_1 and c_2 are not known or functions of parameters and Lagrange multipliers in a non-linear fashion. Moreover, in some cases, the closed form of c_1 and c_2 is not known. In order to deal with this problem, we simply consider the sample counterpart of the moments c_1 and c_2, say, $c_1 = (1/n)\sum_{i=1}^{n} |y_i - \mu|$ and $c_2 = (1/n)\sum_{i=1}^{n}(y_i - \mu)$. Then, it can be easily shown that the profiled objective function is simply the negative log-likelihood function of asymmetric Laplace density, i.e., $h(\lambda_1,\lambda_2,\mu) = -(1/n)L_{2,n}(\mu,\tau,\sigma)$ (see Ebrahimi, Soofi, and Soyer, 2008). In this case, $\hat{\mu}_{ME}$

Figure 1: Linear Combination of $|y-\mu|$ and $(y-\mu)$



and $\hat{\lambda}_{ME}$ satisfy the following first order conditions $\partial h/\partial\mu = 0$, $\partial h/\partial\lambda_2 = 0$ and $\partial h/\partial\lambda_1 = 0$, respectively:

$$-\frac{\lambda_1}{n}\sum_{i=1}^{n}\operatorname{sign}(y_i-\mu)-\lambda_2 = 0. \tag{13}$$

$$\frac{2\lambda_2}{(\lambda_1+\lambda_2)(\lambda_1-\lambda_2)}+\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu) = 0, \tag{14}$$

$$-\frac{1}{\lambda_1}\frac{\lambda_1^2+\lambda_2^2}{\lambda_1^2-\lambda_2^2}+\frac{1}{n}\sum_{i=1}^{n}|y_i-\mu| = 0, \tag{15}$$

Equations (13)-(15) are a re-parameterized version of (4)-(6). In fact, from a comparison of (2) and (11) we can easily see that $\lambda_1 = 1/(2\sigma)$, $\lambda_2 = (2\tau-1)/(2\sigma)$ and $\Omega(\theta) = \sigma/(\tau(1-\tau))$. Given λ_1 the degree of asymmetry is explained by λ_2 that is proportionally equal to $2\tau-1$ in ALPD. Note that $\lambda_2 = 0$ when $\tau = 0.5$, i.e., μ is the median. Thus finding the most appropriate degree of asymmetry is equivalent to estimating τ based on the ML method.

The role of the two moment constraints can be explained by the linear combination of two moment functions, $|y-\mu|$ and $(y-\mu)$. Figure 1 plots $g(y;\lambda_1,\lambda_2,\mu) = \lambda_1|y-\mu|+\lambda_2(y-\mu)$ with three different values of λ_2, $\lambda_1 = 1$, and $\mu = 0$. In general, $g(y;\lambda_1,\lambda_2,\mu)$ can be seen as a loss function. Clearly, this loss function is symmetric when $\lambda_2 = 0$. When $\lambda_2 = 1/3$, $g(\cdot)$ is tilted so that it puts more weight on the positive values in order to attain the maximum of the Shannon's entropy (and the reverse is true for $\lambda_2 = -1/3$). This naturally yields the asymmetric behavior of the resulting ME density.

2.3 Linear Regression Model

Now consider the conditional version of the above, by taking a linear model of the form $y = x'\beta + u$, where the parameter of interest is $\beta \in \mathbb{R}^p$, x refers to a p-vector of exogenous covariates, and u denotes the unobservable component in the linear model. As noted in Angrist, Chernozhukov, and Fernández-Val (2006), QR provides the best linear predictor for y under the asymmetric loss function

$$L_{3,n}(\beta;\tau) = \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta) = \sum_{i=1}^{n} \left((y_i - x_i'\beta)\left(\tau - 1(y_i \leq x_i'\beta)\right)\right), \tag{16}$$

where β is assumed to be a function of the *fixed* quantile τ of the unobservable components, that is $\beta(\tau)$. If u is assumed to follow an ALPD, the log-likelihood function is

$$\begin{aligned} L_{4,n}(\beta,\tau,\sigma) &= n\ln\left(\frac{1}{\sigma}\tau(1-\tau)\right) - \sum_{i=1}^{n}\left(\frac{1}{\sigma}\rho_\tau(y_i - x_i'\beta)\right) \\ &= n\ln\left(\frac{1}{\sigma}\tau(1-\tau)\right) - \frac{1}{\sigma}L_{3,n}(\beta;\tau). \end{aligned} \tag{17}$$

Estimating β in this framework provides the marginal effect of x on the τ-quantile of the conditional quantile function of y.

Computationally, the MLE can be obtained by simulating a grid of quantiles and choosing the quantile that maximizes (17), or by solving the estimating equations, $\nabla L_{4,n}(\beta,\tau,\sigma) = 0$, i.e.,

$$\frac{\partial L_{4,n}(\beta,\tau,\sigma)}{\partial\beta} = \sum_{i=1}^{n}\frac{1}{\sigma}\left(\frac{1}{2}\text{sign}(y_i - x_i'\beta) + \tau - \frac{1}{2}\right)x_i = 0, \tag{18}$$

$$\frac{\partial L_{4,n}(\beta,\tau,\sigma)}{\partial\tau} = \sum_{i=1}^{n}\left(\frac{1-2\tau}{\tau(1-\tau)} - \frac{(y_i - x_i'\beta)}{\sigma}\right) = 0, \tag{19}$$

$$\frac{\partial L_{4,n}(\beta,\tau,\sigma)}{\partial\sigma} = \sum_{i=1}^{n}\left(-\frac{1}{\sigma} + \frac{1}{\sigma^2}\rho_\tau(y_i - x_i'\beta)\right) = 0. \tag{20}$$

As we stated before, $L_{4,n}$ can be written as a penalized QR problem loss function that depends only on (β,τ):

$$-\frac{1}{n}L_{4,n}(\beta,\tau) = \ln\left(\frac{1}{n}L_{3,n}(\beta;\tau)\right) - \ln\left(\tau(1-\tau)\right), \tag{21}$$

and the interpretation is the same as discussed in section 2.1.

3. A Z-ESTIMATOR FOR QUANTILE REGRESSION

In this section we propose a Z-estimator based on the score functions from equations (18)-(20). Thus, although the original motivation for using the estimating equations is based on the ALPD, the final estimator is independent of this requirement. Let $\|\cdot\|$ be the Euclidean norm and $\theta = (\beta, \tau, \sigma)'$. Moreover, define the estimating functions

$$\psi_\theta(y,x) = \begin{pmatrix} \psi_{1\theta}(y,x) \\ \psi_{2\theta}(y,x) \\ \psi_{3\theta}(y,x) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma}\left(\tau - 1(y < x'\beta)\right)x \\ \frac{1-2\tau}{\tau(1-\tau)} - \frac{(y-x'\beta)}{\sigma} \\ -\frac{1}{\sigma} + \frac{1}{\sigma^2}\rho_\tau(y - x'\beta) \end{pmatrix},$$

and the estimating equations

$$\Psi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \frac{1}{\sigma}\left(\tau - 1(y_i < x_i'\beta)\right)x_i \\ \frac{1-2\tau}{\tau(1-\tau)} - \frac{(y_i-x_i'\beta)}{\sigma} \\ -\frac{1}{\sigma} + \frac{1}{\sigma^2}\rho_\tau(y_i - x_i'\beta) \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n}\psi_\theta(y_i, x_i) = 0.$$

A Z-estimator $\hat{\theta}_n$ is the approximate zero of the above data-dependent function that satisfies $\|\Psi_n(\hat{\theta}_n)\| \xrightarrow{p} 0$.

The implementation of the estimator is simple. As discussed in the previous section, an iteration algorithm can be used to solve for the estimates in the estimating equations above. Computationally, the estimates can be obtained by constructing a grid for quantiles τ and solving the QR problem as in (18) and (19) to find $\hat{\beta}(\tau)$ and $\hat{\sigma}(\tau)$. Finally, we estimate the quantile $\hat{\tau}$ that finds an approximate zero in (20). This algorithm is similar to the one proposed in Hinkley and Revankar (1977) and Yu and Zhang (2005) that compute the estimators for MLE under the ALPD. We find that the algorithm converges fast and is very precise.

In the proposed Z-estimator the interpretation of the parameter β is analogous to the interpretation of the location parameter in the QR literature. As in the least squares case, the scale parameter σ can be interpreted as the expected value of the loss function, which in the QR case corresponds to the expectation of the $\rho_\tau(.)$ function. Finally, τ captures a measure of asymmetry of the underline distribution of $y|x$ and also is associated with the most probable quantile. In the Appendix A we discuss the interpretation of these parameters in more detail.

We introduce the following assumptions to derive the asymptotic properties.

Assumption 1. Let $y_i = x_i'\beta_0 + u_i$, $i = 1, 2, ..., n$, where (y_i, x_i) is independent and identically distributed (i.i.d.), and x_i is independent of $u_i, \forall i$.

Assumption 2. The conditional distribution function of y, $G(y|x)$, is absolutely continuous with conditional densities, $g(y|x)$, with $0 < g(\cdot|\cdot) < \infty$.

Assumption 3. Let Θ be a compact set, with $\theta = (\beta, \tau, \sigma)' \in \Theta$, where $\beta \in \mathcal{B} \subset \mathbb{R}^p$, $\tau \in \mathcal{T} \subset (0, 1)$, and $\sigma \in \mathcal{S} \subset \mathbb{R}^+$, and θ_0 is an interior point of Θ;

Assumption 4. $E\|x\|^{2+\epsilon} < \infty$, and $E\|y\|^{2+\epsilon} < \infty$ for some $\epsilon > 0$.

Assumption 5. (i) Define $\Psi(\theta) = E\left[\psi_\theta(y, x)\right]$. Assume that $\Psi(\theta_0) = 0$ for a unique $\theta_0 \in \Theta$. (ii) Define $\Psi_n(\theta) = \mathbb{E}_n[\psi_\theta(y, x)] = \frac{1}{n}\sum_{i=1}^{n} \psi_\theta(y_i, x_i)$. Assume that $\|\Psi_n(\hat{\theta}_n)\| = o_p(n^{-1/2})$.

Assumption A1 considers the usual linear model and imposes i.i.d. to facilitate the proofs. Assumption A2 is common in the QR literature and restricts the conditional distribution of the dependent variable. Assumption A3 is standard in asymptotic theory and imposes compactness of the parameter space, and A4 is important to guarantee the asymptotic behavior of the estimator. The first part of A4 is usual in QR literature and second part in least squares literature. Finally Assumption A5 imposes an identifiability condition and ensure that the solution to the estimating equations is "nearly-zero", and it deserves further discussion.

The first part of A5 imposes a unique solution condition. Similar restrictions are frequently used in the QR literature to satisfy $E[\psi_{1\theta}(y, x)] = 0$ for a unique β and any given τ. This condition also appears in the M and Z estimators literatures. Uniqueness in QR is a very delicate subject and is actually imposed. For instance, Chernozhukov, Fernández-Val, and Melly (2009, p. 49) propose an approximate Z-estimator for QR process and assume that the true parameter $\beta_0(\tau)$ solves $E[(\tau - 1\{y \leq X'\beta_0(\tau)\})X] = 0$. Angrist, Chernozhukov, and Fernández-Val (2006) impose a uniqueness assumption of the form: $\beta(\tau) = \arg\min_\beta E[\rho_\tau(y - x'\beta)]$ is unique (see for instance their Theorems 1 and 2). See also He and Shao (2000) and Schennach (2008) for related discussion.

It is possible to impose more primitive conditions to ensure uniqueness. These conditions are explored and discussed in Theorem 2.1 in Koenker (2005, p. 36). If the y's have a bounded density with respect to Lebesgue measure then the observations (y, x) will be in general position with probability one and a solution exists.[1] However, uniqueness cannot be ensured if the covariates are discrete (e.g. dummy variables). If the x's have a component that have a density with respect to a Lebesgue measure, then multiple optimal solutions occur with probability zero and the solution is unique. However, these conditions are not very attractive, and uniqueness is in general imposed as an assumption.

Note that the usual assumptions of uniqueness in QR described above for $E[\psi_{1\theta}(y, x)] = 0$ guarantee that β is unique for any given τ. Combining these assumptions and bounded moments we guarantee uniqueness for $E[\psi_{3\theta}(y, x)] = 0$, because $\sigma = E[\rho_\tau(y - x\beta)]$ such that for each τ, σ is unique. With respect to the second equation $E[\psi_{2\theta}(y, x)] = 0$, it is satisfied if

$$\frac{1 - 2\tau}{\tau(1 - \tau)} = \frac{E[y - x'\beta]}{\sigma}.$$

[1] See definition 2.1 in Koenker (2005) for a definition of general position.

Therefore, for unique β and σ, the right hand side of the above equation is unique. Since $\frac{1-2\tau}{\tau(1-\tau)}$ is a continuous and strictly decreasing, τ is also unique.[2]

The second part of A5 is used to ensure that the solution to the approximated working estimating equations is close to zero. The solution for the estimating equations, $\Psi_n(\hat{\theta}_n) = 0$, does not hold in general. In most cases, this condition is actually equal to zero, but least absolute deviation of linear regression is one important exception. The indicator function in the first estimating equations determines that it may not have an exact zero. It is common in the literature to work with M and Z estimators $\hat{\theta}_n$ of θ_0 that satisfy $\sum_{i=1}^n \psi(x_i, \hat{\theta}_n) = o_p(\delta_n)$, for some sequence δ_n. For example, Huber (1967) considered $\delta_n = \sqrt{n}$ for asymptotic normality, and Hinkley and Revankar (1977) verified the condition for the unconditional asymmetric double exponential case. This condition also appears in the quantile regression literature, see for instance He and Shao (1996) and Wei and Carroll (2009). In addition, in the approximate Z-estimator for quantile process in Chernozhukov, Fernández-Val, and Melly (2009), they have that the empirical moment functions $\hat{\Psi}(\theta, u) = E_n[g(W_i, \theta, u)]$, for each $u \in T$, the estimator $\hat{\theta}(u)$ satisfies $\|\hat{\Psi}(\hat{\theta}(u), u)\| \leq \inf_{\theta \in \Theta} \|\hat{\Psi}(\theta, u)\| + \epsilon_n$ where $\epsilon_n = o(n^{-1/2})$. For the quantile regression case, Koenker (2005, p. 36) comments that the absence of a zero to the problem $\Psi_{1n}(\hat{\beta}_n(\tau)) = 0$, where $\hat{\beta}_n(\tau)$ is the quantile regression optimal solution for a given τ and σ, "is unusual, unless the y_i's are discrete." Here we follow the standard conditions for M and Z estimators and impose A5(ii). For a more general discussion about this condition on M and Z estimators see e.g. Kosorok (2008, pp. 399-407).

Now we move our attention to the asymptotic properties of the estimator.

Theorem 1 *Under Assumptions A1-A5,* $\|\hat{\theta}_n - \theta_0\| \xrightarrow{p} 0$.

Proof: In order to show consistency we check the conditions of Theorem 5.9 in van der Vaart (1998). Define $\mathcal{F} \equiv \{\psi_\theta(y, x), \theta \in \Theta\}$, and recall that $\Psi_n(\theta) = \frac{1}{n}\sum_{i=1}^n \psi_\theta(y, x)$ and $\Psi(\theta) = E[\psi_\theta(y, x)]$. First note that, under conditions A3 and A5, the function $\Psi(\theta)$ satisfies,

$$\inf_{\theta : d(\theta, \theta_0) \geq \epsilon} \|\Psi(\theta)\| > 0 = \|\Psi(\theta_0)\|,$$

because for a compact set Θ and a continuous function Ψ, uniqueness of θ_0 as a zero implies this condition (see van der Vaart, 1998, p.46).

Now we need to show that $\sup_{\theta \in \Theta} \|\Psi_n(\theta) - \Psi(\theta)\| \xrightarrow{p} 0$. By Lemma A1 in the Appendix B we know that the class $\mathcal{F}$ is Donsker. Donskerness implies a uniform law of large numbers such that

$$\sup_{\theta \in \Theta} |\mathbb{E}_n[\psi_\theta(y, x)] - E[\psi_\theta(y, x)]| \xrightarrow{p} 0,$$

[2]Note that $m(\tau) \equiv \frac{1-2\tau}{\tau(1-\tau)}$ is a continuous function, has a unique zero at $\tau = 1/2$ and $m(\tau) > 0$ for $\tau < 1/2$, $m(\tau) < 0$ for $\tau > 1/2$. As $\tau \to 0$, $m(\tau) \to +\infty$, and as $\tau \to 1$, $m(\tau) \to -\infty$. Finally, $\frac{dm(\tau)}{d\tau} = \frac{-2\tau(1-\tau)-(1-2\tau)^2}{\tau^2(1-\tau)^2} = \frac{-2\tau+2\tau^2-1+4\tau-4\tau^2}{\tau^2(1-\tau)^2} = \frac{-1+2\tau-2\tau^2}{\tau^2(1-\tau)^2} = \frac{-1+2\tau(1-\tau)}{\tau^2(1-\tau)^2} < 0$ for any $\tau \in (0, 1)$.

where $f \mapsto \mathbb{E}_n[f(w)] = \frac{1}{n}\sum_{i=1}^n f(w_i)$. Hence we have $\sup_{\theta\in\Theta} \|\Psi_n(\theta) - \Psi(\theta)\| \xrightarrow{p} 0$.

Finally, from assumptions A1-A5 the problem has a unique root and also we have $\|\Psi_n(\hat{\theta}_n)\| \xrightarrow{p} 0$. Thus, all the conditions in Theorem 5.9 of van der Vaart (1998) are satisfied and $\|\hat{\theta}_n - \theta_0\| \xrightarrow{p} 0$. ■

After showing consistency we move our attention to the asymptotic normality of the estimator. In order to derive the limiting distribution define

$$V_{1\theta} = E\left[\psi_\theta(y,x)\psi_\theta(y,x)'\right], \tag{22}$$

$$V_{2\theta} = \frac{\partial E\left[\psi_\theta(y,x)\right]}{\partial\theta'}. \tag{23}$$

Here,

$$V_{1\theta} = \begin{bmatrix} a & b & c \\ . & d & e \\ . & . & f \end{bmatrix},$$

where

$$a = \frac{1}{\sigma}\tau(1-\tau)E[xx'],$$

$$b = \frac{E\left[\left((1-2\tau)sign(y-x'\beta) - (1-2\tau)^2\right)x'\right]}{2\sigma\tau(1-\tau)} - E\left[\frac{1}{\sigma^2}\rho_\tau(y-x'\beta)x'\right],$$

$$c = \frac{1}{2\sigma^3}E[\rho_\tau(y-x'\beta)\ (sign(y-x'\beta) - (1-2\tau))x'],$$

$$d = \frac{(1-2\tau)^2}{\tau^2(1-\tau)^2} + E[\frac{1}{\sigma^2}(y-x'\beta)^2] - 2\frac{(1-2\tau)}{\tau(1-\tau)}E[\frac{1}{\sigma}(y-x'\beta)],$$

$$e = \frac{1}{\sigma^2}E[\rho_\tau(y-x'\beta)\ (\frac{(1-2\tau)}{\tau(1-\tau)} - \frac{1}{\sigma}(y-x'\beta))],$$

$$f = \frac{1}{\sigma^4}E[\rho_\tau^2(y-x'\beta)] + \frac{1}{\sigma^2} - \frac{1}{\sigma^3}E[\rho_\tau(y-x'\beta)];$$

and

$$V_{2\theta} = \begin{bmatrix} -\frac{E[g(x'\beta|x)xx']}{\sigma} & \frac{1}{\sigma}E[x] & 0 \\ . & \frac{-1+2\tau-2\tau^2}{\tau^2(1-\tau)^2} & \frac{1}{\sigma^2}E[(y-x'\beta)] \\ . & . & -\frac{1}{\sigma^2} \end{bmatrix}.$$

Note that when $y|x \sim ALPD(x'\beta,\tau,\sigma)$, then $V_{1\theta} = V_{2\theta}$.

Assumption 6. Assume that $V_{1\theta_0}$ and $V_{2\theta_0}$ exist and are finite, and $V_{2\theta_0}$ is invertible.

Chernozhukov, Fernández-Val, and Melly (2009) calculated equations (22) and (23) in the quantile process as an approximate Z-estimator.

Now we state the asymptotic normality result.

Theorem 2 *Under Assumptions 1-6,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow N(0, V_{2\theta_0}^{-1} V_{1\theta_0} V_{2\theta_0}^{-1}).$$

Proof: First, combining Theorem 1 and second part of Lemma A1, we have

$$\mathbb{G}_n \psi_{\hat{\theta}_n}(y, x) = \mathbb{G}_n \psi_{\theta_0}(y, x) + o_p(1),$$

where $f \mapsto \mathbb{G}_n[f(w)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(w_i) - Ef(w_i))$. Rewriting we have

$$\sqrt{n}\mathbb{E}_n \psi_{\hat{\theta}_n}(y, x) = \sqrt{n} E\psi_{\hat{\theta}_n}(y, x) + \mathbb{G}_n \psi_{\theta_0}(y, x) + o_p(1). \tag{24}$$

By assumption A5

$$\left\|\mathbb{E}_n \psi_{\hat{\theta}_n}(y, x)\right\| = o_p(n^{-1/2}) \quad \text{and} \quad E[\psi_{\theta_0}(y, x)] = 0.$$

Now consider the first element of the right hand side of (24). By a Taylor expansion about $\hat{\theta}_n = \theta_0$ we obtain

$$E[\psi_{\hat{\theta}_n}(y, x)] = E[\psi_{\theta_0}(y, x)] + \frac{\partial E[\psi_\theta(y, x)]}{\partial \theta'}\Big|_{\theta=\theta_0} (\hat{\theta}_n - \theta_0) + o_p(1), \tag{25}$$

where

$$\frac{\partial E[\psi_\theta(y, x)]}{\partial \theta'}\Big|_{\theta=\theta_0} = \frac{\partial}{\partial \theta'} E \left(\begin{array}{c} \frac{1}{\sigma}\left(\tau - 1(y < x'\beta)\right) x \\ \left(\frac{1-2\tau}{\tau(1-\tau)} - \frac{(y - x'\beta)}{\sigma}\right) \\ -\frac{1}{\sigma} + \frac{1}{\sigma^2}\rho_\tau(y - x'\beta) \end{array} \right) \Bigg|_{\theta=\theta_0}.$$

Since by condition A6, $\frac{\partial E[\psi_\theta(y,x)]}{\partial \theta'}\Big|_{\theta=\theta_0} = V_{2\theta_0}$, equation (25) can be rewritten as

$$E[\psi_\theta(y, x)]\Big|_{\theta=\hat{\theta}_n} = V_{2\theta_0}(\hat{\theta}_n - \theta_0) + o_p(1). \tag{26}$$

Using Assumption A5 (ii), from (24) we have

$$o_p(1) = \sqrt{n} E\psi_{\hat{\theta}_n}(y, x) + \mathbb{G}_n \psi_{\theta_0}(y, x) + o_p(1),$$

and using the above approximation given in (26)

$$o_p(1) = V_{2\theta_0}\sqrt{n}(\hat{\theta}_n - \theta_0) + \mathbb{G}_n \psi_{\theta_0}(y, x) + o_p(1).$$

By invertibility of $V_{2\theta_0}$ in A6,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V_{2\theta_0}^{-1} \mathbb{G}_n \psi_{\theta_0}(y, x) + o_p(1). \tag{27}$$

Finally, from Lemma A1 $\theta \mapsto \mathbb{G}_n \psi_\theta(y, x)$ is stochastic equicontinuous. So, stochastic equicontinuity and ordinary CLT imply that $\mathbb{G}_n \psi_\theta(y, x) \Rightarrow z(\cdot)$ converges to a Gaussian process with

variance-covariance function defined by
$V_{1\theta_0} = E\left[\psi_\theta(y,x)\psi_\theta(y,x)'\right]\Big|_{\theta=\theta_0}$. Therefore, from (27)

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow V_{2\theta_0}^{-1} z(\cdot),$$

so that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow N(0, V_{2\theta_0}^{-1} V_{1\theta_0} V_{2\theta_0}^{-1}).$$

■

4. MONTE CARLO SIMULATIONS

In this section we provide a glimpse into the finite sample behavior of the proposed ZQR estimator. Two simple versions of our basic model are considered in the simulation experiments. In the first, reported in Table 1, the scalar covariate, x_i, exerts a pure location shift effect. In the second, reported in Table 2, x_i has both a location and scale shift effects. In the former case the response, y_i, is generated by the model, $y_i = \alpha + \beta x_i + u_i$, while in the latter case, $y_i = \alpha + \beta x_i + (1+\gamma)u_i$, where u_i are i.i.d. innovations generated according to a standard normal distribution, t_3 distribution, χ_3^2 centered at the mean, Laplace distribution (i.e. $\tau = 0.5$), and ALPD with $\tau = 0.25$.[3] In the location shift model x_i follows a standard normal distribution; in the location-scale shift model, it follows a χ_3^2. We set $\alpha = \beta = 1$ and $\gamma = 0.5$. Our interest is on the effect of the covariates in terms of bias and root mean squared error (RMSE). We carry out all the experiments with sample size $n = 200$ and 5,000 replications. Three estimators are considered: our proposed ZQR estimator, quantile regression at the median (QR), and ordinary least squares (OLS). We pay special attention to the estimated quantile $\hat{\tau}$ in the ZQR.

Table 1 reports the results for the location shift model. In all cases we compute the bias and RMSE with respect to $\beta = 1$. Bias is close to zero in all cases. In the Gaussian setting, as expected, we observe efficiency loss in ZQR and QR estimates compared to that of OLS. Under symmetric distributions, normal, t_3, and Laplace, the estimated quantile of interest $\hat{\tau}$ in the ZQR is remarkably close to 0.5. In the χ_3^2 case, the ZQR estimator performs better than the QR and OLS procedures. Note that the estimated quantile for the χ_3^2 is 0.081, consistent with the fact that the underline distribution is right skewed. Finally, for the ALPD(0.25) case, ZQR produces the estimated quantile ($\hat{\tau} = 0.248$) rightly close to 0.25, and also has a smaller RMSE. Overall, Table 1 shows that the ZQR estimator retains the robustness properties of the QR estimator, although we do not specify a particular quantile of interest.

In the location-scale version of the model we adopt the same distributions for generating the data. For this case the effect of the covariate x_i on quantile of interest response in QR is given by $\beta(\tau) = \beta + \gamma Q_u(\tau)$. In ZQR we compute bias and RMSE by averaging estimated τ from 5,000 replications. The results are summarized in Table 2. The results for the normal,

[3]Although not reported, similar results were obtained for ALPD with $\tau = 0.75$.

Table 1: Location-Shift Model: Bias and RMSE

		ZQR	QR (Median)	OLS
$N(0,1)$	Bias	0.0007	-0.0004	0.0008
	RMSE	0.0904	0.0899	0.0710
	$\hat{\tau}$	0.501	–	–
t_3	Bias	0.0012	-0.0008	0.0014
	RMSE	0.1133	0.0967	0.1217
	$\hat{\tau}$	0.498	–	–
χ^2_3	Bias	-0.0021	0.0024	0.0020
	RMSE	0.1419	0.1892	0.1801
	$\hat{\tau}$	0.081	–	–
ALPD	Bias	0.0001	0.0001	0.0001
($\tau = 0.5$)	RMSE	0.0638	0.0549	0.0710
	$\hat{\tau}$	0.499	–	–
ALPD	Bias	-0.0008	-0.0001	0.0003
($\tau = 0.25$)	RMSE	0.0718	0.0860	0.0917
	$\hat{\tau}$	0.248	–	–

t_3 and Laplace distributions are similar to those in the location model, showing that all point estimates are approximately unbiased. As expected, OLS outperforms ZQR and QR in the normal case, but the opposite occurs in the t_3 and Laplace distributions. In the χ^2_3 case, the estimated quantile is $\hat{\tau} = 0.086$. For the ALPD(0.25) distribution, the best performance is obtained for the ZQR estimator.

5. EMPIRICAL ILLUSTRATION: THE EFFECT OF JOB TRAINING ON WAGES

The effect of policy variables on distributional outcomes are of fundamental interest in empirical economics. Of particular interest is the estimation of the quantile treatment effects (QTE), that is, the effect of some policy variable of interest on the different quantiles of a conditional response variable. Our proposed estimator complements the QTE analysis by providing a parsimonious estimator at the most probable quantile value.

We apply the estimator to the study of the effect of public-sponsored training programs. As argued in LaLonde (1995), public programs of training and employment are designed to improve participant's productive skills, which in turn would affect their earnings and dependency on social welfare benefits. We use the Job Training Partnership Act (JTPA), a public training program that has been extensively studied in the literature. For example, see Bloom, Orr, Bell, Cave, Doolittle, Lin, and Bos (1997) for a description, and Abadie, Angrist, and Imbens (2002) for QTE analysis. The JTPA was a large publicly-funded training program that began funding in October 1983 and continued until late 1990's. We focus on the Title II subprogram, which was offered only to individuals with "barriers to em-

Table 2: Location-Scale-Shift Model: Bias and RMSE

		ZQR	QR (Median)	OLS
$N(0,1)$	Bias	0.0015	0.0036	0.0037
	RMSE	0.2209	0.1461	0.1365
	$\hat{\tau}$	0.499	–	–
t_3	Bias	−0.0005	0.0002	−0.0052
	RMSE	0.2457	0.1460	0.2565
	$\hat{\tau}$	0.501	–	–
χ_3^2	Bias	−0.0004	0.0076	0.0089
	RMSE	0.5087	0.2833	0.3788
	$\hat{\tau}$	0.086	–	–
ALPD ($\tau = 0.5$)	Bias	−0.0010	−0.0001	−0.0013
	RMSE	0.1455	0.0845	0.1459
	$\hat{\tau}$	0.501	–	–
ALPD ($\tau = 0.25$)	Bias	0.0051	0.0004	0.4076
	RMSE	0.1331	0.1429	0.4505
	$\hat{\tau}$	0.248	–	–

ployment" (long-term use of welfare, being a high-school drop-out, 15 or more recent weeks of unemployment, limited English proficiency, phsysical or mental disability, reading proficiency below 7th grade level or an arrest record). Individuals in the randomly assigned JTPA treatment group were offered training, while those in the control group were excluded for a period of 18 months. Our interest lies in measuring the effect of a training offer and actual training on of participants' future earnings.

We use the database in Abadie, Angrist, and Imbens (2002) that contains information about adult male and female JTPA participants and non-participants. Let z denote the indicator variable for those receiving a JTPA offer. Of those offered, 60% did training; of those in the control group, less than 2% did training. For our purposes of illustrating the use of ZQR, we first study the effect of receiving a JTPA offer on log wages, and later we pursue instrumental variables estimation in the ZQR context. Following Abadie, Angrist, and Imbens (2002) we use a linear regression specification model, where the JTPA offer enters in the equation as a dummy variable.[4] We consider the following regression model:

$$y = z\gamma + x\beta + u,$$

where the dependent variable y is the logarithm of 30 month accumulated earnings (we exclude individuals without earnings), z is a dummy variable for the JTPA offer, x is a set of exogenous covariates contaning individual characteristics, and u is an unobservable component. The parameter of interest is γ that provides the effect of the JTPA training offer on wages.

[4]Linear regression models are common in the QTE literature to accomodate several control variables capturing individual characteristics. See for instance Chernozhukov and Hansen (2006, 2008) and Firpo (2007).

Table 3: JTPA offer

	ZQR [$\hat{\tau} = 0.84$]		OLS		Median regression	
Intercept	9.894	(0.059)	8.814	(0.088)	9.188	(0.086)
JTPA offer	0.045	(0.022)	0.075	(0.032)	0.100	(0.033)
FEMALE	0.301	(0.023)	0.259	(0.030)	0.260	(0.031)
HSORGED	0.201	(0.025)	0.267	(0.034)	0.297	(0.037)
BLACK	-0.102	(0.026)	-0.121	(0.036)	-0.175	(0.039)
HISPANIC	-0.032	(0.034)	-0.034	(0.050)	-0.025	(0.051)
MARRIED	0.129	(0.025)	0.242	(0.036)	0.265	(0.034)
WKLESS13	-0.255	(0.023)	-0.598	(0.032)	-0.556	(0.036)
AGE2225	0.229	(0.057)	0.175	(0.084)	0.125	(0.080)
AGE2629	0.285	(0.058)	0.192	(0.085)	0.131	(0.081)
AGE3035	0.298	(0.057)	0.191	(0.084)	0.176	(0.080)
AGE3644	0.320	(0.058)	0.130	(0.085)	0.173	(0.081)
AGE4554	0.267	(0.064)	0.110	(0.094)	0.080	(0.092)
$\hat{\tau}$	0.840	(0.051)			0.500	
$\hat{\sigma}$	0.249	(0.060)			0.538	(0.006)

Notes: 9872 observations. The numbers in parenthesis are the corresponding standard errors.

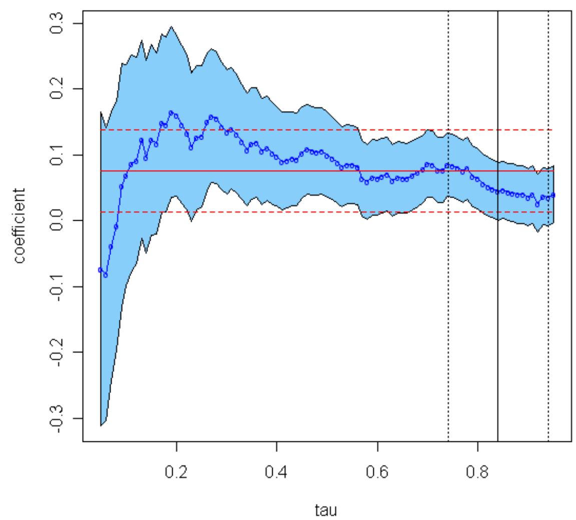
First, we compute the QR process for all $\tau \in (0.05, 0.95)$ and the results are presented in Figure 2. The JTPA effect estimates for QR and OLS appear in Table 3, where the variables are defined as follows:

JTPA offer: dummy variable for individuals that received a JTPA offer;
FEMALE: Female dummy variable;
HSORGED: dummy variable for individuals with completed high school or GSE;
BLACK: race dummy variable; HISPANIC: dummy variable for hispanic;
MARRIED: dummy variable for married individuals;
WKLESS13: dummy variable for individuals working less than 13 weeks in the past year;
AGE2225, AGE2629, AGE3035, AGE3644 and AGE4554 age range indicator variables.

Interestingly, with exception of low quantiles, the effect of JTPA is decreasing in τ, which implies that those individuals in the high quantiles of the conditional wage distribution benefited less from the JTPA training. Second, by solving equation (19) we obtain that the most probable quantile $\hat{\tau} = 0.84$. This is further illustrated in Figure 3, and this means that the distribution of unobservables is negatively skewed. This value is denoted by a vertical solid line, together with the 95% confidence interval given by the vertical parallel dotted lines. From Table 3 we observe that the training effect estimate from mean and median regressions are, respectively, 0.075 (0.032) and 0.100 (0.033) which are similar, however they both are larger than the ZQR estimate of 0.045 (0.022).[5] Figure 2 shows that QR estimates in the upper tail of the distribution have smaller standard errors, which suggests that by choosing the most likely quantile the ZQR procedure implicitly solves for the smallest standard error

[5]The numbers in parenthesis are the corresponding standard errors.

Figure 2: JTPA offer: Quantile regression process and OLS



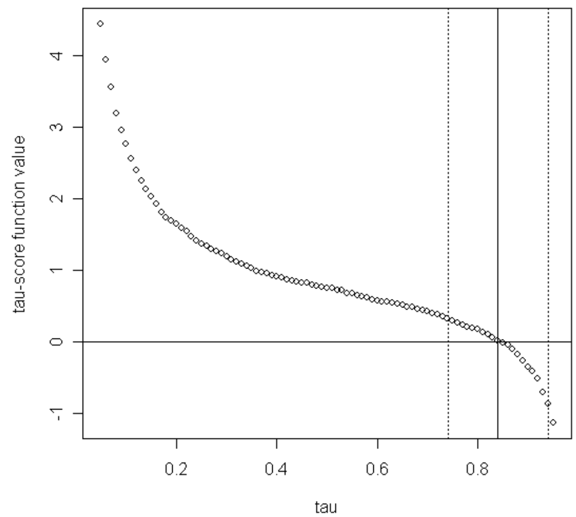
Notes: Quantile regression process (shaded area), OLS (horizonal lines) and estimated most informative quantile (vertical lines) with 95% confidence intervals.

QR estimator. The results show that for the most probable quantile, $\hat{\tau} = 0.84$ (0.051), the effect of training is different from the mean and median effects. From a policy maker perspective, if one is asked to report the effect of training on wage, it could be done through the mean effect (0.075), the median effect (0.100) or even the entire conditional quantile function as in Figure 2; our analysis recommends reporting the most likely effect (0.045) coming from the most probable quantile $\hat{\tau} = 0.84$. Using the above model, the fit of the data reveals that the upper quantiles are informative, and the ZQR estimator is appropriate to describe the effect of JTPA on earnings.

As argued in the Introduction, the ZQR framework allows for a different interpretation of the QR analysis. Suppose that we are interested in a targeted treatment effect of $\bar{\gamma} = 0.1$, and we would like to get the representative quantile of the unobservables distribution that will most likely have this effect. This corresponds to estimating the ZQR parameters for $y - z\bar{\gamma} = x\beta + u$. In this case, we obtain an estimated most likely quantile of $\widehat{\tau(\bar{\gamma})} = 0.85$.

To value the option of treatment is an interesting exercise in itself, but policy makers may be more interested in the effect of actual training rather than the possibility of training. In this case the model of interest is $y = d\alpha + x\beta + u$, where d is a dummy variable indicating if the individual actually completed the JTPA training. We have strong reasons to believe that $cov(d, u) \neq 0$ and therefore OLS and QR estimates will be biased. In this case, while the JTPA offer is random, those individuals who decide to undertake training do not constitute a random sample of the population. Rather, they are likely to be more motivated individuals

Figure 3: JTPA offer: τ-score function



Notes: The τ-score function is $\frac{1-2\tau}{\tau(1-\tau)} - \frac{\sum_{i=1}^{n}(y_i - x_i'\hat{\beta}(\tau))}{n\hat{\sigma}}$.

or those that value training the most. However, the exact nature of this bias is unknown in terms of quantiles. Figure 4 reports the entire quantile process and OLS for the above equation. Interestingly the effect of training on wages is monotonically decreasing in τ. The selection of the most likely quantile determines that as in the previous case $\hat{\tau} = 0.84$.

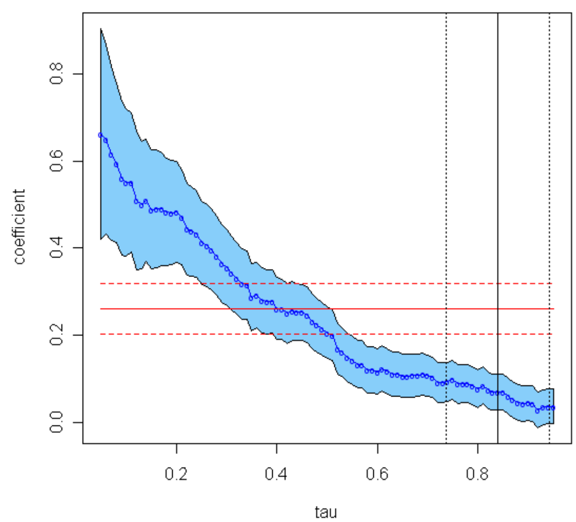
In order to solve for the potential endogeneity, and following Abadie, Angrist, and Imbens (2002), z can be used as a valid instrument for d. The reason is that it is exogenous as it was a randomized experiment, and it is correlated with d (as mentioned earlier 60% of individuals undertook training when they were offered). The IV strategy is based on Chernozhukov and Hansen (2006, 2008) by considering the model $y - d\alpha = x\beta + z\gamma + u$.

The IV method in QR proceeds as follows. Note that z does not belong to the model, as conditional on d, undertaking training, the offer has no effect on wages. Then, we construct a grid in $\alpha \in \mathcal{A}$, which is indexed by j for each $\tau \in (0,1)$ and we estimate the quantile regression model for fixed τ, $y - d\alpha_j(\tau) = x\beta + z\gamma + u$. This gives $\{\hat{\beta}_j(\alpha_j(\tau),\tau), \hat{\gamma}_j(\alpha_j(\tau),\tau)\}$, the set of conditional quantile regression estimates for the new model. Next, we choose α by minimizing a given norm of γ (we use the Euclidean norm), $\hat{\hat{\alpha}}(\tau) = argmin_{\alpha\in\mathcal{A}} \|\hat{\gamma}(\alpha(\tau),\tau)\|$.

Figure 5 shows the values of γ^2 for the grids of α and τ. As a result we obtain the map $\tau \mapsto \{\hat{\hat{\alpha}}(\tau), \hat{\beta}(\hat{\hat{\alpha}}(\tau),\tau) \equiv \hat{\hat{\beta}}(\tau), \hat{\gamma}(\hat{\hat{\alpha}}(\tau),\tau) \equiv \hat{\hat{\gamma}}(\tau)\}$.

Finally, we select the most probably quantile as in the previous case, by using the first

Figure 4: JTPA: Quantile regression process and OLS



Notes: Quantile regression process (shaded area), OLS (horizonal lines) and estimated most informative quantile (vertical lines) with 95% confidence intervals.

Figure 5: JTPA: Minimization of $\|\gamma^2(\tau, \alpha)\|$

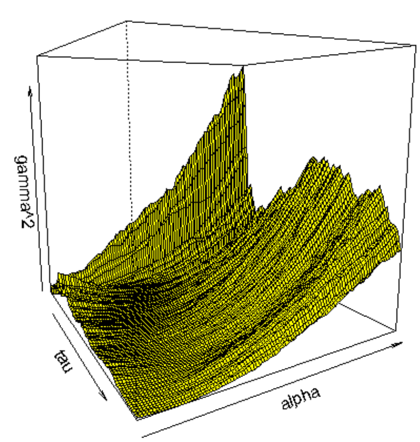
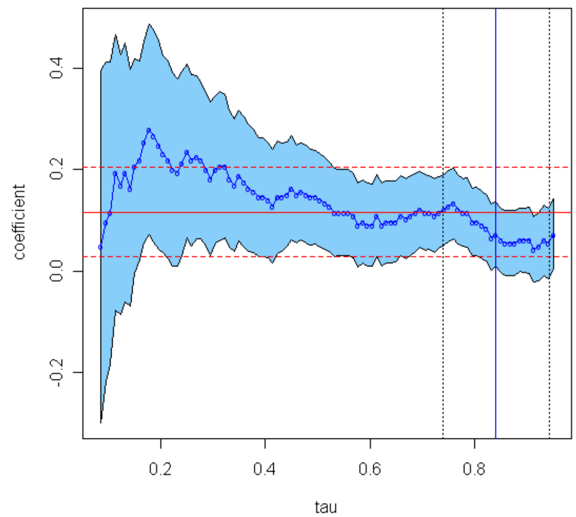


Figure 6: JTPA: IV Quantile regression process and IV OLS



Notes: Quantile regression process (shaded area), OLS (horizonal lines) and estimated most informative quantile (vertical lines) with 95% confidence intervals.

order condition corresponding the selection of τ:

$$\hat{\hat{\tau}} = argmin_{\tau\in(0,1)} \left| \frac{1-2\tau}{\tau(1-\tau)} - \frac{\sum_{i=1}^{n} \hat{\hat{u}}_i(\tau)}{\sum_{i=1}^{n} \rho_\tau \left(\hat{\hat{u}}_i(\tau)\right)} \right|$$

where $\hat{\hat{u}}_i(\tau) = y_i - d_i\hat{\hat{\alpha}}(\tau) - x_i'\hat{\hat{\beta}}(\tau) - z_i\hat{\hat{\gamma}}(\tau)$. Figure 6 reports the IV estimates together with the most likely quantile. Interestingly, the qualitative results are very much alike those of the value of the JTPA training offer. The IV least squares estimator for the effect of JTPA training gives a value of 0.116 (0.045) while IV median regression gives a much higher value of 0.142 (0.047). The most likely quantile continues to be 0.84 (0.053), which has an associated training effect of 0.072 (0.033). The ZQR effect continues to be smaller than the mean and median estimates. Therefore, the upper quantiles are more informative when analyzing the effects of JTPA training on log wages.

6. CONCLUSIONS

In this paper we show that the maximum likelihood problem for the asymmetric Laplace distribution can be found as the solution of a maximum entropy problem where we impose moment constraints given by the joint consideration of the mean and the median. We also propose an approximate Z-estimator method, which provides a parsimonious estimator that complements the quantile process. This provides an alternative interpretation of quantile regression and frames it within the maximum entropy paradigm. Potential estimates from this method has important applications. As an illustration, we apply the proposed estimator to a well-known dataset where quantile regression has been extensively used.

Appendix

A. Interpretation of the Z-estimator

In order to interpret θ_0, we take the expectation of the estimating equations with respect to the unknown true density. To simplify the exposition we consider a simple model without covariates: $y_i = \alpha + u_i$. Our estimating equation vector is defined as:

$$E(\Psi_\theta(y)) = E\begin{pmatrix} \frac{1}{\sigma}\left(\tau - 1(y<\alpha)\right) \\ \left(\frac{1-2\tau}{\tau(1-\tau)} - \frac{(y-\alpha)}{\sigma}\right) \\ -\frac{1}{\sigma} + \frac{1}{\sigma^2}\rho_\tau(y-\alpha) \end{pmatrix} = 0,$$

and the estimator is such that

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_\theta(y_i) = 0$$

Let $F(y)$ be the cdf of the random variable y. Now we need to find $E[\Psi_\theta(y)]$.

For the first component we have

$$\begin{aligned}\frac{1}{\sigma}E[\tau - I(y<\alpha)] &= \frac{1}{\sigma}\left(\int_{\mathbb{R}}(\tau - 1(y<\alpha))dF(y)\right) \\ &= \frac{1}{\sigma}\left(\tau - \int_{-\infty}^{\alpha} dF(y)\right) \\ &= \frac{1}{\sigma}\left(\tau - F(\alpha)\right).\end{aligned}$$

Thus if we set this equal to zero, we have

$$\alpha = F^{-1}(\tau),$$

which is the usual quantile. Thus, the interpretation of the parameter α is analogous to QR if covariates are included.

For the third term in the vector, $-\frac{1}{\sigma} + \frac{1}{\sigma^2}\rho_\tau(y-\alpha)$, we have

$$E\left[-\frac{1}{\sigma} + \frac{1}{\sigma^2}\rho_\tau(y-\alpha)\right] = 0,$$

that is,

$$\sigma = E\left[\rho_\tau(y-\alpha)\right].$$

Thus, as in the least squares case, the scale parameter σ can be interpreted as the expected value of the loss function.

Finally, we can interpret τ using the second equation,

$$E\left[\frac{1-2\tau}{\tau(1-\tau)} - \frac{(y-\alpha)}{\sigma}\right] = 0,$$

which implies that

$$\frac{1-2\tau}{\tau(1-\tau)} = \frac{E[y] - F^{-1}(\tau)}{\sigma}.$$

Note that $g(\tau) \equiv \frac{1-2\tau}{\tau(1-\tau)}$ is a measure of the skewness of the distribution (see also footnote 2). Thus, τ should be chosen to set $g(\tau)$ equal to a measure of asymmetry of the underline distribution $F(\cdot)$ given by the difference of τ-quantile with the mean (and standardized by σ). In the special case of a symmetric distribution, the mean coincides with the median and mode, such that $E[y] = F^{-1}(1/2)$ and $\tau = 1/2$, which is the most probable quantile and a solution to our Z-estimator.

B. Lemma A1

In this appendix we state an auxiliary result that states Donskerness and stochastic equicontinuity. Let $\mathcal{F} \equiv \{\psi_\theta(y,x), \theta \in \Theta\}$, and define the following empirical process notation for $w = (y,x)$:

$$f \mapsto \mathbb{E}_n[f(w)] = \frac{1}{n}\sum_{i=1}^{n} f(w_i) \quad f \mapsto \mathbb{G}_n[f(w)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(w_i) - Ef(w_i)).$$

We follow the literature using empirical process exploiting the monotonicity and boundedness of the indicator function, the boundedness of the moments of x and y, and that the problem is a parametric one.

Lemma A1. Under Assumptions A1-A4 $\mathcal{F}$ is Donsker. Furthermore,

$$\theta \mapsto \mathbb{G}_n\psi_\theta(y,x)$$

is stochastically equicontinuous, that is

$$\sup_{\|\theta-\theta_0\|\le\delta_n} \left\|\mathbb{G}_n\psi_\theta(y,x) - \mathbb{G}_n\psi_{\theta_0}(y,x)\right\| = o_p(1),$$

for any $\delta_n \downarrow 0$.

Proof: Note that a class $\mathcal{F}$ of a vector-valued functions $f : x \mapsto \mathbb{R}^k$ is Donsker if each of the classes of coordinates $f_i : x \mapsto \mathbb{R}^k$ with $f = (f_1, ..., f_k)$ ranging over $\mathcal{F}(i = 1, 2, ..., k)$ is Donsker (van der Vaart, 1998, p.270).

The first element of the vector is $\psi_{1\theta}(y,x) = (\tau - 1(y_i < x_i'\beta))\frac{x_i}{\sigma}$. Note that the functional class $\mathfrak{A} = \{\tau - 1\{y_i < x_i'\beta\}, \tau \in \mathcal{T}, \beta \in \mathcal{B}\}$ is a VC subgraph class and hence also Donsker class, with envelope 2. Its product with x also forms a Donsker class with a square integrable envelope $2 \cdot \max_j |x_j|$, by Theorem 2.10.6 in van der Vaart and Wellner (1996) (VW henceforth). Finally, the class $\mathcal{F}_1$ is defined as the product of the latter with $1/\sigma$, which is bounded. Thus, by assumption A4 $\mathcal{F}_1$ is Donsker. Now define the process $h_1 = (\beta, \tau, \sigma) \mapsto \mathbb{G}_n\psi_{1\theta}(y,x)$. Using the established Donskerness property, this process is Donsker in $l^\infty(\mathcal{F}_1)$.

The second element of the vector is $\psi_{2\theta}(y,x) = \left(\frac{1-2\tau}{\tau(1-\tau)} - \frac{(y_i - x_i'\beta)}{\sigma}\right)$. Define

$$\mathfrak{H} = \{(y_i - x_i'\beta), \beta \in \mathcal{B}\}.$$

Note that

$$|(y_i - x_i'\beta_1) - (y_i - x_i'\beta_2)| = |x_i'(\beta_2 - \beta_1)| \leq \|x_i\|\|\beta_2 - \beta_1\|,$$

where the inequality follows from Cauchy-Schwartz inequality. Thus by Assumptions A3-A4 and Example 19.7 in van der Vaart (1998) the class $\mathfrak{H}$ is Donsker. Moreover, $\mathfrak{H}$ belongs to a VC class satisfying a uniform entropy condition, since this class is a subset of the vector space of functions spanned by $(y, x_1, ..., x_p)$, where p is the fixed dimension of x, so Lemma 2.6.15 of VW shows the desired result. Thus, by Example 2.10.23 (and Theorem 2.10.20) in VW the class defined by $1/\sigma\ \mathfrak{H}$ is Donsker, because the envelope of $\mathfrak{H}$ $(|y| + \mathit{const} * |x|)$ is square integrable by assumptions A3-A4. Thus $\mathcal{F}_2$ is Donsker. Using the same arguments as in the previous case we can define $h_2 = (\beta, \tau, \sigma) \mapsto \mathbb{G}_n\psi_{2\theta}(y,x)$, and by the established Donskerness property, this process is Donsker in $l^\infty(\mathcal{F}_2)$.

The third element of the vector is $\psi_{3\theta}(y,x) = \left(-\frac{1}{\sigma} + \frac{1}{\sigma^2}\rho_\tau(y_i - x_i'\beta)\right)$. Consider the following empirical process defined by $\mathfrak{J} = \{\rho_\tau(y_i - x_i'\beta), \tau \in \mathcal{T}, \beta \in \mathcal{B}\}$. This is Donsker by an application of Theorem 2.10.6 in VW. Finally, as in the previous cases define $h_3 = (\beta, \tau, \sigma) \mapsto \mathbb{G}_n\psi_{3\theta}(y,x)$, and by the established Donskerness property, this process is Donsker in $l^\infty(\mathcal{F}_3)$.

Now we turn our attention to the stochastic equicontinuity. The process $\theta \mapsto \mathbb{G}_n\psi_\theta(y,x)$ is stochastically equicontinuous over Θ with respect to a $L_2(P)$ pseudometric.[6] First, as in Angrist, Chernozhukov, and Fernández-Val (2006) and Chernozhukov and Hansen (2006), we define the distance d as the following $L_2(P)$ pseudometric

$$d(\theta', \theta'') = \sqrt{E\left([\psi_{\theta'} - \psi_{\theta''}]^2\right)}.$$

Thus, as $\|\theta - \theta_0\| \to 0$ we need to show that

$$d(\theta, \theta_0) \to 0, \tag{28}$$

[6] See e.g. Kosorok (2008, p. 405) for a sufficient condition for stochastic equicontinuity.

and therefore, by Donskerness of $\theta \mapsto \mathbb{G}_n \Psi_\theta(y,x)$, we have

$$\mathbb{G}_n \psi_\theta(y,x) = \mathbb{G}_n \psi_{\theta_0}(y,x) + o_p(1),$$

that is

$$\sup_{|\theta-\theta_0|\le\delta_n} \|\mathbb{G}_n \psi_\theta(y,x) - \mathbb{G}_n \psi_{\theta_0}(y,x)\| = o_p(1).$$

To show (28), first note that

$$\begin{aligned}
d(\theta',\theta) &= \sqrt{E\left([\psi_{1\theta'} - \psi_{1\theta}]^2\right)} \\
&= \sqrt{E\left(\left[(\tau' - 1(y - x\beta'))\frac{x}{\sigma'} - (\tau - 1(y - x\beta))\frac{x}{\sigma}\right]^2\right)} \\
&\le \Big[\left(E\Big|\frac{1}{\sigma'}(\tau' - 1(y - x\beta')) - \frac{1}{\sigma}1(\tau - (y - x\beta))\Big|^{\frac{2(2+\epsilon)}{\epsilon}}\right)^{\frac{\epsilon}{(2+\epsilon)}} \cdot \left(E(|x|^2)^{\frac{2+\epsilon}{2}}\right)^{\frac{2}{(2+\epsilon)}}\Big]^{\frac{1}{2}} \\
&= \left(E\Big|(\frac{\tau'}{\sigma'} - \frac{\tau}{\sigma}) + (\frac{1}{\sigma}1(y \le x\beta) - \frac{1}{\sigma'}1(y \le x\beta'))\Big|^{\frac{2(2+\epsilon)}{\epsilon}}\right)^{\frac{\epsilon}{2(2+\epsilon)}} \cdot \left(E(|x|^2)^{\frac{2+\epsilon}{2}}\right)^{\frac{1}{(2+\epsilon)}} \\
&\le \Big[\left(E\Big(\Big|\frac{\tau'}{\sigma'} - \frac{\tau}{\sigma}\Big|\Big)^{\frac{2(2+\epsilon)}{\epsilon}}\right)^{\frac{\epsilon}{2(2+\epsilon)}} \\
&+ \left(E\Big(\Big|\frac{1}{\sigma}1(y \le x\beta) - \frac{1}{\sigma'}1(y \le x\beta')\Big|\Big)^{\frac{2(2+\epsilon)}{\epsilon}}\right)^{\frac{\epsilon}{2(2+\epsilon)}}\Big] \cdot \left(E(|x|^2)^{\frac{2+\epsilon}{2}}\right)^{\frac{1}{(2+\epsilon)}} \\
&\le \Big[\Big|\frac{\tau'}{\sigma'} - \frac{\tau}{\sigma}\Big| + \left(E\Big|\bar{g}\cdot x'(\frac{\beta'}{\sigma'} - \frac{\beta}{\sigma})\Big|\right)^{\frac{\epsilon}{2(2+\epsilon)}}\Big] \cdot \left(E\|x\|^{2+\epsilon}\right)^{\frac{1}{(2+\epsilon)}} \\
&\le \Big[\Big|\frac{\tau'}{\sigma'} - \frac{\tau}{\sigma}\Big| + \left(\bar{g}E\|x\|\Big\|\frac{\beta'}{\sigma'} - \frac{\beta}{\sigma}\Big\|\right)^{\frac{\epsilon}{2(2+\epsilon)}}\Big] \cdot \left(E\|x\|^{2+\epsilon}\right)^{\frac{1}{(2+\epsilon)}},
\end{aligned}$$

where the first inequality is Holder's inequality, the second is Minkowski's inequality, the third is a Taylor expansion as in Angrist, Chernozhukov, and Fernández-Val (2006) where $\bar{g}$ is the upper bound of $g(y|x)$ (using A2), and the last is Cauchy-Schwarz inequality.

Now rewrite $\psi_{2\theta}(y,x) = \left(\sigma\frac{1-2\tau}{\tau(1-\tau)} - (y - x'\beta)\right)$ and

$$
\begin{aligned}
d(\theta',\theta) &= \sqrt{E\left([\psi_{2\theta'} - \psi_{2\theta}]^2\right)} \\
&= \sqrt{E\left(\left[\sigma'\frac{1-2\tau'}{\tau'(1-\tau')} - (y - x'\beta') - \sigma\frac{1-2\tau}{\tau(1-\tau)} + (y - x'\beta)\right]^2\right)} \\
&= \sqrt{E\left(\left|\sigma'\frac{1-2\tau'}{\tau'(1-\tau')} - \sigma\frac{1-2\tau}{\tau(1-\tau)} + (x'(\beta-\beta'))\right|^2\right)} \\
&\le \left(E\left|\sigma'\frac{1-2\tau'}{\tau'(1-\tau')} - \sigma\frac{1-2\tau}{\tau(1-\tau)}\right|^2\right)^{1/2} + \left(E|x'(\beta-\beta')|^2\right)^{1/2} \\
&\le \left(E\left|\sigma'\frac{1-2\tau'}{\tau'(1-\tau')} - \sigma\frac{1-2\tau}{\tau(1-\tau)}\right|^2\right)^{1/2} + \|\beta'-\beta\|\left(E\|x\|^2\right)^{1/2},
\end{aligned}
$$

where the first inequality is given by Minkowski's inequality

$$
(E|X+Y|^p)^{1/p} \le (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}, \text{ for } p \ge 1,
$$

and the second inequality is Cauchy-Schwarz inequality.

Finally, rewrite $\psi_{3\theta}(y,x) = (-\sigma + \rho_\tau(y - x'\beta))$, and thus

$$
\begin{aligned}
d(\theta',\theta) &= \sqrt{E\left([\psi_{3\theta'} - \psi_{3\theta}]^2\right)} \\
&= \sqrt{E\left(\left[-\sigma' + \rho_{\tau'}(y - x\beta') + \sigma - \rho_\tau(y - x\beta)\right]^2\right)} \\
&= \sqrt{E\left(\left[-\sigma' + \sigma + \rho_{\tau'}(y - x\beta') - \frac{1}{\sigma^2}\rho_\tau(y - x\beta)\right]^2\right)} \\
&\le \sqrt{E\left(-\sigma'+\sigma\right)^2} + \sqrt{E\left(\left[\rho_{\tau'}(y - x\beta') - \rho_\tau(y - x\beta)\right]^2\right)} \\
&\le |\sigma-\sigma'| + \sqrt{E\left(\left[\|x(\beta'-\beta)\| + |\tau'-\tau|(y - x\beta)\right]^2\right)} \\
&\le |\sigma-\sigma'| + \left(E\left[\|x\|\|\beta'-\beta\|\right]^2\right)^{1/2} + \left(E\left[|\tau'-\tau|(y - x\beta)\right]^2\right)^{1/2} \\
&= |\sigma-\sigma'| + \|\beta'-\beta\|\left(E\|x\|^2\right)^{1/2} + |\tau'-\tau|\left(E[(y - x\beta)]^2\right)^{1/2} \\
&\le const\cdot(|\sigma-\sigma'| + \|\beta'-\beta\| + |\tau'-\tau|),
\end{aligned}
$$

where the first inequality is is given by Minkowski's inequality, the second inequality is given by QR check function properties as $\rho_\tau(x+y) - \rho_\tau(y) \le 2|x|$ and $\rho_{\tau_1}(y - x't) - \rho_{\tau_2}(y - x't) = (\tau_2 - \tau_1)(y - x't)$. Third inequality is Cauchy-Schwarz inequality. Fourth is Minkowski's inequality. Last inequality uses assumption A4.

Thus, $\|\theta' - \theta\| \to 0$ implies that $d(\theta', \theta) \to 0$ in every case, and therefore, by Donskerness of $\theta \mapsto \mathbb{G}_n \psi_\theta(y, x)$ we have that

$$\sup_{\|\theta - \theta_0\| \leq \delta_n} \|\mathbb{G}_n \psi_\theta(y, x) - \mathbb{G}_n \psi_{\theta_0}(y, x)\| = o_p(1).$$

■

ACKNOWLEDGEMENT

We are very grateful to Arnold Zellner, Jushan Bai, Rong Chen, Daniel Gervini, Yongmiao Hong, Carlos Lamarche, Ehsan Soofi, Zhijie Xiao, and the participants in seminars at University of Wisconsin-Milwaukee, City University London, World Congress of the Econometric Society, Shanghai, August 2010, Latin American Meeting of the Econometric Society, Argentina, October 2009, Summer Workshop in Econometrics, Tsinghua University, Beijing, China, May 2009, South Asian and Far Eastern Meeting of the Econometrics Society, Singapore, July 2008, for helpful comments and discussions. However, we retain the responsibility for any remaining errors.

REFERENCES

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

Bloom, H. S. B., L. L. Orr, S. H. Bell, G. Cave, F. Doolittle, W. Lin, and J. M. Bos (1997): "The Benefits and Costs of JTPA Title II-A programs. Key Findings from the National Job Training Partnership Act Study," *Journal of Human Resources*, 32, 549–576.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2009): "Inference on counterfactual distributions," CEMMAP working paper CWP09/09.

Chernozhukov, V., and C. Hansen (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effects Models," *Journal of Econometrics*, 132, 491–525.

——— (2008): "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics*, 142, 379–398.

Ebrahimi, N., E. S. Soofi, and R. Soyer (2008): "Multivariate Maximum Entropy Identification, Transformation, and Dependence," *Journal of Multivariate Analysis*, 99, 1217–1231.

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259–276.

Geraci, M., and M. Botai (2007): "Quantile regression for longitudinal data using the asymmetric Laplace distribution," *Biostatistics*, 8, 140–154.

He, X., and Q.-M. Shao (1996): "A General Bahadur Representation of M-Estimators and its Applications to Linear Regressions with Nonstochastic Designs," *Annals of Statistics*, 24, 2608–2630.

——— (2000): "Quantile regression estimates for a class of linear and partially linear errors-in-variables models," *Statistica Sinica*, 10, 129–140.

Hinkley, D. V., and N. S. Revankar (1977): "Estimation of the Pareto Law from Underreported Data: A Further Analysis," *Journal of Econometrics*, 5, 1–11.

Huber, P. J. (1967): "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," in *Fifth Symposium on Mathematical Statistics and Probability*, pp. 179–195. Unibersity of California, Berkeley, California.

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

Koenker, R. (2005): *Quantile Regression*. Cambridge University Press, Cambridge.

Koenker, R., and G. W. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–49.

Koenker, R., and J. A. F. Machado (1999): "Godness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 94, 1296–1310.

Koenker, R., and Z. Xiao (2002): "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612.

Komunjer, I. (2005): "Quasi-maximum likelihood estimation for conditional quantiles," *Journal of Econometrics*, 128, 137–164.

——— (2007): "Asymmetric Power Distribution: Theory and Applications to Risk Measurement," *Journal of Applied Econometrics*, 22, 891–921.

Kosorok, M. R. (2008): *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag Press, New York, New York.

KOTZ, S., T. J. KOZUBOWSKI, AND K. PODGÓRSK (2002a): "Maximum Entropy Characterization of Asymmetric Laplace Distribution," *International Mathematical Journal*, 1, 31–35.

——— (2002b): "Maximum Likelihood Estimation of Asymmetric Laplace Distributions," *Annals of the Institute Statistical Mathematics*, 54, 816–826.

LALONDE, R. J. (1995): "The Promise of Public-Sponsored Training Programs," *Journal of Economic Perspectives*, 9, 149–168.

MACHADO, J. A. F. (1993): "Robust Model Selection and M-Estimation," *Econometric Theory*, 9, 478–493.

MANSKI, C. F. (1991): "Regression," *Journal of Economic Literature*, 29, 34–50.

PARK, S. Y., AND A. BERA (2009): "Maximum entropy autoregressive conditional heteroskedasticity model," *Journal of Econometrics*, 150, 219–230.

SCHENNACH, S. M. (2008): "Quantile Regression with Mismeasured Covariates," *Econometric Theory*, 24, 1010–1043.

SOOFI, E. S., AND J. J. RETZER (2002): "Information Indices: Unification and Applications," *Journal of Econometrics*, 107, 17–40.

VAN DER VAART, A. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.

VAN DER VAART, A., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag, New York, New York.

WEI, Y., AND R. J. CARROLL (2009): "Quantile regression with measurement error," *Journal of the American Statistical Association*, 104, 1129–1143.

YU, K., AND R. A. MOYEED (2001): "Bayesian quantile regression," *Statistics & Probability Letters*, 54, 437–447.

YU, K., AND J. ZHANG (2005): "A three-parameter asymmetric Laplace distribution and its extension," *Communications in Statistics - Theory and Methods*, 34, 1867–1879.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 236–245.

M-TEST OF TWO PARALLEL REGRESSION LINES UNDER UNCERTAIN PRIOR INFORMATION

Shahjahan Khan and Rossita M Yunus[1]
Department of Mathematics and Computing
Australian Centre for Sustainable Catchments
University of Southern Queensland
Toowoomba, Q 4350, AUSTRALIA.
E-mail:khans@usq.edu.au and rossita@um.edu.my

ABSTRACT

This paper considers the problem of testing the intercepts of two simple linear models following a pre-test on the suspected equality of slopes. The unrestricted test (UT), restricted test (RT) and pre-test test (PTT) are proposed from the M-tests using the M-estimation methodology. The asymptotic power functions of the UT, RT and PTT are given. The computational comparisons of power function of the three tests are provided. The PTT achieves a reasonable dominance over the other two tests asymptotically.

Keywords: Two parallel regression lines, pre-test, asymptotic power and size, M-estimation, local alternative hypothesis, bivariate non-central chi-square.

1. INTRODUCTION

A researcher may model independent data sets from two random samples for two separate groups of respondents. Often, the researcher may wish to know whether the regression lines for the two groups are parallel (i.e. slopes of the two regression lines are equal) or whether the lines have the same intercept on vertical-axis. An interesting situation would be if the researcher decides to test the equality of the intercepts when the equality of slopes is suspected, but not sure.

Data for this problem can be represented by the following two simple linear regression equations

$$\boldsymbol{X}_{1_{n_1}} = \theta_1 \mathbf{1}_{n_1} + \beta_1 \boldsymbol{c}_1 + \boldsymbol{\varepsilon}_1 \text{ and } \boldsymbol{X}_{2_{n_2}} = \theta_2 \mathbf{1}_{n_2} + \beta_2 \boldsymbol{c}_2 + \boldsymbol{\varepsilon}_2. \qquad (1.1)$$

For the first data set, $\boldsymbol{X}_{1_{n_1}} = (X_{1_1}, \ldots, X_{1_{n_1}})'$ is a vector of n_1 observable response random variables, $\mathbf{1}_{n_1} = (1, 1, \ldots, 1)'$, is an n_1 tuple of 1's, $\boldsymbol{c}_1 = (c_{1_1}, \ldots, c_{1_{n_1}})'$ is a vector of n_1 independent variables, θ_1 and β_1 are the unknown intercept and slope parameters respectively. For the second data set, $\boldsymbol{X}_{2_{n_2}} = (X_{2_1}, \ldots, X_{2_{n_2}})'$ is a vector of n_2 observable response random variables, $\mathbf{1}_{n_2} = (1, 1, \ldots, 1)'$, is an n_2 tuple of 1's, $\boldsymbol{c}_2 = (c_{2_1}, \ldots, c_{2_{n_2}})'$ is a vector of

[1]On leave from Institute of Mathematical Sciences, Faculty of Sciences, University of Malaya, Malaysia.

n_2 independent variables, θ_2 and β_2 are the unknown intercept and slope parameters respectively. Assume the error $\varepsilon_{j_i} = X_{j_i} - \theta_j - \beta_j c_{j_i}$, for $i = 1, \ldots, n_j$ and $j = 1, 2$ are mutually independent and identically distributed with cdf F.

The researcher may wish to test the intercept vector $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ of the two regression lines equal to a fixed vector $\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02})'$ while it not sure if the two slope parameters are equal. In this situation, three different scenarios associated with the value of the slopes are considered: the value of the slopes would either be (i) completely unspecified, (ii) equal at an arbitrary constant, β_0, or (iii) suspected to be equal at an arbitrary constant, β_0. The unrestricted test (UT), the restricted test (RT) and the pre-test test (PTT) are defined respectively for the three scenarios of the slope parameters. Thus, the UT is for testing $H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)} : \boldsymbol{\theta} > \boldsymbol{\theta}_0$ when $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ is unspecified, the RT is for testing $H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)} : \boldsymbol{\theta} > \boldsymbol{\theta}_0$ when $\boldsymbol{\beta} = \beta_0 \mathbf{1}_2$ (fixed vector) and the PTT is for testing $H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)} : \boldsymbol{\theta} > \boldsymbol{\theta}_0$ after pre-testing $H_0^{\star} : \boldsymbol{\beta} = \beta_0 \mathbf{1}_2$ against $H_A^{\star} : \boldsymbol{\beta} > \beta_0 \mathbf{1}_2$ (to remove the uncertainty). The PTT is a choice between the UT and the RT. If the null hypothesis $H_0^{\star}$ is rejected in the pre-test (PT), then the UT is used, otherwise the RT is used.

The inclusion of non-sample prior information (NSPI) in the parameter estimation usually improve the quality of an estimator. In many cases, the prior information is available as a suspected value of the parameter interest. However, such a prior value is likely to be uncertain. This has led to the suggestion of pre-testing the suspected value to remove the uncertainty. The idea of pre-testing by Bancroft (1964) arouses a number of studies in literature. Akritas et al. (1984), Lambert et al. (1985a) and Khan (2003) are among authors who considered the problem of estimating the intercepts parameters when it is *apriori* suspected that the regression lines are parallel.

In literature, the effects of pre-testing on the performance of the ultimate test are studied for some parametric cases by Bechhofer (1951), Bozivich et al. (1956) and Mead et al. (1973). For nonparametric cases, Tamura (1965) investigated the performance of the PTT for one sample and two sample problem while Saleh and Sen (1982, 1983) developed the PTT for the simple linear model and multivariate simple model using nonparametric rank tests. Lambert et al. (1985b) used least-squares (LS) based tests to propose the UT, RT and PTT for the parallelism model. However, LS estimates are non robust with respect to deviation from the assumed (normal) distribution (c.f. Jurečková and Sen, 1996, p.21), so, it is suspected that the UT, RT and PTT defined using the LS based tests are also non robust. In this paper, the M-test which is originally proposed by Sen (1982) to test the significance of the slope is used to define the UT, RT and PTT. Recently, Yunus and Khan (2010) used M-tests to define the UT, RT and PTT for the simple linear regression model.

The comparison of the power of the UT, RT and PTT are studied by Lambert et al. (1985b) analytically. The cdf of the bivariate noncentral chi-square distribution is required to compute the power of the PTT. The bivariate noncentral chi-square distribution function used in their paper is complicated and not practical for computation, so there is no graphical representation on the comparison of the power of the tests provided in their paper. In this paper, Yunus and Khan (2009) is referred for the computation of the cdf of the bivariate noncentral chi-square distribution.

Along with some preliminary notions, the UT, RT and PTT are proposed in Sections 2. In Section 3, the asymptotic power functions for UT, RT and PTT are given. The graphical representation is given in Section 4. The final Section contains comments and conclusion.

2. THE UT, RT AND PTT

2.1 THE UNRESTRICTED TEST (UT)

If $\boldsymbol{\beta}$ is unspecified, ϕ_n^{UT} is the test function of $H_0^{(1)}: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)}: \boldsymbol{\theta} > \boldsymbol{\theta}_0$. We consider the test statistic

$$\left[T_n^{UT} = n^{-1} \frac{\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}})' \boldsymbol{\Lambda}_0^{\star -1} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}})}{{S_n^{(1)}}^2}, \right]$$

where $\tilde{\boldsymbol{\beta}} = \frac{1}{2}\left[\sup\{\boldsymbol{b} : \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{b}) > 0\} + \inf\{\boldsymbol{b} : \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{b}) < 0\}\right]$ is a constrained M-estimator of $\boldsymbol{\beta}$ under $H_0^{(1)}$. For $\boldsymbol{a} = (a_1, a_2)'$ and $\boldsymbol{b} = (b_1, b_2)'$, vectors of real numbers a_j and b_j, $j = 1,\ 2$, $\boldsymbol{M}_{n_1}(\boldsymbol{a}, \boldsymbol{b}) = \left(M_{n_1}^{(1)}(a_1, b_1), M_{n_1}^{(2)}(a_2, b_2)\right)'$ and $\boldsymbol{M}_{n_2}(\boldsymbol{a}, \boldsymbol{b}) = \left(M_{n_2}^{(1)}(a_1, b_1), M_{n_2}^{(2)}(a_2, b_2)\right)'$ where

$$\begin{aligned} M_{n_1}^{(j)}(a_j, b_j) &= \sum_{i=1}^{n_j} \psi\left(\frac{X_{j_i} - a_j - b_j c_{j_i}}{S_n}\right) \text{ and} \\ M_{n_2}^{(j)}(a_j, b_j) &= \sum_{i=1}^{n_j} c_{j_i}\psi\left(\frac{X_{j_i} - a_j - b_j c_{j_i}}{S_n}\right). \end{aligned}$$

Here, S_n is an appropriate scale statistic for some functional $S = S(F) > 0$ and ψ is the score function in the M-estimation methodology. Note, $\boldsymbol{\Lambda}_0^\star = \mathrm{Diag}\left(\frac{\lambda_1 C_1^{\star 2}}{C_1^{\star 2} + \bar{c}_1^2}, \frac{\lambda_2 C_2^{\star 2}}{C_2^{\star 2} + \bar{c}_2^2}\right)$, where $\lambda_j = \lim_{n\to\infty} \frac{n_j}{n}$ $(0 < \lambda_j < 1)$ with $n = n_1 + n_2$. Also, $\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n_j} c_{j_i} = \lambda_j \bar{c}_j$ $(|\bar{c}_j| < \infty)$ and $\lim_{n\to\infty} n^{-1} {C_{n_j}^\star}^2 = \lambda_j C_j^{\star 2}$, where ${C_{n_j}^\star}^2 = \sum_{i=1}^{n_j} c_{j_i}^2 - n_j \bar{c}_{n_j}^2$ and $\bar{c}_{n_j} = n_j^{-1} \sum_{i=1}^{n_j} c_{j_i}$. Let ${S_n^{(1)}}^2 = \frac{1}{n} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \psi^2 \left(\frac{X_{i_j} - \theta_{0j} - \tilde{\beta}_j c_{j_i}}{S_n}\right)$. The asymptotic results in Yunus and Khan (2010) is adapted for the parallelism model. Thus, we find T_n^{UT} is χ_2^2 (chi-square distribution with 2 degrees of freedom).

Let ℓ_{n,α_1}^{UT} be the critical value of T_n^{UT} at the α_1 level of significance. So, for the test function $\phi_n^{UT} = I(T_n^{UT} > \ell_{n,\alpha_1}^{UT})$, the power function of the UT becomes $\Pi_n^{UT}(\boldsymbol{\theta}) = E(\phi_n^{UT}|\boldsymbol{\theta}) = P(T_n^{UT} > \ell_{n,\alpha_1}^{UT}|\boldsymbol{\theta})$, where $I(A)$ is an indicator function of the set A. It takes value 1 if A occurs, otherwise it is 0.

2.2 THE RESTRICTED TEST (RT)

If $\boldsymbol{\beta} = \beta_0\mathbf{1}_2$, the test function for testing $H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)} : \boldsymbol{\theta} > \boldsymbol{\theta}_0$ is ϕ_n^{RT}. The proposed test statistic is

$$\left[T_n^{RT} = n^{-1} \frac{\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_2)'\boldsymbol{\Lambda}_0^{-1}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_2)}{{S_n^{(2)}}^2}, \right]$$

where ${S_n^{(2)}}^2 = \frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\psi^2\left(\frac{X_{i_j} - \theta_{0j} - \beta_0 c_{j_i}}{S_n}\right)$ and $\boldsymbol{\Lambda}_0 = \mathrm{Diag}(\lambda_1, \lambda_2)$. Again, using the asymptotic results in Yunus and Khan (2010) and adapt them for use in the parallelism model, we obtain for large n, T_n^{RT} is χ_2^2 under $H_0^{(2)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\beta} = \beta_0\mathbf{1}_2$. Again, let ℓ_{n,α_2}^{RT} be the critical value of T_n^{RT} at the α_2 level of significance. So, for the test function $\phi_n^{RT} = I(T_n^{RT} > \ell_{n,\alpha_2}^{RT})$, the power function of the RT becomes $\Pi_n^{RT}(\boldsymbol{\theta}) = E(\phi_n^{RT}|\boldsymbol{\theta}) = P(T_n^{RT} > \ell_{n,\alpha_2}^{RT}|\boldsymbol{\theta})$.

2.3 THE PRE-TEST (PT)

For the pre-test on the slope, the test function of $H_0^{\star} : \boldsymbol{\beta} = \beta_0\mathbf{1}_2$ against $H_A^{\star} : \boldsymbol{\beta} > \beta_0\mathbf{1}_2$ is ϕ_n^{PT}. The proposed test statistic is

$$\left[T_n^{PT} = n^{-1} \frac{\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\mathbf{1}_2)'\boldsymbol{\Lambda}_2^{\star -1}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\mathbf{1}_2)}{{S_n^{(3)}}^2}, \right]$$

where $\tilde{\boldsymbol{\theta}} = \frac{1}{2}\left[\sup\{\boldsymbol{a} : \boldsymbol{M}_{n_1}(\boldsymbol{a}, \beta_0\mathbf{1}_2) > 0\} + \inf\{\boldsymbol{a} : \boldsymbol{M}_{n_1}(\boldsymbol{a}, \beta_0\mathbf{1}_2) < 0\}\right]$ is a constrained M-estimator of $\boldsymbol{\theta}$ and ${S_n^{(3)}}^2 = \frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\psi^2\left(\frac{X_{i_j} - \tilde{\theta}_j - \beta_0 c_{j_i}}{S_n}\right)$ and $\boldsymbol{\Lambda}_2^{\star} = \mathrm{Diag}\left(\lambda_1 C_1^{\star 2},\ \lambda_2 C_2^{\star 2}\right)$. It follows that as $n \to \infty$, $T_n^{PT} \xrightarrow{d} \chi_2^2$ under $H_0^{\star}$.

2.4 THE PRE-TEST-TEST (PTT)

We are now in the position to formulate ϕ_n^{PTT} for testing $H_0^{(1)}$ following a pre-test on $\boldsymbol{\beta}$. Since the PTT is a choice between RT and UT, define,

$$\phi_n^{PTT} = I[(T_n^{PT} < \ell_{n,\alpha_3}^{PT}, T_n^{RT} > \ell_{n,\alpha_2}^{RT}) \text{ or } (T_n^{PT} > \ell_{n,\alpha_3}^{PT}, T_n^{UT} > \ell_{n,\alpha_1}^{RT})], \tag{2.1}$$

where ℓ_{n,α_3}^{PT} is the critical value of T_n^{PT} at the α_3 level of significance. The power function of the PTT is given by

$$\Pi_n^{PTT}(\boldsymbol{\theta}) = E(\phi_n^{PTT}|\boldsymbol{\theta}) \tag{2.2}$$

and the size of the PTT is obtained by substituting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ in equation (2.2).

3. ASYMPTOTIC POWER FUNCTIONS

Let $\{K_n^\star\}$ be a sequence of alternative hypotheses, where

$$K_n^\star : (\boldsymbol{\theta}, \boldsymbol{\beta}) = (\boldsymbol{\theta}_0 + n^{-\frac{1}{2}}\boldsymbol{\delta}_1, \beta_0\mathbf{1}_2 + n^{-\frac{1}{2}}\boldsymbol{\delta}_2), \tag{3.1}$$

with $\boldsymbol{\delta}_1 = n^{\frac{1}{2}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) > \mathbf{0}$ and $\boldsymbol{\delta}_2 = n^{\frac{1}{2}}(\boldsymbol{\beta} - \beta_0\mathbf{1}_2) > \mathbf{0}$. Here, $\boldsymbol{\delta}_1 = (\delta_{1_1}, \delta_{1_2})', \boldsymbol{\delta}_2 = (\delta_{2_1}, \delta_{2_2})'$ are vectors of fixed real numbers.

Under $\{K_n^\star\}$, for large sample, asymptotically (T_n^{RT}, T_n^{PT}) are independently distributed as bivariate non-central chi-square distribution with 2 degrees of freedom and (T_n^{UT}, T_n^{PT}) are distributed as correlated bivariate non-central chi-square distribution with 2 degrees of freedom and non-centrality parameters,

$$\begin{aligned}
\theta^{UT} &= (\gamma\boldsymbol{\Lambda}_0^\star\boldsymbol{\delta}_1)'\,\boldsymbol{\Lambda}_0^{\star-1}\,(\gamma\boldsymbol{\Lambda}_0^\star\boldsymbol{\delta}_1)\,/\sigma_0^2 &(3.2)\\
\theta^{RT} &= [\gamma(\boldsymbol{\Lambda}_0\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_2)]'\,\boldsymbol{\Lambda}_0{}^{-1}\,[\gamma(\boldsymbol{\Lambda}_0\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_2)]\,/\sigma_0^2, &(3.3)\\
\theta^{PT} &= (\gamma\boldsymbol{\Lambda}_2^\star\boldsymbol{\delta}_2)'\,\boldsymbol{\Lambda}_2^{\star-1}\,(\gamma\boldsymbol{\Lambda}_2^\star\boldsymbol{\delta}_2)\,/\sigma_0^2, &(3.4)
\end{aligned}$$

where $\boldsymbol{\Lambda}_{12} = \mathrm{Diag}(\lambda_1\bar{c}_1, \lambda_2\bar{c}_2)$, $\sigma_0^2 = \int_{-\infty}^{\infty}\psi^2(z/S)dF(z)$ and $\gamma = \int_{-\infty}^{\infty}\frac{1}{S}\psi'(z/S)dF(z)$.

Thus, under $\{K_n^\star\}$, the asymptotic power functions for the UT, RT and PT which are denoted by $\Pi^h(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ for h any of the UT, RT and PT, are defined as

$$\Pi^h(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \lim_{n\to\infty}\Pi_n^h(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \lim_{n\to\infty}P(T_n^h > \ell_{n,\alpha_\nu}^h|K_n^\star) = 1 - G_2(\chi^2_{2,\alpha_\nu}; \theta^h), \tag{3.5}$$

where is $G_2(\chi^2_{2,\alpha_\nu}, \theta^h)$ is the cdf of a non-central chi-square distribution with 2 degre es of freedom and non-centrality parameter θ^h. The level of significance, $\alpha_\nu,\ \nu = 1, 2, 3$ are chosen together with the critical values ℓ^h_{n,α_ν} for the UT, RT and PT. Here, $\chi^2_{2,\alpha}$ is the upper $100\alpha\%$ critical value of a central chi-square distribution and $\ell^{UT}_{n,\alpha_1} \to \chi^2_{2,\alpha_1}$ under $H_0^{(1)}$, $\ell^{RT}_{n,\alpha_2} \to \chi^2_{2,\alpha_2}$ under $H_0^{(2)}$ and $\ell^{PT}_{n,\alpha_3} \to \chi^2_{2,\alpha_3}$under $H_0^\star$.

For testing $H_0^{(1)}$ following a pre-test on $\boldsymbol{\beta}$, using equation (2.1), the asymptotic power function for the PTT under $\{K_n^\star\}$ is given as

$$\begin{aligned}
&\Pi^{PTT}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)\\
&= \lim_{n\to\infty}P(T_n^{PT} \le \ell^{PT}_{n,\alpha_3}, T_n^{RT} > \ell^{RT}_{n,\alpha_2}|K_n^\star) + \lim_{n\to\infty}P(T_n^{PT} > \ell^{PT}_{n,\alpha_3}, T_n^{UT} > \ell^{UT}_{n,\alpha_1}|K_n^\star)\\
&= G_2(\chi^2_{2,\alpha_3}; \theta^{PT})\{1 - G_2(\chi^2_{2,\alpha_2}; \theta^{RT})\} + \int_{\chi^2_{2,\alpha_1}}\int_{\chi^2_{2,\alpha_3}}\phi^\star(w_1, w_2)dw_1dw_2,
\end{aligned} \tag{3.6}$$

where $\phi^\star(\cdot)$ is the density function of a bivariate non-central chi-square distribution with 2 degrees of freedom, non-centrality parameters, θ^{UT} and θ^{PT} and correlation coefficient

$-1 < \rho < 1$. The probability integral in (3.6) is given by

$$\begin{aligned}
&\int_{\chi^2_{2,\alpha_1}} \int_{\chi^2_{2,\alpha_3}} \phi^\star(w_1, w_2) dw_1 dw_2 \\
= \;& \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{\kappa_1=0}^{\infty} \sum_{\kappa_2=0}^{\infty} (1-\rho^2)^p \frac{\Gamma(1+j)}{\Gamma(1)j!} \frac{\Gamma(1+k)}{\Gamma(1)k!} \rho^{2(j+k)} \\
&\times \left[1 - \gamma^\star\left(1 + j + \kappa_1, \frac{\chi^2_{2,\alpha_1}}{2(1-\rho^2)}\right)\right] \left[1 - \gamma^\star\left(1 + k + \kappa_2, \frac{\chi^2_{2,\alpha_3}}{2(1-\rho^2)}\right)\right] \\
&\times \frac{e^{-\theta^{UT}/2}(\theta^{UT}/2)^{\kappa_1}}{\kappa_1!} \frac{e^{-\theta^{PT}/2}(\theta^{PT}/2)^{\kappa_2}}{\kappa_2!}. \qquad (3.7)
\end{aligned}$$

Let $\rho^2 = \sum_{j=1}^{2} \frac{1}{2}\rho_j^2$ be the mean correlation, where $\rho_j = -c_j/\sqrt{C_j^{\star 2} + \bar{c}_j^2}$ is the correlation coefficient between $\left(M_{n_1}^{(j)}(\theta_{0_j}, \tilde{\beta}_j), M_{n_2}^{(j)}(\tilde{\theta}_j, \beta_0)\right)$. Here, $\gamma^\star(v, d) = \int_0^d x^{v-1}\, e^{-x}/\Gamma(v) dx$ is the incomplete gamma function. For details on the evaluation of the bivariate integral, see Yunus and Khan (2009). The density function of the bivariate noncentral chi-square distribution given above is a mixture of the bivariate central chi-square distribution of two central chi-square random variables (see Gunst and Webster, 1973, Wright and Kennedy, 2002), with the probabilities from the Poisson distribution. Krishnaiah et al., 1963, Gunst and Webster, 1973 and Wright and Kennedy, 2002)

4. ILLUSTRATIVE EXAMPLE

The power functions given in equations (3.5) and (3.6) are computed for graphical view. The non-centrality parameters for UT, RT and PT respectively are

$$\begin{aligned}
\theta^{UT} &= \begin{bmatrix} \xi_{1_1}\lambda_1 \frac{C_1^{\star 2}}{C_1^{\star 2}+\bar{c}_1^2} \\ \xi_{1_2}\lambda_2 \frac{C_2^{\star 2}}{C_2^{\star 2}+\bar{c}_2^2} \end{bmatrix}' \begin{bmatrix} \lambda_1 \frac{C_1^{\star 2}}{C_1^{\star 2}+\bar{c}_1^2} & 0 \\ 0 & \lambda_2 \frac{C_2^{\star 2}}{C_2^{\star 2}+\bar{c}_2^2} \end{bmatrix}^{-1} \begin{bmatrix} \xi_{1_1}\lambda_1 \frac{C_1^{\star 2}}{C_1^{\star 2}+\bar{c}_1^2} \\ \xi_{1_2}\lambda_2 \frac{C_2^{\star 2}}{C_2^{\star 2}+\bar{c}_2^2} \end{bmatrix}, \\
\theta^{RT} &= \begin{bmatrix} \xi_{1_1}\lambda_1 + \xi_{2_1}\lambda_1\bar{c}_1 \\ \xi_{1_2}\lambda_2 + \xi_{2_2}\lambda_2\bar{c}_1 \end{bmatrix}' \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}^{-1} \begin{bmatrix} \xi_{1_1}\lambda_1 + \xi_{2_1}\lambda_1\bar{c}_1 \\ \xi_{1_2}\lambda_2 + \xi_{2_2}\lambda_2\bar{c}_1 \end{bmatrix} \text{ and} \\
\theta^{PT} &= \begin{bmatrix} \xi_{2_1}\lambda_1 C_1^{\star 2} \\ \xi_{2_2}\lambda_2 C_2^{\star 2} \end{bmatrix}' \begin{bmatrix} \lambda_1 C_1^{\star 2} & 0 \\ 0 & \lambda_2 C_2^{\star 2} \end{bmatrix}^{-1} \begin{bmatrix} \xi_{2_1}\lambda_1 C_1^{\star 2} \\ \xi_{2_2}\lambda_2 C_2^{\star 2} \end{bmatrix},
\end{aligned}$$

where $\xi_{k_l} = \delta_{k_l}\gamma/\sigma_0$ for $k, l = 1, 2$ and $\delta_{1_l} = \sqrt{n}(\theta_l - \theta_{0_l})$ and $\delta_{2_l} = \sqrt{n}(\beta_l - \beta_{0_l})$.

A special case of the two sample problem (Saleh, 2006, p.67) is considered with $n_j = n_{j_1} + n_{j_2}$ for $j = 1, 2$, $n_{j_1}/n_j \to 1 - P$, $c_{j_1} = \ldots = c_{j_{n_1}} = 0$ and $c_{j_{n_1+1}} = \ldots = c_{j_n} = 1$. So $\bar{c}_j = 1 - P$ and $C_j^{\star 2} = P(1 - P)$. In this example, let $P = 0.5$ and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$.

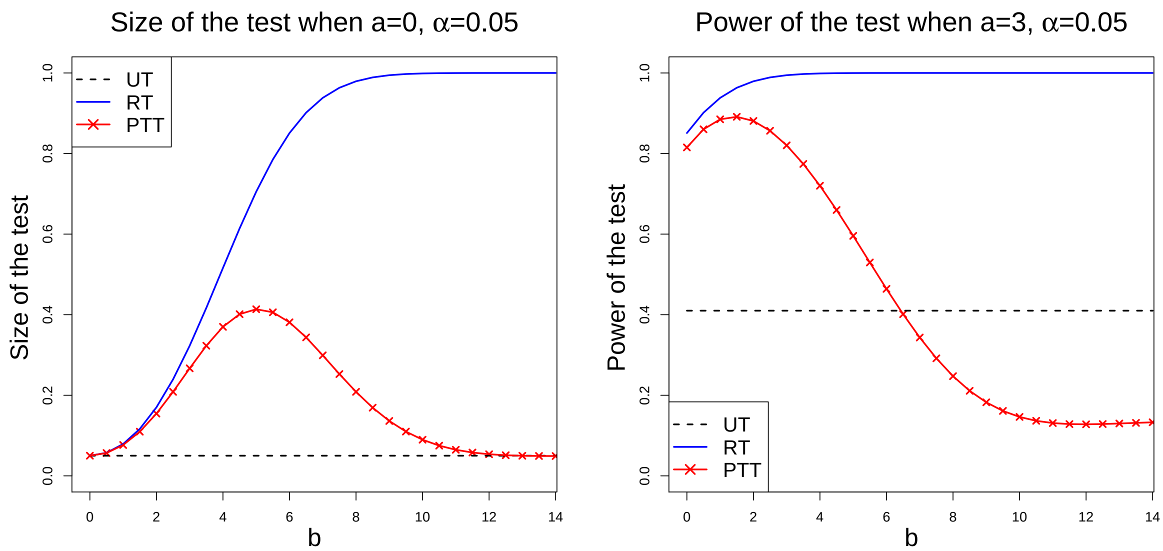


Figure 1: Graphs of power functions as a function of (b) $((=\xi_{2_1}=\xi_{2_2}))$ for selected values of (ξ_{1_1}) and (ξ_{1_2}) with $(\bar{c}>0)$ and $(\alpha_1=\alpha_2=\alpha_3=\alpha=0.05)$. Here, $(\xi_{k_l}=\delta_{k_l}\gamma/\sigma_0, k,l=1,2.)$

Also, let $n_1, n_2 = 50$ so $n = n_1 + n_2 = 100$. As a result, the correlation coefficient, $\rho_j, j = 1, 2$ for both regression lines are the same since $\bar{c}_1^2 = \bar{c}_2^2 = \bar{c}^2$ for both samples, (X_{n_1}, c_1) and (X_{n_2}, c_2) of the two regression lines. Note, in plotting the power functions for the PTT, a bivariate non-central chi-square distribution is used.

Let $\xi_{1_1} = \xi_{1_2} = a$ and $\xi_{2_1} = \xi_{2_2} = b$. Figure 1 shows the power of the test against b at selected values of ξ_{1_1} and ξ_{1_2}. A test with higher size and lower power is a test which makes small probability of Type I and Type II errors. In Figure 1, the UT has the smallest size and the PTT has smaller size than that of the RT. The RT has the largest power as b grows. The PTT has larger power than that of the UT except for large b. In Figure 2, power of the UT, RT and PTT are plotted against a at selected values of ξ_{2_1} and ξ_{2_2}. As a grows large, power of all tests grow large too. Although the power of the UT and RT are increasing to 1 as a is increasing, the power of the PTT is increasing to a value that is less than 1.

Although the NSPI on the slopes parameters may be uncertain, there is a high possibility that the true values are not too far from the suspected values. Therefore, the study on the behaviour of the three tests when the suspected NSPI values is not too far away from that under the null hypothesis is more realistic.

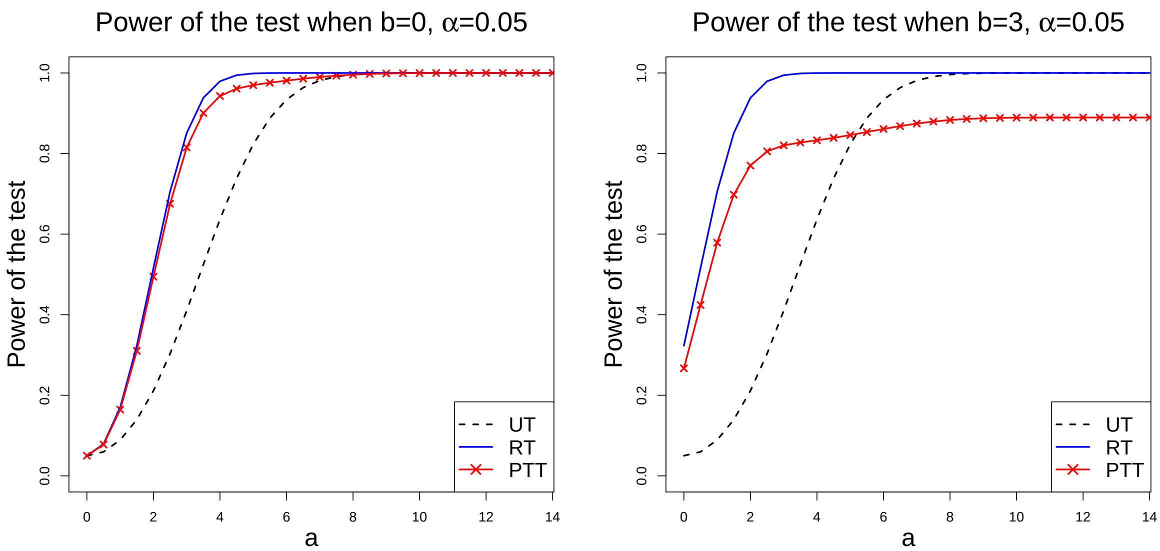


Figure 2: Graphs of power functions as a function of (a) $((= \xi_{1_2} = \xi_{1_2}))$ for selected values of (ξ_{2_1}) and (ξ_{2_2}) with $(\bar{c} > 0)$ and $(\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05)$. Here, $(\xi_{k_l} = \delta_{k_l}\gamma/\sigma_0, k, l = 1, 2.)$

5. COMMENTS AND CONCLUSION

The sampling distributions of the UT, RT and PT follow a univariate noncentral chi-square distribution under the alternative hypothesis when the sample size is large. However, that of the PTT is a bivariate noncentral chi-square distribution as there is a correlation between the UT and PT. Note that there is no such correlation between the RT and PT.

The size of the RT reaches 1 as b (a function of the difference between the true and suspected values of the slopes) increases. This means the RT does not satisfy the asymptotic level constraint, so it is not a valid test. The UT has the smallest constant size; however, it has the smallest power as well, except for very large values of b, that is, when $b > q$, where q is some positive number. Thus, the UT fails to achieve the highest power and lowest size simultaneously. The PTT has a smaller size than the RT and its size does not reach 1 as b increases. It also has higher power than the UT, except for $b > q$.

Therefore, if the prior information is not far away from the true value, that is, b is near 0 (small or moderate), the PTT has a smaller size than the RT and more power than the UT. So, the PTT is a better compromise between the two extremes. Since the prior information comes from previous experience or expert assessment, it is reasonable to expect b should not be too far from 0, although it may not be 0, and hence the PTT achieves a reasonable dominance over the other two tests in a more realistic situation.

ACKNOWLEDGEMENTS

The authors thankfully acknowledge valuable suggestions of Professor A K Md E Saleh, Carleton University, Canada that helped improve the content and quality of the results in the paper.

REFERENCES

Akritas, M.G., Saleh, A.K.Md.E. and Sen, P.K. (1984). Nonparametric Estimation of Intercepts After a Preliminary Test on Parallelism of Several Regression Lines. In P.K. Sen (Ed). Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences (Bernard G. Greenberg Vol.), North Holland.

Bancroft, T.A. (1964). Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. *Biometrics*, 20, 427-442.

Bechhofer, R.E. (1951). The effect of preliminary test of significance on the size and power of certain tests of univariate linear hypotheses. Ph.D. Thesis (unpublished), Columbia Univ.

Bozivich, H., Bancroft, T.A. and Hartley, H. (1956). Power of analysis of variance test procedures for certain incompletely specified models, *Ann. Math. Statist.*, 27, 1017 - 1043.

Gunst, R.F. and Webster, J.T. (1973). Density functions of the bivariate chi-square distribution, *Journal of Statistical Computation and Simulation A*, 2, 275–288.

Jurecčková, J. and Sen, P.K. (1996). *Robust Statistical Procedures Asymptotics and Interrelations.* US: John Wiley & Sons.

Khan S. (2003). Estimation of the parameters of two parallel regression lines under uncertain prior information. *Biometrical Journal*, 45, 73-90.

Lambert A. and Saleh A.K.Md.E. and Sen P.K. (1985a). On least squares estimation of intercepts after a preliminary test on parallelism of regression lines. *Comm. Statist.-theor. meth.*, 14, 793-807.

Lambert, A., Saleh A.K.Md.E. and Sen P.K. (1985b). Test of homogeneity of intercepts after a preliminary test on parallelism of several regression lines: from parametric to asymptotic, *Comm. Statist.*, 14, 2243-2259.

Mead, R., Bancroft, T.A. and Han, C.P. (1975). Power of analysis of variance test procedures for incompletely specified fixed models. *Ann. Statist.*, 3, 797-808. New

Saleh, A.K.Md.E. and Sen, P.K. (1982). Non-parametric tests for location after a preliminary test on regression, *Communications in Statistics: Theory and Methods*, 11, 639-651.

Saleh, A.K.Md.E. and Sen, P.K. (1983). Nonparametric tests of location after a preliminary test on regression in the multivariate case, *Communications in Statistics: Theory and Methods*, 12, 1855-1872.

Saleh, A.K.Md.E. (2006). *Theory of Preliminary test and Stein-type estimation with applications.* New Jersey: John Wiley & Sons.

Sen, P.K. (1982). On M-tests in linear models, *Biometrika*, 69, 245-248.

Tamura, R. (1965). Nonparametric inferences with a preliminary test, *Bull. Math. Stat.*, 11, 38-61.

Yunus, R M and Khan, S (2009). The bivariate noncentral chi-square distribution - a compound distribution approach, Working Paper Series, SC-MC-0902, Faculty of Sciences, University of Southern Queensland, Australia

Yunus, R.M. and Khan, S. (2010). Increasing power of the test through pre-test - a robust method. *Communications in statistics - theory & Method*, to appear.

Wright K. and Kennedy W.J. (2002). Self-validated computations for the probabilities of the central bivariate chi-square distribution and a bivariate F distribution, *J. Statist. Comput. Simul.*, 72, 63–75.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 246–246.

COINCIDENT TEST AND CONVERGENCE HYPOTHESIS: THEORY AND EVIDENCE

Samarjit Das[1]
Indian Statistical Institute

Ajoy Pal
Maulana Azad College

Manisha Chakrabarty
Indian Institute of Management Calcutta

ABSTRACT

The contribution of this paper is both methodological and empirical. It develops methodologies to examine the convergence hypothesis under cross sectional dependence in panel data framework. The proposed tests are shown to follow standard distributions and are free from nuisance parameters. A detailed Monte Carlo study is then carried out to evaluate the performance of these tests in terms of size and power. The tests are then applied to find the nature of convergence in real per capita income across various groups of countries. Findings are stunning in nature. There are no convergences across any group of countries including the EU and the OECD in terms of real per capita income. However, conditional convergence is observed among the members of both EU and OECD in terms of growth rates. Finally, absolute convergence in growth rates is evidenced only for the EU group.

Keywords: Absolute Convergence, coincident test, conditional convergence, cross-sectional dependence, panel unit root tests, common factor

[1]Corresponding author: Samarjit Das, ERU, Indian Statistical Institute, 203 B.T. Road, Kolkata-700108, India, E-mail: samarjit@isical.ac.in

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 247–247.

SESSION 6: BIOSTATISTICS

Chair: Nick Fieller
Department of Probability & Statistics
University of Sheffield
Sheffield, S3 7RH, U.K.
E-mail: n.fieller@sheffield.ac.uk

SUBDISTRIBUTION HAZARD: ESTIMATION AND INFERENCE

Ronald Geskus
Department of Clinical Epidemiology, Biostatistics and Bioinformatics
Academic Medical Center
Meibergdreef 15
1105 AZ, Amsterdam, The Netherlands
E-mail: R.B.Geskus@amc.uva.nl

ABSTRACT

Background: Fine and Gray (JASA, 1999) developed a regression model on the subdistribution hazard (SDH) in a competing risks setting with right censored data.

Objective: To extend estimation and inference on the SDH to left truncated data and time dependent covariates.

Methods: The standard estimator for the cause specific cumulative incidence (CSCI) is shown to have an equal representation as a product limit form based on an estimator of the SDH. This estimator differs from the standard estimator of the cause specific hazard only in the definition of the risk set: individuals with an earlier competing event remain in the risk set with a weight that depends on product-limit estimates of the censoring and truncation distributions. Using the counting process style notation, estimation of any aspect that is directly related to the SDH can be estimated using standard software, if the software allows for the inclusion of time-dependent weights. In a simulation study, results with right censored data only from the survival R package are compared with the cmprsk R package.

Results: Estimates of CSCI and SDH parameters are equal to results from the cmprsk package. Standard errors are different. For CSCI this is explained by a different choice of scale on which standard errors are calculated. For the SDH regression it is explained by the use of a robust estimator of the standard error in the cmprsk package. In our simulation study, the robust estimator performs worse, yielding too small coverage probabilities of the 95% confidence intervals.

Conclusion: No special software is needed for competing risks analyses. The standard survival R package has better performance with respect to standard errors and confidence intervals than the cmprsk package.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 248–254.

APPLICATION OF THE PROSPECTIVE SPACE-TIME SCAN STATISTIC FOR DETECTING MALARIA CASES HOTSPOTS IN BANGKA DISTRICT, INDONESIA

Asep Saefuddin[1] and Etih Sudarnika[2]
[1]Department of Statistics, Faculty of Mathematics and Science, IPB, 16680, Darmaga, Bogor, Indonesia
[2]Laboratory of Epidemiology, Faculty of Veterinary Medicine, IPB, 16680, Darmaga, Bogor, Indonesia
E-mail: [1]asaefuddin@gmail.com, [2]etih23@yahoo.com

ABSTRACT

Malaria is one of the complex health problems in Indonesia. The disease is still listed high priority due to its high mortality rate, especially among children under five years old, and its fatal impact on pregnant woman. The aim of this study was to apply cluster detection method for detecting disesase outbreaks in the malaria surveillance system in Bangka District, Bangka Belitung Province, Indonesia. Bangka is one of the malaria endemic areas in Indonesia. The cluster detection method used is the Prospective Kulldorff's space-time scan statistic. The research was conducted in a one year period, starting from September 2007 to September 2008. The malaria cases data were recorded from eleven public health center facilities in Bangka Districts, based on confirmed lab examinations. The results showed that the primary emerging cluster was Sinar Baru with a cluster period from January 2008 to July 2008. The secondary emerging cluster was Kenanga with a cluster period from January 2008 to July 2008. The conclusions were Sinar Baru and Kenanga were the areas which need further investigation and priority in the disease control and surveillance.

Keywords: malaria, space-time clustering, spatial statistics

1. INTRODUCTION

Background

Malaria is an infectious disease caused by a parasite from the *plasmodium* genus and is spread through the bite of the *Anopheles* mosquito. Its general symptoms include periodic fevers, anemia, bubonic, and other symptoms related to its effect on other organs such as the brain, liver, and kidneys (WHO 2007, Wikipedia 2007).

Indonesia is a country endemic of malaria. Indonesia is a tropical region with high rainfall and swampy topography which then enhance the life cycle of the Anopheles mosquito as the vector of malaria (Jamal, 2009). According to the Indonesian Malaria Endemic Map, it is predicted *that* in 2007 45% of Indonesia's population live in areas endemic with malaria. One of the areas endemic with malaria is the Bangka district, Bangka Belitung province. This area is categorized as medium endemic with malaria with an AMI of 29.3 per 1000 person in the year 2007 (Indonesian Health Department, 2008).

Generally, malaria is endemic in small villages with poor sanitation, bad transportation and communication facilities, hard to access healthcare services, low social economy levels, and unhealthy community lifestyles. Locations and social conditions like these are not uncommon in Indonesia. Aside to activities such as early diagnosis, quick and proper medication, *fogging*, and larva and vector control, surveillance is also an essential activity in the effort to reduce the number of cases and deaths caused by malaria.

Through surveillance, quick preventive actions can be conducted in certain malaria hotspots to better control the disease. Thus, it is necessary to conduct a geospatial statistical analysis to help analyze the surveillance data in order to find out when and where serious actions must be conducted, and so that the area gets first priority in handling malaria.

The objective of this research was to conduct a *space-time scan statistic* to detect malaria hotspots in Bangka district, Indonesia.

2. RESEARCH METHODS

2.1. Study Area

The Bangka district is located on Bangka Island covering an area of approximately 295.068 Ha populated by 237.053 people or 80 people per sq km. It has a tropical climate with rainfalls between 18.5 to 394.7 mm each month, reaching its lowest point in August. The temperature ranges between 26.2^{0}C to 28,3^{0}C and the humidity varies from 71% to 88%. The light intensity is around 18% to 66.1% and the air pressure varies from 1009.1 to 1011.1 mb.
The mining sector is one of Bangka's main sectors. The district is rich with tin and other mining goods and has a relatively high reserve. Thus, Bangka district has many abandoned mines containing ponds and puddles resulting from former mining activities. These ponds and puddles are potential breeding grounds for mosquitoes. The Bangka district has different topography, consisting of 4% hills, 25% swamps, and lowlands for the rest of the areas (Central Statistics Agency (BPS) and the Bangka District Regional Development Agency (BAPEDA) 2007).

Overall, Bangka's climate, topography, and environmental conditions are suitable for sustaining the *Anopheles* mosquito life cycle.

2.2. Data Collection and Management

Data collection was done monthly by collecting lab examination notes from every public health center (*puskesmas*) in the Bangka district. There were 11 *puskesmas* working areas spread throughout 8 sub districts in the Bangka district. They were *Puskesmas* Sungai Liat, Sinar Baru and Kenanga in sub district Sungai Liat, *Puskesmas* Bakam in sub district Bakam, *Puskesmas* Petaling in sub district Mendo Barat, *Puskesmas* Gunung Muda and Belinyu in sub district Belinyu, *Puskesmas* Pemali in sub district Pemali, *Puskesmas* Riau Silip in sub district Riau Silip, *Puskesmas* Puding Besar in sub district Puding Besar, and *Puskesmas* Batu Rusa in sub district Merawang. The definition of a malaria case in the study was a person positively diagnosed with malaria after a *Plasmodium* parasite lab check. Data resulting from clinical examinations and *Rapid Diagnostic Tests* were not considered as cases. If a *puskesmas* had a positive case but the person did not live in the area, then the case was moved to the *puskesmas* where the person lives. Each *puskesmas* location was mapped using Geography Positioning System (GPS) technology. The data used was the malaria cases that occured from June 2007 to July 2008. The demographic data used was obtained from the Central Statistics Agency (BPS) and the Bangka district Regional Development Agency (Bapeda). Every person in the area is assumed possible of being infected by malaria.

2.3. Statistical Analysis

The geospatial analysis used in this research is the *prospective space-time scans statistics* developed by Kulldorff in 1997. The data was assumed to spread in a Poisson spread and was analyzed using *SatScan* 7.0 software. Kulldorff 2001 mentioned that the prospective space-time scan statistics uses a cylindrical window in three dimensions. The base of cylinder represents space, whereas height represents time. The cylinder is flexible its circular geographical base as well as in its starting date, independently of each other. In mathematical notation, let $[Y_1, Y_2]$ be the time interval for which data exist, and let s and t be the start and end dates of the cylinder respectively. We then consider all cylinders for which $Y_1 \leq s \leq t \leq Y_2$ and $t \geq Y_m$, where Y_m is the time periodic surveillance began. Conditioning on the observed total number of cases, N, the definition of the spatial scan statistic S is the maximum likelihood ratio over all possible circles Z,

$$S = \frac{\max_z\{L(Z)\}}{L_0} = \max_z \left\{ \frac{L(Z)}{L_0} \right\} \tag{1}$$

where $L(Z)$ is the maximum likelihood for circle Z, expressing how likely the observed data are given a differential rate of events within and outside the zone, and where L_0 is the likelihood function under the null hypothesis.

Let n_z be the number of cases in circle Z. For the Poisson model, let $\mu(Z)$ be the expected number under the null hypothesis, so that $\mu(A) = N$ for A, the total region under study. It can then be shown that

$$\frac{L(Z)}{L_0} = \left\{ \frac{n_z}{\mu(Z)} \right\}^{n_z} \left\{ \frac{N - n_Z}{N - \mu(Z)} \right\}^{N - n_Z} \tag{2}$$

if $n_Z > \mu(Z)$ and $L(Z)/L_0 = 1$ otherwise.

3. RESULTS AND DISCUSSION

With its swampy and water surrounded ecology and also the presence of palm plantations and abandoned tin mines, Bangka district is a potential breeding ground for the *Anopheles* mosquito. Incidences of malaria vary according to *puskesmas* working area and observation period. Image 1 shows malaria incidence per 10000 people per year in the 11 *puskesmas* monitored from June 2007 until July 2008. The darker the color gradation, the higher the malaria incidence occur. During the observation period, the Bangka district had a total 1.778% malaria incidence. The area with the highest incidence is *puskesmas* Sinar Baru with 3.449% followed by *puskesmas* Kenanga with 3.138%. These two *puskesmas* are surrounded by beaches, an area potential for *Anopheles* breeding.

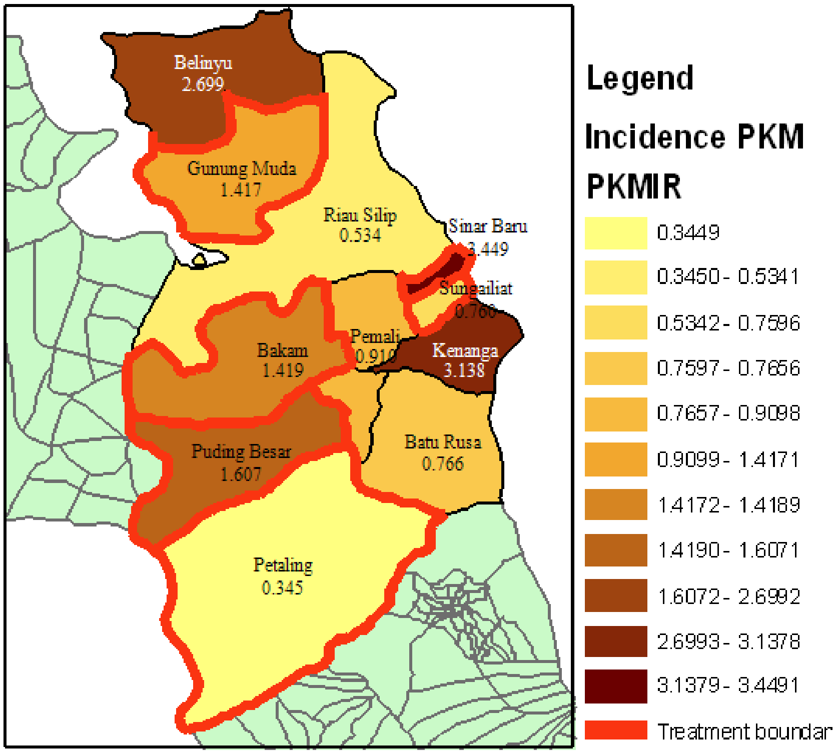


Image 1. Malaria Incidence per 10000 people in Bangka district *Puskesmas*, June 2007 – July 2008

The fluctuation of malaria incidence in every *puskesmas* working area in the Bangka district through June 2007 to July 2008 is shown in image 2. In the beginning of the observation, the malaria incidence in *puskesmas* Belinyu shows a significantly high value compared to the other regions, but the value declines during the rest of the observations. Other areas with high malaria incidence are Sinar Baru and Kenanga.

The result of the surveillance on malaria incidence per month during the observation period is spatially shown in image 3. The darker gradation shows areas with higher malaria incidence. Image 3 shows that spatial and temporal analysis can quickly be done by observing the color changes in each district every month. For example, in the beginning of the observation, Belinyu has the highest malaria incidence value, but shows a decline until the end of the observation period. The opposite occurs in Bakam, having a low value in the beginning but increasing until the end of the observation period.

Aside to descriptive data analysis by visualizing incidence maps, statistical analysis is also needed to determine malaria hotspot areas for immediate preventive actions. The statistical analysis was done by using a prospective space time scan statistic. The scan statistic is a statistical method designed to detect a local excess of events and to test if such an excess can reasonably have occurred by chance. Three basic properties of the scan statistic are the geometry of the area being scanned, the probability distribution generating events under the null hypothesis, and the shapes and sizes of the scanning window (Kulldorff 1997). The spatial scan statistic has the following features, which make it particularly suitable as a tool for surveillance purposes:

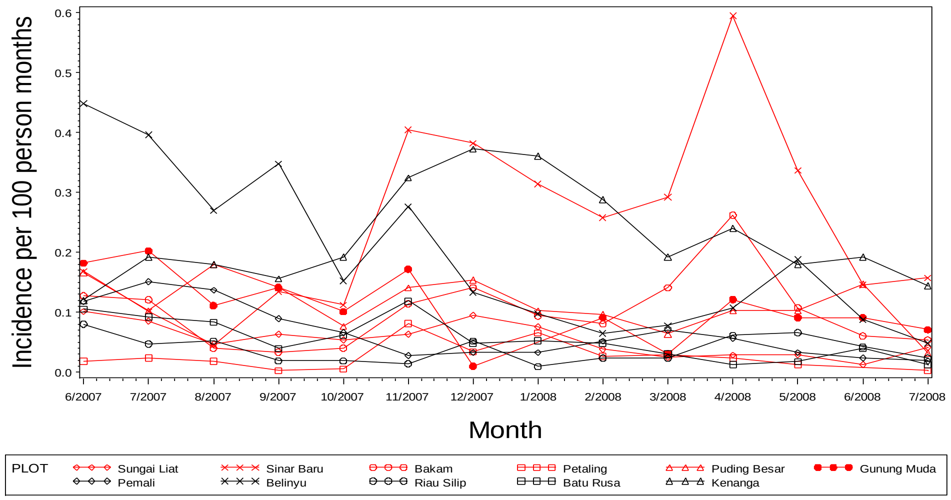


Image 2. Malaria Incidence per 100 people in Bangka District, June 2007 – July 2008

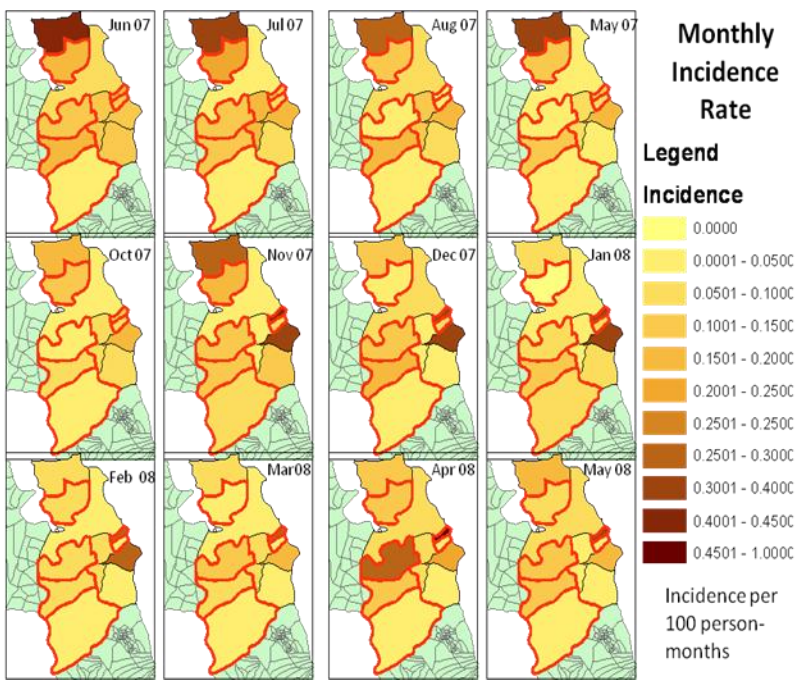


Image 3. Malaria Incidence per Month in every *Puskesmas* in Bangka district

a) it adjusts both for the inhomogeneous population density and for any number of confounding variables;
b) by searching for clusters without specifying their size or location the method ameliorates the problem of pre selection bias;
c) the likelihood-ratio-based test statistic takes multiple testing into account and delivers a single p-value for the test of the null hypothesis;
d) if the null hypothesis is rejected, we can specify the approximate location of the cluster that caused the rejection. (Kulldorff 2001).

