PROCEEDINGS

Eighth International Conference on

RECENT ADVANCES IN STATISTICS

"Statistics, Biostatistics and Econometrics"

on

February 8-9, 2011

In the honour of

Dr. Shahjahan Khan

President, ISOSS

for

Outstanding Contributions to Statistical Research and Leadership in Promoting the ISOSS

Vol. 20

Jointly organized by

Islamic Countries Society of Statistical Sciences
44-A, Civic Centre
Sabzazar, Multan Road
Lahore, Pakistan
Tel: +92-42-37840065
Fax: +92-42-35752547
Email: secretary@isoss.net
URL: http://www.isoss.net

National College of Business Administration & Economics
40-E/1 Gulberg-III
Lahore (Pakistan)
Tel: +92-42-35875853
35753430/35752716-19
Email: drmunir@brain.net.pk
URL: www.ncbae.edu.pk

Published by: ISOSS, Lahore, Pakistan.
Printed by: Izharsons Printers, 9-Rattigan Road, Lahore (Pakistan).

CONTENTS

Address by Dr. Munir Ahmad
Founding President and Patron ISOSS

- **Dr. Shahjahan Khan, Mian Shamim Haider, Dr. Ghapor,**
- **Dr. Ahmed Abdulatif Bahnassy, Dr. Qamar-ul-Islam,**
- **Dr. Ashraf Chaudhry, Delegates from Pakistan,**
- **Ladies and Gentlemen!**

I on my behalf and on behalf of members of ISOSS welcome you to the 8th International Conference on Recent Advances in Statistics being held in the honour of Dr. Shahjahan Khan for his outstanding contribution to the development and leadership in ISOSS.

I thank Dr. Shahjahan Khan in particular, Dr. Ghapor, Ahmad, Qamar-ul-Islam, Ashraf for their presence on this occasion.

I also want to thank Mian Shamim Haider, Chairman, NCBA&E and former Federal Minister for his presence on this occasion, who for the love of education has dedicated himself to the spreading of knowledge. I deeply appreciate his efforts for NCBA&E programs by creating alliances amongst statisticians and encouraging us in organizing such conferences for exchange of knowledge and experiences for better understanding and advancement of knowledge.

I feel proud that our efforts have become fruitful in holding the 8th International Conference when there are many socio-political changes being held in the Islamic World.

The society, which had celebrate its 20 years journey in 2008, was established way back in 1988, during the first Islamic Countries Conference on Statistical Sciences held at Lahore.

I feel really proud of its existence, as the Society had held ten Islamic Countries Conferences on Statistical Sciences at Lahore, Morocco, Malaysia, and Indonesia, and numerous national conferences, seminars and workshops. The 11th Conference will be held in Jordan in 2011.

Prior to this conference, more than 37 countries participated in the conference deliberations, but due to security measures, only a few countries participated. In some cases, Pakistan Embassies did not issue visa to some participants. In fact, we had received papers from 42 foreign participants.

You will be glad to know that ISOSS has now b uilt ISOSS House on the land allotted by Mian Nawaz Sharif when he was the Chief Minister, Punjab on the recommendations of Dr. Abudllah Bin Omar Nasseef who came to Lahore from Saudi Arabia to personally request Mian Nazwaz Sharif, who very gracefully accepted his recommendations. Members and friends of ISOSS had generously contributed and are contributing / finances for the ISOSS House.

We still need help from all friends of Statistics for setting up internationally recognized Research Institution and equipping with necessary facilities. ISOSS activities have been internationally recognized. In this connection, I thank Mian Shahid Ali Haider who had offered full support to make ISOSS activities fully operational.

I strongly believe that planning has to be knowledge based and to be monitored by a strong team of statisticians. At present, there is no interaction between academicians and Statistics Officials at Federal as well as provincial levels. If Official Statistics is to be improved, academia must be involved very effectively and if statistical agencies need qualified incumbents, statistical agencies must be effectively involved in colleges and universities so that statisticians produced by colleges and universities are directly inducted in statistical organizations.

ISOSS has prepared a type of ‘white paper’ on the Statistics and 63-years history of Statistics in Pakistan with a large number of practical proposals and recommendations for the Government which had already been will be mailed to all the concerned quarters.

I foresee ISOSS would develop into as a World Forum that can be managed on a collective vision of its active members. Dignity of top statisticians working both in public and private sectors is a pre-requisite to Society’s strategy and action plans.