By using a prospective space-time scan statistic analysis, observation period from June 2007 to July 2008, time aggregation length 1 month, maximum spatial cluster size 50% of population at risk, and maximum temporal cluster size 50% of study period, the results are shown in table 1.

Table 1. Detection of recently emerging clusters of malaria in Bangka

Most likely cluster	Cluster period	Cases	Expected	Relative risk	*p*
Primary cluster:					
Sinar Baru	2008/1/1 - 2008/7/31	187	56.76	3.455	0.001
Secondary cluster:					
Kenanga	2008/1/1 - 2008/7/31	133	53.05	2.581	0.001
Bakam	2008/3/1 - 2008/7/31	93	68.12	1.378	0.145

Table 1 show that the first emerging cluster is *puskesmas* Sinar Baru with a cluster period from January 1st 2008 to July 31st 2008 and a relative risk value 3.455. The second emerging cluster is *Puskesmas* Kenanga with 133 cases and a relative risk value 2.581. Another emerging cluster is *Puskesmas* Bakam, but this cluster does not have a significant p-value.

The working areas of *puskesmas* Sinar Baru and Kenanga are located in lowlands mainly consisting of coasts and swamps. These areas also have a lot of abandoned and unmaintained unauthorized tin mines that contain many ponds and puddles, the potential breeding ground for mosquitoes. These topographic conditions along with the residential areas neighboring with mosquito breeding grounds are the main cause of the high malaria incidence in these areas. The Bakam *puskesmas* working area shows low malaria incidence rates in the beginning of the observation period. The rates then increase in the following observations reaching its highest in April 2008, but there is insufficient data to categorize the area as an emerging cluster.

Quantitative analysis using prospective space time scan statistics allows us to discover the emerging clusters that need attention based on the surveillance results so that the authority can necessary priority actions in controlling the disease. The methods used to prevent and control malaria Includes installing mosquito nets, covering the body while sleeping, and using natural mosquito repellants. Nets with insecticides are available for children and pregnant woman in the Bangka district. The nets are distributed by UNICEF and the Health Department through *puskesmas*, *posyandu* (Integrated Community Health Service Center), and birth Clinics. Other methods include exterminating mosquito nests by cleaning up puddles and conducting *Indoor Residual Spraying* (RIS).

Kulldorf 2001 states bahwa the space-time scan statistic can serve as an important tool for systematic time periodic geographical disease surveillance. It is possible to detect emerging clusters, and we can adjust for multiple tests performed over time. No a priori hypothesis about cluster location, size or length need to be made. The method can be used at different levels of geographical and temporal aggregation, and for different types of disease. Although it is computer intensive, the method is not overly complex.

4. CONCLUSION

The primary emerging cluster is Sinar Baru with a time frame starting from January 2008 until July 2008. The secondary emerging cluster is Kenanga with a time frame starting from January

2008 until July 2008 and Bakam with a time frame starting from March 2008 until July 2008, but is not significant for Bakam. Sinar Baru and Kenanga were the areas which need further investigation and priority in the control.

ACKNOWLEDGMENTS

We thank the people of Bangka District, District Health Service officers, and their local authorities for their excellent cooperation. We would also like to thank Ministry of Health of Indonesia for the support, Center for Disease Control and Prevention (CDC) and UNICEF for financial support. Last but not least, a word of thanks to all the team members.

REFERENCES

Badan Pusat Statistik Kabupaten Bangka (2007). *Bangka Dalam Angka 2006*. Bangka: BPS & BAPEDA Kab. Bangka.

Jamal S. (2009). Apa yang perlu anda ketahui tentang malaria?

http://www.isfinational.or.id/pt-isfi-penerbitan/124/443-apakah-malaria-itu- [3/16/2009].

Kulldorff, M. (1997). *A Spatial Scan Statistic*. Communs Statist. Med., **26**. 1481-1496.

Kulldorff, M. (2001). *J. R. Statist. Soc. A*, **164**. Part 1. 61-72

Ministry of Health of Republik Indonesia (2008). *Indonesian Health Profile*. Jakarta: MOH RI.

World Health Organization (2007). *Malaria*. http://www.who.int/ topics/malaria/en/ [6/23/22007]

Wikipedia (2007). Malaria. http://en.wikipedia.org/wiki/Malaria [6/23/22007]

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 255–263.

TWO-STAGE TESTING IN THREE-ARM NON-INFERIORITY TRIALS

Nor Afzalina Azmee and Nick Fieller
Department of Probability and Statistics, University of Sheffield, Hicks Building, Hounsfield Road, S3 7RH, UK
E-mail: Afzalina.Azmee@sheffield.ac.uk

ABSTRACT

The aim of a non-inferiority trial is to show that a new experimental treatment is not worse than the reference treatment by more than a certain, pre-defined margin. We consider the design of a 3-arm non-inferiority trial, where the inclusion of a placebo group is permissible. The widely used 3-arm non-inferiority procedure was authoritatively first described by Pigeot et al. (2003), which involved establishing superiority of reference against placebo in the first stage before testing non-inferiority of experimental against reference in the second stage. If this preliminary test fails, the second-stage test has to be abandoned. In such an eventuality, we believe the whole study will be wasted as nothing new could be learnt about the new experimental treatment. Therefore, instead of showing superiority in the first stage, we propose that the reference treatment has to be significantly different from placebo as a pre-requisite before using Fieller's confidence interval to assess non-inferiority. This procedure leads to no peculiar intervals (i.e. exclusive or imaginary) and offers easy interpretation regarding the efficacy of experimental and reference treatments.

Keywords: non-inferiority trials, ratio estimation, Fieller's theorem, phase III trial

1. INTRODUCTION

Non-inferiority trials are considered to be the new emerging area in medical statistics. They were first reported about twenty years ago and since then they have received much attention from researchers from both medical and statistical backgrounds. The first paper related to non-inferiority trial was written by Blackwelder (1982), describing the statistical approach to demonstrate therapeutic equivalence. Given that the reference treatment has already been established, the aim will be to conduct either an active-controlled superiority trial or an active-controlled non-inferiority trial. The trend has seen fewer new drugs emerge and so superiority trials are now being replaced by non-inferiority trials. The aim of a non-inferiority trial is to show that the new treatment is not worse relative to the standard treatment by more than a predefined margin. The ultimate aim is to enable the new treatment with similar efficacy to the standard one to be marketed.

Pigeot et al. (2003) proposed that if the efficacy of a new experimental treatment has to be shown in the situation where an active control exists, it is often sufficient to show non-inferiority of the new experimental treatment with respect to the primary clinical outcome. This idea is generally supported by Koch and Röhmel (2004), although in their paper, the statistical test procedure differs slightly from that proposed by Pigeot et al. (2003). It is important to appreciate that a non-inferiority trial cannot be perceived as an easy gate-way for the pharmaceutical companies to introduce their drugs to the market. If the efficacy of the new treatment is not clearly superior to that of the reference treatment, then the new treatment has to possess other advantages over the reference treatment. For example, the new treatment could be cheaper, easier to be administered or do not exhibit serious adverse effects when compared to the current standard treatment.

A simple non-inferiority trial can be conducted by having just experimental and reference arms. In this 2-arm non-inferiority trial, assay sensitivity poses a particular problem which is not observed in a superiority trial. Assay sensitivity is defined as the ability of a trial to demonstrate the difference between treatments. Without an additional placebo arm, the result stating that the experimental treatment has the same effect with the reference treatment is difficult to interpret. The investigator would not be confident in deciding whether the treatments are truly similarly effective or similarly ineffective. Design issues such as, choosing a non-inferiority margin are constantly faced in 2-arm non-inferiority trials. Issues regarding the choice of non-inferiority margin are discussed by D'Agostino et al. (2003), CHMP (2005), Lange & Freitag (2005), Hung et al. (2005) and Brown et al. (2006).

These problems of interpretation in a 2-arm trial can be eliminated by introducing an extra placebo arm. The inclusion of a placebo arm is discussed in Temple & Ellenberg (2000), in the situations where it is acceptable to administer a placebo instead of the reference drug even though superiority of the reference over placebo had presumably previously been established. Such a circumstance might arise with a non-mortal chronic condition or where general medical procedures have improved so much that there is now uncertainty that the reference is still superior to the current "placebo" treatment of just general medical care.

The arguments justifying the addition of a placebo group are also given in Koch and Röhmel (2004). The inclusion of a placebo group is strongly recommended when the reference is traditional, weak, volatile or the disease has not been understood properly. To make it even more credible, the guideline of CHMP (2005) has declared that the 3-arm non-inferiority trial involving experimental, reference and placebo groups should be used wherever possible.

This paper concentrates on such a design, the conduct of two-stage testing and the implementation of Fieller's theorem in finding the intervals of the ratio. We will start by giving a brief account on the construction of Fieller's interval, followed by an account of our two-stage testing procedure. Our procedure requires unequal means of reference and placebo to be established in the first stage, which is a slight modification from the idea outlined in Pigeot et al. (2003). To assess the behaviour of our procedure, a simulation study is conducted and is detailed in Section 4 of this paper. Summary of the findings is discussed in the last section.

2. RATIO ESTIMATION

Assuming that higher treatment corresponds to better efficacy, the statistical testing problem in the three-arm non-inferiority trial starts with:

$$H_0 : \mu_E - \mu_R \leq \delta \qquad \text{versus} \qquad H_0 : \mu_E - \mu_R > \delta \tag{1}$$

where μ_E and μ_R represent the population means of experimental and reference groups and δ is non-inferiority margin. In practice, δ is usually specified as a fraction of the effect size $(\mu_R - \mu_P)$ where μ_P denotes the population mean for placebo group. Equation (1) can be written as:

$$H_0 : \mu_E - \mu_R \le f\left(\mu_R - \mu_P\right) \quad \text{versus} \quad H_0 : \mu_E - \mu_R > f\left(\mu_R - \mu_P\right) \tag{2}$$

Here f represents the fraction of reduction. In this paper, we consider, $f = -1/5$ to represent a 20% reduction considered tolerable. In practice, the reduction varies based upon statistical reasoning and clinical judgment. A simple algebraic manipulation to tidy up the equation above reveals a ratio form of hypothesis tested.

$$H_0 : \frac{\mu_E - \mu_P}{\mu_R - \mu_P} \le \theta \quad \text{versus} \quad H_0 : \frac{\mu_E - \mu_P}{\mu_R - \mu_P} > \theta \tag{3}$$

where θ = 1 – f. The aims of a study could be switched simply by setting different values of θ (Koch & Tangen, 1999). As an example, when θ is set as 0, it implies testing the superiority of experimental over placebo. On the other hand, if θ is chosen as 1, it corresponds to testing superiority of experimental over reference.

Statistically equivalent to testing the hypothesis on the value of the ratio we can instead consider an estimate of the ratio and an associated confidence interval with coefficient complementary to the specified size of the hypothesis test, e.g. a 95% confidence interval for a size 0.05 test. The null hypothesis is not rejected if and only if the specified value of θ is included in the interval. To find the confidence interval for the ratio, we follow the result discussed in Fieller (1954) and Finney (1971). Here, two parameters are considered, namely α and β such that $\mu = \alpha/\beta$ and $m = a/b$ is the estimate of μ. In this setting, a and b are assumed to be normally distributed with the expectations, variances and covariance, given as follows:

$$E(a) = \alpha\,;\ E(b) = \beta\,;\ \operatorname{var}(a) = v_{11}\hat{\sigma}^2\,;\ \operatorname{var}(b) = v_{22}\hat{\sigma}^2\,;\ \operatorname{cov}(a,b) = v_{12}\hat{\sigma}^2$$

where $\hat{\sigma}^2$ is the unbiased estimator of the common variance σ^2 and v_{11}, v_{22}, v_{12} are the reciprocal of the sample sizes. The lower and upper limits of the interval for a ratio could be obtained by solving the roots of a quadratic equation. The final result of Fieller's theorem is given as follows:

$$\theta_L = \frac{1}{(1-g)}\left[\frac{a}{b} - g\frac{v_{12}}{v_{22}} - \frac{t\hat{\sigma}}{b}\sqrt{v_{11}(1-g) - 2v_{12}\frac{a}{b} + v_{22}\frac{a^2}{b^2} + \frac{v_{12}^2}{v_{22}}g}\right] \tag{4}$$

$$\theta_U = \frac{1}{(1-g)}\left[\frac{a}{b} - g\frac{v_{12}}{v_{22}} + \frac{t\hat{\sigma}}{b}\sqrt{v_{11}(1-g) - 2v_{12}\frac{a}{b} + v_{22}\frac{a^2}{b^2} + \frac{v_{12}^2}{v_{22}}g}\right] \tag{5}$$

where $g = t^2\hat{\sigma}^2 v_{22}/b^2$ and t is the two-sided α percentage point of Student's t-distribution. In the context of three-arm non-inferiority trials, we shall have the followings:

$$\alpha = \mu_E - \mu_P\,;\ \beta = \mu_R - \mu_P\,;\ \ a = \overline{x}_E - \overline{x}_P\,;\ b = \overline{x}_R - \overline{x}_P\,;$$

$$v_{11} = 1/n_E + 1/n_P\,;\ v_{22} = 1/n_R + 1/n_P\,;\ v_{12} = 1/n_P\,;$$

$$\hat{\sigma}^2 = s_E^2\left(n_E - 1\right) + s_R^2\left(n_R - 1\right) + s_P^2\left(n_P - 1\right)\Big/ n_E + n_R + n_P - 3$$

Here $\bar{x}_E$, $\bar{x}_R, \bar{x}_P, s_E^2, s_R^2, s_P^2, n_E, n_R$ and n_P refer to sample means, sample variances and sample sizes of experimental, reference and placebo groups respectively.

Although the implementation of Fieller's theorem is in general considered useful, it does not escape from criticisms. Several authors have noted the erratic behaviour of Fieller's interval. Under certain circumstances, the interval can be exclusive or imaginary. However, the occurrences of unexplainable intervals can be prevented when the two-stage procedure is implemented.

3. INEQUALITY OF MEANS AS A PRE-REQUISITE

The assessment of non-inferiority given by Pigeot et al. (2003) involves a two-stage hierarchical procedure. The first stage testing requires the null hypothesis $H_{01}: \mu_R \leq \mu_P$ to be rejected first at level α. Koch and Röhmel (2004) expressed their concern, where failing to prove superiority of reference against placebo in the first stage, would put the experimental treatment at severe disadvantage. The analysis then would have to be terminated at stage one. We support their view as we feel that something useful should be gained out of expensive Phase III clinical trials. At least, the question regarding the efficacy of the new treatment should be answered. Based on this justification, we propose another type of two-stage procedure, where the pre-requisite is to show that reference treatment and placebo have unequal means. This can be seen as a loose restriction compared to the one recommended by Pigeot et al. (2003).

There are two important grounds as why the decision to introduce preliminary test of unequal means is made. The first reason is attributed to the way Fieller's theorem is constructed. Based on the equation, this theorem cannot be used if the denominator *b* is 0 or more specifically is not significantly different from 0. Our initial exploration has shown that denominator *b* is at the root of all problematic intervals. It affects the behaviour of the term *g* and this in turn affects the occurrence of exclusive or imaginary intervals. Therefore, $\mu_R \neq \mu_P$ has to be established in the first place before testing for non-inferiority. Secondly, by proposing this preliminary test, we have covered scenarios of having the reference that is not superior to placebo. In the situation where the reference is worse but the experimental is better, both with respect to placebo, our two-stage procedure offers an advantage of learning something about the experimental treatment. The procedure also ensures that no peculiar intervals will be obtained in the second stage.

The procedure starts with testing $H_{01}: \mu_R = \mu_P$ against $H_{01}: \mu_R \neq \mu_P$ at a 2-sided, 5% significance level. If the null hypothesis is rejected, then only the procedure continues to testing non-inferiority of the new treatment. Fieller's theorem is used to calculate the lower and upper limits of 95% confidence interval, given by Equation (4) and (5). This is the established result, assuming that the direction of reference is positive (i.e. reference is superior to placebo). The interval is then inclusive as $\theta_L < \theta_U$. However, when the situation is reversed, the lower and upper limits are going to be reversed as well. When the standard result of Fieller's theorem is used in this situation, the confidence interval obtained will be "seem-exclusive" as $\theta_L > \theta_U$.

Pigeot et al. (2003) have shown that performance of an experimental treatment can still be studied even though a non-inferiority trial has failed. This however restricts to situations where the reference is found superior to placebo. On the other hand, our slight modification of Fieller's theorem enable us to make useful conclusion about an experimental treatment, given the reference is superior to placebo or vice versa. The inference is based by looking at the behaviour of Fieller's interval, either inclusive (i.e. $\theta_L < \theta_U$) or seems-exclusive (i.e. $\theta_L > \theta_U$), as summarized in Table I.

Table 1 Once the preliminary test is conducted, the inference is solely based on the behavior of Fieller's interval

Outcomes	Inference
• **Inclusive intervals**	• **Reference is superior to placebo**
○ $\theta_L < 0$	○ Non-inferiority failed; experimental is not superior to placebo
○ $0 < \theta_L < \theta$	○ Non-inferiority failed; experimental is superior to placebo
○ $\theta_L > \theta$	○ Non-inferiority established; experimental is superior to placebo
○ $\theta_L > 1$	○ Non-inferiority established; experimental is superior to reference
• **Seems-exclusive intervals**	• **Reference is not superior to placebo**
○ $\theta_L < 0$	○ Non-inferiority failed; experimental is superior to placebo
○ $\theta_L > 0$	○ Non-inferiority failed; experimental is not superior to placebo

Once θ_L and θ_U have been calculated, the next step is to look at whether they are inclusive or seems-exclusive. The behavior of an experimental treatment with respect to placebo will then be determined by looking at θ_L.

4. EXAMPLES

The results of a simulation study are separated into three different tables; Table II corresponds to situations where the reference is truly superior compared with the placebo, Table III simulates situations where the reference is made to be equal to placebo and Table IV looks at how the procedure behaves when the reference is truly worse than placebo. For each table, different locations of experimental treatment are chosen to reflect situations where experimental is superior, equal or worse than placebo. We compare the use of equal and different allocations of sample size, as suggested by some authors. The effect of using different sample size allocations is prominent in the scenario where experimental treatment is truly non-inferior to reference.

For the purpose of understanding the behaviour of our proposed two-stage procedure, the total sample size is arbitrarily chosen as 450. There are 7 different outcomes wished to be detected by simply looking at Fieller's confidence intervals. The number of occurrences for each category is summarized in the following tables:

Table 2 Configurations chosen are $\mu_R = 4.5$ and $\mu_P = 3$ such that $\mu_R > \mu_P$ with total sample size N = 450 and variances set as 1 for all groups.

μ_E	Sample Size Allocation	First Stage Failed	$\mu_R > \mu_P$ $\mu_E \le \mu_P$	$\mu_R > \mu_P$ $\mu_E > \mu_P$	E non-inferior to R	E superior to R	$\mu_R \le \mu_P$ $\mu_E > \mu_P$	$\mu_R \le \mu_P$ $\mu_E \le \mu_P$
4.5	1:1:1	-	-	184	791	25	-	-
	2:2:1	-	-	165	808	27	-	-
	3:2:1	-	-	135	845	20	-	-
3	1:1:1	-	972	28	-	-	-	-
	2:2:1	-	974	26	-	-	-	-
	3:2:1	-	979	21	-	-	-	-
1.5	1:1:1	-	1000	-	-	-	-	-
	2:2:1	-	1000	-	-	-	-	-
	3:2:1	-	1000	-	-	-	-	-

Because of the relative values of reference and placebo made to be truly superior to the placebo, the null hypothesis of equal means is rejected for all 1000 simulated data sets, for each different setting of μ_E and sample size allocation. For the first case of $\mu_E = 4.5$, we would naturally expect that the procedure will be able to declare superiority of an experimental treatment. This is clearly demonstrated in the third, fourth and fifth columns of the possible outcomes. The sample size chosen is arbitrarily large enough to ensure power of detecting non-inferiority of the new treatment is more than 80%. Since we are also interested in the efficacy of experimental treatment with respect to placebo, a danger arises from making the wrong conclusion.

The second case here where experimental treatment is similar to placebo (i.e. $\mu_E = 3$), at most 28 cases of declaring superiority of an experimental treatment is recorded. Further investigation has shown the mistake however is controlled at the chosen significance level, 2.5%. Note that when an experimental treatment is truly not superior to placebo (i.e. $\mu_E = 1.5$), there is no mistake of declaring superiority of an experimental treatment. The type I error, defined as declaring an inferior new treatment as non-inferior is not observed at all. If the procedure outlined by Pigeot et al. (2003) is used, similar results to the one observed in Table I will be obtained.

Table 3 Configurations chosen are $\mu_R = 3$ and $\mu_P = 3$ such that $\mu_R = \mu_P$ with total sample size N = 450 and variances set as 1 for all groups.

μ_E	Sample Size Allocation	First Stage Failed	$\mu_R > \mu_P$ $\mu_E \le \mu_P$	$\mu_R > \mu_P$ $\mu_E > \mu_P$	E non-inferior to R	E superior to R	$\mu_R \le \mu_P$ $\mu_E > \mu_P$	$\mu_R \le \mu_P$ $\mu_E \le \mu_P$
4.5	1:1:1	952	-	-	-	24	24	-
	2:2:1	958	-	-	-	21	21	-
	3:2:1	950	-	-	-	22	28	-
3	1:1:1	943	15	7	-	-	-	35
	2:2:1	956	14	5	-	-	-	25
	3:2:1	954	15	3	-	-	-	28
1.5	1:1:1	948	25	-	-	-	-	27
	2:2:1	955	23	-	-	-	-	22
	3:2:1	950	29	-	-	-	-	21

Given that reference treatment is similar to placebo (i.e. $\mu_R = \mu_P$), about 5% of the simulated data sets are expected to pass through the first stage. The type I error rate is observed in the first setting of $\mu_E = 4.5$ where at most 24 cases are declared to have successful non-inferiority tests. Since the lower limit of Fieller's interval exceed the non-inferiority margin, set at 0.8, in those 24 cases we conclude that the experimental is superior to the reference treatment, illustrated in the fifth column of possible outcomes. The type I error is shown to be controlled at the chosen significance level.

On the other hand, there is no type I error observed in the second setting of $\mu_E = 3$. However, there is a small possibility for making the wrong conclusion regarding the efficacy of experimental treatment with respect to placebo. Our investigation has revealed that when the reference, experimental and placebo are truly equal (i.e. $\mu_R = \mu_E = \mu_P$), concluding experimental treatment as superior to placebo will be around 1%. It is safe to say that this error will be at most, around 2.5%. In the third setting where $\mu_E = 1.5$ no such mistakes of type I error or drawing superiority of experimental treatment is made.

In the setting described in Table IV, we will miss the chance of drawing a useful conclusion about the efficacy of the experimental treatment when superiority of reference against placebo is made as a pre-requisite. We tackle this issue by imposing a slightly loose restriction in the first stage. As long as the reference treatment is not similar to placebo, the evaluation can still be carried out in the second stage. Fieller's theorem can still be implemented to assess either non-inferiority of experimental treatment with respect to reference or to assess superiority of experimental treatment with respect to placebo.

Table 4 Configurations chosen are $\mu_R = 1.5$ and $\mu_P = 3$ such that $\mu_R < \mu_P$ with total sample size *N* = 450 and variances set as 1 for all groups.

μ_E	Sample Size Allocation	First Stage Failed	$\mu_R > \mu_P$ $\mu_E \le \mu_P$	$\mu_R > \mu_P$ $\mu_E > \mu_P$	E non-inferior to R	E superior to R	$\mu_R \le \mu_P$ $\mu_E > \mu_P$	$\mu_R \le \mu_P$ $\mu_E \le \mu_P$
4.5	1:1:1	-	-	-	-	-	1000	-
	2:2:1	-	-	-	-	-	1000	-
	3:2:1	-	-	-	-	-	1000	-
3	1:1:1	-	-	-	-	-	22	978
	2:2:1	-	-	-	-	-	23	977
	3:2:1	-	-	-	-	-	24	976
1.5	1:1:1	-	-	-	-	-	-	1000
	2:2:1	-	-	-	-	-	-	1000
	3:2:1	-	-	-	-	-	-	1000

The result for $\mu_E = 4.5$ clearly demonstrates the ability of our procedure to detect superiority of the experimental treatment with respect to placebo. In the second setting where $\mu_E = 3$ we expect the error of concluding superiority of experimental treatment is controlled around 2.5%. The third setting where $\mu_E = 1.5$ generated no problem in correctly categorizing the efficacy of experimental and reference treatments. We have also investigated the effect of reducing the sample size, just to ensure that the procedure suggested will not lead to having exclusive or imaginary confidence intervals. The results have shown that by establishing inequality means of reference and placebo, no peculiar intervals will be observed in the second stage.

5. COMMENTS AND CONCLUSIONS

A non-inferiority trial can only be conducted if a good reference treatment exists. In practice, however, we might come across situations where the assumption of a good reference is not met. We might have thought the reference treatment established in the past still has the same effect, when in fact within this time span, the overall medical procedures have improved substantially. The study has shown that a weak reference treatment could cause the two-stage three-arm procedure given by Pigeot et al (2003) to stop prematurely and a good new experimental drug will not even be evaluated. Plausible situations will be when people develop drug resistance, as reported in the bulletin of the World Health Organization (Geerligs, Brabin and Eggelte, 2003).

Therefore a preliminary test to establish unequal means enable useful conclusions to be obtained regarding the efficacy of experimental treatment. When the null hypothesis of equal means is not rejected, nothing could be learnt about the experimental treatment. In this situation, we could consider performing a superiority test of experimental with respect to placebo. This however introduces problem of multiplicity and is not pursued here.

REFERENCES

Blackwelder, W.C. (1982). "Proving the null hypothesis" in clinical trials. *Controlled Clinical Trials*, 3, 345-353.

Brown, D., Volkers, P. and Day, S. (2006). An introductory note to CHMP guidelines: choice of the non-inferiority margin and data monitoring committees. *Statistics in Medicine*, 25, 1623-1627.

CHMP- Committee for Medicinal Products for Human Use (2005). *Guideline on the choice of the non-inferiority margin*. London: European Medicines Agency. Available from: http://www.tga.gov.au/docs/pdf/euguide/ewp/215899en.pdf

D'Agostino, R.B., Massaro, J.M. and Sullivan, L.M. (2003). Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Statistics in Medicine*, 22, 169-186.

Fieller, E.C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society (Series B)*, 16, 175-185.

Finney, D.J. (1971). *Statistical method in biological assay*. 2nd ed. London:Griffin, 83-98.

Geerligs, P.D.P., Brabin, B.J. and Eggelte, T.A. (2003). Analysis of the effects of malaria chemoprohylaxis in children on haematological responses, morbidity and mortality. *Bulletin of the World Health Organization*, 81, 205-216.

Hung, H.M.J., Wang, S.J. and O'Neill, R. (2005). A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials. *Biometrical Journal*, 47, 28-36.

Koch, A. and Röhmel, J. (2004). Hypothesis testing in the “gold standard” design for proving the efficacy of an experimental treatment relative to placebo and a reference. *Journal of Biopharmaceutical Statistics*, 14, 315-325.

Koch, G.G. and Tangen, C.M. (1999). Nonparametric analysis of covariance and its role in noninferiority clinical trials. *Drug Information Journal*, 33, 1145-1159.

Lange, S. and Freitag, G. (2005). Special invited papers section: therapeutic equivalence – clinical issues and statistical methodology in noninferiority trials. *Biometrical Journal*, 47, 12-27.

Pigeot, I., Schäfer, J., Röhmel, J. and Hauschke, D. (2003). Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine*, 22, 883-899.

Temple, R. and Ellenberg, S.S. (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: Ethical and Scientific Issues. *Annals of Internal Medicine*, 133, 455-463.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 264–264.

SESSION 7: ASTROSTATISTICS

Chair: Asis Kumar Chattopadhyay
Department of Statistics, Calcutta University, India
E-mail: akcstat@caluniv.ac.in

STUDY OF NGC 5128 GLOBULAR CLUSTERS UNDER MULTIVARIATE STATISTICAL PARADIGM

Asis Kumar Chattopadhyay
Department of Statistics, Calcutta University, India

Tanuka Chattopadhyay
Department of Applied Mathematics, Calcutta University, India

Emmanuel Davoust
Laboratoire d'Astrophysique de Toulouse-Tarbes, Universite de Toulouse, France

Saptarshi Mondal
Department of Statistics, Calcutta University, India

Margarita Sharina
Special Astrophysical Observatory, Russian Academy of Sciences, Russia

ABSTRACT

An objective classification of the globular clusters of NGC 5128 has been carried out by using a model based approach of cluster analysis. The set of observable parameters includes structural parameters, spectroscopically determined Lick indices and radial velocities from the literature. The optimum set of parameters for this type of analysis is selected through a modified technique of Principal Component Analysis, which differs from the classical one in the sense that it takes into consideration the effects of outliers present in the data. Then a mixture model based approach has been used to classify the globular clusters into groups. On the basis of the above classification scheme three coherent groups of globular clusters have been found. We propose that the clusters of one group originated in the original cluster formation event that coincided with the formation of the elliptical galaxy, and that the clusters of the two other groups are of external origin, from tidally stripped dwarf galaxies on random orbits around NGC 5128 for one group, and from an accreted spiral galaxy for the other.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 265–265.

STATISTICS OF THE BRIGHTEST YOUNG STAR CLUSTERS

Dean E. McLaughlin
Keele University, UK
E-mail: dem@astro.keele.ac.uk

ABSTRACT

There is a well-established correlation between the global star formation rate in local disk galaxies and starbursts, and the luminosity of the brightest young star clusters in the galaxies. I will discuss how this correlation may be used to derive information about the distribution of young cluster masses, its detailed shape and its possible universality, and how this in turn may be used to improve our understanding of the formation of young cluster systems and the long-term dynamical evolution that connects them to the systems of old globular star clusters in galaxy haloes.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 266–279.

CLASSIFICATION OF GAMMA-RAY BURSTS

Tanuka Chattopadhyay
Department of Applied Mathematics,
Calcutta University, 92 A.P.C. Road, Calcutta 700009, India.
E-mail: tanuka@iucaa.ernet.in

ABSTRACT

Two different multivariate clustering techniques, the K-means partitioning method and the Dirichlet process of mixture modeling, have been applied to the BATSE Gamma-ray burst (GRB) catalog, to obtain the optimum number of coherent groups. In the standard paradigm, GRB are classified in only two groups, the long and short bursts. However, for both the clustering techniques, the optimal number of classes was found to be three, a result which is consistent with previous statistical analysis. In this classification, the long bursts are further divided into two groups which are primarily differentiated by their total fluence and duration and hence are named low and high fluence GRB. Analysis of GRB with known red-shifts and spectral parameters suggests that low fluence GRB have nearly constant isotropic energy output of 10^{52} ergs while for the high fluence ones, the energy output ranges from 10^{52} to 10^{54} ergs. It is speculated that the three kinds of GRBs reflect three different origins: mergers of neutron star systems, mergers between white dwarfs and neutron stars, and collapse of massive stars. In this respect classification of Swift data without and with redshift values in three groups show similar properties with the groups found with BATSE data.

1. INTRODUCTION

Although it has now been well established that Gamma-Ray Bursts (GRB) are of cosmological origin, their nature and source still remains a mystery. Detailed observations and studies of their afterglow emission have revealed important information regarding the dynamic features and environments of these explosive events (see Piran (2005) for a review). The detection of supernova light curve in the afterglows of long duration nearby GRB has indicated that a fraction of the GRB occur during the collapse of a massive star (see Woosley & Bloom (2006) for a review). Other mechanism that could produce GRB are the merger of compact objects like a pair of neutron stars or a neutron star with a black hole (e.g. Piran, 1992; Gehrels et al., 2005; Bloom et al., 2006). Thus GRB may be a heterogeneous group and a proper classification of the phenomena is crucial to isolate and identify the possible different sources. Such a classification will also enable the identification of spectral or temporal correlations which may exist only for a particular class of GRB.

In the next section, we briefly describe the two classification schemes and present the result of the analysis on the BATSE catalog. In §3, the classification obtained from the BATSE data are used to classify GRB with known red-shifts and inferences are made on the intrinsic properties of the different GRB groups.In section §4 the Swift data with and without redshift values are classified using K means algorithm assuming K=3. In §5 the work is summarized and the main results are discussed.

2. CLUSTERING ANALYSIS FOR GRB DATA

The BATSE catalog provides temporal and spectral information for more than 1500 GRB. The parameters include, two measures of burst durations, the times within which 50% (T_{50}) and 90% (T_{90}) of the flux arrive, three peak fluxes, $P_{64}, P_{256}, P_{1024}$ measured in 64, 256 and 1024 ms bins respectively, four time integrated fluences $F_1 - F_4$, in the $20-50$, $50-100$, $100-300$ and > 300 KeV spectral channels. Many of the parameters are highly correlated and following previous works (e.g. Mukherjee et al., 1998; Hakkila et al., 2000) we use the following six parameter set: $\log T_{50}$, $\log T_{90}$, $\log P_{256}$, $\log F_T$, $\log H_{32}$, $\log H_{321}$, where $F_T = F_1 + F_2 + F_3 + F_4$ is the total fluence while $H_{32} = F_3/F_2$ and $H_{321} = F_3/(F_1 + F_2)$ are measures of spectral hardness. The sample consists of 1594 GRB that have non-zero detections of these parameters. We have not introduced any completeness criteria (like a lower flux cutoff), since incompleteness primarily affects the short duration bursts and hence is not expected to change the qualitative results obtained. We retain the F_4 flux (in the definition of F_T), despite the uncertainties in its calibration and sensitivity, because as we discuss later in §5, the γ-ray flux $>$ 300 keV is expected to have important spectral information. However, this fluence is not used in the computation of spectral hardness.

The Swift catalog provides $T90$, integrated Fluence in the 15-150 KeV spectral channel and peak photon flux in the 15-150 Kev spectral channel(in ph/cm^2/sec). Both the catalogs with and without red shift(z) values are taken. The sample size are 148 and 452 respectively in the former and latter cases.

2.1 Partitioning (K-means Clustering) Method

In this work, we apply the K- means method of MacQueen (1967) which is probably the most widely technique, to the BATSE catalog. For this method, the optimum value of K can be obtained in different ways (Hartigan , 1975). This is done by computing for each cluster formation (i.e for number of clusters $K = 2, 3, 4..$) a distance measure $d_K = (1/p) min_x E[(x_K - c_K)'(x_K - c_K)]$ which is defined as the distance of the x_K vector (values of the parameters) from the center of a cluster c_K (which is estimated as mean value). p is the order of the x_K vector, i.e. the number of parameters which for our case is six. If d'_K is the estimate of d_K at the K^{th} point, then d'_K is the minimum achievable distortion associated with fitting K centers to the data. A natural way of choosing the number of clusters is to plot d'_K versus K and look for the resulting distortion curve. This curve will monotonically decrease with increasing K, till K is greater than true number of clusters,

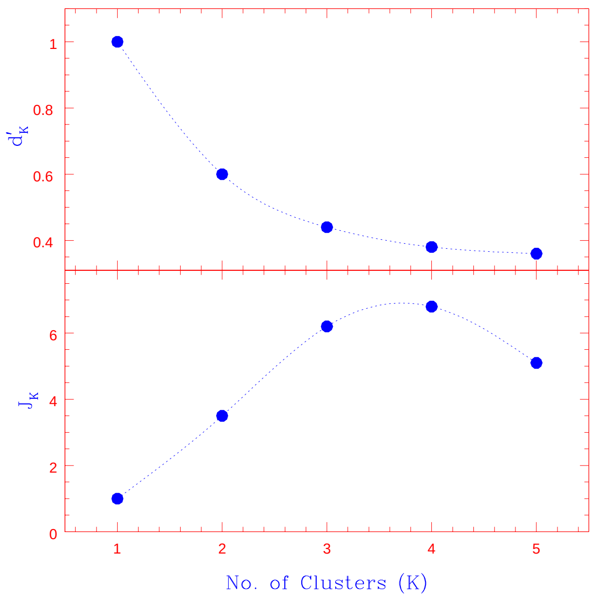


Figure 1: The distortion curve d'_K and the transformed jump curve J_K for different number of clusters. The largest value of J_K and the leveling of d'_K indicates that for the K-means clustering technique the optimal number of clusters is three.

after which the curve will level off with a smaller slope. This is expected since adding more clusters beyond the true number, will simply create partitions within a group. According to Sugar & James (2003) it is more illustrative to consider the transformation of the distortion curve to an appropriate negative power, $J_K = (d'^{-(p/2)}_K - d'^{-(p/2)}_{K-1})$, which will exhibit a sharp "jump" when K equals the true number of clusters. The optimum number of clusters is the value of K at which the distortion curve levels off as well as its value associated with the largest jump for the transformed curve. Figure 1 shows the distortion curve and transformed jump curve for the analysis of the BATSE data. The leveling of the distortion curve and the shape of the jump function strongly suggests that the optimum number of clusters is greater than two and is likely to be three. The group means and the standard errors for the six parameters for three cluster classification, are given in Table 1. Cluster I with 423 members has an average $< T_{90} > \sim 0.5$ s can be clearly identified with the short duration bursts. The long duration bursts are cleanly separated into two clusters (clusters 2 and 3) with 622 and 549 members.

Table 1: Average Cluster properties based on K-means classification

Parameters	Cluster I	Cluster II	Cluster III
T_{50} (sec)	0.19 ± 0.01	5.37 0.25	22.9 1.0
T_{90} (sec)	0.50 0.02	15.85 0.73	63.1 2.9
P_{256} (#/cm^2/sec)	1.66 0.08	1.26 0.06	2.88 0.13
F_T ($\times 10^{-6}$ ergs/sec)	0.62 ± 0.04	2.34 ± 0.11	17.8 ± 0.8
H_{32}	5.50 0.13	2.45 0.06	3.16 0.07
H_{321}	3.39 0.08	1.32 0.03	1.78 0.04

Errors quoted are standard errors. The number of members are 423, 622 and 549 for Clusters I, II and III respectively.

Table 2: Discriminant Analysis for the K-means classification

	Cluster I	Cluster II	Cluster III
Cluster I*	417	28	0
Cluster II*	6	578	23
Cluster III*	0	21	526
Total	423	622	549

Clusters I, II and III, are the clusters obtained from the K-means classification. Clusters I*,II* and III* are the clusters to which the GRB were assigned by the Discriminant analysis.

Once the optimum classification (clustering) is obtained, using a process called Discrimination Analysis Johnson (1996), one can verify the acceptability of the classification by computing classification/misclassification probabilities for the different GRB. Although the K-means clustering method is purely a data analytic method, for classification it may be necessary to assume that the underlying distribution is Multivariate Normal. In this standard procedure, using the probability density functions in parameter space for the different clusters, one can assign an object (in this case a GRB) to be a member of a certain class. If the original classification was robust, then every GRB should be classified again as a member of the same class that it was before. If a significant number of objects are not reclassified then that would mean that the original classification was not stable and hence not trustworthy. Table 2 show the result of a Discrimination Analysis, where the columns represent how the GRB of a cluster were assigned by the analysis. The fraction of correct classification is 0.954 which implies that the classification is indeed robust.

2.2 **Dirichlet Process Model Based Clustering**

The standard approach to model-based clustering analysis, is based on modeling by finite mixture of parametric distributions. For example, Mukherjee et al. (1998) used such a model based approach to analyze GRB data where they assumed that the GRB population consists of mixture of multivariate Gaussian classes. The number of classes is, however, determined from an initial classification method (e.g. via agglomerative hierarchical clustering). The Dirichlet process model based clustering is more general and avoids the assumption of known number of possible classes. Since this method is less commonly used as compared to the K-means technique, we describe here the basic concept on the analysis in more detail.

The Dirichlet process avoids a prior assumption of the number of classes by applying a Bayesian nonparametric modeling of the unknown distribution for the multi component data. In this particular case the six component GRB data can be represented by $x_i = (\log T_{50}, \log T_{90}, \log F_T, \log P_{256}, \log H_{321}, \log H_{32})', i = 1, 2, \ldots, n$. More specifically, x_i is assumed to follow a multivariate normal distribution whose mean vector is generated from a Dirichlet Process (DP). Following Escober & West (1995), the method is best conceptualized by representing the model as

$$\begin{aligned} x_i|\mu_i, \Sigma &\sim \mathrm{MVN}(\mu_i, \Sigma), \quad i = 1, 2, \ldots, n; \\ \mu_i|G &\sim G; \quad G \sim DP(\alpha G_0) \end{aligned} \tag{1}$$

where MVN means multivariate normal distribution, G is a discrete measure of the unknown distribution, α is the precision parameter and G_0 is a known base measure distribution. Since G is discrete, there can be ties among the μ_i's , which can also be seen from Polya urn representation of Blackwell & MacQueen (1973) as

$$\mu_i|\mu_1, \mu_2, \ldots, \mu_{i-1} \sim \frac{\alpha}{\alpha + i - 1} G_0 + \frac{1}{\alpha + i - 1} \sum_{h=1}^{i-1} \delta(\mu_h) \tag{2}$$

where $\delta(x)$ is the distribution concentrated at the single point x. It is evident from Eqn. (2) that μ_i are marginally sampled from G_0 with positive probability and that some of the μ_i's are identical. Thus, a partition of $S = \{1, 2, \ldots, n\}$ can be formed by defining classes under the relation that μ_i belongs to the j^{th} class if and only if $\mu_i = \mu_j, j = 1, 2, \ldots, k$, k being the number of distinct μ_i's , $i = 1, 2, \ldots, n$. This induces a certain posterior distribution of S and a posterior inference can then be used to provide clustering procedure. There are various algorithms available to obtain the posterior partitions of S which are useful for making inferences on clustering of $x_1, x_2, \ldots, x_n$. We implement the independent and identically distributed Weighted Chinese Restaurant (iidWCR) algorithm (see Ishwaran & Takahara, 2002) which comes from its use of the partition distribution of S. Let $p = \{C_1, C_2, \ldots, C_{n(p)}\}$ be a partition of size $n(p)$ of S, where each C_j contains e_j elements. Assuming G_0 as the multivariate normal with mean vector m and covariance matrix B_0 and denoting $N_p(x; \mu, \Sigma) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} exp[-\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)]$ as the density of a p-component multivariate normal distribution, the iidWCR algorithm for inducing posterior partition of S consists of following steps:

Step 1: Assign $p_1 = \{1\}$ and the corresponding importance weight $\lambda(1) = N_6(x_1; m, \Sigma_0 + B_0)$ where Σ_0 is the initial estimate of Σ.

Step r: Given p_{r-1}, compute Σ_{r-1} from $x_1, x_2, \ldots, x_{r-1}$. Create p_r by assigning label r to a new set with probability $\frac{\alpha}{(\alpha+r-1)\lambda(r)} \times N_6(x_r; m, \Sigma_{r-1} + B_0)$. Otherwise, assign label r to an existing set $C_{j,r-1}$ with probability $\frac{e_{j,r-1}}{(\alpha+r-1)\lambda(r)} \times N_6(x_r; \mu_{j,r-1}, \Sigma_{j,r-1})$ where $\Sigma_{j,r-1} = (B_0^{-1} + e_{j,r-1}\Sigma_{r-1}^{-1})^{-1}$ and $\mu_{j,r-1} = \Sigma_{j,r-1}(B_0^{-1} + e_{j,r-1}\Sigma_{r-1}^{-1}\overline{x}_{j,r-1})$. Note that $e_{j,r-1}$ and $\overline{x}_{j,r-1}$ are the number of elements and observed mean in $C_{j,r-1}$ respectively and $\lambda(r)$ is the normalizing constant.

Running step 1 followed by step r for $r = 2, 3, \ldots, n$ gives a draw from posterior partition of S. This $n - step$ draw, in fact, provides an iid sample from WCR density given by

$$g(p) = \frac{f(x|p)\pi(p)}{\Delta(p)}$$

where $\pi(p)$ is the prior density of p, f(.) is the density of x and $\Delta(p) = \lambda(1) \times \lambda(2) \times \ldots \times \lambda(n)$ is the importance weight. Repeating the above algorithm B times, one can obtain $p^1, p^2, \ldots, p^B$ iid sample observations from posterior partition of S. Based on these sample observations, Monte Carlo method can be devised to estimate $E\{n(p)\}$, the expected number of clusters, as

$$\widehat{E\{n(p)\}} \approx \frac{\sum_{b=1}^{B} n(p^b)\Delta(p^b)}{\sum_{b=1}^{B} \Delta(p^b)} \tag{3}$$

The key advantages of using this Dirichlet process model-based clustering are that the underlying distribution of x_i's and the number of clusters are unknown. Moreover, one can provide an estimate of the expected number of clusters by using Eqn. (3).

For fitting the model, we used $\alpha = 1.0$ and a flat prior $G_0 \sim N_6(0, \sigma^2\mathbf{I})$ with $\sigma^2 = 1000$. The initial value of Σ, Σ_0 is obtained as the sample covariance matrix. We applied the iidWCR algorithm for $B = 1000$ to obtain the estimate of the number of clusters. Three classes were obtained consistent with the results from the K-means technique. In Table 3 the mean values of the parameters with errors are tabulated. The values are consistent with those found from the K-means method. The number of members of cluster II, 892 is somewhat larger than what was found by the K-means method, 622, but considering the different nature and approach of the two techniques, such differences are perhaps expected. In summary, these two independent clustering techniques indicate that there are three classes of GRB with qualitatively similar properties.

3. CLASSIFICATION OF GRB WITH KNOWN RED-SHIFT

Although the classification described in the previous section is based on six GRB parameters the segregation of the classes can be visualized using the total fluence, F_T and the duration, T_{90}. This is illustrated in Figure (2) which shows F_T versus T_{90} for the members of the three clusters obtained using the K-means technique.

Table 3: Average Cluster properties based on the Dirichlet Mixture Modeling method

Parameters	Cluster I	Cluster II	Cluster III
T_{50} (sec)	0.31 0.02	6.76 0.31	16.22 1.50
T_{90} (sec)	0.45 0.03	19.05 0.88	43.65 3.02
P_{256} (#/cm^2/sec)	1.66 0.08	1.35 0.03	4.79 0.33
F_T ($\times 10^{-6}$ ergs/sec)	0.89 ± 0.06	3.46 ± 0.08	18.2 ± 0.2
H_{32}	4.68 0.22	2.82 0.06	3.31 0.15
H_{321}	2.75 0.13	1.58 0.04	1.86 0.09

Errors quoted are standard errors. The number of members are 409, 892 and 293 for Clusters I, II and III respectively.

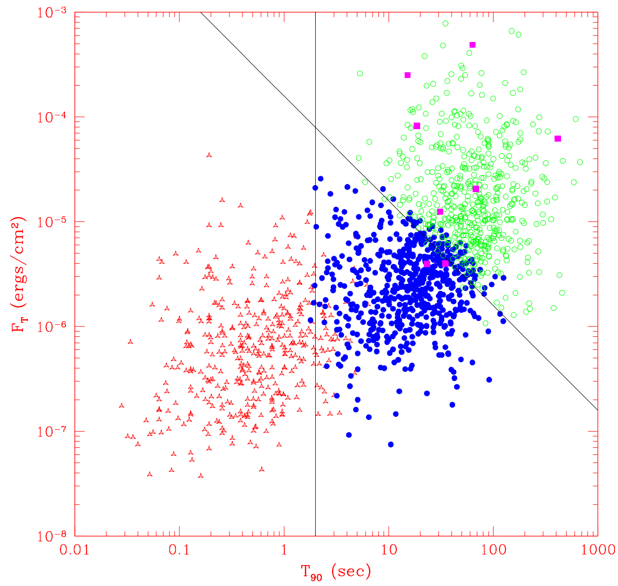


Figure 2: The fluence F_T in ergs/cm^2 versus the duration T_{90} in seconds for members of Cluster I (triangles), II (solid circles) and III (open circles) from the K-means clustering method. The solid lines represent $T_{90} = 2$s and $F_T = 1.6 \times 10^{-4}/T_{90}$ ergs/cm^2 which qualitatively separate the three groups. The solid squares represent eight GRB detected by BATSE for which redshifts are also measured (e.g. Bagoly et al., 2003).

The Dirichlet process model gives qualitatively similar results. The solid line representing $T_{90} = 2$ sec, differentiates the members of Cluster I (marked by triangles) with those of Cluster II (marked by filled circles). Thus Cluster I is consistent with the standard classification of short duration bursts. The standard long duration bursts (with $T_{90} > 2$ s) are further classified into two groups with one of them (members of Cluster III, marked using open circles) having typically higher fluence. Thus we have named members of Cluster II and III as low and high fluence GRB. The solid line representing $F_T = 1.6 \times 10^{-4}/T_{90}$ ergs/cm^2 qualitatively separates two groups. There are eight GRB detected by BATSE for which there are redshift estimates (e.g. Bagoly et al., 2003). These are marked by squares in Figure 2. Six of them are in Cluster III while two are close to the demarking line. One of these GRB (980425) is at a low redshift ($z = 0.0085$) and is associated with a supernova.

To identify a GRB with a known red-shift as a member of a cluster, broad band coverage of the prompt emission is required in order to correctly estimate the total fluence F_T, which BATSE would have observed for the burst. This is particularly important, when the peak of the energy spectrum of a GRB is at high energies > 300 keV. Amati et al. (2002) analyzed GRB with known red-shifts and well constrained spectral parameters over a broad energy range and discovered that the intrinsic (i.e. red-shift corrected) peak of the energy spectrum, E_p correlates with the isotropic energy output, E_{iso}.

Apart from being a stringent condition and test for any theoretical model that describes the GRB prompt emission, this empirical relation highlights the possibility that GRB can be used to probe and constrain the expansion of the Universe at early times. Ghirlanda et al. (2004) added more GRB to the sample and found that the beaming corrected luminosity, E has a tighter correlation with E_p than the isotropic one. Nevertheless, there is still significant dispersion in the relationship which needs to be explained. In order to see how the classification found in this work affects such relationship, we have used 21 GRB listed in Table 1 of Ghirlanda et al. (2004) that have well measured temporal and spectral parameters. The total fluence F_T versus the duration T_{90} for these GRB are plotted in Figure (3) (filled circles). Overlaid on the plot are the two solid lines $F_T = 1.6 \times 10^{-4}/T_{90}$ ergs and $T_{90} = 2$s, which qualitatively segregate the three clusters (Figure 2).

Most of the GRB have high fluence and are of long duration (consistent with them being members of Cluster III) which is probably a selection effect. Two of the GRB have $F_T < 1.6 \times 10^{-4}/T_{90}$ ergs/cm^2 and are probably members of Cluster II. However, there are three GRB with $F_T \sim 1.6 \times 10^{-4}/T_{90}$ ergs/cm^2 and hence there is an ambiguity about their classification. For each GRB, a corresponding dotted line is plotted in Figure (3) which shows the variation of the observed F_T versus T_{90} if the same GRB was located at different red-shifts. The lines are drawn for a red-shift range $0.1 < z < 5.0$ for a Λ CDM cosmology with $\Omega_\Lambda = 0.7$ and Hubble parameter $H = 65$ km/s/Mpc. It is interesting to note that the red-shift trajectories for the high fluence GRB in general do not cross the $F_T = 1.6 \times 10^{-4}/T_{90}$ line. In other words, if these GRB were located at a wide range of red-shifts, their observed fluence would have been $F_T > 1.6 \times 10^{-4}/T_{90}$ and hence they would have been classified as members of Cluster III.

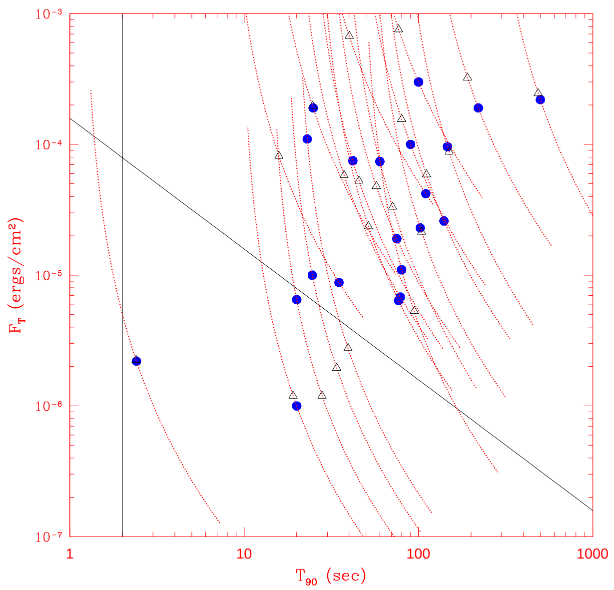


Figure 3: The fluence F_T in ergs/cm^2 versus the duration T_{90} in seconds for 21 GRB (filled circles) with known redshifts taken from Ghirlanda et al. (2004) and references therein. The solid lines represent $T_{90} = 2$s and $F_T = 1.6 \times 10^{-4}/T_{90}$ which qualitatively distinguish the three clusters (Figure 2). For each GRB, the dotted lines represent the predicted fluence and duration if it was located at a redshift range $0.1 < z < 5$. The open triangles represent the predicted fluence and duration if the GRB were located at redshift $z = 1$. These predicted values allow for the classification of the GRB into members of Cluster II and III.

On the other hand there are six GRB whose red-shift trajectories mostly lie below the de-marking line. Their observed fluence would be $F_T < 1.6 \times 10^{-4}/T_{90}$ for a wide range of red-shifts and hence they would be classified as members of Cluster II. The open triangles in the Figure mark the positions of the GRBs if they were all located at a red-shift, $z = 1.0$. In this representation, the GRB are more clearly segregated into Clusters II and III with five of them having a predicted fluence less than $F_T > 1.6 \times 10^{-4}/T_{90}$ with the others being significantly brighter. This strongly suggests that the classification described in this work is not due to observational bias arising from the use of observed parameters instead of intrinsic ones.

We classify the 21 GRB according to their predicted observed fluence and duration if they were situated at $z = 1$. In this scheme, five GRBs are classified as members of Cluster II while the remaining 18 are identified as members of Cluster III. Figure (4) shows the variation of the intrinsic energy peak E_{peak} versus the isotropic energy realized, E_{iso}, which is the correlation discovered by Amati et al. (2002). GRB identified as Cluster I are marked as triangles while those belonging to Cluster III are represented by filled circles. GRB belonging

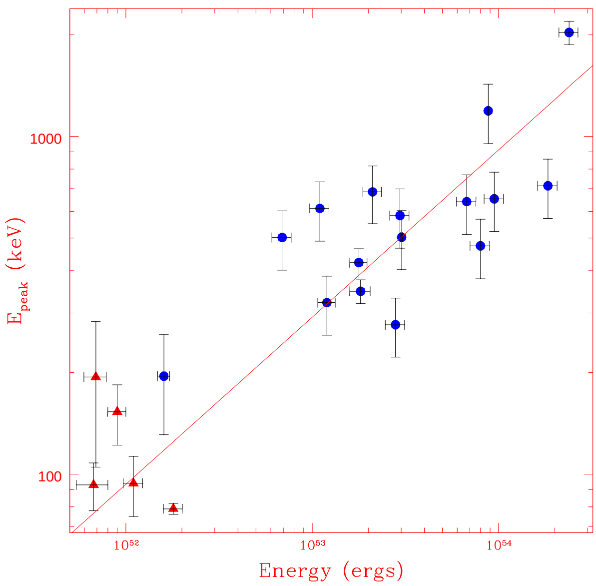


Figure 4: The intrinsic (red-shift corrected) peak of the energy flux, E_{peak} versus the isotropic energy output for 21 GRB with well defined spectral parameters (Amati et al., 2002; Ghirlanda et al., 2004). GRB that are members of Cluster II (triangles) have isotropic energy output of nearly 10^{52} ergs, while members of Cluster III (circles) have a much wider range of energy output.

to Cluster II all have an isotropic energy output close to $\sim 10^{52}$ ergs, while those belonging to Cluster III span a much larger range of energy 10^{52-54} ergs and roughly follow the $E_p - E_{iso}$ correlation. Although the number of GRB in this sample is small, this segregation of the GRB in intrinsic parameter space is a another indication that the classification is robust and perhaps not due to observational bias. Most of the members of Cluster III have a rest frame peak energy $E_{peak} > 300$ keV and hence would have had significant flux in the highest energy channel of BATSE, i.e. F_4. Thus, the retention of F_4 in the classification analysis of §2 is important even though there is systematic uncertainty in the measured value of the fluence. Indeed, if F_4 is not taken into account the evidence for three clusters in the BATSE sample decreases.

4. CLASSIFICATION WITH SWIFT DATA

Since the Swift GRBs have fluence in the lower energy band we first considered the BATSE sample with total fluence as the sum of $F1$, $F2$ and $F3$. Then we classify the sample with respect to $\log T90$ and logF ($F = (F1 + F2 + F3)$) assuming $K = 3$. Then we consider the Swift GRBs without z values and classify them with the parameters $\log T90$ and logF

Table 4: Average Cluster properties based on Swift catalog using K means cluster analysis

Catalog	Groups	No.of members	$\log(T_{90})$ (sec)	$\log(F_T)$ (ergs/sec)
BATSE	Cluster I	420	-0.30 ±0.02	-6.79 ±0.02
	Cluster II	610	1.19±0.02	-5.93 ±0.02
	Cluster III	564	1.78±0.01	-5.06±0.02
SWIFT(without z)	Cluster I	32	-0.49 ±0.09	-7.26 0.06
	Cluster II	146	1.06±0.03	-6.41±0.03
	Cluster III	235	1.96 0.04	-5.56 0.03
SWIFT(with z)	Cluster I	13	-0.24 0.14	-6.99 0.16
	Cluster II	44	1.130.06	-6.27 0.06
	Cluster III	91	1.95 0.04	-5.460.05

assuming $K = 3$. Then we remove the shortest class from the catalog of Swift GRBs and classify them with respect to the same parameters but for K=2. The mean values of $\log T90$ and $\log F$ are given for 3 and 2 groups respectively in Table 4 with the accompanying BATSE GRB classes. It is clear from Table 4 that we have similar distributions of $T90$ and F for both the catalogs.

5. SUMMARY AND DISCUSSION

Two multivariate clustering techniques, the K-means partitioning method and the Dirichlet process of mixture modeling, have been applied for the first time to the BATSE Gamma-ray burst (GRB) catalog. These two schemes do not make any a priori assumptions about the number of clusters, but instead provide quantitative estimate of the optimal number of groups. The jump curve for the K-means partitioning method suggests that this optimal number is three which is further supported with the value of the expected number of clusters, $E\{n(p)\}$, obtained using Dirichlet process of mixture modeling. The two techniques group the GRBs in qualitatively similar classes, which can be described as short bursts ($T_{90} < 2$ s, Cluster I), long duration, low fluence bursts ($F_T < 1.6 \times 10^{-4}/T_{90}$ ergs/cm^2, Cluster II) and long duration, high fluence bursts ($F_T > 1.6 \times 10^{-4}/T_{90}$ ergs/cm^2, Cluster III).

To estimate how such a classification, based on observed spectral and temporal parameters, can arise from intrinsic GRB properties, a sample of 21 GRB with known red-shifts and well constrained spectral parameters, are classified within this scheme. The observed total fluence, F_T and duration T_{90} for these GRBs if they were located at different red-shifts ($0.1 < z < 5$) were estimated. For 16 of the 21 GRB, the estimated fluence would have satisfied $F_T > 1.6 \times 10^{-4}/T_{90}$ ergs/cm^2 for nearly the entire range of red-shift space and hence they were classified as high fluence bursts. This invariance in red-shift, indicates that the classification scheme is not strongly effected by observational bias and by the use of observed parameters instead of intrinsic ones. For five GRB, classified as low fluence bursts, the predicted fluence $F_T < 1.6 \times 10^{-4}/T_{90}$ for a significant fraction of the red-shift space, which again signifies the physical nature of the classification. Based on the classification of these GRB with known red-shift, it can be inferred that the low fluence GRB have a nearly constant isotropic Energy output of 10^{52} ergs and have an intrinsic (red-shift corrected) duration of $T_{90} \sim$ 2-30 secs. On the other hand, the high fluence GRB (Cluster III) have a much wider range of isotropic energy output 10^{52-54} ergs and a corresponding wide range of intrinsic durations 10 - 500 secs.

We note with caution that the number of GRB with known red-shifts, used for this analysis is small and a much larger sample is required before concrete conclusions can be drawn. It is also important, for this analysis, to have well constrained spectral parameters of these GRB. In particular, the peak of the energy fluxes, E_p are required to be well estimated and since for high energetic sources $E_p > 300$ keV, it is imperative to have well calibrated high energy information. Indeed, if the highest energy channel of the BATSE measurement is not taken into account, the significance of the classification is smaller.

The classification presented here, needs to be supported by theoretical considerations. It is tempting to identify the low fluence GRB with neutron star-white dwarf mergers (King et al., 2007) and the higher fluences ones with massive stellar collapse. The near constancy of the isotropic energy output of low fluence bursts, seem to be consistent with them being neutron star-white dwarf mergers. Since both neutron stars and white dwarfs do not have significant mass variations, their initial conditions for the binary merger could be similar, leading to the nearly constant energy output. Moreover, their merger time may also be typically smaller than massive stellar collapse time-scales, which is consistent with the shorter intrinsic duration $2 - 30$ s, found in this work. On the other hand, the energy output and duration of GRB induced by massive stellar collapse may depend on the mass and size of the progenitor which is consistent with the variation inferred for high fluence bursts. The present observational evidence for such a model is not clear. Evidence for supernova light curves have been detected in GRB with different energy output, including some low luminosity ones, e.g. GRB 0311203, $E \sim 3 \times 10^{49}$ ergs (Malesani et al., 2004), the nearby GRB 060614, for which no supernova was detected also had a low isotropic energy output of 10^{51} ergs. Such low energy output are not represented in the 21 GRB with well constrained spectral parameters used in this analysis which all have energies $> 10^{52}$ ergs. These rare GRB (since they have to be located relatively nearby to be detected) may not represent a significant fraction of the BATSE catalogue. BATSE did detect GRB 980425 which is at low redshift ($z = 0.0086$) and is associated with a supernova. At this redshift, the GRB would be a borderline case between the high and low fluence GRB (Figure 2). Thus the interpretation of the two different classes of long bursts as being due to stellar collapse and white dwarf-neutron star mergers, is speculative and more quantitative theoretical predictions and observational evidences are required before a definite conclusion can be made.

ACKNOWLEDGEMENTS

We thank the IUCAA associateship program for support

REFERENCES

Amati, L., et al. 2002, *A&A*, 390, 81

Bagoly, Z., Meszaros, A., Horvath, I., Balazs, L. G., & Meszaros, P. 1998, *Astrophys. J.*, 498, 342

Bagoly, Z., Csabai, I., Mészáros, A., Mészáros, P., Horváth, I., Balázs, L. G., & Vavrek, R. 2003, *A&A*, 398, 919

Balastegui, A., Ruiz-Lapuente, P., & Canal, R. 2001, *MNRAS*, 328, 283

Baumgart,C.W. 1994, in SPIE Proc. 2243, Applications of Artificial Neural Networks V, eds. S. K. Rogers & D. W. Ruck (Bellingham: SPIE), 552.

Blackwell, D. & MacQueen,J. B. 1973, *Annals of Statistics*, 1, 353.

Bloom, J. S., et al. 2006, *Astrophys. J.*, 638, 354

Dezalay, J.-P., Barat, C., Talon, R., Syunyaev, R., Terekhov, O., & Kuznetsov, A. 1992, American Institute of Physics Conference Series, 265, 304

Escobar, M. D. & West, M. 1995, *Journal of the American Statistical Association*, 90, 577.

Feigelson, E. D., & Babu, G. J. 1998, IAU Symp. 179: New Horizons from Multi-Wavelength Sky Surveys, 179, 363

Fynbo, J. P. U., et al. 2006, *Nature*, 444, 1047

Gehrels, N., et al. 2005, *Nature*, 437, 851

Gehrels, N., et al. 2006, *Nature*, 444, 1044

Ghirlanda, G., Ghisellini, G., & Lazzati, D. 2004, *Astrophys. J.*, 616, 331

Hakkila, J., Haglin, D. J., Pendleton, G. N., Mallozzi, R. S., Meegan, C. A., & Roiger, R. J. 2000, *Astrophys. J.*, 538, 165

Hakkila, J., Giblin, T. W., Roiger, R. J., Haglin, D. J., Paciesas, W. S., & Meegan, C. A. 2003, *Astrophys. J.*, 582, 320

Hartigan, J. A. 1975, Clustering Algorithms, New York, Wiley.

Horváth, I. 1998, *Astrophys. J.*, 508, 757

Horváth, I. 2002, *A&A*, 392, 791

Horváth, I., Balázs, L. G., Bagoly, Z., Ryde, F., & Mészáros, A. 2006, *A&A*, 447, 23

Ishwaran, H. & Takahara, G., 2002, *Journal of the American Statistical Association*, 97, 1154.

Johnson, R. A., & Wichern, D.W. 1996, Applied Multivariate Statistical Analysis, Prentice Hall of India.

King, A., Olsson, E., & Davies, M. B. 2007, *MNRAS*, 374, L34.

Kouveliotou, C., Meegan, C. A., Fishman, G. J., Bhat, N. P., Briggs, M. S., Koshut, T. M., Paciesas, W. S., & Pendleton, G. N. 1993, *ApJ*, 413, L101

MacQueen, J., 1967, Fifth Berkeley Symp. Math. Statist. Prob.,1,281.

Malesani, D., et al. 2004, *ApJ*, 609, L5

Mukherjee, S., Feigelson, E. D., Jogesh Babu, G., Murtagh, F., Fraley, C., & Raftery, A. 1998, *Astrophys. J.*, 508, 314

Piran, T. 1992, *ApJ*, 389, L45

Piran, T. 2005, Reviews of Modern Physics, 76, 1143

Sazonov, S. Y., Lutovinov, A. A., & Sunyaev, R. A. 2004, *Nature*, 430, 646

Schaefer, B. E., & Xiao, L. 2006, ArXiv Astrophysics e-prints, arXiv:astro-ph/0608441

Sugar, A. S., & James, G. M. 2003, JASA, 98, 750.

Soderberg, A. M., et al. 2004, *Nature*, 430, 648

Woosley, S. E., & Bloom, J. S. 2006, *Ann. Rev. Astron. Astrophys.*, 44, 507

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 280–292.

ASTROCLADISTICS: MULTIVARIATE EVOLUTIONARY ANALYSIS IN ASTROPHYSICS

Didier Fraix-Burnet
Université Joseph Fourier - Grenoble 1 / CNRS
Laboratoire d'Astrophysique de Grenoble (LAOG) UMR 5571
BP 53, F-38041 GRENOBLE Cedex 09, France
E-mail: fraix@obs.ujf-grenoble.fr

ABSTRACT

The Hubble tuning fork diagram, based on morphology and established in the 1930s, has always been the preferred scheme for classification of galaxies. However, the current large amount of data up to higher and higher redshifts asks for more sophisticated statistical approaches like multivariate analyses. Clustering analyses are still very confidential, and do not take into account the unavoidable characteristics in our Universe: evolution. Assuming branching evolution of galaxies as a 'transmission with modification', we have shown that the concepts and tools of phylogenetic systematics (cladistics) can be heuristically transposed to the case of galaxies. This approach that we call "astrocladistics", has now successfully been applied on several samples of galaxies and globular clusters. Maximum parsimony and distance-based approaches are the most popular methods to produce phylogenetic trees and, like most other studies, we had to discretize our variables. However, since astrophysical data are intrinsically continuous, we are contributing to the growing need for applying phylogenetic methods to continuous characters.

Keywords: Classification; astrophysics; cladistics; multivariate; evolution; galaxies; continuous characters

1. INTRODUCTION

The extragalactic nature of galaxies was discovered by Edwin Hubble less than 90 years ago. He also discovered the expansion of the Universe and established in 1936 the famous tuning fork diagram or Hubble diagram based on the idea that elliptical galaxies should evolve into flattened systems such as ordinary or barred spiral galaxies. Since then, telescopes and the associated detectors have made remarkable progress. We are now able to study galaxies in great detail, identifying individual stars, gas and dust clouds, as well as different stellar populations. Imagery brings very fine structural details, and spectroscopy provides the kinematical, physical and chemical conditions of the observed entities at different locations within the galaxy. For more distant objects, information is scarcer, but deep systematic sky surveys gather millions of spectra for millions of galaxies at various redshifts.

Like paleontologists, we observe objects from the distant past (galaxies at high redshift), and like evolutionary biologists, we want to understand their relationships with nearby galaxies, like our own Milky Way. Strangely enough, the Hubble classification, depicted in the Hubble diagram, is still frequently used as a support to describe galaxy evolution, even though it ignores all observables except morphology (Fraix-Burnet et al., 2006a, Hernandez and Cervantes-Sodi, 2006). Can the enormous amount of very detailed observations and galaxy diversity be solely depicted by few large families characterized only by their global shape? There obviously must be a better way to exploit the data gathered by very large telescopes and their sophisticated detectors, and to account for the complicated physical processes that lead to the diversification of galaxy.

Abandoning the one-parameter classification approach and using all available descriptors means taking a methodological step equivalent to the one biologists took after Adanson and Jussieu in the 18th century. One basic tool, the Principal Component Analysis, is relatively well-known in astrophysics (e.g. Cabanac et al., 2002, Recio-Blanco, 2006). However, only a very few attempts to apply multivariate clustering methods have been made very recently (Chattopadhyay and Chattopadhyay, 2006, 2007; Chattopadhyay et al., 2007, 2008, 2009). Sophisticated statistical tools are used in some areas of astrophysics and are developing steadily, but multivariate analysis and clustering techniques have not much penetrated the community.

A supplementary difficulty is that evolution, an unavoidable fact, is not correctly taken into account in most classification methods. By mixing together objects at different stages of evolution, most of the physical significance and usefulness of the classification is lost. In practice, the evolution of galaxies is often limited to the evolution of the properties of the entire population as a function of redshift (Bell, 2005). Since environment (the expanding Universe) and galaxy properties are so much intricate, this kind of study is relevant to a first approximation. However, recent observations have revealed that galaxies of all kinds do not evolve perfectly in parallel, as illustrated for instance by the so-called downsizing effect which shows that large galaxies formed their stars earlier than small ones (e.g. Neistein, 2006). New observational instruments now bring multivariate information at different stages of evolution, and in various evolutive environments. In this multivariate context, we believe that the notion of "evolution", easy to understand for a single parameter, is advantageously replaced by "diversification".

In this conference, I present our efforts to implement a method to reconstruct the "galacto-genesis", based on tools and concepts largely developed in evolutionary biology and bioinformatics. I first present cladistics and some of our results. Then I replace this approach among multivariate statistical methods and explain why we also need these other tools to explore the parameter space. Finally, I present one of our developments that make the astrophysical problematic to join current bioinformatics and mathematics studies about the use of continuous characters in reconstructing phylogenies.

2. ASTROCLADISTICS

Multivariate clustering methods compare objects with a given measure and then gather them according to a proximity criterion. Distance analyses are based on the overall similarity derived from the values of the parameters describing the objects. The choice of the most adequate distance measure for the data under study is not unique and remains difficult to justify a priori. The way objects are subsequently grouped together is also not uniquely defined. Cladistics uses a specific measure that is based on characters (a trait, a descriptor, an observable, or a property that can be given at least two states characterizing the evolutionary stages of the object for that character) and compares objects in their evolutionary relationships (Wiley, 1991). Here, the “distance” is an evolutionary cost. Groupings are then made on the basis of shared or inherited characteristics, and are most conveniently represented on an evolutionary tree.

Character-based methods like cladistics are better suited to the study of complex objects in evolution, even though the relative evolutionary costs of the different characters are not easy to assess. Distance-based methods are generally faster and often produce comparable results, but the overall similarity is not always adequate to compare evolving objects. In any case, one has to choose a multivariate method, and the results are generally somewhat different depending on this choice (e.g. Buchanan, 2008). However, the main goal is to reveal a hidden structure in the data sample, and the relevance of the method is mainly provided by the interpretation and usefulness of the result.

Multivariate evolutionary classification in astrophysics has been pioneered by the author (Fraix-Burnet et al., 2006a, 2006c, Fraix-Burnet, 2009). Called astrocladistics, it is based on cladistics that is heavily developed in evolutionary biology. Astrocladistics has been first applied to galaxies (Fraix-Burnet et al., 2006b) because they can be shown to follow a transmission with modification process when they are transformed through assembling, internal evolution, interaction, merger or stripping. For each transformation event, stars, gas and dust are transmitted to the new object with some modification of their properties. Cladistics has also been applied to globular clusters (Fraix-Burnet et al., 2009), where interactions and mergers are probably rare. These are thus simpler stellar systems, even though we have firm evidence that internal evolution can create another generation of stars and that globular clusters can lose mass. Basically, the properties of a globular cluster strongly depend on the environment in which it formed (chemical composition and dynamics), and also on the internal evolution which includes at least the aging of its stellar populations. Since galaxies and globular clusters form in a very evolving environment (Universe, dark matter haloes, galaxy clusters, chemical and dynamical environment), the basic properties of different objects are related to each other by some evolutionary pattern.

In both cases, there is a kind of transmission with modification process, which justifies a priori the use of cladistics. It must be clear that this is not a "descent with modification" in the sense that there is no replication. But evolution does nevertheless create diversity. We are dealing with phylogeny (relationships between species), not with genealogy (relationships between individuals). Since a multivariate classification of galaxies is not yet available, we assume that each object represents a “species” that will have to be defined later on. The word ``progenitor", often used for galaxy evolution, is to be understood in this way.

A cladistic analysis works as follows. One first builds a matrix with values of the four parameters for all galaxies. In contrast to multivariate distance methods, undocumented values are not a problem in cladistic analyses. The values for each parameter are discretized into 30 bins representing supposedly evolutionary states. Discretization of continuous variables is quite a complex problem, especially in the evolutionary context (see Section 4). Here, we took equal-width bins. The choice of the number of bins cannot be made in a simple objective way. A priori, one could consider a compromise between an adequate sampling of continuous variables and the uncertainties on the measurements. The first constraint is given by the software (32 in this case). The second one would a priori give a lower limit of something like total range/uncertainty, but Shannon's theorem would multiply this by 2. Hence, 30 bins would account for about 7% measurement errors. Even so, border effects always imply that some objects could belong to a bin or its neighbor, a process that add some more artificial noise. The best way to avoid this effect is to make several analyses with different number of bins and check that the result does not depend on this number. From our experience in astrocladistics, we know that the results are very generally identical between 20 and 30, and can differ with 10 bins. For lower number of bins, the results are very dependent on this number.

Then, all possible arrangements of galaxies on a tree-like structure are constructed, and using the discretized matrix, the total number of state changes is computed for each tree. The most parsimonious tree is finally selected. If several such trees are found, then a consensus (strict or majority rule) tree is built. The whole procedure is computerized since the number of arrangements is very large. The result (depicted on a cladogram) is a diversification scenario that should be confronted to other knowledge and parameters.

Figure 1 shows the cladogram obtained for globular clusters of our Galaxy (Fraix-Burnet et al., 2009). Three groups are identified. The first one (in blue) has on average the lower ratio Fe/H that measures the proportion of heavy atomic elements that are processed within stars. This group is consequently considered as more primitive.

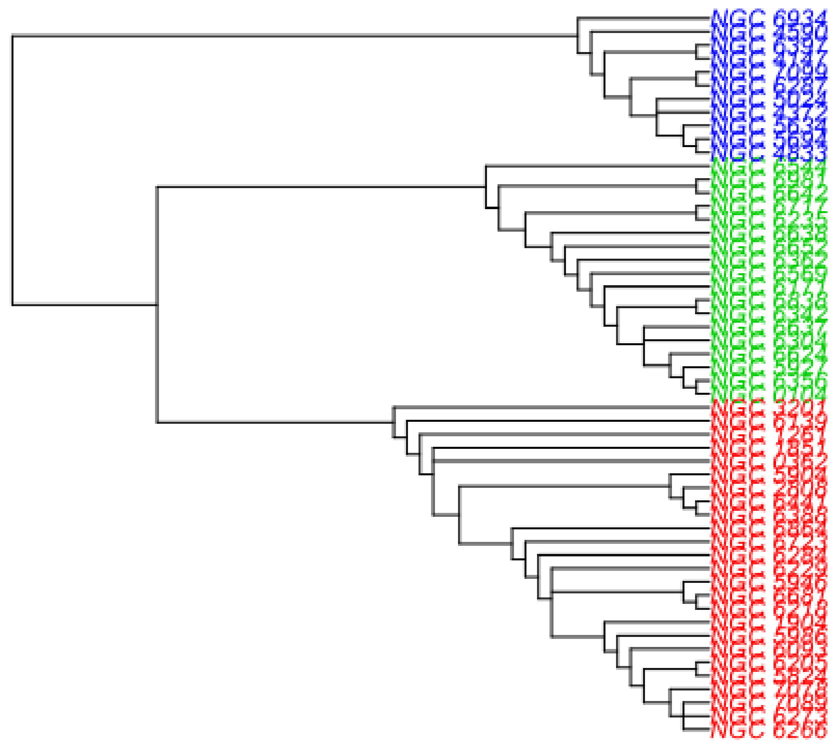


Figure 1. Cladogram of Galactic Globular Clusters.

Figure 2 shows bivariate diagrams for the 4 parameters used in the analysis: logTe, that measures the temperature of stars that are at a specific point in their evolution, Fe/H, MV that is the total visible intensity (magnitude) and roughly indicates the mass of the globular cluster, and Age that can be measured quite precisely because all stars of a given globular cluster are formed nearly at the same time. However Age is not an intrinsic property discriminating evolutionary groups since it evolves in the same way for all. But we gave it a half weight to arrange the objects within each group. Looking at plots like Figure 2 and using other parameters (orbital elements, kinematics, more refined chemical abundances...), we are able to infer that each group formed during a particular stage of the assembly history of our Galaxy. The blue group is the older one. It formed during the dissipationless collapse of the protogalaxy. They are located mainly in the outer halo. The red group belongs to the inner halo and the corresponding clusters formed at a later stage during the dissipational phase of Galactic collapse, which continued in the halo after the formation of the thick disc and its globular clusters. These clusters were very massive before "star evaporation" took place. The latter group (green) formed during an intermediate and relatively short period and comprises clusters of the disk of our Galaxy.

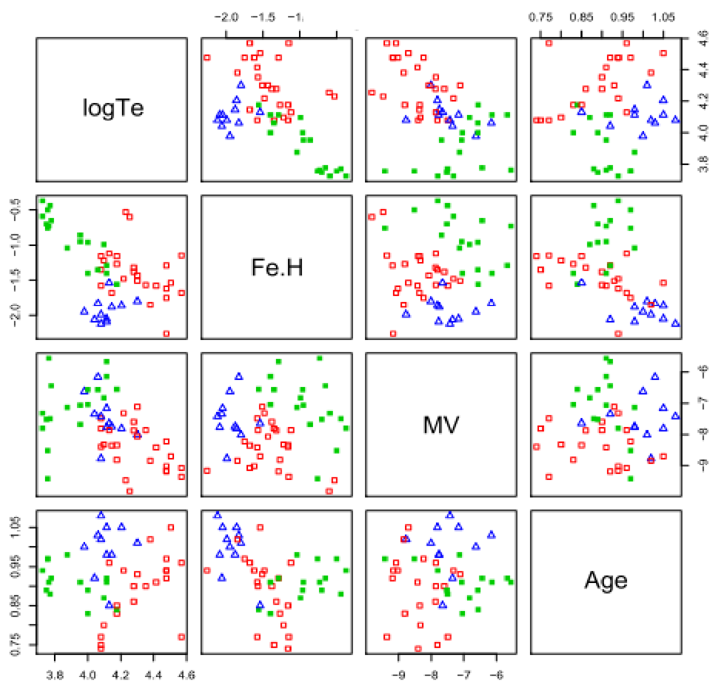


Figure 2. Scatterplots for the 4 parameters used in the analysis, colors corresponding to groups defined on the cladogram of Figure 1.

3. STRUCTURE OF THE PARAMETER SPACE

The spectacular astrophysical results obtained so far were possible because the characters are pertinent from the evolutionary and physical points of view. However, they are too few to obtain a detailed classification of tens of thousands of galaxies. Since astrophysics is a science of observation, not of experimentation, it is difficult to obtain the intrinsic physical and chemical properties of galaxies or globular clusters especially if distant. The spectra carries all the information we gather, so that much data is available. But this information is generally not straightfully pertinent, either because it is not directly connected to physical parameters, or because there are redundancies or incompatibilities that are paticularly annoying in a cladistic analysis. Being as multivariate as possible ensures objectivity, but some care is required in selecting the parameters. Different multivariate tools are thus needed to explore the structure of the data.

The first and obvious approach is to use Pincipal Component Analysis (PCA) to reduce the dimensionality of the parameter space and avoid redundancy. It is possible to perform a cladistic analysis directly with the main components. However, the correlations revealed by such analyses are not necessarily intrinsic nor due to scaling effects. They can be generated by evolution. For instance, in Figure 3, the parameter logs, that measures the central velocity dispersion, is not physically linked to mgbfe, that measures to the global metallic composition of the stars, but they appear "correlated" through their respective evolutions. They possibly characterize two independent facets of galaxy diversification. Such information would be lost with the use of the PCA components.

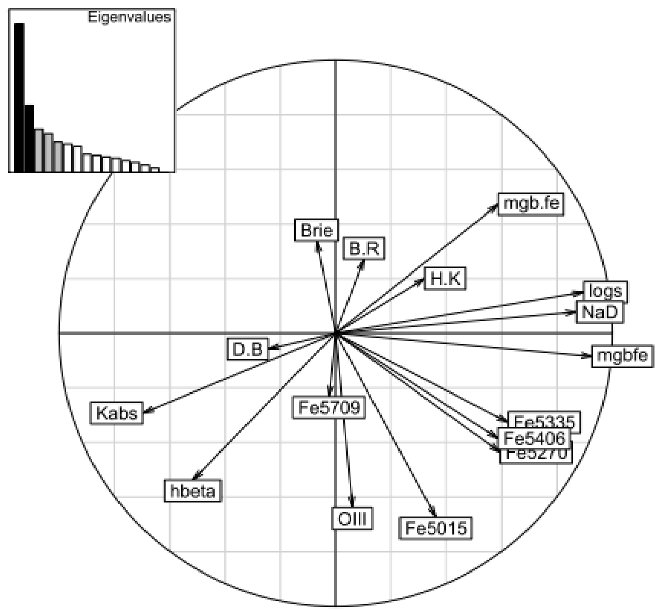


Figure 3. Projection on the two first Principal Components of parameters available for a typical sample of galaxies.

Alternatively, it is possible to select the parameters that appear to have more weight in the PCA. In this approach, we both keep the evolutionary information and gain a precious knowledge on the parameters themselves, which are physical in contrast to the principal components. Parameters for subsequent analyses are thus selected according to their loadings and their physical meaning. At the end, we eliminate true redundancies (parameters that measure the same quantity) and undiscriminating parameters.

There are other techniques that surely could be interesting and that we have not yet explored. Rather, we are investigating approaches more specific to the problem of obtaining a phylogeny with continuous characters (see Section 4). It appears that the structure of the parameter space must be at least grossly understood before embarking on sophisticated clustering tools, especially if the evolutionary information is to be retrieved. One difficulty in multivariate clustering of continuous and evolutive parameters, certainly not specific to astrophysics, is the cosmic variance that makes the groups to largely overlap in the parameter space.

Once the parameters are selected, it is instructive to compare results obtained with several multivariate clustering techniques in several different conditions (sub-samples, subsets of parameters). This provides additional insights on the parameters, and the different groupings can be confronted and analysed. Similarities reinforce the results, and mismatches can be interpreted in light of specificities of the techniques. In this way, we have obtained very convincing results on samples of hundreds of galaxies (Fraix-Burnet, Chattopadhyay, Chattopadhyay, Davoust and Thuillard, in prep). We demonstrate that multivariate clustering techniques and cladistics are the right tools to use. They yield results that are largely consistent, the cladogram providing the additional evolutionary relationshups between the groups. These groups cannot be guessed by any a priori splitting of the data, but we find that their properties happen to delineate different assembling histories, like for the globular clusters above.

4. CONTINUOUS CHARACTERS AND SPLIT NETWORKS

Cladistics is a difficult concept and its use in the case of continuous variables (morphometric data) is under intense investigation in the bioinformatics community (Gonzàles-Rosé et al., 2008). In particular, the formal relationship between distance-based (clustering, continuous variables) and character-based (cladistics, discrete values) is a very interesting topic in itself for which astrophysics brings an entirely new kind of data (Thuillard and Fraix-Burnet, 2009).

Maximum parsimony and distance-based approaches are the most popular methods to produce phylogenetic trees. While most studies in biology use discrete characters, there is a growing need for applying phylogenetic methods to continuous characters. Examples of continuous data include gene expressions (Planet et al., 2001), gene frequencies (Edwards and Cavalli-Sforza, 1964; 1967), phenotypic characters (Oakley and Cunningham, 2000) or some morphologic characters (MacLeod and Forey, 2003; González-José et al., 2008).

The simplest method to deal with continuous characters using maximal parsimony consists of discretizing the variables into a number of states small enough to be processed by the software.

Distance-based methods are applied to both discrete and continuous input data. Compared to character-based approaches, distance-based approaches are quite fast and furnish in many instances quite reasonable results. As pointed out by Felsenstein (2004), the amount of information that is lost when using a distance-based algorithm compared to a character-based approach is often surprisingly small. The use of continuous characters in distance-based methods may at first glance be less problematic than in character-based methods as algorithms like the Neighbour-Joining work identically on discrete or continuous characters. But also here it is often not easy to determine if the data can be described by a tree. In Thuillard and Fraix-Burnet (2009), we investigate the conditions for a set of continuous characters to describe a split network or a X-tree (split networks are generalized trees). We show that a set of m continuous characters can be described by a split network or a weighted X-tree if the m-dimensional space representation of the taxa state values is on an orthogonal convex hull. In that case, there exists an order of the taxa, called perfect order, for which the distance matrix satisfies the Kalmanson inequalities for each character.

In practice, identifying a priori characters that comply with these conditions is difficult. For complex objects in evolution, this requires some good knowledge of the evolution of the characters together with some ideas about the correct phylogeny or at least a rough evolutionary classification. In astrophysics, the study of galaxy evolution has yet not reached this point. However, we can show that the approach presented in this work, based on the Minimum Contradiction Analysis (Thuillard, 2007, 2008) is extremely valuable even in cases with very little a priori hints. Here, we illustrate how this method can be used in practice, in particular to discover structuring characters and objects (more can be found in Thuillard and Fraix-Burnet, 2009). We have taken from Ogando et al. (2008) a sample of 100 galaxies described by some observables and derived quantities.

In this first example, three variables have been chosen: logs, that is the logarithm of the central velocity dispersion, Kabs that is the absolute magnitude in the K band and measures the total quantity of near-infrared K light emitted mainly by the stars, and ldiam that is the diameter of the galaxy. They are all expected to be correlated with the total mass of the galaxy. In addition, Kabs and ldiam both strongly depend on the total number of stars.

Figure 4 shows the three variables logs, Kabs, Idiam after ordering of the galaxies with the Minimum Contradiction Analysis that provides the best order (the closest to the perfect order). The two last characters fulfill quite well the conditions for perfect order and the associated distance matrices are well ordered. The first character is in very first approximation well ordered. The distance matrix is well approximated by an expression characterizing a line tree structure that does not show particular substructures.

As expected, the variables Kabs and ldiam are well correlated as can be seen from the ordered taxa corresponding to the best order. They are essentially responsible for the two groups seen on the distance matrix of Fig. 4 and thus for the line tree structure. The variable logs is only in first approximation correlated to the other variables. It is much noisier and has no obvious extremum.

In the second example, we have chosen two variables possibly very little correlated and with evolution function probably not far from the case yielding a convex hull. D/B is the ratio between the size of the disk to that of the bulge. This is a quantitative and objective measure of the morphology of galaxies, spiral (that are disky) systems having a high D/B and elliptical ones having low D/B. The OIII variable is an intensity measure of a region of the spectrum where the OIII emission or absorption line is present. A strong emission line indicates the presence of interstellar gas and thus probably star formation. It is a priori not directly physically linked to the morphology.

Despite some large deviations to perfect order for a number of galaxies, the level of contradiction is quite low as can be seen in Figure 5. The distance matrix can be described to a good approximation by a split network. A typical split network is shown in Figure 6.

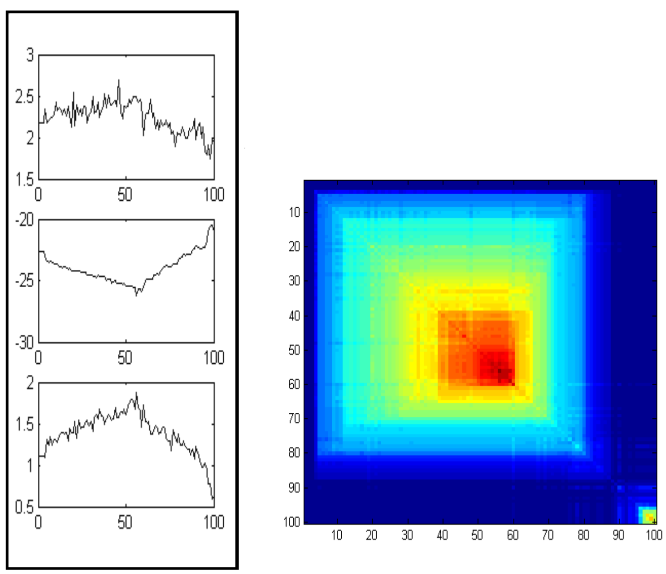


Figure 4. Left: Character values (top: logs, middle: Kabs, bottom: ldiam).vs. taxa arranged according to the best order. Right: Distance matrix on the 3 characters.

Obviously, split networks can represent more complex relationships than trees, but they are very difficult to use for inferring phylogenetic hypotheses, even more than reticulograms that take into account hybridization or horizontal transfer. In Thuillard and Fraix-Burnet (2009), we propose a simple a posteriori discretization of the variables that, in favorable cases, simplifies a split network into an X-tree. This is depicted in Figure 7 where the values of the variables are plotted in the best order of Figure 5.

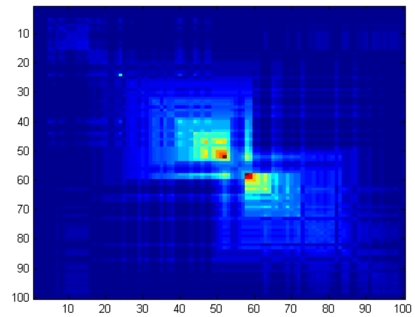


Figure 5. Minimum contradiction matrix for the characters D/B and OIII.

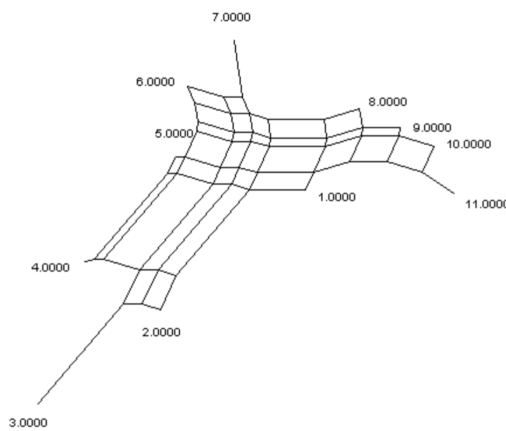


Figure 6. An example of a split network.

Using a threshold value for the two variables OIII and D/B furnishing the discrete characters OIII large, OIII small, D/B large and D/B small. Compared to the peak in the D/B curve, the large peak for OIII is shifted to the right by roughly 5-10 galaxies. This shift generates some structure in the distance matrix (Figure 5). The galaxies can then be grouped into 3 different groups. The first group has large D/B values and large OIII except for 2 or 3 galaxies that show small OIII, the second group has small D/B and small OIII values while the group 3 contains all galaxies with small values of D/B and large OIII. This grouping can be represented on the tree of Figure 8.

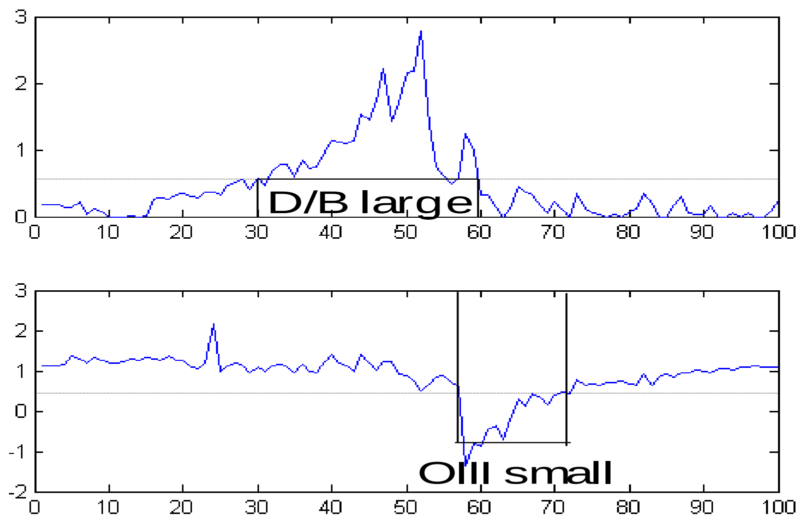


Figure 7. Behavior of the two variables D/B and OIII arranged according to the best order represented in Figure 5. The a posteriori discretization is shown, and yields the tree in Figure 8.

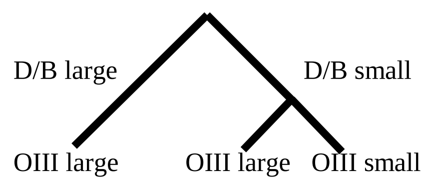


Figure 8. The "phylogenetic" tree obtained from the discretization shown in Figure 7.

The first group is made of disky galaxies, which have generally a high star formation rate. There are only 2 or 3 galaxies of this group that have a small OIII. The second and third groups have more bulgy galaxies. In the latter, some galaxies show an OIII line in emission as strong as in group 1. These galaxies seem to contradict the long known observation, that elliptical galaxies have less gas, and thus form less stars than spiral galaxies. This observation is believed to be due to historical reasons in the evolution of galaxies that are not at all clear. Is it due to a sweeping of the gas in the elliptical galaxies by the intergalactic medium? Or is it due to the way elliptical galaxies formed, exhausting nearly all their gas in very efficient star formation episodes? Even if our sample is relatively small, OIII and D/B are obviously not tightly correlated since the first group contains some disky galaxies with low star formation and group 3 has bulgy galaxies with OIII in emission. Anyhow, this example is only for illustration, and does not pretend to generality. At least, it shows that since the formation scenarios for galaxies are complicated, they cannot be studied with two parameters only, like the present morphologies (like D/B) and the OIII signature.

As a conclusion of this study, the level of contradiction of the minimum contradiction matrix furnishes an objective measure of the deviations to a tree or network structure. This measure can be used as a criterion to select the most important characters. The minimum contradiction matrix can also be of great help to a posteriori discretize continuous characters so that the resulting discrete characters can be well described by a tree or a split network. Quite interestingly, while discretization of continuous characters is often problematic, the posterior discretization after ordering of the taxa can help removing contradictions from a split network or tree structure. We believe there is here a great hope not only for astrophysics but for many evolutionary and multivariate studies.

5. COMMENTS AND CONCLUSIONS

It is now clear that cladistics can be applied and be useful to the study of galaxy diversification. Many difficulties, conceptual and practical, have been solved,. Significant astrophysical results have been obtained and will be extended to larger samples of galaxies and globular clusters. However, many paths remain in the exploration of this new and large field of research.

There are difficulties that seem to be intrinsic to astrophysics. Most notably, we have millions of objects, but a few tens of descriptors. Of course the situation will improve with time, in particular with integral-field spectroscopy (spectra of detailed regions in the galaxy, e.g. Ensellem et al., 2007). Spectra might not be currently employed at their full capacity of description. Here also, sophisticated statistics can help. But it is not clear whether these data would lead to the discrimination of hundreds of classes. Perhaps this is an erroneous target, perhaps galaxies cannot be classified with such a refinement. Nonetheless, this is already a matter of using multivariate clustering methods and interpreting their results usefully. This is probably more than a conceptual question because clustering with continuous data and large intraspecific variance is a very complicated problem in itself. Sophisticated statistical tools must be used, but the question of characterizing the groups in this context requires a different culture that is not yet widespread in astrophysics. We are convinced that improvements can be made here.

Cladistics is supposed to identify clades, that are evolutionary groups, whereas the concept of “species” is not defined at all in astrophysics, and we have even not converged toward groupings based on multivariate analyses. Understanding the formation and evolution of galaxies, depicting a global scenario for galaxy diversification, is a major challenge for the astrophysics of XXIst century.

REFERENCES

Bell, E. F. (2005). Galaxy Assembly. *Planets to Cosmology: Essential Science in Hubble's Final Years*. Ed. M. Livio. Cambridge: CUP (astro-ph/0408023).

Buchanan, B. and Collard, M. (2008). Phenetics, cladistics, and the search for the Alaskan ancestors of the Paleoindians: a reassessment of relationships among the Clovis, Nenana, and Denali archaeological complexes. *Journal of Archaeological Science*, 35, 1683-1694.

Cabanac, R. A., de Lapparent, V. and Hickson, P. (2002). Classification and redshift estimation by principal component analysis. *Astronomy & Astrophysics*, 389, 1090–1116 (astro-ph/0206062).

Chattopadhyay, T. and Chattopadhyay, A. (2006) Objective classification of spiral galaxies having extended rotation curves beyond the optical radius. *The Astronomical Journal*, 131, 2452–2468.

Chattopadhyay, T. and Chattopadhyay, A. (2007). Globular Clusters of LOCAL Group – Statistical Analysis. *Astronomy & Astrophysics,* 472, 131-140.

Chattopadhyay, T., Misra R., Naskar, M. and Chattopadhyay, A. (2007). Statistical evidences of three classes of Gamma Ray Bursts. *Astrophysical Journal,* 667, 1017.

Chattopadhyay, T., Mondal, S. and Chattopadhyay, A. (2008). Globular Clusters in the Milky Way and Dwarf Galaxies - A Distribution-Free Statistical Comparison. *Astrophysical Journal*, 683, 172.

Chattopadhyay, T., Babu, J., Chattopadhyay, A. and Mondal, S. (2009). Horizontal Branch Morphology of Globular Clusters: A Multivariate Statistical Analysis. *Astrophysical Journal*, in press.

Edwards, A. W. F. and Cavalli-Sforza, L. L. (1964). Reconstruction of evolutionary trees. *Phenetic and Phylogenetic Classification.* Ed. V. H. Heywood and J. McNeill. Systematics Association Pub. No. 6, London, pp. 67- 76.

Felsenstein, J. (2004). *Inferring phylogenies*, Sinauer Associates.

Fraix-Burnet, D., Choler, P., Douzery, E. and Verhamme, A. (2006a). Astrocladistics: a phylogenetic analysis of galaxy evolution. I. Character evolutions and galaxy histories. *Journal of Classification*, 23, 31-56 (astro-ph/0602581).

Fraix-Burnet, D., Choler, P. and Douzery, E. (2006b). Towards a Phylogenetic Analysis of Galaxy Evolution: a Case Study with the Dwarf Galaxies of the Local Group. *Astronomy & Astrophysics*, 455, 845-851 (astro-ph/0605221).

Fraix-Burnet, D., Davoust, E. and Charbonnel, C. (2009). The environment of formation as a second parameter for globular cluster classification. *Monthly Notices of the Royal Astronomical Society*, 398, 1706-1714.

Fraix-Burnet, D., Douzery, E., Choler, P. and Verhamme, A. (2006c). Astrocladistics: a phylogenetic analysis of galaxy evolution. II. Formation and diversification of galaxies. *Journal of Classification*, 23, 57-78 (astro-ph/0602580).

Fraix-Burnet D. (2009). Galaxies and Cladistics. *Evolutionary Biology: Concept, Modeling, and Application.* Ed. Pontarotti P. Biomedical and Life Sciences, Springer Berlin Heidelberg, pp. 363–378 (http://fr.arxiv.org/abs/0909.4164).

Gonzàlez-José, R., Escapa, I., Neves, W. A., Cúneo, R. and Pucciarelli, H. M. (2008). Cladistic analysis of continuous modularized traits provides phylogenetic signals in Homo evolution. *Nature*, 453, 775-779.

Hernandez, X. and Cervantes-Sodi, B. (2005). A dimensional study of disk galaxies. *11th Latin-American Regional IAU Meeting* (astro-ph/060225)

MacLeod, N. and Forey, P. L. (2003). *Morphology, Shape and Phylogeny.* Eds. Taylor and Francis Inc., New York.

Neistein, E., van den Bosch, F. C. and Dekel, A. (2006). Natural downsizing in hierarchical galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 372, 933-948 (astro-ph/0605045).

Oakley, T. H. and Cunningham, C. W. (2000). Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. *Evolution*, 54 (2), 397-405.

Ogando, R .L. C., Maia, M. A. G., Pellegrini, P. S. and da Costa, L. N. (2008). Line strengths of early-type galaxies. *The Astronomical Journal*, 135, 2424-2445.http://fr.arxiv.org/abs/0803.3477.

Planet, P. J, DeSalle, R., Siddal, M., Bael, T., Sarkar, I. N. and Stanley, S. E. (2001). Systematic analysis of DNA microarray data: ordering and interpreting patterns of gene expression. *Genome Research*, 11, 1149-1155.

Recio-Blanco, A., Aparicio, A., Piotto, G., De Angeli, F. and Djorgovski, S. G. (2006). Multivariate analysis of globular cluster horizontal branch morphology: searching for the second parameter. *Astronomy & Astrophysics*, 452, 875-884 (http://arxiv.org/abs/astro-ph/0511704).

Thuillard, M. (2007). Minimizing contradictions on circular order of phylogenic trees. *Evolutionary Bioinformatics*, 3, 267-277.

Thuillard, M. (2008). Minimum contradiction matrices in whole genome phylogenies. *Evolutionary Bioinformatics*, 4, 237-247.

Thuillard M. and Fraix-Burnet D. (2009). Phylogenetic Applications of the Minimum Contradiction Approach on Continuous Characters. *Evolutionary Bioinformatics*, 5, 33-46 http://fr.arxiv.org/abs/0905.2481.

Wiley, E. O., Siegel-Causey, D., Brooks, D. R. and Funk, V.A. (1991). *The Compleat Cladist: A Primer of Phylogenetic Procedures.* The University of Kansas, Museum of Natural History special publication N°19.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 293–302.

PROPERTIES OF GLOBULAR CLUSTERS AND THEIR HOST GALAXIES

M. E. Sharina
Special Astrophysical Observatory, Nizhnij Arkhyz, Zelenchukskiy region, Karachai-Cherkessian Republic, Russia 369167
E-mail: sme@sao.ru

ABSTRACT

We consider statistical properties of dwarf galaxies and of globular cluster systems situated within ~10 Mpc, and explain them physically. We analyse ages, metallicities, and light-element abundance ratios obtained from medium-resolution spectra of extragalactic globular clusters at the 6m telescope of the Russian Academy of Sciences and other large telescopes, as well as high-quality structural parameters of globular clusters derived from photometry on Hubble Space telescope images. We consider new multivariate statistical methods for addressing the problem of the origin and evolution of galaxies.

1. INTRODUCTION

The investigation of dwarf galaxies and their globular cluster (GC) systems is an important task for modern astrophysics. The hierarchical structure formation scenario predicts that objects with masses typical for the present-day faintest dwarf galaxies and brightest GCs (~$10^{6\text{-}7}$ $M_{\odot}$) were the first objects to form in the Universe (Peebles & Dicke, 1968), and all structures we observe today were formed by subsequent merging of these smaller sub-units. Studying the chemical composition, mass, luminosity, velocity dispersion, structural parameters of such objects, and possible correlations between different observational properties should shed light on the cosmological structure formation process.

2. PROPERTIES OF DWARF GALAXIES AND THEIR PLACE IN THE COSMOLOGICAL HIERARCHY

The Local Volume (LV), a region of up to ~10 Mpc from the Sun, mainly consists of so-called "poor" groups, formed by one or two bright spiral or elliptical galaxies and a number of fainter dwarfs (e.g. Karachentsev et al., 2004). Voids with diameters ~1-2 Mpc may host only very few faint galaxies. The nearest rich cluster of galaxies in the Virgo constellation is situated at a distance of ~17 Mpc from us and contains roughly two orders of magnitude more galaxies of different luminosities than a typical poor group. The number of known small galaxies grows with the progress of observational facilities. In our days the accurate measurement of distances and photometric properties of individual representatives of the brightest stellar populations is possible up to ~20 Mpc thanks to images of high spatial resolution from the Hubble Space Telescope. Our knowledge of the wide variety of properties of dwarf galaxies is close to be complete only for the objects within the Local Group (LG) (distance from us less than ~1 Mpc), i.e. for the satellites of the Milky Way and M31.

Low surface brightness (LSB) dwarf galaxies are fainter than $M_B = -16^m$. They represent the most common type of objects in the local Universe. The number density of dwarfs observed in nearby rich clusters may approach 50-100 Mpc^{-3}. In distinction to GCs, dwarf galaxies are dark-matter dominated as evidenced by the large velocity dispersions of their stars. Like their normal cousins, dwarfs may be of early or late morphological type. Spheroidals (dSph) are gas-poor, old stellar systems. Irregulars (dIrr) contain young stars and neutral hydrogen, the fuel for star formation activity. Some objects, classified as transitional types (dIrr/dSph), show evidence for a few young and intermediate-age stars but not for any gas, or contain very small amount of it, which is below the detection level of most modern telescopes. They probably lost their gas during some recent powerful star-forming events, because of their shallow potential well. Indeed, radio observations show the presence of HI clumps around some faint galaxies (Begum et al., 2008). The strongest star-forming bursts covering entire galaxies occur in blue compact irregular dwarfs (BCDs). They have therefore a high surface brightness in the blue and ultraviolet spectral ranges.

The origin of dwarf galaxies is under debate. What processes drive morphological transformations? Why do objects lose their gas? Some clues to the solution of the problem are in the observational properties of these objects. Dwarf galaxies may experience ram pressure or tidal stripping when moving through the dense hot gaseous halo of their massive neighbours. Indeed, it is a well-known fact that the majority of dSphs are located at a projected distance of less than ~200 kpc form the central galaxy in groups and clusters (Karachentsev et al., 2005). The starvation mechanism, which may play a role, means that the gravitational potential of a group or cluster creates a Roche lobe around any individual galaxy. The gas tidally stripped from objects in violent star-forming events, like supernovae bursts, or birth of star clusters, is acquired by the galaxy ensemble (Shaya & Tully, 1984). Some observational facts are hard to explain theoretically. For example, the existence of the isolated dSph KKR25 situated in the outskirts of the LG (Karachentsev et al., 2001), and of the isolated low-mass lenticular (S0) NGC404 (Karachentsev & Makarov, 1999). It is important to compare dynamical processes driving the star formation and chemical evolution of dwarf and normal galaxies, in order to understand better the reason for the sharp differences in their properties.

The aforementioned observational facts qualitatively depict the properties of dwarf galaxies and outline problems concerning their formation and evolution. In the following we will examine statistical facts and relations. Are the properties of dSphs, dIrrs, and dIrr/dSphs different? Table 1 summarizes observational data for dwarf galaxies. The information for the LG is taken from Mateo (1998), Grebel et al. (2003), and Karachentsev et al. (2004). The catalog of Karachentsev et al. (2004) and the results presented by Sharina et al. (2008) are used for the objects situated in the LV. We consider (V-I) (visual minus near-infrared) and (B-V) (blue minus visual) mean integrated colors corrected for Galactic extinction, total mass of neutral hydrogen, mean metallicity of red giant branch stars, mean surface brightness in the V band corrected for Galactic extinction, absolute magnitude in the V-band corrected for Galactic extinction. Table 1 shows that dIrrs are bluer, richer in HI, poorer in metals, brighter in the visual band, than dSphs and intermediate-type galaxies. Suppose that our three samples were drawn from the same normally distributed parameter space. To check our hypothesis we consider each property independently, and apply the F-test (comparison of variances) and Student's t-test (comparison of means) (Stuart & Ord, 1994). One can see from Table 2 that the samples of dSphs and dIrrs are different that at a high significance level. However, some of the photometric properties of transitional-type galaxies appear to be similar to the ones of dSphs, or dIrrs. Other characteristics of dSph/dIrr are different from those of early and late-type dwarfs at significance levels of 5, or 10 per cent. So, our simple statistical tests confirm the transitional nature of this rare type of galaxies.

Table 1. Observational properties of LSB dwarf galaxies in the Local Group (LG) and in the Local Volume (LV)

	dSph	dSph/dIrr	dIrr
$(V\text{-}I)_{LG}$	1.2±0.20	1.00 ± 0.10	0.73±0.2
$(V\text{-}I)_{LV}$	1.1±0.12	0.90 ± 0.12	0.74 ±0.2
$(B\text{-}V)_{LG}$	0.7±0.10	0.68 ± 0.10	0.46 ±0.2
$(B\text{-}V)_{LV}$	0.8±0.20	0.57 ± 0.10	0.48 ±0.2
Mass HI $_{LG}$	[1.8 10^3; 1.3 10^5]	[6.3 10^4 ; 1.5 10^5]	[7.2 10^5 ; 1.0 10^7]
Mass HI $_{LV}$	not detected	[8.0 10^5 ; 3.3 10^6]	[2.0 10^6 ; 5.0 10^8]
$[Fe/H]_{LG}$	-1.56 ± 0.31	-1.80 ± 0.14	-1.90 ± 0.31
$[Fe/H]_{LV}$	-1.44 ± 0.23	-1.60 ± 0.31	-1.94 ± 0.32
SB $V_{0, LG}$	24.61 ± 0.96	23.90 ± 0.85	22.43 ±1.14
SB $V_{0, LV}$	23.36 ± 1.09	22.75 ± 0.76	22.36 ±1.10
$M_{V\ LG}$	-11.31 ± 2.17	-10.86 ± 1.18	-12.89 ± 1.81
$M_{V\ LV}$	-11.07 ± 1.66	-12.79 ± 1.99	-12.97 ± 1.73

Note however that this analysis does not allow one to study different observational properties and groups of properties in a multivariate setup which are more responsible for the parent galaxy evolution. This may be carried out by applying methods of Principal Component Analysis (PCA) (Pearson, 1901), and Cluster Analysis (CA)**.** PCA is one of the most common multivariate statistical methods used in astronomy (see e. g. Strader & Brodie, 2004). It is used to reduce the large number of parameters to a smaller one which gives the maximum variation among the objects under consideration. For instance, Prada & Burkert (2002) discovered a new correlation between the mass-to-light ratio and the mean metallicity for the satellites of the LG, using a classical PCA. In the following, we summarize the linear correlations between the properties of dwarf galaxies studied up to date, which are a guide for future work using classical, modified PCA and CA methods. Each relation is physically explained.

There is a linear correlation between *the luminosity* of galaxies and the mean *metallicity* of their old stars for early-type dwarf galaxies (see Sharina et al., 2008 and references therein). Its functional form, $L \sim Z^{2.5}$, coincides with that predicted by the model of Dekel & Silk (1986), according to which the evolution of small galaxies is mainly regulated by supernova-driven winds. A similar relation exists for dIrrs (Grebel et al., 2003). However, dIrrs have old stellar populations of lower mean metallicity at a given optical luminosity in comparison with dSphs, and the scatter in the relation is large. *The surface brightness - luminosity* relation ($L \sim R^4$) is also well explained by the fact that galaxies with a mass below some critical value undergo substantial gas loss as a result of a violent bursts of star formation (Dekel & Silk, 1986). The central and effective surface brightness appears to correlate simultaneously with the absolute magnitudes of galaxies in different photometric bands: $SB \sim 0.5\ M_V$. This relation is equivalent to the aforementioned luminosity – radius relation. On the other hand, the mean surface brightness within the larger isophote, 25 mag./sq.arsec, correlates with the absolute magnitude in a different way: $SB_{25} \sim 0.33\ M_V$, which implies a constancy of galactic spatial density averaged within this isophote.

The exponential scale length – luminosity relation for dwarf galaxies (Sharina et al., 2008) shows a behaviour different from that for galaxies more massive than M_{crit} ~10^8 $Mass_{Sun}$. This limit corresponds to a critical luminosity M_B ~ -12 -13 mag. (Bell & de Jong, 2001). For galaxies more massive than this limit the scale length is proportional to $L^{0.5}$, which implies the constancy of the central surface brightness. The change of slope of this relation is presumably caused by the fact that less massive objects lose their gas more efficiently.

Table 2. Student's t-test statistics quantifying the probability that the photometric properties (Table 1) for the LG and the LV samples of dSphs, dIrrs and transitional type dwarfs were drawn from the same distribution. Numbers of degrees of freedom are given in brackets. Critical values of T-distribution for each photometric parameter at a significance levels of 5 per cent for the three samples taken in pairs are also presented. The critical values at significance levels of 10 per cent (in brackets) are shown for some cases (see text for details).

	DSph -- dSph/dIrr	dSph/dIrr -- dIrr	dIrr -- dSph
T (V-I) $_{LG}$	1.02 (N=15)	1.79 (N=3)	6.21 (N=15)
T (V-I) $_{LV}$	1.76 (N=21)	1.11 (N=3)	7.57 (N=56)
$t_{0.95, LG}$	1.75 (1.34)	1.75	1.70
$t_{0.95, LV}$	1.72	1.67 (1.30)	1.67
T ([Fe/H] $_{LG}$)	0.77 (N=18)	0.31 (N=3)	2.51
T ([Fe/H] $_{LV}$)	0.65 (N=14)	1.47 (N=3)	5.33 (N=14)
$t_{0.95, LG}$	1.73	1.73	1.70 (N=36)
$t_{0.95, LV}$	1.76	1.69 (1.31)	1.68
T (SB V $_{0, LG}$)	1.19 (N=17)	1.97 (N=3)	3.12 (N=8)
T (SB V $_{0, LV}$)	0.75 (N=21)	0.49 (N=3)	3.49 (N=56)
$t_{0.95, LG}$	1.73 (1.33)	1.81	1.71
$t_{0.95, LV}$	1.72 (1.32)	1.67	1.66
T ($M_{V\ LG}$)	0.40 (N=18)	2.00 (N=5)	2.24 (N=15)
T ($M_{V\ LV}$)	1.37 (N=21)	0.14 (N=3)	4.34 (N=56)
$t_{0.95, LG}$	1.72 (1.32)	1.73	1.70
$t_{0.95, LV}$	1.72 (1.32)	1.67	1.66

3. PROPERTIES OF GLOBULAR CLUSTERS

GCs are the brightest representatives of the oldest, simplest stellar populations. Their physical properties are often used for better understanding the evolution of dwarf galaxies and their role as building blocks in the cosmological structure formation. The chemical tagging of hierarchical building blocks in different types of galaxies or, in other words, reconstructing the star formation and assembly histories of galaxies by identifying "tracer" populations has been a popular idea in the scientific community for many years (see for a review West et al., 2004, and references therein). The existence of chemical similarities between GC systems of dwarf and giant galaxies may indicate that a certain fraction of GCs in giant galaxies were accreted from satellite dwarf galaxies.

The metallicity, age, and alpha-element abundance ratio of GCs were derived in a series of papers devoted to the results of medium-resolution spectroscopy obtained at the 6m telescope of the Russian Academy of Sciences and the Very Large Telescope of the European Southern Observatory: Sharina, Afanasiev, and Puzia, 2006a, b; Sharina, Puzia & Krylatyh, 2007; Puzia & Sharina, 2008; Sharina & Davoust, 2009. A large amount of data is still in preparation for publication.

Astronomers often find difficulties in interpreting the “age-metallicity degeneracy problem”, when they deal with fluxes (I), or magnitudes (-2.5 log10(I)) of stellar ensembles integrated within a wide spectral range (~1000 Angströms in the commonly used photometric Johnson-Cousins system), and colors (differences between two magnitudes). Bluer colors may indicate that the maximum of the relative energy distribution is shifted toward the blue. This may be caused by the young age or low metallicity of the object. When we compare our observational colors with the model ones, the same data may be explained by the effect of either metallicity or age. The method of Lick indices is the most commonly used for disentangling age-metallicity degeneracy effects in the integrated spectra of GCs (e.g. Proctor et al., 2004). Metal lines arise from the coolest stars (red giant branch and low Main Sequence stars). Balmer lines of Hydrogen arise from the hottest stars (Main Sequence Turn Over, and Horizontal Branch stars). So, Lick indices centered on different spectroscopic absorption-line features allow one to break the age-metallicity degeneracy. To derive the evolutionary parameters of GCs we measured absorption-line indices in the spectra using the pass-band definitions of the well-known and widely used Lick system (Worthey, 1994, Worthey et al., 1994, and Worthey & Ottaviani, 1997). To bring the measured indices into correspondence with the standard system, we calibrated the instrumental system of our telescope and spectrograph into the standard one by using stars from the list of Worthey et al. (1994) observed during the same night as the program GCs. We compared the measured indices with Simple Stellar Population model predictions from Thomas et al. (2003, 2004). For this a three-dimensional interpolation and chi-square minimization routine was developed (Sharina et al., 2006; see also Sharina & Davoust, 2009). In the following we summarize our results on the chemical composition of GCs in dwarf galaxies presented in the published papers.

We found that the alpha-element abundance ratio tends to be low ([α/Fe]~0.0) in dwarf galaxies in comparison to the corresponding values for GCs in the Milky Way, M31 and other giant galaxies, where [α/Fe]~0.3. This trend was present for each observational dataset presented in the quoted papers and is seen in the cumulative histogram shown in Fig.1. We present here the probability density function (PDF) for all metallicity estimates obtained spectroscopically. The line indicates a non-parametric density estimate using an Epanechnikov kernel (Epanechnikov, 1969). For an explanation of the diagram as a density see Freedman & Diaconis (1981). The peak in the [α/Fe] distribution is defined with an accuracy of 0.08 dex. Alpha-particle capture elements are mainly produced in type II supernovae. The progenitors of SNe II are massive stars with short lifetimes (~10^7 years). Type Ia supernovae, produced by low-mass long-lived stars, eject mainly iron-peak elements. This is why abundance ratios can be used as cosmic clocks indicating the evolutionary timescales.

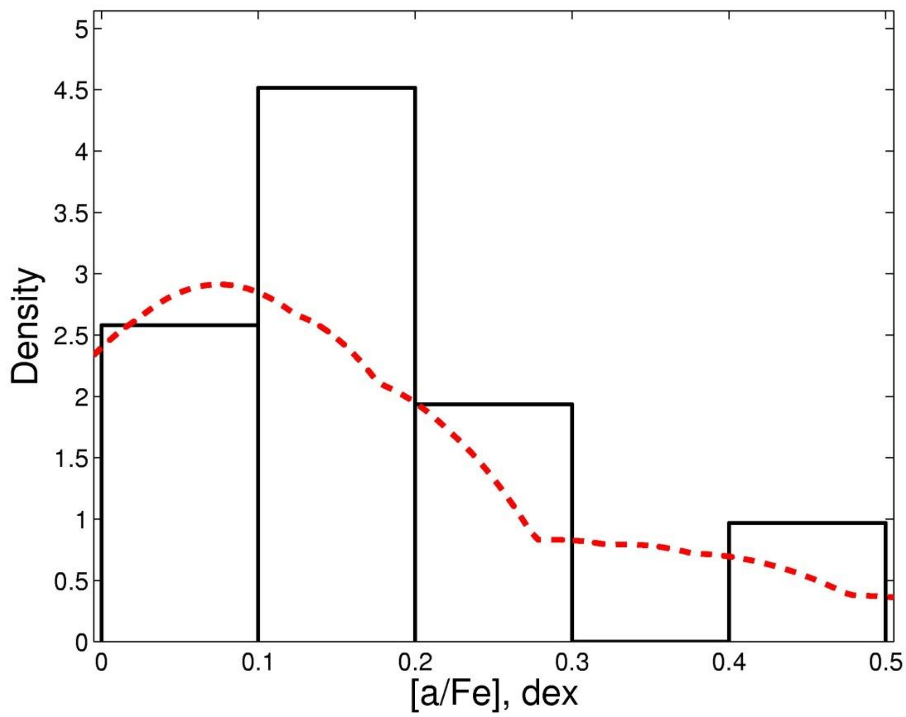


Figure 1. Probability density function of alpha-element abundance ratios for GCs in dwarfs galaxies observed spectroscopically at the 6m telescope and VLT (see text for details).

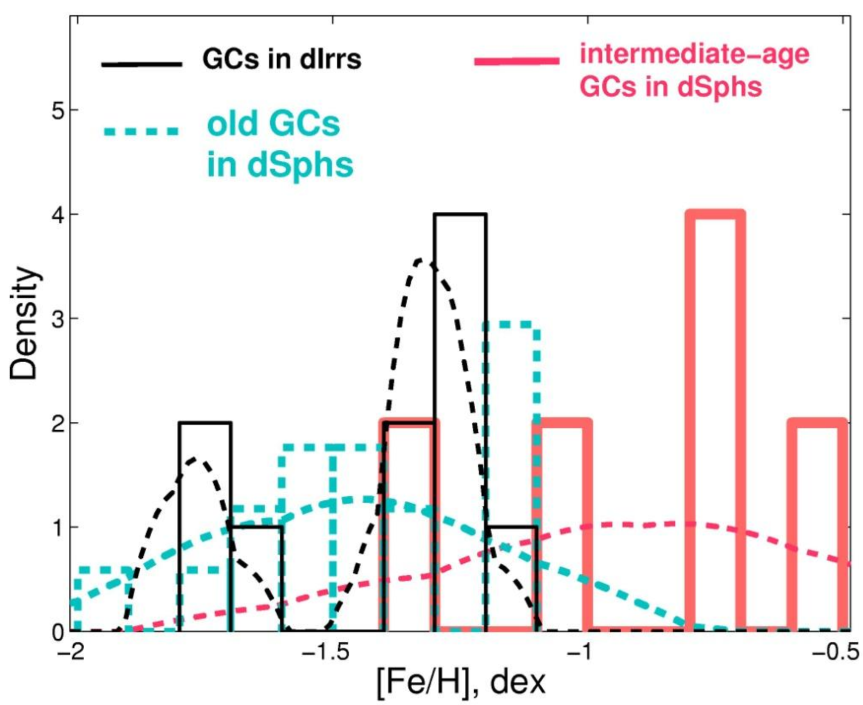


Figure 2. Probability density function of [Fe/H] for old and intermediate-age GCs in dwarfs galaxies observed spectroscopically (see text for details).

Metallicities tend to be low in dwarf galaxies in comparison to the normal ones. PDFs for star clusters in dIrrs, old and intermediate-age GCs in dSph compiled using all the published data are shown in Fig.2. The functions were produced using the same methods as for the [a/Fe] distribution described above. The two peaks for dIrrs near [Fe/H]~-1.3 dex, and -1.77 dex are defined with the accuracies ~0.05 dex using the non-parametric density estimate with an Epanechnikov kernel. The probability-density peaks for old and intermediate-age dSphs are [Fe/H]~ -1.45, and -1.02 dex, respectively. They are defined with the accuracies ~0.2 dex. Note, that [Fe/H] of old and intermediate-age GCs in dSphs is systematically higher than for dIrrs. In general, GCs are representatives of old stellar populations, and their chemical compositions reflect physical parameters and processes in the interstellar medium existing hundreds million years ago. In the case of isolated galaxy, massive enough to keep its stellar ejecta, the metallicity of stars and interstellar medium is well described by the close-box model of continuous star formation (Layden & Sarajedini, 2000). If there are no inflows of fresh gas, the metallicity of youngest stellar populations is expected to be high ([Fe/H]>-0.5). The fact that the interstellar medium of dwarf galaxies is poor in heavy elements probably means that their gravitational potentials are not strong enough to prevent stellar outflows to leave galaxies. On the other hand, systematically higher metallicity of old and intermediate-age GCs in dSphs likely indicates their larger mass in the past. If we compare the metallicity distribution of GCs in dwarf galaxies with the one for the three groups of GCs in NGC5128 selected basing on PCA analysis according to their half-light radii (Chattopadhyay et al., 2009), we can see that the low-metallicity peaks for the two groups of GCs in NGC5128 having presumably external origin lie close to the peak value of [Fe/H] shown in Fig. 2.

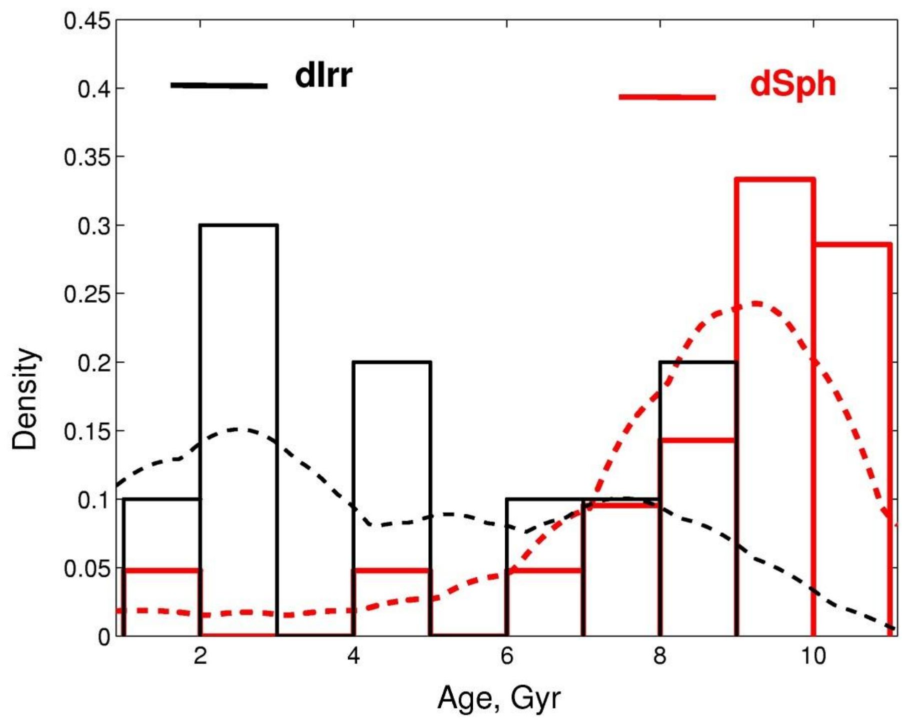


Figure 3. Probability density functions of age for GCs in dSphs and dIrrs observed spectroscopically (see text for details).

Our spectroscopic observations reveal evidence for only intermediate-age and old GCs in dSphs, and for star clusters in a wide range of ages in dIrrs (see Fig. 3). The peaks in the PDFs presented in Fig.3 are ~9 Gyr for dSphs and ~2.5 Gyr for dIrrs. They were derived with uncertainties of 0.8 Gyr and 1.9 Gyr, respectively, using the non-parametric density estimate with an Epanechnikov kernel. In distinction to late-type galaxies, early-type dwarfs often contain super-massive GCs near their optical centers. There are young and embedded star clusters in the centers of most massive higher surface brightness dSphs. These are the satellites of M31 NGC205 and 185 (Sharina et al., 2006). This means that LSB dwarfs had enough gas and specific physical conditions for star cluster formation only in the early stages of their evolution. Star formation is rather stochastic in dIrrs. Star-forming regions are distributed more randomly and chaotically.

The number of massive gravitationally bound star clusters per unit mass and luminosity in dwarf galaxies is much higher than in giant galaxies (Sharina et al., 2005, Georgiev et al., 2008). The formation of a star cluster with a mass 1-10% of the total galaxy stellar mass is a cataclysmic event in the life of a small galaxy that truncates further star formation for some period of time (Silk et al., 1987). This process demands high densities and pressures in the interstellar medium. Star clusters in dwarf galaxies are expected to be more compact in general due to high virial densities in stars, gas, and dark matter. However, they are not all compact (Da Costa et al., 2009). There are two peaks in the distribution of half-light radii at 3 pc and 10 pc. The most promising discoveries in the literature concerning half-light radii (r_h) of GCs are probably the following: 1. There is a correlation between r_h of GCs and their galactocentric distance (van den Bergh, 2000) ; 2. The theoretical prediction that r_h remains constant during the dynamical evolution of GCs has been confirmed by the analysis of a large sample of GCs in early-type galaxies in the Virgo cluster (Jordán et al., 2005); 3. Unusual GCs exist in the outskirts of M31, in M33 and LSB dwarf galaxies with r_h intermediate between star clusters and dwarf galaxies (Huxor et al., 2005, Da Costa et al., 2009).

The structural parameters, luminosities, masses of star clusters in galaxies change in the course of their dynamical evolution in different ways depending on the initial mass and environmental conditions. GCs moving through the dense gaseous parts of giant galaxies destroy the faster the lower their mass is. These parameters are also used by PCA and CA to study galaxy formation process (e.g. Chattopadhyay et al., 2009). The main aim is to derive a main physical parameter or parameter space influenced by evolution. PCA and CA are extremely useful to separate groups of GCs in large galaxies, which show the largest differences in all their properties. Some of the groups may be accreted from dwarf satellites. Chattopadhyay, Chattopadhyay, Davoust, Mondal, & Sharina (2009) discovered three groups of GCs in the giant elliptical NGC 5128 using PCA and CA, and derived the mean properties of the three groups. The three groups differ in their mean masses, metallicities, and origin.

4. COMMENTS AND CONCLUSION

We obtained a large amount of observational data on GCs in dwarf galaxies and complemented them by data published in the literature. The derived parameters are: 1. Lick indices, which are centered on different absorbtion-line features, and therefore sensitive to age, metallicity, and abundances of different chemical elements; 2. colors measured in different pass-bands; 3. evolutionary parameters obtained by comparison of the observed data with Simple Stellar

Population models; 4. luminosities and masses of GCs; 5. structural parameters (radii, central densities, concentration indices, etc.) .

Dwarf galaxies reveal properties which can allow us to further improve their morphological classification. Some of the derived observational parameters (metallicity, mass, luminosity, spatial density, size, velocity dispersion) are correlated. The chemical composition and evolutionary parameters of GCs in dwarf and giant galaxies are used for determining evolutionary timescales for the host galaxies formation, and for testing the hierrarchical structure formation scenario.

The goal of our future studies is to derive components in the parameter space of various observational properties of GCs and their host galaxies, which are responsible for the parent galaxy's evolution.

REFERENCES

Bell E. F., de Jong R. S. (2001) Stellar Mass-to-light Ratios and the Tully-Fisher Relation, The Astrophysical Journal, 550, 212-229.

Begum, A., Chengalur, J. N., Karachentsev, I. D., Sharina, M. E., Kaisin, S. S. (2008). FIGGS: Faint Irregular Galaxies GMRT Survey - overview, observations and first results, Monthly Notices of the Royal Astronomical Society, 2008, 386, 1667-1682.

Chattopadhyay A. K., Chattopadhyay T., Davoust E., Mondal S., Sharina M. E. (2009) Study of NGC5128 Clusters under Multivariate Statistical Paradigm, The Astrophysical Journal 705, 1533-1547.

Da Costa, G. S.; Grebel, E. K.; Jerjen, H.; Rejkuba, M.; Sharina, M. E. (2009), Scl-dE1 GC1: An Extended Globular Cluster in a Low-Luminosity Dwarf Elliptical Galaxy, The Astronomical Journal, 137, 4361-4367.

Dekel A., Silk J. (1989) The Origin of Dwarf Galaxies, Cold Dark Matter, and Biased Galaxy Formation, The Astrophysical Journal, 303, 39-55.

Epanechnikov V. A. (1969) Theory Probab. Appl., 14, 153.

Georgiev, I. Y.; Goudfrooij, P.; Puzia, T. H.; Hilker, M. (2008) Old Globular Clusters in Magellanic-Type Dwarf Irregular Galaxies, The Astronomical Journal, Volume 135, 1858-1876.

Grebel, E. K.; Gallagher, J. S., III; Harbeck, D. (2003) The Progenitors of Dwarf Spheroidal Galaxies, The Astronomical Journal, 125, 1926-1939.

Huxor A. P., et al. (2005) A new population of extended, luminous star clusters in the halo of M31, Monthly Notices of the Royal Astronomical Society, 360, 1007.

Jordán, A., et al. (2005) The ACS Virgo Cluster Survey. X. Half-Light Radii of Globular Clusters in Early-Type Galaxies: Environmental Dependencies and a Standard Ruler for Distance Estimation, The Astronomical Journal, 634, 1002.

Karachentsev I. D., Karachentseva V. E., Sharina M. E. (2005) Dwarf spheroidal galaxies in nearby groups imaged with HST, in Near-Field Cosmology with Dwarf Elliptical Galaxies, ed. H. Jergen and B. Binggeli, IAUS 198, 295-302.

Karachentsev, I. D., Karachentseva V. E., Huchtmeier W. K., Makarov D. I. (2004) A Catalog of Neighboring Galaxies, The Astronomical Journal, 127, 2031-2068.

Karachentsev, I. D., et al. (2001), A new galaxy near the Local Group in Draco, Astronomy & Astrophysics, 379, 407-411.

Karachentsev, I. D., Makarov I. D. (1999) Galaxy Interactions in the Local Volume, in Galaxy interactions at low and high redshifts, ed. J.E. Barnes, D.F. Sanders, IAUS 186, 109-116.

Layden, A. C., Sarajedini, A. (2000) The Astronomical Journal, 119, 1760.

Mateo M. L. (1998) Annual Review of Astronomy and Astrophysics, 36, 435-506.

Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space". Philosophical Magazine 2 (6): 559–572.

Peebles P. J. E., Dicke R. H. (1968) Origin of the Globular Star Clusters, The Astrophysical Journal, 154, 891.

Prada F., Burkert, A. The Fundamental Line of the Local Group Satellites (2002) The Astrophysical Journal, 564, L73-L76.

Proctor, R. N.; Forbes, D. A.; Beasley, M. A. (2004) A robust method for the analysis of integrated spectra from globular clusters using Lick indices, Monthly Notices of the Royal Astronomical Society, Volume 355, 1327-1338.

Puzia T. H., Sharina M. E. (2008) VLT Spectroscopy of Globular Clusters in Low Surface Brightness Dwarf Galaxies, The Astrophysical Journal, 674, 909-926.

Sharina et al. (2008) Photometric properties of Local Volume dwarf galaxies, Monthly Notices of the Royal Astronomical Society, 384, 1544-1562.

Sharina, M.; Davoust, E. (2009) Globular cluster content and evolutionary history of NGC 147, Astronomy and Astrophysics, 497, 65-80.

Sharina, M. E.; Puzia, T. H.; Krylatyh, A. S. (2007) A globular cluster in the dwarf galaxy Sextans B, Astrophysical Bulletin, 62, 209-216.

Sharina, M. E.; Afanasiev, V. L.; Puzia, T. H. (2006) Ages, metallicities and [α/Fe] ratios of globular clusters in NGC 147, 185 and 205, Monthly Notices of the Royal Astronomical Society, 372, 1259-1274.

Shaya E. J., Tully R. B. (1984) The Astrophysical Journal, 281, 56-66.

Silk J., Wyse K. F. G., Shields G. A. (1987) The Astrophysical Journal, 322, L59.

Starder J., Brodie J. P. (2004) A principal component analysis of the Likck indices of Galactic globular clusters, The Astronomical Journal, 128, 1671-1675.

Van den Bergh, S. (2000) The Astrophysical Journal, 530, 777.

West et al., (2004) Nature, 427, 31.

Worthey G., Ottaviani D.L. (1997) ApJS, 111, 377.

Worthey (1994) ApJS, 95, 107.

Worthey G., Faber S. M., Gonzalez J. J., Burstein D. (1994) ApJS, 94, 687.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 303–303.

SESSION 8: DIRECTIONAL STATISTICS

Chair: A. Ghapor Hussein
Centre for Foundation Studies in Science, Universiti of Malaya,
50603 Kuala Lumpur, Malaysia.
E-mail: ghapor@um.edu.my

MIDDLE-CENSORING FOR CIRCULAR DATA

S. Rao Jammalamadaka
Department of Statistics and Applied Probability, University of California, Santa Barbara, CA. 93106 USA
E-mail: rao@pstat.ucsb.edu

ABSTRACT

A generalized censoring scheme in the survival analysis context was introduced by Jammalamadaka and Mangalam (2003, Jour. of Nonparametric Statistics, pp.253-265). In this talk we discuss how such a censoring scheme applies to circular data and in particular when the original data is assumed to come from a parametric model such as the von Mises. Maximum likelihood estimation of the parameters as well as their large sample properties are considered under this censoring scheme. Monte Carlo comparisons are made with alternate estimators of the mean direction and concentration.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 303–304.

OUTLIER DETECTION IN CIRCULAR SAMPLES

Ibrahim Mohamed
Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia
E-mail: imohamed@um.edu.my

ABSTRACT

One of the most common problems that arise in statistical analysis is the existence of some unexpected observations in data set. Beckman and Cook (1983) and Barnett and Lewis (1984) reviewed the literature on outliers in various areas of statistical data. In this presentation, we are interested to study the occurrence of single outlier in directional or circular data. To date, a number of discordancy tests are available for circular samples. Three of them, the C-statistic, D-statistic and M-statistic were suggested by Collett (1980) and the L-statistic was suggested earlier by Mardia (1975). Recently, Abuzaid *et al.* (2009) proposed an alternative statistics called A-statistic. Via simulation studies, they showed that the statistic performs well compared to other statistics for data generated from von Mises distributions. The statistics is further

investigated for other circular distributions. The percentage points are calculated and the performance is examined. We compare the performance with the tests. As an illustration a practical example is presented.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 304–304.

ANALYSIS OF MISSING VALUES FOR CIRCULAR DATA

Yong Zulina Zubairi
Centre for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur, Malaysia
E-mail: yzulina@um.edu.my

ABSTRACT

In this paper, the analysis of missing values for circular data is discussed. Circular data are rather special because they are in a closed form over the range of $[0^o, 360^o)$. Hence, all existing procedures that have been used in treating the missing values in linear data are not longer valid for circular data. Imputation methods for missing values data are proposed in the paper. Two different methods namely circular mean by column and sample mean are used. In this study, missing values were tested in parameter estimation for simultaneous linear functional relationship model for circular variables. From the simulation studies, it shows that the proposed method provide an adequate approach in handling missing values for circular variables.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 305–316.

SIMULTANEOUS LINEAR FUNCTIONAL RELATIONSHIP FOR CIRCULAR VARIABLES

Abdul Ghapor Hussin, Siti Fatimah Hassan and Yong Zulina Zubairi
Centre for Foundation Studies in Science, Universiti of Malaya,
50603 Kuala Lumpur, Malaysia.
E-mail: ghapor@um.edu.my

ABSTRACT

This paper proposed a statistical model to compare or describe the relationship between several circular variables subject to measurements errors. The model is known as the simultaneous linear functional relationship for circular variables, which is an extension of the linear functional relationship model. Maximum likelihood estimation of parameters has been obtained iteratively by assuming the ratios of concentration parameters are known and by choosing suitable initial values. The variance and covariance of parameters have been derived using the Fisher information matrix. The model was applied to the Malaysian wind direction data recorded at various levels.

Keywords and phrases: Simultaneous linear function relationship model, circular variables, von Mises distribution, concentration parameters, wind direction data.

1. INTRODUCTION

Circular variables can be defined as one which takes values on the circumference of a circle and they are in the form of angles in the range $\left(0^0, 360^0\right)$ or $\left(0, 2\pi\right)$ radians. These types of data are widely observed in geological, meteorological, biological and astronomical studies. A simple linear functional relationship model for circular variable refer to a two mathematical circular variables *X* and *Y* which are observed inexactly, linearly related by $Y = \alpha + \beta X$ (mod 2π). This model was first introduced by Hussin (1997). It also assume that the errors of circular variables *X* and *Y* are independently distributed and follow the von Mises with probability density function given by $g\left(\mu_0, \kappa; \theta\right) = \left(2\pi I_0\left(\kappa\right)\right)^{-1} \exp\left\{\kappa \cos\left(\theta - \mu_0\right)\right\}$ where $I_0\left(\kappa\right)$ is the modified Bessel function of the first kind and order zero; Mardia (1972). The parameter μ_0 is the mean direction while the parameter κ is described as the concentration parameter.

An application when both *X* and *Y* are circular, for example in modeling a relationship for calibration between two instruments for wind direction, requires that $\beta \approx 1$. As explained by Hussin et al. (2004), the practical considerations often imply that $\beta \approx 1$ is the most meaningful model, even though a larger integer value (typically very much larger) may give a model which apparently fits the data slightly better. Caires and Wyatt (2003) also proposed similar model by assuming $\beta = 1$ in the study of the assessment of ocean wave measurements.

This study considers a linear functional relationship between all pairs of a set of *X* and circular random variables Y_j (j = 1, …, q). The proposed model is known as simultaneous linear functional relationship model for circular variables which can be used, for example in assessing the relative relationship of a set of q circular random variables. As an illustration, this model will be used in studying the fundamental relationship of meteorological data, which is Malaysian wind direction data which have been recorded at different levels.

2. THE MODEL

Suppose the circular variables $Y_j(j=1,\ldots,q)$ are related to X by the linear functional relationship $Y_j=\alpha_j+\beta_j X$ (mod 2π). Let (X_i, Y_{ji}) be the true values of the circular variables X and Y_j respectively and we assume that the observations x_i and y_{ji} have been measured with errors δ_i and ε_{ji} respectively. Thus, the full model of simultaneous linear functional relationship for circular variable can be written as

$$x_i = X_i + \delta_i \text{ and } y_{ji} = Y_{ji} + \varepsilon_{ji}, \text{ where}$$
$$Y_j = \alpha_j + \beta_j X (\text{mod } 2\pi), \text{ for } j = 1,\ldots,q.$$

We also assume δ_i and ε_{ij} are independently distributed with von Mises distribution that is,

$$\delta_i \sim VM(0,\kappa) \text{ and } \varepsilon_{ji} \sim VM(0,\nu_j) \text{ respectively.}$$

Suppose that we assume the ratio of the error concentration parameters of the model, $\lambda_j = \dfrac{\nu_j}{\kappa}$ is known. Then, the likelihood function is,

$$\begin{aligned}
&L(\alpha_j, \beta_j, \kappa, X_1,\ldots,X_n; \lambda_j, x_1,\ldots,x_n, y_{11},\ldots,y_{qn}) \\
&= (2\pi)^{-2n} I_0^{-n}(\kappa) \sum_{j=1}^{q} I_0^{-n}(\lambda_j\kappa) \sum_{i=1}^{n} \exp\{\kappa\cos(\delta_i)\} \sum_{j=1}^{q} \exp\left\{\lambda_j\kappa\sum_{i=1}^{n}\cos(\varepsilon_{ji})\right\} \\
&= (2\pi)^{-2n} I_0^{-n}(\kappa) \sum_{j=1}^{q} I_0^{-n}(\lambda_j\kappa) \sum_{i=1}^{n} \exp\{\kappa\cos(x_i - X_i)\} \sum_{j=1}^{q} \exp\left\{\lambda_j\kappa\sum_{i=1}^{n}\cos(y_{ji} - \alpha_j - \beta_j X_i)\right\}
\end{aligned}$$

By taking the logarithm to likelihood function, the equation above becomes,

$$\begin{aligned}
&\log L(\alpha_j, \beta_j, \kappa, X_1,\ldots,X_n; \lambda_j, x_1,\ldots,x_n, y_{11},\ldots,y_{qn}) = \\
&-2n\log(2\pi) - n\log I_0(\kappa) - n\sum_{j=1}^{q}\log I_0(\lambda_j\kappa) + \kappa\sum_{i=1}^{n}\cos(x_i - X_i) + \kappa\sum_{j=1}^{q}\lambda_j\sum_{i=1}^{n}\cos(y_{ji} - \alpha_j - \beta_j X_i)
\end{aligned}$$

By assuming that $\lambda_j = 1$, for instance in the relationship calibration of the same characteristic, the equation above can be simplified as,

$$-2n\log(2\pi) - (1+q)n\log I_0(\kappa) + \kappa\sum_{i=1}^{n}\cos(x_i - X_i) + \kappa\sum_{j=1}^{q}\sum_{i=1}^{n}\cos(y_{ji} - \alpha_j - \beta_j X_i).$$

3. PARAMETER ESTIMATION

There are $(2q+n+1)$ parameters to be estimated, i.e. $\alpha_1,\ldots,\alpha_q$, $\beta_1,\ldots,\beta_q$, κ and incidental parameters $X_1,\ldots,X_n$ by using maximum likelihood estimation (MLE). By differentiating log L with respect to the parameters α_j, β_j, κ and X_i, we can obtain the parameters $\hat{\alpha}_j$, $\hat{\beta}_j$, $\hat{\kappa}$ and $\hat{X}_i$ as follows :

3.1 MLE for α_j

The first partial derivative of the log likelihood function with respect to α_j is

$$\frac{\partial \log L}{\partial \alpha_j} = \kappa \sum_i \sin\left(y_{ji} - \alpha_j - \beta_j X_i\right).$$

By setting $\frac{\partial \log L}{\partial \alpha_j} = 0$ and simplifying the equation we get,

$$\hat{\alpha}_j = \begin{cases} \tan^{-1}\left\{\dfrac{S}{C}\right\} & S > 0, C > 0 \\ \tan^{-1}\left\{\dfrac{S}{C}\right\} + \pi & C < 0 \\ \tan^{-1}\left\{\dfrac{S}{C}\right\} + 2\pi & S < 0, C > 0 \end{cases}$$

where $S = \sum_i \sin\left(y_{ji} - \hat{\beta}_j \hat{X}_i\right)$ and $C = \sum_i \cos\left(y_{ji} - \hat{\beta}_j \hat{X}_i\right)$.

3.2 MLE for β_j

The first partial derivative of the log likelihood function with respect to β_j is,

$$\frac{\partial \log L}{\partial \beta_j} = \kappa \sum_i X_i \sin\left(y_{ji} - \alpha_j - \beta_j X_i\right).$$

$\hat{\beta}_j$ cannot be obtained analytically and may be obtained iteratively. By setting $\frac{\partial \log L}{\partial \beta_j} = 0$ and suppose $\hat{\beta}_{j0}$ is an initial estimate of $\hat{\beta}_j$, then we can simplify the above equation and approximately given by:

$$\hat{\beta}_{j1} \approx \hat{\beta}_{j0} + \frac{\sum_i \hat{X}_i \sin\left(y_{ji} - \hat{\alpha}_j - \hat{\beta}_{j0}\hat{X}_i\right)}{\sum_i \hat{X}_i^2 \cos\left(y_{ji} - \hat{\alpha}_j - \hat{\beta}_{j0}\hat{X}_i\right)}$$

where $\hat{\beta}_{j1}$ is an improvement of $\hat{\beta}_{j0}$.

3.3 MLE for X_i

The first partial derivative of the log likelihood function with respect to X_i is,

$$\frac{\partial \log L}{\partial X_i} = \kappa \sin(x_i - X_i) + \kappa \sum_j \beta_j \sin(y_{ji} - \alpha_j - \beta_j X_i).$$

By setting $\frac{\partial \log L}{\partial X_i} = 0$ and suppose X_{i0} is an initial estimate of $\hat{X}_i$, we then get the equation approximately given by:

$$\hat{X}_{i1} \approx \hat{X}_{i0} + \frac{\sin(x_i - \hat{X}_{i0}) + \sum_j \hat{\beta}_j \sin(y_{ji} - \hat{\alpha}_j - \hat{\beta}_j \hat{X}_{i0})}{\cos(x_i - \hat{X}_{i0}) + \sum_j \hat{\beta}_j^2 \cos(y_{ji} - \hat{\alpha}_j - \hat{\beta}_j \hat{X}_{i0})},$$

where $\hat{X}_{i1}$ is an improvement of $\hat{X}_{i0}$.

3.4 MLE for κ

The first partial derivative of the log likelihood function with respect to κ is,

$$\frac{\partial \log L}{\partial \kappa} = -n(1+q)\frac{I_0'(\kappa)}{I_0(\kappa)} + \sum_i \cos(x_i - X_i) + \sum_j \sum_i \cos(y_{ji} - \alpha_j - \beta_j X_i) .$$

By setting $\frac{\partial \log L}{\partial \kappa} = 0$ and after simplification, we get the equation approximately given by,

$$A(\hat{\kappa}) = \frac{1}{n(1+q)} \left\{ \sum_i \cos(x_i - \hat{X}_i) + \sum_j \sum_i \cos(y_{ji} - \hat{\alpha}_j - \hat{\beta}_j \hat{X}_i) \right\}$$

where $A(\hat{\kappa}) = \frac{I_0'(\hat{\kappa})}{I_0(\hat{\kappa})} = \frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})}$.

Then, the equation becomes,

$$\hat{\kappa} = A^{-1}\left[\frac{1}{n(1+q)} \left\{ \sum_i \cos(x_i - \hat{X}_i) + \sum_j \sum_i \cos(y_{ji} - \hat{\alpha}_j - \hat{\beta}_j \hat{X}_i) \right\} \right].$$

Estimation of $\hat{\kappa}$ can be obtained by using the approximation given by Fisher (1993),

$$A^{-1}(w) = \begin{cases} 2w + w^3 + \frac{5}{6}w^5 & w < 0.53 \\ -0.4 + 1.39w + \frac{0.43}{(1-w)} & 0.53 \le w < 0.85 \\ \frac{1}{w^3 - 4w^2 + 3w} & w \ge 0.85 \end{cases}$$

Hence,

$$\hat{\kappa} = A^{-1}(w) \text{ where } w = \frac{1}{n(1+q)} \left\{ \sum_i \cos(x_i - \hat{X}_i) + \sum_j \sum_i \cos(y_{ji} - \hat{\alpha}_j - \hat{\beta}_j \hat{X}_i) \right\}.$$

In this paper, we propose an improvement of the estimate of $\hat{\kappa}$ by multiplying with a correction factor of $\frac{q}{(q+1)}$, where q is the number of equations in the simultaneous relationship. Therefore the estimate of κ is given by $\tilde{\kappa} = \left(\frac{q}{q+1}\right)\hat{\kappa}$.

4. FISHER INFORMATION MATRIX OF PARAMETERS

We will consider the Fisher information matrix of $\hat{\kappa}$, $\hat{\alpha}_j$ and $\hat{\beta}_j$ for the simultaneous circular functional relationship model assuming that $\lambda_j = 1$. By considering the first partial derivatives of the log likelihood function as well as the second derivatives and their negative expected values, we define the Fisher information matrix, F for $\hat{\kappa}, \hat{\alpha}_j$ and $\hat{\beta}_j$, given by

$$F = \begin{bmatrix} R & 0 & W \\ 0 & S & 0 \\ W^T & 0 & U \end{bmatrix}$$

where R is an $n \times n$ matrix given by

$$R = \begin{bmatrix} \hat{\kappa}A(\hat{\kappa}) + \hat{\kappa}\sum_j \hat{\beta}_j^2 A(\hat{\kappa}) & & 0 \\ & \ddots & \\ 0 & & \hat{\kappa}A(\hat{\kappa}) + \hat{\kappa}\sum_j \hat{\beta}_j^2 A(\hat{\kappa}) \end{bmatrix}.$$

W is an $n \times 2$ matrix given by

$$W = \begin{bmatrix} \hat{\kappa}\hat{\beta}_j A(\hat{\kappa}) & \hat{\kappa}\hat{\beta}_j \hat{X}_1 A(\hat{\kappa}) \\ \vdots & \vdots \\ \hat{\kappa}\hat{\beta}_j A(\hat{\kappa}) & \hat{\kappa}\hat{\beta}_j \hat{X}_n A(\hat{\kappa}) \end{bmatrix}.$$

S is given by $S = n(1+q)A'(\hat{\kappa})$ where $A'(\kappa) = 1 - A^2(\kappa) - \frac{A(\kappa)}{\kappa}$ and

U is 2×2 matrix given by

$$U = \begin{bmatrix} n\hat{\kappa}A(\hat{\kappa}) & \hat{\kappa}A(\hat{\kappa})\sum_i \hat{X}_i \\ \hat{\kappa}A(\hat{\kappa})\sum_i \hat{X}_i & \hat{\kappa}A(\hat{\kappa})\sum_i \hat{X}_i^2 \end{bmatrix}.$$

The bottom right minor of the inverse of F with order 3×3 will form the asymptotic covariance matrix of $\hat{\kappa}$, $\hat{\alpha}_j$ and $\hat{\beta}_j$. From the theory of partitioned matrices, this is given by

$$\hat{Var}\begin{bmatrix} \hat{\kappa} \\ \hat{\alpha}_j \\ \hat{\beta}_j \end{bmatrix} = \begin{bmatrix} S^{-1} & 0 \\ 0 & \left(U - W^T R^{-1} W\right)^{-1} \end{bmatrix}.$$

In particular we have the following results,

$$V\hat{a}r(\hat{\kappa}) = \frac{\hat{\kappa}}{n(1+q)\left[\hat{\kappa} - \hat{\kappa}A^2(\hat{\kappa}) - A(\hat{\kappa})\right]}$$

$$V\hat{a}r(\hat{\alpha}_j) = \frac{\left(1 + \sum_j \hat{\beta}_j^2\right)\sum_i \hat{X}_i^2}{\hat{\kappa}A(\hat{\kappa})\left(1 + \sum_j \hat{\beta}_j^2 - \hat{\beta}_j^2\right)\left\{n\sum_i \hat{X}_i^2 - \left(\sum_i \hat{X}_i\right)^2\right\}}$$

$$V\hat{a}r(\hat{\beta}_j) = \frac{\left(1 + \sum_j \hat{\beta}_j^2\right)n}{\hat{\kappa}A(\hat{\kappa})\left(1 + \sum_j \hat{\beta}_j^2 - \hat{\beta}_j^2\right)\left\{n\sum_i \hat{X}_i^2 - \left(\sum_i \hat{X}_i\right)^2\right\}}$$

5. GOODNESS-OF-FIT

In the model checking procedures, two approaches are used namely residual plots and numerical measures of Watson's U^2 test and Kuiper's V tests as available in ORIANA software for directional data. The graphical method of model assessment of von Mises quantiles of residuals is used. In addition, the numerical measure Watson's U^2 and Kuiper's V are used to test the null hypothesis that the errors followed the von Mises distribution as described by Fisher (1993).

6. SIMULATION STUDIES

The simulation studies have been carried out in order to assess the accuracy and measure the biasness of the proposed model. Sample sizes of n were generated and let s be the number of simulations with $q = 2$. In this model, α_1 and α_2 are circular variables while β_1, β_2 and κ are continuous. The following computations were carried out in the simulation study.

(a) Calculation for $\boldsymbol{\alpha_j}$ where $\boldsymbol{j = 1, 2}$

i) Circular Mean,

$C = \sum \cos(\hat{\alpha}_j)$, $\quad S = \sum \sin(\hat{\alpha}_j)$

$$\bar{\hat{\alpha}} = \begin{cases} \tan^{-1}\left(\frac{S}{C}\right) & S > 0, C > 0 \\ \tan^{-1}\left(\frac{S}{C}\right) + \pi & C < 0 \\ \tan^{-1}\left(\frac{S}{C}\right) + 2\pi & S < 0, C > 0 \end{cases}$$

ii) Circular Distance, $d = \pi - \left|\pi - \left|\bar{\hat{\alpha}} - \alpha\right|\right|$

iii) Mean resultant length, $R = \frac{1}{s}\sqrt{\left(\sum \cos(\hat{\alpha}_j)\right)^2 + \left(\sum \sin(\hat{\alpha}_j)\right)^2}$

(b) Calculation for $\boldsymbol{\beta_j}$ and $\boldsymbol{\kappa}$ where $\boldsymbol{j = 1, 2}$

i) Mean, $\bar{\hat{w}} = \frac{1}{s}\sum \hat{w}_j$

ii) Estimated Bias, $EB = \bar{\hat{w}} - w$

iii) Estimated Root Mean Square Errors, $ERMSE = \sqrt{\frac{1}{s}\sum\left(\hat{w}_j - w\right)^2}$

The simulation results with $s = 5000$ for each set of parameter value are shown in Tables 1 to 5. The values of X have been chosen from $VM\left(\frac{\pi}{4}, 3\right)$ and without loss of generality we choose $\alpha_j = 0$ and $\beta_j = 1$, where $j = 1, 2$. Three different choices of concentration parameters $\kappa = 15, .30$ and 50 for circular random error assuming $\kappa = \nu$ and for each of these four choices of sample size $n = 30, 70, 100$ and 500 have been considered. The values of κ cover a more realistic range as we expect the random error of circular variable is not too concentrated (less dispersed). It appears from Tables 1 and 2 that $\hat{\alpha}$ is a good estimator of α. Its circular distance, *d* which represents the biasness generally decreases with the increase of sample size *n*. Similarly, the measure of bias decrease with increasing concentration parameters of circular random error. Similar trend can be observed from the mean resultant length, *R* with represents a good accuracy as value approaches to one. Likewise, the conclusion may also be drawn from Tables 3 and 4 as the mean of $\hat{\beta}$ approaches to the true value of β with the increases of sample size *n* and the concentration parameters, κ of circular random error. The similar trend is also observed from the estimate bias (EB) and root mean square error (ERMSE) measures which suggest that $\hat{\beta}$ is a good estimator of β.

Table 1: Simulation Results for $\hat{\boldsymbol{\alpha}}_{\mathbf{1}}$ (True value of $\boldsymbol{\alpha_j} = \mathbf{0.0}$ and $\boldsymbol{\beta_j} = \mathbf{1.0}$, where $\boldsymbol{j} = \mathbf{1, 2}$).

Performance Indicator	Sample size, *n*	$\kappa = 15$	$\kappa = 30$	$\kappa = 50$
Circular Mean	30	-0.0139	-0.0068	-0.0042
	70	-0.0076	-0.0036	-0.0026
	100	-0.0060	-0.0026	-0.0016
	500	-0.0021	-0.0012	-0.0007
Circular Distance, *d*	30	0.0139	0.0068	0.0042
	70	0.0076	0.0035	0.0026
	100	0.0060	0.0026	0.0016
	500	0.0021	0.0012	0.0007
Mean Resultant Length, *R*	30	0.9956	0.9979	0.9987
	70	0.9983	0.9992	0.9995
	100	0.9988	0.9994	0.9996
	500	0.9998	0.9999	0.9999

Table 2: Simulation Results for $\hat{\boldsymbol{\alpha}}_{\mathbf{2}}$ (True value of $\boldsymbol{\alpha_j} = \mathbf{0.0}$ and $\boldsymbol{\beta_j} = \mathbf{1.0}$, where $\boldsymbol{j} = \mathbf{1, 2}$).

Performance Indicator	Sample size, *n*	$\kappa = 15$	$\kappa = 30$	$\kappa = 50$
Circular Mean	30	-0.0127	-0.0067	-0.0057
	70	-0.0079	-0.0030	-0.0026
	100	-0.0063	-0.0025	-0.0015
	500	-0.0019	-0.0012	-0.0007
Circular Distance, *d*	30	0.0127	0.0067	0.0056
	70	0.0079	0.0030	0.0026
	100	0.0063	0.0025	0.0014
	500	0.0019	0.0012	0.0007
Mean Resultant Length, *R*	30	0.9957	0.9978	0.9988
	70	0.9982	0.9992	0.9995
	100	0.9987	0.9994	0.9996
	500	0.9998	0.9999	0.9999

Table 3: Simulation Results for $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ (True value of $\boldsymbol{\alpha_j} = \mathbf{0.0}$ and $\boldsymbol{\beta_j} = \mathbf{1.0}$, where $\boldsymbol{j} = \mathbf{1, 2}$).

Performance Indicator	Sample size, *n*	$\kappa = 15$	$\kappa = 30$	$\kappa = 50$
Mean	30	1.0089	1.0048	1.0033
	70	1.0040	1.0028	1.0018
	100	1.0038	1.0018	1.0012
	500	1.0012	1.0008	1.0004
Estimate Bias (EB)	30	0.0089	0.0048	0.0033
	70	0.0040	0.0028	0.0018
	100	0.0038	0.0018	0.0012
	500	0.0012	0.0008	0.0004
Estimate Root Mean square Error (ERMSE)	30	0.0469	0.0320	0.0259
	70	0.0269	0.0190	0.0148
	100	0.0228	0.0159	0.0122
	500	0.0101	0.0071	0.0054

Table 4: Simulation Results for $\widehat{\boldsymbol{\beta}}_{\mathbf{2}}$ (True value of $\boldsymbol{\alpha_j} = \mathbf{0.0}$ and $\boldsymbol{\beta_j} = \mathbf{1.0}$, where $\boldsymbol{j} = \mathbf{1, 2}$).

Performance Indicator	Sample size, n	$\kappa = 15$	$\kappa = 30$	$\kappa = 50$
Mean	30	1.0080	1.0041	1.0032
	70	1.0044	1.0025	1.0015
	100	1.0038	1.0017	1.0009
	500	1.0010	1.0009	1.0005
Estimate Bias (EB)	30	0.0080	0.0041	0.0032
	70	0.0044	0.0025	0.0015
	100	0.0038	0.0017	0.0009
	500	0.0010	0.0009	0.0005
Estimate Root Mean square Error (ERMSE)	30	0.0462	0.0343	0.0265
	70	0.0274	0.0193	0.0149
	100	0.0226	0.0159	0.0122
	500	0.0100	0.0071	0.0054

Table 5: Simulation Results for $\hat{\boldsymbol{\kappa}}$ (True value of $\boldsymbol{\alpha_j} = \mathbf{0.0}$ and $\boldsymbol{\beta_j} = \mathbf{1.0}$, where $\boldsymbol{j} = \mathbf{1, 2}$).

Performance Indicator	Sample size, n	$\kappa = 15$	$\kappa = 30$	$\kappa = 50$
Mean	30	16.3839	32.9824	55.1073
	70	15.4297	31.1373	52.0184
	100	15.2777	30.7237	51.4161
	500	14.8989	29.9879	50.1485
Estimate Bias (EB)	30	1.3839	2.9824	5.1073
	70	0.4297	1.1373	2.0184
	100	0.2777	0.7237	1.4161
	500	0.1011	0.0121	0.1485
Estimate Root Mean square Error (ERMSE)	30	3.4950	7.1875	11.8474
	70	1.9190	3.9173	6.6699
	100	1.5696	3.1689	5.3637
	500	0.6595	1.3192	2.2420

Table 5 clearly indicates that multiplication by the correction factor of 2/3 or $q/(q + 1)$, where $q = 2$ to $\hat{\kappa}$ provide an improvement to the maximum likelihood estimator. Please note that in the linear functional relationship model for circular variables (where $q = 1$), an improvement to $\hat{\kappa}$ by multiplying $\hat{\kappa}$ with 1/2 or $q/(q + 1)$, where $q = 1$ was done to get the consistent estimation of κ. In this paper, we extended the idea by assuming that multiplication by $q/(q + 1)$, may improve the maximum likelihood estimation for κ in the simultaneous linear functional model for circular variables with q equations. As shown in the simulation studies, the mean of $\tilde{\kappa}$ approaches the true value of κ with the increases of sample size n and the concentration parameters of circular random error. Thus, it is plausible that the larger κ (less dispersed), the better is the approximation of κ using $\tilde{\kappa}$ instead of $\hat{\kappa}$. In other words, a more appropriate estimate of κ in this simulation studies is using a correction factor, $\tilde{\kappa} = \frac{2}{3}\hat{\kappa}$.

7. APPLICATION TO METEOROLOGICAL DATA

As an illustration for the proposed model, the wind direction data recorded at Bayan Lepas airport on July and August 2005 at time 1200, located at 16.3 m above ground level, latitude 05°18'N and longitude 100°16'E were used in this study. A total of 62 observations have been recorded at three different pressures (and height) which is at pressure 850 Hpa with 5000 m height (denoted by variable *x*), at pressure 1000 Hpa with 300 m height (denoted by variable y_1) and at pressure 500 Hpa with 19,000 m height (denoted by variable y_2).

Figure 1 show the scatter plots of the wind direction data set where we can see there are a few points located at upper and lower bound of *y*-axis. Knowing that our data are in a circular form, then the observations at the top left should be closed to the bottom left or upper right. Similarly apply to those points at the bottom right. Thus, it suggests that there exists a linear relationship between those variables involved in the measurement of this wind direction.

(a)

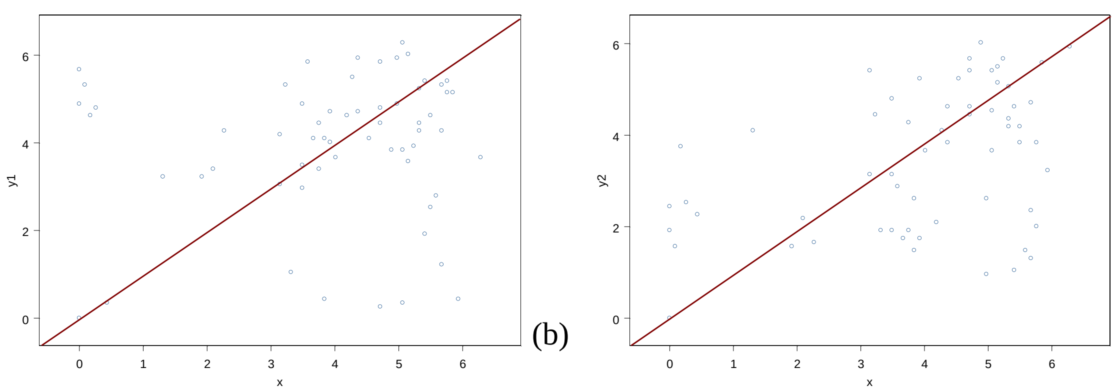


Figure 1: Scatter plot for wind direction recorded at Bayan Lepas airport

(a) X & Y_1 (b) X & Y_2

In this example the maximum likelihood estimates have been obtained by assuming equal error of the concentration parameters. For the convergence criteria, we set that the iteration will stop once our β_j change by no more than 0.0001. The iteration is started by selecting $\hat{X}_{i0} = x_i$ and $\hat{\beta}_0 = 1$ as initial values. These values are chosen because we expect that $\hat{X}_i$ should be close to $x_i, i = 1, \ldots, n$ and $\hat{\beta}_j$ around 1. Our aim is to find α_1, α_2, β_1, β_2 and κ which will maximize the log likelihood function as well as getting the standard error for each parameter estimates by using the Fisher information as described previously.

Table 6 showed the maximum likelihood estimates and the standard error for the data set. The estimate relationship of wind direction data recorded at Bayan Lepas airport may be given by

$$Y_1 = -0.211 + 1.034X \pmod{2\pi} \quad \text{and} \quad Y_2 = -0.174 + 0.912X \pmod{2\pi}$$

Table 6: Parameter estimation for wind direction data recorded at Bayan Lepas airport.

Parameter	Estimated values and its standard error
α_1	-0.211 (2.897×10^{-1})
α_2	-0.174 (2.385×10^{-1})
β_1	1.034 (8.7675×10^{-3})
β_1	0.912 (7.217×10^{-3})
κ	1.026 (1.080×10^{-1})

As for the model checking or goodness-of-fit for our model, we use graphical method which is done by plotting the von Mises quantiles plot for the residuals of data set. From Figure 2, each of the distribution of the residual of the data set are scattered along the straight line. This suggests that each residual are independently distributed and followed the von Mises distribution. Hence, we can say that the wind direction data set was fitted well with the proposed model.

Alternatively, we may also use the Watson U^2 and Kuiper's V tests that are available in the ORIANA software to calculate the value of U^2 and V respectively. The purpose of this tests is in testing the goodness-of-fit against a specified distribution, either uniform or von Mises. In this case, we set our null hypothesis that the data fit the von Mises distribution. Table 6 shows the value obtained from ORIANA software and we can see that the value of U^2 and V respectively for each δ, ε_1 and ε_2 are small and these results suggests that the samples are drawn from the von Mises distribution.

Table 6: Values of Watson U^2 and Kuiper's V test for residuals.

Test		Values
Watson U^2	δ	0.052 ($0.5 > p > 0.25$)
	ε_1	0.033($p > 0.5$)
	ε_2	0.049 ($0.25 > p > 0.15$)
Kuiper's V	δ	1.157($p > 0.15$)
	ε_1	0.926($p > 0.15$)
	ε_2	1.048($p > 0.15$)

8. CONCLUSION

This study proposed the simultaneous linear functional relationship model for circular variables which is an extension of the linear functional relationship model. This model is significant because it allows us to look at the relationship between more than two circular variables simultaneously when only unreplicated data available and all variables are measured inexactly or subject to errors. Using maximum likelihood estimation of the parameter, an improvement of the estimation of the concentration parameter, κ is proposed. The above results suggested that the proposed model can be used in assessing the relative calibrations for any given circular data set in which the measurements are subject to errors. Further, this model can also be applied in other studies which involve the circular variables in which the main interests is to look at the underlying relationship between different circular variables rather than to predict one variable from the other and the variables cannot be recorded exactly.

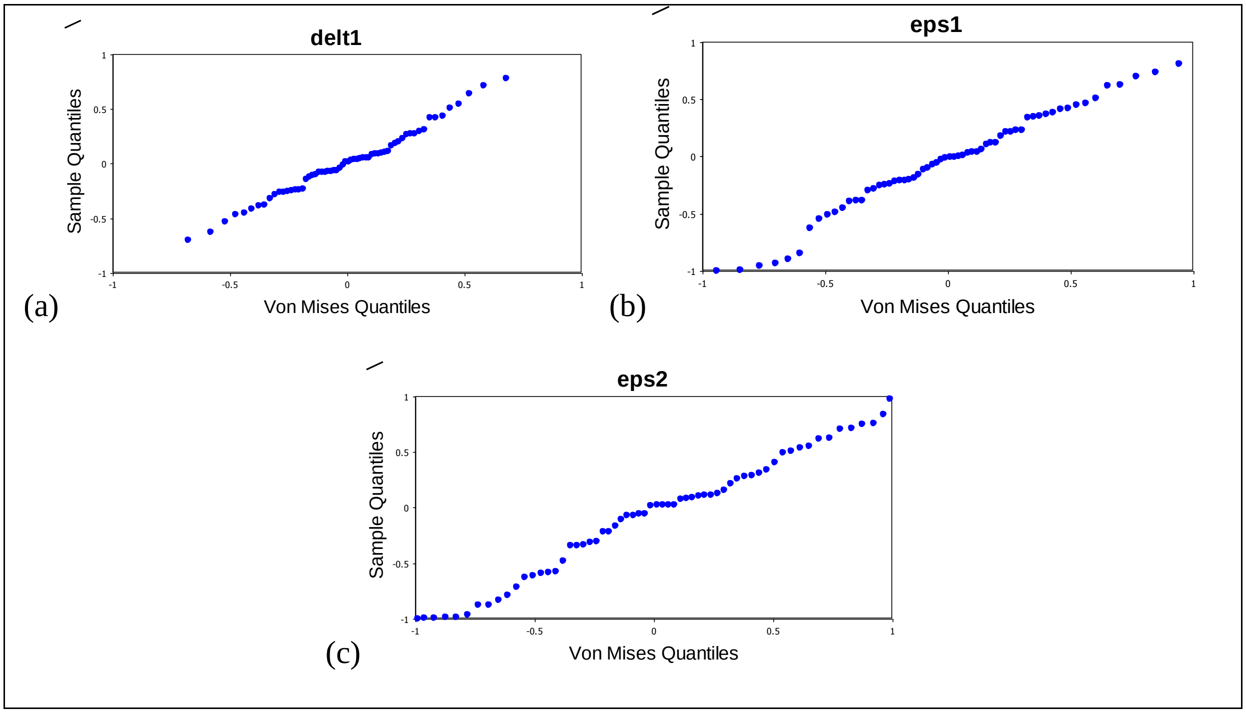


Figure 2: The von Mises quantiles plot of residuals (a) X (b) Y_1 (c) Y_2

REFERENCES

Caires S. and Wyatt L. R. (2003). A linear functional relationship model for circular data with an application to the assessment of ocean wave measurements. *Journal of Agricultural, Biological, and Environmental Statistics,* Vol. 8(2), 153-169.

Fisher N. I. (1993), *Statistical Analysis of Circular Data*. Cambridge University Press.

Hussin, A. G. (1997). *Pseudo-replication in functional relationship with environmental application*, PhD Thesis, University of Sheffield, England.

Hussin A. G., Fieller, N. R. J. and Stillman E. C. (2004). Linear regression for circular variables with application to directional data, *Journal of Applied Science and Technology* (JAST), Vol. 8, Nos. 1 and 2.

Mardia K. V. (1972). *Statistics of Directional Data*. London: Academic Press.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 317–322.

SESSION 9: STATISTICS EDUCATION

Chair: Shahjahan Khan,
Department of Mathematics and Computing
University of Southern Queensland
Toowoomba, Qld 4350, AUSTRALIA
E-mail: khans@usq.edu.au

RECENT ADVANCES IN STATISTICAL EDUCATION

Munir Ahmad and Suleman Aziz Lodhi
National College of Business Administration & Economics, Lahore, Pakistan.
E-mail: drmunir@ncbae.edu.pk

One who takes lessons from the events of life, gets vision, one who acquires vision becomes wise and one who attains wisdom achieves knowledge. Hazrat Ali (R.A).

ABSTRACT

In recent years there has been a tremendous development in statistics and statistical education. The statistical science has become a basic tool for all types of data analysis. It has the capability of transferring raw data to information; statistical tools have transferred information to knowledge, which is the core of decision making and planning. The statistical methodologies are getting popular for conducting cross-discipline research. In this paper, we have identified some recent advances in statistics that contribute towards data management, neural network, social networking, artificial intelligence and cybernetics which are emerging approaches and are useful in decision making and planning process.

1. INTRODUCTION

Statistical sciences have enjoyed a unique position in the domain of knowledge. It has been a subject of special interest since its start in or around the 18th century. The German Statistician, first introduced by Gottfried Achenwall (1749), referred to in Wikipedia (2009) was originally chosen for the analysis of data relating the state, and therefore it was also called the "science of state". The English called it political arithmetic. Thus, the original purpose of Statistic was data management to be used by governmental and administrative bodies (Yale and Kendall, 1964).

Due to its use in government administration and its data-centric perspectives, statistics is not considered to be a subfield of mathematics but rather a distinct field that uses mathematics. Its mathematical linkages were laid in the 17th and 18th centuries with the emergence of probability theory and since then new techniques of probability and statistics are in continual development.

Statistics provided the motivation to administrators to collect data on their people and economies, and it was later in the early 19th century that its use was broadened to include the collection and analysis of data in general. Developments in computer sciences have made possible very-large-scale statistical computation, thus making possible many new methods that would be impractical to perform manually. Today statistics is widely employed in government, business, and the natural and social sciences. The 20th century has witnessed the use of statistics for making critical decisions in agricultural research, public health concerns (epidemiology, biostatistics, etc.), industrial quality control, and economic and applications in social policy making purposes (unemployment rate, econometry, etc.). The use of statistics in today's complex environment has broadened far beyond its origins. Individuals and organizations use statistics to understand data and make informed decisions throughout the natural and social sciences, medicine, business, and other areas. (See also Smith and Reid, 2009, Davis 2007, Columbia Encyclopedia 2008).

2. ISLAMIC EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (ISESCO) ROLE IN STATISTICS EDUCATION

In the past, educationists world-wide prepare curricula to meet the needs of nations. The underlying concepts of the education system are to teach by preparing history and culture. Islamic history in education and research is very rich. The current generation of members of OIC states has to be motivated to embark on new areas of research. One has to take account of ground realities in development of fundamental courses. It is also necessary to seek information from stakeholders for planning, execution and monitoring of installing systems in education. Statistics was introduced in high schools curricula quite early and later it was incorporated in school courses as a part of mathematics. In Schools it is still a part of mathematics. The under-graduate level, a committee was constituted by ISESCO in 1980s at King Fahd University of Petroleum and Minerals (KFUPM) Dharan, Saudi Arabia for preparing statistics courses for six years program from class VII to Class XII. The curriculum was designed with latest statistical methods that could play an important role in economic planning for OIC members' states. Later, four year degree program was devised for OIC members' states. In some of OIC countries, the syllabi were adopted. Copies of the syllabi are available with ISESCO, Rabat, Morocco. Since then, world has seen a tremendous development in statistical methodologies. It is now essential that the new emerging statistical techniques are incorporated in the syllabi.

3. EMERGING STATISTICAL TECHNIQUES IN MULTIVARIATE DATA ANALYSIS

The traditional model of education divides the system into three layers. These layers are represented by a pyramid, in which the pre-high schools form the base of the pyramid. Next layer which comes over it is the high school layer and at the top of the pyramid is the graduate school. The pyramid shape shows that traditionally the pre-school students and teachers are highest in numbers, followed by a relatively lesser number of high school students and teachers and only a few students and teachers are at the top. The modern paradigm has increased the number of high school students to make the pyramid more peaked and more responsive; but still, with the universal free education at pre-school level, the pyramid is relatively thinner at the base.

The information requirement of each category of education is different, and depends on the responsibility that each category performs in an institution. The pre-school system is more concerned with operational decisions. The statistical techniques used by teachers at this level are mostly concerned with basic descriptive statistics like means, dispersion and charts for data collection from a field. In the higher school category, instructors/ students are responsible for the data collection. They are also responsible for summarizing the data and presenting the information in an understandable form. Therefore the statistical techniques used by high school are focused on summarizing of data, its presentation and preliminary analysis. The graduate school is responsible for use of advanced statistical techniques for generation of new knowledge. They decide on the future directions a school would move in the coming years. The graduate school faculty needs to have all possible information regarding the latest statistical technologies. They need to know possibilities that the school has to face in future.

Fig. 1 presents the Statistical Techniques used by different categories of schools. It is seen that as we move from the base of the pyramid to its top, the use of statistical techniques is shifted from data orientation at the base, to knowledge extractions and new knowledge generation at the top level.

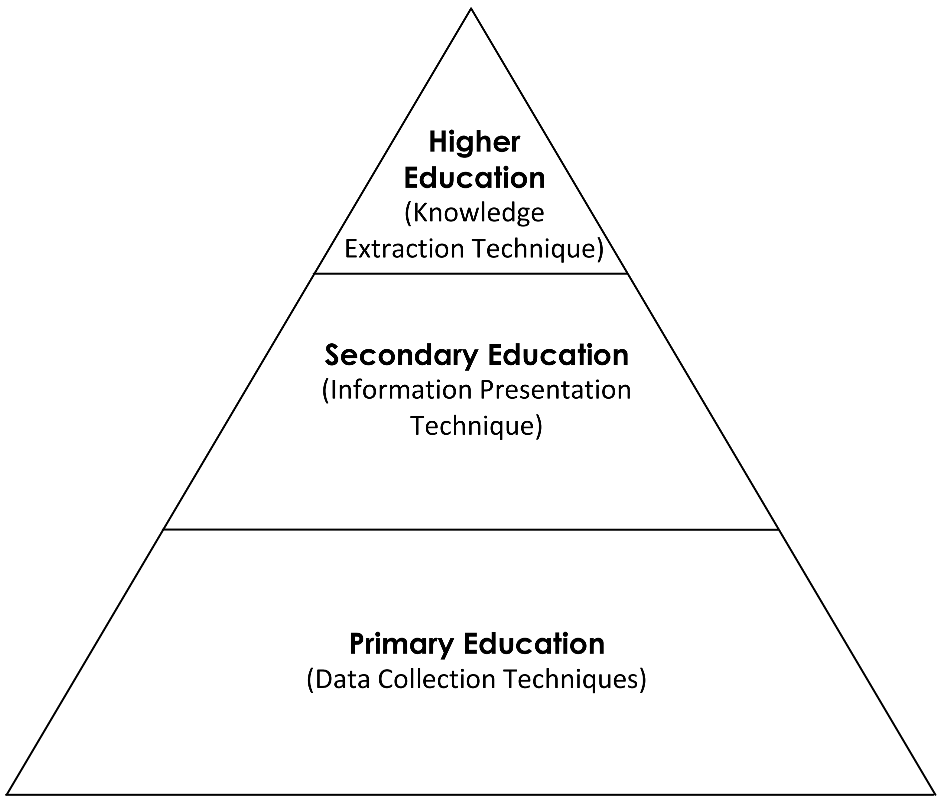


Fig. 1: Statistical Techniques used at different levels of Education

The emerging statistical techniques in multivariate data analysis are i) data ware- housing (ii) data mining (iii) neural network (iv) resampling and (v) social network methods. The statistical data analysis involving four steps viz. (i) identifying data, (ii) data collection (iii) analyzing data and (iv) reporting the results. Most of the tools of the method involve use of mathematical functions to describe structures and their sub-structures.

Further to these new areas of data analysis we have (i) canonical correlation analysis (ii) cluster analysis (iii) correspondence analysis, (v) meta analysis (vi) content analysis (vii) panel analysis (viii) conjoint analysis (ix) structural equation modeling (x) multidimensional scaling and (xi) correspondence analysis. Statisticians have not been concerned with these methods as these are widely used by management and medical scientists. In categorical data analysis, statisticians have now tried to update these areas and devise new methods.

4. RECENT DOMAINS IN STATISTICAL THEORY

Inter-disciplinary relations of various sciences have been developed recently. Before we state the enumeration of latest new areas in the discipline, Zacharias et al. (2008) has provided interrelation of various disciplines in Fig. (2).

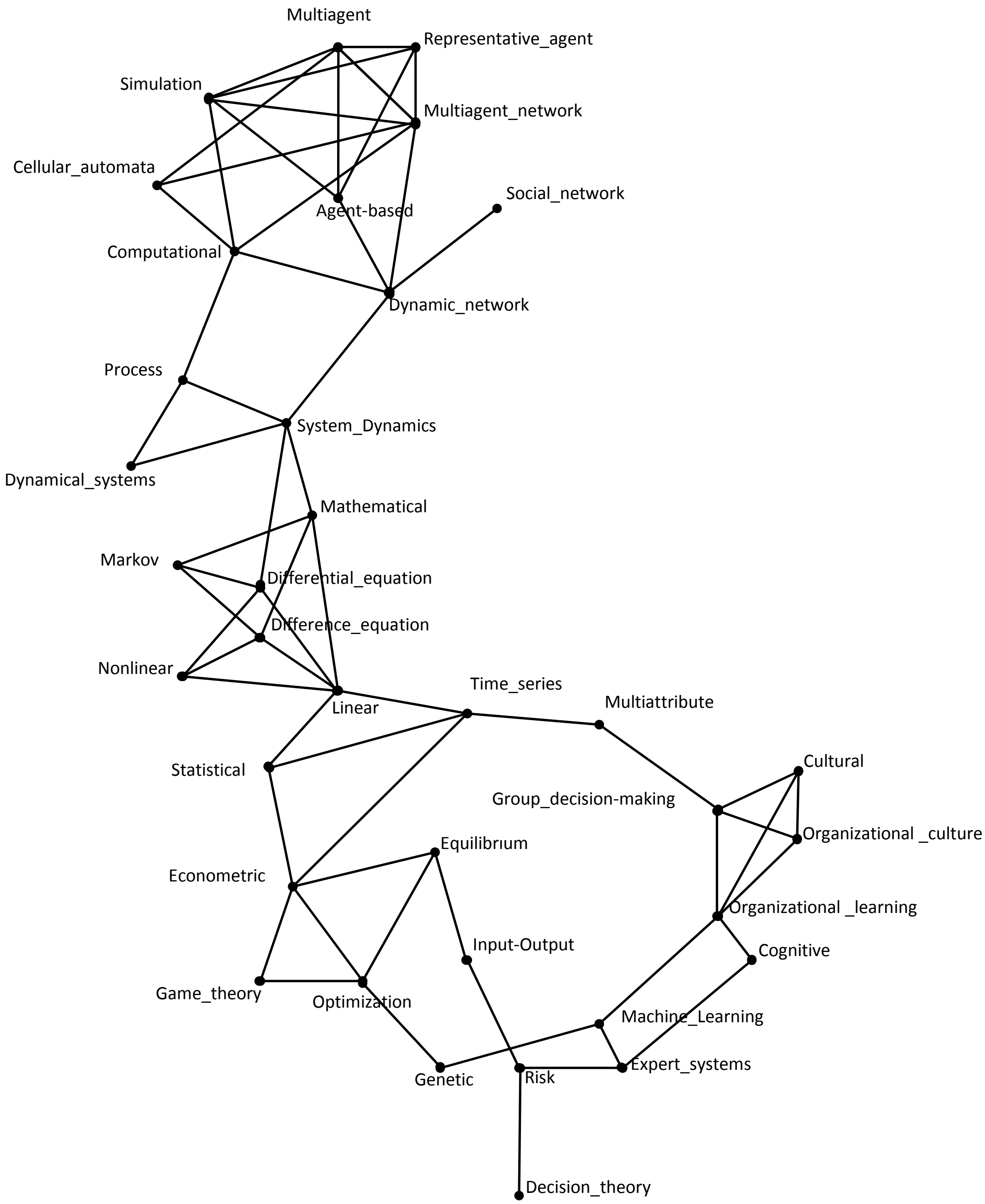


Fig. (2): Relationships among regular research approaches
[adopted from Zacharias et al. (2008)]

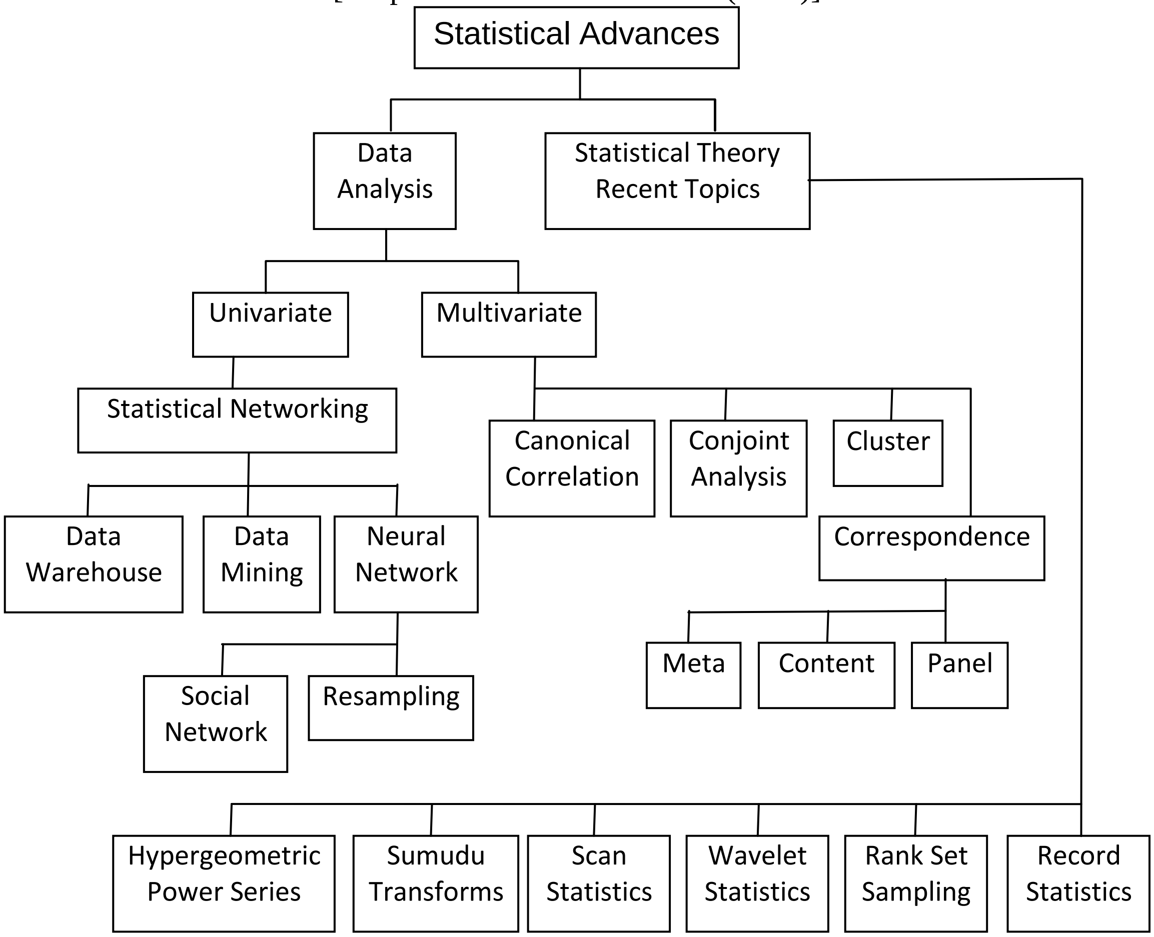


Fig. 3: Division of Statistical Theory with recent developments

Some of the recent advances in statistical theory and practice are

1. Record values and Record Statistics
2. Rank Set Sampling
3. Theory of Inversion
4. Wavelets and its Applications to Statistics
5. Scan Statistics
6. Sudoku Design of Experiments
7. Statistical Neural Network
8. Hypergeometric Power Series Functions

The statisticians from Muslim countries have a leading role in contribution to Record Statistics (Ahsanullah 1976, 1978, 2005), Rank Set Sampling (Muttlak 2001, 2003; Muttlak and Al-Sabah 2003; Smawi 1996; Abu-Dayyeh 1996, 2002; Al-Saleh, 1998 and Rahimov and Muttlak 2001), Theory of Inversion (Ahmad 2007; Habibullah and Ahmad 2006) and Hypergeometric Power Series Function (Ahmad 2007). We have developed a flow chart of relationships with respect to statistics, with inclusion of further sub-disciplines of statistics and are shown in Fig. (3). Literature on each topic available for further research. At the National College of Business Administration and Economics (NCBA&E), Lahore we have started incorporating these new emerging disciplines in to our M.Phil./Ph.D. (Statistics and Management) programs. Later, Board of Studies in various disciplines, are considering introducing these courses at M.Sc. and B.Sc. level. Some universities around the world have already introduced these areas of research to their Ph.D. program.

REFERENCES

Abu-Dayyeh, W. and Muttlak, H. A. (1996). Using ranked set sampling for hypothesis tests on the scale parameter of the exponential and uniform distributions. *Pak. J. Stat*., 12, 131-138.

Abu-Dayyeh, W. and Muttlak, H. A. (2002). Variance estimation for the location-scale family distributions using ranked set sampling. *Pak. J. Statist*., 18(1), 89-100.

Ahmad, M. (2007). On the theory of inversion. *Int. J. Statist. Sciences*, 6, 43-53.

Ahsanullah, M. (1976). On a characterization of the exponential distribution by order statistics, *J. Appl. Probab.*, 13, 818-822.

Ahsahullah, M. (1978). A characterization of the exponential distribution by spacings. *J. Appl. Probab*., 15, 650-653.

Ahsanullah, M. (2005). On lower generalized order statistics and a characterization of power function distribution. *Statistical Methods*, 7(1), 16-28.

Al-Saleh, M. F. and Muttlak, H. A. (1998). A note on Bayesian estimation using ranked set sample. *Pak. J. Statist*., 14(1), 49-56.

Columbia Encyclopedia (2008). Sixth Edition, Columbia, USA.

Davis, N. (2007). Developments in promoting the improvement of statistical education. *Int. Statist. Inst.*, 56th Session.

Habibullah, Saleha N. and Ahmad, M. (2006). On a new class of univariate continuous distributions that are closed under inversion. *Pak. J. Statist. Oper. Res.*, II(2), 151-158.

Muttlak, H. A. (2001). Extreme ranked set sampling: a comparison with regression and ranked set sampling estimators. *Pak. J. Statist*., 17(3), 187-204.

Muttlak, H. A. (2003). Modified Ranked Set Sampling Methods. *Pak. J. Statist*., 19(3), 315-323.

Muttlak, H. A. and Al-Sabah, W.S. (2003). Statistical quality control based on pair and selected ranked set sampling. *Pak. J. Statist*., 19(1), 107-128.

Rahimov, I. and Muttlak, H. A. (2001). Random ranked set sampling. *Pak. J. Statist*., 17(1), 51-66.

Smawi, Hani M. (1996). Stratified Ranked Set Sample. *Pak. J. Statist*., 12(1), 9-16.

Smith, N. and Reid, A. (2009). Representations of internationalization in statistics education. *J. Statist. Education*, 17(1), 1-17.

Wikipedia, Wikipedia Foundation Inc. available at

http://en.wikipedia.org/wiki/History_of_statistics (site accessed on 09.09.2009).

Yale, G. U and Kendall, M. G. (1964). *An introduction to theory of statistics*. 14th Ed. And 4th Impression. Charles Griffin & Co. Ltd, London. UK.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 323–332.

TEACHING PRINCIPLES OF STATISTICS: AN ISSUE-BASED APPROACH

Ismail, Mohamed A.
Statistics Professor,
Cairo University
and Consultant at Egyptian
Cabinet's Information and
Decision Support Center

Abdelmageed, Samar M. M.
Statistical Researcher, Egyptian
Cabinet's Information and Decision
Support Center

ABSTRACT

This paper presents an issue-based approach for teaching the principles of statistics. The proposed issue-based approach relies on selecting one central issue that is subsequently divided into a group of sub-issues or sub-questions through which the syllabus contents are presented. This teaching method replaces the traditional one in which the learning materials are given in a disconnected way. To apply the proposed issue-based approach, a group of stages is required. These stages include preparation, developing skills, leading discussion, and evaluation. The paper designs a syllabus for teaching the contents of the principles of statistics course using subsidies as the main issue along with its sub-issues. The paper also suggests a timetable for the proposed syllabus.

Keywords: Issue-Based Approach, principles of statistics, subsidies issue

1. INTRODUCTION

Different countries in the world now increasingly pay great attention to human resources development due to its crucial role in achieving growth and progress. Education is considered to be the main and most important player in building and developing the human resources of any country. One of the serious problems that confront the educational process and may hinder human resources development in Egypt is the reliance on the traditional approach for designing educational syllabuses and materials whether in social or applied sciences. Syllabuses usually take the form of a group of daily lessons that deal with a number of different disconnected facts about a certain subject or event or time period. This traditional approach of teaching suffers from several flaws that lead to a superficial disconnected education and limit the students' skills to theoretical concepts and discussions that leave them uninterested in taking a positive and effective participation role with the teacher in the learning process, for more details please refer to Garratt and Taylor (2004).

Boyle (1999) pointed out that authors of educational theories showed that knowledge can be recalled and retained for a longer time and applied in an easier way if the educational process is connected with the way the knowledge is to be used and applied. In line with this thinking, the Issue-Based Approach is considered to be one of the most important teaching approaches that have been focused on recently. This teaching approach relies on the existence of a central real-life issue through which the educational curriculum is taught. Garratt and Taylor (2004) emphasized that the issue-based approach helps students to achieve higher levels of learning and to perceive their skills in the practical application of theoretical concepts through a series of real-life issues related to the teaching subjects and curriculums.

The issue-based approach has been discussed in teaching different subjects. In the field of natural sciences, Poon (1998) dealt with the issue-based approach in occupational safety and health education. Moreover, Lewis (2002 and 2003) talked about using the issue-based approach in teaching environmental sciences. And, Curriculum Development Institute (2007) tackled the issue-based approach in teaching science and technology. In the field of social sciences, Onosko and Swenson (1996) discussed the issue-based approach in teaching social studies. Holbrook et al. (2003) indicated that using the issue-based approach in designing and teaching social sciences' curriculums supports students' creativity especially in what concerns their ability to put questions and predict results. And, Garratt and Taylor (2004) pointed out to the importance of applying the issue-based approach in teaching economics for the first-year undergraduate economics students by formulating the economic concepts in real-life frameworks, connecting economics with other disciplines, and applying economic ideas through a wide range of issues

The Issue-Based Approach sometimes appears under the name Case-Based Learning especially in the fields of business administration and law, or under the name Problem-Based Learning as in medicine and engineering schools; whereas the name Issue-Based Approach appears more in the social sciences that deal with debatable issues, see Curriculum Development Institute (2007). Ward and Lee (2002) mentioned that the problem-based learning approach was mainly developed in medicine schools. Moreover, some other studies discussed using the problem-based learning approach in teaching a number of different courses. In this regard, the reader can refer to Stinson and Milter (1996), Hadgraft (2000), and Mimbs (2005) for more details

The objective of this paper is to present a proposal for teaching the principles of statistics course using the issue-based approach. The proposed approach depends on raising the subsidies issue through which the syllabus or the curriculum contents are taught in an applicable way without delving too much in mathematical details. The paper also presents an allocation of the course contents on the sub-issues or the sub-questions of the main subsidies issue and a timetable for teaching the suggested curriculum

The paper consists of five sections. The second section explains the issue-based approach and its application stages. The third section discusses the advantages of the issue-based approach against the traditional approach of designing educational curriculums. The fourth section uses the issue-based approach to propose a syllabus for the principles of statistics course by applying on the subsidies issue and it also presents the timetable required for teaching the proposed syllabus. And finally, the fifth section draws some conclusions.

2. THE ISSUE-BASED APPROACH

The idea behind the issue-based approach in teaching comes from a main central issue that acts as a spine holding the daily lessons of the curriculum tightly together; see Onosko and Swenson (1996). The issue-based approach aims at engaging the students and motivating them to learn by applying on the reality they live in because when the students realize the importance of the scientific material they learn and their ability to use it in reality their interest and motivation to learn increase, see Garratt and Taylor (2004). Lewis (2003) denoted that the application of the issue-based approach consists of different stages, which include: (1) preparation, (2) developing skills and leading discussion, and (3) evaluation. An explanation of each one of these stages is given as follows:

2.1 Preparation

Onosko and Swenson (1996) indicated that the preparation stage of the issue-based approach has different steps, which include: choosing of a central issue to draw the students' attention and interest, putting an attractive introduction, connecting the detailed lessons with the main central issue, and preparing the required detailed educational materials. A detailed overview of each one of these steps is given as follows:

2.1.1 Choosing a central issue

An issue means any inquiry or question that has no unique output or answer, so that people are expected to have different opinions about it, see Lewis (2003). Issues can be classified using several different methods. For example, according to time; there are historical issues, current issues, and future issues, while according to specialization; there are economic issues, social issues, political issues, or even multi-disciplinary issues. A central issue guarantees the coherence of the curriculum and the interdependence of its contents. It also challenges the students mentally and motivates them to think as Poon (1998) pointed out to the importance of issues that act as a main motivation to learn. In this context, the teachers guide the inquiries of the students and help them to build their knowledge rather than just giving them some ready-made concepts. A central issue also helps the teachers to control directing the students' minds towards a big number of side issues that may be related to the main concerning issue and due to time limitations cannot be dealt with, the thing that leads to a complete dedication to the treatment of the central issue in a comprehensive way.

Onosko and Swenson (1996) determined the following criteria to choose a suitable central issue:

1. To be a controversial issue: This criterion may seem unnecessary due to the definition of an issue as a controversial question without a single answer, however, in some cases it may be unclear whether the chosen question is an issue or not? If there are conflicting points of view about how to answer this chosen question due to the nature of the issue itself and not due to the students' or the teachers' unfamiliarity and lack of knowledge about it, then the chosen question can be considered to be a good issue that is worth of discussion.
2. To be an important issue: To agree on what are the important issues and what are not is an extremely difficult if not impossible matter. The questions that can be used to determine the importance of the issue under consideration include: whether this issue has been discussed before in the past or not and if it has been discussed before is the debate still going on at the present time? Another important question is whether this issue is an issue of public opinion that needs some decision or judgment? It is also crucial to ask if the experts in the field of this issue agree on its importance or not? And finally if understanding this issue will help the students to improve their capabilities or not?
3. To be an interesting issue: If the number of important issues exceeds what can be studied, then it is necessary to choose the interesting issues to both the teachers and the students since the teachers' enthusiasm substantially affects the students' learning. The students' interest can

be aroused by putting an interesting attractive title to the issue instead of just a regular title or by making the issue easy to remember. For instance, it is better to put an issue title as: "What would you do for the poor if you were the president?" instead of the title: "The government's efforts to help the poor".

4. To be possible to investigate the issue efficiently: There are some important and interesting issues which cannot be dealt with due to the lack of written educational sources that suit the students. The availability of educational materials that reflect different points of view concerning the issue of interest, and not just the point of view of one or a few sides, contributes to increasing the students' interest and providing them with different facts and ideas which may help to achieve a good and successful analysis for the issue. These educational materials may include active reading or pictures or graphs. Before beginning to investigate the issue, it needs some effort from the teacher to be sure that there are enough materials to study it, since often books alone are not sufficient and some other additional educational tools and materials may be required.
5. To be presented in a way that suits the students: The students' different capabilities, backgrounds, and extent of knowledge about the selected issue should be considered in determining the appropriate style to present this issue.

2.1.2 Putting an attractive introduction

After selecting the central issue for investigation, an attractive introduction about it should be given using a movie scene, primary document, short story, slide show, some data, song, brainstorming session with the students, simulation model, political cartoon, writing activity, field trip, guest speaker,…etc. The importance of this introduction, even if it is a short one and regardless of its form, lies in its ability to stimulate the students' interest and reveal the teacher's enthusiasm about what is going to be taught.

The introduction includes recognizing some of the points of view or the potential positions that may be adopted concerning the selected central issue. The students themselves may seek to adopt an initial position or point of view that can change if new information shows up. The opening lessons of the curriculum content represent one of the critical factors for success as the failure to attract the students' attention and interest may lead to a weak marginal interaction from their side all along until the end of the curriculum content.

2.1.3 Connecting the detailed lessons with the central issue

The total benefit that results from connecting the detailed lessons with the main central issue exceeds the sum of benefits from each one of all the lessons or the components of the curriculum. Some of the methods cited in previous literature for connecting the detailed lessons with the central issue are described as follows:

1. Determining competing points of view: For instance, the students of psychology may raise the question "Why do we dream?", in this case, one or more lessons can be dedicated to present Freud's theory about how dreams express the unconsciousness, and after that, a number of lessons is allocated to explore other points of view like the one that sees dreams as symbolic expressions.

2. Determining the main concepts, events, people, and any other terms that may be needed by the students to enhance their understanding of the central issue.
3. Determining various sub-issues and sub-questions that need analysis in order to deal with the central issue efficiently.

There are some suggestions and advices about the method of teaching. Among the most important ones come: the need to present and always mention the central issue in the lecture room, the need to remind the students with the main objective at the beginning of every lesson, the importance of illustrating the relationship between the daily lessons and the main central issue and explaining at the end of each lesson how it contributes to a more understanding of the fundamental issue with a brief overview of the next lesson and its relationship with the central issue, and the need to clarify how homework and extra readings help in studying the central issue of interest.

2.1.4 Preparing the educational materials

In many cases, textbooks offer just little details and their contents rarely revolve around important issues. Moreover, the way of presenting the curriculums often lacks coherence or logical order within or among different chapters of the textbooks. Textbooks usually tend to build assumptions to provide certain conclusions with little logical or practical support, and due to that, they do not significantly help in enhancing the students' capabilities of critical thinking. A lot of concepts provided in these textbooks lack accurate definitions, and even if they were appropriately defined no illustrative examples are there to support them. Moreover, these textbooks also suffer from a superficial writing style free from controversy that provokes thinking. And, even when some issues are mentioned, the conflicting points of view are not discussed in a brief integrated way.

On the other hand, detailed educational materials play an important role in motivating the students to think and improving their experience and knowledge about the issue of interest, as they get familiar with different points of view concerning the issue and the logic behind each one of them. In addition to that, important facts and events are presented and explored, and concepts are dealt with expansively using agreeing and disagreeing examples. These educational materials range widely to include active reading, pictures, statistics, visual materials, and any other materials that help the students to understand and absorb the issue and stay interested with exploring the issue and building an opinion about it.

2.2 Developing skills and leading discussion

Other elements or requirements of the issue-based approach include developing skills, leading discussion, and evaluation. Developing skills include critical reading, the ability to differentiate between facts and points of view, the ability to differentiate between what is known and what is unknown, exploring the sources of evidences and evaluating these evidences, presenting available data in effective ways, understanding the scientific method, and identifying the points of weaknesses that can exist in scientific research design. The teacher should build the students' skills in these areas by putting a clear model for treating issues that can be followed by the students, and by gradually increasing the complexity of tackled issues as much as the students' skills increase.

It is necessary to focus on developing the students' skills in thinking more than just transferring knowledge to them, and also on enhancing their capabilities of dealing with different data and points of view. The teacher should also pay attention to the values and beliefs adopted by the students during the process of making their final decision about the issue under investigation. This shows the students' abilities to interpret evidences and how they construct their opinions based on the new gained information and knowledge, see Lewis (2003) and Curriculum Development Institute (2007) for more details.

Leading the discussion can be achieved by drawing a roadmap about how to find the necessary information for discussing the issue, making sure of achieving the desired objectives, and giving the students some freedom to deal with the issue in their own ways. Sometimes, students who are not very good in analyzing challenging data may focus more on the social and political sides of the issue. In this situation, the teacher has to guide the students in order to achieve the objectives of the curriculum and not just directly transfer knowledge to them. It is necessary for the teachers to monitor the learning process and shed more light on the main characteristics and steps of the scientific research in order to develop the students' skills and capabilities to deal with real-life issues and problems, see Lewis (2003) and Curriculum Development Institute (2007) for more details.

2.3 Evaluation

The evaluation methods should reflect the progress and improvement in the students' skills of critical thinking, creativity, effective and collective learning, presentation, and identification of new learning patterns. Methods for evaluation vary to include written exams, presentations, written articles, case studies of issues that have not been discussed, a panel discussion that involve all the students inside the lecture room, or presenting suitable articles from scientific periodicals.

These non-traditional methods work on encouraging and increasing the students' interaction with each other and develop their creativity skills. They also give the students the chance to express their opinions and not the opinion of the teacher or the authority in what concerns the central issue under discussion. In addition to that, the students realize through these methods that the final result is not just a written exam corrected by the teacher and given back to them but a chance for them to show their comprehension and mental cleverness compared with their peers. These activities are considered as provocative tools to the students who most often prefer to work with each other rather than working alone. This teamwork also permits the students to have a look at each other's ideas in a way that serves their understanding of the issue and to verify their thinking about it.

3. THE ADVANTAGES OF THE ISSUE-BASED APPROACH IN TEACHING COMPARED WITH THE TRADITIONAL APPROACH

The issue-based approach in teaching avoids the problems that accompany the traditional approach. These problems arise from the way in which the teacher covers all different sides and dimensions of the educational content in one curriculum without a central issue that connects these sides, which makes the students lose their ability to concentrate and link all the different parts together. The issue-based approach also avoids the contradiction that may exist between the contents of the traditional curriculums. To clarify more the difference between the issue-based approach and the traditional approach, it can be said that the latter may get interested in studying certain topics and put them within a wide framework and general main titles such as immigration, environmental pollution, or women's liberation, whereas, the issue-based approach tends to focus on certain specific issues within the context of these topics such as questions about who should be allowed to immigrate and why?, what should be done to deal with the environmental pollution?, or did the liberation of the Egyptian women during the last three decades positively or negatively affect the Egyptian society? Table (1) summarizes the advantages of the issue-based approach in teaching compared with the traditional approach.

Table (1): The advantages of the issue-based approach in teaching compared with the traditional approach

Traditional Approach	Issue-Based Approach
The topics are not linked with a certain issue.	The topics are linked with a certain issue.
The topics lack coherence and internal consistence.	The topics are characterized by coherence and internal consistence.
Dismantled, superficial, ineffective, and negative learning	Coherent, deep, effective, and positive learning
Individuality	Teamwork spirit
Dull non-motivating presentation methods	Attractive motivating presentation methods (Come alive)
Absence of critical analysis and life-long learning	Development of critical analysis and life-long learning
No familiarity with modern specialized methods	Getting familiar with modern specialized methods
Absence of dealing with practical and real-life problems such as data unavailability problems	Dealing with practical and real-life problems
Difficulty to deal with future issues	Dealing with future issues
The final benefit is a group of certain concepts	The final benefit is greater than the sum of that of the detailed lessons due to the multi-dimensional nature of the issue

4. A PROPOSED SYLLABUS FOR TEACHING THE PRINCIPLES OF STATISTICS COURSE USING THE SUBSIDIES ISSUE

The issue of subsidies is considered to be one of the current hot issues on the Egyptian scene. Many are calling for the need to reconsider the entire subsidies system in Egypt due to its defects, among which is the heavy burden on the governments' budget and the low access to subsidies for those who really deserve them. From here emerges the importance of studying the subsidies issue and trying to find answers to the questions posed by it such as whether it is necessary to change the current subsidies system or not.

At the beginning, the subsidies issue has to be raised through a number of sub-questions or sub-issues as follows:

- Why should the subsidies issue be addressed? The importance of this issue can be inferred from the continuous deficiencies the subsidies system suffers from and it's increasing bill endured by the society.
- Where is the subsidies issue located on the Egyptian citizen's agenda of interests? Through this question, the importance and rank of the subsidies issue among other issues that concern the Egyptian citizens' minds can be identified. As a consequence, the priority of working on and dealing with this issue among other issues that also need investigation can be determined as well.
- How do the citizens see the current subsidies system? Through this question, the level of awareness among the citizens about the existence of subsidies and their different types can be discussed, their views about the current subsidies' sufficiency and the sources of their knowledge can be determined, and the degree of their satisfaction about the present subsidies system and their view about the eligibility of its current beneficiaries can be identified.
- How is poverty distributed in Egypt? Through this question, the development of poverty rates and its geographical distribution in Egypt can be investigated.
- What are the characteristics of the poor in Egypt? Through this question, some topics can be discussed such as household characteristics, housing characteristics, and the geographical dimension.
- How can the poor be targeted in Egypt? Through this question, data sources and the wealth index are reviewed.

The syllabus or the curriculum of the principles of statistics (1) course given at the faculty of economics and political science in Egypt already contains the following topics:

1. The first unit: Definition of statistics, its branches, and data collection sources and methods
2. The second unit: One variable's data description (tabulation and graphs)
3. The third unit: Measures of location
4. The fourth unit: Measures of dispersion
5. The fifth unit: Measures of skewness and outliers detection
6. The sixth unit: Two variables' data description (tabulation and graphs)
7. The seventh unit: Correlation between two variables
8. The eighth unit: Regression and time series.

Table (2) shows the allocation of the principles of statistics course topics on the sub-issues of the subsidies issue and the proposed number of hours to cover each topic of the curriculum topics.

5. CONCLUSION

This paper presented a proposal for teaching the principles of statistics course using the issue-based approach through the subsidies issue. The curriculum topics were divided and allocated on the sub-issues of the main subsidies issue and a timetable for teaching this syllabus was also proposed. It is important to note that the way of presenting the subsidies issue is not unique, as it can be presented in various other ways. Moreover, other issues can be also discussed. It is also worth mentioning that the idea of using the issue-based approach is a general one that is applicable on all curriculums and specializations other than statistics. This paper can help in:

1. Designing selective curriculums
2. Designing training courses
3. And, Designing capstone courses like what exists in the statistics departments of some universities such as the statistics department at the United Arab Emirates University.

Table (2): The allocation of the principles of statistics course topics on the sub-issues of the subsidies issue and the proposed number of hours to cover each topic of the curriculum topics

	Why should the subsidies issue be addressed ?	Where is the subsidies issue located on the Egyptian citizen's agenda of interests?	How do the citizens see the current subsidies system?	How is poverty distributed in Egypt?	What are the characteristics of the poor in Egypt?	How can the poor be targeted in Egypt?	Proposed time in hours
1. Data collection sources and methods		4	2	1	1.5	1.5	**10**
2. One variable's data description	1	2	1	1	2	1	**8**
3. One variable's descriptive statistics (measures of location, dispersion, and skewness)		3	3	2	2	2	**12**
4. Two variables' data description		1.5	1	0.5	1		**4**
5. Correlation between two variables		2	1		1		**4**
6. Regression and time series	1			2		1	**4**
Total	**2**	**12.5**	**8**	**6.5**	**7.5**	**5.5**	**42**

REFERENCES

Boyle, C. R. (1999). A problem-based learning approach to teaching biostatistics. *Journal of Statistics Education*. *7*(1). Retrieved February 23, 2009, from http://www.amstat.org/publications/jse/secure/v7n1/boyle.cfm

Curriculum Development Institute (2007). *Issue-based learning for science and technology curriculum*. Presented at Issue-based Learning for "Environmental Science and Sustainable Development" Seminar, Hong Kong. Retrieved February 23, 2009, from http://cd1.edb.hkedcity.net/cd/science/S&T/training/briefing/Rationale_for_IssueBased_Learning.doc

Garratt, D., and R. Taylor (2004). *Issue-based teaching in economics*. Discussion Papers in Applied Economics and Policy. The Nottingham Trent University, Nottingham, UK.

Hadgraft, R. (2000). A personal journey towards a problem-based approach to university teaching: Case studies in flexible learning number 2. Centre for Higher Education Development, Monash University, Clayton, Australia.

Holbrook, J., A. Laius, and M. Rannikmäe (2003). *The influence of social issue-based science teaching materials on student's creativity*. Paper presented at the International Conference on E-learning in Science and Environmental Education, Tartu, Estonia.

Lewis, S. E. (2002). Investigating reformulated gasoline in an issue-based environmental science course. *Bioscene, 28*(2), 15-20.

Lewis, S. E. (2003). *Issue-based teaching in science education*. Retrieved February 23, 2009, from http://www.actionbioscience.org/education/lewis.html

Mimbs, C. A. (2005). Teaching from the critical thinking, problem-based curricular approach: Strategies, challenges, and recommendations. *Journal of Family and Consumer Sciences Education, 23*(2), 7-18.

Onosko, J., and L. Swenson (1996). Designing issue-based unit plans. In R. W. Evans and D. W. Saxe (Eds.), *Handbook on teaching social issues*. National Council for the Social Studies, Washington, DC.

Poon, S. K. (1998). *An issue-based framework for open learning in safety education and professional development*. Paper presented at the 3rd International Conference on Open Learning, Brisbane, Australia.

Stinson, J. E., and R. G. Milter (1996). Problem-based learning in business education: Curriculum design and implementation issues. *New Directions for Teaching and Learning, 1996*(68), 33-42.

Ward, J. D., and C. L. Lee (2002). A review of problem-based learning. *Journal of Family and Consumer Sciences Education, 20*(1), 16-26.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 333–340.

BRIDGING THE DIVIDE BY SCREENCASTING IN AN INTRODUCTORY STATISTICS CLASS AT AN AUSTRALIAN UNIVERSITY

Shahjahan Khan, Birgit Loch and Christine McDonald
Department of Mathematics and Computing
University of Southern Queensland
Toowoomba, Qld 4350, AUSTRALIA
E-mail: khans@usq.edu.au

ABSTRACT

Emerging technologies are being used increasingly to improve the communication and quality of instruction in the statistics classroom. Whereas a decade ago the move commenced away from blackboard writing with chalk towards computer assisted lecture delivery via PowerPoint or Beamer presentations, we are now seeing increasing use of Tablet technology to create dynamic and engaging lectures. While usually only students physically attending lectures would benefit, combined with screencasting – video capturing of the full computer screen along with audio recording – a near live lecture experience can be provided to students who cannot attend lectures. This is particularly valuable at universities with students enrolled in distance mode. This study investigates the benefits and downsides of providing lecture screencasts to oncampus and distance students of an introductory statistics class at the University of Southern Queensland (USQ). After technical issues were overcome, the recorded lectures were well received by the students, and this anecdotal evidence is supported by positive student evaluations.

Keywords: Tablet technology; Camtasia Studio; Online and distance teaching; screencast

1. INTRODUCTION

The work presented in this paper has developed from a University of Southern Queensland (USQ) Learning and Teaching Fellowship which provided Tablet PCs to academics in all faculties to evaluate benefits to lecturers and effect on student learning. The Tablet PCs were also equipped with screen recording and PDF annotation software. The project described in this paper investigates the benefits and downsides of providing screencasts of live lectures to students in an introductory statistics class.

About three quarters of students enrolled at USQ study via distance mode, which means they can complete their degree without physically setting foot on campus. Many of these students may be located close to the university campuses and yet simply choose to study in this mode for family or work related reasons. However, most students are located too far away to attend classes; in fact quite a few USQ students study from overseas. When the dual mode (oncampus and distance) was introduced in 1977 long before the availability of personal computers for teaching purposes, the university started its distance teaching via printed mailed out materials. To meet the instructional needs of these students, the university has traditionally been one of the early adopters and developers of new teaching technologies, for instance, when trialing video conferencing and audiographic tutorials in the late 1990s (Harman & Dorman, 1998), or when moving to an online offer in the early years of the Internet (Reushle & McDonald, 2000).

Today, study material is sent to students via CD (or DVD, as appropriate), including multimedia learning objects, and also made available through the Open Source Learning Management System (LMS) Moodle. The university has even commenced to move away from providing printed material to students.

To decrease the gap between services provided to oncampus students (lectures, tutorials) and distance students, the university developed lecture recording software ("IPLOD"), which produced an audio recording and still screen images of a live lecture. This tool, however, did not allow recording of whiteboard writing in class, was cumbersome to use, and has since been phased out. Macromedia Breeze (now Adobe Presenter) was made available in 2005, to record audio on top of PowerPoint slides (Birch, 2009). However, due to its user interface, this tool can only be used for off-line recordings in an office situation rather than for recording live lectures.

Through the fellowship, a Tablet PC was made available to the first author with Camtasia Studio screen recording software and PDF Annotator software to write on PDF documents. The Tablet PC was used to annotate Beamer slides in a large introductory statistics course servicing students from a variety of different disciplines. Screen movement and audio were recorded in all lectures and uploaded to the course LMS site for student access.

This paper is one in a series of investigations into the use of Tablet PCs and related technologies in higher education to promote student engagement and effective learning and teaching practices at USQ. It describes results from a study on lecture screencasts in an introductory statistics course with annotations facilitated by a Tablet PC where appropriate.

2. BACKGROUND

Tablet PCs and dedicated tablet operating systems taking advantage of the special ink-enabled features have been available since the beginning of this decade, however their uptake has been slower than one might have expected. In the last couple of years, a steady increase in adoption of tablet technology has been observed in Australia, partly due to the reduced cost to purchase a Tablet PC, but also because of word of mouth about its usefulness in higher education contexts.

For instance, in the mathematical sciences where it is important to write by hand when explaining the pathway to the solution of a problem, and where PowerPoint-only or PDF/Beamer presentations are usually complemented by whiteboard writing, the Tablet PC can become a tool that removes the need for the whiteboard when writing is shifted to the tablet screen, as evaluated for teaching introductory mathematics courses for engineering and computer sciences students (Loch, 2005; Loch & Donovan, 2006) and for a number of mathematics courses (Fister & McCarthy, 2008). Another advantage of writing on the screen in class is that a record of the writing can be kept and annotations can be made available afterwards to students for revision purposes, if they have missed a class, or if they are enrolled in distance mode. However, as with all new technologies, the computer technology (hardware or software) may not be reliable and the level of comfort of the lecturer translates to students' perceived effectiveness of the tool (Loch & Donovan, 2006).

Screencasting has been available for a number of years, reaching back to 1993 according to Budgett, Cumming and Miller (2007). However, not many accounts of the effective use of screencasting in statistics teaching and the impact on student learning can be found in the literature. Moving on from pre-recorded PowerPoint and audio lectures, Budgett et al. (2007)

tried Tablet PC annotated lecture screencasts recorded in BB-Flashback across a number of courses, one of which was a first year introductory data analysis course similar to the course evaluated in this study. The lecturers wrote on PowerPoint slides in this course, and recordings were regarded as complementary to lectures, not to replace them. The most important reasons for providing the recordings were so students could revise at their own pace for reinforcement, or catch up when they had missed a class. In addition, the recordings could be played back several times, which benefits for example international students who may not have a good grasp of the English language.

Bilder and Malone (2008) give a brief overview of a number of applications of the Tablet PC in teaching statistics courses, for instance writing in lectures, recording screencasts, and broadcasting live classes via web conferencing software. Further applications of Tablet PCs in mathematics or statistics teaching can be found in (Loch & McDonald, 2007; Reushle & Loch, 2008). These also include synchronous handwritten chat tutorials and web conferencing sessions.

3. INK ANNOTATIONS AND SCREENCASTING IN A STATISTICS COURSE

At USQ Data Analysis is a first year introductory statistics course offered to oncampus and distance students across the university. While taken by mathematics major students, it mainly caters for students not majoring in Mathematics or Statistics who quite often have weak knowledge of the elementary mathematics required to successfully complete the course. This issue is magnified for distance students who do not have the option available to oncampus students to attend lectures or a lecturer's face-to-face office hours. Ink annotations on a Tablet PC and screencasting of these annotations has opened up a variety of opportunities for distance students to engage more effectively with the content of a course which is perceived by many to be difficult.

One of the requirements of the Data Analysis course is that students use the SPSS statistical software package for displaying and analyzing real-life data. Students were provided with a booklet of instructions, including screenshots, on how to use SPSS to perform specific analyses required in the course. Oncampus students were offered classes in computer laboratories where considerable assistance was available. With only these static forms of support, distance students struggled to become competent in using SPSS and often required lengthy email explanations or telephone assistance. In 2007 a member of the teaching team prepared screencasts using Camtasia Studio to capture click-by-click steps performing these analyses and discussing the output produced in SPSS. Students have commented on how helpful these have been and that they have found it easy to learn SPSS using the recorded materials. There have been far fewer calls for assistance since the screencasts were made available.

In 2008 short screencasts explaining particular concepts were produced. These screencasts were targeted at concepts that students found particularly difficult such as probability and sampling distributions. Annotation on blank PowerPoint slides using a Tablet PC allowed the tutor to develop explanations by sketching graphs and writing formulae in "real-time". These screencasts have allowed students to revisit these concepts throughout the semester as necessary. On request short screencasts of ink annotated worked exercises have been emailed to distance students to supplement explanation of the workings of a problem. These screencasts need to be very short to allow them to be made available via email. This has allowed distance students to receive the type of assistance that until now has only been available to oncampus students in a face-to-face consultation with a lecturer.

In Semester 1 2009, all lectures in Data Analysis were recorded live on a Tablet PC using Camtasia Studio. During the lectures the lecturer remained close to the computer and talked into the in-built microphone which resulted in some static being recorded. Using PDF Annotator he highlighted important concepts presented on Beamer slides with the ink annotating functionality of the Tablet PC. He expanded on explanations of statistical and mathematical concepts, for instance providing additional examples that previously would have been written on a whiteboard in class and not captured. Recordings were edited after class where required, and made available in Adobe Flash format for online play back in students' web browsers. A typical screenshot of the recorded lectures using Camtasia Studio is shown in Figure 1.

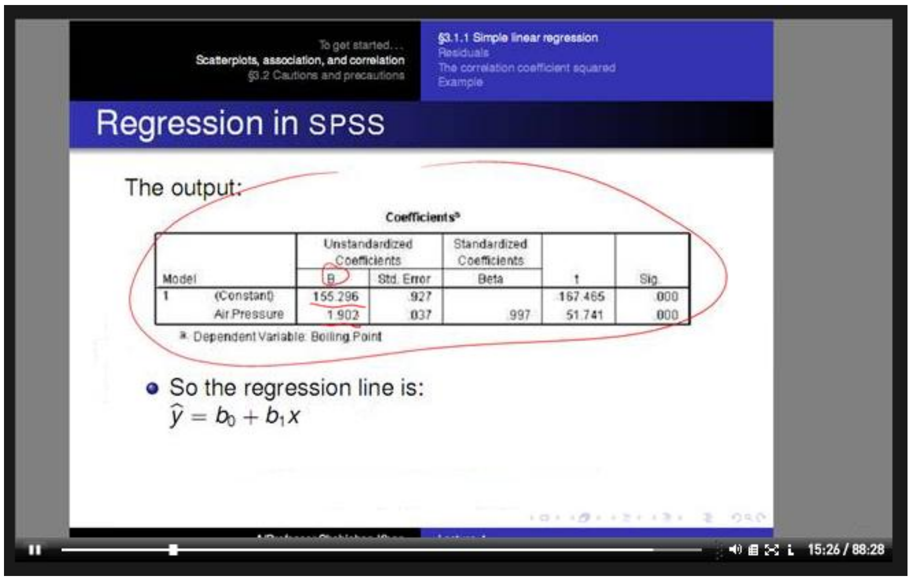


Figure 1: Example of an annotated Beamer slide using PDF Annotator.

Since the Tablet PC allows writing on the screen while presenting lectures in PowerPoint or PDF format, highlighting of the text and modifications to formulae along with drawing of graphs can be incorporated "on the fly", as needed. This flexibility of writing on the computer screen makes the Tablet PC very suitable to capture the contents of any live lecture. Along with the voice recording, the screencast is almost like a live lecture for distance students except for the lack of opportunity to ask questions or to make comments during the class time. This shortcoming can be offset by the opportunity to ask questions, including any related to the recorded lecture, via the course discussion forum in the Moodle LMS.

Since the introduction of the Moodle LMS at USQ in 2007, all course materials and related resources in Data Analysis have been made available from the course site, accessible to all enrolled students. This includes static content such as an Introductory Book, Study Book, assignment instructions, revision problems and past examination papers with answers, SPSS instructions booklet, lecture slides and tutorial worksheets. Each week as the semester unfolds, the lecture notes and tutorial worksheets are posted in advance on the homepage so that students may download and copy them to use in class. This significantly reduces the need for hand-copying the notes, and thereby allows students much needed time to think about and to grasp the new ideas/concepts and understand the methods, rather than busily copying the notes. It also gives the students a clear idea on the materials to be covered with an opportunity to prepare well before participating in the actual lecture and tutorial.

Since late 2007 these static resources have been complemented by the addition of more dynamic presentations such as screencasts of a variety of activities such as SPSS tutorials, explanations of difficult concepts and now complete class lectures. The provision of these screencasts allows students to view them immediately following a lecture and access them again, as needed, prior to the submission of an assignment or in preparation for an examination. Concepts that were not readily assimilated at the time they were presented in the live lecture can be revisited as often as needed until they have been understood.

To round off the wide assortment of resources offered in Data Analysis, the availability of discussion forums has led to a significant amount of student activity on the course site, where students post any problems and discuss any issues that may arise from their studies. Lecturers and tutors take the opportunity to reinforce concepts and methods by way of answering postings of the students.

4. STUDENT ACCESS AND PERCEPTION

Throughout the semester lecture notes were posted on the course LMS at least 3 days prior to the class in each of the 13 weeks of teaching. Many students, both oncampus and distance, accessed the lecture notes before the delivery of the lecture. Recorded screencasts for each lecture were posted to the course LMS site immediately after the delivery of lecture. Detailed data on the use of the teaching resources on the course site including the number of times the lecture notes and screencasts were accessed were extracted from the site. The data shows very strong correlation (r=0.92) between the number of times the lecture notes and screencasts were accessed (see Figure 2). This very high positive correlation reinforces the fact that most of the students who accessed lecture notes also accessed the screencast of the lecture.

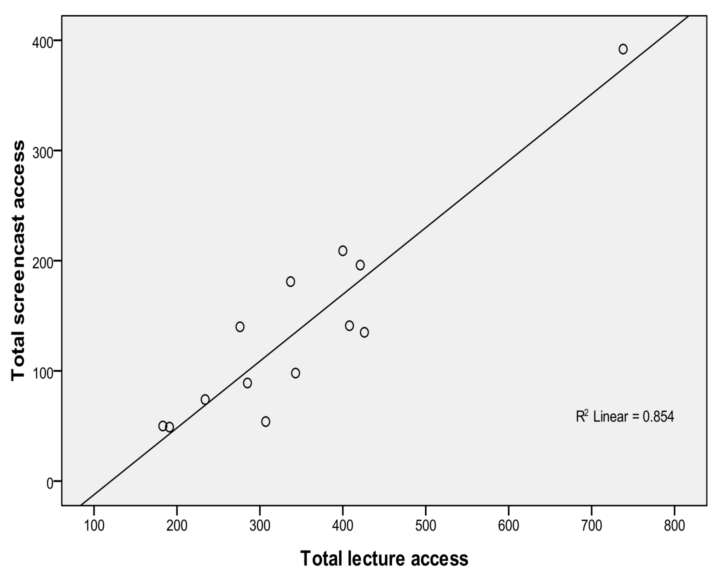


Figure 2: Scatterplot of the relationship between access to lecture slides and access to lecture screencasts across the 13 weeks of the semester.

The significant slope (see Table 1) indicates that one access to the lecture notes corresponds to an increase in the access to the recorded lecture of 0.606. The large value of R-square (0.854) suggests that 85.4% of the variation in the number of accesses to the screencasts is accounted for by the number of accesses to the lecture notes. Not all students who accessed the lecture notes accessed the lecture screencasts and this could have been due to the size of the download required to access a whole lecture. Students in regional and remote areas of Australia still have limited access to broadband internet and rely on slow dial-up connections, however this is improving.

Table 1: Regression analysis of the relationship between access to lecture slides and access to lecture screencasts across the 13 weeks of the semester.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-72.924	28.420		-2.566	.026
Total lecture access	.606	.076	.924	8.014	.000

As with the introduction of any new technology, some difficulties with the recording of lectures due to lack of experience in using the Tablet PC and the screen recording software occurred in the first week of teaching. On some occasions, technical difficulties caused freezing of the screen forcing the lecturer to stop the recording and restart the process again. As well as some students' uneasy feelings about any such disruptions, this kind of technical difficulty intermittently created unexpected delays which decreased the time available to cover the planned content of the lecture and material had to be held over to the following week. One student requested that the recorded lectures be posted in a downloadable format instead of the online playback format, so that it could be saved for future use. Some students complained about the voice level in a few of the initial recordings, but this was quickly addressed by using a better quality microphone.

Some typical student comments on the usefulness of the recorded screencasts and lecture notes from the end of semester Student Evaluation report are listed below:

Student A: *That the lectures were posted on StudyDesk so I could go back and look up anything I didn't understand.*

Student B: *As an external student I found the recorded lectures and lecture notes to be most helpful.*

Student C: *Shahjahan and his team did an exceptional job this semester, i.e. recorded lectures and notes, prompt marking and returning of our assignments.*

Student D: *The lecturer was excellent -the audios of the lectures available on the internet were really helpful.*

Student E: *The recorded lectures. Please continue with them. As an external student, I ended up only focusing on the lecture notes.*

Student F: *The recorded lectures were the most helpful resource I had.*

5. DISCUSSION

With the introduction of screencasts of lectures there is the question of whether the live lecture will continue to be supported by oncampus students. The oncampus lecture provides more than just an avenue for dispensing content. Students attend oncampus lectures for a variety of reasons, including social interaction, collaborating with peers, keeping on track and the opportunity to ask questions and get immediate answers. Only time will tell if attendance at oncampus lectures will continue at the same level. For the moment there does not appear to be a drop in turnout but it may be that lecturers will need to rethink the purpose of the lecture format and offer a different experience to the traditional lecture presentation for oncampus students. It may be that there will be an expectation that students watch a lecture screencast before coming to class and then participate more collaboratively in a discussion of the content of the screencast in the oncampus "lecture". For distance students the introduction of screencasts has meant that they can now access a more "oncampus experience". Recorded lectures provide equity for all students. Screencasting of lectures in visually rich content such as found in a statistics course is integral to bridging the divide between the oncampus and distance students' experiences.

One of the advantages of recording lectures in Camtasia Studio is that a variety of output formats can be generated; catering for a mixture of student needs and circumstances. Some students will want audio only so that they can listen to the commentary while looking at a printout of the lecture slides on the train to work. This version would also reduce the size of the download. Others may want to play the video on a portable multimedia player such as an iPod or mobile phone. Providing the screencasts in all of the available formats will give students choice.

We agree that one of the benefits of providing screencasts reported in Budgett et al. (2007) is that students, for whom English is not a first language, may be able to replay a lecture as often as needed to understand the material. However, we found in our study that an additional benefit was also in allowing all students to replay lecture recordings by a lecturer whose first language is not English and whose accent may be unfamiliar to most and therefore not as easily understood.

6. CONCLUSION

This paper has reported on one of many courses at USQ where Tablet PCs are used to make lectures more dynamic and interactive. Being able to capture this dynamic, using software such as Camtasia Studio, has gone some way towards bridging the divide between oncampus and distance students' experiences of learning introductory statistics since recordings are accessible to all students.

The importance of providing screencasts for distance students has been recognized across USQ, and following trials in 2008 and early 2009, the university has adopted Camtasia Relay in semester 2 2009. Camtasia Relay is the lecture theatre equivalent of Camtasia Studio screen recording software. The software is installed on all lecture theatre computers and lecturers can record lectures whenever required, without the need to purchase an individual license. This has expanded the opportunities that lecturers have for providing distance students with a live lecture experience. As technological advances are taken up by more and more academics in higher education, students will come to expect that all courses should provide such support. The challenge is there to be met.

REFERENCES

Bilder, C. R. and Malone, C. J. (2008). Tablet PC Applications in Statistics Education, Part I. In *JSM Proceedings*, Statistical Education Section. Alexandria, VA: American Statistical Association.

Birch, D. (2009). PowerPoint with audio: a breeze to enhance the student learning experience. *e-Journal of Business Education and Scholarship of Teaching*, 3 (1), 36-42.

Brodie, L. and Loch, B. (2009). Annotations with a Tablet PC or typed feedback: does it make a difference? *Proceedings of the 2009 AaeE Conference*, Adelaide, December 2009.

Budgett, S., Cumming, J. and Miller, C. (2007). The role of screencasting in statistics courses. *Proceedings of the International Statistical Institute, 56th Session*. Lisbon, Portugal. www.stat.auckland.ac.nz/~iase/publications/isi56/CPM82_Budgett.pdf

Fister, K. R. and McCarthy, M. L. (2008). Mathematics Instruction and the Tablet PC. *International Journal of Mathematical Education in Science and Technology*, 39 (3), 285-292.

Harman, C., & Dorman, M. (1998). Enriching distance teaching and learning of undergraduate mathematics using videoconferencing and audiographics, *Distance Education*, 19 (2), 299-318.

Loch, B. (2005). Tablet Technology in First Year Calculus and Linear Algebra Teaching. In *Proceedings of the Fifth Southern Hemisphere Conference on Undergraduate Mathematics and Statistics Teaching and Learning (Kingfisher Delta'05)*, 231-237.

Loch, B. and Donovan D. (2006). Progressive Teaching of Mathematics with Tablet Technology. *e-JIST, e-Journal of Instructional Science and Technology*, 9 (2).

Loch, B., and McDonald, C. (2007). Synchronous chat and electronic ink for distance support in mathematics. *Innovate - Journal of Online Education, 3*(3).

Reushle S. and Loch B. (2008). Conducting a Trial of Web Conferencing Software: Why, How and Perceptions from the Coalface. *Turkish Online Journal for Distance Education*, 9(3). http://eprints.usq.edu.au/4254/

Reushle, S. E., and McDonald, J. (2000). Moving an Australian dual mode university to the online environment: A case study. *Proceedings of ED-MEDIA2000*, June 26-July 1, Canada: Montreal, 907-12.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 341–346.

MODES OF READING AND INTERPRETING STATISTICAL GRAPHS AMONG SECONDARY SCHOOL STUDENTS

Omar ROUAN
GREDIM
Ecole Normale Supérieure
Marrakech- Maroc

1. INTRODUCTION

There are two key factors behind the interest we attach to the teaching and the learning of statistical graphics (SG). The first is the increasing importance given to them. On the one hand, they are special in the sense that they differ from other kinds of graphs by the situations they represent, by the principle of their construction and also by the reasoning they generate. On the other hand, they are ubiquitous in various scientific and technological fields and in different school subjects. Finally, they are frequently encountered in the social environment, through newspapers, magazines, journals and other kinds of publications.

The second factor is related to the importance we attach to teaching and learning them. Because of its holistic way of representing statistical situations, the SG helps in making statistical concepts and functions more palpable and more accessible to students. Mac Gregor and Slodvic (1986) discuss some kinds of SG that permit and facilitate statistical judgments better than others.

Moreover, the introduction of computers in schools makes the interpretation of statistical graphics more urgent than their construction**.** Finally, the different properties of SG, their functions and their interpretation are rarely discussed in statistical books and manuals**.** This leads us to the following questions:

- Does our teaching of SG accord enough importance to their graphical analysis, to their reading and their interpretation?
- What type of graphical activities does it allow?
- Does it permit a better understanding of these graphics?
- Does it enable students to better control the components, the characteristics and the functions of these graphics?
- Does it enable students to use them in a better way?

2. RESEARCH PROBLEM

For the sake of a profound study, we sent a questionnaire to 61 secondary school mathematics teachers. The questionnaire contains items about their conception of Statistics, their vision of the objectives of teaching SG, the difficulties of teaching and learning these graphics, and finally the ways they use them in their classes.

One of the most important results highlighted by the survey was the fact that construction activities dominate those of reading and interpretation activities which seem to be completely absent. The difficulties associated with construction are more numerous and more detailed than those associated with the two other operations. Finally, any definition of reading or interpretation has been given by teachers. Then we can ask: How do students conceive the studied SG? What types of difficulties in reading and interpreting SG, should we expect from students?

3. LITERATURE REVIEW

Works on graph classification allowed us to compare the criteria and the methodology used by different authors to classify these graphs. The statistical graphs we intend to study associate values of some variables to those of another and are thus among the graphics that Bertin (1967) classifies as diagrams.

Works on understanding graphs presented different models of understanding. These models are related to different purposes such as assessing the effectiveness of a graphic, the elaboration of a software program that can read and interpret a graph, measuring the effect of subjects' knowledge on their ability to read or interpret a graph, etc. They all come to define the general components of a graph and the aspects (or criteria) under which the chart will be examined or analyzed. Depending on the purpose and the problem, the model gives more weight to one side than another. Among these models, we adopted Lamrabet's (1993.1999) which brings the understanding of a chart to the following aspects: the structural, the operative, the semiotics, the descriptive, the functional and the epistemological aspect.

Works on the interpretation of graphs have reported a series of interpreting difficulties. According to Duval (1993), the main reason behind the difficulties related to reading and interpreting Cartesian graphs, is not to search in the mathematical concepts related to affine functions, but in the ignorance of the rules of the semiotic correspondence between the graphical register and the algebraic notation.

As for Janvier (1993), he describes a difficulty which results from the interference of the "source mode" in the "target mode". The drawings and diagrams ("source mode") representing a situation will interfere with future production in the "target mode" (determined by context). When this problem occurs in the translation "Graphic / verbal description", it is because the iconic and the spatial aspect of the chart outweigh its symbolic aspect. In Janvier's words (1993), it is a visual destroyer belonging to the source mode.

Works on the historical development of statistical graphics have enabled us to identify their historical functions. At the end, the works on teaching and learning statistical graphics brought some research on the characteristics of some graphs and the conditions for their proper use.

4. TERMINOLOGICAL FRAMEWORK AND RESEARCH OBJECTIVE

Our literature review helped us clarifying the meaning of the terminology used. Thus, reading a statistical graph consists of decoding signs, symbols and syntactic conventions that constitute it. Its interpretation consists of extracting information that is not directly readable on it, and relies on its characteristics, on the assumptions behind its construction and on its links with statistical and mathematical concepts.