In the end, I thank my team of volunteers, mostly students and professors from National College of Business Administration and Economics, especially Mian Shahid Ali Haider, Prof. Shumas-ur-Rehman, Major Ehsan, Dr. Muhammad Hanif, Prof. Akhlaq Ahmad and ISOSS Secretariat staff Muhammad Iftikhar, Muhammad Imtiaz and Saif-ur-Rehman and others for their untiring work. There is a long list of students who made this conference a success.

I am again grateful to Dr. Shahjahan Khan for sparing their precious time for the inauguration ceremony of the conference.

I thank you Sir, and thank you all.

KEY NOTE SPEECH
on
A STATISTICAL JOURNEY – MOVING FORWARD ISOSS
by

Dr. Shahjahan Khan
President, Islamic Countries Society of Statistical Sciences
Department of Mathematics and Computing,
University of Southern Queensland, Toowoomba, Australia

Like many others I came to study statistics by chance. In the admission test for BSc entry at the Jahangirnagar University, Bangladesh I was on the top of the list. Although the test had nothing to do with statistics, as such, I thought I might do well in the subject. When Professor Kazi Saleh Ahmed, the Chairman of the Department asked me if would study statistics, my response was 'yes'. As an undergraduate student I found the subject interesting and challenging, but not exciting – too may jargons. It is only when I was a research student at the University of Western Ontario, Canada I discovered that statistics can be exciting. The real meaning of statistics and its unique role in the modern science and contemporary society were unfolded to me much later. The scientific community now knows that statistics is capable of help solving most of the measure problems – starting from issues under the sea to those above the sky, let alone on the earth.

As an undergraduate student of statistics I thought it was a good subject to get jobs in the academia, government and industry. At this stage, to me, statistics was more of a tool to be used by others, when needed. In my early career in statistics I found it useful to design studies, analyze data, and interpret results to be used in real life. With the passage of time and growing experience and interaction with colleagues in various professions I realized the extent and depth of the diverse use of statistics. At this stage, I started viewing statistics more than a tool. In fact it is a science of making decision in the face of uncertainty. Further engagement with a wide range of applications of statistics and its ability to address complex problems, which otherwise can't be solved, now I believe that the main task of statistics is to formulate strategies. Statisticians must be strategists to secure their rightful position in the scientific community and at the high echelon of the socio-political structure of the contemporary world. No one would offer it to statisticians, rather we will have to achieve and earn it overcoming all hurdles.

ISOSS has played a crucial role in shaping my professional position and global perspectives on statistics. It provided me a platform to implement my professional vision to improve the quality of statistics, enhance collaboration and interaction among statisticians of diverse background, improve the quality of government/official statistics, and promote state of the art statistical research. I am grateful to ISOSS, its members and management, for entrusting me with the biggest challenge of leading ISOSS at a very crucial time of our time. The unfailing support and rock solid unity of the members of ISOSS helped me to move forward to regain the declining glory of

ISOSS and take it from a regional institution to a prestigious international professional organization.

In 1994, I participated my first ISOSS Conference in the Avari Hotel, Lahore, Pakistan. This is the first time I met Dr. Munir Ahmad and Dr. Mohammad Hanif Mian along with many other pioneering members of ISOSS. I was inspired by them, and committed myself to the service of this professional body. Ever since, I have been able to participate in most of the Islamic Countries Conference on Statistical Sciences (ICCS). I proposed to honour statisticians of high achievements by ISOSS in 1998, I am glad it was implemented in 1999. The first two recipients of the ISOSS Gold Medal that year were Professor A K Md Ehsanes Saleh and Professor Emad Eldin Aly. I had the honour to edit a special volume of Pakistan Journal of Statistics in 2000 in their honour.

The 8th Islamic Countries Conference on Statistical Sciences (ICCS) held in December 2005 in the FAST University, Lahore. To my surprise I was elected as the President of ISOSS in the business session of the conference. Although I was not quite prepared to take the heavy responsibilities, I didn't want to disappoint the delegates, especially Dr. Munir and his team, by any means. I simply knew how difficult it was to fill the shoes of Dr. Munir. Some of you may remember, on accepting the position of the President of ISOSS, in my first speech in Lahore, I proposed to build ISOSS Head Quarters in Lahore and name it after Dr. Munir. I also announced to create a true sense of belongingness to all ISOSS members and engage them in the activities of ISOSS.