Concerning the notion of difficulty, it differs from that of the epistemological obstacle in the sense that it can be exceeded without a radical change of the cognitive structure of the subject. To be remediated, it may be a matter of time, of information or of instruction. It has also allowed us to develop a model for understanding graphs, useful for the analysis of our charts and for the categorization of the difficulties that we can reach. It has also allowed us to identify the different functions of the statistical graphics, through their historical development. Among these functions are the following: description, comparison, inference, modelling, and exploration. Finally, it allowed us to restate our research objectives and questions.

Our objective is on the one hand to identify the difficulties of reading and interpretation of statistical graphs among secondary school students. On the other hand, to precise the understanding aspects to which these difficulties are linked. What are the difficulties of reading and interpreting among students relating to various graphs selected? What understanding aspects are related to these difficulties?

5. RESEARCH METHODOLOGY

This exploratory research adopts the questionnaire as methodological tool. The studied graphs were those treated with the secondary curriculum in Morocco, namely the histogram, the pie chart and bar graph. The data represented by these graphs are taken from the high school math textbook. To achieve our objective, we used questionnaires covering each one of the charts reviewed. When the issue is an interpretation of the graph, a justification is required. The questionnaires were administrated to students in first year of the high school. The students were selected so as to represent all levels (low, medium and strong). 147 have responded to the questionnaire on the histogram, 125 to the one on the bar graph and 141 stick to the one on the pie chart.

6. DATA ANALYSIS & RESULTS

For the questionnaires, we conducted percentages analysis associated with each issue. We also analyzed the given justifications. Our analysis of the data enables us to identify different modes of reading and of interpretation of the studied graphs. These modes generate comprehension difficulties that are related either to the nature of the graphics or to the statistical concepts used to construct it.

The discrete mode of reading

This mode of reading was especially relied to the histogram. It considers the classes' extremities as the terms of the statistical variable in a histogram. This mode is characterized by the absence of the concept of class and the dominance of a discrete argument. It ignores the concept of class as well as the conventions or assumptions relied to them and the different criteria used for grouping data into classes.

The ordinal mode of reading

Especially related to statistical graphics using Cartesian axes, this mode of reading conceives the ordinal numbers on the vertical axis as the reference numbers ordering the observed individuals on this axis. In this mode, the values of the variable are kept on the horizontal axis, while individuals are placed on the vertical axis

The functional mode of reading

Related to the histogram, this mode considers this graphic as a step function (a function which is constant by intervals). It associates to all elements of the same class, the class size. Note that for the last two modes, the concept of statistical distribution is completely absent

The analogical mode of reading

This mode of reading is based on a perfect analogy between the graph and the represented situation. It supposes that the graph reproduces the whole reality. This seems to imply that one can read only what is actually written on the chart without leaving an occasion for the integration of the implicit statistical and graphical conventions and assumptions.

7. CONCLUSIONS

These results show to difficulties related to the statistical concepts, to the characteristics of the studied graphics, to the relationship between these graphics and the components of the statistical reasoning, and also to the functions of the statistical graphics. These results mean the need to develop a new vision of statistics among both teachers and students. The latter must be developed in close conjunction with the applications of statistics in various fields such as demography, economics and medicine, as it must give great importance to the question of the meaning of statistical concepts and parameters. We believe that the interpretation of these concepts and parameters is the best way to achieve this purpose.

We also believe that this vision has to attach great importance to the analysis of SG, to their exploitation (to highlight their functions), to their reading and interpretation. We believe that the understanding model adopted by our research provides a framework for analyzing them. In addition, for teachers training, we believe that the focus has to be on the epistemological and didactical analysis of statistical knowledge (heuristic aspects, question of the meaning). This means to provide a great importance to the analysis of statistical concepts and tools. Additionally, to show the various utilities of descriptive statistics such as data organization, graphical representation of data, reduction of data; and the different inductive uses such as prediction, modelling, estimation, decision making, sampling, analysis of series, chronological analysis of multidimensional data.

8. RESTRICTIONS AND NEW PATHS OF RESEARCH

Our research has proved too ambitious, it has touched many aspects: relationship between statistical graphics and statistical or mathematical concepts, different statistical heuristics, students' difficulties of reading or interpreting a graphic, etc. Therefore, we believe it highlights other research questions on each of these components. For example:

1. The meaning we can associate to the different configurations of a histogram or of a bar chart
2. The estimation of the arithmetic mean on the base of those different configurations.

We detected many cases of students' confusion about the average with other parameters such as the maximum, the mode or the medium. But we did not have sufficient evidence explaining this confusion. We think that this may open the ground for a new research, which will focus on the different conceptions of the average and their relation with the mentioned concepts. As we reported, the modes of reading and interpretation may constitute obstacles to reading and interpreting these SG. Verification of this hypothesis and determination of the nature of the obstacles may also be subject to further research.

REFERENCES

Belisle Claire and Guy Jouannade (1988). "La communication visuelle", Les éditions d'organisation.

Bell A. and Claude Janvier (1981). "The interpretation of Graphs Representing situations", For the learning of Mathematics, 2, 1.

Beniger J. R. & Robyn D. L. (1978). "Quantitative graphics in Statistics: a brief history", The American Statistician, vol 32 numéro 1, pp 1-11.

Bertin J. (1967). "Sémiologie Graphique" Editions Mouton &Cie.

Bibby John (1991). "Karl Pearson and the history of teaching Statistics", Volume 13, N°2.

Biehler, R. (1990), "Changing Conceptions Of Statistics: A Problem Area For Teacher Education", From: Hawkins, a. (ed): Training Teachers To Teach Statistics. Voorburg: Intenational Statistical Institute, pp 20-38.

Bouroche J. M. & Saporta G. (1987). "L'analyse des données", Que sais-je.

Cheysson M. E. (1887) "la statistique géométrique, Méthode pour la solution de problèmes commerciaux et industriels", Publications du journal la génie Civil.

Cleveland, W. S. (1985). “The Elements of Graphing Data”. Monterey, California, Wadsworth Health Science.

Costigan-Eaves Patricia and Micheal Macdonald-Ross (1990). "William Playfair (1759-1823)", Statistical science, vol 5, No 3, pp 318-326.

Delacroix M. (1983). "Histogrammes et estimation de la densité", Que sais-je, Presses universitaires de France.

Du Toit, S. H. C. et al. (1986). "Graphical Exploratory Data Analysis", Editions Springer-Verlag.

Fienberg S. E. (1979). "Graphical Méthods in Statistics", The American Statistician, vol 33, pp 165-178.

Graham A. (1987). "Statistical Investigations in The Secondary School", The Open University, Centre For Mathematics Education (PM646), Cambridge University Press, Cambridge.

Groupe Statistique de l'IREM de Rouen (1991). "Graphiques au Collège".

Jambu M. (1999). "Méthodes de base de l'analyse des données", Editions Eyrolles.

Janvier, C. (1978). "The interpretation of complex Cartesian graphs representing situations - studies and teaching experiment", Thesis submitted to the university of Nottingham for the degree of Doctor of Philosophy.

Lohse et al. (1991). "Classifying graphical information", Behavior & information technology, vol 10, N° 5 pp419-436.

Mac Gregor and Slodvic (1986). "Graphic Representation of Judgmental information", human computer interaction, Volume 2, pp 179-200.

Nisbett et al. (1983). "The use of Statistical Heuristics in Everyday Inductive Reasoning", Psychological review, vol 90, N° 4, pp 339-363.

Preece Jenny and Claude Janvier (1992). "A study of interpretation of trends in multiple curve graphs of ecological situations", School Science and Mathematics, Vol 92 (6).

Radhakrishna C. Rao (1989). "Statistics and Truth: Putting chance to work", Council of scientific & Industrial research.

Rouan O. (2001). "Lecture et interprétation des représentations graphiques des données statistiques chez les élèves et les enseignants du secondaire" Thèse présentée à l'université Mohamed V de Rabat pour l'obtention du Doctorat d'état en sciences de l'éducation, spécialité didactique des mathématiques.

Royston Erica (1956). "Studies in the History of Probability and Statistics: A note on the history of the graphical représentation of data" Biometrika, Volume 43, parties 3 et 4, pp 241-247.

Volle M. (1993). "Analyse des données", collection Economie et statistiques avancées, Les éditions Economica.

CONTRIBUTED PAPER SESSIONS

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010) 349–356.

BAYESIAN ESTIMATION FOR THE EXPONENTIATED WEIBULL SHAPE PARAMETER UNDER LINEAR AND ZERO-ONE LOSS FUNCTIONS

Amina I. Abo-Hussien[1], Abeer A. El-helbawy[2]
and Eman A. Abd El-Aziz[3]

Department of Statistics
Faculty of Commerce, Al-Azhar University (Girls' Branch)
E-mail: [1] aeabohussien@yahoo.com, [2] a_elhelbawy@hotmail.com
[3] be123en@hotmail.com

ABSTRACT

In this paper, we derive Bayes estimates for one of the shape parameters θ and the reliability function of the two shape parameter exponentiated Weibull distribution under type II censoring. The analysis is carried out using two types of loss functions and three types of priors.

Keywords: Exponentiated Weibull family; Bayes estimator; Noninformative prior; Uniform prior; Natural conjugate prior; Linear loss function; Zero-one loss function.

1. INTRODUCTION

Mudholkar and Srivastava (1993) introduced the exponentiated Weibull (EW) distribution which is an extension of the Weibull family. The EW distribution plays an important role as a model for many lifetime data of bathtub shape or upside-down bathtub shape failure rates. Thus the EW distribution as a failure model is more realistic than that of monotone failure rates for representing such data. In other words the EW distribution can be widely used in reliability applications because its variety of shapes in its density and failure rate functions, so it is useful for fitting many types of data. Its applications in reliability and survival studies are illustrated in Mudholkar *et al*. (1995) using data on bus–motor failures and data on head and neck cancer. Its properties have been studied in more detail by Mudholkar and Huisen (1996) and Nassar and Eissa (2003). These authors have presented useful applications of the distribution in the modeling of flood data and in reliability. Sing *et al* (2002) have discussed the classical and Bayesian methods of parameter estimation for complete sample case. Also Sing *et al* (2004) obtained estimators of the parameters under type II censoring. Nassar and Eissa (2004**)** derived Bayes estimates considering conjugate prior for the shape parameter under squared error and linear exponential (LINEX) loss functions. The two shape parameter EW distribution has a probability density function (pdf) of the form

$$f(t)=\alpha\theta\, t^{\alpha-1} e^{-t^{\alpha}} (1-e^{-t^{\alpha}})^{\theta-1} \qquad t>0,\ \alpha>0,\ \theta>0 \tag{1}$$

and a cumulative distribution function

$$F(t) = (1-e^{-t^{\alpha}})^{\theta} \qquad t > 0 \tag{2}$$

where α and θ are the two shape parameters. When θ=1 the EW pdf is that of the Weibull distribution. For θ>1, the EW distribution has a unique mode, $\left[2\left(\alpha\theta-1\right)/\alpha(\theta+1)\right]^{1/\alpha}$. The median of the distribution is $\left[-\ln(1-2^{-1/\theta})\right]^{1/\alpha}$. The reliability and failure rate functions of the EW distribution are given, respectively, by

$$R(t) = 1-(1-e^{-t^{\alpha}})^{\theta} \qquad t > 0 \tag{3}$$

$$h(t) = \alpha\theta \; t^{\alpha-1} \; e^{-t^{\alpha}} (1-e^{-t^{\alpha}})^{\theta-1} \left[1-(1-e^{-t^{\alpha}})^{\theta}\right]^{-1} \quad , \quad t > 0 \tag{4}$$

In this article, Bayes estimates for the shape parameter θ when (the shape parameter) α is assumed known are developed under type II censoring. Here we use the linear and the zero-one loss functions assuming different priors.

The linear loss function has the following form:

$$L(\theta,\hat{\theta}) = \begin{cases} c_1\left|\hat{\theta}-\theta\right| & if \;\; \hat{\theta} < \theta, \\ c_2\left|\hat{\theta}-\theta\right| & if \;\; \hat{\theta} > \theta, \end{cases} \tag{5}$$

where c_1 and c_2 can be chosen to reflect the relative importance of under estimation (when $\hat{\theta} < \theta$) and over estimation (when $\hat{\theta} > \theta$). Generally c_1 and c_2 will be unequal. If $c_1 = c_2 = c$, then equation (5) reduces to the following form $L(\theta,\hat{\theta}) = c\left|\theta-\hat{\theta}\right|$, which is called the absolute error loss function. If c_1 and c_2 are functions of θ , the loss function is called the weighted linear loss function. Let

$$\gamma = c_1/(c_1+c_2) \tag{6}$$

Then under any prior distribution g (θ) , the Bayes estimator $\hat{\theta}$ of θ is the γ-th quantile of the corresponding posterior distribution g (θ/ T). Following Box and Tiao (1973**,** p.309), we will consider the following 0-1 (zero-one) loss function

$$L(\hat{\theta},\theta) = \begin{cases} 1 & if \left|\hat{\theta}-\theta\right| > \varepsilon, \\ 0 & if \left|\hat{\theta}-\theta\right| < \varepsilon, \end{cases} \tag{7}$$

where ε is any small positive number.

The objective of this paper is to develop Bayes estimates of the shape parameter (θ) and the reliability function of the two shape parameter EW distribution for the case of type II censoring data. The linear and the zero–one loss functions assuming three prior distributions [the noninformative prior, the uniform prior and the natural conjugate prior] are used. Bayes estimates under the noninformative prior, the uniform prior and the natural conjugate prior assuming the linear and zero-one loss functions are derived in Section 2.

2. BAYESIAN ESTIMATION FOR θ

Assuming that α is known (say $\alpha = \alpha_o$) then the p.d.f of the failure time t depends only on the single unknown parameter θ and then $f(t) = \alpha_o \theta\, t^{\alpha_o - 1} e^{-t^{\alpha_o}} (1 - e^{-t^{\alpha_o}})^{\theta - 1}, \; t > 0, \alpha_o > 0, \theta > 0.$

Using the observed m values: $t_1 < t_2 < \ldots < t_m$ of a type II (right) censored sample of size n, where t_i is the time of the i-th component to fail) , the likelihood function is :

$$L(T/\theta) = \alpha_o^r\, \theta^r\, e^{-Z}\, (1 - \nu^{\theta})^{n-m} \tag{8}$$

where

$$Z = \sum_{i=1}^{m} [t_i^{\alpha_o} - (\alpha_o - 1)\ln t_i - (\theta - 1)\ln u_i]$$

$$u_i = 1 - e^{-t_i^{\alpha_o}} \quad and \quad \nu = 1 - e^{-t_m^{\alpha_o}}$$

$$and$$

$$T = (t_1, t_2, \ldots\ldots, t_m)$$

We can simplify the likelihood function by using the binomial expansion theorem as shown below

$$(1 - \nu^{\theta})^{n-m} = (1 - e^{\theta \ln \nu})^{n-m},$$

$$(1 - e^{\theta \ln \nu})^{n-m} = (1 - e^{\theta p})^{n-m} = \sum_{j=0}^{n-m} (-1)^j \binom{n-m}{j} e^{j\theta p} \tag{9}$$

The posterior p.d.f of θ given the data T is given by

$$g_i(\theta/T) = \frac{L(T/\theta) g_i(\theta)}{\int_{\theta} L(T/\theta) g_i(\theta) d\theta} \tag{10}$$

which can be obtained for different priors $g_i(\theta)$, i = 1, 2, 3.

2.1 Noninformative Prior

2.1.1 Bayesian Estimation Under the Linear Loss Function

Considering the following noninformative prior distribution

$$g_1(\theta) = 1/\theta \quad , \quad \forall \;\; \theta \tag{11}$$

and under the assumption that the parameter α is known (say $\alpha = \alpha_\circ$), then the Bayes estimate $\hat{\theta}_{\gamma_1}$ can be obtained from solving the following equation

$$\int_0^{\hat{\theta}_{\gamma_1}} g_1(\theta/T)\; d\theta = \gamma \quad , \tag{12}$$

where $g_1(\theta/T)$ is the posterior density of θ which can be obtained by applying Bayes theorem as follows

$$g_1(\theta/T) = \frac{\theta^{m-1}\, e^{-\theta q}\,(1-v^{\theta})^{n-m}}{\int_0^{\infty} \theta^{m-1}\, e^{-\theta q}\,(1-v^{\theta})^{n-m}\, d\theta} \tag{13}$$

Using equations (8), (9), (10) and (11) it follows that the posterior p.d.f. is given by

$$g_1(\theta/T) = \frac{\sum_{j=0}^{n-m} c_j \theta^{m-1}\; e^{-\theta[q-jp]}\,[q-jp]^m}{\sum_{j=0}^{n-m} c_j\, \Gamma(m)} \quad , \qquad \theta > 0 \tag{14}$$

where $\quad c_j = (-1)^j \binom{n-m}{j} \quad , q = -\sum_{i=0}^{m} \ln u_i \quad , p = \ln v.$ (15)

$$u = \prod_{i=1}^{m} (1 - e^{-t_i^{\alpha}}) \quad and \quad v = 1 - e^{-t_m^{\alpha}} \quad .$$

From equations (12) and (14) one can obtain

$$\frac{\sum_{j=0}^{n-m} c_j \int_0^{y_{j,\gamma_1}} y^{m-1}\, e^{-y}\; dy}{\sum_{j=0}^{n-m} c_j \Gamma(m)} = \frac{\sum_{j=0}^{n-m} c_j IG\left[y_{j,\gamma_1}, m\right]}{\sum_{j=0}^{n-m} c_j \Gamma(m)} \tag{16}$$

where $y_{j,\gamma_1} = \hat{\theta}_{\gamma_1}(q - jp)$, and $IG(c, n)$ is the incomplete gamma function defined as

$$IG(c,n) = \int_0^{c} x^{n-1}\, e^{-x} dx$$

The values of y_{j,γ_1} can be obtained from solving equation (16) numerically and then the Bayes estimator

$$\hat{\theta}_{\gamma_1} = \frac{y_{j,\gamma_1}}{q - jp} \tag{17}$$

Note that If $c_1 = c_2 = 1$ and putting $\gamma = 0.5$ then the Bayes estimator $\hat{\theta}_{0.5}$ of θ is the median of $g_1(\theta/T)$ and can be obtained just by putting $\gamma = 0.5$ in equations (16) and (17).

Estimation of the Reliability Function

The Bayes estimator $\hat{R}_1(t)$ of the reliability function R (t) is obtained by solving the following equation $\Pr[R(t) < \hat{R}_\gamma(t)] = \gamma$, where R(t)=1-$\left[1-e^{-t^{\alpha_o}}\right]^{\theta}$, t >0. Noting that R (t) is a decreasing function in θ, hence

$$\Pr[R(t) \le \hat{R}_\gamma(t)] = \Pr[\theta > \hat{\theta}_{1-\gamma}] = \gamma \tag{18}$$

and the Bayes estimator is

$$\hat{R}_1(t) = 1 - \left[1-e^{-t^{\alpha_o}}\right]^{\hat{\theta}_{1-\gamma}} \tag{19}$$

where $\hat{\theta}_{1-\gamma}$ is obtained from equation (17) with γ replaced by 1-γ .

2.1.2 Bayesian Estimation Under the Zero-One Loss Function

Considering the noninformative prior given by (11) then the Bayes estimator of θ is the mode of the corresponding posterior density defined in equation (14).Taking logarithms of both sides of (14) yields the following equation

$$\ln g_1(\theta/T) = (m-1)\ln\theta + \ln\left(\sum_{j=0}^{n-m} c_j e^{-\theta(q-jp)} (q-jp)^m\right) - \ln\left(\sum_{j=0}^{n-m} c_j \Gamma(m)\right) \tag{20}$$

The first derivative for $\ln g_1(\theta/T)$ with respect to θ is

$$\frac{d\ln g_1(\theta/T)}{d\theta} = \frac{m-1}{\theta} - \left[\sum_{j=0}^{n-m} c_j e^{-\theta(q-jp)} (q-jp)^m\right]^{-1} \left[\sum_{j=0}^{n-m} c_j e^{-\theta(q-jp)} (q-jp)^{m+1}\right] \tag{21}$$

Equating the above equation to zero and solving it numerically for θ the Bayes estimator can be obtained.

2.2 Uniform Prior On θ

2.2.1 Bayesian Estimation Under the Linear Loss Function

Considering the uniform prior distribution

$$g_2(\theta) = \frac{1}{b-a} \qquad , a < \theta < b \tag{22}$$

Then the Bayes estimator $\hat{\theta}_{\gamma_2}$ is the posterior γ-th quantile of θ . Hence

$$\int_a^{\hat{\theta}_{\gamma_2}} g_2(\theta/T)\, d\theta = \gamma \tag{23}$$

where $g_2(\theta/T)$ the posterior density function .It can be obtained in parallel with the procedures used in Subsection (2.1) . Hence

$$g_2(\theta/T) = \frac{\sum_{j=0}^{n-m} c_j \theta^m \, e^{-\theta[q-jp]} [q-jp]^{m+1}}{k_1 - k_2} \tag{24}$$

where

$$k_1 = \sum_{j=0}^{n-m} c_j IG[b_j, m+1]$$
$$k_2 = \sum_{j=0}^{n-m} c_j IG[a_j, m+1] \tag{25}$$

From equations (23) and (24) one obtains

$$\sum_{j=0}^{n-m} c_j \left[IG(y_{j,\gamma_2}, m+1) - IG(y_{j,a}, m+1) \right] = (k_1 - k_2)\gamma \tag{26}$$

where $y_{j,\gamma 2} = \hat{\theta}_{\gamma 2}(q-jp)$ and $y_{j,a} = a(q-jp)$. The values y_{j,γ_2} can be obtained from solving (26) numerically and $\hat{\theta}_{\gamma_2}$ can be evaluated as

$$\hat{\theta}_{\gamma_2} = \frac{y_{j,\gamma_2}}{q-jp} \tag{27}$$

If $\gamma = 0.5$ in equation (23) then the Bayes estimator $\hat{\theta}_{0.5}$ of θ is the median of the posterior density function $g_2(\theta/T)$ .Then $\hat{\theta}_{0.5}$ can be evaluated as

$$\hat{\theta}_{0.5} = \frac{y_{j,0.5}}{q-jp} \tag{28}$$

where $\hat{\theta}_{1-\gamma}$ is obtained from equation (27) with γ replaced by 1-γ .

2.2.2 Bayesian Estimation Under Zero-One Loss Function

Using the informative uniform prior distribution $g_2(\theta)$ defined by (22), the Bayes estimator is the mode of the posterior density function $g_2(\theta/T)$ given in equation (24).Taking logarithms of both sides of equation (24) yields the following equation

$$\ln g_2(\theta/T) = m \ln \theta + \ln\left(\sum_{j=0}^{n-m} c_j e^{-\theta(q-jp)} (q-jp)^{m+1}\right) - \ln(k_1 - k_2) \tag{29}$$

where k_1 *and* k_2 are defined by equation (25). The first derivative of $\ln g_2(\theta/T)$ with respect to θ is

$$\frac{d \ln g_2(\theta/T)}{d\theta} = \frac{m}{\theta} - \left[\sum_{j=0}^{n-m} c_j e^{-\theta(q-jp)} (q-jp)^{m+1} \right]^{-1} \left[\sum_{j=0}^{n-m} c_j e^{-\theta(q-jp)} (q-jp)^{m+2} \right] \tag{30}$$

Equating the above equation to zero and solving it numerically for θ, the Bayes estimator can be obtained.

2.3 Natural Conjugate Prior On θ

2.3.1 Bayesian Estimation Under the Linear Loss Function

Under the assumption that the parameter α is known, the natural conjugate prior for θ is a gamma distribution (as considered by Nassar and Eissa (2004)) with density function

$$g_3(\theta)=\frac{\delta^{\upsilon}}{\Gamma(\upsilon)}\theta^{\upsilon-1}e^{-\delta\theta} \quad , \quad \theta>0,\upsilon>0,\delta>0 \tag{31}$$

Hence the Bayes estimator $\hat{\theta}_{\gamma_3}$ is the posterior γ-th quantile which can be obtained by solving the following equation

$$\int_0^{\hat{\theta}_{\gamma_3}} g_3(\theta/T)\,d\theta=\gamma \tag{32}$$

Applying Bayes theorem, one can obtain the posterior density of θ as

$$g_3(\theta/T)=\frac{\sum_{j=0}^{n-m} c_j\,\theta^{m+\upsilon-1}\,e^{-\theta[q+\delta-jp]}\left[q+\delta-jp\right]^{m+\upsilon}}{\sum_{j=0}^{n-m} c_j\Gamma(m+\upsilon)} \tag{33}$$

Using equations (32) and (33), one obtains

$$\sum_{j=0}^{n-m} c_j IG\left[z_{j,\gamma_3},(m+\upsilon)\right]=\gamma\sum_{j=0}^{n-m} c_j\Gamma(m+\upsilon) \tag{34}$$

where $z_{j,\gamma_3}=\hat{\theta}_{\gamma_3}(q+\delta-jp)$ The values of z_{j,γ_3} can be obtained numerically and the Bayes estimator

$$\hat{\theta}_{\gamma_3}=\frac{z_{j,\gamma_3}}{(q+\delta-jp)} \tag{35}$$

If $\gamma=0.5$ in equation (32) then the Bayes estimator $\hat{\theta}_{0.5}$ is the median of the posterior density function $g_3(\theta/T)$ which can be evaluated numerically from equation (34) by putting $\gamma=0.5$.

2.3.2 Bayesian Estimation Under the Zero-One Loss Function

Using the natural conjugate prior distribution $g_3(\theta)$ defined by (31), the Bayes estimator is the mode of the posterior density function $g_3(\theta/T)$ given in equation (33).Taking logarithms of both sides of equation (33) yields the following equation

$$\ln g_3(\theta/T) = (m+\upsilon-1)\ln\theta + \ln(\sum_{j=0}^{n-m} c_j e^{-\theta(q+\delta-jp)}(q+\delta-jp)^{m+\upsilon}) - \ln\sum_{j=0}^{n-m} c_j \Gamma(m+\upsilon) \tag{36}$$

The first derivative of $\ln g_3(\theta/T)$ with respect to θ is

$$\begin{aligned}\frac{d\ln g_3(\theta/T)}{d\theta} &= \frac{m+\upsilon-1}{\theta} - \left[\sum_{j=0}^{n-m} c_j e^{-\theta(q+\delta+jp)}(q+\delta-jp)^{m+\upsilon}\right]^{-1} \\ &\quad * \sum_{j=0}^{n-m} c_j (q+\delta-jp) e^{-\theta(q+\delta-jp)}(q+\delta-jp)^{m+\upsilon+1}\end{aligned} \tag{37}$$

for θ to obtain the Bayes estimator. Note that when *m*=*n* all the results obtained for type II censoring will approach those of the complete sample.

REFERENCES

Box, G. E. P and Tiao, G. C. (1973): Bayesian Inference in Statistical Analysis, Addison-Wesley.

Mudholkar, G. S. and Srivastava, D. K. (1993): Exponentiated Weibull family for analyzing bathtub failure-rate data. IEEE Transactions on Reliability, 42, (2), 299-302.

Mudholkar, G. S., Srivastava, D. K. and Freimer, M. (1995): The Exponentiated Weibull family: A reanalysis of the bus-motor-failure data. Technometrics, 37, (4), 436-445.

Mudholkar, G. S. and Huisen, A. D. (1996): The Exponentiated Weibull family: Some properties and a flood data application. Commun. Statist.-Theory Meth., 25(12), 3059-3083.

Nassar, M. M. and Eissa, F. H.(2003): On the Exponentiated Weibull Distribution. Commun. Statist.-Theory Meth., 32(7), 1317- 1336.

Nassar, M. M. and Eissa, F. H.(2004): Bayesian Estimation for the Exponentiated Weibull Model. Commun. Statist.-Theory Meth., 33(10), 2343-2362.

Singh, U., Gupta, P. K. and Upadhyay, S. K.(2002): Estimation of Exponentiated Weibull shape parameters under linex loss function. Commun. Statist.-Theory Meth., 31(4),523-537.

Singh, U., Gupta, P. K. and Upadhyay, S. K.(2004): Estimation for the three parameter Exponentiated Weibull under type-II censoring. Journal of Statistical Planning and Inference 134, 350-372.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 357–366.

GENERALIZED MULTI-PHASE RATIO ESTIMATORS USING MULTI-AUXILIARY VARIABLES

Zahoor Ahmad[1], Muhammad Hanif[2] and Munir Ahmad[3]
[1]University of Gujrat, Gujrat.
E-mail: zahoor_ahmed_stat@yahoo.com
[2]Lahore University of Management Sciences (LUMS), Lahore.
E-mail: hanif@lums.edu.pk
[3]National College of Business Administration and Economics, Lahore.
E-mail: drmunir@brain.net.pk

ABSTRACT

In this paper we propose a pair of generalized multi-phase ratio estimators using multi-auxiliary variables for estimating population mean of a study variable. The expressions for mean square error are also derived for suggested estimators and empirical comparisons have also been made.

Keywords: Multi-Phase Ratio Estimator, Multi-Phase Sampling, Multi-Auxiliary Variables

1. INTRODUCTION

The estimation of the population mean is an unrelenting issue in sampling theory and several efforts have been made to improve the precision of the estimates in the presence of multi-auxiliary variables. A variety of estimators have been proposed following different ideas like ratio, product or regression estimators.

Olkin (1958) was the first author to deal with the problem of estimating the mean of a survey variable when auxiliary variables are made available. He suggested the use of information on more than one auxiliary variable, positively correlated with the study variable analogously to Olkin; Singh (1967a) gave a multivariate expression of Murthy's (1964) product estimator, while Raj (1965) suggested a method for using multi-auxiliary variables through a linear combination of single difference estimators. Moreover, Singh (1967b) considered the extension of the ratio-cum-product estimators to multi-auxiliary variables Shukla (1965) suggested a multiple regression estimator while Rao and Mudholkar (1967) proposed a multivariate estimator based on a weighted sum of single ratio and product estimators.

John (1969) suggested two multivariate generalizations of ratio and product estimators which actually reduce to the Olkin's (1958) and Singh's (1967a) estimators. Srivastava (1971) proposed a general ratio-type estimator which generates a large class of estimators including most of the estimators up to that time proposed.

Ceccon and Diana (1996) provided a multivariate extension of the Naik and Gupta (1991) univariate class of estimators. Agarwal, et al. (1997), moving from Raj (1965), illustrated a new approach to form a multivariate difference estimator which does not require the knowledge of any population parameters. Pradhan (2005) suggested a chain regression estimator for two-phase sampling using three auxiliary variables when the population mean of one auxiliary variable is unknown and other auxiliary population means are known.

In practical surveys, the problem is to estimate population means of variables of interest. For example, in a typical socio-economic survey conducted in rural areas in Indo-Pak subcontinent, the multiple variables of interests may be size of household, monthly income and expenditure of the household, number of unemployed persons, number of illiterates, number of persons engaged in agriculture, amount of land owned, leased and leased out, number of cattle owned etc. In some situations the auxiliary information may be available through the past census data or conveniently collected. For example in a village land survey, the information on the variables such as area of the village, cultivable area, grazing grounds etc. may be easily obtained through the past census data and may be used to estimate the means of variables of interest.

Ahmad (2008) proposed univariate and multivariate pair-wise ratio estimators for estimating population mean and mean vector in the presence of multi-auxiliary variables for multi-phase sampling. Moeen (2009) suggested a multi-phase ratio estimator using consecutive phases in the presence of single auxiliary variable. He (2009) further investigated that for a single auxiliary variable, consecutive phases are not suitable for the case of single auxiliary variable than pair-wise phases as were suggested by Ahmad (2008). Obviously if we select a large sample for a variable for first phase and then small samples for the remaining successive phase there will be no contribution to the efficiency of the estimator because all the successive sample are the subset of first sample. Motivated from this investigation we propose a pair of multi-phase ratio estimators using multi-auxiliary variables in such a way that the sample will be selected for each variable for two-phases and order of the use of variable will depend upon the cost of the units of that variable. We suggested these two estimators also in such a way that their comparison will help us to investigate that what is the role of no of phases and no of auxiliary variables in the construction of estimator for multi-phase sampling using multi-auxiliary variables.

Before suggesting the estimators we provide multi-phase sampling scheme and some useful notations and results in the following section.

2. MULTI-PHASE SAMPLING USING MULTI-AUXILIARY VARIABLES

Consider a population of N units. Let Y be the variable for which we want to estimate the population mean and $X_1, X_2, \ldots, X_q$ are q auxiliary variables. For two-phase sampling design let $n_1, n_2, \cdots n_k$ ($n_1 > n_2 > \ldots > n_{k-1} > n_k$) sample sizes for k phases, respectively. Let $X_{(1)j}, X_{(2)j}, \ldots, X_{(k)j}$ $(j = 1, 2, \cdots, q)$ denotes the j^{th} auxiliary variables form for k phases respectively. and y_k denote the variable of interest from k^{th} phase. $\bar{X}_j$ and C_{x_j} denotes the population means and coefficient of variation of j^{th} auxiliary variable respectively and ρ_{yx_j} denotes the population correlation coefficient of Y and X_j. Further let $\theta_i = \frac{1}{n_i} - \frac{1}{N}$, $(i = 1, 2, \cdots, k)$. Further let $y_{(k)} = Y + e_{y_{(k)}}$, $x_{(i)j} = X_j + e_{x_{(i)j}}$; where $e_{y_{(k)}}$ and $e_{x_{(i)j}}$ are sampling errors and are of very small quantities. We assume that $E_2\left(e_{y_{(k)}}\right) = E_1\left(e_{x_{(i)j}}\right) = 0$. Then for simple random sampling without replacement we can write:

$$\mathrm{E}\left(\bar{\mathrm{e}}_{\mathrm{y_k}}\right)^2 = \theta_k \bar{Y}^2 C_y^2 \,, \mathrm{E}\left(\bar{\mathrm{e}}_{x_{\mathrm{(i)j}}}\right)^2 = \theta_i \bar{X}_j^2 C_{x_j}^2 \,,\ E\left(\bar{\mathrm{e}}_{\mathrm{y_k}} \bar{\mathrm{e}}_{x_{\mathrm{(i)j}}}\right) = \theta_i \overline{YX}_j C_y C_{x_j} \rho_{yx_j} \,,$$

$$E\left(\bar{e}_{y_k}\left(\bar{e}_{x_{(i)j}}-\bar{e}_{x_{(i+1)j}}\right)\right)=\left(\theta_i-\theta_{i+1}\right)\overline{Y}\overline{X}_j C_y C_{x_j}\rho_{yx_j},$$

$$E\left(\bar{e}_{x_{(i+1)j}}\left(\bar{e}_{x_{(i)j}}-\bar{e}_{x_{(i+1)j}}\right)\right)=E\left(\bar{e}_{x_{(i)j}}\bar{e}_{x_{(i+1)j}}\right)-E\left(\bar{e}_{x_{(i+1)j}}\right)^2=\left(\theta_i-\theta_{i+1}\right)\overline{X}_j^2 C_{x_j}^2,$$

$$E\left(\bar{e}_{x_{(i)j}}-\bar{e}_{x_{(i+1)j}}\right)^2=E\left(\bar{e}_{x_{(i)j}}\right)^2-E\left(\bar{e}_{x_{(i)j}}\bar{e}_{x_{(i+1)j}}\right)+E\left(\bar{e}_{x_{(i+1)j}}\right)^2$$

$$=\left(\theta_i-2\theta_i+\theta_{i+1}\right)\overline{X}_i^2 C_{x_i}^2=\left(\theta_{i+1}-\theta_i\right)\overline{X}_j^2 C_{x_j}^2,$$

$$E\left(\bar{e}_{x_{(i)j}}\left(\bar{e}_{x_{(i)j}}-\bar{e}_{x_{(i+1)j}}\right)\right)=E\left(\bar{e}_{x_{(i)j}}\right)^2-E\left(\bar{e}_{x_{(i)j}}\bar{e}_{x_{(i+1)j}}\right)=\left(\theta_i-\theta_i\right)\overline{X}_j^2 C_{x_j}^2=0$$

and

$$E\left(\left(\bar{e}_{x_{(i)j}}-\bar{e}_{x_{(i')j}}\right)\left(\bar{e}_{x_{(i+1)j'}}-\bar{e}_{x_{(i'+1)j'}}\right)\right)$$

$$=E\left(\bar{e}_{x_{(i)j}}\bar{e}_{x_{(i+1)j'}}\right)-E\left(\bar{e}_{x_{(i)j}}\bar{e}_{x_{(i'+1)j'}}\right)-E\left(\bar{e}_{x_{(i')j}}\bar{e}_{x_{(i+1)j'}}\right)+E\left(\bar{e}_{x_{(i')j}}\bar{e}_{x_{(i'+1)j'}}\right)$$

$$=\left(\theta_i-\theta_i-\theta_{i'}+\theta_{i'}\right)\overline{X}_j\overline{X}_{j'}C_{x_j}C_{x_{j'}}\rho_{x_jx_{j'}}=0\ ;\ \left(i<i'\right)\left(i=1,3,5,...,k-1;i'=2,4,6,...,k\right)$$

3. GENERALIZED MULTI-PHASE RATIO ESTIMATORS

In the following sections we propose two multi-phase ratio estimators and derive their expression of mean square error

3.1 Generalized Multi-Phase Ratio Estimator-I

We propose the following generalized multi-phase ratio estimator for estimating population mean in the presence of multi-auxiliary variables. The auxiliary variables utilized in a way that the data on most expensive variable will be collected in the last phase.

$$t_1=\bar{y}_k\left(\frac{\bar{x}_{(1)1}}{\bar{x}_{(2)1}}\right)^{\alpha_1}\left(\frac{\bar{x}_{(3)2}}{\bar{x}_{(4)2}}\right)^{\alpha_2}\cdots\left(\frac{\bar{x}_{(k-1)q}}{\bar{x}_{(k)q}}\right)^{\alpha_q} \tag{3.1}$$

or

$$t_1=\bar{y}_k\prod_{i=1,i'=2,j=1}^{k-1,k,q}\left(\frac{\bar{x}_{(i)j}}{\bar{x}_{(i')j}}\right)^{\alpha_j},\ \left(i=1,3,5,...,k-1;i'=2,4,6,...,k\right),\ \left(j=1,2,\cdots q\right)$$

Using the substitutions and $\bar{y}_{(k)}=\overline{Y}+\bar{e}_{y_{(k)}}$, $\bar{x}_{(i)j}=\overline{X}_j+\bar{e}_{x_{(i)j}}$ we get

$$t_1=\left(\overline{Y}+\bar{e}_{y_k}\right)\prod_{i=1,i'=2,j=1}^{k-1,k,q}\left(\frac{\overline{X}_j+\bar{e}_{\bar{x}_{(i)j}}}{\overline{X}_j+\bar{e}_{\bar{x}_{(i')j}}}\right)^{\alpha_j}$$

or

$$t_1 = \left(\bar{Y} + \bar{e}_{y_k}\right) \prod_{i=1, i^{'}=2, j=1}^{k-1,k,q} \left(1 + \frac{\bar{e}_{\bar{x}_{(i)j}}}{\bar{X}_j}\right)^{\alpha_j} \left(1 + \frac{\bar{e}_{\bar{x}_{\left(i^{'}\right)j}}}{\bar{X}_j}\right)^{-\alpha_j}$$

Ignoring the second and higher terms

$$t_1 = \left(\bar{Y} + \bar{e}_{y_k}\right) \prod_{i=1, i^{'}=2, j=1}^{k-1,k,q} \left(1 + \alpha_j \frac{\bar{e}_{\bar{x}_{(i)j}}}{\bar{X}_j}\right)\left(1 - \alpha_j \frac{\bar{e}_{\bar{x}_{\left(i^{'}\right)j}}}{\bar{X}_j}\right)$$

Again ignoring the second and higher terms

$$t_1 = \left(\bar{Y} + \bar{e}_{y_k}\right) \prod_{i=1, i^{'}=2, j=1}^{k-1,k,q} \left\{1 + \alpha_j \frac{1}{\bar{X}_j}\left(\bar{e}_{\bar{x}_{(i)j}} - \bar{e}_{\bar{x}_{\left(i^{'}\right)j}}\right)\right\}$$

or

$$t_1 = \left(\bar{Y} + \bar{e}_{y_k}\right) \left\{1 + \sum_{i=1, i^{'}=2, j=1}^{k-1,k,q} \alpha_j \frac{1}{\bar{X}_j}\left(\bar{e}_{\bar{x}_{(i)j}} - \bar{e}_{\bar{x}_{\left(i^{'}\right)j}}\right)\right\}$$

or

$$t_1 = \bar{Y} + \bar{e}_{y_k} + \sum_{i=1, i^{'}=2, j=1}^{k-1,k,q} \frac{\bar{Y}}{\bar{X}_j} \alpha_j \left(\bar{e}_{\bar{x}_{(i)j}} - \bar{e}_{\bar{x}_{\left(i^{'}\right)j}}\right)$$

or

$$t_1 = \bar{Y} + \bar{e}_{y_k} + \bar{Y} d^{'}_{1\times q} a_{q\times 1} \tag{3.2}$$

where

$a^{'}_{1\times q} = \begin{bmatrix} \frac{\alpha_1}{\bar{X}_1} & \frac{\alpha_2}{\bar{X}_2} & \cdots & \frac{\alpha_q}{\bar{X}_q} \end{bmatrix}$ and $d^{'}_{1\times q} = \left[\left(\bar{e}_{\bar{x}_{(1)1}} - \bar{e}_{\bar{x}_{(2)1}}\right) \quad \left(\bar{e}_{\bar{x}_{(3)2}} - \bar{e}_{\bar{x}_{(4)2}}\right) \quad \cdots \quad \left(\bar{e}_{\bar{x}_{(k-1)q}} - \bar{e}_{\bar{x}_{(k)q}}\right)\right]$

From (3.2), we can write

$$MSE\left(t_1\right) = E\left[\bar{e}_{y_k} + \bar{Y} d^{'}_{1\times q} a_{q\times 1}\right]^2 \tag{3.3}$$

Differentiating (3.3) w.r.t $a_{q\times 1}$, we get:

$$2E\left[\bar{e}_{y_k} + \bar{Y} d^{'}_{1\times q} a_{q\times 1}\right] \bar{Y} d^{'}_{1\times q} = 0$$

or

$$E\left[d_{q\times 1} \bar{e}_{y_k} + \bar{Y} d_{q\times 1} d^{'}_{1\times q} a_{q\times 1}\right] = 0$$

or

$$E\left(d_{q\times 1} \bar{e}_{y_k}\right) + \bar{Y} E\left(d_{q\times 1} d^{'}_{1\times q}\right) a_{q\times 1} = 0$$

$$S^{*}_{yx} + \bar{Y} S^{*}_{xx} a_{q\times 1} = 0$$

The expressions of S^{*}_{yx} and S^{*}_{xx} are given in Appendix.

$$a_{q\times 1}=\frac{1}{\bar{Y}}\left(S_{xx}^{*}\right)^{-1}S_{yx}^{*} \tag{3.4}$$

After simplification we can write (3.4) as:

$$\alpha_j=\frac{C_y}{C_{x_j}}\rho_{yx_j}\ ,\ \left(j=1,2,\cdots q\right)$$

Now (3.3) can be written as:

$$MSE\left(t_1\right)=E\left[\bar{e}_{y_k}\left(\bar{e}_{y_k}+\bar{Y}d'_{1\times q}a_{q\times 1}\right)\right]$$

$$=E\left(\bar{e}_{y_k}\right)^2+\bar{Y}E\left(\bar{e}_{y_k}d'_{1\times q}\right)a_{q\times 1}$$

$$MSE\left(t_1\right)=\theta_k\bar{Y}^2C_y^2+\bar{Y}S_{yx}^{*'}\left(S_{xx}^{*}\right)^{-1}S_{yx}^{*}$$

Putting the values of S_{yx}^{*} and S_{xx}^{*} from Appendix-B, mean square error can be simplified as:

$$MSE\left(t_1\right)=\bar{Y}^2C_y^2\left[\theta_k+\left(\theta_1-\theta_2\right)\rho_{yx_1}^2+\left(\theta_3-\theta_4\right)\rho_{yx_2}^2+\cdots+\left(\theta_{k-1}-\theta_k\right)\rho_{yx_q}^2\right]$$

or

$$MSE\left(t_1\right)=\bar{Y}^2C_y^2\left[\theta_k-\left(\theta_2-\theta_1\right)\rho_{yx_1}^2-\left(\theta_4-\theta_3\right)\rho_{yx_2}^2-\cdots-\left(\theta_k-\theta_{k-1}\right)\rho_{yx_q}^2\right]$$

3.2 Generalized Multi-Phase Ratio Estimator-II

We propose the second generalized multi-phase ratio estimator for estimating population mean in the presence of multi-auxiliary variables. The auxiliary variables utilized in a way that the data on most expensive variable will be collected in the last phase.

$$t_2=\bar{y}_l\left(\frac{\bar{x}_{(1)1}}{\bar{x}_{(2)1}}\right)^{\alpha_1}\left(\frac{\bar{x}_{(2)2}}{\bar{x}_{(3)2}}\right)^{\alpha_2}\cdots\left(\frac{\bar{x}_{(l-2)q-1}}{\bar{x}_{(l-1)q-1}}\right)^{\alpha_{q-1}}\left(\frac{\bar{x}_{(l-1)q}}{\bar{x}_{(l)q}}\right)^{\alpha_q} \tag{3.5}$$

or

$$t_2=\bar{y}_l\prod_{i=2,\,j=1}^{l,q}\left(\frac{\bar{x}_{(i-1)j}}{\bar{x}_{(i)j}}\right)^{\alpha_j}\ ,\left(i=1,2,\ldots,l\right)\ \text{and}\ \left(j=1,2,\cdots q\right)$$

Using the substitutions and $\bar{y}_{(k)}=\bar{Y}+\bar{e}_{y_{(k)}}$, $\bar{x}_{(i)j}=\bar{X}_j+\bar{e}_{x_{(i)j}}$ we get:

$$t_2=\left(\bar{Y}+\bar{e}_{y_l}\right)\prod_{i=2,\,j=1}^{l,q}\left(1+\frac{\bar{e}_{x_{(i-1)j}}}{\bar{X}_j}\right)^{\alpha_j}\Bigg/\left(1+\frac{\bar{e}_{x_{(i)j}}}{\bar{X}_j}\right)^{\alpha_j}$$

or

$$t_2 = \left(\bar{Y} + \bar{e}_{y_l}\right) \prod_{i=2,\, j=1}^{l,q} \left(1 + \frac{\bar{e}_{x_{(i-1)j}}}{\bar{X}_j}\right)^{\alpha_j} \left(1 + \frac{\bar{e}_{x_{(i)j}}}{\bar{X}_j}\right)^{-\alpha_j}$$

Ignoring the second and higher order terms, we can write

$$t_2 = \left(\bar{Y} + \bar{e}_{y_l}\right) \prod_{i=2,\, j=1}^{l,q} \left(1 + \alpha_j \frac{\bar{e}_{x_{(i-1)j}}}{\bar{X}_j}\right)\left(1 - \alpha_j \frac{\bar{e}_{x_{(i)j}}}{\bar{X}_j}\right)$$

or

$$t_2 = \left(\bar{Y} + \bar{e}_{y_l}\right) \prod_{i=2,\, j=1}^{l,q} \left(1 + \alpha_j \frac{1}{\bar{X}_j}\left(\bar{e}_{x_{(i-1)j}} - \bar{e}_{x_{(i)j}}\right)\right)$$

or

$$t_2 = \bar{Y} + \bar{e}_{y_l} + \sum_{i=2,\, j=1}^{l,q} \alpha_j \frac{\bar{Y}}{\bar{X}_j}\left(\bar{e}_{x_{(i-1)j}} - \bar{e}_{x_{(i)j}}\right)$$

or

$$t_2 = \bar{Y} + \bar{e}_{y_l} + \bar{Y} d'_{1\times q} a_{q\times 1}, \qquad (3.6)$$

where

$$a'_{1\times q} = \begin{bmatrix} \frac{\alpha_1}{\bar{X}_1} & \frac{\alpha_2}{\bar{X}_2} & \cdots & \frac{\alpha_q}{\bar{X}_q} \end{bmatrix} \text{ and } d'_{1\times q} = \left[\left(\bar{e}_{\bar{x}_{(1)1}} - \bar{e}_{\bar{x}_{(2)1}}\right) \quad \left(\bar{e}_{\bar{x}_{(2)2}} - \bar{e}_{\bar{x}_{(3)2}}\right) \quad \cdots \quad \left(\bar{e}_{\bar{x}_{(l-1)q}} - \bar{e}_{\bar{x}_{(l)q}}\right)\right].$$

From (3.6) we can write

$$MSE\left(t_2\right) = E\left[\bar{e}_{y_l} + \bar{Y} d'_{1\times q} a_{q\times 1}\right]^2 \qquad (3.7)$$

Differentiating (3.7) w.r.t $a_{q\times 1}$, we get:

$$2E\left[\bar{e}_{y_l} + \bar{Y} d'_{1\times q} a_{q\times 1}\right] \bar{Y} d'_{1\times q} = 0$$

or

$$E\left[d_{q\times 1}\bar{e}_{y_l} + \bar{Y} d_{q\times 1} d'_{1\times q} a_{q\times 1}\right] = 0$$

or

$$E\left(d_{q\times 1}\bar{e}_{y_l}\right) + \bar{Y} E\left(d_{q\times 1} d'_{1\times q}\right) a_{q\times 1} = 0$$

or

$$S^{**}_{yx} + \bar{Y} S^{**}_{xx} a_{q\times 1} = 0$$

The expressions of S^{**}_{yx} and S^{**}_{xx} are given in Appendix-B.

$$a_{q\times 1}=\frac{1}{\bar{\bar{Y}}}\left(S_{xx}^{**}\right)^{-1}S_{yx}^{**} \tag{3.8}$$

After simplification we can write

$$\alpha_j=\frac{C_y}{C_{x_j}}\rho_{yx_j}\ ,\ \left(j=1,2,\cdots q\right)$$

Now (3.7) can be written as:

$$MSE\left(t_2\right)=E\left[\bar{e}_{y_l}\left(\bar{e}_{y_l}+\bar{Y}d'_{1\times q}a_{q\times 1}\right)\right]$$

$$=E\left(\bar{e}_{y_l}\right)^2+\bar{Y}E\left(\bar{e}_{y_l}d'_{1\times q}\right)a_{q\times 1}$$

$$MSE\left(t_1\right)=\theta_l\bar{Y}^2C_y^2+\bar{Y}S_{yx}^{**'}\left(S_{xx}^{**}\right)^{-1}S_{yx}^{**}$$

Putting the values of S_{yx}^{**} and S_{xx}^{**} from Appendix-B, mean square error can be simplified as:

$$MSE\left(t_2\right)=\bar{Y}^2C_y^2\left[\theta_l+\left(\theta_1-\theta_2\right)\rho_{yx_1}^2+\left(\theta_2-\theta_3\right)\rho_{yx_2}^2+\cdots+\left(\theta_{l-1}-\theta_l\right)\rho_{yx_q}^2\right]\text{or}$$

$$MSE\left(t_2\right)=\bar{Y}^2C_y^2\left[\theta_l-\left(\theta_2-\theta_1\right)\rho_{yx_1}^2-\left(\theta_3-\theta_2\right)\rho_{yx_2}^2-\cdots-\left(\theta_l-\theta_{l-1}\right)\rho_{yx_q}^2\right]$$

4. EMPIRICAL COMPARISON

Description of populations, detail of variables, necessary parameters for calculating MSE'S and MSE's are given in Appendix-A. As we suggested two estimators, we can made comparisons under following two cases:

1. The comparison with same no of auxiliary variables and different no of phases.
2. The comparison with same no of phases and different no of auxiliary variables.

For first case we consider three auxiliary variables and six phases for first and four for second estimator as permissible by the estimators. Table A-5 contains the mean square error of suggested estimators for this case. In this comparison first estimator is less efficient then second because as we increase phase the mean square increases but cost decrease. By increasing the phases the sample size decreases which only leads to increase sampling fraction not the information.

For second case we consider four phases for both and two auxiliary variables for first and three for second estimator as permissible by the estimators. Table A-5 contains the mean square error of suggested estimators for this case. In this comparison again first estimator is less efficient then second because as we increase auxiliary variables the mean square increases.

In both situations the second estimator performs better than first. For same no of phases it uses greater no of auxiliary variables than first and for same no of auxiliary variables it uses less no of phases than first. This paper suggested that we have to use that estimator which uses less no of phases and greater no of auxiliary variables.

APPENDIX A

Table A-1: Detail of Populations

Sr. #	**Source of Populations**
1	Population census report of Jhang district (1998), Pakistan
2	Population census report of Faisalabad district (1998), Pakistan.
3	Population census report of Gujrat district (1998), Pakistan.
4	Population census report of Kasur (1998) Pakistan
5	Population census report of Sialkot district (1998), Pakistan.

Table A-2: Description of variables (Each variable is taken from Rural Locality)

Description of variables			
Y_2	Population of currently married	X_2	Population of primary but below matric
X_1	Population of both sexes	X_3	Population of matric and above

Table A-3: Population and Sample Sizes

Districts	N	n_1	n_2	n_3	n_4	n_5	n_6
Jhang	368	184	92	46	23	12	6
Faisalabad	283	142	71	35	18	9	5
Gujrat	204	102	51	26	13	6	3
Kasur	181	91	45	23	11	6	3
Sailkot	269	135	67	34	17	8	4

Table A-4: Parameters of populations for calculating MSE's of estimators

Districts	$\bar{Y}$	C_y	$\rho_{y_2x_1}$	$\rho_{y_2x_2}$	$\rho_{y_2x_3}$
Jhang	860.11	0.595	.428	.912	.659
Faisalabad	1511.260	0.522	.943	.927	.599
Gujrat	1101.280	0.484	.995	.941	.764
Kasur	1393.200	0.551	.998	.758	.879
Sailkot	1058.740	0.647	.999	.983	.931

Table A-5:- MSE's of suggested estimators

	Same no of variables (6) and different no of phases		Same no of phases (4) and different no of variables (2,3)	
Districts	$MSE(t_1)$	$MSE(t_2)$	$MSE(t_1)$	$MSE(t_2)$
Jhang	22767.02	3719.97	4876.41	3719.97
Faisalabad	69343.69	9813.10	12658.89	9813.10
Gujrat	44083.52	4301.86	7408.46	4301.86
Kasur	79047.15	9646.11	22522.28	9646.11
Sailkot	43868.02	2802.90	8767.40	2802.90

APPENDIX B

$$a'_{q\times 1} = \begin{bmatrix} \frac{\alpha_1}{\bar{X}_1} & \frac{\alpha_2}{\bar{X}_2} & \cdots & \frac{\alpha_q}{\bar{X}_q} \end{bmatrix}$$

$$S_{yx}^{*\prime} = \begin{bmatrix} (\theta_1-\theta_2)S_{yx_1} & (\theta_3-\theta_4)S_{yx_2} & \cdots & (\theta_{k-1}-\theta_k)S_{yx_q} \end{bmatrix}$$

$$S_{xx}^{*} = \begin{bmatrix} (\theta_1-\theta_2)S_{x_1x_1} & 0 & \cdots & 0 \\ 0 & (\theta_3-\theta_4)S_{x_2x_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & (\theta_{k-1}-\theta_k)S_{x_qx_q} \end{bmatrix}$$

$$S_{yx}^{**\prime} = \begin{bmatrix} (\theta_1-\theta_2)S_{yx_1} & (\theta_2-\theta_3)S_{yx_2} & \cdots & (\theta_{l-1}-\theta_l)S_{yx_q} \end{bmatrix}$$

$$S_{xx}^{**} = \begin{bmatrix} (\theta_1-\theta_2)S_{x_1x_1} & 0 & \cdots & 0 \\ 0 & (\theta_2-\theta_3)S_{x_2x_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & (\theta_{l-1}-\theta_l)S_{x_qx_q} \end{bmatrix}$$

REFERENCES

Abu-Dayyeh, W.A., Ahmed, M.S., Ahmed, R.A. and Muttlak, H.A. (2003). Some estimators of finite population mean using auxiliary information. *Applied Mathematics and Computation*, 139, 287-298.

Agarwal, S.K., Sharma, U.K. and Kashyap, S. (1997). A new approach to use multivariate auxiliary information in sample surveys. *J. Statist. Plann. and Infer.*, 60, 261-267.

Arora, S. and Lal. Bansi (1989). *New Mathematical Statistics*. Satya Prakashan, New Delhi.

Ceccon, C. and Diana, G. (1996). Classi di stimatori di tipo rapporto in presenza di pi´u variabili ausiliarie, *Proceedings of Italian Statistical Society*, XXXVIII Scientific Meeting, Rimini, 9-13 April 1996, 119-126.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. Champman and Hall, London.

John, S. (1969). On multivariate ratio and product estimators. *Biometrika*, 56, 533-536.

Kadilar, C. and Cingi, H. (2004). Estimator of a population mean using two auxiliary variables in simple random sampling. *International Mathematical Journal*, 5, 357-367.

Kadilar, C. and Cingi, H. (2005). A new estimator using two auxiliary variables, *Applied Mathematics and Computation*. 162, 901-908.

Lee, Carl M.S. and Peddada, S.D. (1987). Comparisons of some common ratio estimators using Pitman measure of nearess. *Comm. Stat.-Theo Math.*, 16(7), 2017-2027.

Murthy, M.N. (1964). Product method of estimation. *Sankhya*, 26, 294-307.

Naik, V.D. and Gupta, P.C. (1991). A general class of estimators for estimating population mean using auxiliary information. *Metrika*, 38, 11-17.

Olkin, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika*, 45, 154-165.

Perri, P.F. (2004). Alcune considerazioni sull'efficienza degli stimatori rapporto-cum-prodotto. *Statistica & Applicazioni*, 2(2), 59-75.

Perri, P.F. (2005). Combining two auxiliary variables in ratio-cum-product type estimators. *Proceedings of Italian Statistical Society, Intermediate Meeting on Statistics and Environment*, Messina, 21-23 September 2005, 193-196.

Pradhan, B.K. (2000). *Some problems of estimation in multi-phase sampling*. Ph.D. Dissertation (unpublished), Utkal University, Orissa, India.

Pradhan, B.K. (2005). A Chain Regression Estimator in Two Phase Sampling Using Multi-auxiliary Information. *Bull. Malays. Math. Sci. Soc.* (2) 28(1), 81-86.

Raj, D. (1965). On a method of using multi-auxiliary information in sample surveys. *J. Amer. Statist. Assoc.*, 60, 309, 270-277.

Rao, P.S.R.S. and Mudholkar, G.S. (1967). Generalized multivariate estimators for the mean of a finite population. *J. Amer. Statist. Assoc.*, 62, 1009-1012.

Robinson, P.M. (1994). A class of estimators for the mean of a finite population using auxiliary information. *Sankhya*, B, 56, 389-399.

Samiuddin, M. and Hanif, M. (2007). Estimation of population mean in single and two phase sampling with or without additional information. *Pak. J. Statist.*, 23(2), 99-118.

Shukla, G.K. (1965). Multivariate regression estimate. *J. Indian Statist. Assoc.* 3, 202-211.

Singh, M.P. (1965). On the estimation of ratio and product of population parameters. *Sankh*, Ser. C, 27, 321-328.

Singh, M.P. (1967a). Multivariate product method of estimation for finite populations. *J. Ind. Soc. of Agricul. Statist.*, 31, 375-378.

Singh, M.P. (1967b). Ratio cum product method of estimation. *Metrika*. 12, 34-43.

Srivastava, S.K. (1971). A generalized estimator for the mean of a finite population using multi-auxiliary information. *J. Amer. Statist. Assoc.*, 66, 404-407.

Tracy, D.S., Singh, H.P. and Singh, R. (1996). An alternative to the ratio-cum-product estimator in sample surveys. *J. Statist. Plann. and Infer.*, 53, 375-387.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 367–374.

CONSTRUCTION OF DESIGNS BALANCED FOR NEIGHBOR EFFECTS

Munir Akhtar[1], Rashid Ahmed[2] and Furrukh Shehzad[3]
[1] COMSATS Institute of Information Technology, Attock, Pakistan
E-mail: munir_stat@yahoo.com; dir-attock@comsats.edu.pk
[2] Department of Statistics The Islamia University of Bahawalpur, Pakistan
E-mail: rashid701@hotmail.com
[3] National College of Business Administration & Economics, Lahore, Pakistan
E-mail: fshehzad.stat@gmail.com

ABSTRACT

Experiments in agriculture, horticulture and forestry often show neighbor effects and neighbor-balanced designs ensure that treatment comparisons will be as little affected by neighbor effects as possible. In this article, neighbor-balanced designs are constructed using cyclic shifts. Some algorithms are developed to generate the minimal designs balanced for neighbor effects. Designs balanced up to second order neighbors and higher order neighbors are also constructed for many configurations.

Keywords: Neighbor effects; Neighbor balanced designs; Circular binary blocks, Cyclic shifts, Minimal neighbor designs.

1. INTRODUCTION

Rees (1967) introduced neighbor designs in serology. Neighbor design (Neighbor balanced design) is a collection of circular blocks in which any two treatments appear as neighbors equally often. In agro forestry intercropping experiments, there also arises the design problem. As trees are much taller than the crop, there is a neighbor effect through interplant competition (see Monod, 1992). Clearly, in this situation neighbor balance between the treatments must be looked for because neighbor balanced designs ensure that treatment comparisons will be as little affected by neighbor effects as possible. In industrial experiments where blocks correspond to a time period and treatments are applied sequentially to experimental units, different trends often occur within blocks as a result of different rates of aging or equipment wear out, called neighbor effects. Rees (1967) constructed neighbor designs in complete blocks for all odd v (number of treatments). He also presented the initial block(s) to generate the neighbor designs for odd v up to 41 with k (block size) $\leq$ 10. Hwang (1973) constructed some infinite classes of neighbor designs but most of these series are non-binary. Lawless (1971), Das and Saha (1976), and Dey and Chakravarty (1977) also constructed neighbor designs for different configurations. Azais *et al.* (1993) proposed several methods for constructing neighbor designs for $k = v$ and v-1. Iqbal et al. (2006) and Iqbal *et al*. (2009) constructed neighbor designs for first order and second order using cyclic shifts. They constructed second order neighbor designs for $3 \leq k \leq 7$. Ahmed and Akhtar (2008), Akhtar and Ahmed (2009) developed several algorithms to construct neighbor balanced designs and presented several new second and higher order neighbor designs in circular binary blocks using difference method. A block in which first and last units are considered neighbors is called circular block. Circular blocks are called circular binary blocks if no treatment appears more than once in such blocks. Ai *et al.* (2007) constructed all order neighbor balanced designs (ANBD) for $k \leq v$. Ai *et al.* (2007) also suggested that the circular neighbor balanced designs are universally optimal. Recently, optimality of circular neighbor-balanced designs for total effects with autoregressive correlated observations is discussed by Ai *et al.* (2009).

2. METHOD OF CYCLIC SHIFTS

Ahmed and Akhtar (2009) constructed ANBD using cyclic shifts for

(i) v =2m+1and $k = v$; v prime from the following m sets of shifts (Q_i), each of k-1 elements, in m circular blocks with λ'=1, where λ' is number of times, each pair of two distinct treatments appears as neighbors.

$$Q_i = [i, i, \ldots, i](1/v) : i = 1, 2, \ldots, m.$$

(ii) v =2m+1, v prime and k relatively prime to $v(v$-1)/2, from the following m sets of shifts, each of k-1 elements, in mv circular binary blocks with $\lambda'=k$.

$$Q_i = [i, i, \ldots, i] : i = 1, 2, \ldots, m.$$

Ahmed and Akhtar (2009) also developed a methodology to construct ℓ-order neighbor balanced designs using cyclic shifts, which is described below. A design is ℓth- order neighbor balanced if each of 1, 2, … , v-1 appears an equal number of times, say, λ' in a new set of shifts. New set of shifts consists of (i) the sum of every ℓ and (k-ℓ) successive shifts, (ii) the complements of the shifts in (i), where complement of shift q_i is v-q_i. If any set of shifts is with t then a design is ℓth-order neighbor balanced if each of 1, 2, … , v-2 appears an equal number of times, say, λ' in a new set of shifts. In this case, complement of shift q_i is v-1-q_i.

Method of cyclic shifts has edge over difference method due to (i) from the set(s) of shifts a design can be checked whether it is BIBD or not without studying the blocks of designs, (ii) from the set(s) of shifts a non- binary first order neighbor design can be converted into binary, and (iii) from the set(s) of shifts A- optimality of the designs can be measured. In this method a set of shifts with t contains k-2 elements, otherwise k-1 elements (for further details, see Iqbal *et al.*, 2009). The method to obtain the required design from the set(s) of shifts is described with the following examples.

Example 2.1: Set of shifts [2, 3, 8] provides the neighbor design for v = 9 and k = 4 in 9 blocks. Put 0, 1, 2, …, 8 as first entry each of nine blocks (for generation of design we take all or fraction of v blocks If no set of shifts is with t). To get the second entries add 2 mod 9 to each of its corresponding first entries. Similarly add 3 mod 9 to each of its corresponding second entries then 8 to each of third entries. Hence required design is:

B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9
0	1	2	3	4	5	6	7	8
2	3	4	5	6	7	8	0	1
5	6	7	8	0	1	2	3	4
4	5	6	7	8	0	1	2	3

Example 2.2: Set of shifts [1,8,3,6,5,4,7,2,9]t(1/2) provides the complete block neighbor designs for v = 11. Take (1/2) of the v-1 blocks (If any set of shifts is with t, for generation of design we take all or fraction of v-1 blocks) first entry of each block is 0, 1, 2, 3, 4. Second entry of each column will be obtained by adding 1 mod (v-1) to each of its corresponding first entry. Similarly to get third entries, add 8 mod 10 to each of its corresponding second entries and so on. In each column, vth entry will be 10.

$\mathbf{B_1}$	$\mathbf{B_2}$	$\mathbf{B_3}$	$\mathbf{B_4}$	$\mathbf{B_5}$
0	1	2	3	4
1	2	3	4	5
9	0	1	2	3
2	3	4	5	6
8	9	0	1	2
3	4	5	6	7
7	8	9	0	1
4	5	6	7	8
6	7	8	9	0
5	6	7	8	9
10	10	10	10	10

3. CONSTRUCTION OF MINIMAL DESIGNS BALANCED FOR NEIGHBOR EFFECTS

A design in which each pair of treatments appears only once as neighbor i.e. $\lambda' = 1$ is called minimal neighbor designs. In this section, minimal neighbor designs are constructed.

3.1 Complete Block Minimal Nearest Neighbor Balanced Designs (NNBD) for *v* Odd

Theorem 3.1: For v odd, following set of shifts S_1 provides complete block neighbor designs with $\lambda' = 1$ in (v-1)/2 circular blocks.

$$S_1 = [1, v\text{-}3, 3, v\text{-}5, 5, \ldots, v\text{-}(k\text{-}1), v\text{-}2]t(1/2)$$

Proof: New set of shifts S^*_1 = [1, v-3, 3, v-5, 5,…, v-(k-1), v-2, v-2, 2, v-4, 4, … , 1](1/2). Since S^*_1 contains all of 1, 2, 3, …, v-2 only once. Hence the theorem.

Example 3.1: [1,10,3,8,5,6,7,4,9,2,11]t(1/2) provides the complete block neighbor designs for v = 13.

3.2 Minimal NNBD for *v* = 2*mk*+1; *m* >1

Theorem 3.2: For v = 2mk+1; m >1 then m sets of shifts $S_i = [q_{i1}, q_{i2}, \ldots, q_{i(k-1)}]$; i = 1, 2, … , m, provide NNBD with $\lambda' = 1$ in mv circular blocks if $\bigcup_{i=1}^{m} S^*_i$ contains all the elements 1, 2, …, v-1 once, where $S^*_i = [q_{i1}, q_{i2}, \ldots, q_{i(k-1)}, q_{ik}, v\text{-}q_{i1}, v\text{-}q_{i2}, \ldots, v\text{-}q_{i(k-1)}, v\text{-}q_{ik}]$ and $q_{ik} = q_{i1}+q_{i2}+ \ldots +q_{i(k-1)}$ mod v.

Proof: New set of shifts S^*_i contains all of 1, 2, 3, …, v-1 only once. Hence the theorem.

Example 3.2: Following sets of shifts provides NNBD with $\lambda' = 1$ in 132 blocks for v = 33 and k = 4.

$$S_2 = [32, 2, 3] + [28, 6, 7] + [24, 10, 11] + [20, 14, 15] .$$

3.3 NNBD for *v* = 2*k*+1

Let $S_3 = [q_1, q_2, \dots, q_{k-1}]$, $q_i = i$ or $v-i$; i = 1, 2 ... , k-1, such that (i) sum of any two, three, . . ., (k-1) successive elements of S_3 is not zero mod v (If sum of any two, three, . . ., (k-1) successive elements of S_3 is zero mod v then rearrange the assigned elements of S_3), and (ii) $(q_1 + q_2 + \dots + q_{k-1})$ mod v = k or (k+1).

Example 3.3: [1,2,3,4,5,6,7,8,9,13] provides the neighbor designs for v = 23 and k =11 in 23 blocks.

3.4 Minimal Designs Balanced for Second Order Neighbor Effects

Following are the minimal designs balanced up to second order neighbors for $v = 2k+1$, $15 \le v \le 27$ in v circular blocks.

v	*k*	Set of shifts
15	7	[1,2,7,12,10,4]
17	8	[1,3,15,12,10,13,6]
19	9	[1,4,16,7,9,17,11,6]
21	10	[1,2,3,4,6,9,11,14,5]
23	11	[3,2,8,4,22,17,14,5,16,13]
25	12	[3,2,8,4,5,19,24,18,10,9,12]
27	13	[3,2,8,4,5,16,14,21,26,15,10,18]

Furthermore [1,5,6,11,17,9,7,16] provides third-order and [1,9,5,7,6,16,11,4] provides all-order neighbor design with $\lambda' = 1$ in 19 blocks for v = 19 and k = 9.

3.5 Minimal Designs Balanced for ℓ-Order Neighbor Effects

Theorem 3.5: If $k = 2\ell$ then ℓ-order neighbor balanced design never exists with $\lambda' = 1$.

Proof: If $k = 2\ell$ then ℓ-order right neighbor will be the ℓ-order left neighbor. Hence any pair of treatments (which are ℓ-order neighbors) appears at least twice as ℓ-order neighbors. Hence the theorem.

4. CONSTRUCTION OF DESIGNS BALANCED FOR NEIGHBOR EFFECTS WITH $\lambda' > 1$

In this section, neighbor balanced designs (NBD) are constructed with $\lambda' > 1$ using cyclic shifts.

4.1 Complete Block NNBD for $v = 4s$; s Is Integer

Theorem 4.1: For $v = 4s$, $m = 2s-1$, s is an integer, following set of shifts S_4 provides complete block neighbor designs with $\lambda' = 2$ in v-1 circular blocks if the sum of any two, three, … or $2m+1$ successive elements of S_4 is not 0 mod v-1.

$$S_4 = [1, 2, 3, \ldots, (m\text{-}1),\ m,\ m,\ (m\text{-}1), \ldots, 3, 2, 1]t$$

Proof: New set of shifts S^*_4 contains all of 1, 2, 3, …, v-2 twice. Hence the theorem.

Example 4.1: [1,2,3,4,5,6,7,8,9,9,8,7,6,5,4,3,2,1]t provides the complete block neighbor designs for $v = 20$.

4.2 NNBD for v Odd and $k = v$-1

Theorem 4.2: For v odd, S_5 provides nearest neighbor designs with $\lambda' = 2$ in v circular binary blocks if the sum of any two, three, …, successive elements of each of set in S_5 is not 0 mod v-1.

$$S_5 = [1, v\text{-}3, 3, v\text{-}5, 5, \ldots, v\text{-}(k\text{-}1)]t + [1, 1, \ldots, 1]\ 1/(v\text{-}1)$$

Proof: New set of shifts S^*_5 contains all of 1, 2, 3, …, v-2 twice. Hence the theorem

Example 4.2: [1,8,3,6,5,4,7,2]t + [1,1,1,1,1,1,1,1,1](1/10) provides the nearest neighbor designs for $v = 11$ and $k = 10$.

4.3 NNBD for $v = 4s$ And $k = v$-1

Theorem 4.3: For $v = 4s$ and $m = 2s-1$, s is an integer, following set of shifts S_6 provides neighbor designs with $\lambda' = 2$ in v circular binary blocks if the sum of any two, three, … or $2m+1$ successive elements of S_6 is not 0 mod v-1.

$$S_6 = [1, 2, 3, \ldots, (m\text{-}1),\ m,\ m,\ (m\text{-}1), \ldots, 3, 2]t + [1, 1, \ldots, 1]\ 1/(v\text{-}1)$$

Proof: New set of shifts S^*_6 contains all of 1, 2, 3, …, v-2 twice. Hence the theorem

Example4.3:

[1,2,3,4,5,6,7,8,9,9,8,7,6,5,4,3,2]t + [1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1](1/19) provides the nearest neighbor designs for $v = 20$ and $k = 19$.

4.4 NNBD for $v = 2m$, $k = m$

Let $v=2m$; $m = 1, 2, 3, \ldots$, $S_{7,0} = [1, 2, 3, \ldots, m\text{-}1, m]$, $s = m(m+1)/2$ mod v-1 then take the elements of $S_{7,1}$ as:

- If $s < m$, 1,2,3,…, (m-1), (m+1) excluding the element equal to s.
- If $s > m$, 1,2,3,…, (m-1), m then excluding the element equal to $2m+1-s$.
- If $s = m$, 1,2,3,…,(m-2),(m+2) then excluding the element equal to s.

If the sum of any two successive, three successive, … , or all elements of $S_{7,0}$ or $S_{7,1}$ is 0 mod v-1 then rearrange the elements of $S_{7,0}$ or $S_{7,1}$ such that the sum of any two successive, three successive, …, or all elements of rearranged elements is not 0 mod v-1 to obtain the binary block neighbor designs.

Example 4.4: For v=16 and k=8, S_7 = [1,2,3,4,7,6,5] + [1,3,4,6,7,5]t provides the nearest neighbor designs for v = 16 and k = 8.

4.5 Designs Balanced for Second Order Neighbor Effects When *k* = *v*

Following are the set(s) of shifts which generates second order neighbor designs in v-1 blocks for v = k = 6,8,9,10,12 and 15.

v	**Set(s) of shifts**
6	[1,2,1,3]t + [1,2,4,2]t
8	[6,6,4,2,4,2]t + [6,5,4,4,5,6]t
9	[1,3,2,7,6,4,3]t
10	[1,2,3,8,2,4,6,5]t
12	[1,1,6,2,5,3,7,2,4,8]t
15	[1,1,2,6,12,5,10,11,11,8,10,5,7]t

[2,1,1,3,5,10] + [7,9,8,6,2]t provides second order neighbor designs for v = 14, k = 7, λ' = 2.

4.6 Designs Balanced for Second Order Neighbor Effects When *k* = *v*-1

Following are the set(s) of shifts which generates second order neighbor balanced designs in v blocks of size k = v-1 for v = 9 to 21.

v	**Set(s) of shifts**
9	[1,1,3,2,5,3,7]
10	[1,2,3,5,2,4,6]t + [1,1,1,1,1,1,1,1](1/9)
11	[1,6,6,5,8,3,3,2]t + [[1,1,1,1,1,1,1,1,1](1/10)
12	[1,1,2,4,3,7,4,7,10,6]
13	[9,1,1,3,5,2,7,5,11,7,4]
14	[1,5,9,2,8,9,3,7,11,7,10]t + [[1,1,1,1,1,1,1,1,1,1,1,1](1/13)
15	[1,2,3,4,9,6,2,3,7,9,8,10]t + [1,1,1,1,1,1,1,1,1,1,1,1,1](1/14)
16	[1,23,4,7,9,11,6,10,12,7,2,10]t + [1,1,1,1,1,1,1,1,1,1,1,1,1,1](1/15)
17	[1,2,3,4,8,11,7,3,5,2,10,7,10,12]t + [1,1,1,1,1,1,1,1,1,1,1,1,1,1,1](1/16)
18	[1,2,3,4,,5,6,7,11,8,12,4,7,14,8,2]t + [1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1](1/17)
19	[1,3,2,5,4,8,11,15,14,8,9,12,12,11,5,16]t + [1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1](1/18)

4.7 Designs Balanced for Third Order Neighbor Effects When *k* = *v*-1

Following are the set(s) of shifts which generates third order neighbor designs for k v-1 in v blocks.

v	**Set (s) of shifts**
7	[1,1,2,2,4]
8	[2,1,3,5,4]t + [1,1,1,1,1,1](1/7)
9	[1,3,2,5,4,6]t + [1,1,1,1,1,1,1](1/8)
11	[1,1,3,3,9,9,5,5,4]
12	[1,2,4,3,1,10,7,4,6,3]
13	[6,5,9,3,2,6,3,1,5,2,1]
14	[1,9,8,8,7,12,3,2,3,5,5]t + [1,1,1,1,1,1,1,1,1,1,1,1](1/13)

4.8 Designs Balanced for Third Order Neighbor Effects When *k* = *v*

Following are the set(s) of shifts which generates third order neighbor designs for *v* =*k* = 8,10 and 12.

v	**Set (s) of shifts**
8	[1,1,2,1,5,3]t + [2,3,1,4,5,3]t
10	[1,4,6,1,5,7,7,3]t
12	[1,3,6,7,7,3,2,2,10,6]t

4.9 Some other Designs Balanced for Third Order Neighbor Effects

Following are the set(s) of shifts which generates some other third order neighbor designs

v	***k***	**Set (s) of shifts**
9	6	[2,1,3,3]t + [1,4,7,2,4](1/2)
12	6	[1,2,3,10,8] + [4,4,5,5]t

4.10 Designs Balanced for Fourth Order Neighbor Effects

Following are the set(s) of shifts which generates some fourth order neighbor designs.

v	***k***	**Set (s) of shifts**
9	8	[1,3,6,3,2,4]t + [1,1,1,1,1,1,1](1/8)
11	10	[6,1,2,4,8,5,10,9,7]
13	12	[10,2,6,12,4,6,5,10,12,5,2]

REFERENCES

Ahmed, R. and Akhtar, M. (2008). Construction of neighbor balanced block designs. *Journal of Statistical Theory and Practice.* 2(4), 551-558.

Ahmed, R. and Akhtar, M. (2009). Construction of one dimensional all order neighbor balanced designs by cyclic shifts. *Pakistan Journal of Statistics*, 25(2), 121-126.

Ai, M. Y., Gennian Ge and Ling-Yau Chan (2007). Circular neighbor balanced designs universally optimal for total effects. *Science in China Series A: Mathematics* 50, 821-828.

Ai, M. Y., Yu, Y. L. and He, S. Y. (2009). Optimality of circular neighbor balanced designs for total effects with autoregressive correlated observations. *J. Statist. Plann. Inference*, 139, 2293-2304.

Akhtar, M. and Ahmed, R. (2009). Circular binary block second and higher order neighbor designs. *Communications in Statistics – Simulation and Computation*, 38, 821-828.

Azais, J. M., Bailey, R. A. and Monod, H. (1993). A catalogue of efficient neighbour designs with border plots. *Biometrics*, 49 (4), 1252-61.

Das, A. D. and Saha, G. M. (1976). On construction of neighbor designs. *Calcutta Statist. Assoc. Bull.* 25, 151-164.

Dey, A. and Chakravarty, R. (1977). On the construction of some classes of neighbor designs. *J. Indian Soc. Agri. Statist*. 29, 97-104.

Hwang, F. K. (1973). Constructions for some classes of neighbor designs. *Annals of Statistics*, 1(4), 786-790.

Iqbal, I., Amanullah, M. and Nasir, J. A. (2006). The construction of second order neighbor designs. *Journal of Research* (*Science*), *BZU, Multan, Pakistan*, 17 (3), 191- 199.

Iqbal, I., Tahir, M. H., Ghazali, S. S. A. (2009). Circular neighbor-balanced designs using cyclic shifts. *Sci. China Ser. A-Mathematics*, 52 (10), 2243-2256.

Lawless, J. F. (1971) A Note on Certain Types of BIBDs Balanced for Residual Effects. *Annals of Mathematical Statistics*, 42, 1439-1441.

Monod, H. (1992). Two- factor neighbour designs in incomplete blocks for intercropping experiments. *The Statistician*, 41 (5), 487-497.

Rees, D. H. (1967). Some designs of use in serology. *Biometrics*, 23, 779–791.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 375–383.

NEW ESTIMATORS FOR THE POPULATION MEDIAN IN SIMPLE RANDOM SAMPLING

Sibel Al[1] and Hulya Cingi[2]
Hacettepe University, Department of Statistics, Ankara, Turkey
E-mail: [1]sibelal@hacettepe.edu.tr, [2]hcingi@hacettepe.edu.tr

ABSTRACT

In this study, we define a new median estimator utilizing the mean estimator in Searls (1964) and obtain the mean square error (MSE) equation of the suggested estimator. We theoretically show that the suggested estimator is always more efficient than the sample median defined by Gross (1980). Second proposed estimator is calculated by modifying the estimator proposed by Prasad (1989) for the median. When we debate the efficiency of the second proposed estimator, it is always more efficient than the ratio median estimator which is defined by Kuk and Mak (1989). In addition, utilizing the estimator in Singh et al. (2003b), we suggest a new ratio estimator motivated by Searls (1964) in simple random sampling. We get the MSE equation of the proposed estimator and compare it with the MSE of Singh et al. (2003b) estimator. By this comparison, we obtain that the proposed estimator is always more efficient than the Singh et al. (2003b) estimator. Finally we suggest a family of estimators of population median and derive the minimum MSE equation of the proposed estimator. All theoretical comparisons are also performed by using numerical examples.

Keywords: Auxiliary information, mean square error (MSE), median estimation, ratio estimator, simple random sampling.

1. INTRODUCTION

In survey sampling, when variables have a highly skewed distribution such as income, expenditure, production are studied, median is often regarded as a more appropriate measure of location than mean. In literature there have been several studies for mean estimation but few studies for median estimation in simple random sampling. Some researches on median estimation in simple random sampling are Gross (1980), Kuk and Mak (1989), Singh et al. (2003a), Singh et al. (2003b), and Singh (2003).

Let a sample of size n is selected from a finite population of size N by simple random sampling without replacement (SRSWOR). Y_i and X_i, $i=1,2,..N$, denote the values of population units. y_i and x_i, $i=1,2,\ldots,n$, denote the values of sample units for the study variable Y and auxiliary variable X respectively. Let M_Y and M_X be the population medians and $\hat{M}_Y$ and $\hat{M}_X$ are the sample estimators of M_Y and M_X, respectively.

Gross (1980) defined the sample median and showed that $\hat{M}_Y$ is consistent, asymptotically normal with mean M_Y and variance

$$V(\hat{M}_Y) = \lambda\{f_Y(M_Y)\}^{-2} \tag{1}$$

where $\lambda = (1-f)/4n$, $f = n/N$ and $f_Y(\bullet)$ is the density function of Y.

Kuk and Mak (1989) suggested a ratio estimator of median using auxiliary information given in (2)

$$\hat{M}_{YR} = \hat{M}_Y \frac{M_X}{\hat{M}_X}. \tag{2}$$

Let $y_{(1)} \le y_{(2)} \le ... \le y_{(n)}$ be the ordered y values in the sample. Let t be an integer such that $y_{(t)} \le M_Y \le y_{(t+1)}$ and $p = t/n$ be the proportion of the y values in the sample that are less than or equal to M_Y. Thus M_Y is approximately the sample p^{th} quantile $\hat{Q}_Y(p)$. The sample median $\hat{M}_Y$ can be viewed as the special estimator $\hat{Q}_Y(\hat{p})$ with $\hat{p} = 0.5$. Kuk and Mak (1989) defined the two-way classification seen in Table 1, where for instance P_{11} is the proportion of units in the population with $X \le M_X$ and $Y \le M_Y$. Following Kuk and Mak (1989) the MSE of the ratio estimator is given by

$$MSE(\hat{M}_{YR}) \approx \lambda[\{f_Y(M_Y)\}^{-2} + M_Y^2\{M_X f_X(M_X)\}^{-2} - 2M_Y\{M_X f_X(M_X) f_Y(M_Y)\}^{-1}\rho_{XY}]$$

where ρ_{XY} is the correlation coefficient between sampling distribution of $\hat{M}_Y$ and $\hat{M}_X$ which is defined as $\rho_{XY} = 4P_{11} - 1$. If we define $C_{M_Y} = \{M_Y f_Y(M_Y)\}^{-1}$ and $C_{M_X} = \{M_X f_X(M_X)\}^{-1}$ as coefficient of variations of median of study and auxiliary variables respectively, we can write the MSE equation of ratio estimator as following:

$$MSE(\hat{M}_{YR}) \approx M_Y^2[\lambda C_{M_Y}^2 - 2\lambda C_{M_Y} C_{M_X} \rho_{XY} + \lambda C_{M_X}^2]. \tag{3}$$

Table 1 A Matrix of Proportions (P_{ij})

	$Y \le M_Y$	$Y > M_Y$
$X \le M_X$	P_{11}	P_{21}
$X > M_X$	P_{12}	P_{22}

Singh et al. (2003b) suggested a ratio type estimator for population median as:

$$\hat{M}_{YSSP} = \hat{M}_Y \left(\frac{M_X + A}{\hat{M}_X + A} \right) \tag{4}$$

where A which could be the mode or range of the auxiliary variable (Singh et al., 2003a) is a suitably chosen real constant. Singh et al. (2003b) obtained the MSE equation of $\hat{M}_{YSSP}$

$$MSE(\hat{M}_{YSSP}) \approx \lambda[\{f_Y(M_Y)\}^{-2} + \delta^2 M_Y^2\{M_X f_X(M_X)\}^{-2} - 2\delta M_Y\{M_X f_X(M_X) f_Y(M_Y)\}^{-1}\rho_{XY}]$$

$$MSE(\hat{M}_{YSSP}) \approx M_Y^2[\lambda C_{M_Y}^2 - 2\delta\lambda C_{M_Y} C_{M_X}\rho_{XY} + \delta^2\lambda C_{M_X}^2] \tag{5}$$

where $\delta = \dfrac{M_X}{M_X + A}$.

Following Searls (1964), Prasad (1989) and Singh et al. (2003b), we define new median estimators and obtain the MSE equations. Also, we suggest a family of estimators of population median and derive the minimum MSE equation of the proposed estimator. Proposed median estimators are theoretically compared with other estimators. These comparisons are also performed by using numerical examples.

2. SUGGESTED ESTIMATORS

2.1 First Proposed Estimator

Searls (1964) defined a mean estimator using a known coefficient of variation. Following Searls (1964), we adapted this estimator to median estimator. A modified estimator is given in (6)

$$\hat{M}_{YP1} = \kappa_1 \hat{M}_Y \tag{6}$$

where κ_1 is a constant which minimizes the MSE equation. The MSE of $\hat{M}_{YP1}$ can be found as follows:

$$\begin{aligned} E(\hat{M}_{YP1} - M_Y)^2 &= E(\kappa_1\hat{M}_Y - M_Y)^2 \\ &= E(\kappa_1^2\hat{M}_Y^2 - 2\kappa_1\hat{M}_Y M_Y + M_Y^2). \end{aligned}$$

$E(\hat{M}_Y) = M_Y$ can be asymptotically written. So the MSE of $\hat{M}_{YP1}$ can be obtained as

$$\begin{aligned} MSE(\hat{M}_{YP1}) &= \kappa_1^2 E(\hat{M}_Y^2) - 2\kappa_1 E(\hat{M}_Y)M_Y + E(M_Y^2) \\ &= \kappa_1^2 E(\hat{M}_Y^2) - \kappa_1^2[E(\hat{M}_Y)]^2 + \kappa_1^2 M_Y^2 - 2\kappa_1 M_Y^2 + M_Y^2 \\ &= \kappa_1^2 V(\hat{M}_Y) + M_Y^2(\kappa_1 - 1)^2. \end{aligned} \tag{7}$$

We obtain the optimum value of κ_1 which is given in (8) with minimizing the MSE of $\hat{M}_{YP1}$.

$$\begin{aligned} \frac{\partial MSE(\hat{M}_{YP1})}{\partial \kappa_1} &= 2M_Y^2(\kappa_1 - 1) + 2\kappa_1\lambda\{f_Y(M_Y)\}^{-2} \\ &= 0 \end{aligned},$$

$$\kappa_1 = \frac{1}{1 + \lambda\{M_Y f_Y(M_Y)\}^{-2}}. \tag{8}$$

If we define $C_{M_Y} = \{M_Y f_Y(M_Y)\}^{-1}$ as a coefficient of variation of median, we can obtain the optimum value of κ_1 and minimum MSE equation as following:

$$\kappa_1 = (1+\lambda C_{M_Y}^2)^{-1}, \tag{9}$$

$$MSE_{Min}(\hat{M}_{YP1}) = \kappa_1 V(\hat{M}_Y). \tag{10}$$

2.2 Second Proposed Estimator

Following Prasad (1989), we define new ratio estimator as following:

$$\hat{M}_{YP2} = \kappa_2 \hat{M}_Y \frac{M_X}{\hat{M}_X} \tag{11}$$

where κ_2 is a constant which minimizes the MSE equation of $\hat{M}_{YP2}$. If we take $\hat{R}_{M\kappa 2} = \kappa_2 \dfrac{\hat{M}_Y}{\hat{M}_X}$ as an estimator of the $R_{M2} = \dfrac{M_Y}{M_X}$, we can write $MSE(\hat{R}_{M\kappa 2})$ as:

$$\begin{aligned} MSE(\hat{R}_{M\kappa 2}) &= E(\hat{R}_{M\kappa 2} - R_{M2})^2 \\ &= E\left(\frac{\kappa_2 \hat{M}_Y - \hat{M}_X R_{M2}}{\hat{M}_X}\right)^2. \end{aligned} \tag{12}$$

We can write following equation using Taylor series expansion.

$$\frac{1}{\hat{M}_X} = \frac{1}{M_X + \hat{M}_X - M_X} = \frac{1}{M_X}\left(1+\frac{\hat{M}_X - M_X}{M_X}\right)^{-1} \approx \frac{1}{M_X}$$

The large sample approximation up to order n^{-1} to the MSE of the estimator $\hat{R}_{M\kappa 2}$ can be obtained as:

$$\begin{aligned} MSE(\hat{R}_{M\kappa 2}) &= \frac{1}{M_X^2} E(\kappa_2 \hat{M}_Y - \hat{M}_X R_{M2})^2 \\ &= \frac{1}{M_X^2} E[\kappa_2 \hat{M}_Y - M_Y - R_{M2}(\hat{M}_X - M_X)]^2 \\ &= \frac{1}{M_X^2} E[(\kappa_2 \hat{M}_Y - M_Y)^2 - 2R_{M2}(\kappa_2 \hat{M}_Y - M_Y)(\hat{M}_X - M_X) \\ &\qquad + R_{M2}^2(\hat{M}_X - M_X)^2] \\ &= \frac{1}{M_X^2}[\kappa_2{}^2 V(\hat{M}_Y) + M_Y^2(\kappa_2 - 1)^2 - 2R_{M2}\operatorname{cov}(\kappa_2 \hat{M}_Y, \hat{M}_X) \\ &\qquad + R_{M2}^2 V(\hat{M}_X)]. \end{aligned} \tag{13}$$

We can get the MSE of $\hat{M}_{YP2}$ using equation (13)

$$
\begin{aligned}
MSE(\hat{M}_{YP2}) &= E(\hat{M}_{YP2} - M_Y) \\
&= E(\hat{R}_{M\kappa 2} M_X - R_{M2} M_X)^2 \\
&= M_X^2 E(\hat{R}_{M\kappa 2} - R_{M2})^2 \\
&= [\kappa_2{}^2 V(\hat{M}_Y) + M_Y^2(\kappa_2 - 1)^2 - 2R_{M2}\operatorname{cov}(\kappa_2 \hat{M}_Y, \hat{M}_X) + R_{M2}^2 V(\hat{M}_X)]
\end{aligned}
$$

$$
MSE(\hat{M}_{YP2}) = M_Y^2[\kappa_2^2 \lambda C_{M_Y}^2 + (\kappa_2 - 1)^2 - 2\kappa_2 \lambda C_{M_x} C_{M_Y} \rho_{XY} + \lambda C_{M_X}^2] \tag{14}
$$

where $V(\hat{M}_Y) = M_Y^2 \lambda C_{M_Y}^2$, $\operatorname{cov}(\kappa_2 \hat{M}_Y, M_X) = \kappa_2 M_X M_Y \lambda C_{M_Y} C_{M_X} \rho_{XY}$ and $V(\hat{M}_X) = M_X^2 \lambda C_{M_X}^2$.

From (14), the optimum value of κ_2 is obtained as following:

$$
\frac{\partial MSE(\hat{M}_{YP2})}{\partial \kappa_2} = 0,
$$

$$
\kappa_2 = \frac{1 + \lambda C_{M_X} C_{M_Y} \rho_{XY}}{1 + \lambda C_{M_Y}^2}.
$$

Putting the optimum value of κ_2 in (14), the minimum MSE of $\hat{M}_{YP2}$ is obtained as following:

$$
MSE_{Min}(\hat{M}_{YP2}) = M_Y^2 \left(1 - \frac{[1 + \lambda C_{M_X} C_{M_Y} \rho_{XY}]^2}{1 + \lambda C_{M_Y}^2} \right) + M_Y^2 \lambda C_{M_X}^2. \tag{15}
$$

2.3 Third Proposed Estimator

Utilizing the estimator in Singh et al. (2003b), we suggest a new ratio estimator motivated by Searls (1964) as:

$$
\hat{M}_{YP3} = \kappa_3 \hat{M}_Y \left[\frac{M_X + A}{\hat{M}_X + A} \right] \tag{16}
$$

where κ_3 is a constant which minimizes the MSE equation of $\hat{M}_{YP3}$. If we define $\hat{R}_{M\kappa 3} = \kappa_3 \dfrac{\hat{M}_Y}{\hat{M}_X + A}$ as an estimator of the $R_{M3} = \dfrac{M_Y}{M_X + A}$, we can write $MSE(\hat{R}_{M\kappa 3})$ as

$$
\begin{aligned}
MSE(\hat{R}_{M\kappa 3}) &= E(\hat{R}_{M\kappa 3} - R_{M3})^2 \\
&= E\left(\frac{\kappa_3 \hat{M}_Y - (\hat{M}_X + A) R_{M3}}{\hat{M}_X + A} \right)^2.
\end{aligned}
$$

Similarly, using Taylor series expansion we can get

$$\frac{1}{\hat{M}_X + A} \approx \frac{1}{M_X + A}.$$

The large sample approximation up to order n^{-1} to the MSE of the estimator $\hat{M}_{YP3}$ can be similarly written as following:

$$MSE(\hat{M}_{YP3}) = M_Y^2[\kappa_3^2 \lambda C_{M_Y}^2 + (\kappa_3 - 1)^2 - 2\kappa_3 \delta\lambda C_{M_X} C_{M_Y} \rho_{XY} + \delta^2 \lambda C_{M_X}^2] \tag{17}$$

where $\delta = M_X / (M_X + A)$. We obtain the optimum value of κ_3 given in (17) with minimizing the MSE equation of $\hat{M}_{YP3}$

$$\kappa_3 = \frac{1 + \delta\lambda C_{M_X} C_{M_Y} \rho_{XY}}{1 + \lambda C_{M_Y}^2}. \tag{18}$$

Putting optimum value of κ_3 in (17), the minimum MSE equation of $\hat{M}_{YP3}$ can be easily obtained as following:

$$MSE_{Min}(\hat{M}_{YP3}) = M_Y^2 \left(1 - \frac{[1 + \delta\lambda C_{M_X} C_{M_Y} \rho_{XY}]^2}{1 + \lambda C_{M_Y}^2} \right) + \delta^2 M_Y^2 \lambda C_{M_X}^2. \tag{19}$$

Finally, we define a family of estimators of population median seen in (20)

$$T = \kappa \hat{M}_Y \left[\frac{\eta M_X + \tau}{\eta \hat{M}_X + \tau} \right] \tag{20}$$

where κ, η, τ are suitably chosen scalars. Several estimators can be generated from T by putting suitable values of κ, η, τ. These estimators can be seen in Table 2.

Table 2 Members of the family of estimators

	κ	η	τ	MSE
Gross(1980)	1	0	1	$M_Y^2 \lambda C_{M_Y}^2$
Kuk and Mak (1989)	1	1	0	$M_Y^2[\lambda C_{M_Y}^2 - 2\lambda C_{M_Y} C_{M_X} \rho_{XY} + \lambda C_{M_X}^2]$
Singh et al. (2003b)	1	1	A	$M_Y^2[\lambda C_{M_Y}^2 - 2\delta\lambda C_{M_Y} C_{M_X} \rho_{XY} + \delta^2 \lambda C_{M_X}^2]$
Proposed 1	κ_1	0	1	$M_Y^2[\kappa_1^2 \lambda C_{M_Y}^2 + (\kappa_1 - 1)^2]$
Proposed 2	κ_2	1	0	$M_Y^2[\kappa_2^2 \lambda C_{M_Y}^2 + (\kappa_2 - 1)^2 - 2\kappa_2 \lambda C_{M_Y} C_{M_X} \rho_{XY} + \lambda C_{M_X}^2]$
Proposed 3	κ_3	1	A	$M_Y^2[\kappa_3^2 \lambda C_{M_Y}^2 + (\kappa_3 - 1)^2 - 2\kappa_3 \delta\lambda C_{M_Y} C_{M_X} \rho_{XY} + \delta^2 \lambda C_{M_X}^2]$

Minimum MSE equation can be similarly found where $\hat{R}_{M\kappa T} = \kappa \dfrac{\hat{M}_Y}{\eta \hat{M}_X + \tau}$ and $R_{MT} = \dfrac{M_Y}{\eta M_X + \tau}$. Minimum MSE equation is seen in (20)

$$MSE_{Min}(T) = M_Y^2 \left(1 - \frac{[1 + \eta\omega\lambda C_{M_X} C_{M_Y} \rho_{XY}]^2}{1 + \lambda C_{M_Y}^2} \right) + M_Y^2 \eta^2 \omega^2 \lambda C_{M_X}^2 \tag{21}$$

where optimum value of κ is $\kappa = \dfrac{1 + \eta\omega\lambda C_{M_X} C_{M_Y} \rho_{XY}}{1 + \lambda C_{M_Y}^2}$ and $\omega = \dfrac{M_X}{\eta M_X + \tau}$. Note that the optimum value of κ is can not be achieved in applications since C_{M_X}, C_{M_Y} and ρ_{XY} are unknown. Only the estimates of C_{M_X}, C_{M_Y} and ρ_{XY} can be used to determine the optimum value κ.

3. EFFICIENCY COMPARISONS

In this section we debate the efficiencies of proposed estimators with comparing MSE equations. From (1) and (10) we can write following inequalities:

$$MSE_{Min}(\hat{M}_{YP1}) < V(\hat{M}_Y)$$
$$(1 + \lambda C_{M_Y}^2)^{-1} V(\hat{M}_Y) < V(\hat{M}_Y)$$

$\lambda C_{M_Y}^2 > 0$, always. (22)

The efficiency condition is always satisfied so it is clearly seen that the first proposed estimator is always more efficient than the sample median. From (3) and (15) the efficiency condition can be obtained.

$$MSE_{Min}(\hat{M}_{YP2}) < MSE(\hat{M}_{YR})$$
$$M_Y^2 \left(1 - \frac{[1 + \lambda C_{M_X} C_{M_Y} \rho_{XY}]^2}{1 + \lambda C_{M_Y}^2} \right) + \lambda M_Y^2 C_{M_X}^2 < M_Y^2 [\lambda C_{M_Y}^2 - 2\lambda C_{M_Y} C_{M_X} \rho_{XY} + \lambda C_{M_X}^2]$$

$(C_{M_Y} - C_{M_X} \rho_{XY})^2 > 0$, always. (23)

It is clearly seen that second proposed estimator is always more efficient than the ratio estimator. From (5) and (19) we can write these inequalities:

$$MSE_{Min}(\hat{M}_{YP3}) < MSE(\hat{M}_{YSSP})$$
$$M_Y^2 \left(1 - \frac{[1 + \delta\lambda C_{M_X} C_{M_Y} \rho_{XY}]^2}{1 + \lambda C_{M_Y}^2} \right) + M_Y^2 \delta^2 \lambda C_{M_X}^2 < M_Y^2 [\lambda C_{M_Y}^2 - 2\delta\lambda C_{M_Y} C_{M_X} \rho_{XY} + \delta^2 \lambda C_{M_X}^2]$$

$(C_{M_Y} - \delta C_{M_X} \rho_{XY})^2 > 0$, always. (24)

4. NUMERICAL COMPARISON

We use two data sets for comparing efficiencies of estimators. In first data set, we use the data of Cingi et al. (2007). Cingi et al. (2007), compute the development index of all districts in Turkey with regard to educational opportunities by using the data gathered from schools by the Ministry of National Education for the 2006-2007 educational years. Development groups are obtained by clustering the districts with the same development level in the same group. We use the data of number of teachers (as study variable) in elementary schools for 340 medium-developed districts in Turkey in 2007 and number of students as auxiliary variable. We take second data set from Chen et al. (2004). The diameter of conifer trees in centimeters at breast height is chosen as an auxiliary variable, the entire height of conifer trees in feet is taken as a study variable. In application we define new ratio estimators which are members of the family of estimators using known auxiliary information. These estimators can be seen in Table 3.

Table 3 Members of the proposed family of estimators

No	Estimator	κ	η	τ	No	Estimator	κ	η	τ
1	$T_1 = \hat{M}_Y$	1	0	1	9	$T_9 = \kappa_3 \hat{M}_Y \left(\frac{M_X + G_X}{\hat{M}_X + G_X} \right)$	κ_3	1	G_X
2	$T_2 = \frac{\hat{M}_Y}{\hat{M}_X} M_X$	1	1	0	10	$T_{10} = \kappa_3 \hat{M}_Y \left(\frac{M_X + M_0}{\hat{M}_X + M_0} \right)$	κ_3	1	M_0
3	$T_3 = \hat{M}_Y \left(\frac{M_X + \rho_{XY}}{\hat{M}_X + \rho_{XY}} \right)$	1	1	ρ_{XY}	11	$T_{11} = \kappa \hat{M}_Y \left(\frac{\rho_{XY} M_X + G_X}{\rho_{XY} \hat{M}_X + G_X} \right)$	κ	ρ_{XY}	G_X
4	$T_4 = \hat{M}_Y \left(\frac{M_X + G_X}{\hat{M}_X + G_X} \right)$	1	1	G_X	12	$T_{12} = \kappa \hat{M}_Y \left(\frac{G_X M_X + \rho_{XY}}{G_X \hat{M}_X + \rho_{XY}} \right)$	κ	G_X	ρ_{XY}
5	$T_5 = \hat{M}_Y \left(\frac{M_X + M_0}{\hat{M}_X + M_0} \right)$	1	1	M_0	13	$T_{13} = \kappa \hat{M}_Y \left(\frac{\rho_{XY} M_X + M_0}{\rho_{XY} \hat{M}_X + M_0} \right)$	κ	ρ_{XY}	M_0
6	$T_6 = \kappa_1 \hat{M}_Y$	κ_1	0	1	14	$T_{14} = \kappa \hat{M}_Y \left(\frac{M_0 M_X + \rho_{XY}}{M_0 \hat{M}_X + \rho_{XY}} \right)$	κ	M_0	ρ_{XY}
7	$T_7 = \kappa_2 \frac{\hat{M}_Y}{\hat{M}_X} M_X$	κ_2	1	0	15	$T_{15} = \kappa \hat{M}_Y \left(\frac{M_0 M_X + G_X}{M_0 \hat{M}_X + G_X} \right)$	κ	M_0	G_X
8	$T_8 = \kappa_3 \hat{M}_Y \left(\frac{M_X + \rho_{XY}}{\hat{M}_X + \rho_{XY}} \right)$	κ_3	1	ρ_{XY}	16	$T_{16} = \kappa \hat{M}_Y \left(\frac{G_X M_X + M_0}{G_X \hat{M}_X + M_0} \right)$	κ	G_X	M_0

***** G_X : Range of the auxiliary variable

** M_0 : Mode of the auxiliary variable

In Table 4, we observe the statistics about the populations.

Table 4 Data Statistics

	Data 1	**Data 2**		**Data 1**	**Data 2**
N	340	396	$f_Y(M_Y)$	0.00018018	0.011784
n	150	65	P_{11}	0.48	0.46
M_X	3513	14.6	ρ_{XY}	0.92	0.84
M_Y	178	30	G_X	171278	73.6
$f_X(M_X)$	0.00008341	0.02194	M_0	127	2.5

We compute the MSE values of all estimators seen in Table 5. In first data set, T_7 and T_{12} have the smallest MSE value. T_{13} is the most efficient estimator for second data set with having the smallest MSE value. In both data sets, T_6 is more efficient than T_1, T_7 is more efficient than

T_2 and T_8 is more efficient than T_3. In Section 3, the efficiency conditions given in (22), (23) and (24) are always provided. The numerical results support the theoretical results.

Table 5 Mean Square Errors of the Estimators

Estimator	Data 1	Data 2	Estimator	Data 1	Data 2
T_1	28687.06816	23.15123	T_9	14994.71474	16.40236
T_2	23253.17619	8.42478	T_{10}	12942.68342	7.03269
T_3	23254.50891	7.77532	T_{11}	14999.48728	17.15850
T_4	28571.08742	16.81843	T_{12}	**12228.83881**	8.41112
T_5	24635.44323	7.05792	T_{13}	12988.13475	**6.92141**
T_6	15055.57373	22.57063	T_{14}	12228.83935	8.13088
T_7	**12228.83881**	8.42169	T_{15}	12325.44252	11.74148
T_8	12229.52511	7.76648	T_{16}	12228.84438	8.39039

5. CONCLUSION

In this study, we suggest new median estimators using a known constant. We theoretically show that these estimators are always more efficient than classical estimators. In the numerical examples, the theoretical results are also supported. In future works, we hope to adapt the estimators proposed in this study to stratified random sampling.

REFERENCES

Chen, Z., Bai, Z., Sinha, B.K. (2004). *Ranked Set Sampling Theory and Applications*. New York: Springer-Verlag.

Cingi, H., Kadilar, C., Kocberber, G. (2007). Examination of educational opportunities at primary and secondary schools in Turkey suggestions to determined issues. TUBITAK, SOBAG, 106K077. http://yunus.hacettepe.edu.tr/~hcingi/

Gross, T.S. (1980). Median estimation in sample surveys. *Proc. Surv. Res. Meth. Sect. Amer. Statist. Ass.* 181–184.

Kuk, A.Y.C., Mak, T.K. (1989). Median estimation in the presence of auxiliary information. *Journal of the Royal Statistical Society Series, B,* 51, 261–269.

Prasad, B. (1989). Some improved ratio type estimators of population mean and ratio in finite population sample surveys. *Communications in Statistics Theory Methods*, 18, 379–392.

Searls, D.T. (1964). The utilization of a known coefficient of variation in the estimation procedure. *Journal of the American Statistical Association*, 59, 1225–1226.

Singh, S. (2003). *Advanced Sampling Theory with Applications: How Michael 'selected' Amy*. London: Kluwer Academic Publishers.

Singh, H.P., Singh, S., Joarder, A.H. (2003a). Estimation of population median when mode of an auxiliary variable is known. *Journal of Statistical Research*, 37, 1, 57–63.

Singh, H.P., Singh, S., Puertas, S.M. (2003b). Ratio type estimators for the median of finite populations. *Allgemeines Statistisches Archiv*, 87, 369–382.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 384–391.

FORECASTING TOURISM DEMAND OF TURKEY BY USING ARTIFICIAL NEURAL NETWORKS

Cagdas Hakan Aladag[1] and Erol Egrioglu[2]

[1]Hacettepe University, Department of Statistics, Ankara, Turkey
E-mail: aladag@hacettepe.edu.tr
[2]Ondokuz Mayis University, Department of Statistics, Samsun, Turkey
E-mail: erole@omu.edu.tr

ABSTRACT

There is a need to forecast tourism demand accurately so that directors and investors can make operational, tactical and strategic decisions, examples of which are scheduling and staffing, preparing tour brochures and hotel investments. Similarly, government organizations need accurate tourism demand forecasts to plan required tourism infrastructures, such as accommodation site planning and transportation development, among other needs. Therefore, many methods have been proposed in order to forecast tourism time series in the literature. The artificial neural networks (ANN) method, which has been successfully used for time series forecasting in many applications, is one of these methods that have been used to obtain more accurate tourism demand forecasts. However, in the usage of ANN, there are still some problems such as determining the architecture which produces the most accurate forecasts. In this study, to calculate forecasts of international tourism arrivals to Turkey, ANN are employed and to determine the best architecture, the weighted information criterion (WIC) proposed by Egrioglu, Aladag and Gunay is used (Egrioglu et al., 2008). The calculated forecasts are compared to those calculated from other architectures which are determined by frequently used performance criteria and the obtained results are discussed.

Keywords: Artificial neural networks, forecasting, time series, tourism demand, weighted information criterion.

1. INTRODUCTION

Tourism demand forecasting has become an important component in tourism research. Twenty years ago, there were only a handful of academic journals that published tourism related research. Now there are more than 70 journals that serve a thriving research community covering more than 3000 tertiary institutions across five continents (Song and Li, 2008). In addition to academics who have worked on tourism demand modeling and forecasting it has also attracted much attention from practitioners since it is one of the most important areas in tourism research. Li et al. (2005) made an extensive review and found that 420 studies on this topic were published during the years between 1960 and 2002.

Palmer et al. (2006) explained the importance of accurate tourism demand forecast as follows: In the case of tourism demand, better forecasts would help directors and investors make operational, tactical and strategic decisions, examples of which are scheduling and staffing, preparing tour brochures and hotel investments. Similarly, government bodies need accurate tourism demand forecasts to plan required tourism infrastructures, such as accommodation site planning and transportation development, among other needs. Forecasting tourism expenditure is also of value in ascertaining the relative contributions of tourism to production, income and employment in tourist destinations (Bull, 1995).

In the literature, many methods such as Box Jenkins, ANN, fuzzy time series and multivariate time series methods have been used to get more accurate tourism demand forecasts due to the aforementioned reasons. ANN have been successfully used to forecast in many applications since it can model both the linear and non linear parts of time series and do not require satisfaction of any assumption (Aladag et al., 2009a). Although ANN have important advantages in forecasting, there are still some problems with using this method. When ANN are utilized, selection of the optimal components of this method is a vital issue for obtaining good results. Determining the best ANN model can be defined as selection of the components such as architecture structure, learning algorithm and activation function (Egrioglu et al., 2008). Selection of the best model, especially determining the best architecture and weights, remains a problem in ANN applications. Therefore various techniques have been proposed to determine the best ANN architecture. Some of these include constructive and pruning algorithm (Siestema and Dow, 1988), polynomial time algorithm (Roy et al., 1993), network information criterion (Murata et al., 1994), iterative construction algorithm (Rathbun et al., 1997), a method based on Box-Jenkins analysis (Buhamra et al., 2003), genetic algorithms (Dam and Saraf, 2006), the principle component analysis (Zeng et al., 2007), WIC (Egrioglu et al., 2008), a deletion/substitution/addition algorithm (Durbin et al., 2008), an architecture selection strategy for autoregressive seasonal time series (Aladag et al., 2008), design of experiments (Balestrassi et al., 2009), and tabu search (Aladag, 2009).

In this study, the ANN method is applied to a time series of international tourism arrivals to Turkey and the WIC proposed by Egrioglu, Aladag and Gunay (Egrioglu et al., 2008) is used to determine the best architecture. This time series consists of the number of monthly tourists visiting Turkey between January 2001 and December 2005. The forecasts are calculated by using ANN based on WIC. For comparison, the forecasts are also compared with other architectures which are determined by using other criteria such as root mean square error and mean absolute percentage error. In the next section, brief information about the usage of the ANN method in time series forecasting is given. Section 3 presents how the WIC works. The implementation is given in Section 4. The last section provides the obtained results and the discussion.

2. ANN METHOD IN TIME SERIES FORECASTING

ANN are mathematical models that imitate biological neural networks for the purpose of forecasting. ANN models have the ability of modeling both linear and non linear parts of time series. ANN consist of various components such as architecture, learning algorithm and activation function. Determining the components is a vital issue to get accurate forecasts. When ANN are used, one critical decision is to determine the appropriate architecture, that is, the number of layers, number of nodes in each layers and the number of arcs which interconnects with the nodes (Zurada, 1992). The elements of ANN can be briefly described as follows (Aladag et al., 2009b):

There are various types of artificial neural networks. One of these is called feed forward neural networks. Feed forward neural networks have been used successfully in many studies. In feed forward neural networks, there are no feedback connections. Figure 1 depicts the broad feed forward neural network architecture that has a single hidden layer and a single output.

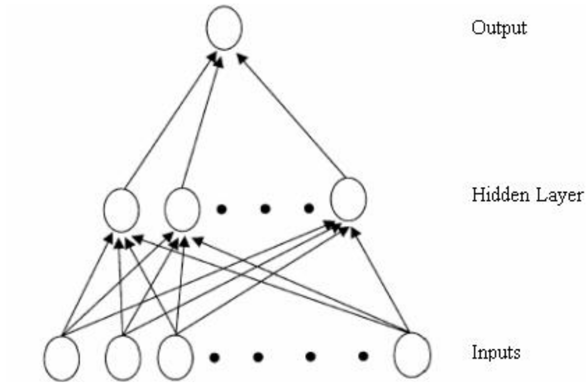


Figure 1. A broad feed forward neural network architecture

Training an artificial neural network for a specific task is equivalent to finding the values of all weights such that the desired output is generated by the corresponding input. Various training algorithms have been used for the determination of the optimal weights values. The most popular training method is the back propagation algorithm presented by Smith (2002). In the back propagation algorithm, training the artificial neural networks consists of adjusting all weights to minimize the error measure between the desired output and actual output (Cichocki and Unbehauen, 1993). Another element of artificial neural networks is the activation function. It determines the relationship between inputs and outputs of a network. In general, the activation function introduces a degree of non linearity that is valuable in most artificial neural networks applications. The most well known activation functions are logistic, hyperbolic tangent, sine (or cosine) and the linear functions. Among these, logistic activation function is the most popular one (Zhang et al., 1998).

3. THE WEIGHTED INFORMATION CRITERION (WIC)

In the literature, various model selection criteria have been employed when ANN models used for forecasting. Some well known criteria include AIC (Akaike, 1974), BIC (Schwarz, 1978), root mean squared error (RMSE), mean absolute percentage error (MAPE), direction accuracy (DA) criterion, and modified direction accuracy (MDA) proposed by Egrioglu et al. (2008). However, all of these criteria measure different properties of forecasting. Therefore, Egrioglu et al. (2008) considered the weighted sum of AIC, BIC, RMSE, MAPE, DA and MDA. Thus, it is possible to measure different properties by using one criterion. The algorithm of the model selection strategy based on WIC is given in Egrioglu et al. (2008) and summarized as follows:

Step 1: The number of possible architecture combinations is determined. For example, if the number of nodes of output layer is 1, number of nodes of input layer is 12 and number of nodes of hidden layer is 12 then the total number of architecture combinations is 144.

Step 2: The Best values of weights are determined by using training data according to the values of AIC, BIC, RMSE, MAPE, DA and MDA for test data.
Step 3: AIC, BIC, RMSE, MAPE, DA and MDA are standardized according to all the possible architecture combinations. For example, the 144 AIC values are standardized as follows:

$$AIC_i = \frac{AIC_i - \min(AIC)}{\max(AIC) - \min(AIC)}$$

Step 4: WIC is computed in the following way.

$$WIC = 0.1(AIC + BIC) + 0.2(RMSE + MAPE) + 0.2\left((1 - DA) + MDA\right) \quad (1)$$

Step 5: The architecture with the minimum WIC is chosen as the best architecture.

WIC criterion, where each criterion appears with its corresponding weight, is computed by using the formula given in Step 4. Since AIC and BIC often select the smallest models in which the number of hidden and input nodes are 1 or 2, their weights are assigned to be 0.1. Egrioglu et al. (2008) considered that the other criteria have same importance for forecasting accuracy so their weights are assigned to be 0.2.

4. THE IMPLEMENTATION

To illustrate this technique, the international tourism arrivals to Turkey time series is forecasted by ANN and the WIC is employed to determine the best ANN architecture. The examined monthly time series observed between January 2001 and December 2005 is shown in Figure 2. The last 6 observations are used for test set and the remaining observations are used for training. Logistic activation function is employed in all of the neurons and the Levenberg Marquardt method (Levenberg, 1944) is used as a training algorithm in the implementation. Both the number of input neurons and the number of hidden layer neurons are changed between 1 and 12 to find best architecture. Thus, AIC, BIC, RMSE, MAPE, DA, MDA and WIC criteria are calculated over the test set for 144 architectures. Then, the best architecture is selected through the WIC, RMSE and MAPE criteria. Other criteria are not employed for the selection since they are insufficient if used alone. Therefore, 3 architectures are determined. The determined architectures and their WIC, RMSE and MAPE values are shown in Table 1. The time series is also forecasted using the Box-Jenkins method and the best model is found to be SARIMA(1,1,1)(0,1,0).

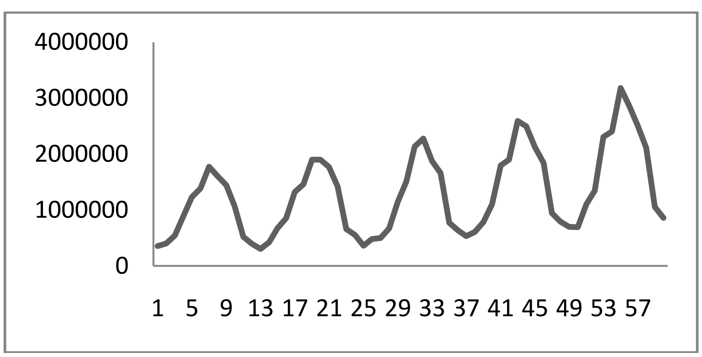


Figure 2. The international tourism arrivals to Turkey.

Table 1. The best architectures.

Used criterion	The best architecture	Criteria values		
		WIC	RMSE	MAPE
WIC	6-2-1	***0.0293***	203563	0.0085
RMSE	11-11-1	0.1681	***178293***	0.0050
MAPE	1-8-1	0.1727	721229	***0.0007***

According to Table 1, when WIC is used as performance measure, the best architecture is determined as 6-2-1 that is 6 and 2 neurons in the input layer and in the hidden layer, respectively and one neuron in the output layer. When this architecture is employed, the calculated WIC, RMSE and MAPE values are 0.0293, 203563 and 0.0085, respectively. In order to show the results visually, the forecasted values obtained from ANN and SARIMA models over the test set and the original values are given in Figures 3-6.

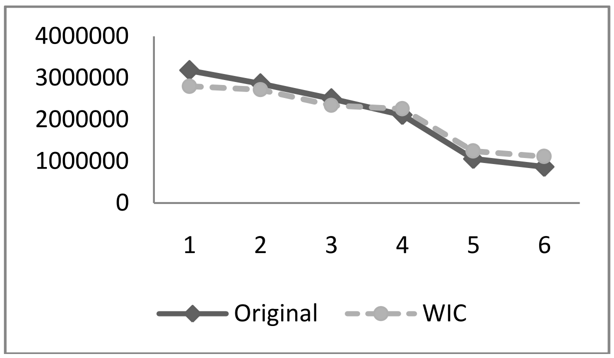


Figure 3. The forecasts obtained from 6-2-1 ANN determined by WIC together with the original values.

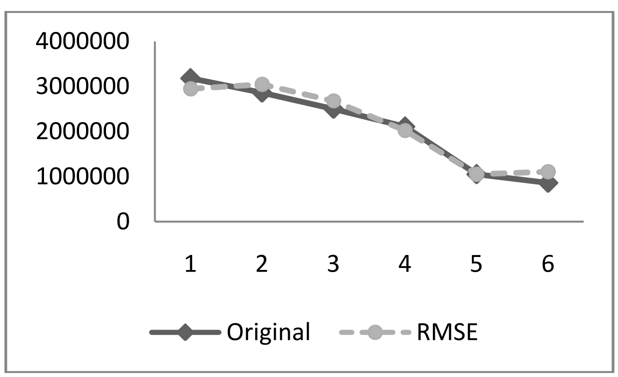


Figure 4. The forecasts obtained from 11-11-1 ANN determined by RMSE together with the original values.

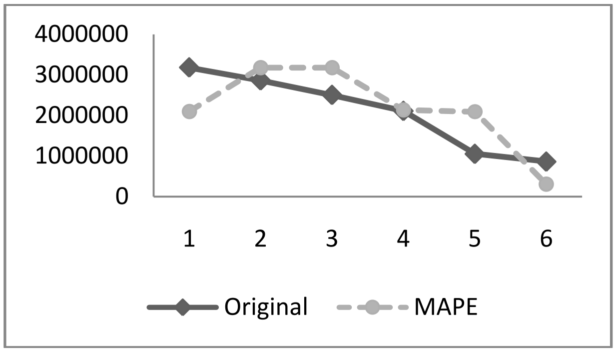


Figure 5. The forecasts obtained from 1-8-1 ANN determined by MAPE together with the original values.

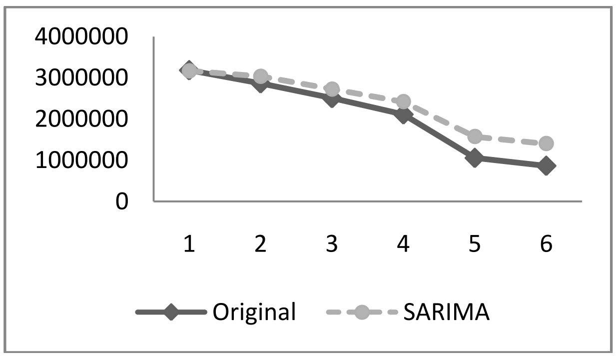


Figure 6. The forecasts obtained from SARIMA model together with the original values.

When these graphs are examined, it can be seen that the best results are obtained by using WIC and RMSE criteria and the third best forecasts are calculated from SARIMA model and the worst forecasts belongs to architecture 1-8-1 determined by MAPE. However, if the forecasts obtained from WIC and RMSE are examined, it is clearly seen that turning points are predicted better when the WIC is utilized to determine the best architecture. Beside this, in the architecture determined by WIC, the number of neurons in the hidden layer is less than the number that belongs to the architecture determined by RMSE so it can be said that WIC has selected architecture with greater generalization ability.

5. COMMENTS AND CONCLUSION

ANN method is an effective forecasting technique since it can model both the linear and non linear parts of time series. Therefore, this method has been successfully used in various forecasting applications in the literature. On the other hand, there are still some problems with using this approach. One of these problems is to determine the best architecture for obtaining accurate forecasts. There are various performance measures such as AIC, BIC, RMSE, MAPE, DA and MDA to find the best architecture. Since all of these criteria measure different properties of forecasting, Egrioglu et al. (2008) considered the weighted sum of AIC, BIC, RMSE, MAPE, DA and MDA. By doing this, Egrioglu et al. (2008) proposed a more consistent criterion for architecture selection. They also show that WIC is more consistent than RMSE criterion in their study. In this study, a time series of international tourism arrivals to Turkey is modeled by using ANN and SARIMA models. When ANN method is used, WIC, RMSE and MAPE are employed to determine the best architecture. All calculated forecasts are visually compared to original values and it is seen that the most accurate results are obtained when WIC and RMSE are used to determine the best architecture. In addition, the turning points are accurately predicted and the selected architecture has more generalization ability when WIC is employed.

REFERENCES

Akaike, H. (1974). A new look at statistical model identification. IEEE Transactions on Automatic Control AC-19, 716-723.

Aladag, C.H., Egrioglu, E., Gunay, S. (2008). A new architecture selection strategy in solving seasonal autoregressive time series by artificial neural networks. Hacettepe Journal of Mathematics and Statistics, 37(2), 185-200.

Aladag, C.H., Egrioglu, E., Kadilar, C. (2009a). Forecasting nonlinear time series with a hybrid methodology. Applied Mathematics Letters 22, 1467-1470.

Aladag, C.H., Basaran, M.A., Egrioglu, E., Yolcu, U. and Uslu, V.R. (2009b). Forecasting in high order fuzzy times series by using neural networks to define fuzzy relations. Expert Systems with Applications 36 (3), 4228–4231.

Aladag, C.H., (2009). Using tabu search algorithm in the selection of architecture for artificial neural networks. Ph.D. thesis, Hacettepe University.

Balestrassi, P.P., Popova, E., Paiva, A.P., Marangon Lima, J.W., (2009). Design of experiments on nn training for nonlinear time series forecasting. Neurocomputing, Volume 72, Issues 4-6, 1160-1178.

Buhamra, S., Smaoui, N., Gabr M., (2003). The box-jenkins analysis and neural networks: prediction and time series modelling. Applied Mathematical Modeling, 27, 805-815.

Bull, A. (1995), The economics of travel and tourism. Melbourne: Longman.

Cichocki, A. and Unbehauen, R. (1993). Neural networks for optimization and signal processing. John Willey & Sons, New York.

Dam, M., Saraf, D.N., (2006). Design of neural networks using genetic algorithm for on-line property estimation of crude fractionator products. Computers and Chemical Engineering, 30, 722–729.

Durbin, B., Dudoit, S., Van der Laan, M.J, (2008). A deletion/substitution/addition algorithm for classification neural networks with applications to biomedical data. Journal of Statistical Planning and Inference, 138, 464 – 488.

Egrioglu, E., Aladag, C.H. and, Gunay, S. (2008). A new model selection strategy in artificial neural network. Applied Mathematics and Computation 195, 591-597.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. The Quarterly of Applied Mathematics 2, 164–168.

Palmer, A., Montano, J.J., Sese, A. (2006). Designing an artificial neural network for forecasting tourism time series. Tourism Management 27, 781–790.

Rathbun, T.F., Rogers, S.K., DeSimo, M.P., Oxley, M.E., (1997). MLP iterative construction algorithm. Neurocomputing 17, 195-216.

Roy, A., Kim, L.S., Mukhopadhyay, S., (1993). A Polynomial time algorithm for the construction and training of a class of multilayer perceptrons. Neural Networks, 6, 535-545.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics 6, 461-464.

Siestema, J., Dow, R., (1988). Neural net pruning: why and how?. In: Procedings of the IEEE International Conferance on Neural Networks, 1, 325-333.

Smith, K.A. (2002). Neural networks in business: techniques and applications. Imprint Info Hershey: Idea Group.

Song, H., Li, G. (2008). Tourism demand modelling and forecasting - A review of recent research. Tourism Management 29 (2008) 203–220.

Zeng, J., Guo, H., Hu, Y., (2007). Artificial neural network model for identifying taxi gross emitter from remote sensing data of vehicle emission. Journal of Environmental Sciences, 19, 427–431.

Zhang, G., Patuwo, B.E. and Hu, Y.M. (1998). Forecasting with artificial neural networks: the state of the art. International Journal of Forecasting, 14, 35-62.

Zurada, J.M. (1992). Introduction of artificial neural systems. St. Paul: West Publishing.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 392–400.

HANDLING OVER-DISPERSION IN THE ANALYSIS OF CONTINGENCY TABLES WITH APPLICATION TO UNEMPLOYMENT RATES IN THE UNITED ARAB EMIRATES

Ibrahim M. Abdalla Al-Faki
College of Business and Economics
United Arab Emirates University
Al-Ain, UAE, P. O. Box 1755
E-mail: i.abdalla@uaeu.ac.ae

ABSTRACT

This paper compares several models for handling over-dispersion in count data arising from contingency tables. To demonstrate different alternatives, unemployment count data from the United Arab Emirates (UAE) is utilized to fit, test and compare competing models. In this context, the paper estimates the rate of unemployment in the UAE and discusses several factors that influence this rate. Counts of the number of times a particular event has occurred arise in many applied research situations and are often displayed as contingency tables of two or more variables and several categories. Inference and uses of categorical data methods are gaining growing importance. Recently, the log-linear modeling approaches are conveniently utilized to determine the nature of the casual relationships among variables in more complex higher dimensions contingency tables. Counts data under these models are often assumed to follow the Poisson distribution. In many situations, such data are sparse in nature and inhibited by over or under-dispersion, with the consequences of underestimating the standard error and type I error. This might produce misleading inference about model parameters. One solution to the problem is achieved by replacing the Poisson by the Negative binomial distribution. More recently, Bayesian methods provided alternatives for tackling the problem by allowing for the presence of over-dispersion by including random effects terms.

Keywords: Contingency table; Log-linear model; Poisson distribution; Random effects

1. INTRODUCTION

Contingency tables; often referred to as cross-classifications or cross-tabulations are tables of counts which describe the relationship between two or more variables/factors in a data-set. They often arise from several data collection efforts and in many applied research scenarios in the behavioral and the natural sciences. Traditionally these tables are analyzed by looking at departures from statistical independence by using the χ^2 test. In recent years, the log-linear models become the most popular approach which is conveniently utilized in the analysis of contingency tables.

The log-linear models are a broad class of methods that enable describing the effects of and interactions between the various factors in multidimensional categorical data. They are a specific type of generalized linear model for discrete valued data whose log-means are expressible as linear functions of parameters (Zelterman, 2002). The distribution for this discrete count data is often assumed to be a Poisson distribution.

A feature of many count data is the presence of over- or under-dispersion, alternatively termed as extra-variation. This is the situation where the variance of the response variable is respectively larger or smaller than the mean. Under the Poisson assumption the mean and the variance of the response variable are constrained to be equal. If the variance is not equal to the mean, the estimates in Poisson models are still consistent but inefficient. Therefore, inference based on the estimated standard errors is no longer valid (Wang and Famoye, 1997).

In practice over-dispersion is the norm and the occurrence of under-dispersion is very rare (Hilbe, 2007). Over-dispersion may arise from omitted covariates or some form of clustering in the original units, individual data may exhibit clustering effects if grouped, for example, by household, Congdon (2001). In behavioral and medical contexts, a generic source of over-dispersion is inter-subject variability in proneness or frailty. Hilbe (2007) discussed the Z and the Lagrange multiplier *post-hoc* tests for ascertaining real over-dispersion and detailed many sources of apparent over-dispersion.

Real over-dispersion in the Poisson model which is the subject of our paper can be accounted for in many ways, mainly by allowing for an independent modeling of the mean and variance by inclusion of an additional parameter. To that end, the most frequently used alternative model to the standard Poisson model is the Negative binomial (NB) (see for example Englin and Shonkwiler 1995, Winkelmann and Zimmermann 1995).

In the circumstances where the count data exhibits under-dispersion; the variance is smaller than the mean, neither the Poisson nor the Negative binomial models provide appropriate solution. Consul and Famoye (1992) proposed a generalized Poisson model (GP) which allows for over- as well as under-dispersion in the data. Other alternative models where the data inhibits no or excessive zeros were also discussed (see for example, Hilbe (2007) for hurdle models, zero-truncated or zero-inflated Poisson and negative binomial models). Recently, Bayesian methods are frequently used to handle over-dispersion by employing finite mixtures of the Poisson distribution with different assumptions concerning the mixing distribution.

To demonstrate different alternative models that capture over-dispersion, the 2005 UAE unemployment count data is exploited to fit, test and compare competing models. For model testing, the likelihood ratio test was employed to test for adequacy of the negative binomial and the generalized Poisson models over the standard Poisson model. The Akaike Information Criteria (AIC) and the Bayesian Schwarz Information Criteria (BIC) were used to compare the performance of alternative maximum likelihood and Bayes fit models.

The unemployment issue in the UAE and the other GCC countries is generally linked with fluctuation in oil prices. Recent drops in oil prices have lead to a decline in the share of non-UAE nationals in the labor force and an increase in the numbers of nationals looking for jobs, placing considerable burden on both the society and the economy (International Monetary Fund Report, 2005; Albuainain, 2005). According to the UAE Ministry of Planning mid-year estimates, the overall unemployment rate in the country stands at 3.0 percent of the total estimated labor force by the end of 2004. The rate is higher for nationals (11.4 percent) compared to non-nationals (2.1 percent), and for females (19.7 percent) compared to males (8.2 percent). These sharp differences across nationality and gender boundaries are elucidated mainly by the increase in the numbers of national graduates particularly females, the quality of education which does not meet the labor market demand, local traditions and social norms which do not accept certain jobs, and the work environment at the private sector with respect to low pay, working hours and close performance measurements which do not attract the locals (Albuainain, 2005).

Unemployment rates are usually estimated by dividing unemployment counts by the size of the economically active population or the work force. However, to account for the influence of other factors, a log-linear model with an *offset* term, the *log* of the size of the economically active UAE population, is used to produce an adjusted unemployment rate by incorporating a number of explanatory/independent factor variables in the model.

This paper is structured as follows. Section 2 outlines alternative models employed in the analysis, including the Poisson, the negative binomial, generalized Poisson, zero-truncated Poisson, zero-truncated negative binomial models and the random effects models within Bayesian framework. Sections 3, criteria for goodness of fit and model comparisons were outlined. In section 4, the presented models were used to analyze UAE unemployment count data. Section 5 gives concluding remarks.

2. MODELS FOR COUNT DATA

2.1 Poisson Model

Let the random variable Y represent the number of occurrences of an event, e.g. unemployment count, in the i^{th} class/cell of a contingency table, and the independent variables or factors X_i, $i = 1,2,\ldots,n$. If Y follows a Poisson distribution, the probability mass function (p.m.f.) for Y given X_i, $i = 1,2,\ldots,n$ is given as,

$$f(y \mid \lambda, x_1,\ldots,x_n) = \frac{\exp(-\lambda)\lambda^y}{y!}, \qquad y = 0, 1, 2\ldots \tag{1}$$

The mean and the variance of the Poisson distribution are equal, $E(Y) = Var(Y) = \lambda$. The Poisson model assumes that the y's are independent.

The Poisson and the negative binomial distributions both include zeros. Inspecting UAE unemployment data at hand (see Table 1), it is evident that the distribution of the data does not include zeros (zero-truncated). Such scenario might call for the utilization of zero-truncated Poisson models ($y > 0$) instead of the standard Poisson distribution ($y = 0,1,2,\ldots$), Zelterman, 2002.

2.2 Zero-Truncated Poisson Model (ZTP)

Suppose that $Y \sim \text{Poisson}(\lambda)$, with p.m.f. given by (1). The probability of zero ($Y = 0$) and positive counts ($Y > 0$) are

$$\Pr(Y = 0 \mid \mu) = \exp(-\lambda) \tag{2}$$
$$\Pr(Y > 0 \mid \mu) = 1 - \exp(-\lambda) \tag{3}$$

The conditional probability of observing y events given that $y > 0$ is obtained as

$$\Pr(Y = y \mid Y > 0) = \frac{\Pr(Y = y \quad \text{and} \quad Y > 0)}{\Pr(Y > 0)} = \frac{f(y \mid \lambda, x_1, \ldots, x_n)}{\Pr(Y > 0 \mid \lambda)}, \qquad (3)$$

hence the p.m.f. of the zero-truncated Poisson distribution is expressed as

$$f(y \mid y > 0, \lambda, x_1, \ldots, x_n) = \Pr(Y = y \mid Y > 0) = \frac{\exp(-\lambda)\lambda^y}{y![1 - \exp(-\lambda)]}, \qquad y = 1, 2, 3\ldots \qquad (4)$$

i.e. the p.m.f. of the zero-truncated Poisson distribution is obtained by normalizing the standard Poisson p.m.f. by the factor $(1 - e^{-\lambda})$ to force the truncated p.m.f. to sum to 1. The mean and variance of the zero-truncated Poisson distribution are given as

$$E(Y \mid Y > 0, \lambda) = \frac{\lambda}{1 - \exp(-\lambda)}, \qquad \mathrm{Var}(Y \mid Y > 0, \lambda) = \frac{\lambda}{1 - \exp(-\lambda)}\left[1 - \frac{\lambda}{\exp(-\lambda) - 1}\right]$$

2.3 Negative Binomial Model (NB)

Following similar lines, suppose that $Y \sim \mathrm{Poisson}(\lambda)$, and that λ itself is a random variable with a gamma distribution,

$$\begin{aligned} Y \mid \lambda &\sim \mathrm{Poisson}(\lambda) \\ \lambda &\sim \mathrm{Gamma}(\alpha, \beta) \end{aligned} \qquad (5)$$

where $\mathrm{Gamma}(\alpha, \beta)$ is the gamma distribution with mean $\alpha\beta$ and variance $\alpha\beta^2$. The marginal or unconditional distribution of Y is therefore the negative binomial,

$$f(y \mid \alpha, \beta) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha) y!}\left(\frac{\beta}{1 + \beta}\right)^y \left(\frac{1}{1 + \beta}\right)^\alpha, \qquad y = 0, 1, 2, \ldots \qquad (6)$$

with mean $E(Y) = \alpha\beta$ and variance $\mathrm{Var}(Y) = \alpha\beta + \alpha\beta^2$.

To build a regression model, the Negative Binomial distribution is usually expressed in terms of the parameters $\lambda = \alpha\beta$ and $\alpha = 1/k$ so that

$$\begin{aligned} E(Y) &= \lambda \quad \text{and} \\ \mathrm{Var}(Y) &= \lambda + k\lambda^2 \end{aligned}$$

The distribution of Y, p.m.f., therefore becomes,

$$f(y \mid \lambda, k, x_1, \ldots, x_n) = \frac{\Gamma(1/k + y)}{\Gamma(1/k) y!}\left(\frac{k\lambda}{1 + k\lambda}\right)^y \left(\frac{1}{1 + k\lambda}\right)^{1/k}, \qquad y = 0, 1, 2, \ldots \qquad (7)$$

which approaches the Poisson distribution with mean λ as $k \to 0$, k denotes the over-dispersion parameter.

2.4 Zero-Truncated Negative Binomial Model (ZTNB)

The logic for the zero-truncated negative binomial is the same as for the zero-truncated Poisson. In addition to allowing for data that excludes zero counts, it also handles over-dispersion. The probability of a zero count is $(1+k\lambda)^{-1/k}$

hence normalizing the p.m.f. of the negative binomial distribution (7) by the factor $[1-(1+k\lambda)^{-1/k}]$ representing the probability of non-zero counts ($Y > 0$), leads to the p.m.f. of the zero-truncated negative binomial

$$f(y \mid \lambda, k, x_1, \ldots, x_n) = \frac{\Gamma(1/k + y)}{\Gamma(1/k) y!} \left(\frac{k\lambda}{1+k\lambda}\right)^y \left(\frac{1}{1+k\lambda}\right)^{1/k} \times \left(\frac{1}{1-(1+k\lambda)^{-1/k}}\right), \qquad (8)$$

$y = 1, 2, 3, \ldots$, with mean $E(Y)$, given as

$$E(Y) = \frac{\lambda}{1-(1+k\lambda)^{-1/k}}$$

2.5 Generalized Poisson Model (GP)

A generalized Poisson distribution can fit both over-dispersion, $Var(Y) > E(Y)$, and under-dispersion, $Var(Y) < E(Y)$. The p.m.f. for the generalized Poisson distribution as given by Wang and Famoye (1997) is

$$f(y \mid \lambda, k, x_1, \ldots, x_n) = \left(\frac{\lambda}{1+k\lambda}\right)^y \frac{(1+ky)^{y-1}}{y!} \exp\left(-\frac{\lambda(1+ky)}{1+k\lambda}\right), \qquad y = 0, 1, 2, \ldots \qquad (9)$$

with mean $E(Y) = \lambda$ and variance $Var(Y) = \lambda(1+k\lambda)^2$. If $k = 0$, the generalized Poisson reduces to the Poisson distribution with $E(Y) = Var(Y) = \lambda$. For $k > 0$, the variance is larger than the mean, $Var(Y) > E(Y)$, and the distribution represents count data with over-dispersion. If $k < 0$, the variance is smaller than the mean, $Var(Y) < E(Y)$, and the distribution represents count data with under-dispersion.

2.6 Random Effects Poisson Models– Bayesian Framework

Over-dispersion due to unobserved confounders will usually not be captured by simple covariate models and often it is appropriate to include some additional terms in a model which can capture such effects (Lawson, 2009). Using a Bayesian framework, over-dispersion can be handled utilizing hierarchical Poisson mixture models by either a) including a prior distribution for the Poisson parameter using for example the Poisson-gamma model given by equation (5) with a marginal likelihood expressed by equation (7), or b) by the extension of the linear predictor term of the standard Poisson model to include an extra random effect, usually formulated by the Poisson-log-normal model with added normally distributed random effect, r,

$$
\begin{aligned}
Y &\sim \text{Poisson}(\lambda) \\
\log(\lambda) &= \mathbf{X}\beta + r \\
r &\sim N(0, \sigma_r^2)
\end{aligned}
$$

3. GOODNESS OF FIT AND MODEL COMPARISONS

The Akaike Information Criteria (AIC) and the Bayesian Schwarz Information Criteria (BIC) are used to compare the performance of alternative maximum likelihood models, even models that are not nested. The smaller AIC and BIC is associated with the better model.

$$AIC = -2L + 2p,$$

$$BIC = -2L + p\,\mathbf{log}(n),$$

where L denotes the log-likelihood evaluated under λ, p is the number of parameters and n is number of categories (classes/cells).

4. ESTIMATING UNEMPLOYMENT RATE IN THE UAE

The models proposed in section (2) will be used in this section to analyze unemployment data in the UAE in 2005. The data (see Table 1) is based on 52159 cases of unemployed individuals officially registered by the Ministry of Economy in the UAE in 2005. Given the number of unemployed individuals together with the size of the economically active population within each age, nationality and education group, the focus is to estimate the rate of unemployment in the UAE in the year 2005. Initial exploration of the cross-classified data reveals a substantial degree of over-dispersion, unemployment counts range from a minimum of 3 to a maximum of 13045. The variance of the data is 6839380, substantially higher than the mean, 1304.

In addition to modeling over-dispersion in the data, unemployment rate within each cell of Table 1 is estimated. Gender, education and age are included as factor covariates in the model and the size of the economically active population in each cell is used as an offset in the analysis. Extra heterogeneity in the data, not explained by these factors is handled by model specific dispersion parameters in the NB and the GP (Gschlobl and Czado, 2006) and the random effects Poisson models, the Poisson log-normal model (PLN) and the Poisson-gamma mixture model (PG). Both the PLN and PG models were fitted using Bayesian methods. Remaining over-dispersion in the NB and the GP models can be interpreted as unobserved heterogeneity among observations. In view of the absence of zero counts in UAE unemployment data (see Table 1), an attempt was made to fit a zero-truncated Poisson (ZTP) and a zero-truncated negative binomial (ZTNB) models to the data, the latter model allows for the absence of zeros as well as for over-dispersion in the data.

Table 1: UAE Unemployment Data 2005 (UAE Ministry of Economy)

						Econ. active	Count
Gender	M	Education	Secondary	Age	15-24	232077	13045
					25-34	813755	8783
					35-49	636606	3516
					50-59	103978	759
					60+	12943	221
			Some post	Age	15-24	9233	478
					25-34	38576	414
					35-49	34811	195
					50-59	7937	38
					60+	961	8
			College/University	Age	15-24	12093	1338
					25-34	102961	1904
					35-49	129101	822
					50-59	35092	176
					60+	5612	38
			Postgraduate	Age	15-24	759	63
					25-34	9871	158
					35-49	18675	111
					50-59	6702	39
					60+	1615	7
	F	Education	Secondary	Age	15-24	61067	5748
					25-34	99659	3372
					35-49	49861	1250
					50-59	4822	235
					60+	699	53
			Some post	Age	15-24	4959	844
					25-34	12426	628
					35-49	8289	189
					50-59	1385	16
					60+	83	6
			College/University	Age	15-24	10737	2812
					25-34	47873	3467
					35-49	29368	796
					50-59	4114	94
					60+	343	10
			Postgraduate	Age	15-24	607	124
					25-34	3915	269
					35-49	4036	112
					50-59	934	18
					60+	120	3

The maximum likelihood approach was used to estimate parameters of the standard Poisson model, using gender, education level and age as factor covariates, including an intercept and an offset represented by the log of the size of economically active population within each group. Analysis results indicate that males have significantly lower unemployment rate compared to females. As the level of education increases, the rate of unemployment decreases. Those with university level or above have lower unemployment rate compared to secondary education level. Age is a significant factor that accounts for heterogeneity in unemployment rate in the UAE, younger individuals have significantly higher unemployment rate compared to older people, fifty years or older. As can be noted from Table 2, the fitted Poisson model is highly over-dispersed (model $\text{d.f.} = 31$ substantially greater than the Pearson chi-square, 2642.4). To account for over-dispersion, the NB, ZTNB, GP, the PLN(Bayes) and PG (Bayes) models were utilized using the same factors used in the standard Poisson model. The NB and GP models resulted in larger standard errors, leading to reduced significance of some factor levels and declaring some education levels as statistically insignificant in determining unemployment rate in the UAE. The likelihood ratio test of the Poisson against the NB and the GP models (Table 2) revealed highly significant test statistics of 2387.6 and 2411.0, respectively, implying that both the NB and GP provide better fit to the data than the standard Poisson model. These results are also apparent in Table 2 when comparisons are based on the AIC and the BIC criteria. Zero-truncated Poisson and negative binomial models (ZTP and ZTNB) provided no better fit compared to the Poisson and the NB models. Hilbe (2007) noted that the theoretical probability of obtaining a zero count is high and the effect of truncation is substantial whenever the mean count is low. For the UAE data, the mean count of unemployed individuals is high, 1304, suggesting a low probability of obtaining a zero count. Hence a truncated model seems to provide no better alternative (see Table 2). The Poisson-gamma mixture Bayes model (PG), on the other hand, seems to outperform all given alternatives, lower AIC and BIC**.**

Table 2: Model Goodness of Fit Measures

Goodness of Fit	**Model**						
	Poisson	**NB**	**GP**	**ZTP**	**ZTNB**	**PLN**	**PG**
Deviance	2559.6	41.3	38.7	2559.6	41.2	40.9	39.5
Chi-square	2642.4	39.8	32.5	2642.5	39.8	41.1	40.0
Log-likelihood	381191.2	382385.0	382396.7	381191.2	382385.1	382500.0	382500.0
AIC	2868.1	482.4	459.0	2868.0	462.2	349.4	348.0
BIC	2883.3	499.3	475.9	2883.2	499.1	364.6	363.2

5. SUMMARY AND CONCLUSION

The paper discussed several alternative models that allow for over-dispersion in unemploment count data in the UAE. Over-dispersion was modeled by adding an extra parameter in the NB, ZTNB and GP models and using hierarchical Poisson mixure models. The maximum likelihood approach was used to estimate parameters of the models and a Bayesian frame work was employed in fitting parameters of the random effects Poisson model. Significantly better fit was obtained using the models allowing for over-dispersion. Among these the random effects Poisson models fitted the data best. Application of suggested models on the UAE unemployment data identified gender, education level and age as significant contributors in estimating unemployment rate in the UAE.

REFERENCES

Albuainain, R. M. (2005). Unemployment rate in the United Arab Emirates: The case of Abu-Dhabi, Working paper 0404, The General Woment Union, Abu-Dhabi, UAE.

Congdon, P. (2001). Bayesian statistical modeling, John Wiley & Sons, LTD, New York.

Consul, P. C. and Famoye, F. (1992). Generalized Poisson regression model, Communication in Statistics, Theory and methods, 21(1): 89-109.

Englin J, Shonkwiler JS (1995) Estimating social welfare using count data: An application to long-run recreation demand under conditions of endogenous stratification and truncation. *Review of Economics and Statistics* 77:104–112.

Gschlobl, S. and Czado, C. (2006). Modeling count data with over-dispersion and spatial effects, *Institute for Statistics*, Sonderforschungsbereich 386, paper 475.

Hilbe, J. M. (2007). Negative Binomial Regression, Cambridge University Press, Cambridge, UK.

International Monetary Fund Report. (2005). United Arab Emirates: Selected Issues and Statistical Appendix, IMF Country Report No. 05/268, Washington, D.C. 20431.

Lawson, A. B. (2009). Bayesian disease mapping: Hierarchical modeling in spatial epidemiology, Chapman & Hall/CRC, Taylor & Francis Group.

Wang, W., Famoye, F. (1997). Modeling household fertility decisions with generalized Poisson regression, Journal of Population Economics 10:273–283.

Winkelmann R, Zimmermann KF (1995) Recent developments in count data modeling: Theory and application. *Journal of Economic Surveys* 9:1–24.

Zelterman, D. (2002). Advanced log-linear models using SAS, Cary, NC; SAS Institute Inc.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 401–408.

DURATION DISTRIBUTION OF CONJUNCTION OF TWO GAUSSIAN PROCESSES: TANGENT LINE AND QUADRATIC METHODS

M. T. Alodat[1], M. Y. Al-Rawwash[2] and M. A. AL-Jebrini[3]
Department of Statistics, Yarmouk University, Irbid, Jordan
E-mail: [1]malodat@yu.edu.jo, [2]rawwash@yu.edu.jo, and [3]mjebrini@hotmail.com

ABSTRACT

The idea of this article is to obtain an approximation to the duration distribution of the conjunction of two Gaussian processes above a threshold u. We use the tangent line method to approximate each process by a Tangent line at $t_0 = 0$. As an alternative approximation technique we use the quadratic polynomial at $t_0 = 0$ to carry out the approximation at a high level u Moreover, we compare our approximations obtained via tangent line method and quadratic polynomial method with the exact one using computational techniques. Finally, we apply our findings to the meteorological data sets.

Keywords: Duration of Conjunction; Gaussian process; Quadratic Method; Tangent Line Method; Up-crossing.

1. INTRODUCTION

Gaussian processes have been used extensively in the literature to model many random responses in various areas of applications such as environment and engineering. The theory developed in this paper is motivated and applied to meteorological data sets collected from National Center for Atmospheric Research where the sea level and the behavior of the wave are good examples of Gaussian processes. As a result, the probability that the wave or the sea level exceeds a given threshold is considered as one of the main and crucial monitoring figures. A stationary and differentiable Gaussian random process $X(t)$ $t \in [0, A]$ with first derivative $\dot{X}(t)$ is said to have an up-crossing of u at $t_0 \in [0, A]$ if $X(t_0) = u$ and $\dot{X}(t_0) > 0$. Also, $X(t)$ is said to have a down-crossing of u at $t_0 \in [0, A]$ if $X(t_0) = u$ and $\dot{X}(t_0) < 0$. The set of all points t for which $X(t)$ exceeds a level u is called the excursion set of $X(t)$ above u. Now, let N_u denote the number of up-crossings of u by the process $X(t)$. Then, the mean of N_u is considered a quite accurate approximation of $P(sup_{t\in[0,A]}X(t) \geq u)$; in other words $E(N_u) \approx P(sup_{t\in[0,A]}X(t) \geq u)$. The length of the interval between an up-crossing and the subsequent down-crossing of a level u is called the duration of the excursion of $X(t)$.

Adler (1981) shows that the durations of a stationary Gaussian process are independently distributed as $Exp(\frac{1}{u})$ when $u \to \infty$. The conjunction of two Gaussian random processes $X_1(t)$ and $X_2(t)$ $t \in [0, A]$ is another Gaussian random process $Y(t)$ defined by $Y(t) = min(X_1(t), X_2(t))$. The excursion set of the conjunction above the threshold u is the intersection of the excursion sets of the processes $X_1(t)$ and $X_2(t)$ above u. Alodat et al. (2010) studied the duration distribution of the conjunction of two independent F processes. They obtained an approximation for the duration distribution of excursion set generated by the minimum of two independent F processes above a high threshold u. In this paper, we plan to obtain an approximation to the duration distribution of conjunction using the tangent line method (TLM) as well as the quadratic polynomial method (QM).

The article is structured as follows. In section 2, we introduce our idea and review some of the work of Adler (1981) and Leadbetter and Spaniolo (2002). In section 3, we derive the distribution formula of the duration and its density and we obtain a closed form of the mean value of the duration. Moreover, an approximation to the duration distribution of two Gaussian random processes using Quadratic Method is obtained in section 4. A simulation study is conducted in section 4 to compare the TLM and QM approximations with the exact one for each approach. In section 5, we apply our approximation TLM to real data sets especially meteorological data sets. Eventually, we present our comments and conclusions of our research in section 6.

2. TANGENT LINE METHOD

Let $X_1(t)$ and $X_2(t)$ be two independent Gaussian processes such that $X_1(t)$ have an up-crossing of u at $t_0 = 0$ and $X_2(t)$ have a down-crossing of u at $t_0 = 0$ i.e. $\dot{X}_1(0) > 0$ and $\dot{X}_2(0) < 0$. Also, assume that Y represents the excess height of $Y(t)$ as shown in Figure 1.

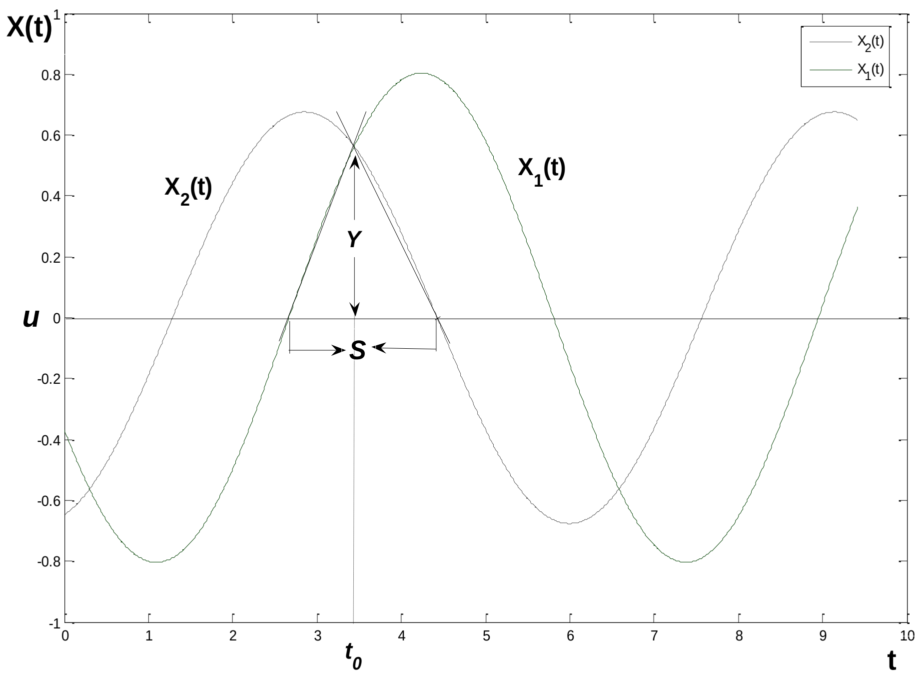


Figure 1: Excursion sets and duration of the conjunction of two Gaussian random processes.

In this section, we approximate the Gaussian processes $X_1(t)$ and $X_2(t)$ at $t_0 = 0$ using TLM as follows $\tilde{X}_1(t) = X_1(0) + t\dot{X}_1(0)$ and $\tilde{X}_2(t) = X_2(0) + t\dot{X}_2(0)$, respectively. Thus, solving the equations $\tilde{X}_1(t) = u$ and $\tilde{X}_2(t) = u$ yields the roots t_1 and t_2 respectively, where $t_1 = -\frac{Y_1}{W_1}$,

$t_2 = \frac{Y_2}{W_2}$, $Y_1 = X_1(0) - u$, $Y_2 = X_2(0) - u$, $W_1 = \dot{X}_1(0)$ and $W_2 = -\dot{X}_2(0)$. According to Adler (1981), the conditional distribution of $u[X(t_0) - u]$ is $exp(1)$ as $u \to \infty$ provided that $X(t)$ is a Gaussian process with a local maximum at $t = t_0$ and height exceeding u. Therefore, we may conclude that Y_1 and Y_2 are independent and identical distributed as exponential with mean $1/u$ as $u \to \infty$. Moreover, assuming that $X(t)$ is a stationary Gaussian process Leadbetter and Spaniolo (2002) introduced the Palm distribution of $\dot{X}(t)$ after an up-crossing of u as $f(\dot{x}) = \frac{\dot{x}}{\lambda} exp(-\frac{\dot{x}^2}{2\lambda})$, for $\dot{x} > 0$ and zero otherwise. The result of Leadbetter and Spaniolo (2002) allows us to conclude that W_1 and W_2 are independent and identical having Rayleigh distribution with probability density function given as

$$f(w_i) = \begin{cases} \frac{w_i}{\lambda} exp(-\frac{w_i^2}{2\lambda}) & w_i > 0. \\ 0 & \text{otherwise.} \end{cases}$$

Now we intend to approximate the duration of the conjunction of $X_1(t)$ and $X_2(t)$ by

$$S = t_2 - t_1 \approx \frac{Y}{W_2} + \frac{Y}{W_1} = Y\left(\frac{1}{W_2} + \frac{1}{W_1}\right) \tag{1}$$

where $Y = sup_{t\in[0,A]} Y(t)$. To find the distribution of S, we need to obtain the conditional distribution of Y given that $Y > 0$ as follows:

$$P[sup_{t\in[0,\,A]} Y(t) > u + v \mid sup_{t\in[0,\,A]} Y(t) > u] = \frac{P(sup_{t\in[0,\,A]} Y(t) > u + v)}{P(sup_{t\in[0,\,A]} Y(t) > u)} \approx \frac{E(N_{u+v})}{E(N_u)}.$$

Independence of $X_1(t)$ and $X_2(t)$ and Rice's formula (see Leadbetter and Spaniolo (2002)) imply that $\frac{E(N_{u+v})}{E(N_u)} \approx \left[\frac{1-\Phi(u+v)}{1-\Phi(u)}\right]^2$, where $\Phi(u)$ and $\phi(u)$ are the cumulative distribution function and the probability density function of *N(0,1)*, respectively. Now, using the fact that $1 - \Phi(u) \approx \frac{\phi(u)}{u}$ for large u, we have that

$$P(sup_{t\in[0,\,A]} Y(t) - u > v \mid sup_{t\in[0,\,A]} Y(t) > u) \approx \left[e^{-\frac{(u+v)^2}{2} + \frac{u^2}{2}}\right]^2 \approx e^{-2uv}.$$

Hence, the cumulative distribution and the probability density function of $(Y - u|Y > u)$ are given by $G(v) = 1 - e^{-2uv}$ and $g(v) = 2ue^{-2uv}$, respectively. To this end, we conclude that $(Y - u)$ is the excess height above u of a local maximum of the conjunction of two independent and identically distributed Gaussian processes $X_1(t)$ and $X_2(t)$. Also, we conclude that $(Y-u)$ is exponentially distributed with parameter $(\frac{1}{2u})$. Thus, we plan to use this result in the sequel to obtain the approximation of the duration distribution of two independent Gaussian processes. Based on equation (1), it is easy to see that the mean and variance of the duration S are $E(S)=2E(Y)E(\frac{1}{W_1})=\frac{\sqrt{2\pi\lambda}}{\lambda u}$ and $Var(S)=\infty.$

3. QUADRATIC METHOD APPROXIMATION

In this section, we follow the footsteps of Aludaat and Alodat (2007) to approximate the duration distribution of the two Gaussian random processes $X_1(t)$ and $X_2(t)$ above a high threshold $u.$ We use the quadratic method to carry out the approximation at some point t_0 that satisfies $X_1(t_0)>u$ and $X_2(t_0)>u.$ Figure 2 clearly illustrates the duration S as the difference between the minimum of the upper roots and the maximum of the lower roots, where R_{11} and R_{12} are the roots of $X_1(t)=u,$ while R_{21} and R_{22} are the roots of $X_2(t)=u,$ respectively when $u\to\infty.$

The behavior of the process $X(t)$, above high threshold, is deterministic in shape and has the representation

$$X(t)\approx X(t_0)+\dot{X}(t_0)(t-t_0)-\frac{u\lambda}{2}(t-t_0)^2, \tag{2}$$

where t_0 is a local maximizer of $X(t)$ and $\lambda=Var(\dot{X}(0))$. Applying equation (2) to $X_1(t)$ and $X_2(t),$ and solving the equations $u\approx X_1(t)$ and $u\approx X_2(t)$ for t, we get

$$R_{i1}\approx\frac{\dot{X}_i-\sqrt{\dot{X}_i^2+2u\lambda Y_i}}{u\lambda} \text{ and } R_{i2}\approx\frac{\dot{X}_i+\sqrt{\dot{X}_i^2+2u\lambda Y_i}}{u\lambda},$$

where $Y_i=X_i(0)-u$, $i=1,2$ are independent and identically distributed as $exp(1/u)$. For each j, $\dot{X}_j(0)$ has a Rayleigh distribution and the duration of the conjunction of $X_1(t)$ and $X_2(t)$ is approximated via $S=min_{i=1}^2(R_{i2})-max_{i=1}^2(R_{i1}).$ To accomplish this mission, we need to derive the joint pdf of the previously obtained roots corresponding to one of the Gaussian processes say $X_1(t)$, i.e. R_{11} and R_{12}.

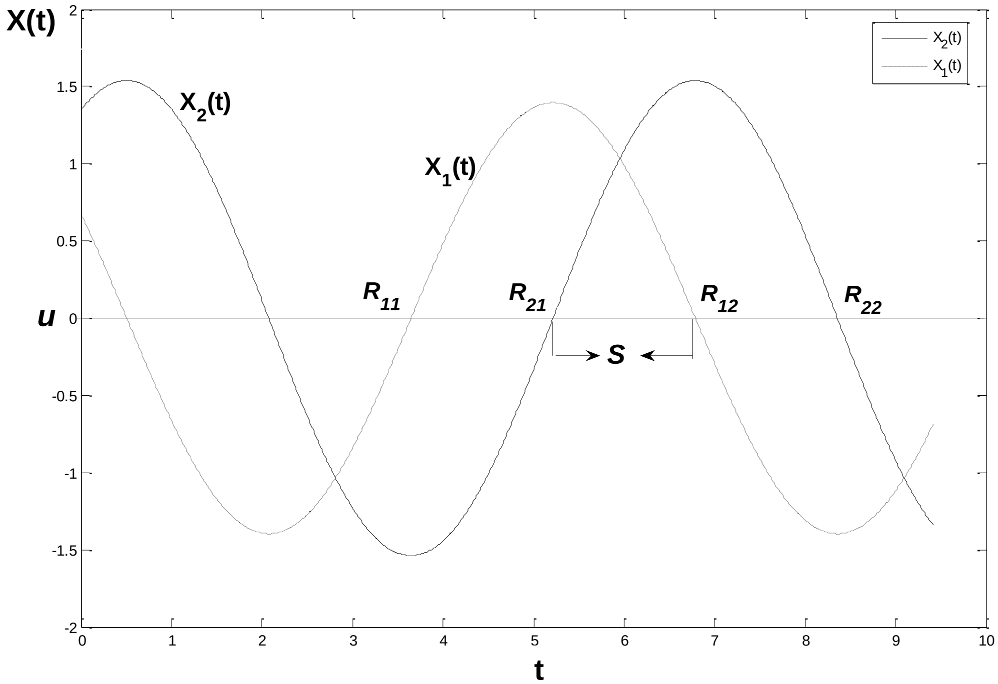


Figure 2: Two Gaussian random processes above high level u.

4. SIMULATION

In this section, we conduct a simulation study to compare our approximations obtained via TLM and QM with the exact ones by considering two processes for each approach. We obtained large samples each of size 5000 observations from the approximate distribution of S for different threshold values $u = 1.0$ and $2.5.$ On the other hand; a large sample from the exact distribution of S is obtained using simulated Gaussian processes. Comparison is conducted based on the cumulative distribution function (CDF) obtained from the exact and approximated distributions for different threshold values. The results concerning our approximation as well as the exact results are shown in Table 1 and Figures 3-4. The results show that the TLM approximation is very closed to the exact one and it works better than QM approximation for different levels of u, however the QM works better than TLM when $u = 2.5$ as shown in Figure 4(a, b). The results are applied to the case of two processes throughout this simulation study.

Table 1: The means for the exact one, TLM and QM approximations

u	Exact	TLM	QM mean
0.5	1.1147	1.6922	2.4585
1.0	0.8486	0.8156	1.2441
1.5	0.6789	0.5722	0.8253
2.0	0.5730	0.5277	0.6270
2.5	0.4896	0.3351	0.4931

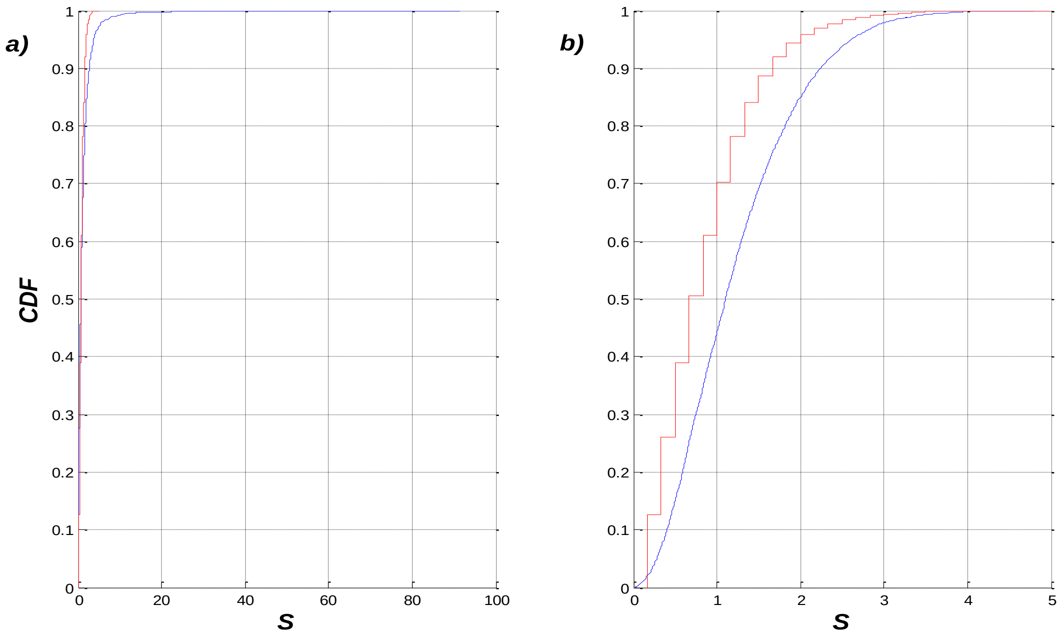


Figure 3: The exact (stairs) CDF of two processes for $u = 1.0$. (a) TLM approximation (smooth) CDF of S. (b) QM approximation (smooth) CDF of S.

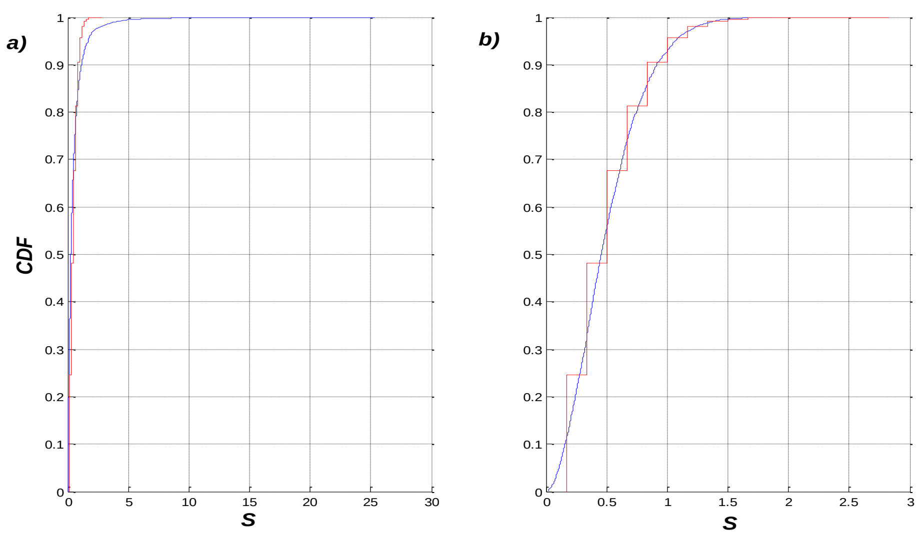


Figure 4: The exact (stairs) CDF of two processes for $u = 2.5$. (a) TLM approximation (smooth) CDF of S. (b) QM approximation (smooth) CDF of S.

5. APPLICATION

In this section, we apply the theory developed in this paper to meteorological data sets collected from National Center for Atmospheric Research. The data represented in the monthly values of the sea level pressure (SLP) during 1866-2007. This series consists of 1692 observations on the SLP during 141 years, computed as "the difference of the departure from the long-term monthly mean sea level pressures" at Tahiti in the South Pacific and Darwin in Northern Australia. The fact that most of the observations in the last part of the series take negative values are related to a recent warming in the tropical Pacific and this pattern has been the cornerstone of many recent studies. A Kolmogrov-Smirnov goodness of it test was conducted for both data sets and showed high p-values. Also Figure 4 gives us the intuition of having stationary data sets. We carry on the mission of extracting the duration of the conjunction of the SLP data sets using a threshold of $u = 1.0$. Since the distribution of the duration of the excursion set of one Gaussian process is the same as the distribution of $2\sqrt{\frac{2Y}{u\lambda}}$ where $Y \sim exp(\frac{1}{u})$, then the duration average is $\frac{\sqrt{2\pi}}{u\sqrt{\lambda}}$. Let $S_{11},\ S_{12},...,S_{1n_1}$ denote the observed durations of the excursion for Tahiti data, then the method of moment estimator of λ is $\hat{\lambda} = n_1\sqrt{2\pi}/uS_1$, where $S_1 = S_{11} + S_{12} + ... + S_{1n_1}$ is distributed as $\text{Gamma}(n_1, \frac{\sqrt{2\pi}}{u\sqrt{\lambda}})$. Moreover, using the δ-method we intend to give an approximation of the variance of $\hat{\lambda}$. To accomplish this, we let $g(x) = n_1\sqrt{2\pi}/ux$ and the approximated variance of $\hat{\lambda}$ will be $Var(\hat{\lambda}) \approx g'(E(S_1))^2 Var(S_1) = \frac{\lambda}{n_1}$. The extreme value of the SLP data sets is an essential random variable in climatology; therefore its probability distribution is highly required. Note that above the threshold $u = 1.0$, the excursion set for Tahiti and Darwin have the following durations: 1 and 2 (in months) with frequencies 23 and 1, respectively. These durations are used to produce the mean duration distribution of

$$E(S) = \frac{\sqrt{\pi}}{u\sqrt{2\hat{\lambda}}} = 0.8080.$$

The smooth parameter happens to be $\sqrt{\hat{\lambda}} = 1.5512$ and its variance is $Var(\hat{\lambda}) = 0.1003$. This means that the SLP process or data set spends (after an up-crossing of $u = 1.0$) on the average 0.8080.

6. CONCLUSION

In this paper, we proposed a new approximation method called the Tangent Line Method which showed a slightly better performance above a high threshold u compared to other approaches derived for Gaussian processes. Another approximation method called the quadratic method is proposed for implementation in the case of at least two processes under investigation. We compared the two approximations with the exact one based on the performance of their CDF's via simulation as shown in the Figures 3-4. The TLM appears to be more accurate than QM when we compare the results with the exact one, while QM works better than TLM at the threshold $u = 2.5$. We derived the distribution of the excess height above a high threshold u when we approximate each process using TLM. Finally, we may conclude that TLM approximation is better than QM approximation for small values of u.

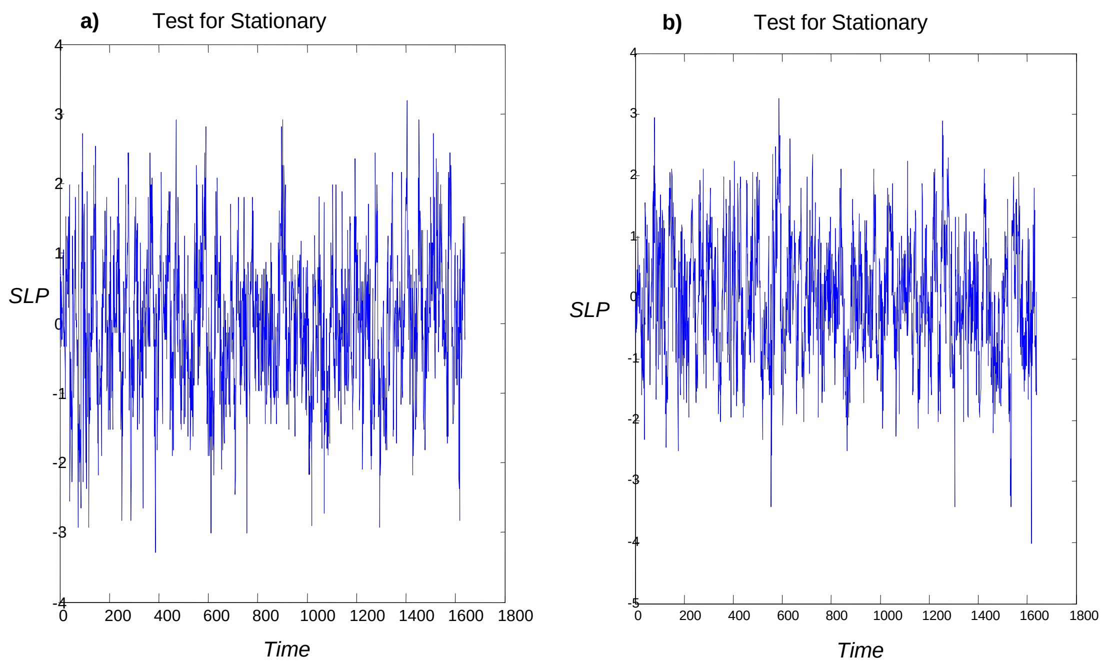


Figure 10: The behavior of sea level pressure: a) At Darwin b) At Tahiti.

REFERENCES

Adler, R. (1981). *The Geometry of Random Fields*. John Wiley and Sons, New York.

Alodat, M. T., Al-Rawwash, M. and Jebrini, M. A. (2010). Duration Distribution of Conjunction of Two Independent *F* Processes. *Journal of Applied Probability*, 47, 1–12.

Aludaat, K. M. and Alodat, M. T. (2007). Duration Distribution of a conjunction. *Applied Mathematical Sciences*, Vol. **1**, 2007, No. **38**, 1879–1890.

Leadbetter, M. R. and Spaniolo, G.V. (2002). On statistics at level crossing by stationary process. *Statistica Neerlandica*, **56**, nr.2, pp.152–164.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 409-423.

ESTIMATION FOR STATE-SPACE MODELS: QUASI-LIKELIHOOD APPROACH

Raed Alzghool
Department of Applied Science
Faculty of Prince Abdullah Ben Ghazi for Science and Information Technology
Al-Balqa' Applied University
Al-Salt, Jordan
E-mail: raedalzghool@bau.edu.jo

Yan-Xia Lin
School of Mathematics and Applied Statistics
University of Wollongong
Wollongong, NSW 2500, Australia
E-mail: yanxia@uow.edu.au

ABSTRACT

The quasi-likelihood approach for estimation of state-space models is proposed in this paper. In the proposed method both state variables and unknown parameters are estimated by quasi-likelihood approach. The procedure is less complex and easily implemented. Two examples are studied in this paper. The simulation study has shown that the performance of the quasi-likelihood approach is as good as that of the maximum likelihood approach. Comparing to the traditional maximum likelihood approach, the quasi-likelihood approach is quite simple and standard and can also be carried out without full knowledge on the probability structure of relevant state-space system.

Keywords: State-Space Model(SSM), Quasi-likelihood(QL), Quasi-score Estimating Function, Quasi-likelihood Estimate(QLE)

1. INTRODUCTION

The class of state space models (SSM) provides a flexible framework for describing a wide range of time series in a variety of disciplines. For extensive discussion on SSM and their applications see Harvey (1989) and Durbin and Koopman (2001). A state-space model can be written as

$$y_t = f(\alpha_t, \theta) + \epsilon_t, \qquad t = 1, 2, \cdots, T \tag{1}$$

where $y_1, \ldots, y_T$ represent the time series of observations; θ is an unknown parameter that needs to be estimated; $f(.)$ is a function of state variable α_t and θ; and $\{\epsilon_t\}$ are uncorrelated disturbances with $E_{t-1}(\epsilon_t) = 0$, $Var_{t-1}(\epsilon_t) = \sigma^2_{\epsilon_t}$; in which E_{t-1}, and Var_{t-1} denote conditional mean and conditional variance associated with past information updated to time t-1 respectively. State variables $\alpha_1, \ldots, \alpha_T$ are unobserved and satisfy the following model

$$\alpha_t = h(\alpha_{t-1}, \theta) + \eta_t, \qquad t = 1, 2 \cdots, T, \tag{2}$$

where $h(.)$ is a function of past state variables and θ; $\{\eta_t\}$ are uncorrelated disturbances with $E_{t-1}(\eta_t) = 0$, $Var_{t-1}(\eta_t) = \sigma^2_\eta$.

One of the special applications that we will consider in detail is the case where the time series $y_1, \ldots, y_T$ consist of counts. Here, it might be plausible to model y_t by a Poisson distribution. Models of this type have been used for rare diseases. (see Zeger (1988); Chan and Ledolter (1995); Davis, Dunsmuir and Wang (1988)).

Another noteworthy application of the SSM that we will consider is Stochastic Volatility Model (SVM), a frequently used model for returns of financial assets. Applications, together with estimation for (SVM), can be found in Jacquier, et al (1994); Briedt and Carriquiry (1996); Harvey and Streible (1998); Sandmann and Koopman (1998); Pitt and Shepard (1999).

There are several approaches in the literature for estimating the parameters in (SSMs) by using the maximum likelihood method when the probability structure of underlying model is normal or conditional normal. For example the EM algorithms, one of the methods, is used to estimate the parameters in SSMs. The EM algorithms involved two steps E-step and M-step. In the E-Step, the expectation of the logarithm of the complete likelihood conditional on the latest estimate of parameters is required. In M-step, the approximation is optimized with respect to the vector of parameters and the procedure will be repeated until estimation is converge. In some applications of the EM algorithm, the E-step is complex and does not admit a closed form solution to the computation of the conditional expectation of the complete data log likelihood. One way to get over this problem is to resort to numerical integration. However, in some cases, especially when the complete-data density does not belong to the exponential family, numerical integration over the missing-data density (state variables) dose not always preserve the function (Meng and Rubin, (1991)). To overcome the complex likelihoods of a time series model with count data, Chan and Ledolter (1995) proposed the Monte Carlo EM algorithm that uses a Markov chain sampling technique in the calculation of the expectation in the the E-step of the EM algorithm. Durbin and Koopman (1997, 2001) obtained accurate approximation of the log-likelihood for Non-Gaussian state space models by using Monte Carlo simulation. The log-likelihood function is maximised numerically to obtain estimates of unknown parameters. Kuk (1999) suggested an alternative class of estimate models based on conjugate latent process and applied it to approximate the likelihood of a time series model for count data. Sandmann and Koopman (1998) introduced the Monte Carlo maximum likelihood method of estimating stochastic volatility models (SVM). Davis and Rodriguez-Yam (2005) proposed an alternative estimation procedure which is based on an approximation to the likelihood function.

In this paper we will follow a different approach, the quasi-likelihood (QL) approach, to estimate the parameters and predictors of state variables in SSMs. In literature, the QL approach has been applied to stochastic volatility models (SVM) by Papanastassiou and Ioannides (2004). They use and extend set of Kalman filter equations in their estimation procedure, and have restricted themselves to a linear state space model. Jorgensen et al. (1999) and Knudsen (2001) estimated parameters from Poisson-gamma and other count data state space models by using estimating functions approach. They use the Kalman filter and smoother in their estimation procedures. The Kalman filter and the smoother are methods used to estimate predictors of state-variables and one-step-ahead predictors of observations. Usually the Kalman filter is derived through the maximum likelihood method. This means that the probability structure of the underlying state space system needs to be known. However, in practice, knowing system probability structure usually is not realistic. Furthermore, the likelihood function is often difficult to calculate. For these reasons, the maximum likelihood method is often difficult to apply in practice. Furthermore, Kalman filter involves many complex matrices calculation, which sometime makes the estimation procedure complex in practice. Not like the QL approach given by Papanastassiou and Ioannides (2004), Knudsen (2001) and Jorgensen et al. (1999), To avoid complexity expression of Kalman filter matrix, We propose to apply the QL method only to the whole estimation procedure of SSMs. We will demonstrate and show the proposed estimating procedure is less complex and easily implemented in estimating the state and parameters in state space models (SSMs). The QL method was first introduced by Wedderburn (1974). Wedderburn's work was mainly based on the generalized linear model. At the same time, a similar technique was also independently developed by Godambe and Heyde. This technique later was called "quasi-likelihood" (see, Godambe and Heyde, (1987)). The later technique is more focused on the applications to the inference of stochastic processes. These two independently developed quasi-likelihood methods are defined in different ways because the original approaches are different. The definition given by Godambe and Heyde (1987) is more general than that given by Wedderburn (1974). For this aspect of discussion see Lin and Heyde (1993). In this paper, we will adopt the definition of the quasi-likelihood given by Godambe and Heyde (1987). For detail knowledge on the quasi-likelihood method see Heyde (1997).

Consider a stochastic process $y_t \in R^r$,

$$y_t = \mu_t(\theta) + m_t, \qquad 0 \le t \le T \tag{3}$$

where $\theta \in \Theta \in R^p$ is the parameter needed to be estimated; μ_t is a function vector of $\{y_s\}_{s<t}$; (in the other words, μ_t is $\mathcal{F}_{t-1}$-measurable); m_t is an error process with $E(m_t|\mathcal{F}_{t-1}) = E_{t-1}(m_t) = 0$. When the following estimating function space

$$\mathcal{G}_T = \{\sum_{t=1}^{T} A_t(y_t - \mu_t) \mid A_t \text{ is a } \mathcal{F}_{t-1}\text{-measurable } p \times r \text{ matrix}\}$$

is considered, the standard quasi-score estimating function in the space has form

$$G_T^*(\theta) = \sum_{i=1}^{T} E_{t-1}(\dot{m}_t)(E_{t-1}(m_t m_t'))^{-1} m_t, \tag{4}$$

where $\dot{m}_t = \frac{\partial m_t}{\partial \theta}$ and "$\prime$" denotes transpose. The solution of $G_T^*(\theta) = 0$ is the quasi-likelihood estimator of θ. For a special scenario, if we only consider sub estimating function spaces of $\mathcal{G}_T$, for example,

$$\mathcal{G}_T^{(t)} = \{A_t(y_t - \mu_t) \mid A_t \text{ is a } \mathcal{F}_{t-1}\text{-measurable } p \times r \text{ matrix}\} \subset \mathcal{G}_T, \qquad t < T,$$

then, the standard quasi-score estimating function in this space is

$$G_{(t)}^*(\theta) = E_{t-1}(\dot{m}_t)(E_{t-1}(m_t m_t'))^{-1} m_t \tag{5}$$

and $G_{(t)}^*(\theta) = 0$ will give the quasi-likelihood estimator based on the information provided by $\mathcal{G}_T^{(t)}$.

In next section, we use these results to develop an inference approach for estimating parameters in SSMs. This approach can be carried out without full knowledge of the probability structure of underlying system y_t.

This paper is organized as follows. In Section 2.1, we apply the QL approach to SSMs. In Sections 2.2 and 2.3 we demonstrate the QL approach via simulation studies. In Section 2.4 we discuss the issue of initial values addressed in the procedure of simulation and estimation. In Section 2.5 we apply the QL approach to real data. A summary is given in Section 3.

2. PARAMETER ESTIMATION

2.1 The quasi-likelihood approach

In this subsection we introduce how to use the QL approach to estimate parameters in SSM without borrowing the transition matrix introduced in the standard Kalman filter method. Consider the following state-space model

$$y_t = f(\alpha_t, \theta) + \epsilon_t, \quad \text{and} \quad \alpha_t = h(\alpha_{t-1}, \theta) + \eta_t, \quad t = 1, 2 \cdots, T, \tag{6}$$

where $\{y_t\}$ represents the time series of observations, $\{\alpha_t\}$ the state variables, θ unknown parameter taking value in an open subset Θ of p-dimensional Euclidean space. Both f and h are functions satisfying certain regularity conditions, and the error terms ϵ_t and η_t are independent. Denote $\delta_t = (\epsilon_t, \eta_t)'$. Then δ_t is a martingale difference with

$$E_{t-1}(\delta_t) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

$$Var_{t-1}(\delta_t) = \begin{pmatrix} \sigma_{\epsilon_t}^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}.$$

Traditionally, normality or conditional normality condition is assumed and the estimation of parameters are obtained by the ML approach. However, in many applications the normality assumption is not realistic. Further more, the probability structure of the model may not be known. Thus the maximum likelihood method is not applicable or it is too complex to estimate parameters through the ML method as the calculation involved is complex sometimes. In the following the QL approach for estimating the parameters in SSM is introduced. This approach can be carried out without full knowledge of the system probability structure. It involves in making decision about the initial values of θ and iterative procedure. Each iterative procedure consists of two steps. The first step is to use the QL method to obtain the optimal estimation for each α_t, say $\hat{\alpha}_t$. The second step is to combine the information of $\{y_t\}$ and $\{\hat{\alpha_t}\}$ to adjust the estimate of θ through the QL method.

In Step 1, assign an initial value to θ and consider the following martingale difference

$$\delta_t = \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} = \begin{pmatrix} y_t - E(y_t|\mathcal{F}_{t-1}) \\ \alpha_t - E(\alpha_t|\mathcal{F}_{t-1}) \end{pmatrix}$$

and estimating function space

$$\mathcal{G}_T^{(t)} = \{A_t\delta_t \mid A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable}\},$$

where α_t is considered as an unknown parameter. As mentioned in (5), a standardized optimal estimating function in this estimating function space is

$$G^*_{(t)}(\alpha_t) = E_{t-1}(\frac{\partial \delta_t}{\partial \alpha_t})[Var_{t-1}(\delta_t)]^{-1}\delta_t.$$

To obtain the QL estimate $\hat{\alpha_t}$ of α_t, we let $G^*_{(t)}(\alpha_t) = 0$ and solve the equation for α_t. This estimation is as same as the estimation given by Kalman filter approach when the underlying system has a normal probability structure. (For detailed discussion see Lin, (2007)).

In Step 2, θ is considered as an unknown parameter and the estimating function space

$$\mathcal{G}_T = \{\sum_{t=1}^{T} A_t\delta_t \mid A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable}\}$$

is considered. Then the standardized optimal estimating function in this estimating function space is

$$G^*_T(\theta) = \sum_{t=1}^{T} E_{t-1}(\frac{\partial \delta_t}{\partial \theta})[Var_{t-1}(\delta_t)]^{-1}\delta_t.$$

To obtain the QL estimate $\hat{\theta}$ for θ we let $G^*_T(\theta) = 0$ and solve the equation while replacing α_t by $\hat{\alpha_t}$ obtained from Step 1. The $\hat{\theta}$ obtained from Step 2 will be used as a new initial value for the θ in Step 1 in the next iterative procedure. These two steps will be alternatively repeated till certain criterion is met. The differences between the QL approach and the EM algorithm described by Harvey (1989), Durbin and Koopman (2001) and Fahrmeir and Tutz (1996) are two. The first is that the QL approach requires less knowledge on the underling probability structure. The second is that the QL approach uses filter states only in estimating parameters.

When $\sigma^2_{\epsilon_t}$ and σ^2_η are unknown, a procedure for estimating $\sigma^2_{\epsilon_t}$ and σ^2_η will be involved. In Step 1, initial value for $\sigma^2_{\epsilon_t}$ and σ^2_η need to be provided. By the end of Step 2, the estimations of $\sigma^2_{\epsilon_t}$ and σ^2_η will be made and will be the new initial value for $\sigma^2_{\epsilon_t}$ and σ^2_η respectively in the next step. For details, see the simulation studies in next sections.

Two simulation studies on this approach are presented below. One is based on Poisson Model and other is based on the basic Stochastic Volatility Model (SVM).

2.2 Poisson model

Let $y_1, y_2, \cdots, y_T$ be observations and $\alpha_1, \alpha_2, \cdots, \alpha_T$ be states. The state-space model is given by

$$y_t \sim \text{Poisson distribution with parameter } e^{\beta+\alpha_t}, \tag{7}$$

where $\alpha_t = \phi\alpha_{t-1} + \eta_t$, and η_t are i.i.d with mean 0 and variance σ^2_η. The study on the generalized form of the above model can be found from Durbin and Koopman (1997), Kuk (1999), and Davis and Rodriguez-Yam (2005). Here the information on η_t is only given by the first two moments. Consider the situation of the above model by assuming that $\{y_t - e^{\beta+\alpha_t}\}$ and η_t are mutually independent; β, ϕ and σ^2_η are unknown. Based on this situation, we consider the following martingale difference

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} = \begin{pmatrix} y_t - e^{\beta+\alpha_t} \\ \alpha_t - \phi\alpha_{t-1} \end{pmatrix}.$$

Our estimation consists of two steps. In Step 1, let α_t act as an unknown parameter. The standard quasi-score estimating function in the estimating function space determined by

$$\mathcal{G} = \{A_t \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \mid A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable } \}$$

is

$$\begin{aligned} G_{(t)}(\alpha_t) &= (-e^{\beta+\alpha_t}, 1) \begin{pmatrix} e^{\beta+\alpha_t} & 0 \\ 0 & \sigma^2_\eta \end{pmatrix}^{-1} \begin{pmatrix} y_t - e^{\beta+\alpha_t} \\ \alpha_t - \phi\alpha_{t-1} \end{pmatrix} \\ &= -y_t + e^{\beta+\alpha_t} + \frac{1}{\sigma^2_\eta}(\alpha_t - \phi\alpha_{t-1}). \end{aligned} \tag{8}$$

To carry out the two-step estimation procedure described in Section 2.1, the starting value $\psi_0 = (\beta_0, \phi_0, \sigma^2_{\eta 0})$, and the initial value for state process α_t are required. Impact of the starting value of ψ_0 and the initial value of α_t on parameter estimation is discussed in Section 2.4. Initially we assign $\alpha_0 = \hat{\alpha}_0 = 0$. Once the optimal estimation of α_{t-1} is obtained, say

$\hat{\alpha}_{t-1}$, the quasi-likelihood estimation of α_t, will be given by solving equation $G_{(t)}(\alpha_t) = 0$ through Newton-Raphson algorithm. It gives

$$\alpha_t^{(k+1)} = \alpha_t^{(k)} - \frac{-y_t + e^{\beta+\alpha_t^{(k)}} + \frac{1}{\sigma_{\eta_0}^2}(\alpha_t^{(k)} - \phi\hat{\alpha}_{t-1})}{e^{\beta+\alpha_t^{(k)}} + \frac{1}{\sigma_{\eta_0}^2}}. \tag{9}$$

It starts with $\alpha_t^{(1)} = \hat{\alpha}_{t-1}$ and will be iterative till it is convergent. Then move to Step 2.

In Step 2, let β and ϕ act as unknown parameters. We apply the QL method to estimate β and ϕ. In this step, the estimating function space is

$$\mathcal{G} = \{\sum_{t=1}^{T} A_t \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \mid A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable } \}.$$

The standard quasi-score estimating function related to $\mathcal{G}$ is

$$G_T(\beta,\phi) = \sum_{t=1}^{T} \begin{pmatrix} -e^{\beta+\alpha_t} & 0 \\ 0 & -\alpha_{t-1} \end{pmatrix} \begin{pmatrix} e^{\beta+\alpha_t} & 0 \\ 0 & \sigma_{\eta_0}^2 \end{pmatrix}^{-1} \begin{pmatrix} y_t - e^{\beta+\alpha_t} \\ \alpha_t - \phi\alpha_{t-1} \end{pmatrix}.$$

Replace α_t by $\hat{\alpha}_t$, $t = 1, 2, \cdots, T$, and the QL estimate of β and ϕ will be given by solving $G_T(\beta,\phi) = 0$. Therefore

$$\hat{\beta} = \ln(\sum_{t=1}^{T} y_t) - \ln(\sum_{t=1}^{T} e^{\hat{\alpha}_t}), \qquad t = 1, 2, \cdots, T, \tag{10}$$

$$\hat{\phi} = \frac{\sum_{t=1}^{T} \hat{\alpha}_t \hat{\alpha}_{t-1}}{\sum_{t=1}^{T} \hat{\alpha}_{t-1}^2}, \qquad t = 1, 2, \cdots, T. \tag{11}$$

and let

$$\hat{\sigma_\eta^2} = \frac{\sum_{t=1}^{T} (\hat{\eta_t} - \bar{\hat{\eta}})^2}{T-1} \tag{12}$$

where $\hat{\eta_t} = \hat{\alpha}_t - \hat{\phi}\hat{\alpha}_{t-1}, t = 1, 2, \cdots, T$, and $\bar{\hat{\eta}} = \frac{\sum_{t=1}^{T} \hat{\eta}_t}{T}$. The above two steps will be iteratively repeated till certain criterion is meet. The $\hat{\psi} = (\hat{\beta}, \hat{\phi}, \hat{\sigma}_\eta^2)$ obtained from previous Step2 will be used as an initial value for Step1 in next iteration.

To demonstrate described estimation procedures we carried out a simulation study on model (7). Our simulation was carried as follows: Firstly, independently simulate 1000 samples with size 500 from (7) based on a true parameter $\psi = (\beta, \phi, \sigma_\eta^2)$. After series $\{y_t\}$ are generated, we pretend that α_t are unobserved and ϕ, β and σ_η^2 are unknown. Then apply the above estimation procedure to y_t only to obtain the estimation of α_t, ϕ, β and σ_η^2. We consider different parameter settings for $\psi = (\phi, \beta, \sigma_\eta^2)$ which are the same as the layout considered in Rodriguez-Yam (2003). For the simulation, we compute mean and root mean squared errors for $\hat{\beta}$, $\hat{\phi}$ and $\hat{\sigma_\eta^2}$ based on N=1000 independent samples. Result are shown in Table 1. In Table 1, QL denotes the quasi-likelihood estimate.

Table 1: QL estimates based on 1000 replication. Root mean square error of estimates are reported below each estimate.

	β	ϕ	σ_η	β	ϕ	σ_η	β	ϕ	σ_η
True	-0.613	0.90	0.6221	-0.613	0.95	0.446	-0.613	0.98	0.284
QL	-0.610	0.888	0.606	-0.615	0.907	0.436	-0.615	0.969	0.287
	0.004	0.096	0.037	0.039	0.056	0.042	0.021	0.016	0.040
True	0.150	0.90	0.312	0.150	0.95	0.223	0.150	0.98	0.142
QL	0.149	0.898	0.312	0.149	0.945	0.217	0.147	0.974	0.137
	0.005	0.021	0.016	0.009	0.017	0.017	0.021	0.012	0.017
True	0.373	0.90	0.111	0.373	0.95	0.079	0.373	0.98	0.051
QL	0.372	0.898	0.114	0.345	0.946	0.082	0.345	0.973	0.052
	0.011	0.019	0.006	0.030	0.015	0.005	0.033	0.013	0.033

2.3 Stochastic volatility model (SVM)

For the second simulation example, we consider the stochastic volatility process, which is often used for modelling log-returns of financial assets, defined by

$$y_t = \sigma_t \xi_t = e^{\alpha_t/2}\xi_t, \qquad t = 1, 2, \cdots, T, \tag{13}$$

and

$$\alpha_t = \gamma + \phi\alpha_{t-1} + \eta_t, \qquad t = 1, 2, \cdots, T, \tag{14}$$

where both ξ_t and η_t i.i.d respectively; η_t has mean 0 and variance σ_η^2. A key feature of the SVM in (13) is that it can be transformed into a linear model by taking the logarithm of the square of observations

$$\ln(y_t^2) = \alpha_t + \ln {\xi_t}^2, \qquad t = 1, 2, \cdots, T. \tag{15}$$

If ξ_t were standard normal, then $E(\ln \xi_t^2) = -1.2704$ and $Var(\ln \xi_t^2) = \pi^2/2$ (see Abramowitz and Stegun (1970), p. 943). Let $\varepsilon_t = \ln \xi^2 + 1.2704$. The disturbance ε_t is defined so as to have zero mean. Based on this situation, we consider the following martingale difference

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} = \begin{pmatrix} \ln(y_t^2) - \alpha_t + 1.2704 \\ \alpha_t - \gamma - \phi\alpha_{t-1} \end{pmatrix}.$$

In Step 1, let α_t act as an unknown parameter. The standard quasi-score estimating function determined by the estimating function space

$$\mathcal{G} = \{A_t \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \mid A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable } \}$$

is

$$G_{(t)}(\alpha_t) = (-1,1)\begin{pmatrix} \frac{\pi^2}{2} & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}^{-1} \begin{pmatrix} \ln(y_t^2) - \alpha_t + 1.2704 \\ \alpha_t - \gamma - \phi\alpha_{t-1} \end{pmatrix}$$

$$= \frac{-2}{\pi^2}(\ln(y_t^2) - \alpha_t + 1.2704) + \sigma_\eta^{-2}(\alpha_t - \gamma - \phi\alpha_{t-1}). \qquad (16)$$

Let $\hat{\alpha_0} = 0$ and initial values $\psi_0 = (\gamma_0, \phi_0, \sigma_{\eta 0}^2)$. Given $\hat{\alpha}_{t-1}$ the optimal estimation of α_{t-1}, the quasi-likelihood estimation of α_t, i.e. the optimal estimation of α_t, will be given by solving $G_{(t)}(\alpha_t) = 0$, i.e.

$$\hat{\alpha_t} = \frac{2\sigma_{\eta_0}^2(\ln(y_t^2) + 1.2704) + \pi^2(\phi\hat{\alpha}_{t-1} + \gamma)}{2\sigma_{\eta_0}^2 + \pi^2}, \qquad t = 1, 2, \cdots, T. \qquad (17)$$

In Step 2, based on $\{\hat{\alpha_t}\}$ and $\{y_t\}$, let γ and ϕ act as unknown parameters, and use the QL approach to estimate them. The standard quasi-score estimating function related to the estimating function space

$$\mathcal{G} = \{\sum_{t=1}^{T} A_t \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \mid A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable } \}$$

is

$$G_T(\gamma,\phi) = \sum_{t=1}^{T} \begin{pmatrix} 0 & -1 \\ 0 & -\alpha_{t-1} \end{pmatrix} \begin{pmatrix} \frac{\pi^2}{2} & 0 \\ 0 & \sigma_{\eta_0}^2 \end{pmatrix}^{-1} \begin{pmatrix} \ln(y_t^2) - \alpha_t + 1.2704 \\ \alpha_t - \gamma - \phi\alpha_{t-1} \end{pmatrix}.$$

Replace α_t by $\hat{\alpha}_t$,, $t = 1, 2, \cdots, T$, the QL estimate of γ and ϕ will be given by solving $G_T(\gamma, \phi) = 0$. Therefore

$$\hat{\phi} = \frac{\sum_{t=1}^{T}\hat{\alpha}_t \sum_{t=1}^{T}\hat{\alpha}_{t-1} - T\sum_{t=1}^{T}\hat{\alpha}_{t-1}\hat{\alpha}_t}{(\sum_{t=1}^{T}\hat{\alpha}_{t-1})^2 - T\sum_{t=1}^{T}\hat{\alpha}_{t-1}^2}, \qquad t = 1, 2, \cdots, T, \qquad (18)$$

$$\hat{\gamma} = \frac{\sum_{t=1}^{T}\hat{\alpha}_t - \hat{\phi}\sum_{t=1}^{T}\hat{\alpha}_{t-1}}{T}, \qquad t = 1, 2, \cdots, T. \qquad (19)$$

and let

$$\hat{\sigma_\eta^2} = \frac{\sum_{t=1}^{T}(\hat{\eta_t} - \bar{\hat{\eta}})^2}{T-1} \qquad (20)$$

where $\hat{\eta_t} = \hat{\alpha}_t - \hat{\gamma} - \hat{\phi}\hat{\alpha}_{t-1}, t = 1, 2, \cdots, T$. The above two steps will be iteratively repeated till certain criterion is meet. As mentioned before, $\hat{\psi} = (\hat{\gamma}, \hat{\phi}, \hat{\sigma}_\eta^2)$ will be used as an initial value for next step in the iterative procedure.

The format for this simulation study is the same as the layout considered by Rodriguez-Yam(2003). From empirical studies (e.g Harvey and Shepard, 1993; Jacquier et, al., (1994)) the values of ϕ between 0.9 and 0.98 are of primary interest. For this simulation study, we consider samples of size T=1000 and compute mean and root mean squared errors for $\hat{\phi}$, $\hat{\gamma}$ and $\hat{\sigma_\eta^2}$ based on N=1000 independent samples. The results are shown in Table 2. QL denotes the quasi-likelihood estimate.

Table 2: QL estimates based on 1000 replication. Root mean square error of estimates are reported below each estimate.

	γ	ϕ	σ_η	γ	ϕ	σ_η	γ	ϕ	σ_η
True	-0.821	0.90	0.675	-0.411	0.95	0.484	-0.614	0.98	0.308
QL	-0.989	0.867	0.370	-0.563	0.921	0.329	-0.213	0.95	0.249
	0.254	0.039	0.293	0.202	0.035	0.159	0.075	0.031	0.0673
True	-0.736	0.90	0.363	-0.368	0.95	0.260	-0.147	0.98	0.166
QL	-0.835	0.898	0.234	-0.416	0.931	0.195	-0.155	0.970	0.158
	0.153	0.015	0.130	0.083	0.022	0.066	0.030	0.012	0.015
True	-0.706	0.90	0.135	-0.353	0.95	0.096	-0.141	0.98	0.061
QL	-0.721	0.891	0.114	-0.353	0.946	0.094	-0.143	0.979	0.060
	0.070	0.014	0.021	0.037	0.007	0.004	0.012	0.002	0.003

2.4 The issues of initial values in the procedures of simulation and estimation

In this subsection, we study the impact of initial values assigned to state variable α_0 and ψ_0.

Firstly, consider the issue of the initial value for state variable α_0. By noting that state variables form a stationary process in both models studied in this paper, therefore, no mater what value has signed to α_0, the force of stationarity will bring the simulated sample path back to the equilibrium described by the model. Therefore, like others (Zivot, et al. (2004); and Durbin and Koopman (2001)), we assign $\alpha_0 = 0$ in our study. In the following, we use simulation study to convince the fact. Consider Poisson Model (7) and

Table 3: QL estimates based on 1000 replication. Root mean square error of estimates are reported below each estimate based on different initial values for α_0.

	α_0	SVM			Poisson Model		
		γ	ϕ	σ_η	β	ϕ	σ_η
True	0	-0.141	0.98	0.061	0.15	0.95	0.223
$\hat{\alpha}_0$	0	-0.144	0.979	0.060	0.139	0.934	0.235
		0.013	0.002	0.003	0.001	0.001	0.019
	1	-0.163	0.979	0.062	0.139	0.933	0.235
		0.025	0.004	0.003	0.001	0.001	0.019
	2	-0.179	0.975	0.063	0.139	0.933	0.235
		0.041	0.006	0.004	0.001	0.001	0.019
	3	-0.194	0.972	0.066	0.139	0.933	0.235
		0.055	0.008	0.006	0.001	0.001	0.019
	4	-0.208	0.970	0.069	0.149	0.946	0.235
		0.068	0.009	0.009	0.001	0.001	0.019
	5	-0.221	0.969	0.073	0.139	0.933	0.235
		0.081	0.012	0.012	0.001	0.001	0.019

SVM (13) and (14). In Table 3, we show how the initial value α_0 effects the final estimation

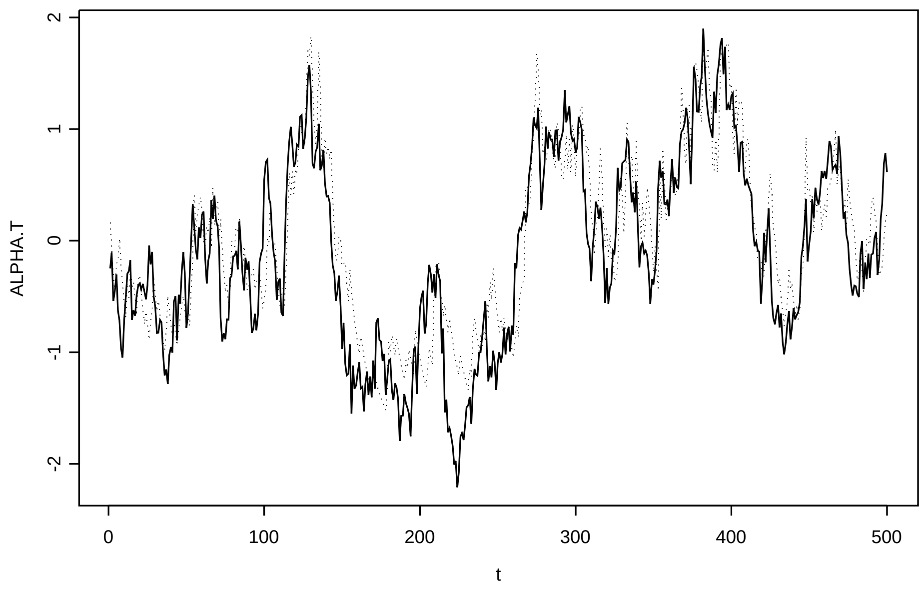


Figure 1: The plot of α_t and $\hat{\alpha}_t$ when $\hat{\alpha}_0$ is assigned as 0.

when the QL approach is applied. The estimate parameters for SVM are presented in the third column using $\psi_0 = (-0.121, 0.99, 1)$ as initial value for all of the estimation procedure and the estimate parameters for Poisson Model are presented in the fourth column using $\psi_0 = (0.14, 0.99, 1)$. The value of α_0 in the true models always assigned to 0. We can see from Table 3 there is no big impact of the choice of the value for $\hat{\alpha}_0$ in estimation procedure.

Consider Poisson Model given in (7) with true parameter $\psi = (0.15, 0.95, 0.223)$. In Figures 1 and 2 using $\psi_0 = (0.14, 0.99, 1)$, we show there is no big impact on the estimate of the path α_t when $\hat{\alpha}_0$ is chosen far away from the true value α_0. The solid line in Figures 1 and 2 shows the true value of state variables and the dotted line shows the optimal estimate of state variables α_t. In both cases, the QL estimate for the state variables are relatively close to the true state, even though the initial value might not very close to the true value $\alpha_0 = 0$.

Secondly, we consider the starting values for system parameters $\psi = (\beta, \phi, \sigma_\eta^2)$. As described in literature, the estimation procedure outputs strongly rely on the appropriate value of the initial parameter ψ_0. Usually ψ_0 should be chosen from the neighborhood of its true value. In practice, how to identify an appropriate ψ_0 is an interesting question.

In their survey article Zivot, et al. (2004) suggest to choose the starting value ψ_0 close to the true value of ψ. Sometime people use different approach to choose ψ_0. For example, Durbin and Koopman (1997) maximised numerically the approximate likelihood for non-Gaussian SSMs to obtain the starting value for ψ. Sandmann and Koopman(1998) used two-dimensional grid search procedure to search for an appropriate starting value for ψ. In

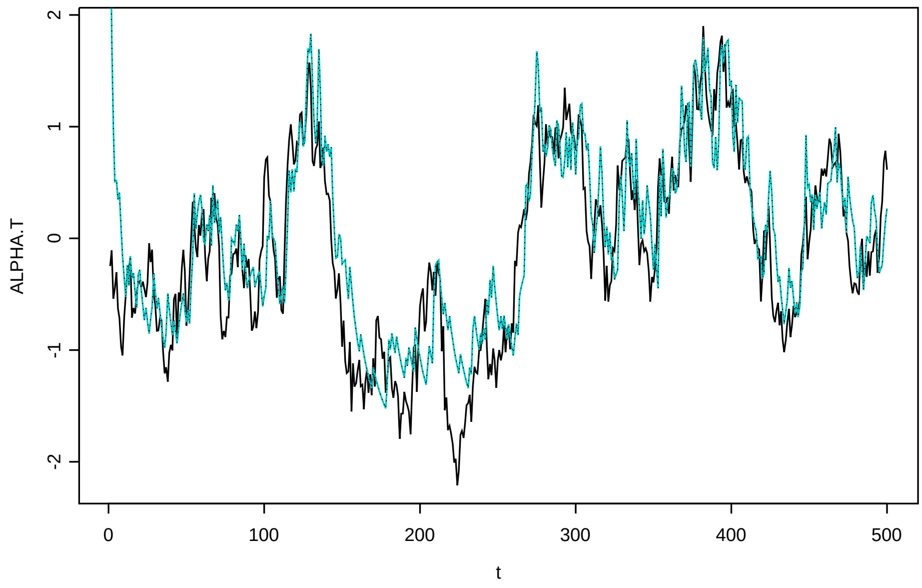


Figure 2: The plot of α_t and $\hat{\alpha}_t$ when $\hat{\alpha}_0$ is assigned as 5.

this paper, we adapt Sandmann and Koopman's idea.

2.5 Application to SVM

We apply the estimation procedure described in previous section to a real case where the observations are assumed to satisfy SVM (13) and (14)(see, Davis and Rodriguez-Yam (2005); Rodriguez-Yam (2003); Durbin and Koopman (2001)). The data are the pound/dollar of the daily observations of weekdays closing pound to dollar exchange rates $x_t, t = 1, \cdots, 945$ from 1/10/81 to 28/6/85 (http://staff.feweb.vu.nl/ koopman/sv/).

In literature, SVM (13) and (14) are used to model $y_t = log(x_t) - log(x_{t-1})$ where in the model, ξ has standard normal distribution. Set parameter $\psi = (\gamma, \phi, \sigma_\eta^2)$.

Table 4: Estimation of γ, ϕ, and σ_η for Pound/Dollar exchange rate data.

	γ	ϕ	σ_η
QL	-0.025	0.929	0.013
AL	-0.0227	0.957	0.0267
MCL	-0.0227	0.975	0.0273

Table 4 shows the estimates of ψ obtains by various method. QL denotes the estimate obtained by quasi-likelihood approach, AL the estimate obtained by maximizing the approximate likelihood proposed by Davis and Rodriguez-Yam (2003) and MCL estimate obtained by maximizing the estimate of the likelihood proposed by Durbin and Koopman (1997).

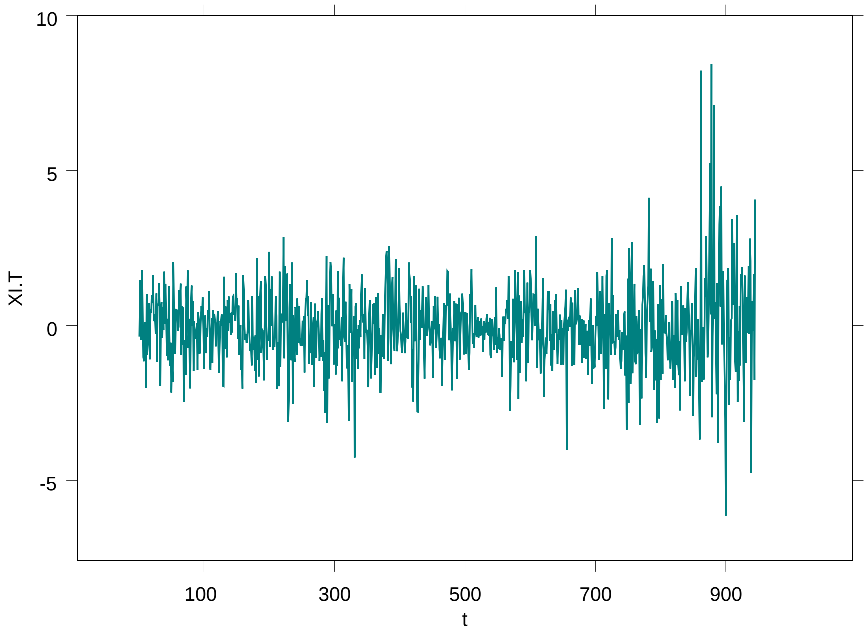


Figure 3: The plot of $\hat{\xi}_t$

Note that AL and MCL outputs are taken from Rodriguez-Yam (2003). The QL estimations are slightly different from the estimation of AL and MCL.

To check if the QL estimation is reasonable, we examine the plot of $\hat{\xi}_t$ given by QL approach and show the plot in Figure 3. Except for some outliers appeared at the end of time period, the plot of $\hat{\xi}_t$ clearly indicate that we can accept that ξ_t are i.i.d. as required.

3. CONCLUSIONS

By using two examples, this paper shows an alternative approach to estimate the parameters in SSMs. Instead of using traditional kalman filter formulae to estimate state variables, this approach use the QL method to estimate the state variables. It turns out the whole estimation processes looks very straightforward and is easily implemental. When the probability structure of underlying systems is complex or unknown, when maximum likelihood or mixture of maximum likelihood is not easily to implemented, the approach proposed in this paper is considerable approach for estimating parameters in SSMs.

REFERENCES

Abramovitz, M., Stegun, N. (1970). *Handbook of Mathematical Functions*, Dover Publication, New York.

Breidt, F.J. and Carriquiry, A.L. (1996). Improved quasi-maximum likelihood estimation for stochastic volatility models. In: Zellner, A. and Lee, J.S. (Eds.), *Modelling and Prediction: Honouring Seymour Geisser*. Springer, New York, 228–247.

Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts, *J. Amer. Statist. Assoc.*, **90**, 242–252.

Davis, R. A., Dunsmuir, W. T. M. and Wang, y. (1988). Modelling time series of count data. *In Asymptotics, Nonparametrics and Time Series* (ed Subir Ghosh), Marcel Dekker.

Davis, R. A. and Rodriguez-Yam, G. (2005). Estimation for State-Space Models: an approximate likelihood approach, *Statistica Sinica*, **15**, 381–406.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford, New York.

Durbin, J. and Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-gaussian state space models, *Biometrika*, **84**, 669–684.

Fahrmeir, L. and G. Tutz (1996). *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer-Verlag.

Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation, *Inter. Statist. Rev.*, **55**, 231–244.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Moodels and the Kalman Filter*, Cambridge University Press.

Harvey, A. C. and Streible, M. (1998). Testing for a slowly changing level with special reference to stochastic volatility, *J. Econometrics*, **87**, 167–189.

Harvey, A. C. and Shepard, N. (1993). Estimation and testing of stochastic variance models. Unpublished manuscript, The London School of Economics.

Hedye, C. C. (1997). *Quasi-likelihood and its Application: a General Approach to Optimal Parameter Estimation*, Springer, New York.

Jacquire, E., Polson, N. G. and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion), *J. Bus. Econom. Statist.*, **12**, 371–417.

Jorgensen, B., Lundbye-Christensen, S., Song X.K., Sun, L. (1999). A state space model for multivariate longitudinal count data, *Biometrika*, **86**, No. 1, 169–181.

Knudsen, S.J.(2001). Bias Adjusted Profile Estimating Functions and State Space Modeling of Air Pollution Data, Dept. of Statistics and Demography, Monographs, University of Southern Denmark

Kuk, A. Y. (1999). The use of approximating models in Monte Carlo maximum likelihood estimation, *Statist. Probab. Letter*, **45**, 325–333.

Lin, Y.-X. (2007). An alternative derivation of the Kalman filter using the quasi-likelihood method, *J. of Statistical Planning and Inference*, **137**, 1627–1633.

Lin, Y.-X. and Heyde, C. C. (1993). Optimal estimating functions and Wedderburn's quasi-likelihood, *Comm. Statist.: Theory and Methods*, **22**, 2341–2350.

Meng, X. L. and Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm, *J.Amer. Statist. Assoc.*, **86**, 899–909.

Papanastastiou, D. and Ioannides, D. (2004). The estimation of a state space model by estimating functions with an application, *Statistica Neerlandica* , **58**, No. 4, 407–427.

Pitt, M. K. and Shepard, N. (1999). Filtering via simulation: auxiliary particle filters, *J.Amer. Statist. Assoc.*, **94**, 590–599.

Rodriguez-Yam, G. (2003). Estimation for State-Space Models and Baysian regression analysis with parameter constraints, Ph.D. Thesis. Colorado State University.

Sandmann, G. and Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *J. Econometrics*, **87**, 271–301.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method, *Biometrika*, **61**, 439–447.

Zeger, S. L. (1988). A Regression Model for Time Series of Counts, *Biometrika*, **75**, 621–629.

Zivot, E., Wang, J. and Koopman, S. J. (2004). State space modelling in macroeconomics and finance using SsfPack in S+Finmetrics. In: Harvey, A., Koopman, S. J. and Shephard, N. (Eds.), *State Space and Unobserved Component Models: Theory and Application*. Cambridge University Press, 284–335.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 424–450.

FITTING WEIBULL DISTRIBUTION TO FAILURE TIME DATA

Zeinab Amin
Department of Mathematics and Actuarial Science,
The American University in Cairo, Egypt, and
Faculty of Economics and Political Science, Cairo University, Egypt.
E-mail: zeinabha@aucegypt.edu

Ali S. Hadi
Department of Mathematics and Actuarial Science,
The American University in Cairo, Egypt, and
Department of Statistical Science, Cornell University, USA.
E-mail: ahadi@aucegypt.edu or ali-hadi@cornell.edu

ABSTRACT

The three-parameter exponentiated Weibull (EW) family not only contains distributions with unimodal and bathtub failure rates, but also allows for a broad class of monotone hazard rates and is computationally convenient. The present article obtains estimators of the three parameters and quantiles of the EW lifetime model. Maximum likelihood and method of moments estimators for the three parameters are developed. Independent non-informative types of priors are considered for the parameters to develop Bayes estimators under the squared error loss function. Estimators of the parameters and quantiles using the elemental percentile method are developed. As opposed to the classical estimators which require multidimensional numerical search, the elemental percentile estimates are easy to derive and to compute and they are unique and well defined for all parameter and sample values. As the presence of censoring creates special problems for the classical estimation methods, an extension of the elemental percentile method is proposed to deal with censored data. Graphical displays for informal checks on the appropriateness of the model as well as formal goodness of fit tests for the three-parameter EW are illustrated. Two real-life data sets with and without censoring are used to illustrate the results.

Keywords: Bayes estimators, Censoring, Elemental percentile method, Goodness of fit, Maximum likelihood estimators, Moments estimators, Non-informative prior, Square error loss function, Three-parameter exponentiated Weibull, Total time on test.

1. INTRODUCTION

Although monotone failure rates have a wide range of applications, data in reliability analysis especially over the life-cycle of the individual or product, can involve decreasing failure rate during the infant mortality phase, constant failure rate during the so-called useful life phase and eventually increasing failure rate due to aging and wear out, indicating a bathtub failure rate. This bath-tub hazard rate is common in the fields of science, engineering, medical, biological, ecological and space explorations; Lawless (2003). Non monotone hazard functions other than the bathtub-shaped ones are not commonly used in practice.

There have been several attempts to answer the need for a family of distributions which allow flexibility in modeling non monotone failure rates. Several models, such as Stacy's generalized gamma Stacy (1962), the mixtures of Weibull distributions proposed by Mann et al. (1974), the generalized F distribution by Prentice (1975), a four-parameter family introduced by Gaver and Acar (1979), a three-parameter family proposed by Hjorth (1980) as a generalization of the Rayleigh distribution which itself is a generalization of the exponential distribution, as well as the two families introduced by Slymen and Lachenbruch (1984) have been introduced for modeling non monotone failure rate data. However, the inferential analysis of these distributions, as studied and discussed by Bain (1974), Gore et al. (1986) and Lawless (2003), often pose much mathematical complexity, especially in the presence of censoring.

The exponentiated Weibull (EW) family was introduced by Mudholkar and Srivastava (1993) as a generalization of the Weibull family. This family not only contains distributions with unimodal and bathtub failure rates but also allows for a broader class of monotone hazard rates and is computationally convenient for censored data. The EW family has been found to be more appropriate than the generalized extreme value family for modeling extremes such as floods, wave heights and wind speed in moderate size samples because of the slow convergence of the sample extremes to their asymptotic distributions (for more applications of the generalized extreme value model, see, for example, Castillo et al. (2005)). This was illustrated by Mudholkar and Hutson (1996) using the flood data for the Floyd River located at James, Iowa. Other applications of the EW family were illustrated in Mudholkar et al. (1995) who used the EW family in analyzing bus-motor-failure data as well as modeling Efron (1988) data on the survival times of head-and-neck cancer patients.

The cumulative distribution function, probability density function and hazard rate of the EW family are given by

$$F(x;\theta) = \left(1 - e^{-(x/\beta)^{\alpha}}\right)^{\lambda}, x > 0,\ \alpha,\ \beta, \lambda > 0, \tag{1}$$

$$f(x;\theta) = \frac{\alpha\lambda}{\beta^{\alpha}} x^{\alpha-1} e^{-(x/\beta)^{\alpha}} \left(1 - e^{-(x/\beta)^{\alpha}}\right)^{\lambda-1}, x > 0,\ \alpha,\ \beta, \lambda > 0, \tag{2}$$

and

$$h(x;\theta) = \frac{f(x;\theta)}{1 - F(x;\theta)}, \tag{3}$$

respectively, where $\theta = \{\alpha, \beta, \lambda\}$, α and λ are shape parameters and β is a scale parameter.

When setting $p = F(x_p; \theta)$, we obtain the pth quantile of X as

$$x_p = F^{-1}(p; \theta) = \beta \left[-\log\left(1 - p^{1/\lambda}\right) \right]^{1/\alpha}, \quad 0 \leq p \leq 1. \tag{4}$$

Mudholkar and Hutson Mudholkar and Hutson (1996) show that the EW family not only includes distributions with bathtub and unimodal failure rates but provides a broad class of monotone failure rates. The hazard function assumes distinctly different shapes over the four regions of the space of the shape parameters $\alpha > 0$ and $\lambda > 0$ separated by the boundary line $\alpha = 1$ and the curve $\alpha\lambda = 1$. Specifically for the EW family the hazard function is

- Monotone increasing for $\alpha \geq 1$ and $\alpha\lambda \geq 1$;
- Monotone decreasing for $\alpha \leq 1$ and $\alpha\lambda \leq 1$;
- Bathtub shaped for $\alpha > 1$ and $\alpha\lambda < 1$;
- Unimodal for $\alpha < 1$ and $\alpha\lambda > 1$.

Mudholkar and Hutson Mudholkar and Hutson (1996) also study the skewness and kurtosis properties, density shapes and tail character and the associated extreme value and extreme spacings for the two parameter EW family which is a special case of (1) when $\beta = 1$.

Life time data usually come in one of three forms:

- Type I Censoring: Under this sampling scheme, the experiment is run over a fixed time period such that the lifetime of the individual or item will be known only if it is less than some predetermined value. A more complicated form of Type I censoring arises when each item has its own specific censoring time, since not all items started to be tested on the same date.
- Type II Censoring: Under this sampling scheme, a total of n items are placed on test, but instead of continuing until all items have failed, the test is terminated at the time of failure of the rth item ($r \leq n$).
- Complete data, where the data are available without censoring.

Singh et al. Singh et al. (2005a,b) state that estimation procedures for the three-parameter EW model in presence of censoring seems to be untouched and that their paper was an attempt in this direction. For simplicity they consider only Type II censored data. Thus, to our knowledge, estimation procedures for the EW model under Type I censoring has not been considered in the literature. In Section 5, we first derive the Elemental Percentile Method (EPM) estimators for the case of complete data. The EPM is proposed by Castillo and Hadi (1995a). We then generalize EPM in Section 6, which makes it applicable to all types of censoring.

Estimation procedures under classical and Bayesian set up for the EW family have been considered by Singh et al. (1999) and Singh et al. (2002, 2005a,b). They develop maximum likelihood estimates (MLE) and Bayes estimates under complete and Type II censored data. However, no theoretical asymptotic variances have been derived. In this paper we derive these asymptotic variances in Section 2, which are required to obtain the MLE using numerical methods such as the Newton Raphson type methods.

In the Bayesian approach, Singh et al. (2005b) derive the joint posterior density of the three parameters, the marginal posterior densities for each parameter, and then state that the Bayes estimates under squared error loss will be the posterior means of these densities. All these densities and estimates involved multidimensional integrals which are not solvable analytically as Singh et al. (2005b) mentioned. In this paper we use Lindley (1980)'s approximation, which leads to Bayes estimators and their posterior risks in closed-forms.

With regards to the method of moments (MOM), Nassar and Eissa (2003) derived the rth moment of the three-parameter EW random variable. Their objective, however, was to study the properties of the distribution but not to find the MOM estimators. In this paper we derive the MOM estimators of the EW model.

The rest of this article is organized as follows: The MLE are given in Section 2 and the approximate estimates of their variances are derived. In Section 3 the method of moments (MOM) estimators are discussed. Bayes estimators are developed in Section 4, where independent non-informative types of priors are considered for the unknown parameters to develop Bayes estimators under the squared error loss function and where Lindley's approximation is used to evaluate the Bayes estimates and their posterior risk. As opposed to the classical estimators which require multidimensional numerical search, the elemental percentile method (EPM) estimators which are easy to derive and to compute are developed in Section 5. An extension of the EPM is proposed in Section 6 to handle Types I and II censored data. Graphical displays for informal checks on the appropriateness of the model as well as formal goodness of fit tests for the three-parameter EW family are illustrated in Section 7. In Section 8 the different methods are applied to two real-life data sets with and without censoring. Finally, a Summary is given in Section 9.

2. THE MAXIMUM LIKELIHOOD ESTIMATORS

Let $\mathrm{X} = \{X_1, X_2, ..., X_n\}$ represent a random sample of size n drawn from $f(x;\theta)$ in (2), the log likelihood function, $l(\theta\,|\mathrm{x})$, is given by

$$\begin{aligned} l(\theta\,|\mathrm{x}) &= n\log\alpha + n\log\lambda - n\alpha\log\beta + (\alpha-1)\sum_{i=1}^{n}\log x_i - \sum_{i=1}^{n}\left(\frac{x_i}{\beta}\right)^{\alpha} \\ &\quad + (\lambda-1)\sum_{i=1}^{n}\log\left(1-e^{-\left(\frac{x_i}{\beta}\right)^{\alpha}}\right). \end{aligned} \tag{5}$$

The MLE, $\hat{\theta} = \left\{\hat{\alpha},\ \hat{\beta},\ \hat{\lambda}\right\}$, are obtained by solving the system of equations $\frac{\partial l(\theta|\mathrm{x})}{\partial\theta_i} = 0$, where $\theta_1 = \alpha$, $\theta_2 = \beta$ and $\theta_3 = \lambda$. These maximum likelihood equations are given by

$$\frac{n}{\hat{\alpha}} - n\log\hat{\beta} + \sum_{i=1}^{n}\log x_i - \frac{1}{\hat{\alpha}}\sum_{i=1}^{n}\psi_i\log\psi_i\left[1-\left(\hat{\lambda}-1\right)\varphi_i\right] = 0, \tag{6}$$

$$\frac{\hat{\alpha}}{\hat{\beta}}\left\{-n + \sum_{i=1}^{n}\psi_i\left[1-\left(\hat{\lambda}-1\right)\varphi_i\right]\right\} = 0, \tag{7}$$

and

$$\hat{\lambda} = -n\left[\sum_{i=1}^{n}\log\left(1-e^{-\psi_i}\right)\right]^{-1}, \tag{8}$$

where $\psi_i = \left(\frac{x_i}{\hat{\beta}}\right)^{\hat{\alpha}}$ and $\varphi_i = \frac{e^{-\psi_i}}{\left(1-e^{-\psi_i}\right)}$, for $i = 1,\ 2,\ ...,\ n$. Clearly (6) and (7) are transcendental equations in α and β and no closed form solutions are known. The Newton-Raphson algorithm can be used to find the maximum of the likelihood function and give the MLE. The optimization algorithms are often sensitive to the choice of starting values. Graphical techniques based on the distribution and cumulative hazard functions can be used to provide starting values for the optimization algorithm. The distribution function $F(x;\theta)$ in (1) can be used for this purpose. Taking the log of both sides of (4), we obtain

$$\log x_p = \log\beta + \alpha^{-1}\log\left[-\log\left(1-p^{1/\lambda}\right)\right], \tag{9}$$

which implies that if $\mathrm{x} = (x_1, x_2, ..., x_n)$ has been drawn from (1), the trend of the graph of $\log x_p$ versus $\log\left[-\log\left(1-p^{1/\lambda}\right)\right]$ would be approximately linear providing evidence that the sample comes from the EW family. However, we need to estimate λ to be able to construct the graph. Since a bathtub shaped hazard function is characterized by $\lambda < 1$, successive values of $\lambda < 1$ can be tried (using, e.g., sliders in some statistical packages) until we get a straight line. By taking a series of λ values, fitting a straight line to the plot, and examining the residual sum of squares as the measure of fit, we have a linear model. A model which also provides preliminary estimates of α and β, α is estimated as the inverse of the slope of the line, whereas the intercept provides an estimate for $\log \beta$. By this procedure one can get $\theta_0 = (\beta_0,\ \alpha_0,\ \lambda_0)$ as a starting value for the algorithm.

The observed information matrix, $\hat{\Sigma}_{\hat{\theta}}$, which under mild conditions is a consistent estimator of the covariance matrix of the MLE, $Cov\left(\hat{\theta}_r, \hat{\theta}_j\right) = \Sigma_{\hat{\theta}}$, can be obtained using the second derivatives of $l\left(\theta\,|\mathrm{x}\right)$, the rj-th entry of which is

$$i_{rj} = \left.\frac{-\partial^2 l\left(\theta\,|\mathrm{x}\right)}{\partial\theta_r\partial\theta_j}\right|_{\theta=\hat{\theta}}, \quad r,\ j = 1,\ 2,\ 3, \tag{10}$$

where

$$\begin{aligned}
\frac{\partial^2 l\left(\theta\,|\mathrm{x}\right)}{\partial\alpha^2} &= \frac{-1}{\alpha^2}\left[n + \sum_{i=1}^{n}\psi_i\left(\log\psi_i\right)^2 - (\lambda-1)\sum_{i=1}^{n}\varphi_i\psi_i\left(\log\psi_i\right)^2\left(1-\psi_i-\varphi_i\psi_i\right)\right],\\
\frac{\partial^2 l\left(\theta\,|\mathrm{x}\right)}{\partial\alpha\partial\beta} &= \frac{-1}{\beta}\left[n - \sum_{i=1}^{n}\psi_i\left(1+\log\psi_i\right)\right.\\
&\quad \left. + (\lambda-1)\sum_{i=1}^{n}\varphi_i\psi_i\log\psi_i\left(\frac{1}{\log\psi_i} + 1 - \psi_i - \varphi_i\psi_i\right)\right],\\
\frac{\partial^2 l\left(\theta\,|\mathrm{x}\right)}{\partial\alpha\partial\lambda} &= \frac{1}{\alpha}\sum_{i=1}^{n}\varphi_i\psi_i\log\psi_i,\\
\frac{\partial^2 l\left(\theta\,|\mathrm{x}\right)}{\partial\beta^2} &= \frac{\alpha}{\beta^2}\left[n - (\alpha+1)\sum_{i=1}^{n}\psi_i + (\lambda-1)\sum_{i=1}^{n}\varphi_i\psi_i\left(1+\alpha-\alpha\psi_i-\alpha\varphi_i\psi_i\right)\right],\\
\frac{\partial^2 l\left(\theta\,|\mathrm{x}\right)}{\partial\beta\partial\lambda} &= -\frac{\alpha}{\beta}\sum_{i=1}^{n}\varphi_i\psi_i,\\
\frac{\partial^2 l\left(\theta\,|\mathrm{x}\right)}{\partial\lambda^2} &= \frac{-n}{\lambda^2}.
\end{aligned}$$

Each component $\hat{\theta}_j$ $(j = 1,\ 2,\ 3)$ is asymptotically normal, that is, $\hat{\theta}_j \to N\left(\theta_j, \sigma^2_{\hat{\theta}_j}\right)$, where $\hat{\sigma}^2_{\hat{\theta}_j}$ is the variance of $\hat{\theta}_j$.

The MLE of the pth quantile x_p in (4), $\hat{x}_p$, can be obtained as

$$\hat{x}_p = \hat{\beta}\left[-\log\left(1 - p^{1/\hat{\lambda}}\right)\right]^{1/\hat{\alpha}} \tag{11}$$

and the delta method given in Castillo et al. (2005) can be used to obtain the variance of $\hat{x}_p$, which is

$$V\left(\hat{x}_p\right) = \left(\nabla_\theta x_p\right)^T \sum_{\hat{\theta}} \nabla_\theta x_p, \tag{12}$$

where $(\nabla_\theta x_p)^T = \left[\begin{array}{ccc} \frac{\partial x_p}{\partial \alpha} & \frac{\partial x_p}{\partial \beta} & \frac{\partial x_p}{\partial \lambda} \end{array}\right]^T$. An estimate of $V(\hat{x}_p)$ can be obtained by replacing θ in (12) by its MLE $\hat{\theta}$, that is, $\hat{V}(\hat{x}_p) = \nabla_{\hat{\theta}}^T \hat{x}_p \hat{\Sigma}_{\hat{\theta}} \nabla_{\hat{\theta}} \hat{x}_p$ where the components of $\nabla_\theta x_p$ are given by

$$\frac{\partial x_p}{\partial \alpha} = \frac{-\beta}{\alpha^2} \left[-\log\left(1 - p^{1/\lambda}\right)\right]^{1/\alpha} \log\left[-\log\left(1 - p^{1/\lambda}\right)\right],$$

$$\frac{\partial x_p}{\partial \beta} = \left[-\log\left(1 - p^{1/\lambda}\right)\right]^{1/\alpha},$$

and

$$\frac{\partial x_p}{\partial \lambda} = \frac{-\beta}{\alpha\lambda^2} \left[-\log\left(1 - p^{1/\lambda}\right)\right]^{(1/\alpha)-1} \left[\frac{p^{1/\lambda}\log p}{\left(1 - p^{1/\lambda}\right)}\right].$$

3. THE METHOD OF MOMENTS ESTIMATORS

The rth moment of the three-parameter EW random variable is given by Nassar and Eissa (2003)

$$E(X^r) = \lambda\beta^r \int_0^\infty t^{r/\alpha} e^{-t} \left(1 - e^{-t}\right)^{\lambda-1} dt.$$

We use this expression to derive the MOM estimators, which are the solutions of the system of equations

$$E(X^r) = \tfrac{1}{n}\textstyle\sum_{i=1}^n X_i^r, \text{ for } r = 1,\ 2,\ 3,$$

In general the moments are analytically intractable, but can be studied numerically. We need to solve the following system of nonlinear equations in α, β and λ:

$$f_r(\theta) = \lambda\beta^r \int_0^\infty t^{r/\alpha} e^{-t} \left(1 - e^{-t}\right)^{\lambda-1} dt - \frac{1}{n}\sum_{i=1}^n X_i^r = 0, \quad r = 1, 2, 3. \tag{13}$$

The Newton-Raphson method for solving a system of nonlinear equations can be used as follows:

1. Take $\theta^{(0)} = \left[\begin{array}{ccc} \alpha^{(0)} & \beta^{(0)} & \lambda^{(0)} \end{array}\right]$ a starting value for $\theta = \left[\begin{array}{ccc} \alpha & \beta & \lambda \end{array}\right]$.

2. Compute $f\left(\theta^{(k)}\right)$, where the superscript k denotes values obtained on the kth iteration,

$$f(\theta) = \left(f_1(\theta), f_2(\theta), f_3(\theta)\right)^T$$

and $f_r(\theta) = E(X^r) - \frac{1}{n}\sum_{i=1}^n X_i^r$, for $r = 1,\ 2,\ 3.$

3. Compute $J\left(\theta^{(k)}\right)$, the Jacobian matrix of $f\left(\theta\right)$ evaluated at $\theta^{(k)}$. The $(r, j)\,th$ element of the Jacobian matrix J is $\frac{\partial f_r(\theta)}{\partial \theta_j}$, for $r,\ j = 1,\ 2,\ 3$.

4. The general expression for Newton-Raphson's method can be written as

$$\theta^{(k+1)} = \theta^{(k)} - J^{-1}\left(\theta^{(k)}\right) f\left(\theta^{(k)}\right).$$

5. Repeat Steps 2 to 4 until convergence. The algorithm is terminated when the magnitude of the computed change in the value of the root, θ, is less than some predetermined quantity, ε.

The Monte Carlo approach will be used to approximate the integrals in the elements of $f\left(\theta^{(k)}\right)$ and $J\left(\theta^{(k)}\right)$ for specific values of α, β and λ as follows:

$$I = \int_0^{\infty} g\left(t\right) dt = \int_0^1 g\left(\frac{1}{u} - 1\right) \frac{1}{u^2} du = \int_0^1 h\left(u\right) du = E\left[h(u)\right].$$

Hence if U_1, U_2, ..., U_k are independent Uniform (0, 1) random variables, it follows that the random variables $h\left(u_1\right),\ h\left(u_2\right),\ ...,\ h\left(u_k\right)$ are independent and identically distributed random variables having mean I. Therefore by the strong law of large numbers it follows that, with probability 1,

$$k^{-1} \sum_{i=1}^{k} h(u_i) \ \rightarrow\ E\left[h\left(u\right)\right] = I, \quad \text{as } k \ \rightarrow \infty.$$

4. THE BAYESIAN APPROACH

In this section the Bayesian approach is utilized to derive estimates of the parameters in (1), assuming we are in a situation where little is known a priori about the values of the parameters. Under this assumption it may be appropriate to resort to use of a diffuse prior. Following the rule of Jeffreys (1961) applied separately to each of the three parameters, and then multiplying the resulting forms we arrive at the overall prior specification which is given by

$$g\left(\alpha,\ \beta,\ \lambda\right) \propto \left(\alpha\beta\lambda\right)^{-1}, \alpha,\ \beta,\ \lambda > 0.$$

If L is the likelihood function indexed by a continuous parameter $\theta = (\alpha, \beta, \lambda)$, with prior density $g\left(\theta\right) = g\left(\alpha,\ \beta,\ \lambda\right)$, then the posterior density for θ is given by

$$g\left(\theta \left| x\right.\right) = \frac{L\left(\theta\right) g\left(\theta\right)}{\int L\left(\theta\right) g\left(\theta\right) d\theta}.$$

In our case it is possible only to know the kernel of the posterior distribution, but not the normalizing constant. In this article, we consider Lindley's approximation to evaluate the Bayes estimate for arbitrary function of θ, say $u(\theta)$, and the corresponding Bayes risk. Lindley's approximation provides a numerical approximation method which is useful when the number of parameters is small (≤ 5); Press (2003). This approximation is generally accurate enough, but requires the evaluation of third derivatives of the likelihood which can be rather tedious in problems with several parameters. Lindley's approximation provides an approximation for

$$E\left[u(\theta)\,|\mathrm{x}\right] = \frac{\int u(\theta)\,L(\theta)\,g(\theta)\,d\theta}{\int L(\theta)\,g(\theta)\,d\theta}. \tag{14}$$

For the special case of three parameters $\theta = (\theta_1,\, \theta_2,\, \theta_3)$, Lindley's approximation of (14) is given by

$$\begin{aligned} E\left[u(\theta)\,|\mathrm{x}\right] \cong u\left(\hat{\theta}\right) + \tfrac{1}{2}\textstyle\sum_{i=1}^{3}\sum_{j=1}^{3}\hat{\sigma}_{ij}\left[\frac{\partial^2 u(\theta)}{\partial\theta_i\partial\theta_j} + 2\left(\frac{\partial u(\theta)}{\partial\theta_i}\right)\left(\frac{\partial \log g(\theta)}{\partial\theta_j}\right)\right]\Bigg|_{\theta=\hat{\theta}} \\ + \tfrac{1}{2}\textstyle\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=1}^{3}\hat{\sigma}_{ij}\hat{\sigma}_{kl}\left(\frac{\partial^3 l(\theta|\mathrm{x})}{\partial\theta_i\partial\theta_j\partial\theta_l}\right)\left(\frac{\partial u(\theta)}{\partial\theta_k}\right)\Bigg|_{\theta=\hat{\theta}}, \end{aligned} \tag{15}$$

where $\hat{\sigma}_{ij}$ denotes the $(i,\, j)$th element of the inverse of the observed information matrix. All functions on the right hand side of (15) are evaluated at $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$, replacing the unknown parameters by their MLE.