My first challenge was to engage with statisticians of the South East Asia, namely Indonesia and Malaysia, and in the Arab World. The second challenge was to create an environment to attract and interact with statisticians outside the Islamic Countries, particularly those in the West. My third challenge was to bring ISOSS in the global stage alongside other international professional statistical organizations. None of the challenges was easy to deal with, but over the years we have been able met them with significant successes. The ICCS-IX in Kuala Lumpur in 2007 and ICCS-X in Cairo have give us the opportunities to demonstrate our commitment to those regions and we are so grateful for the unbelievable support that ISOSS received from the local and international organizers and participants. In increasing interaction and involvement of Western statisticians in ISOSS activities, cooperation with the ISI, and remarkable national and international coverage of major events organized by ISOSS have been instrumental to its global acknowledgement.

On my way back to Australia from Pakistan in January 2006 I had a short stopover in Dhaka, Bangladesh. With the next ICSS in mind I met the President of the People's Republic of Bangladesh, Professor Eajuddin Ahmed, a former senior colleague of mine in the University of Dhaka, for his support and participation in the conference if Dhaka was chosen as the venue. On my next stop in Kuala Lumpur, on the way to Australia, I met Professor Datu Jamil Bin Osman, Dean of Faculty of Business, International Islamic University Malaysia. After my visit he re-organized the Malaysia Institute of Statistics (ISM) and became its new president. A year later I visited several Universities of Malaysia. My first host was University Putra Malaysia (UPM) where I met with the Executive Committee of the ISM to formally host the ICCS-IX. I am very grateful to Professor Datu Kamel Atan, Director of INSPEM, UPM for his generous support during my stay there. Later I was able to secure the support of Professor Nur Hasna Moin, Head, Institute of Mathematical Sciences, University of Malaya to provide us with a conference office as well as a support staff to run it. In that trip I also met the Head, Department of Statistics Malaysia in Putra Jaya. The ISM organized a one-day symposium during my visit to revitalize SIM. This visit was invaluable to lay the foundation of successful organization of ICCS-IX in December 2007. The dedication and contributions of Dr. Ibrahim Mohamed, Dr. Abdul Ghapor Hussin, Dr. Noor Akma Ibrahim and many others were highly appreciated by everyone.

A special session of the conference was dedicated to the memory of Professor M Safiul Haq, University of Western Ontario, Canada who died in 2007. The keynote addresses were made by Ali S Hadi (Egypt/USA), Malay Ghosh (USA), and Kerrie Mengersen (Australia). We were able get funding from the SESRIC, an Ankara based OIC subsidiary through sponsoring of the government statistics session, and several other local sponsors. As the Founding Chief Editor, I was able to include a number of selected articles presented in the conference in a special issue of the Journal of Applied Probability and Statistics (JAPS). The proceedings of the conference were published in due time.

Being invited by the Carlton University, Canada and North South University (NSU), Bangladesh I served as the Chair of the International Scientific Committee of an international statistics conference held at the Bangladesh-China Friendship Convention Centre, Dhaka, Bangladesh in December 2008. Many prominent statisticians from home and abroad participated in the conference to honour Professor A K Md Ehsanes Saleh, Carleton University, Canada. Dr. Abdul Hannan Chowdhury of NSU was the main organizer of the conference. The President of the Bangladesh Academy of Science, Professor M Shamsher Ali, Governor of Bangladesh Bank, Dr. Saleh Uddin, and Chairman of the University Grants Commission, Professor Nazrul Islam highlighted various applications of statistics. Both Dr. Munir and Dr. Hanif participated in the conference as guests of honour. The event received significant media attention.

The search for the host for the next ICCS started from Kuala Lumpur. My target was a venue in the Middle East to take ISOSS at the door step of the Arab world. I met Professor Abdelhamid Elshaarawi and Professor Ali S Hadi in an international conference in Rajshahi University, Bangladesh in December 2006. I approached Ali in Kuala Lumpur to host the ICCS-X in Egypt. Later I contacted Abdel for his support. Both of them were interested but took about a year to finally commit to host ICCS-X. It was the wisdom of Ali to elect Dr. Magued Osman, Head of Egyptian Cabinet Information and Decision Support Centre as the Chair of the Local Organizing Committee. I first met Dr. Osman in Ras AlKhaimah, United Arab Emirates in a conference in November 2008 along with Abdel. In the side line of the conference, we were able to discuss and decide on some of the main aspects of the conference. The hard work of many Egyptian colleagues, especially Dr. Ali Hadi, Dr. Zeinab Amin, Dr. Wafik Younan, and several staff from the office of Dr. Magued Osman and the American University in Cairo were keys to the success of ICCS-X. The Islamic Development Bank made modest financial contribution for the conference. The participation of Jim Berger and Edward Wegman from USA, Jef Teugels, ISI President, and Kaye Basford from Australia as keynote speakers contributed to the success of the conference. The proceedings of the conference have been published in two volumes.