When $u(\theta) = \theta_k$, for $k = 1,\ 2,\ 3$, (15) reduces to

$$E\left(\theta_k\,|\mathrm{x}\right) \cong \hat{\theta}_k + \left.\left(\begin{array}{l} \sum_{i=1}^{3}\left(\frac{\partial \log g(\theta)}{\partial\theta_i}\right)\hat{\sigma}_{ki} \\ +\frac{\hat{\sigma}_{k1}}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial^3 l(\theta|\mathrm{x})}{\partial\theta_i\partial\theta_j\partial\theta_1}\right)\hat{\sigma}_{ij} \\ +\frac{\hat{\sigma}_{k2}}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial^3 l(\theta|\mathrm{x})}{\partial\theta_i\partial\theta_j\partial\theta_2}\right)\hat{\sigma}_{ij} \\ +\frac{\hat{\sigma}_{k3}}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial^3 l(\theta|\mathrm{x})}{\partial\theta_i\partial\theta_j\partial\theta_3}\right)\hat{\sigma}_{ij} \end{array}\right)\right|_{\theta=\hat{\theta}}, \tag{16}$$

which gives the Bayes estimate of θ_k, the mean of the posterior distribution, $E\left(\theta_k\,|\mathrm{x}\right)$, under squared error loss function.

When $u(\theta) = \theta_k^2$, (15) provides the following approximation for $E\left(\theta_k^2\,|\mathrm{x}\right)$

$$E\left(\theta_k^2\,|\mathrm{x}\right) \cong \hat{\theta}_k^2 + \hat{\sigma}_{kk} + \hat{\theta}_k \left.\left(\begin{array}{l} 2\sum_{i=1}^{3}\left(\frac{\partial \log g(\theta)}{\partial\theta_i}\right)\hat{\sigma}_{ki} \\ +\hat{\sigma}_{k1}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial^3 l(\theta|\mathrm{x})}{\partial\theta_i\partial\theta_j\partial\theta_1}\right)\hat{\sigma}_{ij} \\ +\hat{\sigma}_{k2}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial^3 l(\theta|\mathrm{x})}{\partial\theta_i\partial\theta_j\partial\theta_2}\right)\hat{\sigma}_{ij} \\ +\hat{\sigma}_{k3}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial^3 l(\theta|\mathrm{x})}{\partial\theta_i\partial\theta_j\partial\theta_3}\right)\hat{\sigma}_{ij} \end{array}\right)\right|_{\theta=\hat{\theta}}. \tag{17}$$

Under quadratic loss, the Bayes risk is the posterior variance, and the results of (16) and (17) can be used to derive an estimate for the posterior variance of θ_k. This variance is given by

$$V\left(\theta_k\,|\mathrm{x}\right) \cong \hat{\sigma}_{kk} - \left.\left(\begin{array}{l} \sum_{i=1}^{3}\left(\frac{\partial \log g(\theta)}{\partial\theta_i}\right)\hat{\sigma}_{ki} \\ +\frac{\hat{\sigma}_{k1}}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial^3 l(\theta|\mathrm{x})}{\partial\theta_i\partial\theta_j\partial\theta_1}\right)\hat{\sigma}_{ij} \\ +\frac{\hat{\sigma}_{k2}}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial^3 l(\theta|\mathrm{x})}{\partial\theta_i\partial\theta_j\partial\theta_2}\right)\hat{\sigma}_{ij} \\ +\frac{\hat{\sigma}_{k3}}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial^3 l(\theta|\mathrm{x})}{\partial\theta_i\partial\theta_j\partial\theta_3}\right)\hat{\sigma}_{ij} \end{array}\right)^2\right|_{\theta=\hat{\theta}}. \tag{18}$$

Equation (18) could simply be rewritten as $V\left(\theta_k\,|\mathrm{x}\right) \cong \hat{\sigma}_{kk} - \left[E\left(\theta_k\,|\mathrm{x}\right) - \hat{\theta}_k\right]^2$. For the evaluation of (16) and (18) the following third derivatives are obtained from (10):

$$\begin{aligned}
\frac{\partial^3 l\left(\theta\,|\mathrm{x}\right)}{\partial\alpha^3} &= \frac{1}{\alpha^3}\left(2n - \sum_{i=1}^{n}\psi_i\left(\log\psi_i\right)^3\right) \\
&\quad + \frac{\lambda-1}{\alpha^3}\sum_{i=1}^{n}\varphi_i\psi_i\left(\log\psi_i\right)^3\left(1 - 3\psi_i + \psi_i^2 - 3\varphi_i\psi_i + 3\varphi_i\psi_i^2 + 2\varphi_i^2\psi_i^2\right), \\
\frac{\partial^3 l\left(\theta\,|\mathrm{x}\right)}{\partial\alpha^2\partial\beta} &= \frac{1}{\alpha\beta}\sum_{i=1}^{n}\psi_i\left(\log\psi_i\right)\left(2 + \log\psi_i\right) \\
&\quad - \frac{\lambda-1}{\alpha\beta}\sum_{i=1}^{n}\varphi_i\psi_i^2\left(\log\psi_i\right)^2\left(\begin{array}{l} \frac{2}{\psi_i\log\psi_i} - \frac{2(1+\varphi_i)}{\log\psi_i} + \frac{1}{\psi_i} - 3 \\ +\psi_i - 3\varphi_i + 3\varphi_i\psi_i + 2\varphi_i^2\psi_i \end{array}\right), \\
\frac{\partial^3 l\left(\theta\,|\mathrm{x}\right)}{\partial\alpha\partial\beta^2} &= \frac{1}{\beta^2}\left(\begin{array}{l} n - (2\alpha+1)\sum_{i=1}^{n}\psi_i\left[1 - (\lambda-1)\varphi_i\right] \\ -2\alpha\left(\lambda-1\right)\sum_{i=1}^{n}\varphi_i\psi_i^2\left(1+\varphi_i\right) \\ -\left(\alpha+1\right)\sum_{i=1}^{n}\psi_i\log\psi_i\left[1 - (\lambda-1)\varphi_i\right] \\ -\left(3\alpha+1\right)\left(\lambda-1\right)\sum_{i=1}^{n}\varphi_i\psi_i^2\log\psi_i\left(1+\varphi_i\right) \\ +\alpha\left(\lambda-1\right)\sum_{i=1}^{n}\varphi_i\psi_i^3\log\psi_i\left(1 + 3\varphi_i + 2\varphi_i^2\right) \end{array}\right), \\
\frac{\partial^3 l\left(\theta\,|\mathrm{x}\right)}{\partial\beta^3} &= \frac{-2n\alpha}{\beta^3} + \frac{\alpha\left(\alpha+1\right)\left(\alpha+2\right)}{\beta^3}\sum_{i=1}^{n}\psi_i \\
&\quad - \frac{\alpha\left(\lambda-1\right)}{\beta^3}\sum_{i=1}^{n}\varphi_i\psi_i\left(\begin{array}{l} \left(\alpha+1\right)\left(\alpha+2\right) - 3\alpha\left(\alpha+1\right)\psi_i \\ -3\alpha\left(\alpha+1\right)\varphi_i\psi_i + \alpha^2\psi_i^2 \\ +3\alpha^2\varphi_i\psi_i^2 + 2\alpha^2\varphi_i^2\psi_i^2 \end{array}\right), \\
\frac{\partial^3 l\left(\theta\,|\mathrm{x}\right)}{\partial\alpha^2\partial\lambda} &= \frac{1}{\alpha^2}\sum_{i=1}^{n}\varphi_i\psi_i\left(\log\psi_i\right)^2\left(1 - \psi_i - \varphi_i\psi_i\right), \\
\frac{\partial^3 l\left(\theta\,|\mathrm{x}\right)}{\partial\alpha\partial\beta\partial\lambda} &= \frac{-1}{\beta}\sum_{i=1}^{n}\varphi_i\psi_i\log\psi_i\left(\frac{1}{\log\psi_i} + 1 - \psi_i - \varphi_i\psi_i\right),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^3 l(\theta\,|\mathrm{x})}{\partial\alpha\partial\lambda^2} &= \frac{\partial^3 l(\theta\,|\mathrm{x})}{\partial\beta\partial\lambda^2} = 0,\\
\frac{\partial^3 l(\theta\,|\mathrm{x})}{\partial\beta^2\partial\lambda} &= \frac{\alpha}{\beta^2}\sum_{i=1}^{n}\varphi_i\psi_i\left(1+\alpha-\alpha\psi_i-\alpha\varphi_i\psi_i\right),\\
\frac{\partial^3 l(\theta\,|\mathrm{x})}{\partial\lambda^3} &= \frac{2n}{\lambda^3}
\end{aligned}$$

These third derivatives of the likelihood will allow for the calculation of the Bayes estimates of α, β and λ and their Bayes risks.

Equation (15) can be used to derive the Bayes estimate of x_p, where $u(\theta) = x_p$. No much simplification to (15) will take place with this definition of $u(\theta)$. To obtain the respective estimate for the Bayes risk, (15) is used where $u(\theta)$ is defined as x_p^2.

5. THE ELEMENTAL PERCENTILE ESTIMATORS

It is clear from Sections 2–4 that the classical estimation methods encounter some difficulties. From Section 3 the moments are complicated functions of the parameters. The maximum likelihood, the MOM, and Bayes estimators all require multidimensional numerical search.

In this section estimators of the parameters and the quantiles are derived using the *elemental percentile method* (EPM) proposed by Castillo and Hadi (1995a). These estimates are easy to derive and to compute and they are unique and well-defined for all parameter and sample values. Simulation studies by Castillo and Hadi (1995a,b, 1997) indicate that no method is uniformly the best for all $\theta \in \Theta$, but this method performs well compared to all other methods.

The EPM obtains the estimates in two steps: First, some elemental estimates are obtained by solving equations relating the cumulative distribution function to their percentile values for some elemental subsets of the observations. These elemental estimates are then combined to produce final more efficient and robust estimates of the parameters and quantiles.

More specifically, let $\mathrm{X} = \{X_1, X_2, ..., X_n\}$ be independent identically distributed variables having the common cumulative distribution function given in (1). Then we have $F(x_{i:n};\theta) \cong p_{i:n}$, or, equivalently, $x_{i:n} \cong F^{-1}(p_{i:n};\theta)$ for $i = 1,\ 2,\ ...,\ n$, where $X_{i:n}$ is the ith order statistic and $p_{i:n}$ are empirical estimates of $F(x_i;\theta)$ or suitable plotting positions defined by $p_{i:n} = \frac{i-a}{n+b}$ $(i = 1, 2, ..., n)$, for appropriate choices of $a \geq 0$ and $b \geq 0$. Some plotting positions include

$$p_{i:n} = \frac{i}{n+1}, \quad p_{i:n} = \frac{i-0.375}{n+0.25} \quad p_{i:n} = \frac{i-0.5}{n} \quad \text{and} \quad p_{i:n} = \frac{i-0.44}{n+0.12}.$$

For justification of these formulas see, for example, Castillo (1988). A simulation study carried out by Castillo and Hadi (1997) has shown that the choice of $a = 0$ and $b = 1$, which is the choice utilized in this article, gives better results for the proposed method.

From $F(x;\theta)$ in (4) it follows that the $p_{i:n}$-th quantile is

$$x_{i:n} \cong \beta \left[-\log\left(1 - p_{i:n}^{1/\lambda}\right)\right]^{1/\alpha}. \tag{19}$$

Let $x_{i:n}$, $x_{j:n}$, and $x_{k:n}$ be a set of three distinct order statistics in a random sample of size n from $F(x;\theta)$ in (1). This set is referred to as an elemental subset. Then for this elemental subset of order statistics and by replacing the approximation in (19) by an equality, we get the following three equations in three unknowns α, β and λ:

$$x_{i:n} = \beta \left[-\log\left(1 - p_{i:n}^{1/\lambda}\right)\right]^{1/\alpha}, \tag{20}$$

$$x_{j:n} = \beta \left[-\log\left(1 - p_{j:n}^{1/\lambda}\right)\right]^{1/\alpha}, \tag{21}$$

and

$$x_{k:n} = \beta \left[-\log\left(1 - p_{k:n}^{1/\lambda}\right)\right]^{1/\alpha}. \tag{22}$$

The solution of (20) to (22) is obtained by the elimination method as follows. From (20) and (21), we have

$$\log(x_{i:n}/x_{j:n}) = \alpha^{-1} \log\left[\log\left(1 - p_{i:n}^{1/\lambda}\right) / \log\left(1 - p_{j:n}^{1/\lambda}\right)\right]. \tag{23}$$

Similarly, from (20) and (22), we have

$$\log(x_{i:n}/x_{k:n}) = \alpha^{-1} \log\left[\log\left(1 - p_{i:n}^{1/\lambda}\right) / \log\left(1 - p_{k:n}^{1/\lambda}\right)\right]. \tag{24}$$

Dividing (23) by (24), we obtain

$$\frac{\log(x_{i:n}/x_{j:n})}{\log(x_{i:n}/x_{k:n})} = \frac{\log\left[\log\left(1 - p_{i:n}^{1/\lambda}\right) / \log\left(1 - p_{j:n}^{1/\lambda}\right)\right]}{\log\left[\log\left(1 - p_{i:n}^{1/\lambda}\right) / \log\left(1 - p_{k:n}^{1/\lambda}\right)\right]}, \tag{25}$$

which is an equation in only one unknown λ, hence it can be solved easily using several methods. An initial estimate of λ, which depends on $x_{i:n}, x_{j:n}$, and $x_{k:n}$, can be obtained by solving (25). Once an estimate for λ is obtained, estimates for α and β are obtained in closed-forms as follows:

$$\hat{\alpha} = \frac{\log\left[\log\left(1 - p_{i:n}^{1/\lambda}\right) / \log\left(1 - p_{j:n}^{1/\lambda}\right)\right]}{\log(x_{i:n}/x_{j:n})}, \tag{26}$$

and

$$\hat{\beta} = x_{i:n} \left[-\log\left(1 - p_{i:n}^{1/\lambda}\right)\right]^{-1/\alpha}. \tag{27}$$

Hence, the estimates are obtained one at a time solving only one unidimensional equation, instead of solving a complicated multidimensional non-linear system of equations.

The elemental estimate of $\theta = \{\alpha, \beta, \lambda\}$ obtained from these equations depends on an elemental subset with three distinct order statistics $x_{i:n}, x_{j:n}$, and $x_{k:n}$. The computation of all ${}^nC_3$ elemental estimates may not be feasible. Instead a prespecified number, M, of elemental subsets may be selected, where M is a number to be specified by the data analyst. For each of these subsets, an elemental estimate of θ can be computed. If $\hat{\theta}_j = \{\hat{\alpha}_j, \hat{\beta}_j, \hat{\lambda}_j\}$ is the elemental estimate of θ based on the elemental subset j for $j = 1, 2, \ldots, M$, these estimates can then be combined, using some suitable robust functions to obtain an overall final estimate of θ. Examples of robust functions include the median (MED) and the $\gamma-$trimmed mean (TM_γ), where γ indicates the percentage of trimming. The MED estimators are very robust but inefficient while the TM_γ estimators are less robust but more efficient than the MED estimators. The larger the trimming, the more robust and less efficient are the TM_γ estimators.

The estimator of the quantile x_p can then be obtained by substituting the parameter estimates for θ in (4), that is,

$$\hat{x}_p = F^{-1}\left(p; \hat{\theta}\right) = \hat{\beta}\left[-\log\left(1 - p^{1/\hat{\lambda}}\right)\right]^{1/\hat{\alpha}}.$$

The variance of the resultant estimates can be easily obtained using sampling methods such as the bootstrap method; Efron (1979). The bootstrap sampling can be carried out in two ways: The samples can be drawn directly from the data or they can be drawn parametrically from $F(x; \hat{\theta})$. The parametric bootstrap is preferable in obtaining the variance of the estimates of a particular method; Castillo and Hadi (1997).

Although the EPM estimates are easy to derive and to compute and they are unique and well defined for all parameter and sample values, they do not resolve a difficulty frequently associated with life testing. This is the presence of censoring which creates special problems in the analysis of the data. For this reason, we extend the EPM in the next section to deal with censored data.

6. AN EXTENSION OF THE ELEMENTAL PERCENTILE METHOD

Censoring is common in lifetime data because of time limits and other restrictions on data collection. Censoring occurs when exact lifetimes are known for only a portion of the individuals or items under study; the remainder of the lifetimes are known only to exceed certain values. Some work has been done to handle Type II censoring for the EW model; see for example Singh et al. (2005a,b). Under this sampling scheme only the r smallest observations in a random sample of n items are observed ($1 \leq r \leq n$). Fitting the EW model under Type I censoring has not been considered in the literature.

In this section, we propose an extension of the EPM which handles Types I and II censored data. As described in Section 5, the EPM estimates are obtained in two steps. Here we adjust the first to reflect censoring by equating the survivor function $S(x;\theta) = 1 - F(x;\theta)$ at the observed order statistics to their corresponding empirical survivor function, and the resulting equations will then be used as a basis for obtaining initial estimates of the parameters. The second step then follows by combining these estimates in the same manner described earlier to obtain final estimates of the parameters.

The product limit (Kaplan Meier) estimate provides a modification of the empirical survivor function to deal with censored data. For a detailed description of this estimate, see for example Lawless (2003). The product limit estimate of $S(x)$ is defined as

$$\hat{S}(x) = \prod_{j:x_{j:n}<x} \frac{n_j - d_j}{n_j}, \text{ for } j = 1,\ 2,\ ...,\ k,$$

where $x_{1:n} < \ x_{2:n} < ... < x_{k:n}$ are k distinct times at which failures occur ($k = n$ in case of complete sampling), d_j is the number of deaths (failures) at $x_{j:n}$ and n_j is the number of individuals (units) at risk at $x_{j:n}$, that is, the number of individuals (units) alive (unfailed) and uncensored just prior to $x_{j:n}$.

By equating the survivor function $S(x;\theta) = 1 - F(x;\theta)$ at the observed order statistics to their corresponding empirical survivor function we have

$$S(x_{j:n};\theta) \cong \hat{S}(x_{j:n}) = \hat{s}_{j:n},$$

or equivalently

$$x_{j:n} \cong S^{-1}(\hat{s}_{j:n};\theta).$$

From $F(x;\theta)$ in (1) it follows that

$$x_{j:n} \cong \beta\left[-\log\left(1 - (1 - \hat{s}_{j:n})^{1/\lambda}\right)\ \right]^{1/\alpha}. \tag{28}$$

As a first check for the adequacy of the EW model for the censored data, we take the log of both sides of (28) and obtain

$$\log x_{j:n} \cong \log\beta + \alpha^{-1}\log\left[-\log\left(1 - (1 - \hat{s}_{j:n})^{1/\lambda}\right)\ \right], \tag{29}$$

which implies that, given λ, the trend in the graph of

$$\log x_{j:n} \quad \text{versus} \quad w_\lambda = \log\left[-\log\left(1 - (1 - \hat{s}_{j:n})^{1/\lambda}\right)\ \right] \tag{30}$$

would be approximately linear providing evidence that the sample comes from the EW family. The optimal values of λ can be obtained, as in (9), by taking a series of λ values, fitting a straight line to the plot, and examining the residual sum of squares as the measure of fit, we have a linear model.

Note that for the purpose of estimating λ using least squares, it is better to use (29) instead of

$$\log\left[-\log\left(1-(1-\hat{s}_{j:n})^{1/\lambda}\right)\right] \cong \alpha\log x_{j:n} + \alpha\log\beta, \tag{31}$$

This is because it can be shown that, using (31), the sum of squared residual is a monotonically decreasing function of λ and hence the least squares estimate of λ is infinity. On the other hand, there are unique least squares estimates of all three parameters when we use (29).

Now, to obtain the EPM, we replace (20) to (22) by

$$x_{i:n} \cong \beta\left[-\log\left(1-(1-\hat{s}_{i:n})^{1/\lambda}\right)\right]^{1/\alpha}, \tag{32}$$

$$x_{j:n} \cong \beta\left[-\log\left(1-(1-\hat{s}_{j:n})^{1/\lambda}\right)\right]^{1/\alpha}, \tag{33}$$

and

$$x_{k:n} \cong \beta\left[-\log\left(1-(1-\hat{s}_{k:n})^{1/\lambda}\right)\right]^{1/\alpha}. \tag{34}$$

When no censoring occurs and all lifetimes are distinct, $(1-\hat{s}_{i:n})$ reduces to the plotting position $p_{i:n} = (i-a)/(n+b)$, $i = 1, 2, ..., n$, where $a = b = 0$ and the results reduce to the EPM estimates described in Section 5.

7. GOODNESS OF FIT

It is important to check the adequacy of models upon which inferences are based. In this section methods for testing goodness of fit of the EW model are discussed. Some graphical methods as well as more formal analytical procedures are used in assessing the adequacy of the model.

The total time on test (TTT) transform, introduced by Barlow et al. (1972), is a convenient tool for examining the nature of the hazard rate and accordingly checking for the adequacy of a model to represent the failure behavior of the data. Barlow and Campo (1975) illustrate the theoretical basis of the TTT plots and how do they converge to a transform of the underlying probability distribution as the sample size increases. Aarset (1987) presents a test statistic based on the TTT plot for testing exponentiality (constant hazard rate) versus bathtub shaped hazard rate. The TTT transform of a probability distribution with absolutely continuous distribution function $F(\cdot)$ is given in Aarset (1987) as

$$H_F^{-1}(x) = \int_0^{F^{-1}(x)} \left[1 - F(u)\right]\, du.$$

The scaled TTT transform is given by

$$\phi_F(x) = \frac{H_F^{-1}(x)}{H_F^{-1}(1)}.$$

The empirical TTT transform is given by

$$H_n^{-1}\left(\frac{i}{n}\right) = \int_0^{F_n^{-1}(i/n)} [1 - F_n(u)]\, du,$$

while the scaled empirical TTT transform takes the form

$$\phi_n(i/n) = \frac{H_n^{-1}(i/n)}{H_n^{-1}(1)} = \hat{\mu}^{-1}\sum_{j=1}^{i}(n-j+1)(x_{j:n} - x_{j-1:n}), \tag{35}$$

for $i = 1,\ 2,\ ...,\ n$, where $x_{0:n} = 0$ and $\hat{\mu} = \sum_{j=1}^{n}(n-j+1)(x_{j:n} - x_{j-1:n})$ is the total time on test or the total observed lifetime.

A TTT data plot is a graph based on the data which tends to look like the TTT transform corresponding to the underlying distribution, $F(\cdot)$. This scaled TTT plot should be compared with suitably scaled (to 1) TTT transforms for model identification. Hence given a sample from a population, the empirical scaled TTT transform can be used to explore the hazard shape of the population.

Barlow et al. (1972) prove that if $F(\cdot)$ is strictly increasing, $\phi_n(i/n) \to \phi_F(x)$ uniformly on $(0,\ 1)$ with probability one as $(i/n) \to t$ and $n \to \infty$. If the scaled TTT transform $\phi_F(x)$ is a 45-degree straight line inside the unit square, then the hazard function $h(x)$ is constant, demonstrating the appropriateness of the exponential model. If $\phi_F(x)$ is convex, then the hazard function $h(x)$ is decreasing, whereas if it is concave then $h(x)$ is increasing. This suggests the adequacy of the Weibull model with the appropriate shape parameter (less than or greater than 1). If $\phi_F(x)$ is concave-convex, then $h(x)$ is unimodal, and it is convex-concave if $h(x)$ is bathtub shaped, illustrating the adequacy of the EW model; Mudholkar and Hutson (1996).

Jiang and Murthy (1999) provide a parametric characterization for the EW family. For a small x, they show that

$$h(x) \approx \frac{\alpha\lambda}{\beta}\left(\frac{x}{\beta}\right)^{\alpha\lambda-1}$$

and, for a large x, $h(x)$ can be approximated by $h(x) \approx (\alpha/\beta)(x/\beta)^{\alpha-1}$. Thus for small x, $h(x)$ can be approximated by the failure rate of a two-parameter Weibull distribution with shape parameter $\alpha\lambda$ and scale parameter β. For large x, $h(x)$ can be approximated by the failure rate of a two-parameter Weibull distribution with shape parameter α and scale parameter β.

We use this characterization to provide yet another method, based on the cumulative hazard function, $H(x) = -\log S(x)$, for informally checking the appropriateness of the EW model. For small x, $H(x) \approx (x/\beta)^{\alpha\lambda}$ and for large x, $H(x) \approx (x/\beta)^{\alpha}$. We then have

$$\log H(x) \approx \alpha\lambda \log x - \alpha\lambda \log \beta \quad \text{and} \quad \log H(x) \approx \alpha\lambda \log x - \alpha\lambda \log \beta, \tag{36}$$

for small and large x, respectively.

Since the cumulative hazard function is $H(x) = -\log S(x)$, then a natural estimate of it is $\hat{H}(x) = -\log \hat{S}(x)$, where $\hat{S}(x)$ is the product-limit estimate of the survivor function described in Section 6.

An alternate estimate of $H(x)$ is the empirical cumulative hazard function given by

$$\tilde{H}(x) = \sum_{j:x_{j:n}<x} \frac{d_j}{n_j}.$$

Since

$$\hat{H}(x) = -\sum_{j:x_{j:n}<x} \log\left(1 - \frac{d_j}{n_j}\right) = \sum_{j:x_{j:n}<x} \left(\frac{d_j}{n_j} + \frac{d_j^2}{2n_j^2} + ...\right),$$

then $\tilde{H}(x)$ is a first-order approximation for $\hat{H}(x)$ and for continuous models $\hat{H}(x)$ and $\tilde{H}(x)$ are asymptotically equivalent, and do not differ greatly, except for large values of x; Lawless (2003). Plots of $\hat{H}(x)$ and $\tilde{H}(x)$ can be used for informal checks on the appropriateness of EW model. A plot of $\log \hat{H}(x)$ or $\log \tilde{H}(x)$ versus $\log x$ should be roughly linear if the EW model is appropriate.

Checking for the adequacy of the EW family could also be done using formal tests. Mudholkar and Srivastava (1993) show that the EW family can be used to test not only exponentiality but also goodness of fit of the Weibull model. The problem of testing goodness of fit of a Weibull model against the unrestricted class of alternatives is complex. However, by restricting the alternatives to the EW family, we can use the usual likelihood ratio statistic when both the null and alternative hypotheses are composite to test the adequacy of an exponential or a Weibull sub-model. The likelihood ratio statistics are

$$\Lambda_1 = \frac{L(\alpha = 1, \beta_e, \lambda = 1)}{L(\alpha, \beta, \lambda)} \quad \text{and} \quad \Lambda_2 = \frac{L(\alpha_w, \beta_w, \lambda = 1)}{L(\alpha, \beta, \lambda)}, \tag{37}$$

for the null hypotheses corresponding to the exponential and Weibull sub-models respectively. $L(\alpha = 1,\ \beta_e,\ \lambda = 1)$, $L(\alpha_w,\ \beta_w,\ \lambda = 1)$ and $L(\alpha,\ \beta,\ \lambda)$ are the maximum of the likelihood functions replacing α, β and λ by their MLE under the exponential, Weibull and EW distributions respectively. Under the exponential model, the maximum likelihood estimator for β is the sample mean, $\bar{X}$. The MLE of α and β under the Weibull model can be determined by solving the following equations given in Lawless (2003) using an iterative

Table 1: Appliance Failure Data.

14	34	59	61	69	80	123	142	165	210
381	464	479	556	574	839	917	969	991	1064
1088	1091	1174	1270	1275	1355	1397	1477	1578	1649
1702	1893	1932	2001	2161	2292	2326	2337	2628	2785
2811	2886	2993	3122	3248	3715	3790	3857	3912	4100
4106	4116	4315	4510	4584	5267	5299	5583	6065	9701

procedure such as Newton's method:

$$\hat{\beta}_w = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{\hat{\alpha}_w} \right)^{1/\hat{\alpha}_w} \tag{38}$$

and

$$\frac{\sum_{i=1}^{n} x_i^{\hat{\alpha}_w} \log x_i}{\sum_{i=1}^{n} x_i^{\hat{\alpha}_w}} - \frac{1}{\hat{\alpha}_w} - \frac{1}{n} \sum_{i=1}^{n} \log x_i = 0. \tag{39}$$

Small values of the ratios Λ_1 and Λ_2 would lead to rejection of the null hypothesis. Under the null hypothesis $-2\log\Lambda_1$ follows a chi-square distribution with one degree of freedom and $-2\log\Lambda_2$ follows a chi-square distribution with two degrees of freedom.

8. EXAMPLES

Two real-life data sets with and without censoring are used to illustrate the methods proposed in this article. Example 1 deals with a complete data set and MLE, MOM, Bayes, and EPM estimates are developed. Example 2 deals with a general Type I censored data set in which each item has its own censoring time. The extension of the EPM proposed in Section 6 is utilized to derive estimates for the parameters.

8.1 **Example 1**

The data in Lawless (2003) showing the numbers of cycles to failure for a group of 60 electrical appliances in a life test are given in Table 1. As has been shown by Lawless (2003), this data set is actually challenging because the common distributions such as the Weibull, log-normal, and gamma distributions are not appropriate for modeling this data set. Lawless (2003) thus considers a mixture of two Weibull distributions for modeling these data. However, the likelihood function of the model involves five parameters and it is difficult to maximize. Lawless (2003) also discusses the difficulty of handling mixture models with several parameters with regard to formal estimation methods.

A plot of the empirical scaled TTT transform $\phi_n\left(i/n\right)$ as well as the 45-degree straight line inside the unit square are shown in Figure 1(a). It is clear from the graph that the

TTT-transform is first convex and then concave identifying a bathtub failure rate. Figure 1(a) shows that the EW model is found adequate for this data set and thus, as suggested by (9), we consider plotting $\log x_p$ versus $w_\lambda = \log\left[-\log\left(1 - p^{1/\lambda}\right)\right]$, for different values of λ. For $\lambda = 0.2515$, the scatter plot shown in Figure 1(b) exhibits a strong linear relationship ($R^2 = 98.4\%$) with slope 0.3875 and 8.5018. Therefore, the EW assumption seems reasonable. The associated estimates for α and β are 2.5809 and 4923.9681, respectively. These estimates together with $\lambda = 0.2515$ are set as starting values for the Newton Raphson algorithm used in solving the maximum likelihood equations (6)–(8) for α, β and λ and deriving the standard errors of the estimates.

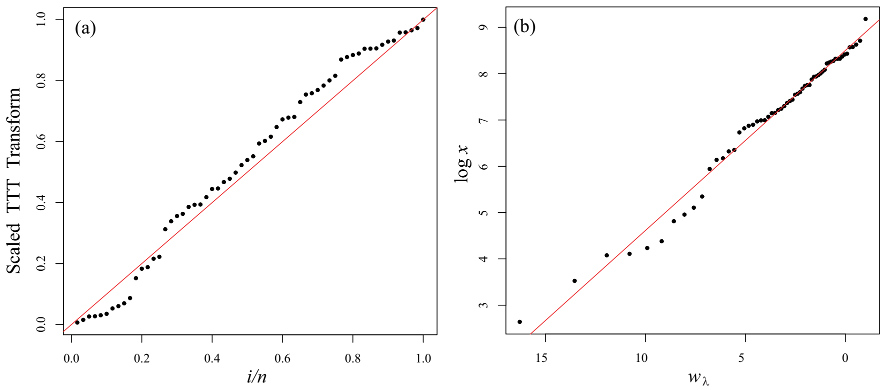


Figure 1: Appliance Failure Data: (a) Empirical scaled TTT transform, and (b) Plot of $\log x_p$ versus $w_\lambda = \log[-\log(1 - p^{1/\lambda})]$, for $\lambda = 0.2515$.

The scatter plots of $y = \log \hat{H}(x)$ and $\log \tilde{H}(x)$ versus $\log x$, in Figure 2, show a strong linear relationship ($R^2 = 98.54\%$ and 98.33%, respectively) supporting the appropriateness of the EW model.

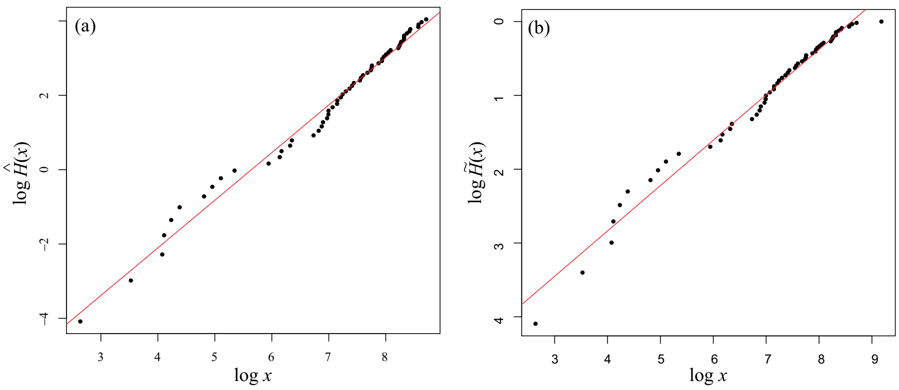


Figure 2: Appliance Failure Data: (a) Plot of $\log \hat{H}(x)$ versus $\log x$, and (b) Plot $\log \tilde{H}(x)$ versus $\log x$.

An EW fit for the data in Table 1 is obtained by solving the maximum likelihood equations (6)–(8) for α, β and λ using the Newton Raphson algorithm with the starting values for α, β

Table 2: Appliance Failure Data: Estimates of the parameters and their standard errors.

Method	$\hat{\alpha}$	$s.e.(\hat{\alpha})$	$\hat{\beta}$	$s.e.(\hat{\beta})$	$\hat{\lambda}$	$s.e.(\hat{\lambda})$
MLE	1.9599	0.6420	4159.80	979.59	0.3716	0.1658
MOM	1.6690	0.6472	3600.54	794.03	0.4955	0.1431
Bayes	1.8999	0.6392	4039.87	972.22	0.4436	0.1492
EPM (MED)	2.4256	0.2719	4538.23	515.69	0.2879	0.0430
EPM (TM)	2.4354	0.2853	4559.30	594.63	0.2918	0.0413

and λ suggested above (from Figure 1(b)). The standard errors of the estimates based on the sample information are obtained from (10). The MLE of the parameters as well as their standard errors are given in Table 2.

Table 3: Appliance Failure Data: Estimates of the parameters and their standard errors when the outlier, $x_{60:60}$, is omitted.

Method	$\hat{\alpha}$	$s.e.(\hat{\alpha})$	$\hat{\beta}$	$s.e.(\hat{\beta})$	$\hat{\lambda}$	$s.e.(\hat{\lambda})$
MLE	4.0427	2.875	4626.39	913.114	0.1720	0.141
MOM	3.5876	0.920	4560.53	513.896	0.2060	0.081
Bayes	4.8765	2.7515	4452.06	896.32	0.2403	0.1234
EPM (MED)	3.6311	0.3828	5021.42	558.59	0.1792	0.0173
EPM (TM)	3.6293	0.3438	5062.65	570.81	0.1830	0.0203

The Monte Carlo approach together with the algorithm described in Section 3 is used to derive the MOM estimates for α, β and λ. Although the algorithm converges fast close to the solution of $f_r(\theta) = 0$, for $r = 1, 2, 3$, there are problems with the approach since even with very small ε of 0.00001 and with 1000 generated Uniform (0, 1) values, the evaluated values of the integrals necessary for evaluating the elements of $f\left(\theta^{(k)}\right)$ and $J\left(\theta^{(k)}\right)$ vary for different sets of Uniform (0, 1) values leading to different values of the estimates of α, β and λ. The resulting MOM estimates for 200 different applications of the algorithm described in Section 3, each application involves 1000 Uniform (0, 1) values are calculated. Table 2 gives the median of these estimates for α, β and λ. Simulated samples are obtained from the parametric cumulative distribution function $F(x; \hat{\theta})$ and the bootstrap method is used to compute the standard errors of the estimates, which are shown in Table 2. Two hundred samples of size 60 are drawn from the EW distribution with $\alpha = 1.669$, $\beta = 3600.539$ and $\lambda = 0.4955$. We then compute the MOM estimates for α, β and λ for each sample. The bootstrap estimate of the variance is the variance of the estimates of these 200 samples. The results are given in Table 2.

Applying the Bayesian approach described in Section 4, the Bayes estimates for α, β and λ in (16) and their standard errors from (18), under the diffuse prior, are given in Table 2.

The results of Section 5 are used to derive the EPM estimates of α, β and λ. The

EW distribution has three parameters which means for the appliance failure data there are ${}^{60}C_3 = 34,220$ elemental subsets each containing three observations. One thousand elemental subsets were randomly selected and Newton's method was used to solve equation (25) for λ for each of these subsets. The corresponding estimates for α and β are derived from (26) and (27), respectively. Table 2 shows the final estimates of α, β and λ are

$$\hat{\theta}_i\,(\text{MED}) = \text{Median}(\hat{\theta}_{i,1},\, \hat{\theta}_{i,2},\, ...,\, \hat{\theta}_{i,1000}) \quad \text{and} \quad \hat{\theta}_i\,(\text{TM}) = \text{TM}(\hat{\theta}_{i,1},\, \hat{\theta}_{i,2},\, ...,\, \hat{\theta}_{i,1000}),$$

for $i = 1, 2, 3$. These estimates and their bootstrap standard errors (s.e.) based on 200 samples of size 60 drawn from the EW distribution with the corresponding values of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\lambda}$ are given in Table 2. We compute the EPM estimates for α, β and λ for each sample. The bootstrap estimate of the variance is the variance of the estimates of the these 200 samples.

All parameter estimates in Table 2 show that $\hat{\alpha} > 1$ and $\hat{\alpha}\hat{\lambda} < 1$, which again indicate a bathtub failure rate. The estimates are consistent with the TTT-transform in Figure 1. It can be seen from Table 2 that the EPM methods (MED and TM) have the smallest standard errors.

To assess the goodness of fit of the five methods of estimation in Table 2 as well as the initial least squares estimates obtained from (29), we measure the amount by which the actual observations x vary around the estimated quantiles $\hat{x}_p = F^{-1}(p_{i:n};\, \hat{\theta})$. For the purpose of comparison, we use the mean absolute prediction error:

$$\text{MAPE}(\hat{\theta}) = n^{-1} \sum_{i=1}^{n} \left| x_{i:n} - F^{-1}\left(p_{i:n};\, \hat{\theta}\right) \right|. \tag{40}$$

The MAPE for all six methods are shown in the second column of Table 4, using all $n = 60$ observations. Both EPM estimates (MED and TM) provide the smallest MAPE, while the Bayes method provides the largest MAPE. It is interesting to see that even the initial least squares (LS) estimates that are used as starting values for the MLE method perform even better than MLE, MOM, and Bayes.

Table 4: Appliance Failure Data: MAPE for six estimation methods.

Method	$n = 60$	$n = 59$
LS	148	82
MLE	155	147
MOM	163	159
Bayes	199	243
EPM (MED)	138	125
EPM (TM)	135	138

All methods indicate that the largest observation, x_{60}, is a clear outlier. This can be seen in all six plots of $\hat{x}_{i:n}$ versus $x_{i:n}$ (two such graphs are shown in Figure 3(a,b), as an example). We computed the MAPE when this observation is omitted. The resulting values of MAPE are shown in the last column of Table 4. Again, both EPM estimates perform better than the MLE, MOM, and Bayes estimates. It is intriguing to see that the Bayes estimates produce a much larger MAPE after omitting the outlier. It is also interesting to note that the LS initial estimates perform rather very well after omitting the outlier. The two graphs in Figure 3(a,b) are reproduced in Figure 3(c,d) after the outlier has been removed.

The values of the likelihood ratio statistics for testing the exponentiality and Weibull goodness of fit hypotheses are determined as described in Section 7. For testing exponentiality versus EW model, under the null hypothesis (where $\lambda = \alpha = 1$) for the complete data, the maximum likelihood estimate for β is $\hat{\beta}_e = 2193.03$. The corresponding likelihood ratio statistic is $-2\Lambda_1 = 3.915$, which is to be compared with the Chi-squared value with 1 degree of freedom.

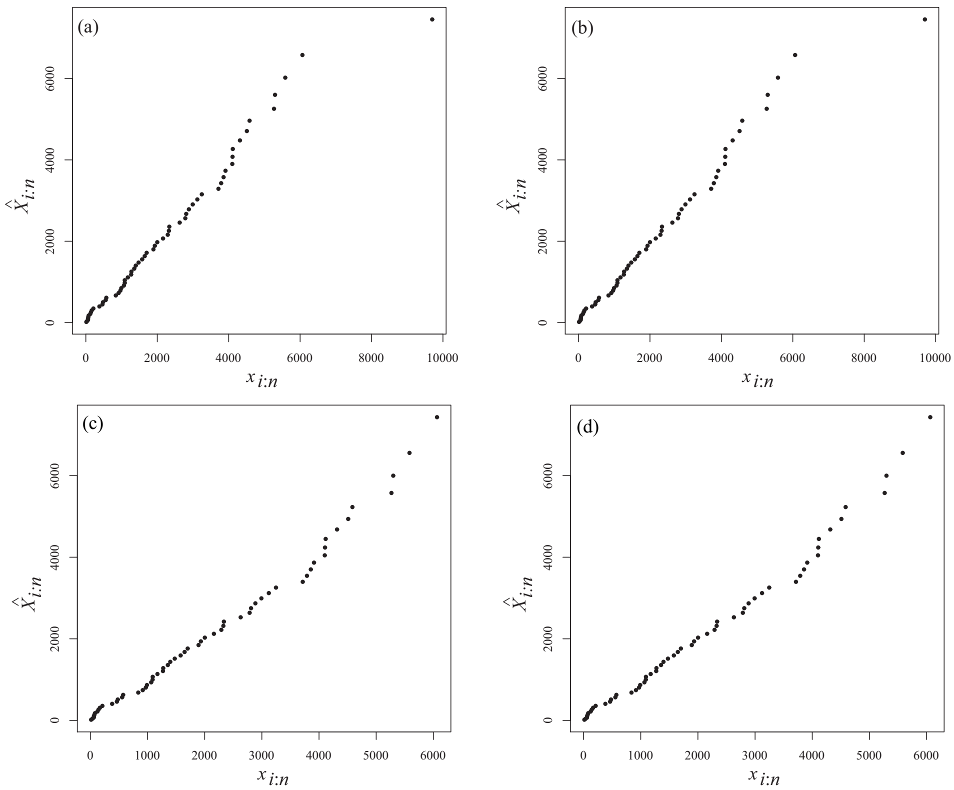


Figure 3: Appliance Failure Data: Plot of $\hat{x}_{i:n}$ versus $x_{i:n}$, for (a) MLE, (b) MED, (c) MLE with the outlier deleted, and (d) MED with the outlier deleted.

To test the Weibull versus EW model (where $\lambda = 1$), we find the MLE for α and β under the Weibull model by fitting a simple linear regression model relating $\log x_p$ to $\log[-\log(1-p)]$ to provide starting values for the iterative procedure. The slope and intercept of the fitted line are 1.2054 and 7.7330, respectively, with $R^2 = 96.79\%$. The associated estimates of α and β are 0.8296 and 2282.38, respectively. Using Newton Raphson's method the MLE are $\hat{\alpha}_w = 1.001$ and $\hat{\beta}_w = 2193.734$. The corresponding likelihood ratio statistic is $-2\Lambda_2 = 3.915$, which is to be compared with the Chi-squared value with 2 degrees of freedom. Accordingly, the exponential model is untenable at the 5% significance level while the Weibull model is untenable at the 20%.

The corresponding estimate of β_e under the exponential model but after deleting the outlier is $\hat{\beta}_e = 2065.78$. The corresponding likelihood ratio statistic is $-2\Lambda_1 = 9.77$, which is to be compared with the Chi-squared value with 1 degree of freedom.

Under the Weibull model the linear regression relating $\log x_p$ to $\log[-\log(1-p)]$ with the outlier omitted gives slope = 1.1914, intercept = 7.6890, and $R^2 = 96.29\%$. The associated estimates of α and β are 0.8394 and 2184.26, respectively. Using Newton Raphson's method the MLE are $\hat{\alpha}_w = 1.040$ and $\hat{\beta}_w = 2093.974$. The corresponding likelihood ratio statistic is $-2\Lambda_2 = 9.65$, which is to be compared with the Chi-squared value with 2 degrees of freedom. Hence, both the exponential and Weibull models are untenable at the 1% significance level.

8.2 **Example 2**

The data in Table 5 are remission times, in weeks, for leukemia patients given two types of treatments, quoted in Lawless (2003). In this study 20 patients were given treatment A and 20 treatment B. Starred observations are censoring times.

Table 5: Leukemia Data: Remission times, in weeks, for leukemia patients given two types of treatments.

Treatment A				Treatment B			
1	3	3	6	1	1	2	2
7	7	10	12	3	4	5	8
14	15	18	19	8	9	11	12
22	26	28*	29	14	16	18	21
34	40	48*	49*	27*	31	38*	44

The product limit estimates of the survivor functions of the two treatments are given in Table 6. To check the adequacy of the EW model, we construct the graph in (30) for the data in each of the two treatments for different values of λ. For Treatment A and with $\lambda = 0.51$, the scatter plot shown in Figure 4(a) exhibits a strong linear relationship ($R^2 = 99.32\%$) with slope 0.592 and intercept 3.451. Therefore, the EW assumption seems reasonable. The associated estimates of α and β are $\hat{\alpha} = 1.6904$ and $\hat{\beta} = 31.53$. Similarly, for Treatment B and $\lambda = 0.85$, the scatter plot shown in Figure 4(b) exhibits a strong linear relationship ($R^2 = 99.57\%$) with slope 1.0433 and intercept 2.7223. Therefore, the EW assumption seems reasonable. The associated estimates of α and β are $\hat{\alpha} = 0.9585$ and $\hat{\beta} = 15.216$.

Since all patients were not started on test on the same date, the extension of the EPM described in Section 6 is the only viable method here. We use it to fit the EW model for the data sets of Treatments A and B.

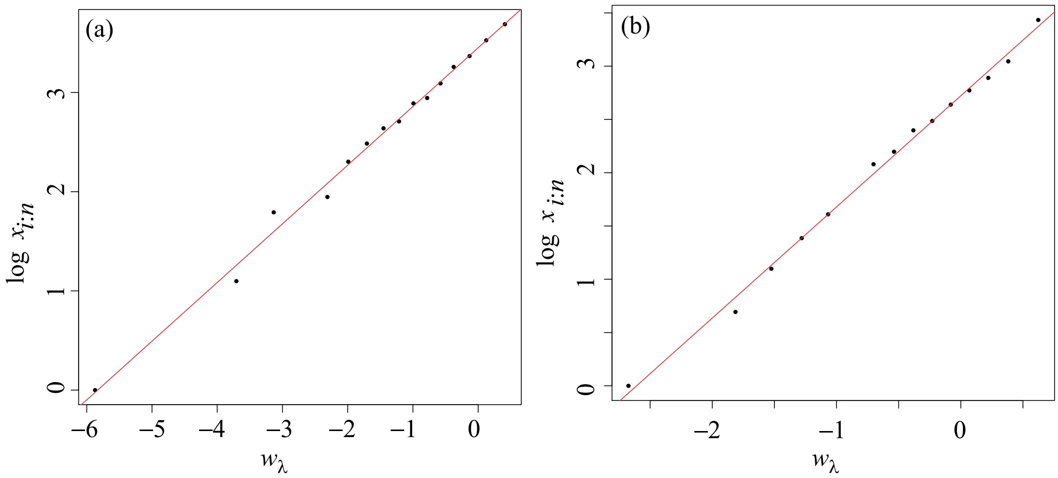


Figure 4: Leukemia Data: The plot in $\log x_{i:n}$ versus w_λ in (30) for (a) Treatment A and (b) Treatment B.

One thousand elemental subsets were randomly selected and the corresponding initial estimates for α, β and λ are derived. The 1,000 initial estimates of α, β and λ are computed and the MED and TM of α, β and λ for Treatment A are:

$$\begin{array}{llll} \text{MED} & \hat{\alpha} = 1.7063 & \hat{\beta} = 31.89 & \hat{\lambda} = 0.5074 \\ \text{TM} & \hat{\alpha} = 1.7055 & \hat{\beta} = 31.56 & \hat{\lambda} = 0.5141 \end{array}$$

The corresponding MED and TM of α, β and λ for Treatment B are:

$$\begin{array}{llll} \text{MED} & \hat{\alpha} = 1.0238 & \hat{\beta} = 16.46 & \hat{\lambda} = 0.7697 \\ \text{TM} & \hat{\alpha} = 1.0446 & \hat{\beta} = 16.30 & \hat{\lambda} = 0.7924 \end{array}$$

Using the MED estimates of the parameters, the EPM estimates of the 25th, 50th and 75th percentiles for Treatment A are $\hat{x}_{0.25} = 6.5577$, $\hat{x}_{0.5} = 15.5778$, and $\hat{x}_{0.75} = 28.7432$ weeks, respectively. The corresponding estimates for Treatment B are $\hat{x}_{0.25} = 3.0915$, $\hat{x}_{0.5} = 8.715$, and $\hat{x}_{0.75} = 19.1118$ weeks, respectively.

9. SUMMARY

In this paper, we present methods for estimating the three parameters and quantiles of the EW distribution for complete data and under two types of censoring. The presence of censoring creates special problems for the classical estimation methods.

First, for complete and Type II censoring, classical estimators, such as the maximum likelihood estimates and Bayes estimates have been previously developed but no theoretical

Table 6: Leukemia Data: Product-limit estimate of the survivor function $\hat{s}_{x_j}$.

Treatment A				Treatment B			
x_j	d_j	n_j	$\hat{s}_{x_j}$	x_j	d_j	n_j	$\hat{s}_{x_j}$
1	1	20	0.95	1	2	20	0.90
3	2	19	0.85	2	2	18	0.80
6	1	17	0.80	3	1	16	0.75
7	2	16	0.70	4	1	15	0.70
10	1	14	0.65	5	1	14	0.65
12	1	13	0.60	8	2	13	0.55
14	1	12	0.55	9	1	11	0.50
15	1	11	0.50	11	1	10	0.45
18	1	10	0.45	12	1	9	0.40
19	1	9	0.40	14	1	8	0.35
22	1	8	0.35	16	1	7	0.30
26	1	7	0.30	18	1	6	0.25
29	1	5	0.24	21	1	5	0.20
34	1	4	0.18	31	1	3	0.13
40	1	3	0.12	44	1	1	0.00

asymptotic variances have been derived. These asymptotic variances are derived in Section 2. Second, we develop the MOM estimators for the case of complete data in Section 3. Fourth, for Bayes estimators, we use Lindley's approximation, which leads to Bayes estimators and their posterior risk in closed-forms. Fifth, we also develop the EPM estimates for the case of complete data in Section 5, which are easy to derive and to compute. Sixth, we generalize the EPM approach to be applicable to all forms of censoring. This is especially useful because fitting the EW model under Type I censoring has not been considered in the literature. Seventh, graphical displays for informal checks on the appropriateness of the model as well as formal goodness of fit tests for the three- parameter EW are proposed and illustrated by examples. Finally, the proposed methods have been successfully applied to estimate the parameters and quantiles of the EW model using two real-life complete and censored data.

REFERENCES

Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Trans. Reliability*, 36:106–108.

Bain, L. J. (1974). Analysis for the linear failure-rate life-testing distribution. *Technometrics*, 16:551–559.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions.* John Wiley & Sons, New York.

Barlow, R. E. and Campo, R. (1975). Total time on test processes and applications to failure data analysis. In Barlow, R. E., Fussel, J., and Singpurwalla, N. D., editors, *Reliability and Fault Tree Analysis*, pages 451–481. SIAM, Philadelphia.

Castillo, E. (1988). *Extreme Value Theory in Engineering.* Academic Press, New York.

Castillo, E. and Hadi, A. S. (1995a). A method for estimating parameters and quantiles of distributions of continuous random variables. *Computational Statistics and Data Analysis*, 20:421–439.

Castillo, E. and Hadi, A. S. (1995b). Modeling lifetime data with applications to fatigue models. *Journal of the American Statistical Association*, 90:1041–1054.

Castillo, E. and Hadi, A. S. (1997). Fitting the generalized pareto distribution to data. *Journal of the American Statistical Association*, 92:1609–1619.

Castillo, E., Hadi, A. S., Balakrishnan, N., and Sarabia, J. M. (2005). *Extreme Value and Related Models in Engineering and Science Applications.* John Wiley & Sons, New York.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7:1–26.

Efron, B. (1988). Logistic regression, survival analysis, and the kaplan-maier curve. *Journal of the American Statistical Association*, 83:414–425.

Gaver, D. P. and Acar, M. (1979). Analytical hazard representations for use in reliability, mortality, and simulation studies. *Communications in Statistics: Simulation and Computation*, 8:91–111.

Gore, A. P., Paranjape, S., Rajarshi, M. B., and Gadgil, M. (1986). Some methods for summarizing survivorship data. *Communications in Statistics: Simulation and Computation*, 28:557–586.

Hjorth, J. (1980). A reliability distribution with increasing, decreasing, constant, and bathtub-shaped failure rates. *Technometrics*, 22:99–108.

Jeffreys, H. (1961). *Theory of Probability (3rd Edition.* University Press, Oxford.

Jiang, R. and Murthy, D. N. P. (1999). The exponentiated weibull family: A graphical approach. *Technometrics*, 48:68–72.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data (2nd Edition).* John Wiley & Sons, New York.

Lindley, D. V. (1980). Approximate bayesian methods. In Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M., editors, *Bayesian Statistics*, pages 223–245. Valencia Press, Valencia, Spain.

Mann, N. R., Schafer, E., R., and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. John Wiley & Sons, New York.

Mudholkar, G. S. and Hutson, A. D. (1996). The exponentiated weibull family: Some properties and a flood data application. *Communications in Statistics: Theory and Methods*, 25:3059–3083.

Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated weibull family for analyzing bathtub failure-rate data. *EEE Trans. Reliability*, 42:299–302.

Mudholkar, G. S., Srivastava, D. K., and Freimer, M. (1995). The exponentiated weibull family: A reanalysis of the bus motor failure data. *Technometrics*, 37:436–445.

Nassar, M. M. and Eissa, F. H. (2003). On the exponentiated weibull distribution. *Communications in Statistics: Theory and Methods*, 32:1317–1336.

Prentice, R. L. (1975). Discrimination among some parametric models. *Biometrika*, 62:607–614.

Press, S. J. (2003). *Subjective and Objective Bayesian Statistics: Principles, Models and Applications*. John Wiley & Sons, New York.

Singh, U., Gupta, P. K., Chaturvadi, M., and Upadhyay, S. K. (1999). Bayesian estimation of exponentiated weibull distribution using squared error loss function. In *Proceeding of National Seminar on Bayesian Analysis Theory and Application*, pages 130–144.

Singh, U., Gupta, P. K., and Upadhyay, S. K. (2002). Estimation of exponentiated weibull shape parameters under linex loss function. *Communications in Statistics: Simulation and Computation*, 31:523–537.

Singh, U., Gupta, P. K., and Upadhyay, S. K. (2005a). Estimation of parameters for exponentiated weibull family under type ii censoring scheme. *Computational Statistics and Data Analysis*, 48:509–523.

Singh, U., Gupta, P. K., and Upadhyay, S. K. (2005b). Estimation of three-parameter exponentiated weibull distribution under type ii censoring. *Journal of Statistical Planning and Inference*, 134:350–372.

Slymen, D. J. and Lachenbruch, P. A. (1984). The exponentiated weibull family: Some properties and a flood data application. *Communications in Statistics: Theory and Methods*, 13:1179–1201.

Stacy, E. W. (1962). A generalization of the gamma distribution. *Annals of Mathematical Statistics*, 33:1187–1192.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 451–451.

REPAIRABLE SYSTEM MODEL WITH TIME DEPENDENT COVARIATE

[1]Jayanthi Arasan, [2]Samira Ehsani and [3]Kaveh Kiani
[1, 2]Department of Mathematics, Faculty of Science, Universiti Putra Malaysia
E-mail: [1, 2]jayanthi@math.upm.edu.my
[3]Applied & Computational Statistics Laboratory, Institute for Mathematical Research, Universiti Putra Malaysia
E-mail: [3]kamakish@yahoo.com

ABSTRACT

In this paper we extend a repairable system model that incorporates both time trend and renewal-type behavior to include a time dependent covariate. We calculated the bias, standard error and rmse of the parameter estimates of this model at different sample sizes using simulated data. Following that, we studied several alternative computer intensive methods of constructing confidence interval estimates for the parameters of the general model. Alternative methods relieve us from making assumptions and having to depend solely on the traditional methods derived from asymptotic statistical theory. In addition, the high capability of modern day computers makes these methods easily applicable and practical. Several parametric bootstrap methods and jackknife confidence interval procedures were compared to the Wald interval via coverage probability study. The results clearly show that the B-t and jackknife techniques work much better than other methods when sample sizes are moderate and low. The Wald intervals was found to be highly asymmetrical and only starts to work when sample sizes are rather large.

Keywords: Bootstrap; jackknife; repairable

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 452–460.

A COMPARISON OF THE PERFORMANCES OF VARIOUS SINGLE VARIABLE CHARTS

Abdu M. A. Atta[1], Michael B. C. Khoo[2] and S. K. Lim
School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia
E-mail: [1]abduatta@yahoo.com and [2]mkbc@usm.my

ABSTRACT

Control charts are used for process monitoring and improvement in industries. Two charts are usually used in the monitoring of both the mean and variance separately. In the past 20 years, numerous control charting approaches that enable a joint monitoring of both the mean and variance on a single chart have been suggested. A joint monitoring of both the mean and variance is more meaningful in a real situation as both the mean and variance may shift simultaneously. Although numerous single variable control charts are available in the literature, not much research is made to compare these charts, in terms of their detection power. This paper compares the performances of several single variable charts, such as the semicircle, MaxEWMA and single MA charts, in terms of their average run length (ARL) results, via a Monte Carlo simulation. The omnibus EWMA and CUSUM charts suggested by Domangue and Patch (1991) and discussed again by Woodall and Mahmoud (2005) are not considered here because unlike the semicircle, MaxEWMA and single MA charts, these charts do not indicate the source of an out-of-control signal as to whether it is caused by a shift in the mean, variance or both, when one occurs. The Statistical Analysis System (SAS) software is employed in the simulation study. This comparison serves as a guide to practitioners by helping them to select a suitable single variable chart for process monitoring.

1. INTRODUCTION

Control charts are used for the purpose of detecting assignable causes that affect process stability. Two control charts, one for monitoring the process mean, such as the $\bar{X}$ chart and the other for monitoring the process variance, such as R chart or S chart, are usually run simultaneously. Most charts for variable data found in the literature monitor the process mean and variance separately. As shown by Reynolds and Stoumbos (2004) and mentioned again by Costa and Rahim (2006), running two charts, one for the mean and the other for the variance, may not always be reliable in identifying the nature of the change. Recently, control charts that can simultaneously monitor both the process mean and the procecss variance have been proposed. These charts are called single variable control charts, and are classified as the Shewhart-type charts, CUSUM-type charts and EWMA-type charts.

For the Shewhart-type single variable chart, White and Schroeder (1987) first introduced the use of one control chart to monitor both process mean and variance on the same chart. This chart was designed using resistant measure and a modified box plot display. Chao and Cheng (1996) proposed a single control chart, called the semicircle (SC) control chart. This chart uses a semicircle to plot a single plotting statistic to indicate the position of the mean and standard deviation, by plotting the standard deviation on the *y*-axis and the mean on the *x*-axis. When a point plots ouside of the semicircle indicating an out-of-control signal, the chart shows whether the mean, the variance or both parameters have shifted. The disadvantage of this chart is that it loses track of the time sequence of the plotted points. Chen and Cheng (1998) proposed a single Shewhart-type control chart, called the Max chart. This Max chart plots the maximum absolute value of the standardized mean and standard deviation. This chart performs like the combined Shewhart charts for the mean and standard deviation, i.e., the combined $\bar{X} - S$ charts. Spiring and Cheng (1998) developed a single variable chart that monitors both the process mean and standard deviation. This chart also plots two variables at the same time and has the advantage of performing equally well for both large and small subgroup sizes. Gan et al. (2004) proposed a single control chart based on the interval approach that combines both $\bar{X}$ and S charts into one scheme. Wu and Tian (2006) suggested a single weighted loss function chart (WL chart) for a simultaneous monitoring of the process mean and variance. İt was shown that the WL chart is significantly more effective than the unadjusted loss function chart and joint $\bar{X} - S$ charts, as well as the other charts.

For the CUSUM-type single chart, the CUSUM M-chart and CUSUM V-chart for detecting small shifts in the process mean and process variance, respectively, were proposed by Yeh et al. (2004). Because these charts have the same distribution when the process is in-control, they can be effectively combined into a single chart, thus, enabling a simultaneous monitoring of the mean and variance to be made on the same chart. A weighted loss function CUSUM (WLC) scheme with variable sampling interval (VSI) that enables a simultaneous monitoring of both the mean shift and an increasing variance shift by using a CUSUM chart was suggested by Zhang and Wu (2006).

For the EWMA-type single chart, numerous single EWMA charts for a simultaneous monitoring of the process mean and variance have been proposed. Domangue and Patch (1991) suggested some omnibus EWMA schemes based on the exponentiation of the absolute value of the standardized sample mean of the observations for a joint monitoring of the mean and variance. Gan (2000) proposed a simultaneous EWMA chart that was developed by combining a chart for the mean and a chart for the variance into one chart by plotting the EWMA of log(S^2) against the EWMA of $\bar{X}$. The control limit of this chart is formed by either using a rectangle or an ellipse. Morais and Pacheco (2000) considered a joint monitoring of the process mean and variance using a combined EWMA (CEWMA) scheme, where the average run length, percentage points of the run length and probability of a misleading signal were investigated. By using a two dimensional Markov chain approximation, these three performance measures are obtained. The MaxEWMA chart which combines the EWMA charts for the process mean and process variance into a single chart was developed by Chen et al. (2001). This chart extends and improves upon the earlier work of Chen and Cheng (1998) on the Max chart. Chen et al. (2004) proposed the EWMA-SC chart by applying the EWMA technique to the statistics employed in the semicircle chart. This proposed chart provides a better detection ability with regards to small shifts in the mean and/or variance in comparison to the SC chart. Costa and Rahim (2004) suggested the use of a single non-central chi-square chart to monitor both the process mean and variance simultaneously. Costa and Rahim (2004) also found that the EWMA chart based on the non-central chi-square statistic has a similar performance to the MaxEWMA chart proposed by Chen et al. (2001). A single EWMA chart which is an extension of the EWMA-SC chart studied by Chen et al. (2004) was suggested by Costa and Rahim (2006).

This paper compares the performances of three single control charts, namely the semicircle (SC), MaxEWMA and single moving average (MA) charts, in terms of thier average run lengths (ARLs), via a Monte Carlo simulation. A simulation study conducted using the Statistical Analysis System (SAS) software shows that the MaxEWMA chart gives the best performance, while the SC chart has the poorest performance. This comparison assist practitioners in selecting a suitable single variable chart for process monitoring.

This paper is organized as follows: Section 2 reviews the semicircle (SC) chart. A review of the MaxEWMA chart is made in Section 3, while Section 4 reviews the single moving average (MA) chart. In Section 5, a performance comparison is made to compare the performances of the SC, MaxEWMA and single MA charts, in terms of their average run length (ARL) profiles. Finally, conclusions are drawn in Section 6.

2. SEMICIRCLE (SC) CHART

Chao and Cheng (1996) proposed a semicircle control chart, where a semicircle is used to plot a single plotting statistic to represent the position of the mean and standard deviation, where the standard deviation is plotted on the *y*-axis and the mean on the *x*- axis. When a point plots outside of the simicircle indicating an out-of-control signal, the chart can easily tell whether the mean, the variance or both the parameters have changed. With its straight forward calculations, this chart can be considerd as a new alternative to the combination of the $\bar{X}$ and *R* charts.

The plotting statistic proposed by Chao and Cheng (1996) is

$$T = \left(\bar{X} - \mu\right)^2 + S^{*2}. \tag{1}$$

The plotting statistic is based on the sample mean, $\bar{X}$, and the root mean square, $S^* = \left(\frac{n-1}{n}\right)^{\frac{1}{2}} S$, where $S = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}}$ is the sample standard deviation. Under the normality assumption, $\left(\frac{n}{\sigma^2}\right)T$ is distributed exactly as a χ_n^2 random variable for any sample size *n*. The statistic T in Equation (1) defines a circle, but because $S^* \geq 0,$ when plotting $\left(\bar{X}, S_i^*\right)$ for each sample on the $\left(\bar{X}, S^*\right)$ plane, a semicircle will be sufficient.

İn order to construct the semicircle chart, the following formula can be used in determining the radius *r*, where α is the size of the Type-I error (Chao and Cheng, 1996):

$$\begin{aligned} P\left(T < r^2\right) &= P\left(\frac{n}{\sigma^2}T < \frac{n}{\sigma^2}r^2\right) \\ &= 1 - \alpha, \end{aligned}$$

Since $\frac{n}{\sigma^2}T \sim \chi_n^2,$ we have $\frac{n}{\sigma^2}r^2 \sim \chi^2_{n,\,(1-\alpha)},$ i.e.,

$$r = \left(\frac{\chi^2_{n,(1-\alpha)}}{n}\right)^{1/2}\sigma. \tag{2}$$

Here, $\chi^2_{n,(1-\alpha)}$ is a 100(1–α)% percentile of the χ^2_n distribution. Note that if parameters are unknown, $\bar{\bar{X}}$ is used to estimate μ and $\bar{S}^*$ to estimate σ. Then we have (Chao and Cheng, 1996)

$$\hat{r} = q\bar{S}^* \tag{3}$$

where $q = \left(\frac{\chi^2_{n,(1-\alpha)}}{n}\right)^{1/2}$.

3. MaxEWMA CHART

An exponentially weighted moving average (EWMA) chart is a control chart for variable data. İt plots weighted moving averages. A weighting factor is chosen by the user to determine how older data points affect the mean value compared to more recent ones. Because the EWMA chart uses information from all samples, it detects smaller process shifts quicker than the Shewhart control chart. The MaxEWMA chart is constructed as follows (Chen et al., 2001):

Assume that a sequence of individual measurements, X_{ij}, in sample *i*, follow a *N*(μ,σ) distribution, for *i* = 1, 2, ... and *j* = 1, 2, ..., n_i. Let μ_0 be the nominal process mean and σ_0 be a known value of the process standard deviation. Assume that the process parameters μ and σ can be expressed as $\mu = \mu_0 + a\sigma_0$ and $\sigma = b\sigma_0$, where *a* and *b* (>0) are constants. The process is in-control when *a* = 0 and *b* = 1; otherwise the process has changed.

Let $\bar{X}_i = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}$ be the i^{th} sample mean and $S_i^2 = \frac{\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2}{(n_i - 1)}$ be the i^{th} sample variance. Define (Chen et al., 2001)

$$U_i = \frac{\left(\bar{X}_i - \mu_0\right)}{\sigma_0/\sqrt{n_i}} \tag{4}$$

and

$$V_i = \Phi^{-1}\left\{H\left(\frac{(n_i-1)S_i^2}{\sigma_0^2}; n_i - 1\right)\right\}, \tag{5}$$

where $\Phi(z) = P(Z \le z)$, for *Z* ~ *N*(0, 1), $\Phi^{-1}(\cdot)$ is the inverse function of $\Phi(\cdot)$ and $H(\omega;v) = P(W \le \omega)$, where *W* follows a chi-square distribution with *v* degrees of freedom.

It is known that U_i and V_i are independent when *a* = 0 and *b* = 1 because $\bar{X}_i$ and S_i^2 are independent. İt can be shown that $U_i \sim N(0,1)$ and $V_i \sim N(0,1)$ (Chen et al., 2001). The distributions of U_i and V_i are both independent of the sample size n_i, when *a* = 0 and *b* = 1, therefore the variable sample size problem can be handled easily by the MaxEWMA chart. Since both U_i and V_i have the same distribution, a single variable chart to monitor both the process mean and process variability can be constructed (Chen et al., 2001). First, define

$$Y_i = \lambda U_i + (1-\lambda)Y_{i-1}, \quad 0 < \lambda \le 1, \quad \text{for } i = 1, 2, \ldots \tag{6a}$$

and

$$Z_i = \lambda V_i + (1-\lambda)Z_{i-1}, \quad 0 < \lambda \le 1, \quad \text{for } i = 1, 2, \ldots, \tag{6b}$$

with Y_0 and Z_0 as the starting values, respectively. Then, the above two EWMA statistics are combined into a single chart by defining a new statistic M_i given by

$$M_i = \max\{|Y_i|, |Z_i|\}. \tag{7}$$

The statistic M_i will be large when the process mean has shifted away from μ and/or when the process variability has increased or decreased. On the other hand, the statistic M_i will be small when the process mean and process variabiity stay close to their respective targets.

Since M_i is non-negative, only a UCL is needed. The UCL is given by

$$\text{UCL} = E(M_i) + K\sqrt{Var(M_i)}, \tag{8}$$

where $E(M_i)$ is the mean of M_i and $Var(M_i)$ is the variance of M_i, when a = 0 and b = 1. Here, K is a multiplier, which together with λ, controls the performance of the new chart. Because this chart is based on M_i, the maximum of $|Y_i|$ and $|Z_i|$, it is called the MaxEWMA chart.

4. SINGLE MOVING AVERAGE (MA) CHART

Khoo and Yap (2005) suggested the use of a joint moving average control chart for a simultaneous monitoring of the process mean and variance. Besides being efficient in detecting increases and decreases in the process mean and/or variability, the joint MA chart is also able to indicate the source and direction of a shift.

Let X_{ij} for i =1, 2, ..., and j = 1, 2, ..., n_i, be observations from subgroups of size n_i, with i representing the subgroup number. It is assumed that $X_{ij} \sim N(\mu + a\sigma, b^2\sigma^2)$, where a = 0 and b = 1 indicate that the process is in-control; otherwise, the process has shifted. Let $\bar{X}_i = \frac{(X_{i1} + X_{i2} + \cdots + X_{in_i})}{n_i}$ be the i^{th} sample mean and let $S_i^2 = \frac{\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2}{n_i - 1}$ be the i^{th} sample variance. Khoo and Yap (2005) define the following statistics:

$$U_i = \frac{(\bar{X}_i - \mu)}{\sigma/\sqrt{n_i}} \sim N(0,1), \qquad \text{for } i = 1, 2, \ldots \tag{9}$$

and

$$V_i = \Phi^{-1}\left\{H\left(\frac{(n_i - 1)S_i^2}{\sigma^2}; n_i - 1\right)\right\} \sim N(0,1), \quad \text{for } i = 1, 2, \ldots, \tag{10}$$

where $\Phi^{-1}(\cdot)$ and $H(\cdot)$ denote the inverse standard normal distribution function and the chi-square distribution function with $n_i - 1$ degrees of freedom, respectively. Because the sample mean, $\bar{X}_i$ and sample variance, S_i^2, are independent, U_i and V_i are also independent. The sample grand average, $\bar{\bar{X}}$ and $\bar{R}/d_2$ or $\bar{S}/c_4$ are substituted for μ and σ, respectively, if the target values of these parameters are unknown, where d_2 and c_4 are the control chart constants.

The plotting statistic, K_i, of the chart can then be defined as (Khoo and Yap, 2005)

$$K_i = \max\{|L_i|, |M_i|\}, \tag{11}$$

where

$$L_i = \frac{U_i + U_{i-1} + \cdots + U_{i-w+1}}{w} \quad \text{and} \quad M_i = \frac{V_i + V_{i-1} + \cdots + V_{i-w+1}}{w},$$

for $i \geq w$, while for $i < w$, say $w = 3$,

$$L_1 = \frac{U_1}{1},\ L_2 = \frac{U_1 + U_2}{2},\ L_3 = \frac{U_1 + U_2 + U_3}{3}$$

and

$$M_1 = \frac{V_1}{1}, M_2 = \frac{V_1 + V_2}{2}, M_3 = \frac{V_1 + V_2 + V_3}{3}.$$

Note that w is the span of the moving average statistic. The statistic K_i will be large when the process mean has shifted away from its target value and/or when the process variance has increased or decreased. Only the upper control limit, UCL is applied on the joint MA chart as K_i is non-negative. The density function of K_i, for the in-control case is (Khoo and Yap, 2005)

$$f(k) = 4\sqrt{w}\phi\left(k\sqrt{w}\right)\left\{2\Phi\left(k\sqrt{w}\right) - 1\right\}, \quad \text{for } k \geq 0. \tag{12}$$

Here, $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and distribution functions of a standard normal random variable, respectively. Suppose that the desired Type-I error set by management based on some predetemined factors is α, then UCL is obtained by solving the following definite integral:

$$\int_{UCL}^{\infty} f(k)\,dk = \alpha \tag{13}$$

For a MA control chart, the control limits for periods $i < w$ are wider than their steady-state value. Besides having the desirable properties of the moving average chart, the variable sample size problem can also be handled automatically with the application of the joint MA chart. Therefore, the joint MA chart can be considered as an attractive alternative to the combined $\bar{X} - R$ or $\bar{X} - S$ charts (Khoo and Yap, 2005).

5. PERFORMANCE COMPARISON

A simulation study is conducted using SAS version 9 to study the performances of the SC, MaxEWMA and single MA charts. The sample size of $n = 5$ is considered. The shifts in the mean and variance considered are $\mu_1 = \mu_0 + a\sigma_0$ and $\sigma_1 = b\sigma_0$, respectively, where $a \in \{0, 0.25, 0.5, 1, 2\}$ and $b \in \{0.25, 0.5, 1, 1.5, 2\}$. Note that when the process is in-control, $a = 0$ and $b = 1$. The in-control ARL (ARL_0) is fixed as 185. For the SC chart, the radius of the chart for $n = 5$ is $r = 1.8174$. For the MaxEWMA chart, $\lambda \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.8, 1\}$ are considered and their corresponding K values are determined. The values of the moving span $w \in \{2, 3, 4, 5\}$ are employed and their corresponding UCLs are determined for the single MA chart.

The ARL profiles for the SC, single MA and MaxEWMA charts are given in Tables 1, 2 and 3, respectively. Generally, the results show that the MaxEWMA chart is superior to the other two charts. The SC chart is found to have the poorest performance. Note that for an arbitrary combination of (a, b), where $a > 0$ and $b \neq 1$, the MaxEWMA chart has the lowest out-of-control ARL, followed by the single MA chart. On the contrary, the SC chart always has the highest out-of-control ARL among the three charts.

6. CONCLUSIONS

İn this paper, the performances of three single variable control charts are compared based on their ARLs. Overall, the MaxEWMA chart provides a better performance than the other two charts, while the SC chart gives the poorest performance.

ACKNOWLEDGEMENT

This research is supported by the Universiti Sains Malaysia, Research University (RU) grant, no. 1001/PMATHS/811024.

APPENDIX

Table 1. ARL profiles for the SC chart with ARL_0 =185 and n =5

	a					
	b	0.00	0.25	0.50	1.00	2.00
r = 1.8174	0.25	1.3	1.5	1.0	1.0	1.0
	0.50	8.7	1.1	1.0	1.0	1.0
	1.00	185.0	123.7	52.0	8.3	1.2
	1.50	5.0	4.7	4.0	2.4	1.2
	2.00	1.9	1.8	1.7	1.5	1.1

Table 2. ARL profiles for the single MA chart with ARL_0 =185 and n =5

	a					
(w, UCL)	*b*	0.00	0.25	0.50	1.00	2.00
w = 2 UCL= 2.1233	0.25	1.3	1.3	1.3	1.1	1.0
	0.50	11.5	11.5	11.0	1.8	1.0
	1.00	185.0	64.7	14.3	2.2	1.0
	1.50	5.0	4.5	3.4	1.8	1.1
	2.00	1.7	1.7	1.6	1.4	1.1
w = 3 UCL= 1.7358	0.25	1.0	1.0	1.0	1.0	1.0
	0.50	5.2	5.2	4.5	1.1	1.0
	1.00	185.0	46.7	9.0	1.6	1.0
	1.50	3.8	3.5	2.7	1.5	1.0
	2.00	1.5	1.4	1.4	1.2	1.0
w = 4 UCL= 1.5066	0.25	1.0	1.0	1.0	1.0	1.0
	0.50	3.3	3.3	2.6	1.0	1.0
	1.00	185.0	36.9	6.4	1.4	1.0
	1.50	3.1	2.8	2.2	1.3	1.0
	2.00	1.3	1.3	1.2	1.1	1.0
w =5 UCL= 1.3540	0.25	1.0	1.0	1.0	1.0	1.0
	0.50	2.4	2.4	1.9	1.0	1.0
	1.00	185.0	29.8	5.0	1.3	1.0
	1.50	2.6	2.4	1.8	1.2	1.0
	2.00	1.2	1.2	1.2	1.1	1.0

Table 3. ARL profiles for the MaxEWMA chart with ARL_0 =185 and n =5

	a					
	b	0.00	0.25	0.50	1.00	2.00
λ=0.05 *K*=3.3833	0.25	2.0	2.0	2.0	1.9	1.0
	0.50	3.5	3.5	3.4	2.2	1.0
	1.00	185.5	12.1	5.1	2.3	1.2
	1.50	4.1	3.9	3.3	2.2	1.2
	2.00	2.1	2.1	2.0	1.7	1.2
λ=0.10 *K*=3.4028	0.25	1.8	1.8	1.8	1.8	1.0
	0.50	3.3	3.3	3.1	2.0	1.0
	1.00	185.0	12.3	4.8	2.1	1.1
	1.50	3.7	3.5	3.0	2.0	1.2
	2.00	1.9	1.9	1.8	1.6	1.2
λ=0.15 *K*=3.2254	0.25	1.8	1.8	1.8	1.8	1.0
	0.50	3.3	3.3	3.1	2.0	1.0
	1.00	185.0	14.7	4.0	2.1	1.1
	1.50	3.7	3.5	2.9	1.9	1.2
	2.00	1.9	1.8	1.8	1.5	1.2
λ=0.20 *K*=3.0996	0.25	1.8	1.8	1.8	1.8	1.0
	0.50	3.5	3.5	3.3	2.0	1.0
	1.00	185.1	18.4	5.4	2.1	1.1
	1.50	3.9	3.6	3.0	2.0	1.1
	2.00	1.9	1.9	1.8	1.5	1.2
λ=0.25 *K*=3.0415	0.25	1.8	1.8	1.8	1.8	1.0
	0.50	3.7	3.7	3.5	2.1	1.0
	1.00	185.1	23.1	6.0	2.2	1.1
	1.50	4.1	3.8	3.1	2.0	1.2
	2.00	1.9	1.9	1.8	1.6	1.2
λ=0.30 *K*=3.0278	0.25	1.8	1.8	1.8	1.8	1.0
	0.50	4.1	4.0	3.8	2.1	1.0
	1.00	185.0	28.5	6.6	2.2	1.1
	1.50	4.3	4.0	3.2	2.0	1.2
	2.00	1.9	1.9	1.8	1.6	1.2
λ=0.40 *K*=3.0499	0.25	1.9	1.9	1.9	1.8	1.0
	0.50	4.9	4.9	4.7	2.2	1.0
	1.00	185.0	38.3	8.3	2.3	1.1
	1.50	4.7	4.3	3.4	2.1	1.2
	2.00	2.0	1.9	1.8	1.6	1.2
λ=0.50 *K*=3.0884	0.25	1.9	1.9	1.9	1.8	1.0
	0.50	6.5	6.5	6.2	2.3	1.0
	1.00	185.0	47.7	10.25	2.5	1.1
	1.50	5.0	4.6	3.6	2.1	1.2
	2.00	2.0	2.0	1.9	1.6	1.2
λ=0.80 *K*=3.0019	0.25	2.6	2.6	2.6	2.4	1.0
	0.50	21.4	21.4	21.1	4.2	1.0
	1.00	185.0	76.8	19.9	3.3	1.1
	1.50	6.5	5.9	4.5	2.4	1.2
	2.00	2.2	2.1	2.0	1.7	1.2
λ=1.00 *K*=2.6918	0.25	4.8	4.8	4.8	4.7	1.0
	0.50	50.4	50.4	50.2	12.1	1.0
	1.00	185.0	96.8	30.3	4.4	1.1
	1.50	7.3	6.7	5.1	2.7	1.2
	2.00	2.3	2.2	2.1	1.7	1.2

REFERENCES

Chao, M.T. and S.W. Cheng (1996). Semicircle control chart for variables data. Quality Engineering, 8, 441-446.

Chen, G. & S.W. Cheng (1998). Max chart: combining *X*-bar chart and *S* chart. *Statistica Sinica*, 8, 263-271.

Chen, G., S.W. Cheng & H. Xie (2001). Monitoring process mean and variability with one EWMA chart. *Journal of Quality Technology*, 33, 223-233.

Chen, G., S.W. Cheng & H. Xie (2004). A new EWMA control chart for monitoring both location and dispersion. *Quality Technology & Quantitative Management*, 1, 217-231.

Costa, A.F.B. & M.A. Rahim (2004). Monitoring process mean and variability with one non-central chi-square chart. *Journal of Applied Statistics*, 31, 1171-1183.

Costa, A.F.B. & M.A. Rahim (2006). A single EWMA chart for monitoring process mean and process variance. *Quality Technology & Quantitative Management*, 3, 295-305.

Domangue, R. & S.C. Patch (1991). Some omnibus exponentially weighted moving average statistical process monitoring schemes. *Technometrics*, 33, 299-313.

Gan F.F. (2000). Joint monitoring of process mean and variance based on the exponentially weighted moving average. *Statistical Process Monitoring and Optimization.* Marcel Dekker, New York.

Gan, F.F., K.W. Ting & T.C. Chang (2004). Interval charting schemes for joint monitoring of process mean and variance. *Quality and Reliability Engineering International*, 20, 291-303.

Khoo, M.B.C. & P.W. Yap (2005). Joint monitoring of process mean and variability with a single moving average control chart. *Quality Engineering*, 17, 51-65.

Morais, M.C. & A. Pacheco (2000). On the performance of combined EWMA schemes for μ and σ: A Markovian approach. *Communications in Statistics – Simulation and Computation*, 29, 153-174.

Reynolds, M.R. & Z.G. Stoumbos (2004). Control charts and efficient allocation of sampling resources. *Technometrics*, 46, 200-214.

Spiring, F.A. & S.W. Cheng (1998). An alternate variables control chart: The univariate and multivariate case. *Statistica Sinica*, 8, 273-287.

White, E.M. & R. Schroeder (1987). A simultaneous control chart. *Journal of Quality Technology*, 19, 1-10.

Wu, Z. & Y. Tian (2006). Weighted-loss-function control chart. *International Journal of Advanced Manufacturing Technology*, 31, 107-115.

Yeh, A.B., D.K.J. Lin & C. Venkataramani (2004). Unified CUSUM charts for monitoring process mean and variability. *Quality Technology & Quantitative Management*, 1, 65-86.

Zhang, S. & Z. Wu (2006). Monitoring the process mean and variance using a weighted loss function CUSUM scheme with variable sampling intervals. *IIE Transactions*, 38, 377-387.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 461–482.

TOURISM SECTOR DYNAMICS IN EGYPT IN LIGHT OF THE GLOBAL FINANCIAL CRISIS

Ahmed Badr[1], Enas Zakareya[2] and Mohamed Saleh[3]
[1,2] Economic Issues Program (EIP), Information and Decision Support Center (IDSC), Egyptian Cabinet.
[3] Cairo University, Egypt
University of Bergen, Norway.
E-mail: [1] amabadr@idsc.net.eg, [2] enabd@idsc.net.eg, [3] saleh@salehsite.info

ABSTRACT

As the tourism sector is very sensitive to economic uncertainty; this study aims at investigating the likely impact of the Global Financial Crisis (GFC) on the tourism sector in Egypt. The analysis will be conducted mainly through an examination of the internal structure of the tourism sector in Egypt and the relationship between key variables based on The Travel & Tourism Competitiveness Index (TTCI), 2009 published by the World Economic Forum (World Economic Forum (WEF), 2009) that determines the competitiveness position of the Egyptian tourism sector. This paper will focus on six vital TTCI sub-indices to analyze their effects on the Egyptian market share. Our study objective is twofold: Firstly, analyzing the impacts of the GFC through exploring the dynamic behavioral relationships between variables. Secondly, simulating these effects through carrying out scenario analysis to determine the impact of the crisis on tourism and testing several hypotheses. Policy recommendations will be tested with the aim of mitigating the negative implications of the crisis on the tourism sector in Egypt.

1. INTRODUCTION

The global financial crisis (GFC) presents significant challenges for world tourism growth. GFC has also exposed weaknesses in the functioning of the global economy and led to calls for the reform of the international financial architecture. Although the crisis was triggered by events in the United States housing market, it has spread to all regions of the world with dire consequences for global trade, investment and growth. The crisis represents a serious setback for the tourism industry because it is taking place at a time when the tourism business is making progress in the growth rate of tourist arrivals.

According to the United Nations World Tourism Organization (UNWTO), international tourist arrivals grew at a global average of 5% for the first quarter of the year 2008 compared with the same period in 2007. Growth was fastest in the Middle East region, (12.5%). World tourism demand sustained in the first months of 2008 and witnessed a decline in the last months of 2008. According to the UNWTO World Tourism Barometer, the negative trend in international tourism that emerged during the second half of 2008 intensified in 2009. International tourist arrivals have suffered a sharp drop between January and April 2009 (about 8.2%, a drop of -8.2% implies a gain for the first quarter of 2009). This implies a drop in the global tourist arrivals of about 247 million, in the first quarter this year, down from 269 million in 2008 (UNWTO World Tourism Barometer, June 2009). International tourism is expected to decline between 6% and 4% percent during the rest of 2009 compared to the same period in 2008. Sub-Saharan Africa is the only bright spot on the horizon with arrivals forecast to grow by 1-5% for all regions; growth is projected to be negative: Asia and the Pacific (4-1%), Americas (6-3%), Europe (8-5%) and the Middle East (10-5%)[1].

In light of the GFC, consumer demand is falling in both the business and leisure tourism sectors, according to the UNWTO (June 2009). The government of Egypt (GOE) assigns a high priority to the tourism sector, however, its market share of the worldwide tourism market is very low (1.4 %) and is ranked 64th out of 133 countries (WEF, 2009). As the Egyptian government aims at enhancing the TTCI, testing different policy scenarios concerning the impact of the GFC on the tourism sector is very important for decision makers to envisage the likely effects of the crisis and hence to make appropriate adjustments to current strategies.

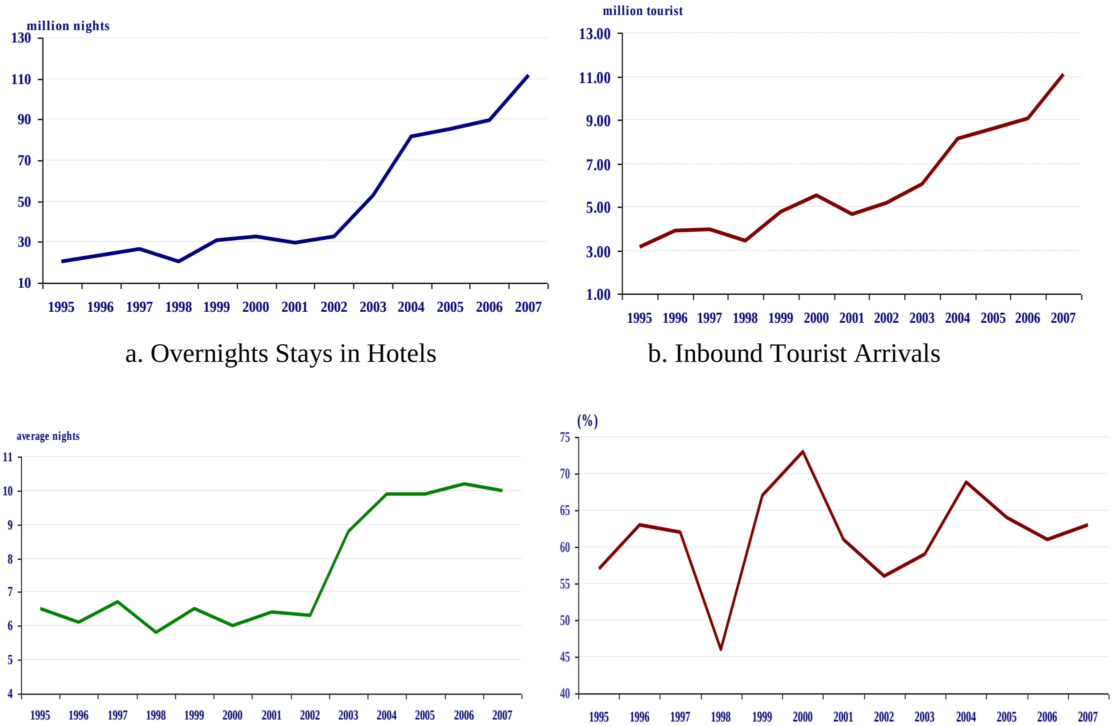
The objective of this paper is to analyze the interactions within the tourism sector in Egypt, and to test policies to mitigate the impact of the global financial and economic crisis on the tourism sector. We develop a dynamic model to simulate these interactions and to analyze the effects of different scenarios. This dynamic model is used to determine the impact over time, including delayed effects. The tourism dynamics model is calibrated according to the historical time series of the key variables in the tourism sector during the period between 1995 and 2007. We use this model to assess the effects of the GFC and to test different mitigation policies over the period 2008-2015.

The remainder of the paper proceeds as follows. Section 2 presents a brief historical overview that delineates the developments in the Egyptian tourism sector. Section 3 describes the TTCI which represents the core foundation of the model. Section 4 focuses on previous research concerning tourism in Egypt and describes how system dynamics can be useful in representing tourism dynamics. Section 5 describes the system dynamics model. Section 6 illustrates the model behavior under different conditions and presents the comparative analysis of four selected scenarios. Section 7 concludes our investigation and suggests possible future work.

2. THE TOURISM SECTOR IN EGYPT: AN OVERVIEW

The Tourism sector in Egypt has witnessed significant developments in the last decade. A review of the evolution of the tourism sector since 1995 provides a background to the assessment of the sector's prospects in future. In the last thirteen years, tourism arrivals increased more than three-fold from 3.1 million in 1995 to 11.1 million in 2007, recording an average annual growth about 12.3% (Figure 1). The average duration of stay of visitors increased from 6 nights to 10 nights over the same period (Figure 2). Overnight tourist hotel stays grew from 20.4 million in 1995 to about 111.5 million nights in 2007 implying an annual increase of about 17.8% on average (figure 1c). Average room occupancy rate increased from 57% in 1995 to 63% in 2007 (figure 1d). It is worth mentioning that the tourist arrivals have increased from 20-28% for the whole period 1995-2007 (Sakr *et al.*, 2009).

[1] UNWTO market monitoring indicates that the plummeting results of international tourism during the last part of 2008 have continued during the first months of 2009. International tourist arrivals are estimated to have declined by as much as 8% in the first two months of 2009, bringing overall international tourism to the level of 2007.



a. Overnights Stays in Hotels

b. Inbound Tourist Arrivals

c. Average length of Stay

d. Occupancy Rate

Figure1. Main Tourism Indicators 1995-2007
Source: World Tourism Organization (UNWTO), 2008

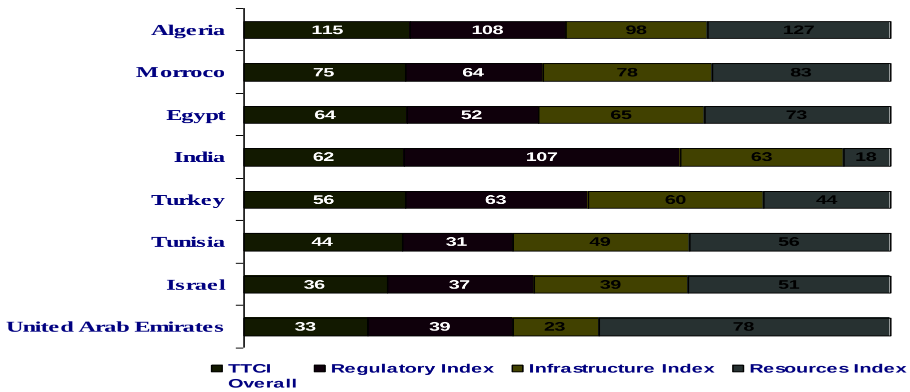
3. TTCI DESCRIPTION

We used the most recent Travel and Tourism Competitiveness Index (TTCI) (WEF, 2009) that evaluates and ranks the competitive performance of 133 countries and suggests possible improvements for decision makers wishing to improve their travel and tourism sectors. The TTCI[2] is composed of 14 pillars grouped into three groups, namely, (a) the travel and tourism regulatory framework, (b) business environment and infrastructure and (c) human, cultural and natural resources. Each pillar is also contains several indicator variables. For example, pillar one – policy rules and regulations – includes eight components; namely, prevalence of foreign ownership, property rights, impact of business rules on foreign direct investment (FDI), visa requirement, openness of bilateral air service agreements, transparency of government policies, time required to start a business, and cost to start a business. TTCI provides a comprehensive tool for measuring the factors and policies that make the Travel and Tourism sector attractive (WEF, 2009). TTCI is quite highly correlated[3] with both the number of tourists traveling to various countries and the annual income generated from Travel & Tourism. This supports the idea that the TTCI captures factors that are important for developing the Travel and Tourism

[2] More details about TTCI methodology can be found in (Appendix 1).
[3] For further details about the correlation between TTCI, international tourist arrivals and receipts, See figure 2 & 3 P.8,(WEF, 2009).

(T&T) industry (WEF, 2009). Our model relies heavily on some of the different pillars constituting the TTCI that can help policy makers identify the weak and strong areas related to the competitiveness of the sector and hence identify the necessary needed adjustments. Analyzing the impact of the GFC can be established through exploring the dynamic behavioral relationships between the main variables used in computing the TTCI. Figure 2 demonstrates the favorable position of Egypt relative to its competitors in the Middle East and North Africa (MENA) region. Egypt occupies the 64th position among 133 countries according to the global TTCI. In terms of the regulatory index, it is ranked 52nd, and in the infrastructure index it is ranked 65th. According to the resources index it is ranked 73rd (WEF, 2009). A detailed description of the TTCI index and its components are shown in Table (1) in the appendix.

Figure2. Egypt's TTCI Ranking among Selected Countries in MENA Region



Source: WEF (2009).

4. LITERATURE REVIEW

Most empirical studies estimate the tourism demand function on the basis of econometrics techniques. Sakr et al (2009) estimated a demand function for Egypt based on Giacomelli (2006a) using panel data including both price and non price determinants such as income, cost of living, nominal exchange rate, transport cost, a measure of infrastructure, and other factors that determine the arrivals of tourists (Crouch, 1995; Eugenio, 2002; Hellstrom, 2002). Following the three-stage estimation procedure of Eilat and Einav (2009), estimates of Sakr et al (2009) for the tourism demand function for Egypt revealed that the rule of law and regulations tended to be the most important factor that positively affects tourism demand[4]. The price of tourism services is the most significant factor that negatively impacts tourist arrivals.[5]

Sakr and Masoud (2003) studied the determinants of volatility of the tourism sector in Egypt found out that not all types of shocks exert the same effect on the tourism sector or last for the same period of time – this is to distinguish between domestic and external shocks –, based on their estimation of the tourism demand function for Egypt for the period 1986-2001. Tourists' responsiveness to different events was examined. Both tourism arrivals and receipts have a positive relation with the GDP in the countries of origin, as well as with the depreciation in the

[4] A one percent improvement in the rule of law resulted in about 5.98 percent increase in tourist arrivals in Egypt (Eugenio, 2002).

[5] Every one percent increase in the price is accompanied by 5.36 percent reduction in arrivals.

effective exchange rate. The most important factor determining the demand for tourism in Egypt is the extent of international competitiveness as expressed by the effective exchange rate. Their study highlighted the importance of developing international competitiveness to broaden Egypt's tourism market. Their study also revealed that the fall in tourism flows and receipts in the aftermath of domestic shocks were more severe compared to the corresponding fall resulting from external shocks[6].

Some of the previous studies primarily focused on analyzing the historical behavior. In this study we aim to construct various scenarios for the future of the sector, taking into consideration the impact of GFC. Given the complexity and dynamic nature of our objective, we view system dynamics as the best approach to address this problem, since it accounts for feedback loops, time delays between causes and consequences and nonlinear relationships between variables. To our knowledge, this is the first study to investigate the impacts of GFC on the tourism sector in Egypt using system dynamics.

5. MODEL STRUCTURE OVERVIEW

The model contains twelve sub-models, eight stocks[7] (world tourists, Egyptian recent tourists, labor, rooms constructed, room capacity, air infrastructure index, ground infrastructure index and health index) and eleven balancing[8] and five reinforcing[9] feedback loops[10]. This section focuses on presenting the dynamic structure of the model. The main sub-models (briefly described below) are as follows:

1- Annual Arrivals
2- Tourism Revenues
3- Global Economic Crisis
4- Employment
5- Hotel Rooms Capacity
6- Travel and Tourism Competitiveness Index
7- Security and Safety Index
8- Tourism Infrastructure Index
9- Air Transport Infrastructure Index
10- Ground Transport Infrastructure Index
11- Health care Index
12- Price Competitiveness in the Travel and Tourism Industry

[6] The global and regional external shocks that affected the perception of safety are mainly the September 11th attacks, the Gulf war and the domestic shock that took place in 1997, namely Luxor attacks. In the aftermath of September 11th attacks, the number of arrivals dropped by 0.4 and 0.6 percent respectively, but the corresponding decrease in real tourism receipts were more striking totaling 10.6 and 5.2 percent, respectively (WTO, 2002c; Sakr and Masoud, 2003).

[7] A stock is a system dynamics concept that represents an entity that you're keeping track of over time

[8] Produce stable, balance, equilibrium and goal-seeking behavior

[9] Generate behaviors of growth, amplify, deviation, and reinforce

[10] Cause and effect linkages among system components that affect system behavior

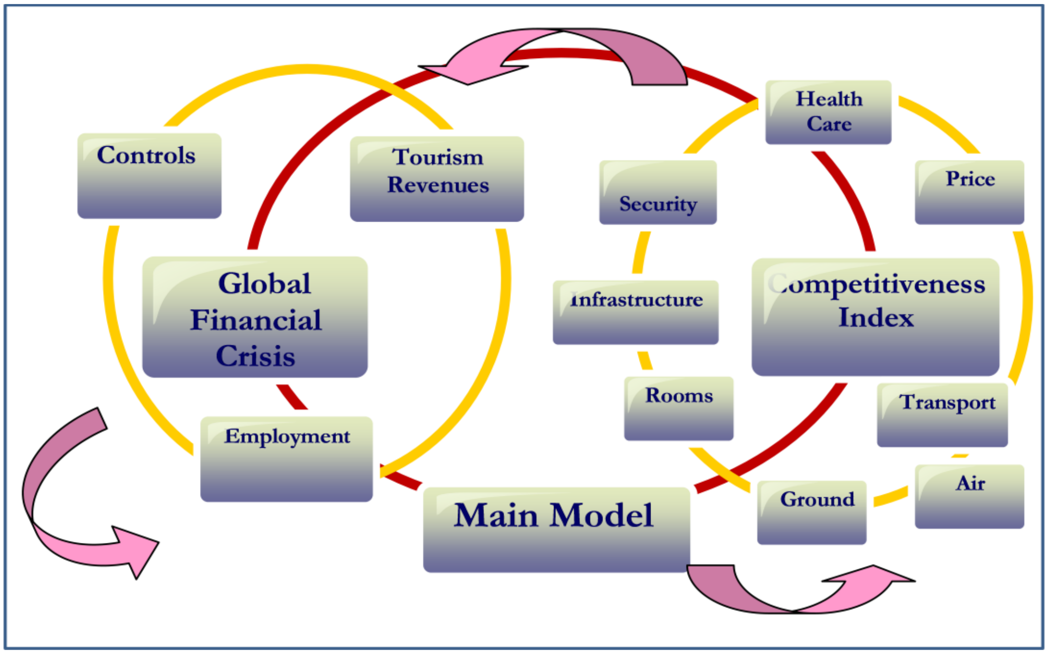


Figure 3. The General Model Structure
Source: Schematic presentation for the system dynamics model.

5.1 Annual Arrivals

This sub-model depicts the flow of international tourists to Egypt each year and illustrates its determinants. As Egypt receives a significant share (4%) from international tourists, Egypt's annual arrivals (an inflow to the Egyptian stock of tourists and at the same time outflow of worldwide stock of tourist arrivals) are defined as the number of international tourists visit Egypt every year. The Egyptian market share is mainly determined by the current value of the Egyptian TTCI. A stock of world tourists represents all international tourists worldwide. This stock is increased by the accumulation of newly attracted tourists according to a growth rate that is assumed to be constant (excluding any possible external effects that can hit the tourism sector). Tourist inter-travel period is the average duration between two consecutive visits to the same country for a tourist. Moreover, a tourist who visits Egypt in any given year is assumed to have a specific probability to revisit Egypt after a specific period (the re-potential period)[11]. Last year tourists that visit Egypt are accumulated in a stock which increases by the inflow of Egypt annual arrivals and decreases by the re-potential rate.

5.2 Global Economic Crisis

The GFC hit the tourism sector and has a significant effect through different variables. This sub-model describes these variables affected by the crisis. GFC affected world tourism growth and hence the Egyptian tourism growth as well. Average nights per tourist, average tourism expenditure and inter-travel period are among the key variables that are affected by GFC. To

[11] A tourist who visits Egypt in 2009 returns to the world tourists stock and has a probability to visit Egypt again in 2011.

incorporate these effects we use four multiplier coefficients that are multiplied by certain reference values. The analysis of the effects of GFC on these variables will be presented in the next section.

5.3 Travel and Tourism Competitiveness Index

As we mentioned before, the TTCI is composed of 14 important pillars. We focus on the six most important pillars out of fourteen since we expect only these six pillars will change significantly during the study period (2008 to 2015). These six pillars are price competitiveness index, tourism infrastructure index, security index, health index, air infrastructure index and ground infrastructure index. The remaining 8 pillars are treated within our model as constants. Unweighted average of the 14 variables (pillars) is used to calculate the Egyptian competitiveness index (WEF, 2009). Our philosophy behind utilizing the components of TTCI is mainly driven by the large number of empirical studies that have investigated the determinants of a country's competitiveness and attractiveness to tourists (for example Giacomelli, 2006a). These studies have shown TTCIto be suitable for measuring and capturing most of the determinants of tourism and travel competitiveness.

5.3.1 Tourism Infrastructure Index

One of the key building blocks in the tourism sector dynamics is the tourism infrastructure which represents the supply side of the model. This sub-model describes the tourism infrastructure index and its determinants the effect of occupancy ratio, the effect of car rental companies and the effect of ATMs availability and accessibility. We use the elasticity concept to model the effect of occupancy ratio on the infrastructure index. Occupancy ratio represents the balance between supply (i.e. room capacity) and demand (i.e. tourist nights). Increasing the demand of tourist nights based on a higher occupancy ratio which is considered a good indicator. On the other hand, increasing capacity given a fixed demand resulting in lower occupancy ratio is considered a bad indicator due to the over-supply without any consideration of demand.

5.3.2 Air Transport Infrastructure Index

This sub-model is inspired by the generic model described by Sterman (Sterman 2000, P.264) where the stock values exhibits a goal seeking behavior to reduce the gap between its current value and a certain desired value. Air index represents one of the main policy intervention points in the model. First in the base case scenario (flags are off[12]) the air index is constant. However during policy analysis, the air index is assumed to exhibit a goal seeking behavior in order to reduce the gap between its current value and the desired value set by the policy maker. The enhancement of the index can be achieved via different ways, for example enhancing the quality of existing airports, building new airports, extending international air transport network, increasing the number of operating airlines, etc.

[12] Default values are used

5.3.3 Ground Transport Infrastructure Index

Similar to the air index, we use the same sub-model for, ground infrastructure index except that the enhancement of the index occurs through increasing the number of roads, extending ground infrastructure network, enhancing quality of roads, etc.

5.3.4 Health Care Index

Health care is one of the critical factors upon which tourists make decisions for visiting a specific destination. The same sub-model structure used in air and ground indices will be customized to represent the dynamics of the health care index structure. The change, from air and ground indices, will be through intervention from both public and private sectors by increasing physician density per population, facilitating access to improved drinking water and increasing available hospital beds for people, etc.

5.3.5 Security and Safety index

One of the crucial factors contributing to tourism arrivals is the security and safety index. Tourists' personal safety is a function of their ability to travel without threats. Therefore, stimulating inbound tourism Can be increased by improving confidence in Egypt as a safe and destination. This index is composed of four indexes capturing the effects of crime, terrorism, road accidents and police services.

5.3.6 Price Competitiveness

This sub-model describes how the Tourism and Travel price been constructed. Price competitiveness is determined through five main sub-indexes, namely ticket taxes and airport charges, purchasing power parity, taxation, fuel price levels and hotel prices. Unweighted Average of the five sub-indexes is used to calculate the price index. It is worth mentioning that occupancy ratio is one of the determinants of hotel price. Hotel price is determined through a ratio between supply (i.e. room capacity) and demand (i.e. tourism nights). The relevance of these measures of price competitiveness is being recognized in studies by the tourism industry. Studies of tourist demand go beyond using the nominal exchange rate as a crude measure of price competitiveness and the use of real exchange rates (nominal rates adjusted for changes in the general level of prices). These studies attempt to determine a more relevant exchange rate, which is adjusted for changes in the prices that identifies tourist bundles of goods and services (Martin and Witt, 1987). However, these do not go beyond trends of prices and do not determine whether a country is more or less competitive than another at a particular point of time. To measure tourist services prices, as opposed to simply trends in tourism prices, cross-sectional studies using the prices paid by tourists in different countries are needed.

5.4 Annual Tourism Revenues

Tourism receipts are one of the main variables used to judge the improvement or recession of the tourism industry. Only foreign tourism nights are used to calculate Egypt annual tourism revenues. Foreign tourism nights are calculated according to the multiplication of the average nights a tourist spends in Egypt by the number of international tourists that visit Egypt annually. Egypt annual tourism revenues are similarly calculated by multiplying foreign tourism nights by the average tourist expenditure per day.

5.5 Tourism Employment

The labor force for the tourism sector is aggregated into a single stock. Tourism employment captures the labor dynamics surrounding the tourism sector and is increased by the hiring rate and decreased by the attrition rate. Because of the seasonality of the tourism industry, hotel managers are willing to eliminate surplus 'temporary' employees when not needed ,which decreases the labor stock by the firing rate (negative inflow).The attrition rate (outflow) consists of voluntary terminations and normal retirements, and is modeled as a first-order process[13] in which employees remain with the tourism industry for the average duration of employment.

Desired labor force is the number of employees hotel managers are willing to employ to serve the expected number of tourists. Labor stock is increased by the extra employees needed[14]. Hotel managers can't equate current labor force to the desired labor force without a delay to announce vacancies, interviews and hire the appropriate candidates to fill these positions, etc. We used annual arrivals instead of delayed annual arrivals because the employment sector usually responds quickly to any changes or fluctuations in the tourism sector.

5.6 Egypt's Hotel Rooms Supply Chain

This sub-model is based on a generic supply chain model (Sterman 200015, Chapters. 17 & 19) representing the hotels capacities in Egypt. Like all countries, Egypt's hotels and tourism villages utilize a set of processes that include order fulfillment, requesting new hotel rooms and capacity adjustment. The full hotel rooms supply chain is simulated by the following sub-model, starting with ordering new hotel rooms[16] from suppliers as requested by tourists.

This sub-model contains two stocks. The first stock represents the supply chain governing the provision of hotel rooms for the tourism sector and is comprised of the stock of unfulfilled orders for new hotel rooms, i.e. orders that have been placed with manufacturers but not yet received. This stock is increased by the accumulation of yet more orders for rooms (inflow), is affected by the start construction rate and depleted by an outflow (finished rooms ready for service).Note that the MAX[17] function is used for both the start and finish construction rates to ensure nonnegative values each.

Another stock consists of all Egypt's hotel rooms and is used to represent the capacity of these hotel rooms This stock increases by receiving new rooms from suppliers that are ready for service (outflow from construction stock and inflow to capacity stock) and is diminished by depreciated rooms that are no longer available for use. This stock is also assumed to depreciate at a constant rate; the depreciation rate is calculated by dividing the number of rooms by the average lifetime of the rooms.

[13] The relationship between remaining employees and attrition rate is linear.
[14] Difference between desired number of employees and the actual number exist in the system.
[15] Business Dynamics: System Thinking and Modeling for a Complex World, John D. Sterman. 2000
[16] Hotel rooms represent all available rooms that can be used by tourists (e.g. hotels, tourism villages, etc.)
[17] Max between zero and construction rates.

Hotel manager can't add as many rooms as they wish; there are money constraints and delays that limit their desires. The construction of new hotels, for example, requires labor, equipment and time. Hotel managers adjust the current stock every year to meet the desired stock. The desired number of rooms is determined by two main factors. Firstly, they aim to replace the depreciated rooms by ordering new rooms. Secondly, they consider perceived tourism nights[18] and the occupancy ratio they wish to achieve. Similarly, the suppliers of hotel rooms adjust the construction of new rooms analogously to hotel managers adjusting their room capacity by taking into consideration the construction time lag between having a job order for a new room and delivering that room to be used.

6. SCENARIO ANALYSIS AND SIMULATION RESULTS

The impact of the GFC on the Egyptian tourism sector can be analyzed using the system dynamics model that helps the decision maker understand the dynamics within the tourism sector, and by simulating the likely repercussions and impacts of the GFC on the tourism sector. Different simulation runs can be conducted to test the behavior of the key tourism indicators and interactions of the variables within the suggested framework in order to achieve the objective of this study. Figure 4 demonstrates four different scenarios for the decision maker to carry out the simulation of the impact of GFC on the tourism sector. These four scenarios the base, pessimistic, mitigation, and optimistic scenarios and are illustrated in Figure 4. The policy scenarios depend mainly on the extreme values of two major uncertainties, namely, the severity of the impact of GFC, and the change of the enabling environment. We will briefly explain each scenario and then perform a comparative analysis of all four scenarios. Each scenario is illustrated through a movement on a scale depicting the severity of GFC which affects the change in world growth rate of tourism arrivals, tourist spending in Egypt, average tourist nights and finally the change in inter-travel period (figure 5).

Figure 4. Scenario Cross Analysis

6.1 Base Scenario

In the base scenario we assume, other things being equal, the current policies will remain in the future. Low impact of the GFC and no significant change will take place. The main purpose of

[18] Perceived tourism nights are the actual tourism nights delayed by a specific period of time until hotel managers comprehend the change in values.

this run is to evaluate model behavior and outputs under the normal state during the study period (2008-2015).

6.2 Pessimistic Scenario

In contrast to the base scenario, the pessimistic scenario, presumes a severe impact of the GFC owing to a sharp decline in the world growth rate of international tourists, average tourist expenditure, average length of stay, and a significant increase in the inter-travel period of international tourists. Therefore, further investment will be needed through intervening in one or more areas of tourism competitiveness sub-indices (health, air, ground, etc.) to mitigate the effect of GFC.

6.3 Mitigation Scenario

A third scenario can be envisaged in the wake of GFC. The severe impact of GFC would induce the Government of Egypt (GOE) to intervene with a set of stimulating policies that could enhance the environment for the tourism and travel sector. The GOE efforts may include encouraging and accelerating investment in specific areas[19] that could result in a major improvement in the sector. The system dynamics model simulates such policies by turning on the flags for air infrastructure index, ground infrastructure index as well as the health care index.

6.4 Optimistic Scenario

Finally, an optimistic scenario representing a low impact of the GFC associated with a significant intervention from the GOE is considered. In this regard, the optimistic scenario presumes a small decline in the world growth rate of international tourists, a small decrease in the average tourist spending, a small decrease in the average length of days spent in Egypt and a slight increase in the inter-travel period of international tourists.

6.5 Comparative Analysis of the Scenarios for Global Financial Crisis

Based on the model described section 6, we simulate the four main scenarios 6.1-6.4 to explore different types of tourism policies and conditions for the period 2008-2015, which allows sufficient time for the adjustments to take place given the dynamics of the tourism sector. Four main variables were investigated under the four scenarios; the results are shown in Figures 6, 7, 8 and 9. The four variables are TTCI, market share, annual arrivals and employment. Below we describe the behavior of each variable under the four scenarios.

According to the optimistic scenario, the competitive ranking of the Egyptian tourism sector can be dramatically increased by 15 positions to occupy a ranking like 49th instead of 64th out of 133 countries. This position will make the tourism sector in Egypt more competitive than other competitors in the Middle East and North Africa (MENA) region such as Turkey (56) and Jordan (54).This would mainly achieved as a matter of changing the basic blocks of the tourism dynamics to restore the environment for the development of the tourism sector . The pessimistic scenario forecasts an increase of Egypt’s ranking by 8 places to occupy the 56th position, a tie with Turkey. The mitigation scenario (together with the optimistic scenario) forecast a TTCI

[19] Air infrastructure , ground infrastructure and health service

index higher than the TTCI forecast under the base scenario (Fig. 6). The mitigation scenario assumes that the TTCI improvement will target and enhance the Egyptian competitiveness index during a period of five years.

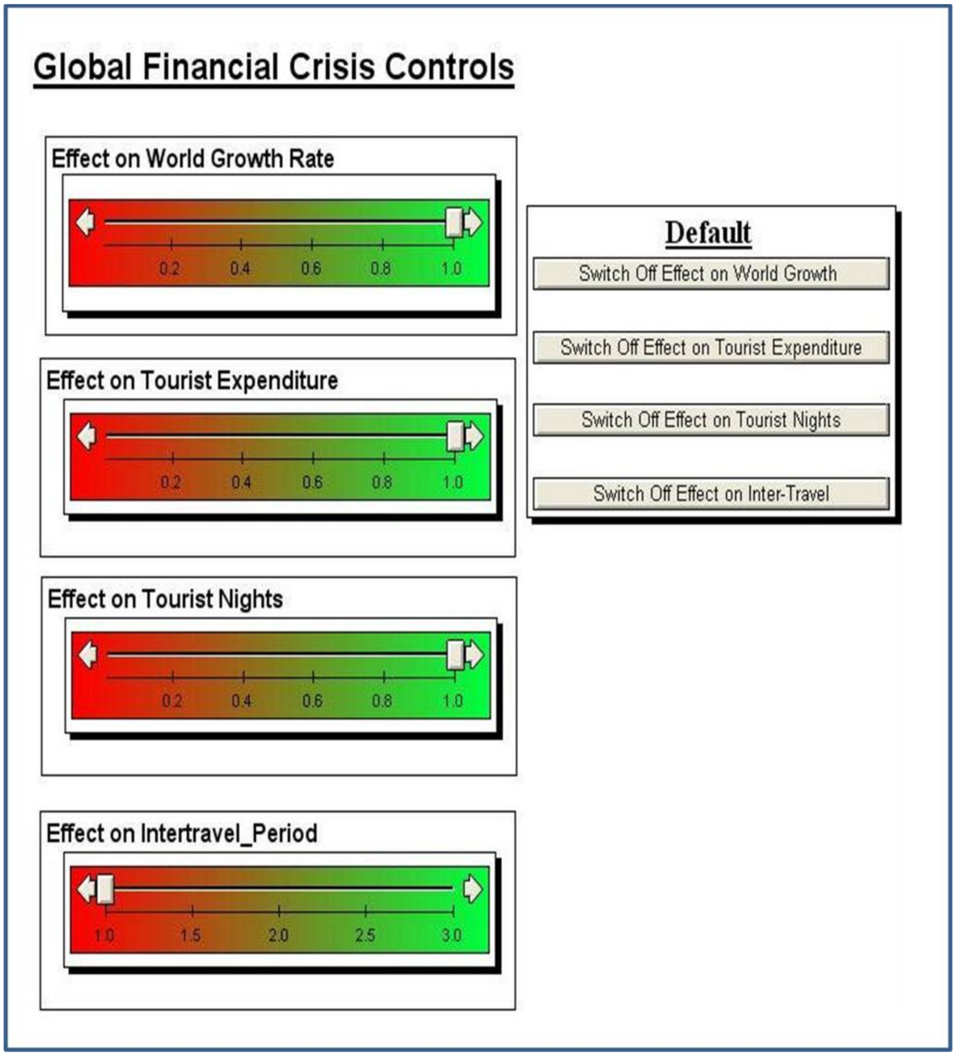


Figure 5. Decision Maker's Simulation for Scenario Crisis effects[20]

[20] Red and green regions represent high and low effects of crisis, respectively.

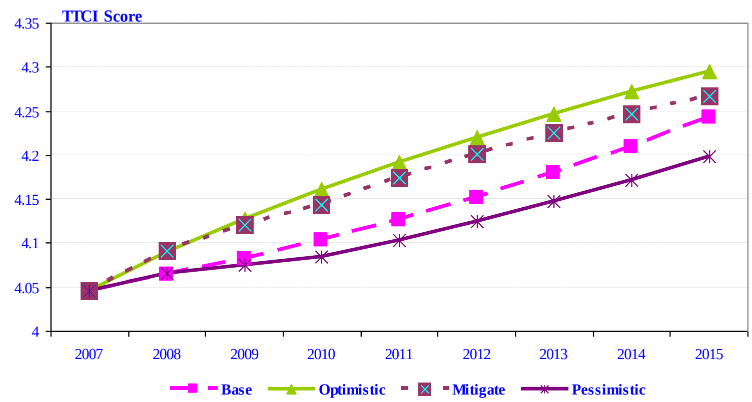


Figure 6. TTCI

With regard to the Egyptian market share of international tourists (Figure 7), the results reveal that market share is forecasted to reach an unprecedented level by 2015 according to the optimistic scenario (1.7%)[21]. As per our assumption that the market share is determined using the Egyptian TTCI, the behavior of the market share follows the behavior of the TTCI. It is worth noting that enhancing the enabling environment in the tourism sector will increase the distance between both the optimistic scenario (Low GFC effect) and the mitigate scenario (Significant GFC effect) with the base scenario (No significant change).

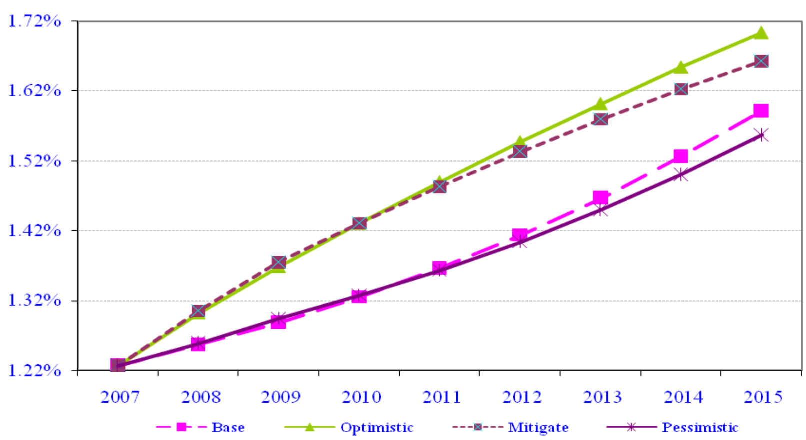


Figure 7. Egyptian Market Share

Compared to market share, annual tourism arrivals behave differently across the 4 scenarios. Although, Egypt is forecast to achieve a higher market share under the mitigation scenario (Figure 7) than under the base scenario; annual arrivals are forecast to be lower under the mitigation scenario than under the base scenario (Figure 8). This can be explained by the GFC causing a decrease in the arrivals of tourists worldwide and hence a decrease in Egyptian tourist arrivals as well. The highest forecasted average growth rate of tourism arrivals is 8.2 %, under the optimistic scenario compared with 4.8 % under the mitigation scenario and 3.8 % under the

[21] 0.1% of international tourists equals 922000 tourist in 2008 (Crouch, 1995).

pessimistic scenario. The corresponding increase in forecasted tourist arrivals varies from 11.1 million in 2008 to 16.9 million in 2012 and 20.8 million tourists by 2015 under the optimistic scenario, while the lowest forecasted number of tourist arrivals occurs under the pessimistic scenario and is predicted to be 14.8 million in 2015. The mitigation scenario lies forecasts the number of tourist arrivals to be between the predicted values of the pessimistic and optimistic scenarios. Because of external shocks that affect the tourism sector under the mitigation scenario; forecasted annual arrivals under the mitigation scenario are less than the forecasted arrivals under the base scenario during 2008-2015 (Fig 8).

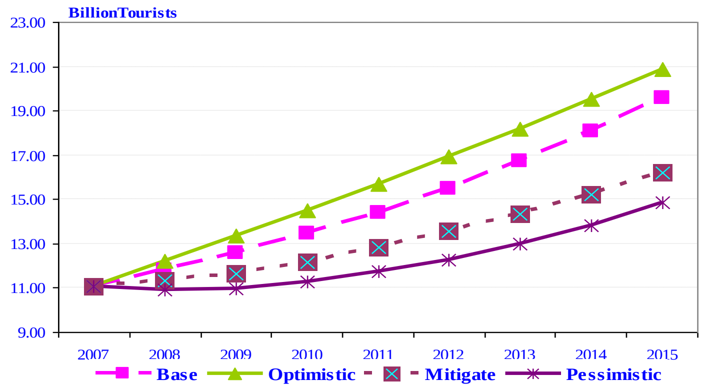


Figure 8. Annual Arrivals

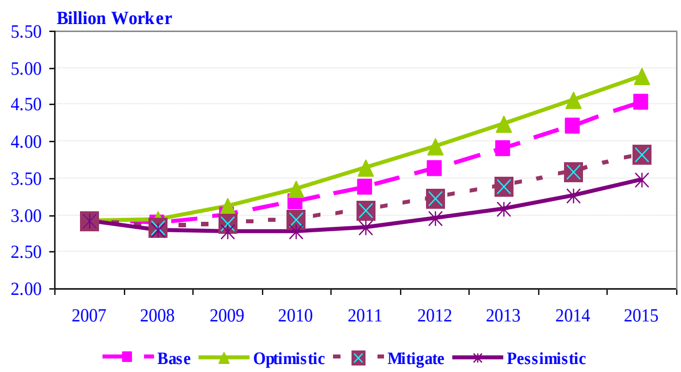


Figure 9. Employment

Tourism employment is forecasted to achieve rapid growth under the optimistic scenario, and is forecast to attain an annual average growth rate of 6.7% during the study period 2008-2015 (Fig. 9). The average growth rate of employment is forecast to be 3.49 under the mitigation scenario. The lowest forecasted growth rate occurs under the pessimistic scenario, namely 2.31. The least average growth rate could be achieved under the pessimistic scenario. It is worth noting that tourist employment follows a similar pattern as tourist arrivals (Figures 8, 9) because tourism labor is highly correlated with the number of tourists. Finally, increasing the Egyptian TTCI would be a matter of interest to Egypt in order to enhance the tourism sector, attract more tourists, increase market share and increase the labor force.

7. CONCLUSION AND POLICY IMPLICATIONS

Constructing a system dynamics model in analyzing the impacts of the GFC and simulating these effects to determine the impact of the crisis on tourism sector in Egypt has many advantages. First, regarding the competitive position of the Egyptian tourism sector, the theoretical model and resulting simulations based on this model increases our knowledge about the interactions of both supply and demand factors driving the competitive position of the tourism sector. Second, the model can be used to assess the impacts of external shocks such as GFC as well as domestic shocks (terrorism attacks and political instability), although the former is the primary focus of the paper while the latter falls out of the scope of this study. Third, this model provides the decision maker with a policy framework that is capable of introducing specific interventions in the areas that are crucial to the tourism sector corresponding to different levels of the severity of the shock. At the same time the decision maker can easily monitor the behavior of key indicator variables in the tourism sector and design targeted policies to influence the specific needs and areas. Fourth, the results of the simulation exercise are useful in suggesting recovery strategies[22] to enhance the sustainability of the tourism sector as well as its contribution to the growth and development of the economy. Finally, the model captures both demand and supply determinants of tourism and travel, thereby more space for policy makers to choose which area that could be enhanced in the planning for the future of the sector development.

Among the areas that can be influenced23 is the safety and security index. This implies that sustainable efforts should be directed towards regaining a positive image and restoring confidence in Egypt as a tourist destination. Another promising area24 is to improve tourist health security and tourism infrastructure for both air and ground facilities. Should Egypt strive for a significant market share of world tourism it has to exert significant effort to improve the score and ranking of TTCI among competitors, especially in the Middle East and North Africa (MENA) region.

For possible future research directions, the model can be expanded to distinguish between domestic and foreign inbound tourism. Moreover, the model could be enhanced to distinguish between domestic and external shocks. Furthermore, the model could also be extended to include the competitors of Egypt.

ACKNOWLEDGEMENT

The authors would like to thank Magued Osman, IDSC chairman, and Professor of Statistics at Cairo University, who gave us the support to complete this research. We are extremely indebted to Tarek Moursi, Professor of Economics at Cairo University and an economic consultant at IDSC for suggesting the topic and for valuable comments.

[22] Enhancing air and ground infrastructure and health service
[23] Areas influenced by government or private sector to enhance the competitiveness index.
[24] As per model results, enhancing these sub-indexes will increase Egypt TTCI.

APPENDIX 1

TTCI METHODOLOGY

This section briefly summarizes the methodology of constructing the Travel and Tourism Competitiveness Index (TTCI). TTCI is composed of three sub-indices: the T&T regulatory framework, the T&T business environment, and infrastructure and the T&T human, cultural, and natural resources. These sub-indices form the fourteen pillars of the TTCI described in Table (1). All sub-indices include both, hard and soft variables which are defined according to the underlying method of data collection:

A) *Soft data* come from the World Economic Forum "Executive Opinion Survey (EOS)". This Survey is conducted every year and asks top management business leaders to compare their own operating environment with global standards on a wide range of dimensions. All EOS questions ask the user to provide a response on a scale between 1-7.

B) *Hard data* indicators are obtained from a variety of sources such as international organizations and leading businesses in the Travel and Tourism industry (for example, IATA, the IUCN, the UNWTO, the WTTC,UNCTAD, and UNESCO)). These data are scaled between 1 to 7 so as to be aligned with the soft data using the following formula:

$$6*\left(\frac{Country\ Score - Sample\ Minimum}{Sample\ Maximum - Sample\ Minimum}\right)+1$$

Clearly, the sample minimum and maximum are the lowest and highest scores of the overall samples, respectively. This formula is used only when a higher score indicates a more favorable outcome (number of available hospitals, ability to invest, etc.). On the other side, another formula needs to be used for cases in which higher score represents a worse outcome (pollution, number of road accidents, etc. ,…)

$$6*\left(\frac{Country\ Score - Sample\ Minimum}{Sample\ Maximum - Sample\ Minimum}\right)+7$$

Each pillar is calculated as the unweighted average of the included variables(soft and hard). Finally, the overall country TTCI is calculated as an unweighted average of the country’s fourteen pillar values.

Table (1) TTCI Description

Sub-index		Pillar		Variable	Description
T&T regulatory framework	1	Policy rules and regulations	1.1	Prevalence of foreign ownership	Foreign ownership of companies in your country is (1 = rare and limited, 7 = prevalent and encouraged)
			1.2	Property rights	Property rights in your country, including over financial assets, are (1 = poorly defined and not protected by law, 7 = clearly defined and well protected by law)
			1.3	Business impact of rules on FDI	In your country, rules governing foreign direct investment (1 = discourage foreign direct investment, 7 = encourage foreign direct investment)
			1.4	Visa requirements	Number of countries whose citizens are exempt from obtaining a visa (= 1) or able to obtain one upon arrival (= 0.5) out of all UN countries | 2008
			1.5	Openness of bilateral Air Service Agreements	Index of openness of bilateral Air Service Agreements | 2005
			1.6	Transparency of government policymaking	Are firms in your country usually informed clearly by the government of changes in policies and regulations affecting your industry? (1 = never informed, 7 = always informed)
			1.7	Time required to start a business	Number of days required to start a business | 2008
			1.8	Cost to start a business	Cost to start a business as a percentage of GNI per capita | 2008
	2	Environmental sustainability	2.1	Stringency of environmental regulation	How stringent is your country's environmental regulation? (1 = lax compared with most countries, among the world's most stringent)
			2.2	Enforcement of environmental regulation	Enforcement of environmental regulation in your country is (1 = lax, 7 = rigorous)
			2.3	Sustainability of T&T industry development	Is your country's government taking the necessary steps to ensure that the Travel & Tourism sector is being developed in a sustainable way? (1 = no, development of the sector does not take into account issues related to environmental protection and sustai
			2.4	Carbon dioxide emissions	Carbon dioxide emissions per capita in metric tons | 2004

Sub-index		Pillar		Variable	Description
			2.5	Particulate matter concentration	Urban population–weighted PM10 micrograms per cubic meter | 2005
			2.6	Threatened species	Threatened species as a percentage of total species (mammals, birds, amphibians) | 2008
			2.7	Environmental treaty ratification	Total number of ratified environmental treaties | 2008
	3	Safety and security	3.1	Business costs of terrorism	The threat of terrorism in your country (1 = imposes significant costs on businesses, 7 = does not impose significant costs on businesses)
			3.2	Reliability of police services	Police services in your country (1 = cannot be relied upon to enforce law and order, 7 = can be relied upon to enforce law and order)
			3.3	Business costs of crime and violence	The incidence of common crime and violence in your country (1 = imposes significant costs on businesses, 7 = does not impose significant costs on businesses)
			3.4	Road traffic accidents	Estimated deaths per 100,000 population due to road traffic accidents | 2002
	4	Health and hygiene	4.1	Physician density	Physician density per 1,000 people | 2006 or most recent year available
			4.2	Access to improved sanitation	Access to adequate sanitation as a percentage of total population | 2006
			4.3	Access to improved drinking water	Access to safe drinking water as a percentage of total population | 2006
			4.4	Hospital beds	Hospital beds per 10,000 population | 2006
	5	Prioritization of Travel & Tourism	5.1	Government prioritization of the T&T industry	The development of the Travel & Tourism industry is a priority for the government of your country (1 = strongly disagree, 7 = strongly agree) SOURCE: World
			5.2	T&T government expenditure	T&T government expenditure as a percentage of total budget | 2008
			5.3	Effectiveness of marketing and branding to attract tourists	Does your country carry out effective marketing and branding to attract tourists? (1 = no, tourism marketing is nonexistent or completely ineffective, 7 = yes, tourism marketing is excellent, and is very effective in attracting tourists)

Sub-index		Pillar		Variable	Description
			5.4	T&T fair attendance	Index of country presence at 13 major T&T fairs, 2007–2008
T&T business environment and infrastructure	1	Air transport infrastructure	1.1	Quality of air transport infrastructure	Passenger air transport in your country is (1 = underdeveloped, 7 = extensive and efficient by international standards)
			1.2	Available seat kilometers, domestic	Scheduled available seat kilometers per week originating in country (in millions) | January to July 2008 average
			1.3	Available seat kilometers, international	Scheduled available seat kilometers per week originating in country (in millions) | January to July 2008 average
			1.4	Departures per 1,000 population	Number of departures per 1,000 population | 2006
			1.5	Airport density	Number of airports per million population | 2007
			1.6	Number of operating airlines	Number of airlines with scheduled flights originating in country | January 2008 and July 2008 average
			1.6	International air transport network	Does the air transport network in your country provide good connections to the overseas markets offering the greatest potential to your business? (1 = no, not at all, 7 = yes, to all of my key business markets)
	2	Ground transport infrastructure	2.1	Quality of roads	Roads in your country are (1 = underdeveloped, 7 = extensive and efficient by international standards)
			2.2	Quality of railroad infrastructure	Railroads in your country are (1 = underdeveloped, 7 = extensive and efficient by international standards)
			2.3	Quality of port infrastructure	Port facilities and inland waterways in your country are (1 = underdeveloped, 7 = extensive and efficient by international standards)*
			2.4	Quality of domestic transport network	
			2.5	Road density	Kilometers of road per 100 square kilometers of land | 2005 or most recent year available
	3	Tourism infrastructure	3.1	Hotel rooms	Number of hotel rooms per 100 population | 2007 or most recent year available
			3.2	Presence of major car rental companies	Index of presence of major car rental companies | 2008
			3.3	ATMs accepting Visa cards	Number of automated teller machines (ATMs) accepting Visa credit cards per million population | 2007

Sub-index		Pillar		Variable	Description
	4	ICT infrastructure	4.1	Extent of business Internet use	Companies within your country use the Internet extensively for buying and selling goods, and for interacting with customers and suppliers (1 = strongly disagree, 7 = strongly agree)
			4.2	Internet users	Internet users per 100 population | 2007 or most recent year available
			4.3	Telephone lines	Telephone lines per 100 population | 2007 or most recent year available
			4.4	Broadband Internet subscribers	Broadband Internet subscribers per 100 population | 2007 or most recent year available
			4.5	Mobile telephone subscribers	Mobile telephone subscribers per 100 population | 2007 or most recent year available
	5	Price competitiveness in the T&T industry	5.1	Ticket taxes and airport charges	Index of relative cost of access (ticket taxes and airport charges) to international air transport services | (0 = highest cost, 100 = lowest cost) | 2008
			5.2	Purchasing power parity	Ratio of purchasing power parity (PPP) conversion factor to official exchange rate | 2007
			5.3	Extent and effect of taxation	The level of taxes in your country (1 = significantly limits the incentives to work or invest, 7 = has little impact on the incentives to work or invest)
			5.4	Fuel price levels	Retail diesel fuel prices (US cents per liter) | 2006
			5.5	Hotel price index	Average room rates calculated for first-class branded hotels for calendar year in US$ | 2007
T&T human, cultural, and natural resources	1	Human resources	1.1	Primary education enrollment	Net primary education enrollment rate | 2006 or most recent year available
			1.2	Secondary education enrollment	Gross secondary education enrollment rate | 2006 or most recent year available
			1.3	Quality of the educational system	The educational system in your country (1 = does not meet the needs of a competitive economy, 7 = meets the needs of a competitive economy)
			1.4	Local availability of specialized research and training services	In your country, specialized research and training services are (1 = not available, 7 = available from world-class local institutions)

Sub-index		Pillar		Variable	Description
			1.5	Extent of staff training	The general approach of companies in your country to human resources is (1 = to invest little in training and employee development, 7 = to invest heavily to attract, train, and retain employees)
			1.6	Hiring and firing practices	The hiring and firing of workers is (1 = impeded by regulations, 7 = flexibly determined by employers)
			1.7	Ease of hiring foreign labor	Labor regulation in your country (1 = prevents your company from employing foreign labor, 7 = does not prevent your company from employing foreign labor)
			1.8	HIV prevalence	HIV prevalence as a percentage of adults aged 15–49 years | 2007 or most recent year available
			1.9	Business impact of HIV/AIDS	How serious do you consider the impact of HIV/AIDS on your company in the next 5 years? (1 = extremely serious, 7 = not a problem)
			1.10	Life expectancy	Life expectancy at birth (years) | 2006
	2	Affinity for Travel & Tourism	2.1	Tourism openness	
			2.2	Attitude of population toward foreign visitors	Are foreign visitors welcome in your country (1 = totally unwelcome, 7 = very welcome)
			2.3	Extension of business trips recommended	When senior executives visit your country for the first time for business purposes, how likely are you to recommend extending their trip for leisure purposes? (1 = I would never recommend it, 7 = I would always recommend it)
	3	Natural resources	3.1	Number of World Heritage natural sites	Number of World Heritage natural sites in the country | August 2008
			3.2	Protected areas	Protected areas as a percentage of total land area | 2007
			3.3	Quality of the natural environment	The natural environment in your country is: (1 = among the most polluted in the world, 7 = as clean as the least polluted countries in the world)
			3.4	Total known species	Total known species (mammals, birds, amphibians) in the country

Sub-index		Pillar		Variable	Description
					| 2008
	4	Cultural resources	4.1	Number of World Heritage cultural sites	Number of World Heritage cultural sites and Oral & Intangible Heritage | August 2008
			4.2	Sports stadiums	Sports stadium capacity per million population | 2008
			4.3	Number of international fairs and exhibitions	Number of international fairs and exhibitions held in the country annually | 2005–2007 average
			4.4	Creative industries exports	Exports of creative industries products as a share of world total in such exports | 2006 or most recent year available

Source: (WEF, 2009).

REFERENCES

Crouch, G., 1995, "A Meta-Analysis Study of International tourism Demand," Annals of Tourism Research.

Eilat, Y., and L. Einav, (2004), "Determinants of International Tourism: A Three Dimensional Panel Data Analysis,"

Eugenio, J., (2002), "A Primer on Modeling International Tourism Demand with Discrete Choice Models," University of York.

Giacomelli, A., (2006a). "Modeling International Tourism Demand," Second International Conference on Tourism Economics, Spain.

Hellstrom, J., 2002, "Count Data Modeling and Tourism Demand," Economic Studies, no. 584. Sweden.

John D.Sterman, (2000), "Business Dynamics: System Thinking and Modeling for a Complex World" Irwin Professional Pub.

Sakr M. and Nada Massoud, (2003), "International Tourism Volatility with Special Reference to Egypt," The Egyptian Center for Economic Studies (ECES). Working Paper No.83.

Sakr M., Nada Massoud and Hala Sakr, (2009), "Tourism in Egypt: An Unfinished Business," The Egyptian Center for Economic Studies (ECES).Working Paper No. 147.

World Economic Forum (WEF), (2009) "The Travel & Tourism Competitiveness Report", 2009.

World Tourism Organization, (2009) "UNWTO World Tourism Barometer", Vol. 7, No. 2, June 2009.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 483–493.

MODELLING OF GROUNDWATER BY USING FINITE DIFFERENCE METHODS AND SIMULATION

Adam Baharum, Hessah Faihan AlQahtani, Zalila Ali, Habibah Lateh, and Koay Swee Peng
School of Mathematical Sciences,
Universiti Sains Malaysia
E-mail: adam@cs.usm.my

ABSTRACT

In active landslide, the prediction of acceleration of movement is crucial issue for the design and performance of warning systems. Landslide occurs when a sudden increase beyond the critical level of groundwater. This is especially true in tropical weather during the wet season. The purpose of this study is to use numerical model to simulate groundwater flow. The goal of this modelling is to predict the value of unknown nodal points in the groundwater piezometric head. The numerical technique used is the finite difference method. The finite difference method is one of the oldest, most general applicable and most easily understood methods of obtaining numerical solution to steady and unsteady groundwater problems. After we obtain the algebraic approximation equations for each node in solution boundary domain, we solve them with digital computer program. Our research presents a broad, comprehensive overview of the fundamental concepts and applications of computerized groundwater modelling. The research covers finite difference method and includes simulation runs to demonstrate theoretical points described. Our model is able to predict the value of aquifer parameters in particular slope.

1. INTRODUCTION

Groundwater is water located under the ground surface in soil poor spaces and in the fractures of lithologic formations. In the wet season in countries with tropical rainforests, the rate of accumulation of groundwater is greater than normal. Groundwater levels may accelerate as rainfalls increase both in terms of frequency and in amount Figure (1.1). This leads to sudden and unpredicted landslides. Most landslides are preceded by a saturation of a slope. This is commonly caused by heavy rainfall events, periods of extended precipitation that saturate the upper part of the slope, or increase in groundwater levels.

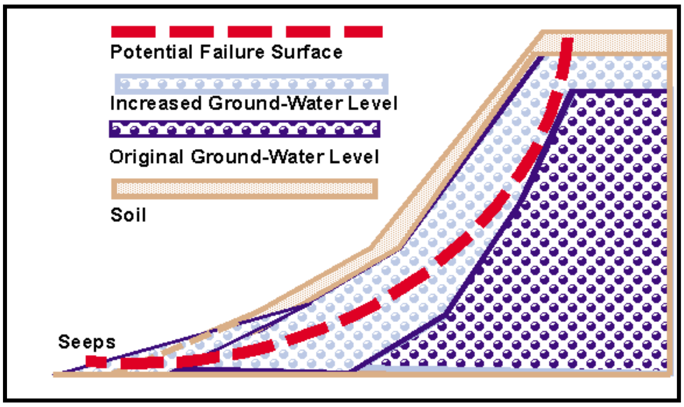


Figure (1.1): The potential for landslide increases when groundwater levels rise due to rain, or landscape irrigation

Landslides are a recurring hazard in Malaysia. They occur in a majority of states in Malaysia. Additionally, landslides and the many other ground failures result in many direct and indirect expenses to society. Some of these direct costs include lost of life and the actual physical damage which runs the gamut from cleanup and repair to replacement. Indirect costs are harder to measure and include business disruption, loss of tax revenues, reduced property values, loss of productivity, losses in tourism, and losses from litigation. Therefore, we need to study the effects of landslides. , And, as we mentioned earlier, all previous studies about landslide have considered the increase in groundwater level caused by slope collapse. In our research we study a model of groundwater to measure the groundwater in a particular slope in order to predict the amount of water and if it has reached a supercritical level. Groundwater models may be used to predict the effects of hydrology on the behaviour of the aquifer. These are often called groundwater simulation models. As the calculations are based on mathematical equations, often with approximate numerical solutions, these models are also called mathematical/numerical groundwater models. A mathematical model consists of a set of differential equations that are known to govern the flow of groundwater. Mathematical models of groundwater flow have been in use since the late 1880's.

As (Wang and Anderson, 1982) state in their book, "Introduction to Groundwater Modelling," they consider two types of models: finite difference models and finite element models. The finite difference models described are based on Laplace's equation and the finite element models described are based on Poisson's equation. Both of them compare and test the numerical solutions by a classical analytical solution of a similar problem. After one year, (Hunt, 1983) wrote in his book, "Mathematical Analysis of Groundwater Recourses," He considers the finite difference method as one of the approximate solutions for boundary-value problems because it is an easily understood method that can provide an approximate solution under very general circumstances. On the other hand, Hunt says, analytical solution is usually easier and more economical to use and interpret then numerical solution. Additionally, they are often useful as standards to test the accuracy of numerical models. (Corominas, Moya, Ledesma, Lloret, and Gili, 2004) investigate the prediction of ground displacements and velocities from groundwater level changes at the Vallcebre landslide (Eastern Pyrenees ,Spain). Their model to predict both landslide displacements and velocities was performed at Vallcebre by solving the momentum equation in which a viscous term (Bingham and Power Law) was added. They found that the landslide is very sensitive to rainfall, cracks, and drainage pathways. These models used in the time-dependent simulations are based on this "dynamic approach" rather than just considering "static" limit equilibrium.

The subject of this research is using the numerical methods to solve mathematical model that simulate groundwater flow and contaminant transport. We consider the finite difference method to be the numerical technique to solve our model. The finite difference method is an easily understood method that can provide approximate solutions under general circumstances. A mathematical groundwater model for steady flow conditions consists of a governing equation and boundary conditions which simulate the flow of groundwater in a particular problem domain. To solve our model we have to calculate the value of head at each point in system. The numerical techniques that we consider are finite difference methods which provide a rationale for operating on the differential equations. Using a computer, one can solve large number of algebraic equations by iterative solution as in our model. We solved 29 algebraic equations by using Gauss Seidel iterative methods.

2. THE FINITE – DIFFERENCE METHOD

The finite difference method is considered as the most applicable and easily understood methods of obtaining numerical solutions to steady and unsteady groundwater flow problems. The general method consists of superimposing a finite – deference grid of nodes upon the solution domain. Actually, each node is given a global identification number and its surrounding of each of these nodes, where the dependent variable is approximated with a finite – degree polynomial whose coefficients are written in forma of the unknown values of the dependent variable at the surrounding nodes. So this polynomial is used to obtain an algebraic approximation for the partial differential equation for each internal node beside an algebraic approximation for boundary condition at each node site upon or near the solution domain boundary.

After we obtain the algebraic equations for each node we can solve the equation simultaneously to obtain the unknown value of the dependent variable at all nodes.

3. SOLUTION OF THE FINITE-DIFFERENCE EQUATIONS

There are two method to solved the system of simultaneous equation that is generated by writing equation at each interior node and equation at each boundary node we can solved by direct elimination method such as Gaussian elimination or iterative method such as the Gauss –Seidel iteration. We can say the direct method more efficient under certain circumstances. However, some of applications consider here may need simultaneous solution of one or two thousand equations with sparse matrix and relatively large diagonal terms in the coefficient matrix. For these conditions, the iterative methods are easier to cod for computer, where it's take less computer storage and less computational time for that we will consider the iterative methods.

4. ANALYSIS AND DISCUSSION

4.1 The Model of Groundwater

The groundwater models are representations of reality and, if properly constructed, can be valuable predictive tools for the management of groundwater as one of water resources and a tool to predict the effects of groundwater on the movement of landslide as our case study. Of course, the validity of the predictions is predicated on how well the model approximates filed conditions.

The development and building of models for groundwater levels have to follow a specific number of well defined steps:

4.1.1- Identify the Particular Problem

For our research we want to know how much water can be in a particular area in the slope or the groundwater level at the slope. Therefore, one of the main factors during landslide is the level of the forces associated with groundwater. If we can predict the groundwater when it's reaching the critical level in the slope this will be useful as an early warning message to the people who could be affected by the landslide.

4.1.2- Formulate the Boundary Value Problem

The problem of formulation assists us in organizing our thinking about a particular problem. We will give each node a global identification number, as well as in the neighbourhood of each of these nodes. The dependent variable h (piezometric head) is approximated with a finite degree polynomial whose coefficients are written in terms of unknown values of the dependent variable at the surrounding nodes.

We will use this polynomial to obtain two algebraic equations

i. An algebraic approximation for the partial differential equation at each interior node.

ii. An algebraic approximation for a boundary condition at each node that lies upon or beside the solution domain boundary.

The equation for a steady two dimensional flow is.

$$\bar{\nabla}.(T\bar{\nabla}h) = \left[\frac{\grave{k}}{\grave{B}}\right]\left(h - \grave{h}\right) + \frac{Q_0}{A} - R \tag{1.1}$$

where

$\bar{\nabla}$:Vector operator del (L-1)

T: Transmissivity (L2 T-1)

h: piezometric head

$\grave{k}$: aquifer permeability (L T-1)

$\grave{B}$: aquifer thickness (L)

Q_0: discharge form a well at node 0.

R: rainfall recharge.

By simplifying this equation and putting in more convenient forms by collection terms which are coefficients of the unknown values of h:

$$A_1 h_1 + A_2 h_2 + A_3 h_3 + A_4 h_4 - A_0 h_0 = -B_0 \tag{1.2a}$$

$$A_i = \frac{1}{2}(T_0 - T_i) \tag{1.2b}$$

$$A_0 = A_1 + A_2 + A_3 + \left[\frac{\dot{k}}{\dot{B}}\right] \tag{1.2c}$$

$$B_0 = \left[\left[\frac{\grave{k}}{\grave{B}}\right]\left(\grave{h}_0\right) + \frac{Q_0}{A} - R\right]\Delta^2 + Q_0 \tag{1.2d}$$

We conclude from Equation (1.2) the approximate algebraic for the partial differential equation at each interior node. With this equation we will solve it in our code because it gives us the values of dependent variable h at every interior node that lies inside the solution domain boundary. After we obtain the algebraic approximation for each interior node .We have to obtain the algebraic approximation for each interior node which lies upon or beside the solution domain.

Along the boundaries, the condition to obtain an algebraic approximation is

$$\alpha \frac{\mathrm{dh}}{\mathrm{dn}} + \beta h = F \tag{1.3}$$

where

α : Aquifer bulk coefficient of compressibility.

β: Bulk coefficient of compressibility for water .

F: a prescribed function equal to the elevation of the water surface above the coordinate origin.

In general α, β and F are discontinuous functions. We can put $\beta = 1, \alpha = 0$ and F=f or$\beta = 1,\ \alpha = 1$ and F=0.

We will approximated the equation (1.3) in the neighbourhood of the three node grid show in figure (1. 2a) by using the following first –degree polynomial:

$$h(x, y) = h_0 + \frac{h_0 - h_2}{\Delta} x + \frac{h_0 - h_2}{\Delta} y + 0(\Delta^2) \tag{1.4}$$

Depending upon whether node 0 lies within or outside of the solution domain, respectively, For that equation(1. 4)gives

$$\left[\frac{dh}{dn}\right]_p = (\overline{\nabla} h . \widehat{e_n})_p = \frac{h_0 - h_2}{\Delta} N_x + \frac{h_0 - h_1}{\Delta} N_y \tag{1.5}$$

$$(h)_p = h_0 + \frac{h_0 - h_2}{\Delta} N_x \delta + \frac{h_0 - h_1}{\Delta} N_y \ \delta \tag{1.6}$$

$$A_0 h_0 - A_1 h_1 - A_2 h_2 = F_p \tag{1.7}$$

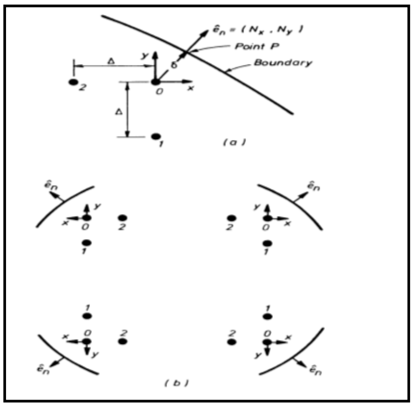


Figure (1.2) a & b: A three node grid along a boundary

In which F_p denotes the value of F at point P and coefficients, A_1, A_2, A_0 are given by

$$A_1 = (\alpha + \beta\delta)\frac{N_y}{\Delta} \tag{1.7a}$$

$$A_2 = (\alpha + \beta\delta)\frac{N_x}{\Delta} \tag{1.7b}$$

$$A_0 = A_1 + A_2 + \beta \tag{1.7c}$$

Equation (1. 7) is the algebraic approximation to equation (1. 3) that is written for each node that lies either on or adjacent to the solution domain boundary. Therefore, we can state that we obtain the two algebraic approximations. The first one is for the interior nodes and the second one is for the nodes which lie upon or adjacent to the solution domain boundary.

$$A_1 h_1 + A_2 h_2 + A_3 h_3 + A_4 h_4 - A_0 h_0 = B_0 \qquad (interior\ nods) \tag{1.2}$$

$$A_0 h_0 - A_1 h_1 - A_2 h_2 = F_p \qquad (boundry\ nodes) \tag{1.7}$$

These two equations are the formulation of the boundary value problem and we will solve these two equations in step four by using the finite difference method after we blog our data.

4.1.3- Collect Field Data

Prior to starting step three, we need to know that step two and step three are serial because a correctly posed boundary value problem cannot be constructed without first knowing these basics, such as the location and the type of aquifer boundaries, the sources of recharge, etc. Additionally, collecting large amounts of field data before the problem is defined mathematically can lead to large expenditures of time and money in collecting data that is unnecessary.

This step is usually the most expensive and time consuming part of any investigation in that we have to use the skills and experiences of a large number of people in various fields. Actually, we need different experts for different aspects. We obtain the recharge rates from rainfall for the slope which is in our study. The location of this slope is in Jalan Tun Sardon, Pulau Pinang, Malaysia; see Figure (1.3).

Figure (1.3): General view of Jalan Tun Sardon slope

We determined the accumulation of rainfall for the period from 30 August 2008 until 13September 2008. We will use the two weeks to demonstrate how much the accumulation of rainfall changes when the weather is taken into consideration for this area of Malaysia which is located in the tropical rainforest. Figure (1.4) illustrates the water level Vs time.

We have selected the 30th of August, 2008, because there is an increasing accumulation of rainfall. The recharge of rainfall is significant data for our study because it is one of the coefficients that will be used in our algebraic equation (1.3) and R will create a specified maximum for h of the accumulation of rainfall R for 30th August 2008 is 0.8×10^{-6} m.

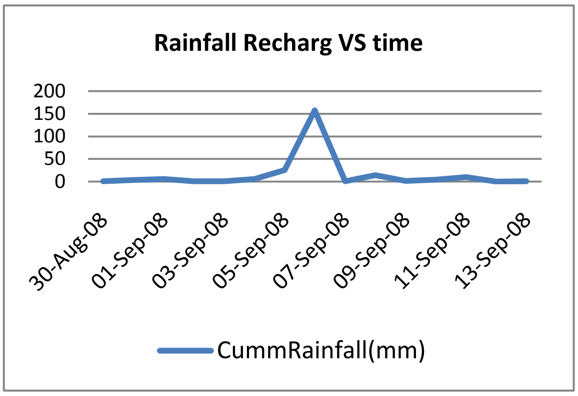


Figure (1.4): Rainfall data

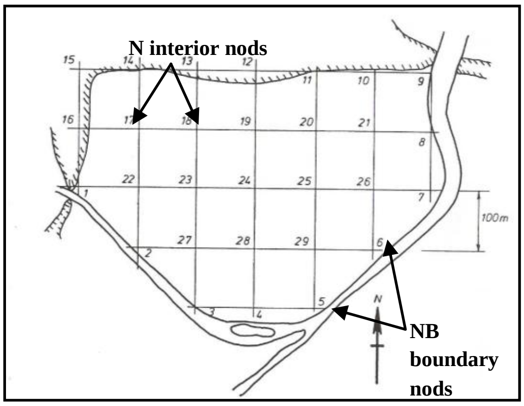


Figure (1.5): Rainfall data

We now need to determine the piezometric head for each interior point or for each point that lies on or next to the boundary domain. Since we don't have the kind of tools needed, which are difficult to obtain and which are too costly, we will use the same piezometric head that Hunt (1983) employed as well as our recharge of rainfall. Let us assume the solution domain boundary as shown in figure (1.5) and NB the number of boundary points, and N, the number of nodes in the interior grid.

Probably the main difficulty is to relate the global and the local numbering schemes in a way which has enough flexibility to solve problems on solution domain with curved boundaries. We can overcome this problem with the flowing scheme:

1. Each point that lies on or next to the boundary is assigned a global number between 1 and NB. Also, each interior point is assigned a global number between NB+1 and N.

2. Each of these points is assigned a series of either four or two integers that give the global number of surrounding nods. For the interior nodes N these integers will be called in our cod I D(I,J) where I goes from I through 4. For the boundary nodes NB I D(I,J) where J goes from 1 through 2. These integers and global number of the nodes corresponds with local numbering.

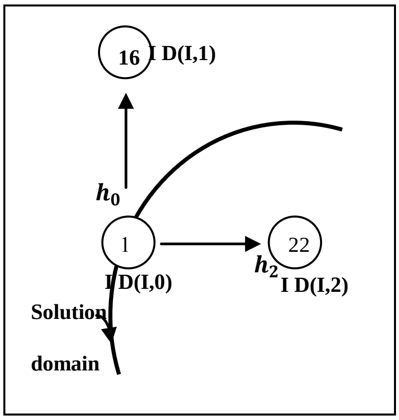


Figure (1.6): A typical three node boundary node

For example, to make this clear we take: Boundary node 1, I D(1,j) where J = 1and 2, I D (1, 1) = 16, and I D(1,2)=22. As Figure (1.6) and this describe our input data where we take the NB=16

4.1.4- Solve the Boundary Value Problem That Was Formulated In Step Two

In step four we calculate answers to the equations that were posed in step two and the data gathered in step three is incorporated directly into the formulation and solution of the problem. In our study to solve the system of simultaneous equations that are generated by writing equation (1.7) at each interior node and equation (1.2) at each boundary node, we can solve them through various methods. We must first choose the finite difference method because it is an easily understood method that can provide approximate solutions under very general circumstances. In short, iterative methods consist of guessing and adjustments. There are three iterative techniques: Jacobi iteration, Gauss Seidel iteration, and Successive over Relaxation (SOR). Jacobi iteration is the least efficient and is seldom used. Gauss Seidel iteration can be considered to be a special case of Successive over Relaxation.

The change between two successive and Gauss Seidel iterations is called" residual c. In the method of SOR, the Gauss Seidel residual multiplied by relaxation factor $\omega \geq 1.$ For the SOR method $0 < \omega < 1.$For the Gauss Seidel method $\omega = 1.$

However, the applications considered here require the simultaneous solution of twenty nine equations with a sparse matrix and relatively large diagonal terms in the coefficient matrix. With these conditions, our choice is the Gauss Seidel iterative method since it is easier to code for a computer ,it requires considerable less computer storage, and it uses less computational time. The Gauss Seidel method will be the last step for the code. After calculating the coefficient of two previous equations to obtain the matrix, we will solve it by Gauss Seidel iteration to estimate for (piezometric head) h. Then we will run the program by using our data with recharge of rain full R=0.0000008m/s=2104mm/month, the output as below in Figure (1.8).

5. ANALYSIS FOR OUTPUT

From the Figure (1.7) above we note that the aquifer has been assumed to be homogeneous, that h will be the change in water level created by the uniformly distributed recharge rate R and that the problem is inverse problem in which the unknown is the value for R that will create a specified maximum for h. However, this inverse problem is easily to solved because the problem is liner in both hand R. Thus, multiplying all of the equations by the same constant λ, shows that if a recharge rate of R creates a water level change of h. Then a recharge rate λR will create a water level change of λh. The output for the computer program gives the water table rise for a recharge rate of R (Hunt data) R = $0.1 * 10^{-6}$ m/s = 263 mm/month. Since the maximum computed rise is 0.315m at the node 12.

The value for R that will give a maximum water level rise of 1 meter is

$R = \frac{1m}{0.315} * 263$ mm/ month=835 mm/month .

This of course, is the maximum rate at which recharge water could actually be allowed to enter the saturated portion of the aquifer. For that we conclude the increase of recharge of the rainfall as show Figure (1.9) especially for the countries with tropical rainforests as our case this leads to increasing in the water table then unpredicted landslide.

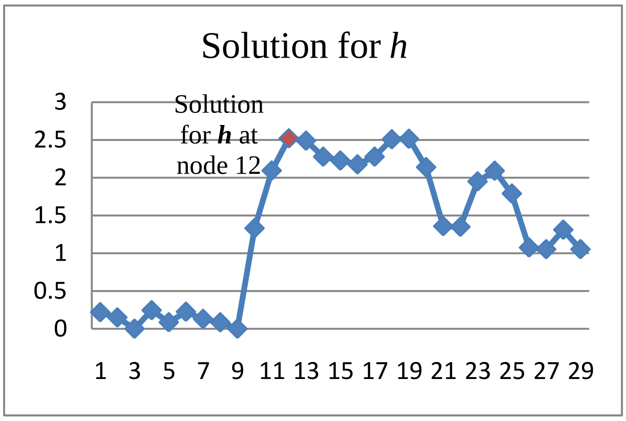


Figure 1.7: the values of the h at each node in solution boundary domain

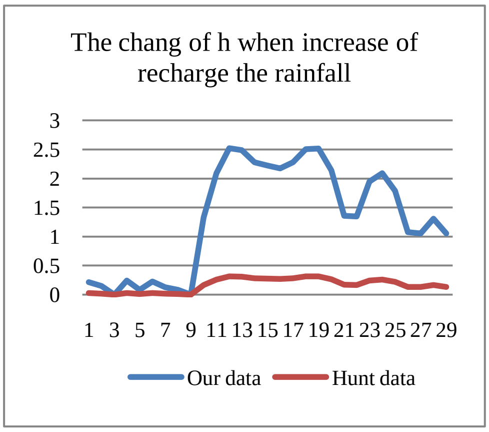


Figure 1.8: The change of h when increase of recharge of the rainfall

6. CONCLUSION

Groundwater flow models have a very long history and come in many diverse forms. Early flow models were based primarily on the finite difference method of the approximation of governing field equations. Finite difference models, both simple in their concept and computationally efficient, found broad acceptance by the general groundwater community.

Each model should be constructed to answer specific questions. Indeed, the detail and accuracy of a model depends on the question it is designed to answer. Our model is designed to predict the amount of water in a particular solution boundary domain in a slope that possibly can collapse. The answers generated using mathematical models are dependent on the quality and the quantity of the field data available that is used to define the input parameters and boundary conditions. Collection of this field data is the most difficult, the most expensive, and the most time consuming part of modelling, as mentioned in chapter two. The success of any modelling largely depends upon the availability and reliability of measured or recorded data required for the study. Usually, in most modelling projects, the time and effort spent on the pre-processing and post-processing of data far exceeds the time spent on other project activities. The reliability of predictions from groundwater models depends on how well the model approximates the field situation. Inevitably, simple assumptions must be made in order to construct a model because the field situation is too complex to be simulated exactly. To deal with more realistic situations like ours, we have to solve the mathematical models approximately using a numerical technique. From many studies in groundwater modelling we found the most easily understood and most applicable method in obtaining numerical solutions to steady groundwater flow to be the finite difference method. Therefore, the finite difference method is our choice to obtain the algebraic approximation equations for the solution boundary domain we have established. When we have to obtain two kinds of algebraic approximation equations, the first one for the nodes lies upon or beside the boundary while the second one is the partial differential equation for the interior nodes. After obtaining the algebraic equations for each node we can solve the equations simultaneously to obtain the unknown value of the dependent variables at all nodes.

The dramatic advances in the efficiency of digital computers during this past decade have provided hydrologists with a powerful tool for numerical modelling of groundwater systems. We used the computer program in Matlab 7.0 to solve a large number of algebraic equations by iterative techniques as in the Gauss Seidel method. The result of the program is the groundwater peizomtric head. This helps us to predict the amount of water in ground surfs or in the aquifer. It also gives us a clear idea of the effect of a recharge of rainfall and how much it will create a specific maximum for h (piezometric head). Additionally, we obtain the value of R (rainfall recharge) that will give a maximum water level rise of 1 meter. The model can be used to predict the acceleration of movement for landslide.

REFERENCES

Baharum, A. (1997). "Prototype *Modelling of Landslide Prevention*", Operational Research Society Annual Conference, U K. 3-5 September 1997.

Biavati, G. J., W. Godt, et al. (2006). "*Drainage effects on the transient, near -surface hydrologic response of a steep hill slope to rainfall: implications for slope stability, Edmonds, Washington, USA* ", Natural Hazards and Earth System Sciences.

Corominas, J., Moya. J, et al. (2005). "*Prediction of ground displacements and velocities from groundwater level changes at the Vallcebre landslide (Eastern Pyrenees, Spain)*." Landslides , Springer 2(2): 83-96.

Hunt, B. (1983). "*Mathematical Analysis of Groundwater Recourses*". Butterworth, Butterworth &CO.

Koay, P. S, L. Habibah, Baharum. A. et al. (2008). "*The preliminary study on landslide prediction model in Malaysia*", ICL conference, Tokyo, Japan.

Ghosh, N. C. and Sharma, K. D. (2006). "*Groundwater Modelling and Management*". New Delhi, Capital Publishing Company

Pinder, G. F. (2002). "*Groundwater modelling using geographical information systems*", Wiley, New York, USA.

Wang, H. F. and M. P. Anderson (1995). "*Introducing to groundwater modelling. Finite difference and finite element methods*." Academic Press, 6277 Sea Harbor Dr., Orlando, Florida 32887 USA, 1995, 237.

WEBSITE: Landslide groundwater: http://www.wgwa.org/articles/landslidesandgroundwater.pdf

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 494–505.

TESTS OF CAUSALITY BETWEEN TWO INFINITE-ORDER AUTOREGRESSIVE SERIES

Chafik Bouhaddioui and Jean-Marie Dufour
Department of Statistics, United Arab Emirates University, Al Ain, UAE.
E-mail: ChafikB@uaeu.ac.ae

ABSTRACT

We propose a test for non-causality at various horizons for infinite-order vector autoregressive. We first introduce multiple horizon infinite-order vector autoregressions which can be approximated by a multiple horizon finite-order vector autoregressions. The order is assumed to increase with the sample size. Under some regularity conditions, we study the estimation of parameters obtained from the approximation of the infinite-order autoregression by a finite-order autoregression. The test can be implemented simply through linear regression methods and do not involve the use of artificial simulations. The asymptotic distribution of the new test statistic is derived under the hypothesis of non-causality at various horizons. An asymptotic power of the test will be studied. Bootstrap procedures are also considered. The methods are applied to a VAR model of the US economy.

1. INTRODUCTION

The concept of causality introduced by Wiener (1956) and Granger (1969) is now a basic notion for studying dynamic relationships between time series. The literature on this topic is considerable; see, for example, the reviews of Pierce and Haugh (1977), Newbold (1982), Geweke (1984), Lütkepohl (2005) and Gourieroux and Monfort (1997)[chapter 10]. The original definition of Granger (1969), which is used or adapted by most authors on this topic, refers to the predictability of a variable $X(t)$, where t is an integer, from its own past, the one of another variable $Y(t)$ and possibly a vector $Z(t)$ of auxiliary variables, *one period ahead*: more precisely, we say that Y causes X in the sense of Granger if the observation of Y up to time t ($Y(\tau):\tau \le t$) can help one to predict $X(t+1)$ when the corresponding observations on X and Z are available ($X(\tau), Z(\tau):\tau \le t$); a more formal definition will be given below.

Recently, however, Lütkepohl (1993) and Dufour and Renault (1998) have noted that, for multivariate models where a vector of auxiliary variables Z is used in addition to the variables of interest X and Y, it is possible that Y does not cause X in this sense, but can still help to predict X *several periods ahead*; on this issue, see also Sims (1980), Renault, Sekkat and Szafarz (1998) and Giles (2002). For example, the values $Y(\tau)$ up to time t may help to predict $X(t+2)$, even though they are useless to predict $X(t+1)$. This is due to the fact that Y may help to predict Z one period ahead, which in turn has an effect on X at a subsequent period. It is clear that studying such indirect effects can have a great interest for analyzing the relationships between time series. In particular, one can distinguish in this way properties of ``short-run (non-)causality" and ``long-run (non-)causality".

In that case, Dufour, Pelletier and Renault (2006) studied the problem of testing non-causality at various horizons as defined in Dufour and Renault (1998) for finite-order vector autoregressive (VAR) models. However, it is usually known that a finite-order VAR process is a rough approximation to the true data generation process (DGP) of a given multivariate time series. Different tests were developed to study causality and orthogonality of two infinite-order autoregressive vector series, see Bouhaddioui and Roy (2006), Bouhaddioui and Dufour (2008, 2009b).

In this paper, we develop a test of non-causality at various horizons as defined in Dufour and Renault (1998) for infinite-order autoregressive (VAR(∞)) series. For a given sample size N, the true VAR(∞) model is approximated by a finite order VAR(p) model where the order p is a function of the sample size N that tends, at some rate, to infinity as N increases.

In such models, the non-causality restriction at horizon one takes the form of the relatively simple zero restrictions on the coefficients of the VAR [see Boudjellaba, Dufour and Roy (1992) and Dufour and Renault (1998)]. However, as cited in Dufour, Pelletier and Renault (2006), non-causality restrictions at higher horizons (greater than or equal to 2) are generally nonlinear, taking the form of zero restrictions on multilinear forms in the coefficients of the VAR. When applying standard test statistics such as Wald-type test criteria, such forms can easily lead to asymptotically singular covariance matrices, so that standard asymptotic theory would not apply to such statistics. Further, calculation of the relevant covariance matrices (which involve the derivatives of potentially large numbers of restrictions) can become quite awkward.

Consequently, we propose simple tests for non-causality restrictions at various horizons [as defined in Dufour and Renault (1998)] which can be implemented only through linear regression methods and do not involve the use of artificial simulations [e.g., as in Lütkepohl-Burda (1997)]. This will be done, in particular, by considering multiple horizon infinite-order vector autoregressions [called *(∞,h)-autoregressions*] which can be approximated by a multiple horizon finite-order vector autoregressions [called *(p,h)-autoregressions*]. The order p obeys to some constraints and is a function of the sample size. Under some regular conditions, the parameters of interest can be estimated by usual linear methods.

In section 2, we describe the model considered and introduce the notion of infinite-order autoregression at horizon h [or (∞, h)-autoregression] which will be the basis of our method. We study the estimation of parameters obtained from the approximation of the (∞, h)-autoregression by a finite (p, h)-autoregression. In section 3, we study the testing of non-causality at various horizons for infinite-order stationary processes. In section 4, we study the asymptotic power of the test under a specific local alternatives. Finally, in section 5, we conduct a small Monte Carlo experiment in order to study the exact level and power of the test for finite samples.

2. MULTIPLE HORIZONS INFINITE-ORDER AUTOREGRESSIONS

Consider a m -dimensional infinite-order autoregressive process of the form

$$\mathbf{X}_t = \boldsymbol{\mu}_t + \sum_{k=1}^{\infty} \boldsymbol{\pi}_k \mathbf{X}_{t-k} + \mathbf{a}_t \,, \;\; t \in \mathbf{Z}, \tag{1}$$

where $\mathbf{X}_t = (X_{1t}, X_{2t}, \ldots, X_{mt})'$ is a $m \times 1$ random vector, $\boldsymbol{\mu}_t$ is a deterministic trend, $\sum_{k=1}^{\infty} \|\boldsymbol{\pi}_k\| < \infty$ and $\|.\|$ is the Euclidean matrix norm defined by $\|A\| = tr(\mathbf{A}'\mathbf{A})$. We suppose that $\mathsf{E}[\mathbf{a}_s \mathbf{a}_t'] = \delta_{ts} \boldsymbol{\Sigma}$, $\forall t, s \in \mathsf{Z}$ with $\delta_{ts} = 1$ if $t = s$ and 0 elsewhere and $\boldsymbol{\Sigma}$ is positive definite matrix. The innovation process $\{\mathbf{a}_t\}$ satisfies the following assumption:

Assumption 1: The m-dimensional strong white noise $\mathbf{a} = \{\mathbf{a}_t = (a_{1t}, \ldots, a_{mt})'\}$ is such that $\mathsf{E}(\mathbf{a}) = 0$, its covariance regular matrix $\boldsymbol{\Sigma}$ and finite fourth moments. This usual representation, 1, is an autoregression at horizon 1. At time $t+h$, where h is an integer, we can write

$$\mathbf{X}_{t+h} = \boldsymbol{\mu}_t^{(h)} + \sum_{k=1}^{\infty} \boldsymbol{\pi}_k^{(h)} \mathbf{X}_{t+h-k} + \sum_{j=0}^{h-1} \boldsymbol{\psi}_j \mathbf{a}_{t+h-j}, \; t \in \mathbf{Z},$$

where $\boldsymbol{\psi}_0 = \mathsf{I}_m$ is the $(m \times m)$ identity matrix. The appropriate formulas for the coefficients $\boldsymbol{\pi}_k^{(h)}$, $\boldsymbol{\mu}_t^{(h)}$ and $\boldsymbol{\psi}_j$ are given in Dufour and Renault (1998) and Dufour, Pelletier and Renault (2006). We can also write that equation in the following way

$$\mathbf{X}_{t+h}' = \boldsymbol{\mu}_t^{(h)'} + \mathbf{X}(t)' \boldsymbol{\pi}^{(h)} + \mathbf{u}_{t+h}', \; t \in \mathbf{Z}, \tag{2}$$

where $\mathbf{X}(t)' = [\mathbf{X}_t', \mathbf{X}_{t-1}', \ldots]$, $\boldsymbol{\pi}^{(h)} = [\boldsymbol{\pi}_1^{(h)}, \boldsymbol{\pi}_2^{(h)}, \ldots]'$ and $\mathbf{u}_{t+h}' = \sum_{j=0}^{h-1} \boldsymbol{\psi}_j \mathbf{a}_{t+h-j}$. We call 2 an "*infinite-order autoregression at horizon h*". We suppose that the deterministic part of such autoregression is a linear function of a finite-dimensional parameter vector, *i.e.*

$$\boldsymbol{\mu}_t^{(h)} = \gamma(h) D_t^{(h)} \tag{3}$$

where $\gamma(h)$ is a $m \times n$ coefficient vector and $D_t^{(h)}$ is a $n \times 1$ vector of deterministic regressors. If $\boldsymbol{\mu}_t$ is a constant vector, then $\boldsymbol{\mu}_t^{(h)}$ is also a constant vector which can be denoted by $\boldsymbol{\mu}_t^{(h)} = \mu_h$. Based on realization $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_T$ of length T, we fit an autoregression model of order p, whose coefficients are denoted by $\boldsymbol{\pi}_{1,p}^{(h)}, \ldots, \boldsymbol{\pi}_{p,p}^{(h)}$ and we can write

$$\boldsymbol{\pi}^{(h)}(p) = \left[\boldsymbol{\pi}_{1,p}^{(h)}, \ldots, \boldsymbol{\pi}_{p,p}^{(h)}\right]. \tag{4}$$

If we define by

$$\boldsymbol{\Pi}^{(h)}(p) = \begin{bmatrix} \gamma(h)' \\ \boldsymbol{\pi}^{(h)}(p) \end{bmatrix} = \left[\Pi_1^{(h)}(p), \Pi_2^{(h)}(p), \ldots, \Pi_p^{(h)}(p)\right]$$

the corresponding least square estimator $\hat{\boldsymbol{\Pi}}^{(h)}(p)$ is given by

$$\hat{\boldsymbol{\Pi}}^{(h)}(p) = \left[\overline{\mathbf{X}}_p(h)' \overline{\mathbf{X}}_p(h)\right]^{-1} \overline{\mathbf{X}}_p(h)' \mathbf{x}_h^{(h)} \tag{5}$$

where

$$\mathbf{x}_h^{(k)} = \begin{bmatrix} \mathbf{X}_{0+h}^{'} \\ \mathbf{X}_{1+h}^{'} \\ \vdots \\ \mathbf{X}_{T-k+h}^{'} \end{bmatrix}, \ \overline{\mathbf{X}}_p(k) = \begin{bmatrix} \mathbf{X}(0,p)^{'} \\ \mathbf{X}(1,p)^{'} \\ \vdots \\ \mathbf{X}(T-k,p)^{'} \end{bmatrix}_{(T-k+1)\times(n+mp)},$$

$$\mathbf{X}(t,p) = \begin{bmatrix} \mathbf{D}_t^{(h)^{'}} \\ \mathbf{X}_t(p) \end{bmatrix}_{(n+mp)\times 1}, \qquad \mathbf{X}_t(p)^{'} = \left[\mathbf{X}_t^{'}, \mathbf{X}_{t-1}^{'}, \ldots, \mathbf{X}_{t-p+1}^{'}\right].$$

In the sequel, we state the following assumptions.

Assumption 2: p is a function of T such that

$$p \to \infty \ , \ \frac{p^3}{T} \to 0 \ and \ \sqrt{T} \sum_{j=p+1}^{\infty} \left\|\boldsymbol{\pi}_j\right\| < \infty \ as \ T \to \infty. \tag{6}$$

To derive the asymptotic properties of $\hat{\boldsymbol{\pi}}^{(h)}(p)$, we can also express the OLS estimator as

$$\hat{\mathbf{\Pi}}^{(h)}(p) = \hat{\mathbf{\Gamma}}_{1,p}^{'} \hat{\mathbf{\Gamma}}_p^{-1}, \tag{7}$$

where

$$\hat{\mathbf{\Gamma}}_{1,p} = (T-p)^{-1} \sum\nolimits_{t=p+1}^{T-h} \mathbf{X}_t(p) \mathbf{X}_t^{'}$$

and

$$\hat{\mathbf{\Gamma}}_p = (T-p)^{-1} \sum\nolimits_{t=p+1}^{T-h} \mathbf{X}_t(p) \mathbf{X}_t(p)^{'}.$$

The consistency rate of $\hat{\mathbf{\Pi}}^{(h)}(p)$ is given in the following proposition. Let $\{\mathbf{X}_t\}$ be a process given by (2) and satisfying the Assumption 2. Under the Assumption 2, the estimator $\hat{\mathbf{\Pi}}^{(h)}(p)$ defined by (7) is such that

$$\left\|\hat{\mathbf{\Pi}}^{(h)}(p) - \mathbf{\Pi}^{(h)}(p)\right\| = O_p\left(\frac{p^{1/2}}{T^{1/2}}\right).$$

Now, if we suppose that the process $\mathbf{X_t}$ is second-order stationary, we define the autocovariance matrix at lag k by $\mathbf{\Gamma}(k) = \mathsf{E}(\mathbf{X}_t \mathbf{X}_{t+k}^{'})$, $k \in \mathsf{Z}$.

3. TESTS OF CAUSALITY FOR AN INFINITE-ORDER STATIONARY h-AUTOREGRESSIONS

Let us now consider the following hypothesis

$$\mathsf{H}_0(h): \mathbf{R}\Pi_i^{(h)} = 0, \ \ i = 1, 2, \ldots$$

where $\mathbf{R}$ is a $q \times (n+mp)$ matrix of rank q. In particular, as mentioned in Dufour, Pelletier and Renault (2006), if we wish to test the hypothesis that x_{jt} does not cause x_{it} at horizon h, the restriction would take the form

$$\mathsf{H}_{j½i}^{(h)}: \pi_{ijk}^{(h)} = 0, \ \ k = 1,2,\ldots,$$

where $\boldsymbol{\pi}_k^{(h)} = \left[\pi_{ijk}^{(h)}\right]_{i,j=1,\ldots,m}$, $k = 1,2,\ldots$. From the model (2) and from the equation (7), we have

$$vec(\hat{\mathbf{\Pi}}^{(h)}(p)) = vec\left\{\left[(T-p)^{-1}\sum_{t=p+1}^{T-h}\mathbf{X}_t(p)\mathbf{X}_t^{'}\right]\left[(T-p)^{-1}\sum_{t=p+1}^{T-h}\mathbf{X}_t(p)\mathbf{X}_t(p)^{'}\right]\right\}.$$

If we denote by

$$\begin{aligned}\mathbf{W}_{1T} &= vec[(T-p)^{-1}\sum_{t=p+1}^{T-h}\mathbf{u}_{t+h}\mathbf{X}_t^{'}(p)]\\ \mathbf{W}_{2T} &= vec[(T-p)^{-1}\sum_{t=p+1}^{T-h}\{\mathbf{u}_{t+h}(p)-\mathbf{u}_{t+h}\}\mathbf{X}_t^{'}(p)],\end{aligned}$$

where $\mathbf{u}_{t+h}(p) = \mathbf{X}_{t+h}^{'} - \mathbf{X}_t^{'}(p)\mathbf{\Pi}^{(h)}(p)$, and under the hypothesis H_0, we have

$$(\mathsf{I}_p \otimes \mathbf{R})vec(\hat{\mathbf{\Pi}}^{(h)}(p)) = (\mathsf{I}_p \otimes \mathbf{R})(\hat{\mathbf{\Gamma}}_p^{-1} \otimes \mathsf{I}_m)(\mathbf{W}_{1T} + \mathbf{W}_{2T}),$$

which can be also decomposed as

$$\begin{aligned}(\mathsf{I}_p \otimes \mathbf{R})vec(\hat{\mathbf{\Pi}}^{(h)}(p)) &= (\mathsf{I}_p \otimes \mathbf{R})(\mathbf{\Gamma}_p^{-1} \otimes \mathsf{I}_m)\mathbf{W}_{1T} + (\mathsf{I}_p \otimes \mathbf{R})(\mathbf{\Gamma}_p^{-1} \otimes \mathsf{I}_m)\mathbf{W}_{2T}\\ &\quad + (\mathsf{I}_p \otimes \mathbf{R})\{(\hat{\mathbf{\Gamma}}_p^{-1} - \mathbf{\Gamma}_p^{-1}) \otimes \mathsf{I}_m\}(\mathbf{W}_{1T} + \mathbf{W}_{2T})\\ &= \mathbf{A}(p) + \mathbf{B}_1(p) + \mathbf{B}_2(p)\\ &= \mathbf{A}(p) + \mathbf{B}(p).\end{aligned} \tag{8}$$

Also, we denote by $\mathbf{\Sigma}_u^{(h)}$ the covariance matrix of $\{\mathbf{u}_t^{(h)}\}$ and $\hat{\mathbf{\Sigma}}_u^{(h)}$ is its estimator such that

$$\sqrt{T}(\hat{\mathbf{\Sigma}}_u^{(h)} - \mathbf{\Sigma}_u^{(h)}) \xrightarrow{L} N(0, \mathbf{\Omega}_h) \tag{9}$$

where $\xrightarrow{L}$ denotes convergence in law and $\mathbf{\Omega}_h$ is a given matrix. The test statistic is based on the following quadratic form

$$\mathsf{T}(\hat{\mathbf{u}}^{(h)}, \hat{\mathbf{\Sigma}}_u^{(h)}) = T\hat{\boldsymbol{\pi}}^{(h)}(p)'(\mathsf{I}_p \otimes \mathbf{R})'\hat{\mathbf{\Delta}}_{p,h}^{-1}(\mathsf{I}_p \otimes \mathbf{R})\hat{\boldsymbol{\pi}}^{(h)}(p) \tag{10}$$

where $\hat{\boldsymbol{\pi}}^{(h)}(p) = vec(\hat{\mathbf{\Pi}}^{(h)}(p))$ and $\hat{\mathbf{\Delta}}_{p,h} = (\mathsf{I}_p \otimes \mathbf{R})(\hat{\mathbf{\Gamma}}_p^{-1} \otimes \hat{\mathbf{\Sigma}}_u^{(h)})(\mathsf{I}_p \otimes \mathbf{R}')$. Now, we consider the following statistic test which is a standardized version of $\mathsf{T}(\hat{\mathbf{u}}^{(h)}, \hat{\mathbf{\Sigma}}_u^{(h)})$ given by

$$\mathsf{Q}_T^{(h)} = \frac{\mathsf{T}(\hat{\mathbf{u}}^{(h)}, \hat{\mathbf{\Sigma}}_u^{(h)}) - pq}{\sqrt{2pq}} \tag{11}$$

The main result to test the non-causality is stated in the following theorem.

Theorem 1: Let $\{\mathbf{X}_t\}$ be a process given by (2) and satisfying the Assumption 2. If the order p satisfies the Assumption 2 and under any hypothesis of the form $\mathsf{H}_0(h)$, we have

$$\mathsf{Q}_T^{(h)} \xrightarrow{L} N(0,1).$$

Proof. Using the decomposition (8), we start by defining the pseudo-statistic

$$\mathsf{T}_T^{(h)} = T\mathbf{A}(p)'\mathbf{\Delta}_{p,h}^{-1\,'}\mathbf{A}(p)$$

where $\mathbf{\Delta}_{p,h} = (\mathsf{I}_p \otimes \mathbf{R})(\mathbf{\Gamma}_p^{-1} \otimes \mathbf{\Sigma}_u^{(h)})(\mathsf{I}_p \otimes \mathbf{R}')$. Thus, we can write

$$\mathsf{Q}_T^{(h)} = \frac{\mathsf{T}_T^{(h)} - pq}{\sqrt{2pq}} + \frac{T\mathsf{Z}_p^{(h)}}{\sqrt{2pq}}, \tag{12}$$

where

$$\mathsf{Z}_p^{(h)} = \mathbf{A}(p)'\{\hat{\mathbf{\Delta}}_{p,h}^{-1} - \mathbf{\Delta}_{p,h}^{-1}\}\mathbf{A}(p) + 2\mathbf{A}(p)'\hat{\mathbf{\Delta}}_{p,h}^{-1}\mathbf{B}(p) + \mathbf{B}(p)'\hat{\mathbf{\Delta}}_{p,h}^{-1}\mathbf{B}(p).$$ The

asymptotic distribution of $\mathsf{Q}_T^{(h)}$ follows from the next two propositions. The proof is given in the appendix of a technical report available from the author (Bouhaddioui and Dufour (2006b)).

Proposition 1: Under the assumptions of Theorem 3, we have that

$$\mathsf{Z}_p^{(h)} = o_p\left(\frac{p^{1/2}}{T}\right).$$

Proposition 2: Under the assumptions of Theorem 3, we have that $\dfrac{\mathsf{T}_T^{(h)} - pq}{\sqrt{2pq}} \xrightarrow{L} N(0,1).$

4. LOCAL ASYMPTOTIC POWER

Let consider the series of local alternative hypothesis

$$H_{\varphi_p(T)}(h): \mathbf{R}\Pi_i^{(h)} = \phi_i(T),\ i = 1, 2, \ldots,$$

where $\varphi_p(T) = vec(\phi_1(T), \ldots, \phi_p(T)) = T^{\kappa} vec(\phi_1, \ldots, \phi_p) = \varphi_p$ and $\|\phi_i(T)\| \to 0$ as $T \to \infty$. We can give the following theorem which characterizes the local power behavior of the test for the particular class of local alternatives designed by $H_{\varphi_p(T)}(h)$. The proof is given in the appendix of a technical report available from the author (Bouhaddioui and Dufour (2006b)).

Let $\{\mathbf{X}_t\}$ be a process given by (2) and satisfying the Assumption 2. If the order p satisfies the Assumption 2 and under the local alternative hypothesis of the form $H_{\varphi_p(T)}(h)$ with $\varphi_p(T) = T^{-1/2}\varphi_p$, we have

$$\frac{\mathrm{Q}_T^{(h)} - \boldsymbol{\mu}_p^{(h)}}{\boldsymbol{\sigma}_p^{(h)}} \xrightarrow{L} N(0,1)$$

where

$$\boldsymbol{\mu}_{\varphi_p}^{(h)} = \frac{\varphi_p^{'} \boldsymbol{\Delta}_{p,h}^{-1} \varphi_p}{\sqrt{2pq}} \text{ and } \boldsymbol{\sigma}_p^{(h)} = \sqrt{1 + \boldsymbol{\mu}_p^{(h)}}.$$

5. EMPIRICAL ILLUSTRATION AND SIMULATION

In this section, we present an application of these causality tests at various horizons to macroeconomic time series. The data set considered is the one used Bernanke and Mihov (1998) and Dufour, Pelletier and Renault (2006) in order to study United States monetary policy. The data set considered consists of monthly observations on nonborrowed reserves (NBR, also denoted ω_1), the federal funds rate (r, ω_2), the GDP deflator (P, ω_3) and real GDP (GDP, ω_4). The monthly data on GDP and GDP deflator were constructed by state space methods from quarterly observations [for more details, see Bernanke and Mihov (1998)]. The sample goes from January 1965 to December 1996 for a total of 384 observations. In what follows, all variables were first transformed by a logarithmic transformation.

Before performing the causality tests, we must specify the order of the VAR model. First, in order to get apparently stationary time series, all variables were transformed by taking first differences of their logarithms. In particular, for the federal funds rate, this helped to mitigate the effects of possible break in the series in the years 1979-1981. Once the data is made stationary, we use a nonparametric approach for the estimation and Akaike's information criterion to specify the orders of the long VAR(k) models. Using Akaike's criterion for the unconstrained VAR model, which corresponds to four variables, we observe that it is minimized at $k = 16$.

Vector autoregressions of order p at horizon h were estimated as described in section 2 and the matrix $\hat{V}_{ip}^{NW}$, required to obtain covariance matrices, were computed. On looking at the values of the test statistics $Q_T^{(h)}$ and their corresponding p-values at various horizons it quickly becomes evident that the $N(0,1)$ asymptotic approximation of the statistic $Q_T^{(h)}$ is very poor. Now, let consider the GDP as the VAR(16) estimated with our data in first difference but we impose that some coefficients are zero such that the federal funds rate does not cause GDP up to horizon h and then we test the $r \overset{h}{\not\to} GDP$ hypothesis. The constraints of non-causality from j to i up to horizon h that we impose are

$$\hat{\pi}_{ijl} = 0,\ 1 \le l \le p, \tag{13}$$

$$\hat{\pi}_{ikl} = 0,\ 1 \le l \le h, 1 \le k \le m. \tag{14}$$

Rejection frequencies for this case are given in Table 1.
In light of these results we computed the p-values by doing a parametric bootstrap, i.e. doing an asymptotic Monte Carlo test based on a consistent point estimate [see Dufour (2006)]. The procedure to test the hypothesis $\omega_j \overset{h}{\not\to} \omega_i \mid I_{(j)}$ is the following:

1. An unrestricted VAR(p) model is fitted for the horizon one, yielding the estimates $\hat{\Pi}^{(1)}$ and $\hat{\Sigma}$ for $\Pi^{(1)}$ and Σ. The autoregressive order was obtained by minimizing the AIC criterion for $p \le P$, where P was fixed to 24
2. An unrestricted (p, h)-autoregression is fitted by least squares, yielding the estimate $\hat{\Pi}^{(h)}$ of $\Pi^{(h)}$
3. The test statistic $Q_T^{(h)}$ for testing non-causality at the horizon h from ω_j to ω_i is computed. We denoted by $Q_{T, j \not\to i}^{(h)}(0)$ the test statistic based on the actual data
4. N simulated samples from (2) are drawn by Monte Carlo methods, using $\Pi^{(h)} = \hat{\Pi}^{(h)}$ and $\Sigma = \hat{\Sigma}$. We impose the constraints of non-causality $\hat{\pi}_{ijk} = 0$, $k = 1, \dots, p$. Estimates of the impulse response coefficients are obtained from $\hat{\Pi}^{(1)}$ through the relations described in Eq. (4). We denote by $Q_{T, j \not\to i}^{(h)}(0)$ the test statistic for $H_{j \not\to i}^{(h)}$ based on n th simulated sample ($1 \le n \le N$).

5. The simulated p-value $\hat{p}[Q_{T, j \not\to i}^{(h)}(0)]$ is obtained, where

$$\hat{p}[x] = \{1 + \sum_{n=1}^{N} \mathrm{I}[Q_{T, j \not\to i}^{(h)}(n) - x]\} / (n+1),$$

   where $\mathrm{I}[x] = 1$ if $x \ge 0$ and $\mathrm{I}[x] = 0$ if $x < 0$.

6. The null hypothesis $H_{j \not\to i}^{(h)}$ is rejected at level α if $\hat{p}[Q_{T, j \not\to i}^{(h)}(0)] \le \alpha$.

7. Finally, for each nominal level $\alpha = 1\%, 5\%$ and 10%, we obtained from the 2000 realizations (with $N = 999$), the empirical frequencies of rejection of the null hypothesis of non-correlation. The power analysis was conducted in the similar way using the model VAR$_{\delta}$ for different values of δ.

From looking at the results in Table 1, we see that we get a much better size control by using this bootstrap procedure. The rejection frequencies over 1000 replications (with $N = 999$) are very close to the nominal size. Although the coefficients ψ_j's are functions of π_j's we do not constrain them in the bootstrap procedure because there is no direct mapping from $\pi_k^{(h)}$ to π_k and ψ_j. This certainly produces a power loss but the procedure remains valid because the ψ_j's are computed with the $\hat{\pi}_k$, which are consistent estimates of the true π_k both under the null and alternative hypothesis. To illustrate that our procedure has power for detecting departure from the null hypothesis of non-causality at a given horizon we ran the following Monte Carlo experiment. We again took a VAR(16) fitted on our data in first differences and we imposed the constraints (13)-(14) so that there was no causality from r to GDP up to horizon 12. Next the value of one coefficient previously set to zero was changed to induce causality from r to GDP at horizon 4 and higher: $\pi_3(1,3) = \theta$. As θ increases from zero to one the strength of the causality from r to GDP is higher. Under this setup, we could compute the power of our simulated test procedure to reject the null hypothesis of non-causality at a given horizon. The power curves are plotted as a function of θ for the various horizons. The level of the tests was controlled through the bootstrap procedure. In this experiment we took again $N = 999$ and we did 1000 simulations. These curves are plotted in Bouhaddioui-Dufour (2009b). As expected, the power curves are flat at around 5% for horizons one to three since the null is true for these horizons. For horizons four and up we get the expected result that power goes up as θ moves from zero to one, and the power curves gets flatter as we increase the horizon.

Now that we have shown that bootstrap procedure does have power we present causality tests at horizon one to 24 for every pair of variables in Tables 2 and 3 presented in Bouhaddioui-Dufour (2009b). For every horizon we have 12 causality tests and we group them by pairs. When we say that a given variable cause or does not cause another, it should be understood that we mean the growth rate of the variables. The p-values are computed by taking $N = 999$. Table 4 summarize the results by presenting the significant results at the 5% and 10% levels.

The first thing to notice is that we have significant causality results at short horizons for some pairs of variables while we have it at longer horizons for other pairs. This is an interesting illustration of the concept of causality at horizon h of Dufour and Renault (1998).

The instrument of the central bank, the nonborrowed reserves, cause the federal funds rate at horizon one, the prices at horizons 1, 2, 3 and 9 (10% level). It does not cause the other two variables at any horizon and except the GDP at horizon 12 and 16 (10% level) nothing is causing it. We see that the impact of variations in the nonborrowed reserves is over a very short term. Another variable, the GDP, is also causing the federal funds rates over short horizons (one to five months).

An interesting result is the causality from the federal funds rate to the GDP. Over the first few months the funds rate does not cause GDP, but from horizon 3 (up to 20) we do find significant causality. This result can easily be explained by, e.g. the theory of investment. Notice that we have following indirect causality. Nonborrowed reserves do not cause GDP directly over any horizon, but they cause the federal funds rate which in turn causes GDP. Concerning the observation that there are very few causality results for long horizons, this may reflect the fact that, for stationary processes, the coefficients of prediction formulas converge to zero as the forecast horizon increases.

Table 1: Rejection using the asymptotic distribution N(0,1) with VAR(16)

	Asymptotic						Bootstrap				
h	α=5%	α=10%	h	α=5%	α=10%	h	α=5%	α=10%	h	α=5%	α=10%
1	22.2	32.2	7	36.9	52.3	1	5.2	9.6	7	4.3	10.8
2	23.4	38.6	8	39.4	55.6	2	4.8	9.3	8	4.1	9.6
3	27.6	40.5	9	42.8	58.4	3	4.9	10.5	9	4.8	9.1
4	31.2	44.2	10	46.6	61.5	4	5.8	11.0	10	5.6	10.7
5	32.7	46.4	11	49.5	63.4	5	6.0	9.8	11	5.4	10.4
6	34.6	48.8	12	53.7	66.8	6	5.3	10.2	12	4.1	9.1

Table 2: Summary of Causality relations at various horizons for series in first difference.

h	1	2	3	4	5	6	7	8	9	10	11	12
$NBR \nrightarrow r$	**	*										
$r \nrightarrow NBR$								**		*		
$NBR \nrightarrow P$	**	**	*			**	*					
$P \nrightarrow NBR$												
	*											
$GDP \nrightarrow NBR$												
$r \nrightarrow P$	*	*										
$P \nrightarrow r$												
$r \nrightarrow GDP$			*	**	*	**	**	**	**	**	**	**
$GDP \nrightarrow r$	**	**	**	**	**	*						
$P \nrightarrow GDP$												
$GDP \nrightarrow P$					*	**	*	*	**			
h	13	14	15	16	17	18	19	20	21	22	23	24
$NBR \nrightarrow r$												
$r \nrightarrow NBR$												
$NBR \nrightarrow P$												
$P \nrightarrow NBR$												
$GDP \nrightarrow NBR$			*	*								
$r \nrightarrow P$												
$P \nrightarrow r$												
$r \nrightarrow GDP$	**	**	**	**	**	**	**	*	*			
$GDP \nrightarrow r$												
$P \nrightarrow GDP$												
$GDP \nrightarrow P$												

Note: The symbols $*$ and $**$ indicate rejection of the non-causality hypothesis at the 10% and 5% levels, respectively.

Using the results of Proposition 4.5 in Dufour and Renault (1998), we know that for the highest horizon that we have to consider is 33 since we have a VAR(16) with four time series. Causality tests for horizons 25 through 33 were also computed but are not reported. Some p-values smaller or equal to 10% are scattered over horizons 30 to 33 but no discernible pattern emerges.

REFERENCES

Bernanke B. and Mihov I., (1998). Measuring monetary policy, *The Quarterly Journal of Economics* **113**(3), 869-902.

Boudjellaba H., Dufour J. M., Roy R. (1992). Testing causality between two vectors in multivariate ARMA models, *Journal of the American Statistical Association* **87**, 1082-1090.

Bouhaddioui C. and Roy R. (2006). A generalized portmanteau test for independence of two infinite-order vector autoregressive series, *Journal of Time Series Analysis* **27**(4), 505-544.

Bouhaddioui C. and Dufour J. M. (2008). Tests for Non-Correlation of two infinite-order cointegrated vector autoregressive series, *Journal of Applied Probability and Statistics* **3 (1)**, 77-94.

Bouhaddioui C. and Dufour J. M. (2009a). Semiparametric tests of orthogonality and causality between two cointegrated series with application to Canada/US monetary interactions, *Discussion Paper*, CIRANO and CIREQ, 35 pages.

Bouhaddioui C. and Dufour J. M. (2009b). Short run and long run causality in an infinite-order autoregressive vector, *Technical Report*, CIRANO and CIREQ, 27 pages.

Dufour J. M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics, *Journal of Econometrics* **133**(2) 443-477.

Dufour J. M., Pelletier D., Renault É. (2006). Short run and long run causality in time series: Inference, *Journal of Econometrics* **132**(2), 337-362.

Dufour J. M. and Renault É. (1998). Short run and long run causality in time series: Theory, *Econometrica* **66**, 1099-1125.

Geweke J. (1984). Inference and causality in economic time series, *in* Z. Griliches, M. D. Intrilligator, eds, `Handbook of Econometrics, Volume 2', North-Holland, Amsterdam, pp. 1102-1144.

Giles J. A. (2002). Time series analysis testing for two-step Granger noncausality in trivariate VAR models, *in* A. Ullah, A. K. Wan, A. Chaturvedi, eds, `Handbook of Applied Econometrics and Statistical Inference', Marcel Dekker, New York, chapter 18, pp. 371-399.

Gouriéroux C. and Monfort A. (1997). *Time Series and Dynamic Models*, Cambridge University Press, Cambridge, U.K.

Granger C. (1969). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* **37**, 424-459.

Lewis R. and Reinsel G. C. (1985). Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis* **16**, 393-411.

Lütkepohl H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Lütkepohl H. (1993). Testing for causation between two variables in higher dimensional VAR models, *in* H. Schneeweiss, K. Zimmermann, eds, `Studies in Applied Econometrics', Springer-Verlag, Heidelberg.

Lütkepohl H. and Burda M. M. (1997). Modified Wald tests under nonregular conditions, *Journal of Econometrics* **78**, 315-332.

Newbold P. (1982). Causality testing in economics, *in* O. D. Anderson, ed., Time Series Analysis: Theory and Practice 1, North-Holland, Amsterdam.

Pierce D. A. and Haugh L. D. (1977). Causality in temporal systems: Characterizations and survey, *Journal of Econometrics* **10**, 257-259.

Renault E. , Sekkat K. , Szafarz A. (1998). Testing for spurious causality in exchange rates, *Journal of Empirical Finance* **5**, 47-66.

Sims C. (1980). Macroeconomic and reality, *Econometrica* **48**, 1-48.

Wiener N. (1956). *The Theory of Prediction*, New York, chapter 8.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 506–506.

ESTIMATING TRANSITION INTENSITY MATRICES TO DOCUMENT DISEASE PROGRESSION IN MULTI-STATE MARKOV MODELS

Mohammad Ashraf Chaudhary[1] and Elamin H. Elbasha[2]
Mail Stop UG1C-60, Merck & Co., Inc.,
351 North Sumneytown Pike,, North Wales PA19454 USA
E-mail: [1]Mohammad_Chaudhary@Merck.Com, [2]Elamin_Elbasha@Merck.Com

ABSTRACT

Multi-state Markov models are commonly used to predict the long term clinical and economic outcomes associated with different treatment strategies for a disease, and to evaluate their relative cost-effectiveness. Disease progression is commonly stratified into heath states with varying degrees of disease severity. In discrete-time Markov models, state transitions only occur at fixed times. Despite its clear advantages in terms of realistically modeling transitions as happening at any time, the multi-state Markov modeling in continuous time is not as common. The clinical outcome of different treatments over a period of time would be estimated by transition rates (or transition probabilities for discrete time Markov models) of movement of patients from one heath state to the next. If there are *k* health states, then a matrix of size $k(k\text{-}1)$ distinct transition rates needs to be estimated. For models with relatively large number of health states, estimating transition rates using the prevalence of patients in different heath states at different time points becomes prohibitively cumbersome. (Kalbfleisch and Lawless 1985) developed a general method for evaluating the likelihood for a general multi-state model in continuous time applicable to any form of transition matrix. The *msm* package for *R* (Jackson 2009) implements this method to fit multi-state model to the longitudinal data consisting of observations of the process at arbitrary times. We illustrate the application of this method by estimating the transition matrices of a continuous-time HIV Markov model, and demonstrate their use to predict long term clinical and economic outcomes and evaluate the cost-effectiveness of raltegravir for the treatment of anti-retroviral therapy experienced HIV/AIDS patients.(Chaudhary, Moreno et al. 2009; Elbasha, Szucs et al. 2009).

REFERENCES

Chaudhary, M. A., S. Moreno, et al. (2009). "Cost-effectiveness analysis of raltegravir in treatment-experienced HIV type 1-infected patients in Spain." AIDS Res Hum Retroviruses **25**(7): 679-89.

Elbasha, E., T. Szucs, et al. (2009). "Cost-effectiveness of raltegravir in antiretroviral treatment-experienced HIV-1-infected patients in Switzerland." HIV Clinical Trials **(in press)**.

Jackson, C. (2009). Multi-state Modeling with R: the msm package, Version 0.7.6 (December 10, 2007):, The R Foundation for Statistical Computing.

Kalbfleisch, J. D. and J. F. Lawless (1985). "The analysis of panel data under a Markov assumption." Journal of the American Statistical Association **80**: 863-871.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 507–536.

ANALYTIC INFERENCE OF COMPLEX SURVEY DATA UNDER INFORMATIVE PROBABILITY SAMPLING

Abdulhakeem Abdulhay Eideh
Department of Mathematics
Faculty of Science and Technology
Al-Quds University, Abu-Dies Campus – Palestine
P.O. Box 20002, Jerusalem
E-mail: msabdul@science.alquds.edu

ABSTRACT

Most of the empirical research based on data from social science and health surveys requires special analytic inference for populations that have a complex structure, and when data are collected from complex sampling designs, standard analysis of complex survey data often fails to account for the complex nature of the sampling design such as the use of unequal probability of selection, clustering, post-stratification and other kinds of weighting used for the treatment of non-response. The effect of sampling on the analysis is because the models in use typically do not incorporate all the design variables determining the sample selection, either because there may be too many of them or because they are not of substantive interest. However if the sampling design is informative in the sense that the outcome variable is correlated with the design variables not included in the model, even after conditioning on the model covariates, standard analysis of the population model parameters can be severely biased, leading possibly to false inference. In the literature there are three methods dealing with the effect of unequal probability of selection and informative sampling, which are: probability weighting, pseudo likelihood estimation, and the third one is the method based on the concept of using the sample distribution (the distribution of the observed outcomes given the selected sample) for inference under informative sampling. In this paper we discuss theses methods, and we introduce the Kullback-Leibler information measure as a measure of distance between the sample and population distributions.

Keywords: Informative sampling; Probability weighting; Pseudo Likelihood; Sample likelihood; Kullback-Leibler.

1. INTRODUCTION

Survey data may be viewed as the outcome of two processes: the process that generates the values of a random variable for units in a finite population, often referred to as the superpopulation model, and the process of selecting the sample units from the finite population values, known as the sample selection mechanism. Most of the empirical research based on data from social science and health surveys requires special analytic inference for populations that have a complex structure, and when data are collected from complex sampling designs. Standard analysis of complex survey data often fails to account for the complex nature of the sampling design such as the use of unequal probability of selection, clustering, post-stratification and other kinds of weighting used for the treatment of non-response. The effect of sampling on the analysis is due to the fact that the models in use typically do not incorporate all the design variables determining the sample selection, either because there may be too many of them or because they are not of substantive interest. However if the sampling design is informative in the sense that the outcome variable is correlated with the design variables not included in the model, even after conditioning on the model covariates, standard analysis of the population model parameters can be severely biased, leading possibly to false inference.

In the literature three methods dealing with the effect of unequal probability of selection and informative sampling are discussed. These methods are: probability weighting, pseudo likelihood estimation and the most recent one is the method based on the concept of using the sample distribution for inference under informative sampling.

This paper contains a summary of previous results, related new results, examples and illustrations. Sections 2 and 3 define the methods of probability weighting and pseudo likelihood estimation. Section 4 defines the sample distribution and its relationship to the population distribution. In Sections 5 and 6 we treat the sample likelihood function and the sample distribution is derived under different models for sample selection probabilities. Section 7 gives results on the relationships between moments under the sample and the population distributions. Section 8 outlines estimation methods based on the sample likelihood. Finally in Section 9 we propose the Kullback- Leibler information as a distance measure between the sample and the population distributions.

2. PROBABILITY WEIGHTING

Let $U = \{1,...,N\}$ denote a finite population consisting of N units. Let y be the target variable of interest and let y_i be the value of y for the ith population unit. At this stage the values y_i are assumed to be fixed unknown quantities. Suppose that an estimate is needed for the population total of y, $T = \sum_{i=1}^{N} y_i$. A probability sample s is drawn from U according to a specified sampling design. The sample size is denoted by n. The sampling design induces inclusion probabilities for the different units of U. Let $\pi_i = \Pr(i \in s)$ be the first order inclusion probability of the ith population unit. The Horvitz-Thompson (1952) estimator or probability-weighted estimator of the population total of y, $T = \sum_{i=1}^{N} y_i$ is given by: $\hat{T} = \sum_{i=1}^{n} w_i y_i$, where $w_i = 1/\pi_i$ is the sampling weight of unit $i \in U$. That is we weight each sample observation i by the sampling weight, w_i. This estimator is design-unbiased, that is $E_D(\hat{T}) = \sum_{i=1}^{N} y_i$, where E_D denotes the expectation under repeated sampling. The attractive feature of the probability-weighted estimators is that they are unbiased or approximately unbiased and consistent with respect to the randomization distribution induced by the random selection of the sample, irrespective of the distribution of the y-values in the population. Another advantage of these estimators is that for the majority of the designs, the randomization variance associated with the use of these estimators can be estimated relatively simply utilizing available techniques for finite population sampling. A major drawback of probability-weighted estimators is that their variance is generally higher than the variance of the corresponding unweighted estimators, particularly for small samples and large variation of the weights.

3. PSEUDO LIKELIHOOD ESTIMATION

We now consider the population values $y_1,...,y_N$ as random variables, which are independent realizations from a distribution with probability density function (pdf) $f_p(y_i \mid \theta)$, indexed by a vector parameter θ. We now consider the estimation of the superpopulation parameter, θ, rather than the prediction of the (random variable) total T. Let $l(\theta \mid y_1,...,y_N)=\sum_{i=1}^{N}\log f_p(y_i \mid \theta)$ be the census log-likelihood that would be obtained in the case of a census. The maximum likelihood estimator of θ solves the population likelihood equations, which in our case are:

$$U(\theta)=\sum_{i=1}^{N}\frac{\partial\left(\log f_p(y_i \mid \theta)\right)}{\partial\theta}=0 \tag{1}$$

Following Binder (1983) the pseudo maximum likelihood (PML) estimator is defined as the solution of: $\hat{U}(\theta)=0$, where $\hat{U}(\theta)$ is a sample estimator of the estimating function $U(\theta)$. For example the probability-weighted estimator of $U(\theta)$ is:

$$\hat{U}_w(\theta)=\sum_{i\in s}w_i\frac{\partial\left(\log f_p(y_i \mid \theta)\right)}{\partial\theta}, \text{ where } w_i=1/\pi_i \tag{2}$$

That is when the explicit form of the population likelihood is not available, we weight instead the sample likelihood and solve the weighted equations.
The use of this approach requires a full specification of the population distribution, unlike simple probability weighting. On the other hand, it permits the incorporation of the sampling weight. More detail on this approach and the properties of PML estimators, see for example Binder (1983, 1996), and Skinner (1989).

4. SAMPLE DISTRIBUTION

In recent years an alternative approach to the probability weighting method and to the pseudo likelihood approach were introduced in the literature, based on the idea of the sample distribution. In recent articles by Krieger and Pfeffermann (1992, 1997), Pfeffermann, Krieger, and Rinott (1998), Pfeffermann and Sverchkov (1999, 2003, 2005), Pfeffermann (2002) Eideh and Nathan (2006a, 2006b, 2009a, 2009b), Nathan and Eideh (2004, 2008), and Eideh (2007, 2008, 2009) the authors introduced an alternative model-based approach to analytic likelihood-based inference from complex survey data and specially dealing with the effect of informative sampling or size-biased sampling. Their basic idea is to derive or extract the distribution of the sample data by modeling the population distribution and the conditional expectation of the first order sample inclusion probabilities. Once this sample distribution is extracted, standard likelihood-based inferential methods can be used to obtain estimates of the parameters of the population model of interest. In order to describe the fundamental idea behind this approach, we assume full response. In case of nonresponse, see Eideh (2009). Let us first present notations relating to the population and sample. Following the notations in the previous sections, let

$\mathbf{x}_i = (x_{i1},...,x_{ip})'$, $i \in U$ be the values of a vector of auxiliary variables, $x_1,...,x_p$, [the prime on a vector or a matrix denotes the transpose], and $\mathbf{z} = \{z_1,...,z_N\}$ be the values of a known design variable, used for the sample selection process but not included in the working model under consideration. In what follows we consider a sampling design with selection probabilities $\pi_i = \Pr(i \in s)$ where $i = 1,...,N$. In practice, the π_i's may depend on the population values $(\mathbf{x}, \mathbf{y}, \mathbf{z})$. We express this by writing: $\pi_i = \Pr(i \in s \mid \mathbf{x}, \mathbf{y}, \mathbf{z}) = g_i(\mathbf{x}, \mathbf{y}, \mathbf{z})$, for some function g_i and all units $i \in U$. Since $\pi_1,...,\pi_N$ are defined by the realizations $(\mathbf{x}_i, y_i, \mathbf{z}_i), i = 1,...,N$, therefore they are random realizations defined on the space of possible populations. The sample s consists of the subset of U selected at random by the sampling scheme with inclusion probabilities $\pi_1,...,\pi_N$. We assume probability sampling, so that $\pi_i > 0$ for all units $i \in U$.

The sample distribution refers to the superpopulation distribution model of the sample measurements, as induced by the population model and the sample selection scheme, with the selected sample of units held fixed. Before defining the sample distribution mathematically, let us introduce the following notations: f_p denotes the pdf of the population distribution, f_s denotes the pdf of the sample distribution, $E_p(\cdot)$ denotes the mathematical expectation under the population distribution, and $E_s(\cdot)$ denotes the mathematical expectation under the sample distribution. Now assume that the population pdf depends on known values of the auxiliary variables $\mathbf{x}_i$ and on θ, so that $y_i \sim f_p(y_i \mid \mathbf{x}_i, \theta)$.

Definition 1: (Sample distribution: Pfeffermann, Krieger and Rinott, 1998). The (marginal) sample pdf of y_i is defined as:

$$
\begin{aligned}
f_s(y_i \mid \mathbf{x}_i, \theta, \gamma) &= f_p(y_i \mid \mathbf{x}_i, \theta, \gamma, i \in \mathrm{s}) \\
&= \frac{\Pr(i \in s \mid \mathbf{x}_i, y_i, \gamma)\, f_p(y_i \mid \mathbf{x}_i, \theta)}{\Pr(i \in s \mid \mathbf{x}_i, \theta, \gamma)}, \qquad (3)
\end{aligned}
$$

where θ is the parameter of the population distribution, γ is the parameter indexing $\Pr(i \in s \mid \mathbf{x}_i, y_i, \gamma)$ and

$$
\begin{aligned}
\Pr(i \in s \mid \mathbf{x}_i, \theta, \gamma) &= \int \Pr(i \in s \mid \mathbf{x}_i, y_i, \gamma)\, f_p(y_i \mid \mathbf{x}_i, \theta) dy_i \\
&= E_p\big(\Pr(i \in s \mid \mathbf{x}_i, y_i, \gamma) \mid \mathbf{x}_i\big)
\end{aligned}
$$

Note that the marginal sample pdf is different from the superpopulation pdf generating the finite population values, unless $\Pr(i \in s \mid \mathbf{x}_i, y_i, \gamma) = \Pr(i \in s \mid \mathbf{x}_i, \theta, \gamma)$ for all possible values y_i, in which case the sampling process or scheme is noninformative or can be ignored conditional on $\mathbf{x}_i$. Also note that the marginal sample distribution is a function of the population distribution and of the first order sample inclusion probabilities.

Comment 1: When the population pdf does not depend on auxiliary variables $\mathbf{x}_i$, so that $y_i \sim f_p(y_i \mid \theta)$, then the sample pdf of y_i is given by:

$$f_s(y_i \mid \theta, \gamma) = \frac{\Pr(i \in s \mid y_i, \gamma) f_p(y_i \mid \theta)}{\Pr(i \in s \mid \theta, \gamma)}, \tag{4}$$

where

$$\Pr(i \in s \mid \theta, \gamma) = \int \Pr(i \in s \mid y_i, \gamma) f_p(y_i \mid \theta) dy_i .$$

Note that $E_p(\pi_i \mid y_i) = E_{\mathbf{z}_i \mid y_i} E_p(\pi_i \mid y_i, \mathbf{z}_i)$, so that $\mathbf{z}_i$ is integrated out in (4).
In what follows, we give the relationship between selection probabilities and conditional expectation of inclusion probabilities.

Theorem 1: (Pfeffermann, Krieger and Rinott, 1998).

$$\Pr(i \in s \mid \mathbf{x}_i, y_i, \gamma) = E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma). \tag{5}$$

Comment 2: In general, $\Pr(i \in s \mid \mathbf{x}_i, y_i, \gamma) \neq \Pr(i \in s \mid \mathbf{x}, \mathbf{y}, \mathbf{z}) = g_i(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \pi_i$, but from Theorem 1 we see that: $\Pr(i \in s \mid \mathbf{x}_i, y_i, \gamma) = E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma)$. Now substituting equation (5) in (3), we have corollary 1.

Corollary 1: An alternative representation of the marginal sample pdf is:

$$f_s(y_i \mid \mathbf{x}_i, \theta, \gamma) = \frac{E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma) f_p(y_i \mid \mathbf{x}_i, \theta)}{E_p(\pi_i \mid \mathbf{x}_i, \theta, \gamma)}, \tag{6}$$

where $E_p(\pi_i \mid \mathbf{x}_i, \theta, \gamma) = E_p(E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma))$.

The advantage of Corollary 1 is that: given the population pdf, $f_p(y_i \mid \mathbf{x}_i, \theta)$, the sample pdf, $f_s(y_i \mid \mathbf{x}_i, \theta, \gamma)$, is completely determined by the population conditional expectation of the first order sample inclusion probabilities, given the outcome and auxiliary variables, $E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma)$. Note that the use of (6) does not require the specification of the full distribution of the first order selection probabilities π_i, which is often intractable.

Comment 3: The sample pdf contains the population parameter, θ, that indexes, $f_p(y_i \mid \mathbf{x}_i, \theta)$, and the informativeness parameter, γ, that indexes, $E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma)$. Thus, the sample pdf may contain more parameters than the population pdf. The following example is of practical use.

Example 1: (Polynomial regression population model). Let the population distribution be:

$$y_i \mid x_i \underset{p}{\sim} N\left(\sum_{j=0}^{J} \delta_j \beta_j x_i^j, \sigma^2 v(x_i)\right), i = 1,\ldots, N, \tag{7}$$

where x_i, s are assumed fixed (non-stochastic) quantities, β_j, σ^2 are unknown quantities, $v(x_i)$ is a known function of x_i, and $\delta_j = 1(0)$ if the term x_i^j is present (absent) in $\mu_i = E_p(y_i \mid x_i)$. This model was discussed by Royall and Herson (1973) under noninformative probability sampling design.

Suppose that the sample inclusion probabilities have conditional expectations:

$$E_p(\pi_i \mid x_i, y_i, \gamma) = \exp(\gamma y_i + h(x_i)) \tag{8}$$

for some function $h(x)$. Now substituting (7) and (8) in (6), the marginal sample pdf of y_i is obtained as:

$$f_s(y_i \mid x_i, \boldsymbol{\beta}, \gamma) = \frac{\exp\left(\frac{-1}{2\pi\sigma^2 v(x_i)}\left(y_i - \sum_{j=0}^{J} \delta_j \beta_j x_i^j - \gamma\sigma^2 v(x_i)\right)^2\right)}{\left(2\pi\sigma^2 v(x_i)\right)^{0.5}}. \tag{9}$$

Thus the polynomial regression of y on x in the sample is the same as the polynomial regression in the population, except that the mean is shifted by a constant: $\gamma\sigma^2 v(x_i)$. That is:

$$\begin{aligned} E_s(y_i \mid x_i) &= \sum_{j=0}^{J} \delta_j \beta_j x_i^j + \gamma\sigma^2 v(x_i) \\ &= E_p(y_i \mid x_i) + \gamma\sigma^2 v(x_i) \end{aligned} \tag{10}$$

and

$$Var_s(y_i \mid x_i) = \sigma^2 v(x_i) = Var_p(y_i \mid x_i). \tag{11}$$

According to (10), we have:

1. If $\gamma = 0$, that is, the sampling design is ignorable (noninformative), then $E_s(y_i \mid x_i) = E_p(y_i \mid x_i)$.
2. If $\gamma > 0$, that is, the sample distribution produces larger values of y_i more often than the population distribution, and produces its smaller values less often, then

   $$E_s(y_i \mid x_i) > E_p(y_i \mid x_i).$$

3. If $\gamma < 0$, that is, the sample distribution produces smaller values of y_i more often than the population distribution, and produces its larger values less often, then $E_s(y_i \mid x_i) < E_p(y_i \mid x_i)$. In particular, if $\delta_1 = \delta_0 = 1, \delta_j = 0$ for $j = 2,..., J$ and if $v(x_i) = x_i$, that is we have a simple linear regression model with variance of y_i proportional to x_i, then by substitution in (9) we get:

$$y_i \mid x_i \underset{s}{\sim} N\left(\beta_0 + \left(\gamma\sigma^2 + \beta_1\right)x_i, \sigma^2 x_i\right), i = 1,..., n. \qquad (12)$$

Hence for this particular case, the regression of y on x in the sample is the same as in the population, except that the slope of the regression line is shifted by a constant, $\gamma\sigma^2$, while the intercept and the variance do not change. However if the population variance is constant, i.e., $v(x_i) = 1$, so that $y_i \mid x_i \underset{p}{\sim} N\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$, then the regression line in the sample is $y_i \mid x_i \underset{s}{\sim} N\left(\beta_0 + \gamma\sigma^2 + \beta_1 x_i, \sigma^2\right)$. That is, the regression of y on x in the sample is the same as in the population, except that the intercept of the regression line is shifted by a constant, $\gamma\sigma^2$, while the slope and the variance do not change.

Comment 4: Under informative sampling, we see that the regression relationship for the selected sample is, in general, different from that of the population. Similar results are obtained in the area of econometrics, when treating the problem of missing values where we observe only a subsample of a larger random sample. This leads to the sample selection bias problem, as discussed in detail by Heckman (1976, 1979), Nathan and Smith (1989), Little (1982), Hausman and Wise (1977, 1979), Ruud (2000) and Greene (2000).

Let us now consider the following example where no auxiliary variables are available.

Example 2: (Exponential population distribution). Assume that the population distribution of the study variable y_i, $i \in U$ is exponential with parameter θ, so that the pdf of y_i is given by:

$$f_p(y_i \mid \theta) = \theta \exp(-\theta y_i), \; \theta > 0 \text{ and } y_i > 0.$$

Suppose that the sample inclusion probabilities have conditional expectations:

$$E_p(\pi_i \mid y_i, \gamma_0, \gamma_1) = \exp(\gamma_0 + \gamma_1 y_i).$$

Then according to (6), we can show that the sample pdf of y_i is given by:

$$f_s(y_i \mid \theta, \gamma_0, \gamma_1) = (\theta - \gamma_1)\exp(-(\theta - \gamma_1)y_i), \; \theta > \gamma_1 \text{ and } y_i > 0. \qquad (13)$$

Let us compute the sample cumulative distribution function of y_i:

$$F_s(t) = \Pr(y_i \le t \mid i \in s) = 1 - \exp(-(\theta - \gamma_1) y_i).$$

So that, under the sample distribution, the $\alpha - th$ percentile, $y_{\alpha s}$, is given by:

$$\begin{aligned} y_{\alpha,s} &= -\frac{1}{\theta}\log(1-\alpha) + \gamma_1 \log(1-\alpha) \\ &= y_{\alpha,p} + \gamma_1 \log(1-\alpha) \end{aligned}$$

where $y_{\alpha,p} = -(1/\theta)\log(1-\alpha)$ is the $\alpha - th$ percentile under the population distribution.

In Section 1 we mentioned that the sampling design is informative if the outcome variable is correlated with design variables not included in the model. To make this statement more clear let us continue with Example 2.

Example 3: (Exponential population distribution under linear design variable). Let z_i, $i \in U$ be a design variable used for the selection process correlated with the outcome variable, y_i, but not included in the working model, which is in our case an exponential distribution. Assume that: $z_i = \gamma_0 + \gamma_1 y_i + u_i$, $u_i \underset{p}{\sim} U(-1,1)$ and u_i and y_i are uncorrelated. Select a sample of size n from this population using probability proportional to size sample, with size variable z_i. We assume that the distribution of y_i and the values of γ_0 and γ_1 ensure that $z_i > 0$ with probability one. Thus the first order sample selection probabilities are given by: $\pi_i = f\, z_i / \bar{z}$, where $f = n/N$ is the sampling fraction and $\bar{z} = \sum_{i=1}^{N} z_i \big/ N$ is the population mean of the design variable, z_i. In order to find the sample pdf of y_i, we need first to compute $E_p(\pi_i \mid y_i, \gamma_0, \gamma_1)$ and $E_p(\pi_i \mid \theta, \gamma_0, \gamma_1)$.

Now

$$E_p(\pi_i \mid y_i, \gamma_0, \gamma_1) = E_p\left(f\, \frac{z_i}{\bar{z}} \mid y_i, \gamma_0, \gamma_1 \right).$$

To simplify calculation of this conditional expectation we assume that N is large enough so that, approximately:

$$E_p\left(\frac{z_i}{\bar{z}} \mid y_i, \gamma_0, \gamma_1 \right) = \frac{E_p(z_i \mid y_i, \gamma_0, \gamma_1)}{E_p(\bar{z})}.$$

Hence, we have, approximately:

$$E_p(\pi_i \mid y_i, \gamma_0, \gamma_1) = (f/\mu_z)(\gamma_0 + \gamma_1 y_i), \tag{14}$$

where $\mu_z = E_p(\bar{z})$ and

$$E_p(\pi_i \mid \theta, \gamma_0, \gamma_1) = E_p(E_p(\pi_i \mid y_i, \gamma_0, \gamma_1)) \\ = (f/\mu_z)(\gamma_0 + \gamma_1 E_p(y_i)) \tag{15}$$

Thus for the exponential distribution (Example 2) the sample pdf of y_i is given by:

$$f_s(y_i \mid \theta, \gamma_0, \gamma_1) = \frac{(\gamma_0 + \gamma_1 y_i)\theta \exp(-\theta y_i)}{\left(\gamma_0 + \frac{\gamma_1}{\theta}\right)} \tag{16}$$

This sample pdf can be written as:

$$f_s(y_i \mid \theta, \gamma_0, \gamma_1) = a_0 f_p(y_i \mid \theta) + a_1 f_p^w(y_i \mid \theta), \tag{17}$$

where $f_p^w(y_i \mid \theta) = y_i \theta^2 \exp(-\theta y_i)$ is the weighted population pdf , $a_0 = \dfrac{\gamma_0}{\gamma_0 + (\gamma/\theta)}$ and $a_1 = 1 - a_0$. Thus in our case, the sample pdf is a mixture of the population pdf, $f_p(y_i \mid \theta)$, and the weighted population distribution, $f_p^w(y_i \mid \theta)$, with weights a_0 and a_1. Thus it is not always the case that the sample and population pdf's belong to the same family.

Referring to (15), we can see that the normalizing factor, f/μ_z , in the numerator and denominator cancels out, so that this normalizing factor is irrelevant when modeling the conditional expectation of the first order sample selection probabilities, given the outcome variable, $E_p(\pi_i \mid y_i, \gamma_0, \gamma_1)$. Also notice that (15) is equivalent to $E_p(\pi_i \mid \gamma_0, \gamma_1) = f$, the sampling fraction, which is constant.

5. SAMPLE LIKELIHOOD

Having derived the sample distribution, the question that arises is whether under commonly used sampling methods the sample measurements are approximately independent. This question is fundamental as many of the standard inference procedures assume independence of the observation. Thus, for example, classical likelihood inference usually assumes that the sample likelihood is computed as a product of pdf's.

Pfeffermann, Krieger and Rinott (1998) proved the following asymptotic independence theorem related to probability proportional to size (PPS) sampling with replacement. The same independence property holds for various other sampling schemes of selection without replacement. The following theorem introduces the asymptotic independence of sample measurements when auxiliary variables are available.

Theorem 2: (Asymptotic independence of sample measurements: Pfeffermann, Krieger and Rinott, 1998). Under the regularity conditions below, the following is true: if s consists of n distinct units then, asymptotically, as $N \to \infty$ (with n fixed):

$$f_s(y_1,...,y_n \mid \mathbf{x}_i,\theta,\gamma)=\prod_{i=1}^{n} f_s(y_i \mid \mathbf{x}_i,\theta,\gamma) \tag{18}$$

The prominent feature of this theorem is that, if the population measurements are independent, then as $N \to \infty$ $(\text{with } n \text{ fixed})$ the sample measurements are asymptotically independent, so we can apply standard inference procedures to complex survey data by using the marginal sample distribution for each unit. This leads to the following sample likelihood for θ and γ:

$$L_s(\theta,\gamma)=\prod_{i=1}^{n} f_s(y_i \mid \mathbf{x}_i,\theta,\gamma)=\prod_{i=1}^{n}\frac{E_p(\pi_i \mid \mathbf{x}_i,y_i,\gamma)}{E_p(\pi_i \mid \mathbf{x}_i,\theta,\gamma)} f_p(y_i \mid \mathbf{x}_i,\theta).$$

and the logarithm of the sample likelihood for θ and γ is given by:

$$\begin{aligned} l_s(\theta,\gamma)&=\sum_{i=1}^{n}\log f_s(y_i \mid \mathbf{x}_i,\theta,\gamma)\\ &=l_{srs}(\theta)+\sum_{i=1}^{n}\log E_p(\pi_i \mid \mathbf{x}_i,y_i,\gamma)-\sum_{i=1}^{n}\log E_p(\pi_i \mid \mathbf{x}_i,\theta,\gamma), \end{aligned} \tag{19}$$

where $l_{srs}(\theta)=\sum_{i=1}^{n}\log\left(f_p(y_i \mid \mathbf{x}_i,\theta)\right)$ is the classical log-likelihood obtained by ignoring the sample design.

The function given in equation (19) can be maximized with respect to θ and γ to obtain the maximum sample likelihood estimates of these parameters. Maximum sample likelihood estimators of other parameters, which are the parameters of interest, (e.g. the parameter θ characterizing the population distribution of y) are defined using the invariance properties of the maximum likelihood (ML) approach.

The sample likelihood function, $L_s(\theta,\gamma)$, can be interpreted as a weighted likelihood, where the weights are ratios of the population conditional expectations of the inclusion probabilities, given the values of y_i , and their unconditional expectations.

Standard estimation processes consider the case where the sampling scheme is ignored and base the inference on the classical log-likelihood function, $l_{srs}(\theta)$. However, analysis using standard estimation methods, which ignores the last two terms of (19), leads to inconsistent estimates of θ. Thus the effect of the sampling scheme must be taken into account.

Comment 5: The joint sample pdf of $\mathbf{y}=(y_1,...,y_n)$ is given by:

$$f_s(\mathbf{y})= f_s(y_1,...,y_n)\frac{\Pr(s \mid y_1,...,y_n)}{\Pr(s)} f_p(y_1,...,y_n).$$

If $\Pr(s \mid \mathbf{y}) = \prod_{i=1}^{n} \exp(a_i y_i)$, then $f_s(\mathbf{y}) = \prod_{i=1}^{n} f_s(y_i)$, where $f_s(y_i) = \dfrac{\exp(a_i y_i) f_p(y_i)}{E_p(\exp(a_i y_i))}$.

Thus the sample measurements are independent.

Comment 6: The sample likelihood function given in (19) is a sum of two components, the first component is the classical likelihood function which ignores the method of selection and just treats the sample values of y as independent draws from the population distribution of y, while the second component reflects the effect of the method of sample of selection.

Example 4: (MLE of θ- exponential design variable). Let us find the MLE of θ based on the sample pdf given in (13). In this case the parameters of the sample pdf contains the parameters of the population pdf, θ, and the informativeness parameter, γ_1. So that the MLE of θ and γ_1 is the solution of the sample likelihood equations:

$$\begin{aligned} \frac{\partial l_s(\theta, \gamma_1)}{\partial \theta} &= \frac{n}{\theta - \gamma_1} - n\bar{y} = 0 \\ \frac{\partial l_s(\theta, \gamma_1)}{\partial \gamma_1} &= \frac{-n}{\theta - \gamma_1} + n\bar{y} = 0 \end{aligned} \tag{20}$$

Solving this system of sample likelihood equations for θ and γ_1 gives: $(\theta - \gamma_1)^{-1} = \bar{y}$.

Since we have two unknowns and one equation, we have infinitely many solutions. Hence the parameters are not identifiable from the sample observations of y alone. (A model is said to be nonidentifiable if it contains parameters that cannot be estimated uniquely or, to put in another way, that have standard errors of infinity). The identifiability problem occurs here because we have only one sufficient statistic, which is $\bar{y}$, for two parameters. To solve this problem we consider two-step estimation, see Section 9. For illustration assume that the informative parameter is known, say $\gamma_1 = \gamma_1^0$, so that if this is the case, the MLE of θ is: $\hat{\theta} = (1/\bar{y}) + \gamma_1^0 = \hat{\theta}_{srs} + \gamma_1^0$. Here $\hat{\theta}_{srs}$ underestimates the true value of θ if $\gamma_1^0 > 0$ and overestimates the true value of θ if $\gamma_1^0 < 0$.

6. MODELING OF THE CONDITIONAL EXPECTATIONS OF SAMPLE INCLUSION PROBABILITIES

In the discussion so far, we mentioned that, for a given population distribution, the sample distribution is completely determined by the specification of the conditional expectations of sample inclusion probabilities, $E_p(\pi_i \mid y_i, \mathbf{x}_i, \gamma)$. So in order to obtain the sample pdf of y_i, we need to model these population conditional expectations. Pfeffermann, Krieger, and Rinott (1998) introduced two alternative approximation models for this population conditional expectation:

6.1 Exponential Model

Suppose that the sample inclusion probabilities have conditional expectations:

$$E_p(\pi_i \mid y_i, \mathbf{x}_i) \approx \exp\left(\sum_{j=0}^{J} a_j y_i^j + h_1(\mathbf{x}_i)\right) \tag{21a}$$

for some function $h_1(\mathbf{x})$, where $\{a_j, j = 0,..., J\}$ are unknown parameters to be estimated from the sample, see Section 7. In particular, if $J = 1$, then (21a) becomes:

$$E_p(\pi_i \mid y_i, \mathbf{x}_i) \approx \exp(a_0 + a_1 y_i + h_1(\mathbf{x}_i)). \tag{21b}$$

We interpret exponential model approximation in the spirit of probability proportional to size sampling as follows, let the size measure be:

$$z_i = \exp\left(\sum_{j=0}^{J} a_j y_i^j + h_1(\mathbf{x}_i) + u_i\right),$$

where $u_i \underset{p}{\sim} U(a,b)$. Let $\pi_i = nz_i \Big/ \sum_{i=1}^{N} z_i$. Assume N is large enough so that the difference between $N\bar{z}$ and $E_p\left(\sum_{i=1}^{N} z_i\right)$ can be ignored. Under these assumptions we can show that $E_p(\pi_i \mid y_i, \mathbf{x}_i)$ has the form of (21a).

Comment 7: Under (21b), the marginal effect of y_i on $E_p(\pi_i \mid y_i, \mathbf{x}_i)$ is given by:

$$\frac{\partial E_p(\pi_i \mid y_i, \mathbf{x}_i)}{\partial y_i} = a_1 E_p(\pi_i \mid y_i, \mathbf{x}_i).$$

1. If $a_1 = 0$, then $E_p(\pi_i \mid y_i, \mathbf{x}_i)$ does not depend on y_i, so that the sampling design is noninformative.
2. If $a_1 > 0$, then $E_p(\pi_i \mid y_i, \mathbf{x}_i)$ is an increasing function of y_i, so that larger values are more likely to be in the sample than smaller values.
3. If $a_1 < 0$, then $E_p(\pi_i \mid y_i, \mathbf{x}_i)$ is a decreasing function of y_i, so that smaller values are more likely to be in the sample than larger values.

6.2 Polynomial Model

Suppose that the sample inclusion probabilities have conditional expectations:

$$E_p(\pi_i \mid y_i, \mathbf{x}_i) \approx \left(\sum_{j=0}^{J} b_j y_i^j + h_2(\mathbf{x}_i) \right) \tag{22a}$$

for some function $h_2(\mathbf{x})$, where $\{b_j, j = 0,..., J\}$ are unknown parameters to be estimated from the sample, see Section 7. In particular, if $J = 1$, then (22a) becomes:

$$E_p(\pi_i \mid y_i, \mathbf{x}_i) \approx (b_0 + b_1 y_i + h_2(\mathbf{x}_i)). \tag{22b}$$

We interpret the polynomial model approximation in the spirit of probability proportional to size sampling as follows, let the size measure be:

$$z_i = \left(\sum_{j=0}^{J} b_j y_i^j + h_2(\mathbf{x}_i) + u_i \right),$$

where $u_i \underset{p}{\sim} U(a,b)$. Let $\pi_i = nz_i \big/ \sum_{i=1}^{N} z_i$. Assume N is large enough so that the difference between $N\bar{z}$ and $E_p\left(\sum_{i=1}^{N} z_i\right)$ can be ignored. Under these assumptions we can show that $E_p(\pi_i \mid y_i, \mathbf{x}_i)$ has the form of (22a), see Example 3.

Comment 8: Under (22b), the marginal effect of y_i on $E_p(\pi_i \mid y_i, \mathbf{x}_i)$ is given by:

$$\frac{\partial E_p(\pi_i \mid y_i, \mathbf{x}_i)}{\partial y_i} = b_1 .$$

Similarly to the situation in Example 2, if $b_1 = 0$, the sampling design is noninformative, if $b_1 > 0$ larger values are more likely to be in the sample than smaller values and vice versa if $b_1 < 0$.

Other standard ways of modeling first order inclusion probabilities are obtained by the spirit of generalized linear models, via the logit and probit models, see McCullagh and Nelder (1989) and Eideh and Nathan (2004).

6.3 Logit Model

Suppose that the sample inclusion probabilities have conditional expectations:

$$\begin{aligned} E_p(\pi_i \mid y_i, \mathbf{x}_i) &= \Pr(i \in s \mid y_i, \mathbf{x}_i) \\ &\approx \frac{\exp\left(\sum_{j=0}^{J} c_j y_i^j + h_3(\mathbf{x}_i) \right)}{1 + \exp\left(\sum_{j=0}^{J} c_j y_i^j + h_3(\mathbf{x}_i) \right)} \end{aligned} \tag{23a}$$

for some function $h_3(\mathbf{x})$, where $\{c_j, j = 0,..., J\}$ are unknown parameters to be estimated from the sample, see Section 7. In particular, if $J = 1$, then (23a) becomes:

$$E_p(\pi_i \mid y_i, \mathbf{x}_i) = \Pr(i \in s \mid y_i, \mathbf{x}_i) \approx \frac{\exp(c_0 + c_1 y_i + h_3(\mathbf{x}_i))}{1 + \exp(c_0 + c_1 y_i + h_3(\mathbf{x}_i))}. \tag{23b}$$

Here we interpret this approximation by taking:

$$z_i = \frac{\exp\left(\sum_{j=0}^{J} c_j y_i^j + h_3(\mathbf{x}_i) + u_i\right)}{1 + \exp\left(\sum_{j=0}^{J} c_j y_i^j + h_3(\mathbf{x}_i) + u_i\right)}$$

and by using Taylor approximation we can show that $E_p(\pi_i \mid y_i, \mathbf{x}_i)$ has the form of (23a), where $u_i \underset{p}{\sim} U(a,b)$.

6.4 Probit Model

Suppose that the sample inclusion probabilities have conditional expectations:

$$E_p(\pi_i \mid y_i, \mathbf{x}_i) = \Pr(i \in s \mid y_i, \mathbf{x}_i) \approx \Phi\left(\sum_{j=0}^{J} d_j y_i^j + h_4(\mathbf{x}_i)\right), \tag{24a}$$

where Φ denotes the cumulative distribution function of the standard normal distribution, for some function $h_4(\mathbf{x})$ and $\{d_j, j = 0,..., J\}$ are unknown parameters to be estimated from the sample, see Section 7. In particular, if $J = 1$, then (24a) becomes:

$$E_p(\pi_i \mid y_i, \mathbf{x}_i) = \Pr(i \in s \mid y_i, \mathbf{x}_i) \approx \Phi(d_0 + d_1 y_i + h_4(\mathbf{x}_i)). \tag{24b}$$

Here also we interpret this approximation by taking:

$$z_i = \Phi\left(\sum_{j=0}^{J} d_j y_i^j + h_4(\mathbf{x}_i) + u_i\right)$$

and by using a Taylor approximation we can show that $E_p(\pi_i \mid y_i, \mathbf{x}_i)$ has the form of (245a), where $u_i \underset{p}{\sim} U(a,b)$.