I met Mr. Rashed Al Sywaidi, Director General of National Bureau of Statistics, UAE in the second Arab Statistics Conference in the city of Surt, Libya along with Dr. Haidar Fraihat, Director General of Department of Statistics, Government of Jordan. During our discussions in Surt, Mr. Al Sywaidi invited me to visit his office in Abu Dhabi. We met in his office in November 2009 about hosting ICCS-XI in 2011. Although initially he was quite interested in the idea, due to the commitment for the national census of UAE he finally decided not to proceed with the proposal. Our initial hope of Indonesia to host ICCS-XI disappeared when Professor Asef Saifuddin and Professor Maman Djouhari expressed their concerns about the unfavourable political situation to host ISOSS conference.

My next options to explore were Jordan and Qatar. I contacted Dr. Fraihat who took some time to make a final decision. In the meantime, I approached Dr. Ayman Baklizi of Qatar University. With initial support from the Department a formal application to the University authority was lodged for final approval. While the approval process was taking longer than anticipated time, I was contacted by Dr. Fraihat informing their willingness to host ICCS-XI. Later we agreed on the proposal of Dr. Hilal Albayyati to organize a joint conference of ISOSS and UAS (Union of Arab Statisticians) in Jordan in December 2011. I met Dr. Albayyati in the first Arab Statistics Conference in Amman, Jordan in 2007, and he is the main organizer of the UAS. Weeks after our agreement on the venue of the ICCS-XI, Dr. Baklizi confirmed me that Qatar University authority had approved only the application for ISOSS conference and rejected application of several other conferences. Fortunately, Qatar University is flexible about the date of the conference so we are able to plan to hold ICCS-XII in Qatar. This is the first time in the history of ISOSS we have a host for an ICCS in advance.

This is the first time a government department is hosting any ISOSS conference in its 23 year history. I hope it would be an eye opener to other government statistics offices in the OIC member states to come forward to host future ISOSS conferences. Obviously this phenomenon would contribute to the direct interaction of the primary producers of statistics and research statisticians with high potential to improve the quality of statistics. It may be noted here that the Department of Statistics, Government of Jordan received the best statistics office award in the Second Arab Statistics Conference in Libya in November 2009.

Finally, the dream to build ISOSS House has also come true. The life long efforts of Dr. Munir Ahmed and his team have produced its fruits. Financial contributions and support from everyone are highly appreciated by ISOSS family. We need young generation of statisticians who would be able to utilize the facility to benefit of the Ummah. I intend to use this building for further development of statistics in the Islamic countries and beyond. Initially we would like to foster statistical research and training activities through the ISOSS House. In addition to promote the value of statistics we hope to build it as an institute for excellence of statistical learning and teaching. We may do so by producing high quality statisticians through undergraduate and postgraduate programs in statistics.

I always consulted ISOSS office, particularly Dr. Munir Ahmad and Dr. M. Hanif Mian, before taking any major decision. It is true that we did not agree on everything all the time, but we did not disagree on any serious matters. Their unqualified support and guidance have been a source of inspiration for me.

Every informed statistician now knows about ISOSS and the biennial ICCS. The ISOSS activities on the first Word Statistics Day attracted significant global publicity of ISOSS via ISI. In the Islamic Countries ISOSS is a household name among the statisticians and statistical organizations. We need to attract more and regular sponsors to move ISOSS forward. The good wills and standing of ISOSS should be utilized to its benefit. We need to cooperate with all sister organizations and individuals who are willing to work with us. We would continue doing good and promote good in the scientific community and conscious members of the Ummah. We have travelled a long journey but a lot more to be done. Statistics is growing faster and ISOSS must be in the forefront to lead it to our benefit and the benefit of the mankind.

I thank each and every one of you who helped me and ISOSS over the years. I am really very grateful to all of you. Without your help we would not be able to achieve any successes. Certainly I did not work hard to gain any recognition from anyone other than being rewarded by Almighty Allah. I ask Him to reward you all unboundedly for your good deeds.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 1-8

INTRICACIES AND PITFALLS OF MODEL SELECTION STRATEGIES USED IN TIME SERIES MODELS

M. Qamarul Islam[1] and **Mehmet Yazici**[2]
Department of Economics, Cankaya University, Ankara, Turkey.
Email: [1]islammq@cankaya.edu.tr; [2]myazici@cankaya.edu.tr

ABSTRACT

Selecting the correct lag order is necessary in order to avoid model specification errors in autoregressive (AR) time series models. Here we explore the problem of lag order selection in such models. This study provides an in-depth understanding of the model selection mechanism to the practitioners in various fields of applied research. Several interesting findings are reported and through these the intricacies and pitfalls of the model selection strategies are exposed.

KEYWORDS

AR models; Lag order; Model selection; Model selection criteria.

JEL Classification: C01; C22; C52.

1. INTRODUCTION

Selecting a suitable lag length for an autoregressive (AR) time series model is very important in order to avoid model specification errors. More specifically, such errors, if present, may cause the regression estimates to become biased and inefficient. It is known in the literature (see, for example, Lutkepohl (1993), Hafer and Sheehan (1989) and Granger and Jeon (2004)) that selecting a lag higher than the true lag order causes an increase in the mean forecast errors, while selecting a lower lag than the true lag order generates autocorrelated errors. Applied researchers in various fields are using a number of model selection procedures with the aim of detecting and estimating the true model using the available data. The aim of this study is to explore the intricate aspects of data-based model selection.

There are a number of studies investigating the model selection problem, mostly through simulation work, and reporting the probabilities of selecting the true model under various model setups using different selection procedures. A recent study by Hacker and Hatemi (2008) compares the performance of different information criteria in lag length selection in vector autoregressive (VAR) models in the presence of heteroscedasticity (ARCH). Hatemi and Hacker (2007) discusses the role of likelihood ratio (LR) test in choosing optimal lag order in VAR models. Kose and Ucar (2006) investigates the finite sample properties of several model selection criteria in bivariate VAR models. Liew and Chong (2005) studies the effects of ARCH errors on the performance of commonly used lag selection criteria. Basci and Zaman (1998) considers the behavior of model selection criteria in AR models when the error terms are not normally distributed. Koreisha and

Pukkila (1993) compares the performance of several methods used for identifying the order of VAR processes when the number of component series is large.

The main objective of the above mentioned studies is to compare the performances of different model selection procedures. However, these studies do not explore the mechanism of model selection itself. There are some articles theoretically exploring the problem of model selection but they are few in number (see, for example, Leeb and Potscher (2008), (2006) and (2005)). In our study, we work with simple AR models and rely on simulation in order to bring into light several interesting findings related with the model selection and thus provide an insight into the problem of model selection to the researchers in applied fields.

2. METHODOLOGY

We consider the autoregressive time series model with true lag length p ($p \geq 1$),

$$y_t = b_0 + \sum_{j=1}^{p} b_j y_{t-j} + u_t \,, \; t = 1, 2, \ldots\ldots, T \,, \tag{1}$$

with the usual assumptions about the errors u_t (i.i.d. $N(0, \sigma^2)$). Furthermore, the regression parameters, b_j ($j = 1, 2, \ldots, p$), take values that make the series { y_t } stationary. The stationarity can be assured by selecting b_j in such a way that the characteristic equation $\left|\lambda^p - \sum_{j=1}^{p} b_j \lambda^{p-j}\right| = 0$ has characteristic roots λ_j such that $\left|\lambda_j\right| < 1$.

All models with lag order p^* ($0 \leq p^* \leq M$, $M \geq p$) are estimated using ordinary least squares (OLS) method, imposing a reasonably selected maximum prior lag order M . Among these models, we select the model that is optimal according to certain optimality criterion. In this study we use several well known and widely used model selection optimality criteria in applied work such as Akaike Information Criterion (AIC) (Akaike, 1973), Schwarz Information Criterion (SIC) (Schwarz, 1978), and Hannan-Quinn Criterion (HQC) (Hannan and Quinn, 1979), etc. However, due to similar results obtained, we only report here the findings with respect to AIC.