From now on, we use the terms *'linear model' or 'linear sampling'* or *'linear inclusion probabilities'* to denote that the conditional expectation of the first order inclusion probabilities is a linear function of the response variable and the available auxiliary variables; see equation (22b). Similarly, we use the terms 'exponential model' or *'exponential sampling'* or *'exponential inclusion probabilities'* to denote that the conditional expectation of the first order inclusion probabilities is an exponential function of the response variable and the available auxiliary variables; see equation (21a).

Now we have the following theorem, which gives the sample pdf's and their moments under the special cases of the different models for the conditional expectations considered above. In the following, we suppress the notation relating to the dependence of the pdf's on the unknown parameters.

Theorem 3: (Sample distribution under exponential, linear, logit and probit models).

I. Under the exponential model (21b), the sample pdf of y_i is given by:

$$f_s(y_i \mid \mathbf{x}_i) = \frac{\exp(a_1 y_i) f_p(y_i \mid \mathbf{x}_i)}{M_p(a_1)}, \tag{25a}$$

where $M_p(a_1) = E_p(\exp(a_1 y_i) \mid \mathbf{x}_i)$ is the moment generating function (mgf) of the population pdf of y_i.

Also the mgf and the mean of the sample pdf of y_i are given by:

$$\begin{aligned} M_s(t) &= \frac{M_p(t+a_1)}{M_p(a_1)} \\ E_s(y_i \mid \mathbf{x}_i) &= \frac{\partial \log M_p(a_1)}{\partial a_1} \end{aligned} \tag{25b}$$

II. Under the polynomial model (22b), the sample pdf of y_i is given by:

$$f_s(y_i \mid \mathbf{x}_i) = \alpha_0 f_p(y_i \mid \mathbf{x}_i) + \alpha_1 f_p^w(y \mid \mathbf{x}_i), \tag{26a}$$

where,

$$\alpha_0 = \frac{b_0^*}{b_0^* + b_1 E_p(y_i \mid \mathbf{x}_i)}, b_0^* = b_0 + h_2(\mathbf{x}_i), \alpha_0 + \alpha_1 = 1, \text{ and } f_p^w(y \mid \mathbf{x}_i) = \frac{y_i f_p(y_i \mid \mathbf{x}_i)}{E_p(y_i \mid \mathbf{x}_i)}.$$

Also the mgf of sample pdf of y_i is given by:

$$M_s(t)=\frac{\left(b_0^* M_p(t)+b_1 M'_p(t)\right)}{b_0^*+b_1 E_p(y_i)} \tag{26b}$$

and

$$E_s(y_i \mid \mathbf{x}_i)=E_p(y_i \mid \mathbf{x}_i)+\frac{b_1 Var_p(y_i \mid \mathbf{x}_i)}{b_0^*+b_1 E_p(y_i \mid \mathbf{x}_i)}. \tag{26c}$$

III. Under the logit model (23b), the sample pdf of y_i is given by:

$$f_s(y_i \mid \mathbf{x}_i)=\frac{\exp(c_0+c_1 y_i+h_3(\mathbf{x}_i))f_p(y_i \mid \mathbf{x}_i)}{(1+\exp(c_0+c_1 y_i+h_3(\mathbf{x}_i)))E_p\left(\frac{\exp(c_0+c_1 y_i+h_3(\mathbf{x}_i))}{(1+\exp(c_0+c_1 y_i+h_3(\mathbf{x}_i)))} \mid \mathbf{x}_i\right)} \tag{27a}$$

IV. Under the probit model (24b), the sample pdf of y_i is given by:

$$f_s(y_i \mid \mathbf{x}_i)=\frac{\Phi(d_0+d_1 y_i+h_4(\mathbf{x}_i))f_p(y_i \mid \mathbf{x}_i)}{E_p(\Phi(d_0+d_1 y_i+h_4(\mathbf{x}_i)) \mid \mathbf{x}_i)}. \tag{27b}$$

Proof: Using equation (6).

The importance of this theorem is that, it gives the relationship between the sample and population pdf's under different models for the conditional expectation of the sample selection probabilities and also provides us with the relationship between the moments of the sample and population distributions, which can be calculated by routine differentiation.
The following theorem gives the sample distributions of the exponential family of distributions under the exponential model of conditional expectation of the sample selection probabilities.

Theorem 4: (Generalized linear models – exponential sampling). Let the population distribution be a member of the exponential family of distributions:

$$f_p(y_i \mid \mathbf{x}_i)=\exp\left(\frac{y_i\theta-\Psi(\theta)}{\phi}+c(y_i,\phi)\right), \tag{28a}$$

where $\psi(\cdot)$ and $c(\cdot)$ are known functions. The parameter θ is known as the natural parameter. Assume that $\theta=g(\mathbf{x}_i'\boldsymbol{\beta})$ where $g(\cdot)$ is a known increasing differentiable function and $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)'$ is a vector of parameter. Suppose the sample inclusion probabilities have expectations:

$$E_p(\pi_i \mid y_i,\mathbf{x}_i)\approx\exp(a_0+a_1 y_i+h_1(\mathbf{x}_i)).$$

Then

$$f_s(y_i \mid \mathbf{x}_i) = \exp\left(\frac{y_i(a_1\phi + \theta) - \Psi(a\phi_1 + \theta)}{\phi} + c(y_i, \phi) \right) \tag{28b}$$

and

$$E_s(y_i \mid \mathbf{x}_i) = \frac{\partial \Psi(a_1\phi + \theta)}{\partial a_1}. \tag{29}$$

Hence the distribution of y_i in the sample is the same as that of the population, which is a member of exponential family of distributions, except that the natural parameter θ changes to $\theta + a_1\phi$ and the function $\Psi(\theta)$ changes to $\Psi(\theta + a_1\phi)$.

Proof:
The moment generating function of the population pdf of y_i is given by:

$$M_p(a_1) = \int \exp(a_1 y_i) \exp\left(\frac{y_i\theta - \Psi(\theta)}{\phi} + c(y_i, \phi) \right) dy_i$$
$$= \exp\left(\frac{\Psi(a_1\phi + \theta) - \Psi(\theta)}{\phi} \right).$$

Now substituting $M_p(a_1)$and (28a) in (25a), the sample pdf of y_i is obtained as:

$$f_s(y_i \mid \mathbf{x}_i) = \frac{\exp(a_1 y_i) \exp\left(\frac{y_i\theta - \Psi(\theta)}{\phi} + c(y_i, \phi) \right)}{\exp\left(\frac{\Psi(a_1\phi + \theta) - \Psi(\theta)}{\phi} \right)},$$
$$= \exp\left(\frac{y_i(a_1\phi + \theta) - \Psi(a\phi_1 + \theta)}{\phi} + c(y_i, \phi) \right)$$

The moment generating function of y_i is given by:

$$M_s(t) = \exp\left(\frac{\Psi(t\phi + a_1\phi + \theta) - \Psi(\theta + a_1\phi)}{\phi} \right).$$

Then using (25b), we have:

$$E_s(y_i \mid \mathbf{x}_i) = \frac{\phi \partial \Psi(a_1\phi + \theta)}{\phi \partial a_1} = \frac{\partial \Psi(a_1\phi + \theta)}{\partial a_1} \neq \frac{\partial \Psi(\theta)}{\partial \theta} = E_p(y_i \mid \mathbf{x}_i).$$

7. RELATIONSHIP BETWEEN MOMENTS

In practice, the conditional expectations of the sample inclusion probabilities, $E_p(\pi_i \mid y_i, \mathbf{x}_i)$ are not known. Assuming that the available data to the analyst is $\{y_i, \mathbf{x}_i, w_i;\ i \in s\}$, which is the case in secondary analysis, the question that arises is: how can we identify and estimate, $E_p(\pi_i \mid y_i, \mathbf{x}_i)$, based only on the sample data? The following theorem answers this question.

Theorem 5: (Relationship between moments: Pfeffermann and Sverchkov, 1999). For pairs of random variables $(y_i, \mathbf{x}_i)$, the following relationships hold:

1. $$E_s(w_i \mid y_i) = \frac{1}{E_p(\pi_i \mid y_i)}, \tag{30a}$$

2. $$E_p(y_i \mid \mathbf{x}_i) = \frac{E_s(w_i y_i \mid \mathbf{x}_i)}{E_s(w_i \mid \mathbf{x}_i)}, \tag{30b}$$

3. $$E_p(y_i) = \frac{E_s(w_i y_i)}{E_s(w_i)}, \tag{30c}$$

4. $$E_s(w_i) = \frac{1}{E_p(\pi_i)}. \tag{30d}$$

Proofs:

1. By definition of conditional expectation, using (12) and (5), we have:

$$E_s(w_i \mid y_i) = \int w_i f_s(w_i \mid y_i) dw_i = \int w_i \frac{E_p(\pi_i \mid w_i, y_i) f_p(w_i \mid y_i)}{E_p(\pi_i \mid y_i)} dw_i$$
$$= \int \frac{f_p(w_i \mid y_i)}{E_p(\pi_i \mid y_i)} dw_i = \frac{1}{E_p(\pi_i \mid y_i)}.$$

2. By definition of conditional expectation and using (30a), we have:

$$E_p(y_i \mid x_i) = \int y_i f_p(y_i \mid x_i) dy_i = \int y_i \frac{E_s(w_i \mid x_i, y_i) f_s(y_i \mid x_i)}{E_s(w_i \mid x_i)} dy_i$$
$$= E_s\left(\frac{y_i E_s(w_i \mid x_i, y_i)}{E_s(w_i \mid x_i)} \mid x_i \right) = \frac{E_s(w_i y_i \mid x_i)}{E_s(w_i \mid x_i)}.$$

3. By setting $x_i = 1$ in (30b).

4. By setting $y_i = \pi_i$ in (30c).

The prominent feature of this theorem is that the expectations of the population conditional sample inclusion probabilities can be identified and estimated from the sample data.

Example 5: Estimation of $E_p(\pi_i \mid y_i, \mathbf{x}_i) \approx \exp(a_0 + a_1 y_i)$. Using (30a) and by expanding $\log(y)$ in a Taylor series expansion about $E(y)$ yields $\log E(y) \approx E[\log(y)]$. So that we can estimate a_0 and a_1 by regressing $(-\log(w_i))$ against y_i, $i \in s$. Thus we get the following OLS estimators of a_0 and a_1:

$$\begin{aligned} \tilde{a}_1 &= -\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)^{-1}\left(\sum_{i=1}^{n}(y_i - \bar{y})(W_i - \overline{W})\right), \\ \tilde{a}_0 &= -(\overline{W} - \tilde{a}_1 \bar{y}_1) \end{aligned} \tag{31}$$

where $W_i = \log(w_i)$. We could also estimate a_0 and a_1, directly, using nonlinear regression techniques.

8. ESTIMATION

One of the main advantages of basing the inference on the sample distribution is that it permits the use of standard inference procedures like those based on the likelihood principle. But as mentioned before, when the conditional expectation of the first order sample selection probabilities is exponential, see (20) and Example 4, the problem of identifiability appears, while under the linear model this problem does not appear. Also there are cases where the number of parameters indexing the sample distribution (the informativeness parameters that index the conditional expectation of the first order sample selection probabilities and the parameters that characterize the population) is large. In these cases it is often computationally easier to estimate these parameters in two steps, see Pfeffermann, Krieger and Rinott (1998) and Pfeffermann and Sverchkov (1999). In the first step, the informativeness parameters are estimated from $\{\mathbf{x}_i, y_i, w_i; i \in s\}$ using Theorem 5. In the second step the population parameters are estimated by ML or other standard methods, with the estimates of the informativeness parameter held fixed. Now according to (19) and using Theorem 5, if the conditions of Theorem 2 hold, then the log-likelihood of the sample data can be written as:

$$l_s(\theta, \gamma) = l_{srs}(\theta) - \sum_{i=1}^{n} \log E_s(w_i \mid \mathbf{x}_i, y_i, \gamma) + \sum_{i=1}^{n} \log E_s(w_i \mid \mathbf{x}_i, \theta, \gamma). \tag{32}$$

where $l_{srs}(\theta) = \sum_{i=1}^{n} \log(f_p(y_i \mid \mathbf{x}_i, \theta))$ is the classical log-likelihood obtained by ignoring the sample design. This function can be maximized with respect to the parameters θ and γ. But as discussed above, in some cases this sample log-likelihood function is not identifiable, for example when the conditional expectation of the first order sample selection probabilities is of exponential form. To solve this problem a two-step estimation method is adopted. Based on the sample data $\{y_i, \mathbf{x}_i, w_i;\ i \in s\}$ we can estimate the parameters of the population model in two steps:

Step-one: Estimate the informativeness parameters γ using the following relationship:

$$E_s\left(w_i \mid \mathbf{x}_i, y_i, \gamma\right) = \frac{1}{E_p\left(\pi_i \mid \mathbf{x}_i, y_i, \gamma\right)}.$$

Thus the informativeness parameters can be estimated using regression analysis. Denoting the resulting estimate of γ by $\tilde{\gamma}$.

Step-two: Substitute $\tilde{\gamma}$ in the sample log-likelihood function, (32), and then maximize the resulting sample log-likelihood function with respect to the population parameters, θ:

$$l_{rs}(\theta, \tilde{\gamma}) = l_{srs}(\theta) - \sum_{i=1}^{n} \log E_s\left(w_i \mid \mathbf{x}_i, y_i, \tilde{\gamma}\right) + \sum_{i=1}^{n} \log E_s\left(w_i \mid \mathbf{x}_i, \theta, \tilde{\gamma}\right), \tag{33a}$$

where $l_{rs}(\theta, \tilde{\gamma})$ is the sample log-likelihood after substituting $\tilde{\gamma}$ in the sample log-likelihood function, (32). But the second component of this sample log-likelihood function does not contain θ, so we can just maximize:

$$\begin{aligned} l_{rs}(\theta, \tilde{\gamma}) &= l_{srs}(\theta) - \sum_{i=1}^{n} \log E_p\left(\pi_i \mid \mathbf{x}_i, \theta, \tilde{\gamma}\right) \\ &= l_{srs}(\theta) + \sum_{i=1}^{n} \log E_s\left(\pi_i \mid \mathbf{x}_i, \theta, \tilde{\gamma}\right) \end{aligned} \tag{33b}$$

The following example illustrates the two-step estimation method.

Example 6: (Bernoulli distribution). Let the population distribution be:

$$f_p\left(y_i \mid \theta\right) = \theta^{y_i}(1-\theta)^{1-y_i},\, i = 1,\ldots,N,\, \theta \in (0,1);\;\; y_i = 0,1. \tag{34a}$$

Assume that $y_1,\ldots,y_N$ are independent.

8.1 Exponential Model

Suppose that the sample inclusion probabilities have expectations:

$$E_p\left(\pi_i \mid y_i\right) = \exp\left(a_0 + a_1 y_i\right),\, a_0, a_1 \neq 0$$

1. According to Theorem 4, we can show that, the sample distribution of y_i is:

$$f_s\left(y_i \mid \theta, a_0, a_1\right) = \left(\theta^*\right)^{y_i}\left(1-\theta^*\right)^{1-y_i},\;\; y_i = 0,1. \tag{34b}$$

   Thus the distribution of $y_i, i \in s$ is the same as the distribution of $y_i, i \in U$, except that the population parameter θ changes to $\theta^* = \dfrac{\theta \exp\left(a_1\right)}{1-\theta+\theta\exp\left(a_1\right)}$. Notice that if $a_1 = 0$, that is the sampling design is ignorable then the sample and population distributions are the same.

2. Estimation of θ, a_0 and a_1

First-step: Estimation of informativeness parameters, a_0 and a_1:
According to Example 5, the least squares estimations of a_0 and a_1 are given in (31).

Second-step: Estimation of the population parameter θ:
According to (33b), the sample log-likelihood function to be maximized is given by:

$$l_{rs}(\theta,\tilde{a}_1) = n\bar{y}\log(\theta) + (n - n\bar{y})\log(1-\theta) - n\log(1-\theta+\theta\exp(\tilde{a}_1)). \tag{35}$$

Now differentiating (35a) with respect to θ and equating it to zero, we have the following MLE of θ which is defined by:

$$\hat{\theta} = \frac{\bar{y}}{\exp(\tilde{a}_1) + \bar{y} - \bar{y}\exp(\tilde{a}_1)}. \tag{36}$$

If $\tilde{a}_1 = 0$, that is the sampling design is estimated as noninformative, then $\hat{\theta} = \bar{y} = \hat{\theta}_{srs}$.

8.2 Linear Model

Now suppose that the sample inclusion probabilities have expectations:

$$E_p(\pi_i \mid y_i) = b_0 + b_1 y_i,\ b_0, b_1 \neq 0.$$

1. According to Theorem 4, the sample distribution of y_i is:

$$f_s(y_i \mid \theta, b_0, b_1) = \frac{b_0 + b_1 y_i}{b_0 + b_1\theta}(\theta)^{y_i}(1\text{-}\theta)^{1-y_i},\ y_i = 0,1. \tag{37}$$

2. Estimation of θ, b_0 and b_1. In this case, we can show that, the MLE's of θ, b_0 and b_1 are defined by:

$$\begin{aligned}
\frac{\partial l_s(\theta,b_0,b_1)}{\partial\theta} &= \frac{n\bar{y}}{\theta} - \frac{n-n\bar{y}}{1-\theta} - \frac{nb_1}{\theta} = 0,\\
\frac{\partial l_s(\theta,b_0,b_1)}{\partial b_0} &= \sum_{i=1}^{n}\left(\frac{1}{b_0 + b_1 y_i}\right) - \frac{n}{(b_0 + b_1\theta)} = 0,\\
\frac{\partial l_s(\theta,b_0,b_1)}{\partial b_1} &= \sum_{i=1}^{n}\left(\frac{y_i}{b_0 + b_1 y_i}\right) - \frac{nb_1}{(b_0 + b_1\theta)} = 0.
\end{aligned} \tag{38}$$

This system of equations can be solved numerically.

Let us now consider the following theorem which is related to the effect of the normalizing factor on the estimation process, when modeling the population conditional expectation of the first order sample selection probabilities, given the outcome variable and possibly auxiliary variables.

Theorem 6: (The effect of normalizing factor on estimation process). Under the two-step estimation method:

1. If $E_p(\pi_i \mid y_i) = k_e \exp(a_0 + a_1 y_i),\ a_0, a_1 \neq 0$, where k_e is some constant, then:

$$l_{rs}(\theta, \tilde{a}_0, \tilde{a}_1) = l_{srs}(\theta) - n \log M_p(\tilde{a}_1).$$

2. If $E_p(\pi_i \mid y_i) = k_l (b_0 + b_1 y_i),\ b_0, b_1 \neq 0$, where k_l is some constant, then:

$$l_{rs}(\theta, \tilde{a}_0, \tilde{a}_1) = l_{srs}(\theta) - n \log\left(b_0 + b_1 E_p(y_i)\right).$$

Proof:

1. $E_p(\pi_i \mid y_i) = k_e \exp(a_0 + a_1 y_i)$ can be written as:

$$E_p(\pi_i \mid y_i) = \exp\left(a_0^* + a_1 y_i\right),$$

where $a_0^* = a_0 + \log(k_e)$. So that:

$$E_p(\pi_i) = \exp\left(a_0^*\right) M_p(a_1) \quad \text{and} \quad \log E_p(\pi_i) = a_0^* + \log M_p(a_1).$$

Hence the estimated sample likelihood, using estimates of the informativeness parameters a_0^* and a_1, is given by:

$$l_{rs}(\theta, \tilde{a}_0, \tilde{a}_1) = l_{srs}(\theta) - (\tilde{a}_0 + \log k_e) - n \log M_p(\tilde{a}_1).$$

Now since $(\tilde{a}_0 - \log k_e)$ does not depend on the parameters of the population distribution, therefore it just can be omitted from the likelihood. Thus the estimated sample likelihood is:

$$l_{rs}(\theta, \tilde{a}_0, \tilde{a}_1) = l_{srs}(\theta) - n \log M_p(\tilde{a}_1).$$

which is free of the normalized factor k_e. Also the estimate of the informativeness parameter, a_1 is not affected by the estimate of the informativeness parameter, a_0^*.

2. The second part is proved similarly

Comment 9: Using Theorem 5, the sample pdf $y_i \mid \mathbf{x}_i$ can be written as:

$$f_s(y_i \mid \mathbf{x}_i) = \frac{E_s(w_i \mid \mathbf{x}_i) f_p(y_i \mid \mathbf{x}_i)}{E_s(w_i \mid y_i, \mathbf{x}_i)}. \tag{39a}$$

The application of this sample pdf requires the estimation of the conditional expectations: $E_s(w_i \mid y_i, \mathbf{x}_i)$ and $E_s(w_i \mid \mathbf{x}_i)$. These conditional expectations can be estimated from the sample data, using Theorem 5 as follows:

1. Estimate $E_s(w_i \mid y_i, \mathbf{x}_i)$ by regressing w_i against $(y_i, \mathbf{x}_i)$, $i \in s$.
2. Estimate $E_s(w_i \mid \mathbf{x}_i)$ in two steps:

Step-one: Using Theorem 5 and the estimate of $E_s(w_i \mid y_i, \mathbf{x}_i)$ obtained above, we can estimate $E_p(\pi_i \mid \mathbf{x}_i)$ as follows:

$$\begin{aligned} E_p(\pi_i \mid y_i, \mathbf{x}_i) &= \int E_p(\pi_i \mid y_i, \mathbf{x}_i) f_p(y_i \mid \mathbf{x}_i) dy_i \\ &= \int \frac{1}{E_s(w_i \mid y_i, \mathbf{x}_i)} f_p(y_i \mid \mathbf{x}_i) dy_i . \end{aligned} \tag{39b}$$

Step-two: Using Theorem 5 we get:

$$E_s(w_i \mid \mathbf{x}_i) = \frac{1}{E_p(\pi_i \mid \mathbf{x}_i)}. \tag{39c}$$

See Pfeffermann and Sverchkov (2003) for examples and discussion.

The prominent feature of (39a) is that, in order to fit a population model for survey data, obtained under an informative sampling scheme, we need only the sample data $\{(\mathbf{x}_i, y_i, w_i), i \in s\}$ and to specify the underlying population model of $y_i \mid \mathbf{x}_i$. However we do not need to specify the population conditional expectation of sample selection probabilities.

9. MEASURE OF DISTANCE BETWEEN THE SAMPLE AND POPULATION DISTRIBUTIONS

Kullback and Leibler (1958) defined and studied the properties of a measure of information, which is nowadays called Kullback-Leibler information measure. In the previous sections we defined and studied the properties of the sample distribution. The major question underlying the use of the sample distribution is its robustness or sensitivity to wrong specification of the conditional expectations of the sample selection probabilities that determine the sample model. In this section we propose to deal two issues: the sensitivity problem and the amount of information we gain by basing the inference on the sample distribution.

Definition.2: (The Kullback-Leibler information measure). If $H_k, k = 1,2$ is the hypothesis that y is from the probability distribution with pdf $f_k(y)$, then the *Kullback-Leibler (K-L) information measure* for discrimination in favor of H_1 against H_2 is defined as:

$$I(f_1 : f_2) = E_1\left(\log \frac{f_1(y)}{f_2(y)}\right) = \begin{cases} \int_{-\infty}^{\infty} f_1(y) \log \frac{f_1(y)}{f_2(y)} dy_i & \text{if } y \text{ is continuous} \\ \sum_{-\infty}^{\infty} f_1(y) \log \frac{f_1(y)}{f_2(y)} & \text{if } y \text{ is discrete} \end{cases} \tag{40}$$

$I(f_1 : f_2)$ can be interpreted as the mean information per observation from $f_1(y)$ for discrimination in favor of H_1 against H_2, or a measure of distance between $f_1(y)$ and $f_2(y)$. Notice that the expected value is taken under $f_1(y)$ which means that we are assuming that y has pdf $f_1(y)$.

Theorem 7: (Properties of the Kullback-Leibler information measure).
1. Nonnegativity, [Kullback (1978), p14, Theorem 3.1]: $I(f_1 : f_2) \geq 0$ with equality if and only if $f_1(y) = f_2(y)$.
2. Additivity, [Kullback (1978), p12, Theorem 2.1] : If y_1 and y_2 are independent random variables under both $f_1(y)$ and $f_2(y)$, then the K-L information measure in y_1 and y_2 is given by:

$$I(f_1 : f_2; y_1, y_2) = I(f_1 : f_2; y_1) + I(f_1 : f_2; y_2). \tag{41a}$$

3. Minimum discrimination information, [Kullback (1978), p38, Theorem 2.1]: If $f_1(y)$ and $f_2(y)$ are pdf's, $T(y)$ is a statistic such that $\theta = E_1(T(y)) = \int T(y) f_1(y) dy$ exists, the moment generating function of $T(y)$, i.e., $M_2(\tau) = \int \exp(\tau T(y)) f_2(y) dy$ exists for $|\tau| < h$, where h is a positive number and $M_2(\tau)$ is differentiable with respect to τ, then:

$$I(f_1 : f_2) \geq \theta\tau - \log M_2(\tau) = I(f_1^* : f_2), \tag{41b}$$

where $\theta = \frac{d \log M_2(\tau)}{d\tau}$ and equality of (41b) holds if and only if:

$$f_1(y) = f_1^*(y) = \frac{\exp(\tau T(y)) f_2(y)}{M_2(\tau)}.$$

$f_1^*(y)$ is called the least informative distribution (the conjugate distribution) of $f_2(y)$, and $I(f_1^* : f_2)$ is called the minimum discrimination information. Note that τ is a function of θ, therefore when necessary, we shall write $\tau(\theta)$ instead of θ.

4. Estimation of $I\left(f_1^* : f_2\right)$, [Kullback (1978), p94]: Notice that $\tau\left(\hat{\theta}\right)$ is the MLE estimate of $\tau(\theta)$ as a parameter of pdf $f_1^*(y)$ and the MLE $\hat{I}\left(f_1^* : f_2\right)$ of the minimum discrimination information $I\left(f_1^* : f_2\right)$ is given by:

$$\hat{I}\left(f_1^* : f_2\right) = \hat{\theta}\tau\left(\hat{\theta}\right) - \log M_2\left(\tau\left(\hat{\theta}\right)\right). \tag{41c}$$

Proofs: Proofs are given in Kullback (1978). The following corollary relates to the sample and the population distribution, under the exponential sampling, i.e., when the conditional expectation of the sample selection probabilities is exponential.

Corollary 2: For a given population distribution of y_i, with pdf $f_p(y_i)$, and under the conditions that $\int y_i f(y_i) dy_i$ exists, and $M_2(\tau) = \int \exp(\tau y_i) f_2(y_i) dy_i$ exists for τ in some interval, then the least informative distribution corresponding to $f_p(y_i)$ is given by:

$$f_{es}(y_i) = \frac{\exp(\tau y_i) f_p(y_i)}{M_p(\tau)}. \tag{42}$$

This is the sample distribution of y_i when the conditional expectation of the sample selection probabilities is exponential, that is: $E_p(\pi_i \mid y_i) = \exp(\tau_0 + \tau y_i)$. In this case the minimum discrimination information is:

$$I\left(f_{es} : f_p\right) = \theta\tau - \log M_p(\tau), \quad \text{where } \theta = \frac{\partial \log M_p(\tau)}{\partial \tau} \tag{43}$$

Thus for a given population distribution, $f_p(y_i)$, we can measure the degree of informativeness (or nonignorability) by the amount of information we gain by basing the inference on the sample distribution, $f_{es}(y_i)$, rather than on the population distribution $f_p(y_i)$.

In general, for a given $f_p(y_i)$ and $E_p(\pi_i \mid y_i)$, using Theorem 5, the K-L information for discrimination between the population and the sample distributions, $f_p(y_i)$ and $f_s(y_i)$, is given by:

$$\begin{aligned} I\left(f_s : f_p\right) &= E_s \log\left(\frac{E_p(\pi_i \mid y_i)}{E_s \log E_p(\pi_i)}\right) \\ &= E_s \log\left(\frac{E_s(w_i)}{E_s(w_i \mid y_i)}\right) \end{aligned} \tag{44}$$

Notice that the expected value is taken under $f_s(y_i)$ which means that we are assuming that y_i has pdf $f_s(y_i)$. Also $I\left(f_s : f_p\right)$ is completely determined by the population distribution and by the conditional expectation of the sample selection probabilities.

Now since $I\left(f_s : f_p\right) \geq 0$, therefore:

$$E_s \log f_s\left(y_i\right) \geq E_s \log f_p\left(y_i\right). \tag{45}$$

Thus one way to interpret the nonnegativity property is that, in the average, over the sample distribution the log-likelihood of the sample distribution tends to be larger than the log-likelihood of a population distribution, provided that the sample distribution is the true distribution of sample value y_i.

The following example gives the distance between the sample and population distributions under different modeling of the conditional expectation of sample selection probabilities.

Example 7. Let the population distribution be exponential with parameter θ. That is,

$$f_p\left(y_i \mid \theta\right) = \theta \exp\left(-\theta y_i\right),\ \theta > 0 \text{ and } y_i > 0$$

A. Assume that the conditional expectation of the sample selection probabilities is exponential; see equation (21b). Denote by f_{es} the sample distribution.

1. In the analysis of social data, researchers often ignore the sampling design and analyse the data as if they come from a simple random sample. In this case, the K-L information measure is:

$$\begin{aligned} I\left(f_{es} : f_p\right) &= E_{es}\left(\log \frac{f_{es}\left(y_i\right)}{f_p\left(y_i\right)}\right) = E_{es}\left(\log \frac{\left(\theta - a_1\right)\exp\left(-\left(\theta - a_1\right)y_i\right)}{\theta \exp\left(-\theta y_i\right)}\right) \\ &= \log\left(\frac{\theta - a_1}{\theta}\right) + \left(\frac{a_1}{\theta - a_1}\right). \end{aligned} \tag{46a}$$

2. If we assume a linear model for the conditional expectation of the sample selection probabilities; see equation (22b). Denote by f_{ls} the sample distribution, then the K-L information measure is:

$$I\left(f_{es} : f_{ls}\right) = \log\left(\frac{\left(\theta - a_1\right)\left(b_0 + b_1\theta^{-1}\right)}{\theta}\right) + \left(\frac{a_1}{\theta - a_1}\right) - E_{es} \log\left(b_0 + b_1 y_i\right). \tag{46b}$$

No analytical expression for $E_{es} \log\left(b_0 + b_1 y_i\right)$ is available, so general analytical evaluation of $I\left(f_{es} : f_{ls}\right)$ can be complicated, but a Monte Carlo estimate can be computed by generating $y_i \sim f_{es}\left(y_i\right)$ and taking the sample average of $\log\left(b_0 + b_1 y_i\right)$.

B. Assume the sample data are generated under the linear model.

1. If we ignore the sampling design, then the K-L information measure is:

$$I\left(f_{ls} : f_p\right) = E_{ls} \log\left(b_0 + b_1 y_i\right) - \log\left(b_0 + b_1\theta^{-1}\right). \tag{47a}$$

2. If we assume the exponential model whereas the true model is linear, then the K-L information measure is:

$$I\left(f_{ls} : f_{es}\right) = \log\left(\frac{\theta}{(\theta - a_1)\left(b_0 + b_1\theta^{-1}\right)}\right) - \frac{a_1\left(b_0\theta^{-1} + b_1\theta^{-2}\right)}{b_0 + b_1\theta^{-1}} + E_{ls}\log\left(b_0 + b_1 y_i\right). \qquad (47b)$$

C. Assume the sample data were generated by simple random sample design.
1. If we model the sample data as if they were generated by the exponential model, then the K-L information measure is:

$$I\left(f_p : f_{es}\right) = \log\left(\frac{\theta}{\theta - a_1}\right) - \frac{a_1}{\theta}. \qquad (48a)$$

2. If we model the sample data as if they were generated by the linear model, then the K-L information measure is:

$$I\left(f_p : f_{ls}\right) = \log\left(b_0 + b_1\theta^{-1}\right) - E_p \log\left(b_0 + b_1 y_i\right). \qquad (48b)$$

The above cases (A, B and C) lead to the Table 1 which summarizes the obtained results.

Theorem 8: (K-L Distance, exponential family). For the generalized linear model (28a).
1. Suppose the sample inclusion probabilities have expectations:

$$E_p\left(\pi_i \mid y_i, \mathbf{x}_i\right) \approx \exp\left(a_0 + a_1 y_i + h_1(\mathbf{x}_i)\right).$$

Then the K-L distance between f_{es} and f_p is:

$$I\left(f_{es} : f_p\right) = a_1 \frac{\partial\Psi\left(a_1\phi + \theta\right)}{\partial a_1} - \frac{\Psi\left(a_1\phi + \theta\right) - \Psi(\theta)}{\phi}. \qquad (49a)$$

2. Suppose the sample inclusion probabilities have expectations:

$$E_p\left(\pi_i \mid y_i, \mathbf{x}_i\right) \approx \left(b_0 + b_1 y_i + h_1(\mathbf{x}_i)\right).$$

Then the K-L distance between f_{ls} and f_p is:

$$I\left(f_{ls} : f_p\right) = E_{ls}\log\left(b_0 + b_1 y_i + h(\mathbf{x}_i)\right) - \log b_0 + b_1 E_p\left(y_i \mid \mathbf{x}_i\right) + h(\mathbf{x}_i). \qquad (49b)$$

Proof: Using Theorem 4 and Definition 2.

ACKNOWLEDGEMENTS

The author is grateful to the referee, to Gad Nathan, and to my friend Ibrahim Almassri for their valuable comments.

Table 1

The K-L information measure under ignorable[*], exponential, and linear models when the population distribution is exponential: $f_p(y_i \mid \theta) = \theta \exp(-\theta y_i)$, $\theta > 0$ and $y_i > 0$.

True Model	Assumed Model	Kullback-Leibler Information Measure
Exponential	Ignorable	$\log\left(\frac{\theta - a_1}{\theta}\right) + \left(\frac{a_1}{\theta - a_1}\right)$
Exponential	Linear	$\log\left(\frac{(\theta - a_1)(b_0 + b_1\theta^{-1})}{\theta}\right) + \left(\frac{a_1}{\theta - a_1}\right) - E_{es} \log(b_0 + b_1 y_i)$
Linear	Ignorable	$E_{ls} \log(b_0 + b_1 y_i) - \log(b_0 + b_1\theta^{-1})$
Linear	Exponential	$\log\left(\frac{\theta}{(\theta - a_1)(b_0 + b_1\theta^{-1})}\right) - \frac{a_1(b_0\theta^{-1} + b_1\theta^{-2})}{b_0 + b_1\theta^{-1}} + E_{ls} \log(b_0 + b_1 y_i)$
Ignorable	Exponential	$\log\left(\frac{\theta}{\theta - a_1}\right) - \frac{a_1}{\theta}$
Ignorable	Linear	$\log(b_0 + b_1\theta^{-1}) - E_p \log(b_0 + b_1 y_i)$

[*] By *'ignorable model'* we mean that the conditional expectation of the first order inclusion probabilities given the response variable is constant.

The following theorem gives the K-L distance for an exponential family of population distributions, when the conditional expectations of the sample selection probabilities follow the exponential and the linear models.

REFERENCES

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review* 51: 279-292.

Binder, D.A. (1996). Linearization methods for single phase and two phase samples: a cookbook approach. *Survey Methodology* 22: 17-22.

Eideh A.H. (2009). On the use of the Sample Distribution and Sample Likelihood for Inference under Informative Probability Sampling. *DIRASAT(Natural Science), Volume 36 (2009), Number 1*, pp18-29.

Eideh, A. H. and Nathan, G. (2009a). Two-Stage Informative Cluster Sampling with application in Small Area Estimation. *Journal of Statistical Planning and Inference.139*, pp 3088-3101.

Eideh, A. H. and Nathan, G. (2009b). Joint treatment of nonignorable dropout and informative sampling for longitudinal survey data, Chapter 14. *In: Methodology of Longitudinal Surveys*. (P. Lynn - Editor). New York: Wiley, pp. 251 – 263.

Eideh A.H. (2008). Estimation and Prediction of Random Effects Models for Longitudinal Survey Data under Informative Sampling. *Statistics in Transition – New Series .Volume 9, Number 3, December 2008,* pp 485 – 502.

Eideh, A. H. (2007). A Correction Note on Small Area Estimation. *International Statistical Review*. *Volume 75, Issue 1, pp 122-123*.

Eideh, A. H. and Nathan, G. (2006a). The Analysis of Data from Sample Surveys under Informative Sampling. *Acta et Commentationes Universitatis Tartuensis de Mathematica*. *Tartu 2006. Volume 10, pp 41-51.*

Eideh, A. H. and Nathan, G. (2006b). Fitting Time Series Models for Longitudinal Survey Data under Informative Sampling. *Journal of Statistical Planning and Inference* 136**,** 9, pp 3052-306. [Corrigendum, 137 (2007), p 628].

Greene, W.H. (2000). *Econometrics analysis* (4th Ed), London: Prentice-Hall.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample survey methods and theory* (Vol.2)**,** New York: Wiley.

Hausman, J.A. and Wise, D.A. (1977). Social experimentation, truncated distributions and efficient estimation. *Econometrica* 45: 319-339.

Heckman, J.J., (1976). The common structure of statistical models of truncation, sample selection and limited variables, and a simple estimation of such models. *Annals of Econometric and Social Measurement* 54: 475-492.

Heckman, J.J. (1979). Sample selection bias as a specification error. *Econometrica*, 47: 153-161.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 64: 175-195.

Krieger, A.M, and Pfeffermann, D. (1992). Maximum likelihood estimation from complex sample surveys. *Survey Methodology* 18: 225-239.

Krieger, A.M, and Pfeffermann, D. (1997). Testing of distribution functions from complex sample surveys. *Journal of Official Statistics*. 13: 123-142.

Kullback, S. (1978). *Information theory and statistics*. Gloucester, Mass: Peter Smith.

Kullback, S., and Leibler, R.A. (1951). On information and sufficiency. *Ann. Math. Statist*., Vol.22, pp. 79-86.

Little, R.J.A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association* 77: 237-250.

McCullagh, P. and Nelder, J.A (1989). *Generalized linear models*. New York: Chapman and Hall.

Nathan, G and Eideh, A. H. (2008). Le traitement combiné des effets de non-réponse non-ignorable et de sondage informatif dans l'analyse des données issues des enquêtes longitudinales. *In: Méthodes de Sondage : Applications aux Enquêtes Longitudinales, à la Santé, aux Enquêtes Electorales et aux Enquêtes dans les Pays en Développement*. Edited by P. Guilbert, D. Haziza, A. Ruiz-Gazen and Y. Tillé. Paris: Dunod, pp 37-45.

Nathan, G. and Eideh, A. H. (2004). L'analyse des données issues des enquêtes longitudinales sous un plan de sondage informatif. in**:** *Échantillonage et Méthodes d'Enquêtes*. (P. Ardilly – Ed.) Paris: Dunod, pp 227-240.

Nathan, G. and Smith, T.M.F. (1989). The effect of selection on regression analysis. In *Analysis of complex surveys*, eds. C.J. Skinner, D. Holt and T.M.F. Smith, New York: Wiley, pp. 149-163.

Patil, G.P and Rao, C.R. (1978). Weighted distributions and size biased sampling with applications to wildlife populations and human families. *Biometrics* 34: 179-189.

Pfeffermann, D. (2002). Small area estimation - new developments and directions. *International Statistical Review* 70: 125-143.

Pfeffermann, D., Krieger, A.M, and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica* 8: 1087-1114.

Pfeffermann, D., Moura, F. A. S. and Nascimaneto-Silva, P.L. (2001). Multilevel modeling under informative sampling. *Proceedings of the Invited Paper Sessions Organized by the International Association of Survey Statisticians (IASS) for the 53rd session of the International Statistical Institute*, pp. 505-532.

Pfeffermann, D. and Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya***,** 61, Ser. B, Pt. 1: 166-186.

Pfeffermann, D., and Sverchkov, M. (2003). Fitting generalized linear models under informative probability sampling. In *Analysis of Survey Data*, eds. C. Skinner and R. Chambers, New York: Wiley, in press, pp. 175-195.

Ruud, P.A. (2000). *An introduction to classical econometric theory*. New York: Oxford University Press, Inc.

Skinner, C.J. (1989). Domain means, regression and multivariate analysis. In *Analysis of Complex Surve*ys, eds. C.J. Skinner, D. Holt and T.M.F. Smith, New York: Wiley, pp. 59-88.

Skinner, C.J. (1994). Sample models and weights. *American Statistical Association Proceeding of the Section on Survey Research Methods***,** 133-142.

Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I,
The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 537–557.

ON ELICITING EXPERT OPINION IN GENERALIZED LINEAR MODELS

Fadlalla G. Elfadaly[1] and Paul H. Garthwaite[2]
Department of Mathematics and Statistics,
The Open University, Walton Hall,
Milton Keynes, UK.MK7 6AA

[1]E-mail: f.elfadaly@open.ac.uk [2]E-mail: p.h.garthwaite@open.ac.uk

ABSTRACT

Suitable elicitation methods play a key role in Bayesian analysis of generalized linear models (GLMs) by obtaining and including expert knowledge as a prior distribution for the model parameters. Some elicitation methods for GLMs available in the literature focus mainly on logistic regression. A more general elicitation method of quantifying opinion about any GLM was developed in Garthwaite and Al-Awadhi (2006). The relationship between each continuous predictor and the dependant variable was modeled as a piecewise-linear function and each of its dividing points is accompanied with a regression coefficient. However, a simplifying assumption was made regarding independence between these coefficients, in the sense that regression coefficients were *a priori* independent if associated with different predictors. In this current research we relax the independence assumption between coefficients of different variables. In this case the variance-covariance matrix of the prior distribution is no longer block-diagonal. A method of elicitation for this more complex case is given and it is shown that the resulting covariance matrix is positive-definite. The method was designed to be used with the aid of interactive graphical software. It has been used in practical case studies to quantify the opinions of ecologists and medical doctors (Al-Awadhi and Garthwaite (2006); Garthwaite, Chilcott, Jenkinson, and Tappenden (2008)). The software is being revised and extended further in this research to handle the case of GLM with correlated pairs of covariates.

Keywords: Elicitation Methods; Expert Opinion; Assessment Task; Prior Distribution; Generalized Linear Model; Interactive Graphical Software.

1. INTRODUCTION

In many situations there is a substantial amount of information that is only recorded in the experience and knowledge of experts. To efficiently use this knowledge as an input to a statistical analysis, the experts must be asked meaningful questions whose answers determine a probability distribution. This process is referred to as *elicitation* and different forms of probability model require different elicitation methods. Reviews of methods of eliciting probability distributions are given in Garthwaite, Kadane, and O'Hagan (2005) and O'Hagan, Buck, Daneshkhah, Eiser, Garthwaite, Jenkinson, Oakley, and Rakow (2006).

Bayesian statistics offers an approach in which data and expert opinion are combined at the modeling stage, yielding probabilities that are a synthesis of the survey data and the expert's opinion. To incorporate expert opinion into a Bayesian analysis, it must be quantified as a prior distribution. In the work of interest here, this is accomplished through an interactive computer program that asks the expert to perform various assessment tasks. For the most part, the tasks require the expert to represent his opinion by drawing graphs on a computer screen, a judgemental problem that people perform relatively well because it is visual (Hammond (1971); Chaloner, Church, Louis, and Matts (1993); Garthwaite and Al-Awadhi (2006)).

Generalized linear models (GLMs) constitute a natural generalization of the classical linear models, where the linear predictor part is linked to the mean of the dependent variable through some link function. The distribution of the dependent variable is not necessarily assumed to be normal. The model is determined by the combination of the link function and the family of distributions to which the dependent variable belongs (see McCullagh and Nelder (1989) for an introduction to GLMs). Being very common in both frequentist and Bayesian data analysis, GLMs have attracted much research attention.

An important task in the Bayesian analysis of GLMs is to specify an informative prior distribution for model parameters. Suitable elicitation methods play a key role in this specification by obtaining and including expert knowledge as a prior distribution (see, for example, Bedrick, Christensen, and Johnson (1996) and O'Leary, Choy, Murray, Kynn, Denham, Martin, and Mengersen (2009)).

A method of quantifying opinion about a generalized linear model (GLM) was developed in Garthwaite and Al-Awadhi (2006). The method has been used to quantify the opinions of ecologists (Al-Awadhi and Garthwaite (2006)) and medical doctors (Jenkinson (2007); Garthwaite, Chilcott, Jenkinson, and Tappenden (2008)). The method makes a simplifying assumption regarding independence between the regression coefficients. One purpose of the current research work is to extend the elicitation method so that these assumptions are unnecessary. This will significantly increase the range of situations where the method is useful.

Available software written in Java by Jenkinson (2007) for expert opinion elicitation in GLMs is being modified. The software is interactive, requiring the expert to either type in assessments or plot points on graphs and bar-charts using interactive graphics. The previous version is available as an open source software at http://statistics.open.ac.uk/elicitation/. The current version of the software is meant to be more flexible in determining the options available for the user, especially for data input and results output. Some important modifications involve broadening the scope of the available models with different link functions, and giving the user many suggestions, help notices, questions, warning messages and directions aimed at making the software more interactive and easy to use for non-statistical experts. An executable stand-alone version of the software is available as a Java executable (jar) file. Another executable file format is available as a Windows executable file (with *.exe* extension). The software is aimed to be executable on any machine regardless of its operating systems and with no need for any other software packages. The most important modification in the current version is a new section for assessing expert knowledge about correlated covariates.

Section 2 of this paper gives a brief review of the main theoretical methods found in the literature for eliciting prior distributions for a GLM. Interactive computer software for these purposes is also listed with some of the different applications for which they have been used. In section 3, the piecewise-linear model of Garthwaite and Al-Awadhi (2006) is discussed together with the assessment tasks that the expert performs to quantify his opinion. The case of eliciting the covariance matrix when opinion about the covariates in a GLM does not permit the simplifying independence assumption is discussed in section 4, where assessment tasks needed for elicitation are also given. Some further discussion and concluding comments are given in section 5.

2. LITERATURE REVIEW

Relatively recent comprehensive reviews of eliciting probability distribution in its theory, methods, techniques, software, and applications and case studies are found in Garthwaite, Kadane, and O'Hagan (2005), O'Hagan, Buck, Daneshkhah, Eiser, Garthwaite, Jenkinson, Oakley, and Rakow (2006) and Jenkinson (2007).

Starting from the idea that it is more efficient and easier to elicit expert opinion about observable quantities, rather than about parameter values, Bedrick, Christensen, and Johnson (1996) were the first to elicit priors for some arbitrary generalized linear models. Their work switched from normal linear regression elicitation (Kadane, Dickey, Winkler, Smith, and Peters (1980); Garthwaite and Dickey (1988); Garthwaite and Dickey (1992)) into GLM. Their specification of informative prior distributions for the regression coefficients of a GLM is based on expanding the idea of conditional means priors (CMP). The expert is asked to give his assessment of the mean of potential observations conditional on given values at some carefully chosen points in the explanatory variable space. This information is used to specify a prior distribution at each location point. A prior distribution for the regression coefficient vector is then induced from the CMP. They gave examples where their method is applicable to common GLMs including normal and gamma regression models for which dispersion parameters are assumed to be known. The use of data augmentation priors (DAP) was also proposed for GLM. They showed that DAP, which gives priors in the same form as the likelihood at some prior data set, can be included by particular CMP's.

Garthwaite and Al-Awadhi (2006) developed an elicitation method for piecewise-linear logistic regression. The method is also valid for other GLMs and Garthwaite and Al-Awadhi (2009) extends the idea to GLMs with any link function. They assumed a multivariate normal distribution for the regression coefficients; its parameters can be determined from the expert assessments. The model was designed to be used with the aid of interactive graphical software written for this purpose. The software was used in a practical case study for threatened species in Al-Awadhi and Garthwaite (2006).

Dynamic graphical displays of probability distributions that can be freehand adjusted were used by Chaloner, Church, Louis, and Matts (1993) to help quantify opinion in the form of a prior distribution about regression coefficients in a proportional hazards regression model. In a clinical trial, prior distributions from five AIDS experts were elicited.

Interactive graphical software was given by Kynn (2005) to elicit expert opinion for the Bayesian logistic regression model. The software is called ELICITOR and appeared as an add-in to WinBUGS. Kynn extended a program written by Garthwaite (1990) and rewrote it in a more robust programming language. The software was originally developed as a user friendly tool for quantifying environmental experts' knowledge while studying the presence or absence of endangered species. It adopted the same approach of Al-Awadhi and Garthwaite (2006). For more details on ELICITOR see Kynn (2005); Kynn (2006) and O'Leary, Choy, Murray, Kynn, Denham, Martin, and Mengersen (2009), although the software and its documentation no longer seem to exist as an open source on the web.

Denham and Mengersen (2007) introduced a method and developed software to elicit expert opinion based on maps and geographic data for logistic regression models. Although eliciting information on observable quantities, such as values of the dependant variable at given values of the predictors, (referred to as structural procedure) is usually preferred and easier than direct assessment of the regression parameters (predictive procedure), Denham and Mengersen (2007) proposed a new approach that combines both strategies. In their combination approach, the expert may use either method simultaneously with each variable, according to his preference and background. They developed elicitation software under a Geographic Information System (GIS), in which design points were actual location on interactive maps. The software was applied in two case studies for modeling the median house prices in an Australian city and for predicting the distribution of an endangered species in Queensland. Although the software is specially designed for geographical data elicitation of a logistic regression model, they indicated that the concepts can be generalized to any GLM. Denham and Mengersen (2007) used the R language to code statistical functions, with Visual Basic and other software for interactive graphs embedded in the GIS system. The latter limits the usability of their software.

Jenkinson (2007) re-wrote the software of Garthwaite and Al-Awadhi (2006) in Java to provide a more transportable and stable version. He gave a detailed description and documentation to both the software and the piecewise-linear theoretical model behind it (Jenkinson (2007), p.215-251). Further modifications on the theoretical model and the software is given in this current paper in section 4.

A real medical application of the GLM elicitation software is given in a case study reported in Garthwaite, Chilcott, Jenkinson, and Tappenden (2008). Aiming to estimate the costs and benefits of current and alternate bowel cancer service in England, a pathway model was developed, whose transition parameters depend on covariates such as patient characteristics. Data to estimate some parameters was lacking and for these expert opinion was elicited, using the indicated software and under the assumption that the quantity of interest was related to covariates by the generalized piecewise-linear model given by Garthwaite and Al-Awadhi (2006). The assessments were then used to determine a multivariate normal distribution to represent the expert's opinions about the regression coefficients of that model. One conclusion of this work was that quantifying and using expert judgement can be acceptable in real problems of practical importance, provided that the elicitation is carefully conducted and reported in detail.

A thorough detailed comparison has been conducted by O'Leary, Choy, Murray, Kynn, Denham, Martin, and Mengersen (2009) for three relatively recent elicitation tools for logistic regression. The comparison included the interactive graphical tool of Kynn (2005) and Kynn (2006), the geographically assisted tool under GIS of Denham and Mengersen (2007) and a third simple direct questionnaire tool with no software. These tools were compared in an elicitation workshop (see O'Leary, Choy, Murray, Kynn, Denham, Martin, and Mengersen (2009) for more details on the third method). The paper discusses and gives a detailed description for each of the three methods used, showing advantages and disadvantages of each of them. Methods were compared according to their differences in the type of elicitation, the proposed prior model, the elicitation tool and the requirement of a facilitator to help the expert. Prior knowledge of two experts was elicited to model the habitat suitability of the endangered Australian brush-tailed rock-wallaby. The comparison revealed that the elicitation method influences the expert-based prior, to the extent that the three methods gave different priors for one of the experts. Some guidelines were also given for proper selection of the elicitation method. This work of O'Leary, Choy, Murray, Kynn, Denham, Martin, and Mengersen (2009) is part of a large body of applied research which shows the importance of eliciting expert knowledge when modeling rare event data, see also Kynn (2005); Al-Awadhi and Garthwaite (2006) or Choy, O'Leary, and Mengersen (2009).

Although they are interested mainly in designing the elicitation process in ecology, Choy, O'Leary, and Mengersen (2009) give a framework for statistical design of expert elicitation processes for informative priors which may be valid for Bayesian modeling in any field. The proposed design consists of six steps, namely, determining the purpose and motivation for using prior information; specifying the relevant expert knowledge available; formulating the statistical model; designing effective and efficient numerical encoding; managing uncertainty; and designing a practical elicitation protocol. Other important stages in the elicitation process may be found in Garthwaite, Kadane, and O'Hagan (2005), Jenkinson (2007) and Kynn (2008). Choy, O'Leary, and Mengersen (2009) validated these six steps in a detailed discussion and comparison of five case studies, revisiting the principles of successful elicitation in a modern context.

3. ELICITING A PRIOR DISTRIBUTION FOR GLMS

For quantifying expert's opinion about GLMs, Garthwaite and Al-Awadhi (2009) proposed a method to elicit opinion about the prior distribution of regression coefficients and its hyperparameters. This method, which will be referred to here as GA, is a generalization of the same authors' piecewise-linear model that they used for quantifying opinion for logistic regression (Garthwaite and Al-Awadhi (2006)).

In their work, the relationship between each continuous predictor variable and the dependant variable (assuming all other variables are held fixed) was modeled as a piecewise-linear function. Figure 1 illustrates a piecewise-linear relationship between the quantity of interest Y, and a continuous covariate "Weight"; the relationship correspondence to a sequence of straight lines that form a continuous line. Places where the slope of the line changes are refereed to as knots.

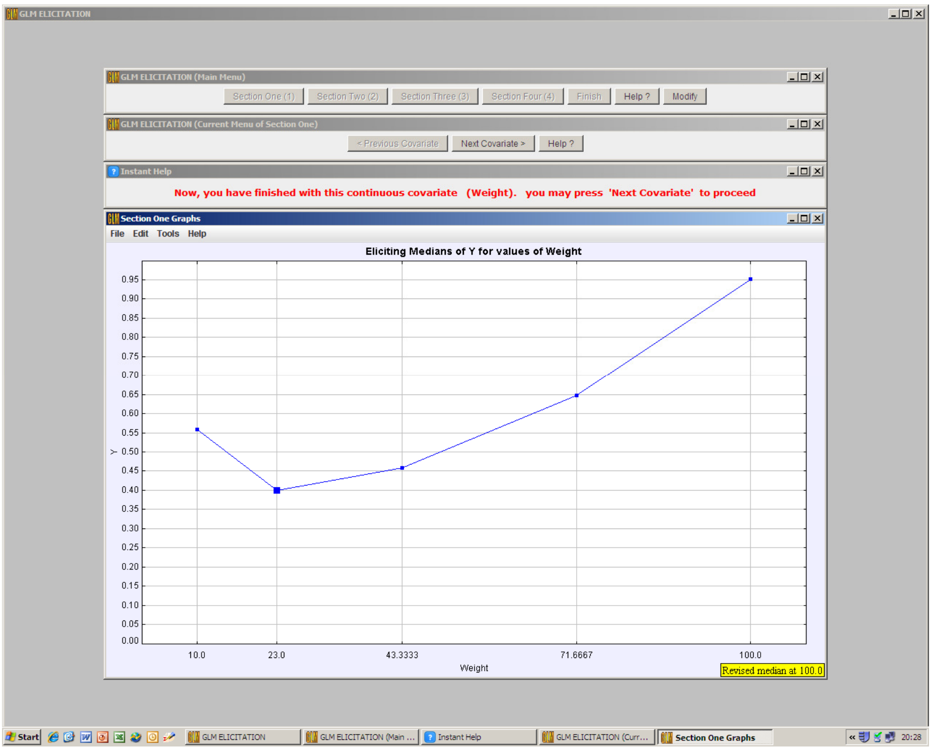


Figure 1: A piecewise-linear relationship given by median assessments

If a covariate is a factor, then its relationship with Y corresponds to a bar chart as in Figure 2, where X1 takes 4 levels Very large, Large, Normal and Small.

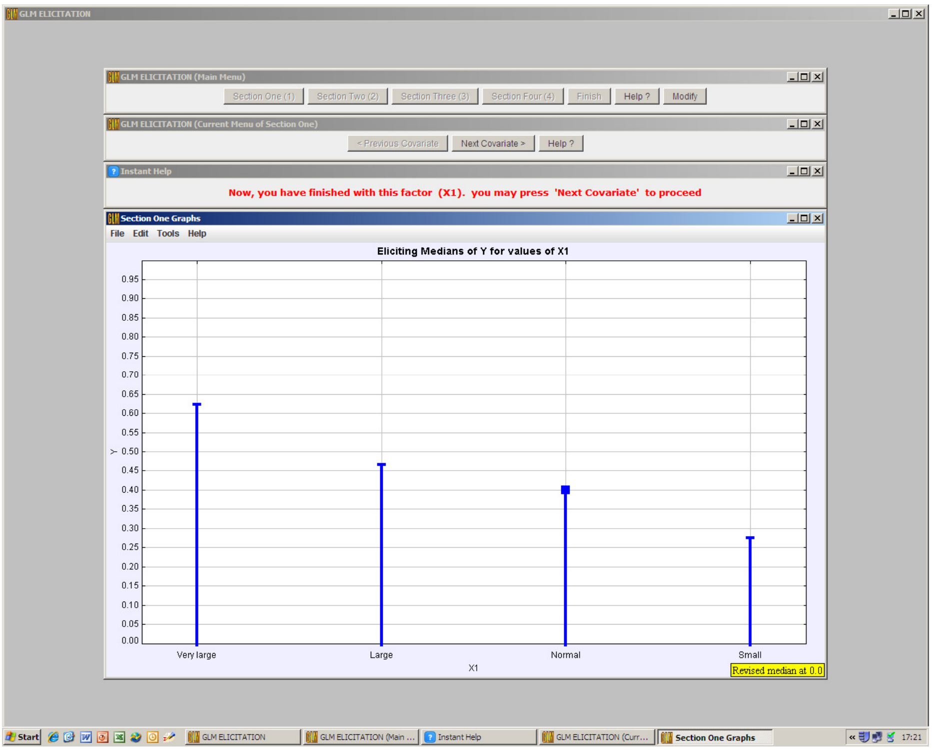


Figure 2: A bar chart relationship for a factor given by median assessments

The aim of elicitation is to quantify opinion about the slopes of the straight lines (for continuous variables) and the heights of the bars (for factors). In the GA method, a multivariate normal distribution was used to represent the prior knowledge about the regression coefficients. These coefficients were allowed to be dependant if associated with a single variable. A detailed discussion of their model is given hereafter.

Consider a response variable ζ, with m continuous covariates $R_1, R_2, \cdots, R_m$ and n categorical variables (factors) $R_{m+1}, R_{m+2}, \cdots, R_{m+n}$. Each variable R_i has $\delta(i)+1$ knots, $r_{i,0}, r_{i,1}, \cdots, r_{i,\delta(i)}$, where $r_{i,j-1} < r_{i,j}$ for $j = 1, 2, \cdots, \delta(i)$ and $i = 1, 2, \cdots, m+n$. These knots represents the dividing points of the piecewise relation for the continuous variables or levels for factors, with $r_{i,0}$ taken as the reference point of R_i.

For the response variable ζ, the expert is asked about its mean values given some points on the space of the explanatory variables, i.e about

$$\mu(r) = E(\zeta | R = r). \tag{1}$$

Let

$$Y = g[\mu(r)] = \alpha + \underline{\beta}_1' \mathbf{X}_1 + \underline{\beta}_2' \mathbf{X}_2 + \cdots + \underline{\beta}_{m+n}' \mathbf{X}_{m+n}, \tag{2}$$

where $g(.)$ is any monotone increasing link function, and

$$\mathbf{X}_i = (X_{i,1} \quad X_{i,2} \quad \cdots \quad X_{i,\delta(i)})', \qquad i = 1, 2, \cdots, m+n, \tag{3}$$

$$\underline{\beta}_i = (\beta_{i,1} \quad \beta_{i,2} \quad \cdots \quad \beta_{i,\delta(i)})', \qquad i = 1, 2, \cdots, m+n. \tag{4}$$

The relation between R_i and $\mathbf{X}_i$, for continuous covariates is that:

$$X_{i,j} = \begin{cases} 0 & \text{if } R_i \leq r_{i,j-1} \\ R_i - r_{i,j-1} & \text{if } r_{i,j-1} < R_i \leq r_{i,j} \\ d_{i,j} & \text{if } r_{i,j} < R_i, \end{cases} \tag{5}$$

for $i = 1, 2, \cdots, m$, and $j = 1, 2, \cdots, \delta(i)$, with $d_{ij} = r_{i,j} - r_{i,j-1}$.
For factors, $X_{i,j}$ are defined by:

$$X_{i,j} = \begin{cases} 1 & \text{if } R_i = r_{i,j} \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

for $i = m+1, m+2, \cdots, m+n$, and $j = 1, 2, \cdots, \delta(i)$.
Note that, if $R_i = r_{i,0}$, $\mathbf{X}_i$ are defined to be zero vectors for $i = 1, 2, \cdots, m+n$.

The prior distribution of α and $\underline{\beta} = (\underline{\beta}'_1 \quad \underline{\beta}'_2 \quad \cdots \quad \underline{\beta}'_{m+n})'$ is assumed to be multivariate normal with mean vector $\underline{b} = (b_0 \quad b_1 \quad \cdots \quad b_{m+n})'$ and variance-covariance matrix

$$\Lambda = \begin{pmatrix} \sigma_{0,0} & \underline{\sigma}'_1 \\ \underline{\sigma}_1 & \Sigma \end{pmatrix}, \tag{7}$$

where $\underline{b}$, $\sigma_{0,0}$, $\underline{\sigma}_1$ and Σ were estimated in GA. The needed assessment tasks in order to estimate these hyperparameters are given in details with the software description in the next subsection.

3.1 Assessment tasks and software description for GA method

The assessment procedure divides naturally into five stages, which are described in turn. A description of the method and theory for using the assessments to estimate the hyperparameters of the prior distribution is given in Garthwaite and Al-Awadhi (2006).

Defining the model The expert has different options for the model to be fitted, the choices available are ordinary linear regression, logistic regression, Poisson regression and any other user defined model. Ordinary linear regression assumes a normal distribution for the response variable with the identity link function. For the logistic regression the assumed distribution is Bernoulli with the logit link function. Poisson regression assumes a Poisson distribution with the logarithm link function.

The expert can choose to define any other model, in which case he will be asked to give a distribution and a link function. Available distributions are the normal, Poisson, binomial, gamma, inverse normal (inverse Gaussian), negative binomial, Bernoulli, geometric and exponential. The user is also asked for some parameters of the selected distribution where appropriate. Available link functions are canonical, identity, logarithm, logit, reciprocal, square root, probit, log-log, complementary log-log, power, log ratio and user defined link function. For a detailed definition of these link functions see McCullagh and Nelder (1989). For the power link function the software expects the exponent of the power function to be entered by the expert, a value of (-2) is suggested as a default. On choosing the distribution the software suggests the suitable canonical link function so as to help the expert (see Figure 3).

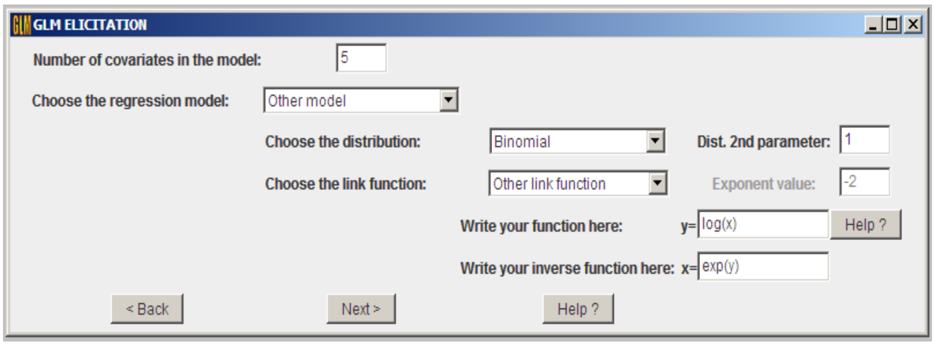


Figure 3: The dialogue box for defining the model

An important modification to the software is that it offers a large range of GLM's. It also lets the expert write his own link function and its inverse. The programm can parse both formulas and check their validity as mathematical expressions. Moreover, the program can help by checking whether the functions are valid inverses for each other.

Defining the response variable and covariates The expert determines the dependant variable with its minimum and maximum value in a dialogue box. The modified version of the software suggests the maximum and minimum values of the response variable whenever possible. The expert may still change them, but, in the light of the chosen model with the specified link function, invalid values are not accepted, and the expert is confronted with a warning message (For example, the range for a binomial proportion must not extend outside the interval (0,1)).

A set of explanatory variables (covariates) are chosen by the expert. Each covariate is treated as either a continuous random variable or a factor. Continuous covariates are specified with their minimum and maximum, factors are specified with their levels. For each continuous covariate, knots are chosen by the expert or suggested by the software. A reference point is chosen for each covariate, while the origin is the setting for which every covariate is at its reference point. After determining the number, names and types (continuous covariate or categorical factor) of the variables, the expert has only to give the maximum and minimum for each of his continuous covariates together with the value of its reference knot, and the modified software then suggests a suitable number of knots and the position of the reference knot relative to the other knots. The software can then divide the range and gives the value of each knot. This process is done automatically to reduce the burden of data entry, but, again, the expert can change any of these. The fractional part of each single numeric value is always being rounded to four decimal places, so as to avoid large decimal numbers which are not easily readable nor suitable for graph axis. If higher precision is to be used, measurement units can be modified to use data values of no more than four decimal places. For categorical factors, the expert gives the value of each level. In some cases, when the factor levels are ordinal data, for example, the expert may wish to keep the order of the factor levels, while still being able to select any level as the reference level. The current modification of the software gives an option to select the reference level of each factor without restricting it to be the first knot (see Figure 2).

Using a dialogue box, the median, lower and upper quartiles of Y at the origin are assessed. These values must be inside the previously specified range of the response variable; if not, the software warns the expert and asks him to resolve this conflict. In the expert's opinion, the true value of Y is equally likely to be bigger or smaller than the assessed median. Together with the median, these quartiles should divide the range into four equally likely intervals. The expert is encouraged to modify his median and quartile assessments until they divide the range into four intervals that each seem equally likely to him.

Medians assessments In the remainder of the elicitation procedure, the expert is separately questioned about each covariate in turn. He is asked to assume the other covariates are at their reference values/levels and forms a piecewise-linear graph or bar chart to represent his opinion about the remaining covariates.

The previous stage elicited the expert's median estimate of Y at the origin. The software plots this value on the reference vertical line and the expert is told to treat it as being correct. The expert then plots his median estimates of Y to form the remainder of the graph. He does this by using the computer mouse to 'click' points on the vertical lines. Straight lines are drawn by the computer between the 'clicked' points, which the expert can change until he feels the graph corresponds to his opinions.

As an illustration, Figure 1 shows a software graph for the variable "Weight". The horizontal axis gives values for the variable and the vertical axis gives values of Y. Thus the graph plots the effect on Y as the value of "Weight" varies. The experts is told that, if the graph is fairly flat, then the variable has less influence on Y than if the graph is more curved. The axes and vertical lines are drawn by the software.

For factors, bar charts are formed to represent the expert's opinion. The value of Y has been elicited earlier for the reference level and this gives the height of the reference bar. The expert is told to assume that this bar is correct and to judge the appropriate heights for other bars relative to it. These heights give the value of Y for each level when the other covariates are at their reference values/levels. The software draws thin vertical lines for each level and the expert specifies the height of a bar by clicking on the line with the mouse. This is illustrated in Figure 2 where all bars have been specified.

The expert could change an assessment by re-clicking on a line. These median assessments for the continuous covariates and factors yield estimates of the mean of the hyperparameter $\underline{\beta}$. Theoretical derivation of this estimation is given in details in Garthwaite and Al-Awadhi (2006).

Conditional medians assessments During this stage the expert is asked to assess his conditional medians for each covariate in turn. This is done by changing the conditioning value at the reference point from the median to the upper quartile. See Figure 4 in which median assessments made in the previous stage are given together with the upper quartile at the reference point. The expert assumes that the true value of Y at the reference point is the given upper quartile and he is asked to change the median values at other points in the light of this new conditioning value. Conditional medians for all values have been assessed by the expert in Figure 4.

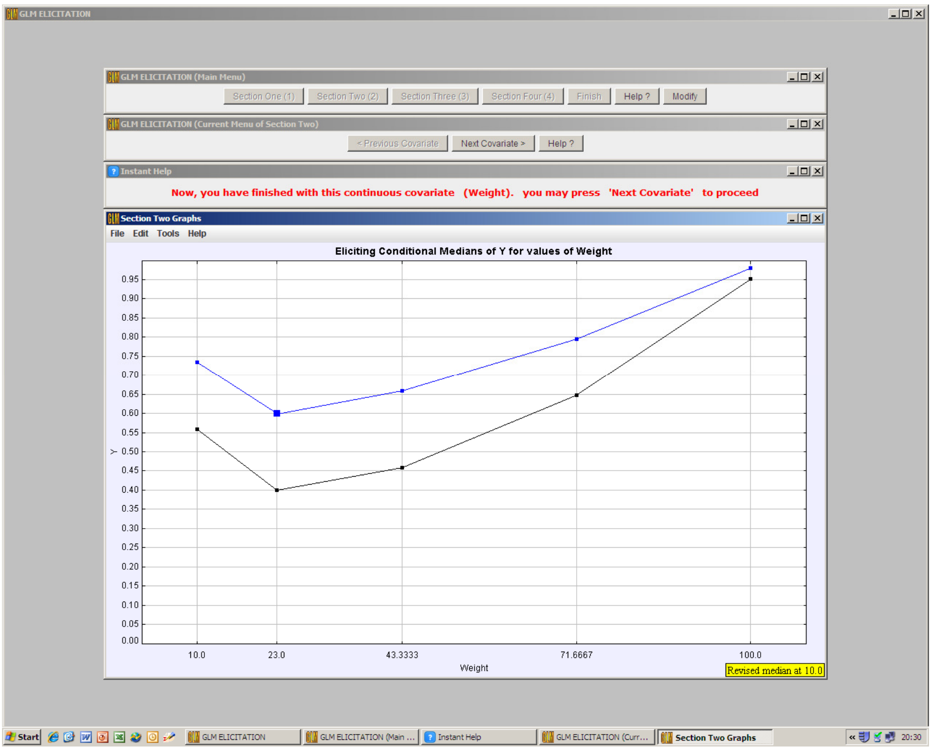


Figure 4: Conditional median assessments for the continuous covariate “Weight”

These assessments are needed to elicit part of the covariance matrix Λ, namely, the covariances between α and each of the components of $\underline{\beta}$. Suggested values of these conditional medians are given by the software, assuming that α and $\underline{\beta}$ components are independent. The expert can change these suggested values if he wishes.

Conditional quartiles assessments The median assessments provide point estimates of the relationship between different covariates and the variable Y. The remaining task is to quantify the expert's confidence in these estimates and their interrelationship. *i.e.* how accurate he believes the estimates to be and the correlations between them for each covariate individually. Correlations between coefficients of different covariates are estimated in the way proposed in section 4.

To achieve this, assessments of lower and upper quartiles are elicited. Assessing quartiles is a harder task for an expert than assessing medians, and quite a large number of quartile assessments are required. To assist the expert, the software suggests some quartile values by extrapolating from other quartile assessments of the expert. The theoretical procedure for getting these suggested values, given in Garthwaite and Al-Awadhi (2006) and Garthwaite and Al-Awadhi (2009), was programmed into the software to effectively help the expert during these two sections. The expert can change these assessments and commonly does so but, even then, a starting value to consider seems to make the task easier.

For each continuous covariate in turn, the software displays the graph of the medians that had been assessed earlier and then sets of *conditional* quartile assessments are elicited. For the first set of assessments, the condition is that the value of Y at the reference value/level equals the median assessment.

In an interactive graph like Figures 5 and 6, the expert is asked to give his lower and upper quartiles for Y at one point on each side of the medians for each value/level of the covariate except for the reference value/level. The lines joining quartiles look similar to confidence intervals and it is emphasized to the expert that there should only be a 50% chance that the value of Y is between the lines at any point. The expert uses the computer mouse to make assessments or change values suggested by the software.

For the second set of conditional assessments, the expert is asked to assume that the median estimates of Y are correct at both the reference value/level and the nearest points on each side of it. The expert gives lower and upper quartiles at another point and the software suggests quartiles for the remaining points. In Figure 7 lower quartiles have been assessed while upper quartiles are to be assessed. The expert modifies quartile values so as to represent his opinion, subject to the restriction that the current values must be within the previous set of quartile assessments. The idea is that as conditions increase, uncertainty should reduce. Garthwaite and Al-Awadhi (2006) showed that this condition guarantees that the covariance matrix of correlation coefficients is positive definite.

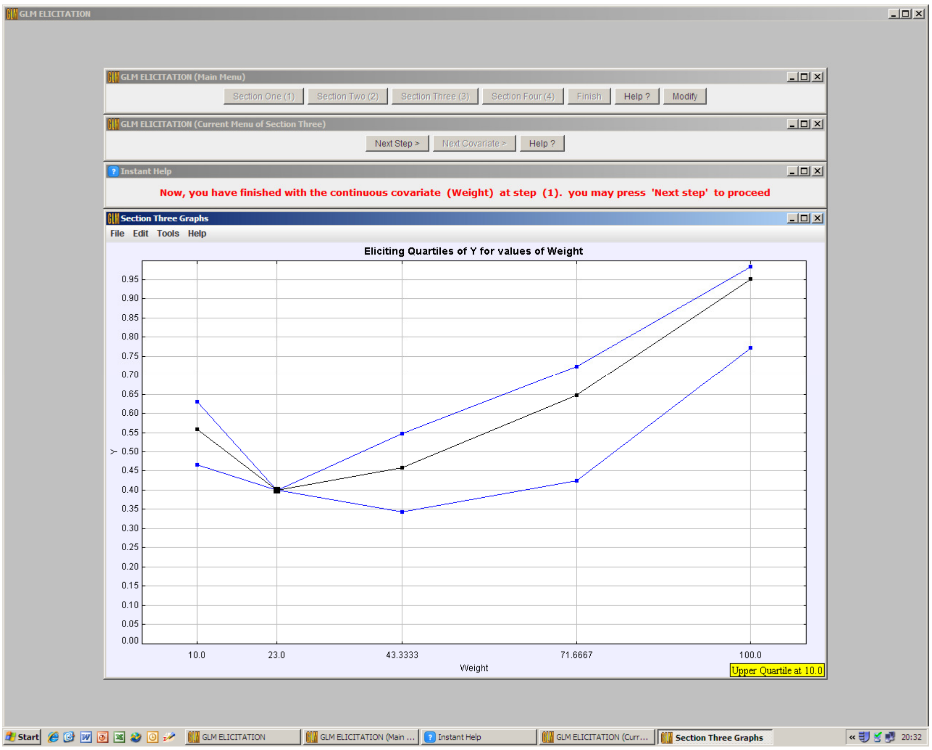


Figure 5: Quartile assessments for a continuous covariate

Figure 7 illustrates the graph formed at that stage. The two red lines (the outer lines) represent the previous set of quartile assessments, the second highest (black) line gives the median assessments, and the second lowest (blue) line joins the new lower quartile assessments. The black line joining the median at the right two bold points represents the condition that these medians should be treated as being correct. In assessing quartiles, the expert is told to consider the points to which he thinks the blue line may reasonably extend. Conditional assessments are also needed for factors. The software displays the bar chart that was formed during the assessment of medians. Conditional on the value of the bar at the reference level being correct, the expert assesses an upper and a lower quartile for other factor levels.

For each further set of conditional assessments, for both continuous covariates and factors, the expert is asked to assume that a further median given by another value/level was correct and to give his opinion about quartiles for the remaining values/levels. This is continued until the condition includes all but one of the values/levels at one side or one at both sides, when the expert gives his opinion about just the last one or two values/levels (see Figure 8).

As in other parts of the elicitation procedure, the expert uses the mouse to make assessments. Figure 8 illustrates the bar chart when conditioning values are specified (indicated by the solid squares); quartiles for the last level are marked with short horizontal blue lines

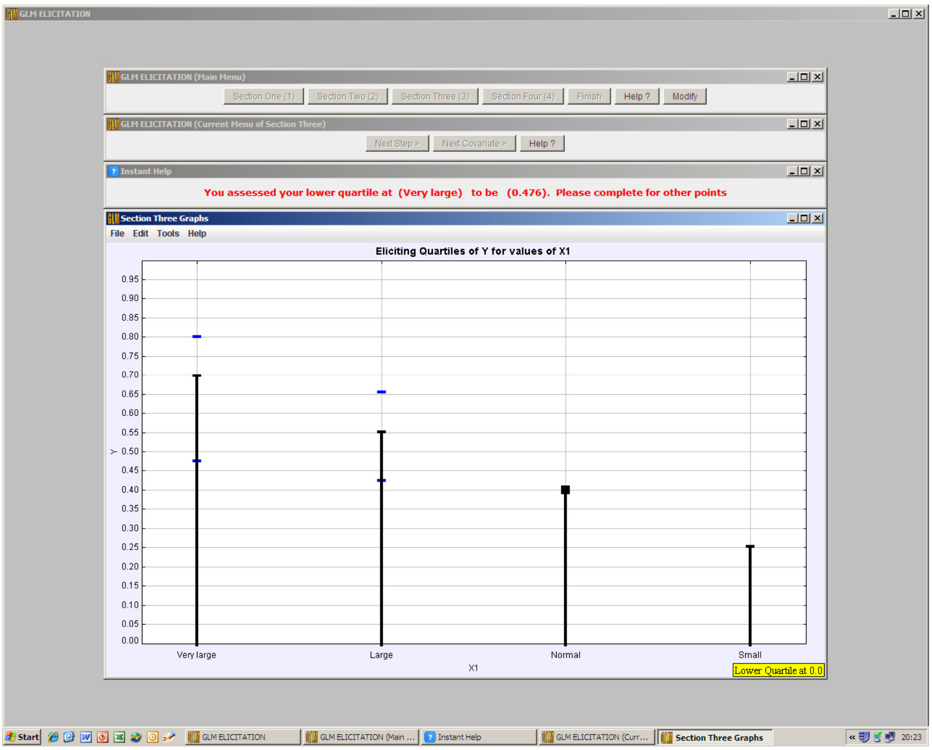


Figure 6: Quartile assessments for a factor

(the inner two lines), while the highest and lowest (red) lines represent the previous quartiles conditioning on fewer medians. Again, current conditional quartiles are not allowed to lay outside these red lines. The conditional assessments complete the elicitation procedure for the case of independent coefficients given by Garthwaite and Al-Awadhi (2006). The quartile assessments and conditional assessments yield estimates of the variance of the hyperparameter $\underline{\beta}$ which is assumed, under GA method to be a block-diagonal matrix Σ of the form

$$\Sigma = \begin{pmatrix} \Sigma_1 & O & \cdots & O \\ O & \Sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & \Sigma_{m+n} \end{pmatrix}. \tag{8}$$

Here, Σ_i is the covariance matrix of $\underline{\beta}_i = (\beta_{i,1} \quad \cdots \quad \beta_{i,\delta(i)})'$, $i = 1, 2, \cdots, m+n$.

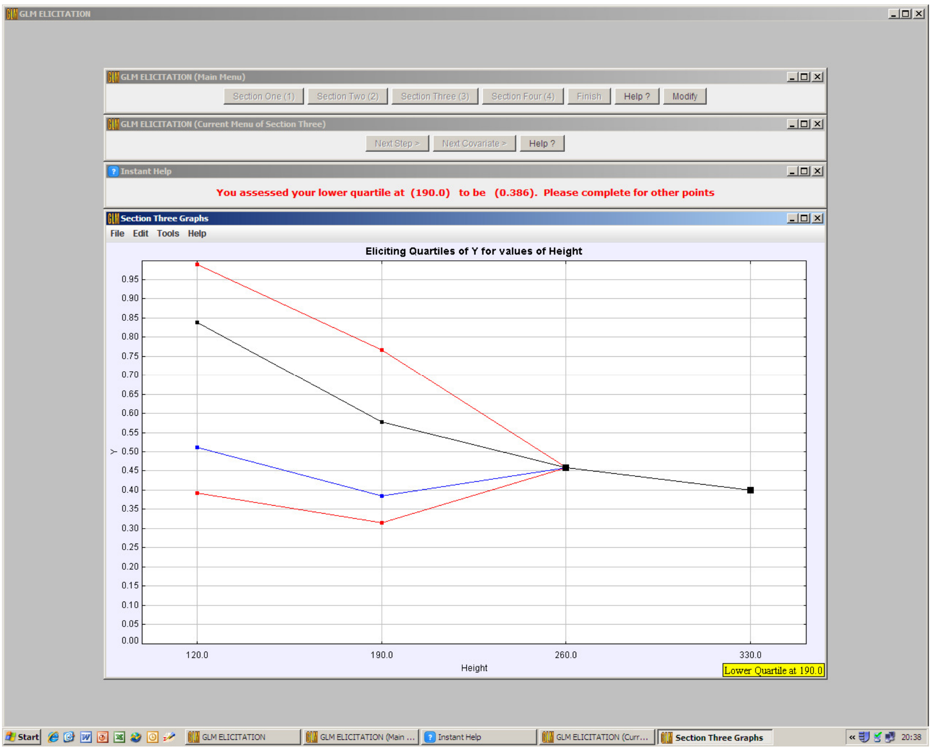


Figure 7: Assessing quartiles conditioning on two fixed points

4. THE CASE OF CORRELATED COEFFICIENTS

The previous block-diagonal structure of the covariance matrix assumes no interaction between any variables, in the sense that regression coefficients were *a priori* independent if associated with different covariates. Our aim now is to relax the independence assumption between the coefficients of different covariates. In fact, in many practical situations, it may be thought that regression coefficients of different variables should be related in the prior distribution, if the prior distribution is to give a reasonable representation of the expert's opinion. Although many of the correlations will be close to zero and can be represented by a zero correlation matrix, the expert may feel that some other covariates are strongly correlated and a correlation matrices should be elicited for them. The expert may be asked to determine which variables this applies to. Then, after assessing some more conditional quartiles, as will be described, GA's method of estimating the variance-covariance matrix can be generalized for this more complex case.

Instead of the previous block-diagonal structure of the matrix Σ in equation (8), it will be assumed that Σ can be conformally partitioned as

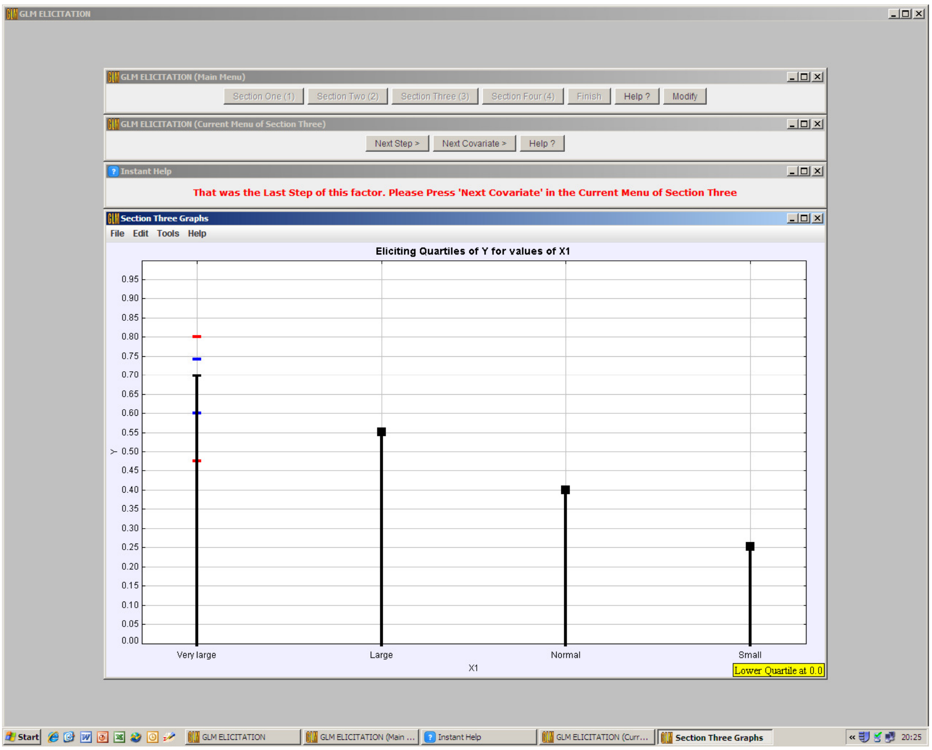


Figure 8: Assessing conditional quartiles for the last level of a factor

$$\Sigma = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \cdots & \Sigma_{1,m+n} \\ \Sigma_{2,1} & \Sigma_{2,2} & \cdots & \Sigma_{2,m+n} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m+n,1} & \Sigma_{m+n,2} & \cdots & \Sigma_{m+n,m+n} \end{pmatrix}, \tag{9}$$

where each submatrix $\Sigma_{i,i} = \Sigma_i$, $i = 1, 2, \cdots, m+n$, and the submatrices $\Sigma_{s,t}$ are not necessarily zero matrices for $s = 1, 2, \cdots, m+n$, $t = 1, 2, \cdots, m+n$ and $s \neq t$.

We estimate $\Sigma_{s,t}$ $(s \neq t)$ by generalizing the way Σ_i, $i = 1, 2, \cdots, m+n$ were estimated in GA. Each correlated pair of correlated vectors will be treated in turn. Assume that the expert sees that $\underline{\beta}_s$ and $\underline{\beta}_t$ are correlated. For $s < t$, we are trying to estimate the upper diagonal covariance submatrix $\Sigma_{s,t}$ of V,where,

$$V = Var(\underline{\beta}'_s \quad \underline{\beta}'_t)' \equiv \begin{pmatrix} \Sigma_{s,s} & \Sigma_{s,t} \\ \Sigma_{t,s} & \Sigma_{t,t} \end{pmatrix}. \tag{10}$$

As a variance-covariance matrix, V is symmetric, and hence $\Sigma_{t,s} = \Sigma'_{s,t}$.

In what follows we describe the assessment tasks that are used to elicit $\Sigma_{s,t}$ in the modified software. The idea is that further conditional quartiles must be assessed, where the set of

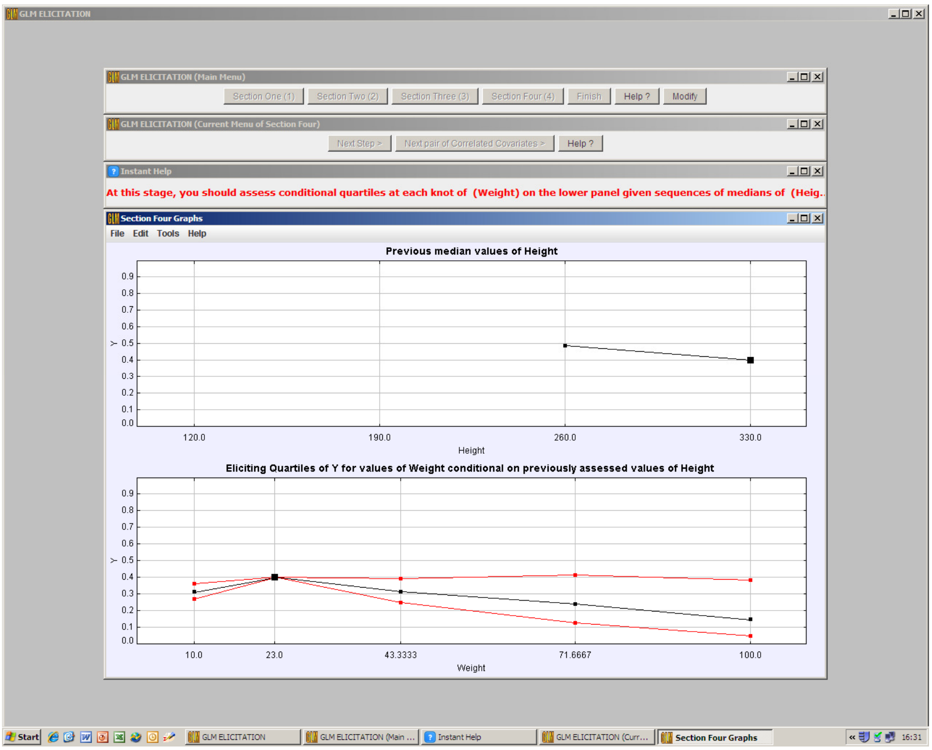


Figure 9: Assessments needed in the first phase for correlated covariates

conditions not only relate to knots of the same covariate, say $\underline{\beta}_t$, but also to knots of the other covariate, $\underline{\beta}_s$.

4.1 Assessment tasks and software description for correlated coefficients

The modified software elicits the expert's assessments of conditional quartiles that are needed to estimate the covariance matrix of correlated pair of covariates. The expert is asked whether he sees dependence between the regression coefficients of any pair of covariates. If not, he may go directly to the last stage to get the results. If yes, he will be asked to say which two variables have such dependence and then he will get a panel of two simultaneous graphs (see Figure 9 or Figure 10).

The upper graph is for one variable of the correlated pair. It shows the previously assessed median values for that variable and these are used as conditioning values for assessing quartiles of the other variable. The expert is asked to assume that these median values are correct. That is, they are accurate estimates of the mean response for the specified covariate values. Conditional on this information, the expert clicks on the lower interactive graph to assess new conditional quartile values given the median values shown in the upper graph.

The procedure consists of two phases; in the first phase the expert assesses quartile values for the variable in the lower graph given sets of medians for the variable in the upper graph.

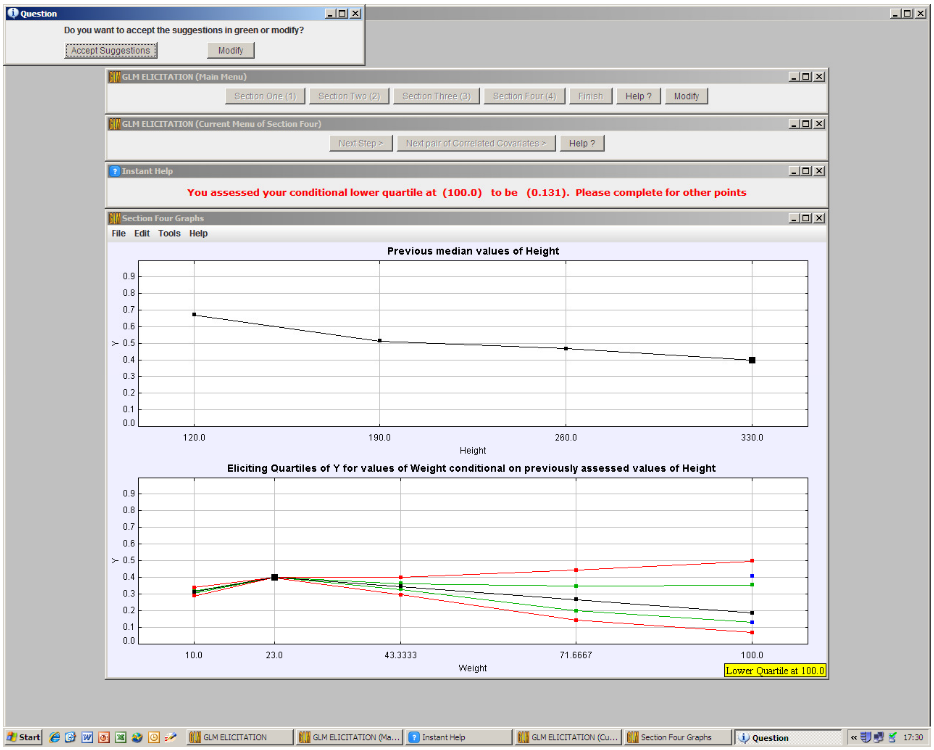


Figure 10: Assessments needed in the second phase for correlated covariates

The set of conditioning values of the first variable in the upper graph are incremented by one extra value at each further step. The expert is asked to re-assess conditional quartiles at each step. The initial step of the first phase is shown in Figure 9, where the expert is asked to assess conditional quartiles for different knots of the "Weight" variable in the lower graph conditioning on the previously assessed medians of the "Height" variable at its reference and only one other knot which are connected by the rightmost (black) line in the upper graph. The conditioning set includes also the median of the "Weight" variable at its reference knot (23.0).

The second phase starts after conditioning on all median values at all knots in the top graph. Each further step in this phase adds an extra median value from the lower graph to the conditioning set. Further conditional quartiles are assessed in the lower graph.

This phase is very similar to the assessment of conditional quartiles in the GA method where incremented sets of medians of the same variable are used as conditioning sets for assessing conditional quartiles. However, in this phase previously assessed median values at knots for a different variable are also taken into consideration when assessing conditional quartiles.

One of the steps of the second phase is shown in Figure 10. In this step, the expert is asked to assess conditional quartiles for different knots of the "Weight" variable in the lower graph while conditioning on the previously assessed medians of the "Height" variable at all of its four knots that are connected by the black line in the upper graph. The conditioning set includes also the median of the "Weight" variable at its reference knot (23.0). Suggested conditional quartiles are computed by extrapolating from other quartile assessments in the same manner as in GA method. The middle (green) lines in the lower graph in Figure 10 represent these suggested values.

On finishing all phases of the assessment for this pair of explanatory variables, the user is asked about other correlated pairs, and the process starts again for the new pair, if any. The modified software outputs data in three different files, one containing the basic setup data, the second containing all assessments made by the expert, and the third containing the resulting mean vector and covariance matrix of the hyperparameter vector for further Bayesian analysis. More graphical output is also being considered to give the expert useful feedback.

5. CONCLUSION AND FUTURE WORK

The number of conditional quartiles assessed here is sufficient for estimating the covariance matrix Σ when off-diagonal submatrices are not necessarily all zero. The theoretical details of the method proposed here will be given in a separate paper (in preparation). Conditions needed to guarantee positive definiteness of the matrix V have been investigated. Results show that the only substantive constraint needed is that the additional condition at each further step must reduce the value of the conditional variance. The expert's uncertainty must therefore reduce as the elicitation process progresses. This means that his assessed interquartile ranges must be less than his previous assessments. To fulfil this theoretical requirement of the estimated covariance matrix, red lines are shown to the expert at each step and his new quartile assessments must be within these red lines (see Figure 10 or Figure 10).

Although each 2×2 covariance matrix V has been shown to be positive-definite, some extra conditions must be imposed for the whole covariance matrix Σ to be positive-definite. Further work is still being done to investigate such conditions. One of the possible drawbacks of this proposed elicitation method is that the number of conditional quartiles needed to be assessed by the expert would become uncomfortably large if many pairs of covariates are thought to be correlated.

Another method is being examined to elicit the off-diagonal covariance matrices using a small number of coefficients that reflect the pattern of correlation of each pair. These coefficients reduce the number of assessments needed from the expert and can be used to induce all the elements of the covariance matrix. Moreover, under suitable conditions, the whole resulting covariance matrix Σ is positive-definite. Although this method has not been programmed yet, a very brief description of the assessments follows.

For a pair of correlated vectors of coefficients $(\underline{\beta}_s, \underline{\beta}_t)$, the expert will be asked to determine the conditional median of $\underline{\beta}_t$ given a specific value of $\underline{\beta}_s$. The new feature here is that the expert can only change the vertical position of the piecewise-linear curve of $\underline{\beta}_t$ given a fixed vertical change of $\underline{\beta}_s$, no changes of the single slopes at any knots are allowed. For $n > 2$ correlated vectors of coefficients, the process will consist of $k-1$ steps, at the i^{th} step, the expert will be asked to assess the conditional median of $(\underline{\beta}_k|\underline{\beta}_1, \underline{\beta}_2, \cdots, \underline{\beta}_i)$ given a set of i graphs, each of which shows a change with a different fixed value for each $\underline{\beta}_j$, $j = 1, 2, ..., i$. It can be shown that the full covariance matrix Σ is positive definite under a set of $n(n-1)/2$ conditions. In which case, the expert will be asked to assess a number of $n(n-1)/2$ values. Conditions needed for positive definiteness can be translated into allowable ranges shown to the expert on the interactive graph and he will be asked to restrict his assessments so that conditional medians lie inside these ranges.

REFERENCES

Al-Awadhi, S. A., and P. H. Garthwaite (2006): "Quantifying Expert Opinion for Modelling Fauna Habitat Distributions," *Computational Statistics*, 21, 121–140.

Bedrick, E. J., R. Christensen, and W. Johnson (1996): "A New Perspective on Priors for Generalized Linear Models," *Journal of the American Statistical Association*, 91, 1450–1460.

Chaloner, K., T. Church, T. A. Louis, and J. P. Matts (1993): "Graphical Elicitation of a Prior Distribution for a Clinical Trial," *Statistician*, 42, 341–353.

Choy, S. L., R. O'Leary, and K. Mengersen (2009): "Elicitation by Design in Ecology: Using Expert Opinion to Inform Priors for Bayesian Statistical Models," *Ecology*, 90, 265–277.

Denham, R., and K. Mengersen (2007): "Geographically Assisted Elicitation of Expert Opinion for Regression Models," *Bayesian Analysis*, 2, 99–136.

Garthwaite, P. H. (1990): "Implementation, user guide and computer programme for a method of eliciting opinion about a regression model in variable-selection problems," Discussion Paper 12, Department of Mathematical Sciences, University of Aberdeen.

Garthwaite, P. H., and S. A. Al-Awadhi (2006): "Quantifying Opinion About a Logistic Regression Using Interactive Graphics," Discussion Paper 06/07, Statistics Group, The Open University, UK.

——— (2009): "Quantifying Subjective Opinion About Generalized Linear and Piecewise-linear Models," In preparation.

Garthwaite, P. H., J. B. Chilcott, D. J. Jenkinson, and P. Tappenden (2008): "Use of Expert Knowledge in Evaluating Costs and Benefits of Alternative Service Provisions: A Case Study," *International Journal of Technology Assessment in Health Care*, 24, 350–357.

Garthwaite, P. H., and J. M. Dickey (1988): "Quantifying Expert Opinion in Linear Regression Problems," *Journal of the Royal Statistical Society, Series B*, 50, 462–474.

——— (1992): "Elicitation of Prior Distributions for Variable Selection Problems in Regression," *Annals of Statistics*, 20, 1697–1719.

Garthwaite, P. H., J. B. Kadane, and A. O'Hagan (2005): "Statistical Methods for Eliciting Probability Distributions," *Journal of the American Statistical Association*, 100, 680–701.

Hammond, K. R. (1971): "Computer Graphics as an aid to Learning," *Science*, 172, 903–908.

Jenkinson, D. J. (2007): "Quantifying Expert Opinion as a Probability Distribution," Ph.D. thesis, Open University, UK.

Kadane, J. B., J. M. Dickey, R. Winkler, W. Smith, and S. Peters (1980): "Interactive Elicitation of Opinion for a Normal Linear Model," *Journal of the American Statistical Association*, 75, 845–854.

Kynn, M. (2005): "Eliciting Expert Knowledge for Bayesian Logistic Regression in Species Hapitat Modelling in Natural Resources," Ph.D. thesis, Queensland University of Technology, Australia.

——— (2006): "Designing ELICITOR: Software to Graphically Elicit Priors for Logistic Regression Models in Ecology," Available at http://www.winbugs-development.org.uk/elicitor/files/designing.elicitor.pdf [Accessed 10th October 2008].

——— (2008): "The 'Heuristtics and Biases' Bias in Expert Elicitation," *Journal of the Royal Statistical Society, Series A*, 171, 239–264.

McCullagh, P., and J. A. Nelder (1989): *Generalized Linear Models*. Chapman and Hall, London, second edn.

O'Hagan, A., C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow (2006): *Uncertain Judgements: Eliciting Expert Probabilities*. John Wiley, Chichester.

O'Leary, R. A., S. L. Choy, J. V. Murray, M. Kynn, R. Denham, T. G. Martin, and K. Mengersen (2009): "Comparison of Three Expert Elicitation Methods for Logistic Regression on Predicting the Presence of the Threatened Brush-tailed Rock-wallaby *Petrogalae peicillata*," *Environmmetrics*, 20, 379–398.

List of Contributors

University of Botswana
Natal.Ayiga@mopipi.ub.bw

Afzalina.Azmee 255
Department of Probability & Statistics
University of Sheffield
Afzalina.Azmee@sheffield.ac.uk

Ahmed Badr 461
Economic Issues Program (EIP),
Information and Decision Support Center
Egyptian Cabinet.
amabadr@idsc.net.eg

Adam Baharum 483
School of Mathematical Sciences,
Universiti Sains Malaysia
adam@cs.usm.my

K.E. Basford 1
The University of Queensland, School of
Land, Crop and Food Sciences,
Australian Centre for Plant Functional
Genomics, Brisbane 4072, Australia
k.e.basford@uq.edu.au

Jan Beirlant 2
University Center of Statistics,
Katholieke Universiteit Leuven
Goedele Dierckx, HUBrussel
jan.beirlant@wis.kuleuven.be

Abdelhafid Belarbi 805
Faculty of Economics and Administrative
Sciences,
Al-Zaytoonah University of Jordan,
P.O.Box 130, Amman 11733, Jordan

Anil K. Bera 207
Department of Economics, University of
Illinois, 1407 W. Gregory Drive,
Urbana, IL 61801
abera@illinois.edu

Jim Berger 2
Department of Statistical Science
Duke University, Durham,
NC 27708-0251, USA
Statistical and Applied Mathematical
Sciences Institute , P.O. Box 14006,
Research Triangle Park,
Durham, NC 27709-4006
berger@samsi.info, berger@stat.duke.edu

Paul Bigala 7
Population Studies and Demography
North West University (Mafikeng Campus)
South Africa
paulgigs@yahoo.com

Atanu Biswas 205
Applied Statistics Unit, Indian Statistical
Institute
203 B. T. Road, Kolkata 700 108, India
atanu@isical.ac.in

Chafik Bouhaddioui 494
Department of Statistics,
United Arab Emirates University,
Al Ain, UAE.
ChafikB@uaeu.ac.ae

Jennifer Bremer 4
Public Policy and Administration
Department, School of Public Affairs,
The American University in Cairo,
P.O.Box 74, New Cairo 11835, Egypt
jbremer@aucegypt.edu

Matthew John Burstow 98, 106
Department of Surgery, Ipswich Hospital,
Queensland, Australia
mjburstow@gmail.com

Manisha Chakrabarty 246
Indian Institute of Management,
Calcutta, India

Asis Kumar Chattopadhyay 264
Department of Statistics, Calcutta
University, India
akcstat@caluniv.ac.in

Cedex 09, France
fraix@obs.ujf-grenoble.fr

May Gadallah 584
Department of Statistics,
Faculty of Economics and Political Science,
Cairo University, Cairo, Egypt
mayabaza@hotmail.com

Hesham F. Gadelrab 607
Mansoura University, Faculty of Education,
Psychology Department
Mansoura, Egypt 35516
The British University in Egypt (BUE),
Business Administration Department
Sherouk City, Cairo,
Postal No. 11837, P.O. Box 43
heshfm@mans.edu.eg,
hesham.gadelrab@bue.edu.eg

Antonio F. Galvao Jr. 207
Department of Economics,
University of Wisconsin-Milwaukee,
Bolton Hall 852, 3210 N. Maryland Ave.,
Milwaukee, WI 53201
agalvao@uwm.edu

Paul H. Garthwaite 537
Department of Mathematics and Statistics,
The Open University, Walton Hall,
Milton Keynes, MK7 6AA, UK
p.h.garthwaite@open.ac.uk

Ronald Geskus 247
Department of Clinical Epidemiology,
Biostatistics and Bioinformatics
Academic Medical Center
Meibergdreef 15
1105 AZ, Amsterdam, The Netherlands
R.B.Geskus@amc.uva.nl
E.Hogervorst@lboro.ac.uk

Malay Ghosh 157, 159
Department of Statistics, University of
Florida, Gainesville, FL 32611
ghoshm2000@yahoo.com

Pulak Ghosh 119
Department of Quantitative sciences, Indian
Institute of Management,
Bangalore, India
pulakghosh@gmail.com

Suryo Guritno 1079
Mathematics Department,
Gadjah Mada University, Yogyakarta,
Indonesia
Guritno0@mailcity.com

Hyung-Tae Ha 634
Department of Applied Statistics,
Kyungwon University, Sungnam-ci,
Kyunggi-do South Korea, 461-701
htha@kyungwon.ac.kr

Saleha Naghmi Habibullah 646
Kinnaird College for Women, Lahore,
Pakistan
salehahabibullah@hotmail.com

Ali S. Hadi 6, 424
Department of Mathematics and Actuarial
Science, The American University in Cairo,
Egypt, and Department of Statistical
Science, Cornell University, USA.
ahadi@aucegypt.edu

Ramadan Hamed 652
Faculty of Economics and Political Science,
Cairo University, Cairo, Egypt
ramadanh@aucegypt.edu

Tengku-Aizan Hamid 17
Institute of Gerontology
Universiti Putra Malaysia, Malaysia
tengkuaizan06@gmail.com

Muhammad Hanif 357, 664, 676
Lahore University of Management Sciences,
Lahore, Pakistan
hanif@lums.edu.pk

Muna F. Hanoon 805
Faculty of Economics and Administrative Sciences,
Al-Zaytoonah University of Jordan,
P.O.Box 130, Amman 11733, Jordan

Toni Hartono 70
National Commission for Older Persons
Indonesia

Siti Rahayu Mohd. Hashim 685
Department of Probability and Statistics,
Hicks Building,
Hounsfield Road, S3 7RY, University of Sheffield, UK.
stp08sm@sheffield.ac.uk

Anwar Hassan 693
PG Department of Statistics
University of Kashmir, Srinagar-India
anwar.hassan5@gmail.com,
anwar.hassan2007@gmail.com

Siti Fatimah Hassan 305
Centre for Foundation Studies in Science,
Universiti of Malaya,
50603 Kuala Lumpur, Malaysia.

Christian Heumann 1094
Department of Statistics,
Ludwig Strasse 33, 80539.
Ludwig-Maximilians University Munich,
Germany
christian.heumann@stat.uni-muenchen.de

Rafiq H. Hijazi 701
Department of Statistics
United Arab Emirates University
P. O. Box 17555, Al-Ain, UAE
rhijazi@uaeu.ac.ae

Eef Hogervorst 70
Department of Human Sciences
Loughborough University, UK

Zahirul Hoque 205, 206
Department of Statistics
College of Business and Economics
United Arab Emirates University
zahirul.hoque@uaeu.ac.ae

Md Belal Hossain 132, 140, 148
Department of Mathematics and Computing,
Australian Centre for Sustainable Catchments,
University of Southern Queensland,
Toowoomba, Queensland, Australia
hossainm@usq.edu.au

Shereen Hussein 25
Social Care Workforce Research Unit King's College London Melbourne House, 5th Floor Strand, London, UK, WC2R 2LS
shereen.hussein@kcl.ac.uk

Osama Abdelaziz Hussien 708, 733
Department of Statistics,
Faculty of Commerce,
Alexandria University, Egypt
osama52@gmail.com,
ossama.abdelaziz@alexcommerce.edu.eg

Abdul Ghapor Hussin 303, 305
Centre for Foundation Studies in Science,
Universiti of Malaya,
50603 Kuala Lumpur, Malaysia.
ghapor@um.edu.my

K. Ibrahim 759
School of Mathematical Sciences,
Universiti Kebangsaan Malaysia,
Selangor, Malaysia
Kamarulz@ukm.my

Noor Akma Ibrahim 747
Institute for Mathematical Research
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor
Malaysia
nakma@putra.upm.edu.my

Thomas Mathew 120
Department of Mathematics and Statistics,
University of Maryland Baltimore County,
Baltimore, Maryland 21250, USA
mathew@umbc.edu

Christine McDonald 333
Department of Mathematics and Computing
University of Southern Queensland
Toowoomba, Qld 4350, AUSTRALIA
Christine.McDonald@usq.edu.au

Dean E. McLaughlin 265
Keele University, UK
dem@astro.keele.ac.uk

P. J. McLellan 1038
Department of Chemical Engineering
Queen's University, Kingston, Ontario,
Canada, K7L 3N6
mclellnj@chee.queensu.ca

Ahmed Zogo Memon 646
National College of Business Administration
& Economics
Lahore, Pakistan

Breda Memon 98, 106, 140, 148
Department of Surgery, Ipswich Hospital,
Queensland, Australia
bmemon@yahoo.com

Muhammed Ashraf Memon 98, 106, 121, 132, 140, 148
Department of Surgery, Ipswich Hospital,
Queensland, Australia
Department of Surgery, University of
Queensland, Herston, Queensland, Australia
Faculty of Medicine and Health Sciences,
Bond University, Gold Coast, Queensland,
Australia, Faculty of Health Science, Bolton
University, Bolton, Lancashire, UK
mmemon@yahoo.com

Sherzod M. Mirakhmedov 881
Ghulam Ishaq Khan Institute of Engineering
Sciences & Technology,
Topi-23460,Swabi ,NW.F.P. Pakistan
shmirakhmedov@yahoo.com

Saidbek S. Mirakhmedov 881
Institute of Algoritm and Engineering,
Fayzulla Hodjaev-45, Tashkent -700149.
Uzbekistan
saeed_0810@yahoo.com

Ibrahim Mohamed 303
Institute of Mathematical Sciences,
University of Malaya, 50603 Kuala Lumpur,
Malaysia
imohamed@um.edu.my

Saptarshi Mondal 264
Department of Statistics, Calcutta
University, India

Gabriel V. Montes-Rojas 207
Department of Economics, City University
of London, 10 Northampton Square, London
EC1V 0HB, U.K
Gabriel.Montes-Rohas.1@city.ac.uk

Ghada Mostafa 898
Central Agency For Public Mobilization and
Statistics
Salah Salem St. Nasr City
ghadaabd@yahoo.com

G. M. Nair 920
School of Mathematics and Statistics,
The University of Western Australia,
Perth, WA 6009,
Australia
gopal@maths.uwa.edu.au

Kenneth Nordstrom 120
Department of Mathematical Sciences,
University of Oulu, Finland

D. Nur 920
School of Mathematical and Physical
Sciences,
The University of Newcastle
Callaghan, NSW 2308, AUSTRALIA
Darfiana.Nur@newcastle.edu.au

Y. Nurizan 17
Institute of Gerontology
Universiti Putra Malaysia, Malaysia

Teresa Azinheira Oliveira 82
CEAUL and DCeT, Universidade Aberta,
rua Fernão Lopes nº9, 2ºdto, 1000-132
Lisboa, Portugal
toliveir@univ-ab.pt
nurizan@putra.upm.edu.my

M. F. Omran 921
Business School, Nile University, Egypt
mfomran@nileuniversity.edu.eg

Emma Osland 121, 132
Dept of Nutrition and Dietetics, Ipswich
Hospital, Ipswich, Queensland, Australia
Department of Mathematics and Computing,
Australian Centre for Sustainable
Catchments, University of Southern
Queensland, Toowoomba, Queensland,
Australia
Emma_Osland@health.qld.gov.au

Magued Osman 4
Chairman, Information and Decision
Support Center
The Egyptian Cabinet
magued_osman@idsc.net.eg

G. N. Osuafor 931
Department of Statistics and Demography,
University of the Western Cape,
X17 7535 Bellville, South Africa
gnosuafor@gmail.com

Pinakpani Pal 206
Indian Statistical Institute, Calcutta, India
pinak@isical.ac.in

Sung Y. Park 207
Department of Economics, University of
Illinois,
1407 W. Gregory Drive, Urbana, IL 61801,
and The Wang Yanan Institute for Studies in
Economics,
Xiamen University, Xiamen, Fujian 361005,
China
sungpark@sungpark.net

Ajoy Paul 246
Bidhan Nagar Govt. College

Koay Swee Peng 483
School of Mathematical Sciences,
Universiti Sains Malaysia

Danny Pfeffermann 186
Southampton Statistical Sciences Research
Institute, University of Southampton, SO17
1BJ UK, Department of Statistics, Hebrew
University of Jerusalem, 91905, Israel
msdanny@soton.ac.il

H. N. Phua 817
School of Mathematical Sciences,
Universiti Sains Malaysia, Penang, Malaysia

Hasih Pratiwi 950
Sebelas Maret University, Surakarta,
Indonesia
Gadjah Mada University, Yogyakarta,
Indonesia
hasihpratiwi@ymail.com

Tri Budi W. Rahardjo 70
Center for Ageing Studies Universitas
Indonesia
Center for Health Research Universitas
Indonesia

National Commission for Older Persons
Indonesia
tri.budi.wr@gmail.com

Mohamed Ramadan 652
Population Council WANA Regional Office,
59 Misr Agricultural Road, Maadi, Cairo, Egypt
mramadan@popcouncil.org

Omar Rouan 341
GREDIM, Ecole Normale Supérieure
Marrakech- Maroco
omarrouan@gmail.com

Umu Sa'adah 1079
Mathematics Department,
Gadjah Mada University, Yogyakarta, Indonesia,
Mathematics Department,
Brawijaya University, Malang, Indonesia
umusaadah@yahoo.com

Ishmael Kalule-Sabiti 7
North West University (Mafikeng Campus)
South Africa
Ishmael.KaluleSabiti@nwu.ac.za

Asep Saefuddin 248, 1087
Department of Statistics, Faculty of Mathematics and Science,
IPB, 16680, Darmaga, Bogor, Indonesia
asaefuddin@gmail.com

Yasmin H. Said 963
Department of Computational and Data Sciences, George Mason University,
Fairfax, VA, USA
ysaid99@hotmail.com

Mohamed Saleh 461
Cairo University, Egypt
University of Bergen, Norway
saleh@salehsite.info

Mamadou-Youry Sall 977
Unit of Formation and Research in Economic Sciences and Management at Gaston Berger University, Saint-Louis, Senegal, BP 234
sallmy@ufr-seg.org

Mohd Sahar Sauian 989
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA (UiTM),
40450 Shah Alam, Malaysia
mshahar@tmsk.uitm.edu.my

Hussein Abdel-Aziz Sayed 998
Faculty of Economics, Cairo University
husseinsayed@hotmail.com

Kamal Samy Selim 1030
Department of Computational Social Sciences,
Faculty of Economics and Political Science
Cairo University, Cairo, Egypt
kselim9@yahoo.com

Muhammad Qaiser Shahbaz 676
Department of Mathematics,
COMSATS Institute of Information Technology,
Lahore, Pakistan
qshahbaz@gmail.com

Qi-Man Shao 80, 97
Department of Mathematics
Hong Kong University of Science and technology
Clear Water Bay, Kowloon
Hong Kong, China
maqmshao@ust.hk

M. E. Sharina 264, 293
Special Astrophysical Observatory, Nizhnij Arkhyz,
Zelenchukskiy region, Karachai-Cherkessian Republic, Russia 369167
sme@sao.ru

Furrukh Shehzad 367
National College of Business Administration & Economics, Lahore, Pakistan
fshehzad.stat@gmail.com

R. Steorts 159
US Bureau of the Census

N. Stiegler 931
Department of Statistics and Demography, University of the Western Cape,
X17 7535 Bellville, South Africa
nstiegler@uwc.ac.za

Andrew Stone 5
World Bank

John Stufken 97
Professor and Head, University of Georgia
jstufken@uga.edu

Subanar 950, 1079
Mathematics Department,
Gadjah Mada University, Yogyakarta, Indonesia
subanar@yahoo.com, subanar@ugm.ac.id

Subarkah 70
Center for Health Research Universitas Indonesia

Manjunath S Subramanya 140, 148
Department of Surgery, Mount Isa Base Hospital, Mount Isa, Queensland, Australia
manjunathbss9@yahoo.com

Etih Sudarnika 248
Laboratory of Epidemiology, Faculty of Veterinary Medicine, IPB, 16680, Darmaga, Bogor, Indonesia
etih23@yahoo.com

Suhartono 845, 1079
Perum ITS U-71, Jl. Teknik Komputer II, Keputih Sukolilo, Surabaya, Indonesia
suhartono@ statistika.its.ac.id

Suliadi 747
Dept. of Statistics, Bandung Islamic University
Jl. Tamansari No. 1 Bandung
Indonesia
suliadi@gmail.com

H. Sulieman 1038
Department of Mathematics and Statistics
American University of Sharjah,
P.O.Box 26666, Sharjah, U.A.E.
hsulieman@aus.ae

Khalaf S. Sultan 1050
Department of Statistics and Operations Research,
College of Science, King Saud Universit,
P.O.Box 2455, Riyadh 11451, Saudi Arabia
ksultan@ksu.edu.sa

Yusep Suparman 1070
Statistics Department, Padjadjaran University
Jl. Ir. H. Juanda no. 4, Bandung 40115
Indonesia
yusep.suparman@unpad.ac.id

Jef L. Teugels 2
Katholieke Universiteit Leuven & EURANDOM, Eindhoven
Jan Beirlant, University Center of Statistics, Katholieke Universiteit Leuven
Goedele Dierckx, HUBrussel
jef.teugels@wis.kuleuven.be
jan.beirlant@wis.kuleuven.be

Inam-Ul-Haq 664, 676
National College of Business Administration & Economics, Lahore, Pakistan
inam-ul-haq786@hotmail.com

J. A. M. Van der Weide 950
Delft University of Technology, Delft, The Netherlands
jamvanderweide@tudelft.nl

Edward J. Wegman 3, 963
Center for Computational Statistics
George Mason University
368 Research I, Ffx, MSN: 6A2
ewegman@gmu.edu

Yekti Widyaningsih 1087
Department of Statistics, Bogor Institute of Agriculture, Indonesia
yekti@ui.ac.id

Wafik Youssef Younan 1030
Department of Economics
The American University in Cairo
Cairo, Egypt
wyounan@aucegypt.edu

Yudarini 70
Center for Health Research Universitas Indonesia

Rossita Mohamad Yunus 98, 106. 121, 236
Department of Mathematics and Computing, Australian Centre for Sustainable Catchments,
University of Southern Queensland, Toowoomba, Queensland, Australia
Institute of Mathematical Sciences, University of Malaya, Kuala Lumpur, Malaysia
Rossita.MuhamadYunus@usq.edu.au

Anis Y. Yusoff 5
Institute of Ethnic Studies (KITA),

National University of Malaysia
anis.yusoff@gmail.com

Faisal Maqbool Zahid 1094
Department of Statistics, Ludwig Strasse 33, 80539.
Ludwig-Maximilians University Munich, Germany
faisalmz99@yahoo.com

Enas Zakareya 461
Economic Issues Program (EIP),
Information and Decision Support Center (IDSC),
Egyptian Cabinet.
enabd@idsc.net.eg

Lu Zou 1109
Hicks Building,
Sheffield University,
Sheffield S3 7RF
heron_20012003@hotmail.com

Yong Zulina Zubairi 304, 305
Centre for Foundation Studies in Science,
University of Malaya, 50603 Kuala Lumpur, Malaysia
yzulina@um.edu.my

Organized by:

الجمعية الاحصائية لدول العالم الاسلامي
Islamic Countries Society of Statistical Sciences

Cosponsored by:

The Egyptian Cabinet Information and Decision Support Center (IDSC)

American University in Cairo

Islamic Development Bank
Together We Build a Better Future

ISBN: 978-977-416-365-8

The American University in Cairo
P.O.box 74, New Cairo 11835, Egypt
tel. 20.2.2615.2720
Fax 20.2.2795.7565
iccs-x@aucegypt.edu
www.iccs-x.org.eg/
www.isoss.com.pk/

The Islamic Countries Society of Statistical Sciences (ISOSS)
Address: Plot No. 44-A, Civic Centre, Liaqat Chowk, Sabzazar Scheme, Multan Road, Lahore, PAKISTAN.
Tel: +92 -42- 3784 0065, +92 -42- 3587 8583 Fax: +92 -42- 575 2547 E-Mail: secretary@isoss.com.pk