3. SIMULATION STUDY

We conduct a simulation study using $[100000/T]$ (integer) Monte Carlo runs for a $AR(p)$ model (1) with true lag order $p = 4$. We search for the most parsimonious model over the span of lag order from 0 to 7 ($M = 7$) using the model selection procedures mentioned in the previous section while incorporating different sample sizes. In order to make the results comparable, we use an adjusted sample size $T^* = T - M$ as suggested by Ng and Perron (2005). Furthermore, we take σ^2 and b_0 each equal to 1 without loss of generality and the stationarity requirement restricts us to choose b_j - values ($j = 1, 2, \ldots, p$) in the range from -1 to 0.50, end points exclusive.

3.1 Model Selection Uncertainties

The first interesting finding is that the probability of getting the most parsimonious model the same as true model, henceforth calling it 'true selection', depends upon the true value of the parameter b_j ($j = 1, 2, \ldots\ldots, p$). For illustration we have tabulated in Table 1 the probabilities (%) of model selections by the selection criteria (AIC) while varying b_4 from 0 to 0.48 ($b_1 = b_2 = b_3 = -0.50$) and for different sample sizes ($T = 30, 50, 100$). It is observed that when the values of b_4 are reasonably large, i.e., away from zero (e.g., 0.48, relative to the working range from 0 to 0.50), the probabilities of true selection are substantially high particularly in large samples. However, the performance of the criteria in revealing the true model steadily decline with values of b_4 approaching to zero. More disturbing is the fact that enlarging the sample in such situations does not increase the probability of true selection to a comfortable level. Even for moderate sample sizes (30 or 40) that are generally available in most of the practical researches the selection criteria show a tendency of selecting models towards the lower end. These findings reveal the fact that the convergence in probability of detecting the true model is non-uniform with respect to b_4. Similar patterns are observed while considering b_4 having values in the negative range; i.e., from –1 to 0.

In order to provide more insight into the process of model selection, we report (Table 2) the weighted average (weights are the probabilities of model selections using AIC) lag lengths while varying b_4. It is clear that the lag length is increasingly underestimated (towards the lags lower than the true lag 4) with decreasing b_4.

Table 1:
Probabilities (%) of selecting models of different Lag Lengths (LL)

	$T = 30$				$T = 50$				$T = 100$			
	LL				LL				LL			
b_4	< 3	$= 3$	$= 4$	> 4	< 3	$= 3$	$= 4$	> 4	< 3	$= 3$	$= 4$	> 4
0.00	34.08	55.32	6.81	3.78	10.25	78.00	7.80	3.95	0.00	90.40	7.30	2.30
0.04	31.72	58.16	6.03	4.11	8.05	82.24	6.15	3.60	0.30	89.99	6.51	3.31
0.08	27.04	63.17	5.49	4.32	5.15	82.99	8.15	3.75	0.00	83.89	12.31	3.90
0.12	23.29	65.75	6.24	4.74	4.50	81.89	8.30	5.35	0.00	75.38	18.91	5.80
0.16	21.52	66.89	7.44	4.17	3.35	77.29	14.46	4.95	0.00	63.97	28.72	7.41
0.20	18.52	68.87	7.47	5.16	2.60	73.74	17.96	5.75	0.00	48.86	41.83	9.41
0.24	16.21	67.61	10.83	5.37	2.20	67.84	24.41	5.60	0.00	37.15	54.74	8.21
0.28	12.69	66.80	13.38	7.14	1.15	59.83	31.41	7.65	0.00	25.23	63.96	10.91
0.32	10.74	65.96	16.23	7.08	0.95	53.33	37.52	8.25	0.00	15.52	75.67	8.91
0.36	8.79	63.62	20.14	7.47	0.55	42.62	47.97	8.90	0.00	8.91	81.78	9.41
0.40	7.44	58.70	24.85	9.03	0.60	35.52	53.63	10.30	0.00	4.31	83.69	12.11
0.44	7.11	54.83	28.18	9.90	0.40	27.07	61.93	10.65	0.00	2.00	85.78	12.31
0.48	4.89	50.51	35.08	9.54	0.25	20.51	67.98	11.31	0.00	0.60	88.19	11.31

Table 2:
Average estimated lag lengths

b_4	0.00	0.04	0.08	0.12	0.16	0.20	0.24	0.28	0.32	0.36	0.40	0.44	0.48
$T = 30$	2.62	2.68	2.75	2.83	2.85	2.94	3.01	3.14	3.19	3.27	3.38	3.45	3.54
$T = 50$	3.03	3.03	3.10	3.16	3.22	3.29	3.36	3.48	3.57	3.69	3.80	3.89	3.96
$T = 100$	3.13	3.14	3.22	3.33	3.48	3.65	3.76	3.91	3.98	4.05	4.13	4.16	4.15

3.2 Serial Correlation

At this point one may argue that in practice a model selection procedure is not required to necessarily identify the unknown true model but to provide a most parsimonious model close to the true model with substantially high probability of its selection (e.g., 90% or more). However, it is worth to remind here that using a model with a lag order smaller than the true lag may encounter the problem of serial correlation, and hence, making the use of OLS inappropriate. In order to illustrate it, we have reported (Table 3) the chances (%) of observing serial correlation, by rejecting the null hypothesis of no serial correlation at 10 % level of significance, for models of different orders. It is to be noted that for the model of true lag length, even with low probability of its selection, the chances of encountering serial correlation remain negligible. However, the chances of observing serial correlation rise with the probability of selecting the model lower than the true lag. Particularly, when the probability of selecting a lower lag length model is moderate (e.g., 50 to 60%) the chances of having serial correlation are unreasonably high (e.g., 20 to 40%). Furthermore, the situations worsen with increase in the sample size.

Table 3:
Chances (%) of observing Serial Correlation

	$T = 30$		$T = 50$		$T = 100$	
b_4	LL = 3	LL = 4	LL = 3	LL = 4	LL = 3	LL = 4
0.00	0.93	0.42	0.45	0.00	0.40	0.00
0.04	0.78	0.15	0.60	0.05	0.50	0.00
0.08	1.14	0.39	1.55	0.10	1.70	0.00
0.12	2.07	0.60	2.20	0.05	4.70	0.00
0.16	3.27	0.51	4.05	0.10	9.21	0.00
0.20	3.96	0.36	7.25	0.20	20.11	0.10
0.24	5.46	0.60	11.60	0.00	32.52	0.00
0.28	7.74	0.45	16.76	0.10	50.14	0.10
0.32	9.81	0.57	24.56	0.25	66.96	0.00
0.36	14.73	0.60	36.92	0.15	79.57	0.10
0.40	19.47	0.90	46.82	0.30	89.89	0.40
0.44	25.09	1.11	57.48	0.25	97.09	0.10
0.48	31.96	1.29	69.08	0.70	98.80	0.20

3.2 Estimation Bias and Efficiency of Estimates

In the following Table (Table 4) we give the bias in estimating the parameters b_1, b_2, and b_3 by taking the models of Lag Length 3 and Lag Length 4 (the true model) into consideration. Again for true model (of lag 4) the estimation biases are negligible. However, when the model with lag 3 is selected and estimated, the biases in estimation sharply increase with increasing value of b_4.

Table 4:
Estimation Biases in the estimates $\hat{b}_1$, $\hat{b}_2$, and $\hat{b}_3$

	$T=30$						$T=50$					
	LL = 4			LL = 3			LL = 4			LL = 3		
b_4	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$
0.00	-0.02	-0.05	0.01	0.01	-0.01	0.04	-0.01	-0.03	0.01	0.01	-0.00	0.03
0.04	-0.03	-0.06	0.00	-0.01	-0.03	0.03	-0.01	-0.03	0.01	-0.01	-0.02	0.01
0.08	-0.03	-0.06	-0.00	-0.02	-0.05	0.00	-0.02	-0.03	0.00	-0.03	-0.04	-0.01
0.12	-0.03	-0.06	-0.00	-0.04	-0.07	-0.01	-0.01	-0.03	0.00	-0.05	-0.06	-0.03
0.16	-0.04	-0.07	-0.01	-0.06	-0.09	-0.04	-0.01	-0.03	0.01	-0.07	-0.09	-0.05
0.20	-0.03	-0.06	-0.01	-0.08	-0.11	-0.06	-0.01	-0.03	0.00	-0.09	-0.11	-0.08
0.24	-0.03	-0.07	-0.01	-0.10	-0.14	-0.08	-0.02	-0.03	0.00	-0.12	-0.14	-0.10
0.28	-0.03	-0.07	-0.01	-0.12	-0.16	-0.11	-0.02	-0.04	-0.00	-0.15	-0.17	-0.13
0.32	-0.04	-0.07	-0.02	-0.16	-0.20	-0.14	-0.02	-0.04	-0.01	-0.18	-0.20	-0.17
0.36	-0.05	-0.07	-0.02	-0.18	-0.23	-0.17	-0.02	-0.04	-0.00	-0.21	-0.23	-0.20
0.40	-0.05	-0.08	-0.03	-0.21	-0.26	-0.20	-0.02	-0.04	-0.01	-0.24	-0.28	-0.24
0.44	-0.04	-0.08	-0.02	-0.23	-0.29	-0.23	-0.02	-0.04	-0.01	-0.28	-0.31	-0.27
0.48	-0.04	-0.08	-0.03	-0.26	-0.33	-0.27	-0.02	-0.04	-0.01	-0.32	-0.36	-0.32

	$T=100$					
	LL = 4			LL = 3		
b_4	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$
0.00	-0.00	-0.01	0.01	-0.01	-0.02	-0.00
0.04	-0.00	-0.01	0.01	-0.01	-0.02	-0.00
0.08	-0.01	-0.01	0.00	-0.04	-0.04	-0.03
0.12	-0.00	-0.01	0.00	-0.06	-0.06	-0.05
0.16	-0.01	-0.01	0.00	-0.08	-0.09	-0.07
0.20	-0.01	-0.01	0.00	-0.11	-0.11	-0.10
0.24	-0.01	-0.01	-0.00	-0.14	-0.15	-0.13
0.28	-0.01	-0.02	0.00	-0.16	-0.18	-0.16
0.32	-0.01	-0.02	-0.00	-0.21	-0.22	-0.20
0.36	-0.01	-0.02	-0.00	-0.24	-0.25	-0.24
0.40	-0.01	-0.02	-0.00	-0.28	-0.30	-0.28
0.44	-0.01	-0.02	-0.00	-0.33	-0.35	-0.33
0.48	-0.01	-0.02	-0.00	-0.38	-0.40	-0.38

In order to study the precision of the estimates we report in Table 5 below the efficiencies (%) of the estimates ($\hat{b}_j$, $j=1,2,3$) based upon the mean square errors (MSE) calculated for the model of lag length 3 and lag length 4 and used as follows:

$$Efficiency(\hat{b}_j \mid Model\,3) = 100 \times \frac{MSE(\hat{b}_j \mid Model\,4)}{MSE(\hat{b}_j \mid Model\,3)}$$

Table 5:
Efficiencies (%) of estimates

	$n=30$			$n=50$			$n=100$		
b_4	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$
0.00	133.05	144.23	131.51	131.18	138.95	129.00	131.65	135.71	130.35
0.04	137.38	149.86	138.38	135.10	140.60	137.33	132.46	132.50	137.80
0.08	139.31	148.47	143.93	133.27	134.18	140.11	119.45	115.77	128.45
0.12	138.72	143.46	147.19	128.08	125.76	136.81	102.78	97.98	109.56
0.16	135.78	136.55	147.49	115.02	107.67	129.11	79.64	72.37	87.70
0.20	131.22	128.09	142.01	101.36	95.05	113.00	59.61	56.48	64.56
0.24	123.17	116.64	136.28	85.31	76.31	96.61	42.39	39.42	45.75
0.28	111.39	102.81	124.34	69.14	60.76	78.15	31.99	28.04	34.90
0.32	96.91	89.02	109.57	55.07	48.90	60.58	21.95	20.59	23.52
0.36	84.27	75.97	95.63	42.50	38.67	47.63	16.26	15.61	17.12
0.40	73.26	64.73	80.95	34.08	29.89	36.65	12.03	11.09	12.52
0.44	64.66	55.65	71.39	27.08	24.20	28.76	8.70	8.24	9.01
0.48	53.86	45.65	56.87	20.68	18.19	21.59	6.48	6.06	6.64

Clearly, these efficiencies vary with the values of b_4. Although, as expected, the precision of estimates for model of lag length 3 are high in the beginning they steadily decrease with increase in b_4. The disturbing fact is that the efficiencies of these estimates are low (considerably lower than 100%) for moderate values of b_4 where the chances of selecting model of length 3 are considerably high. Here again, the disconcerting feature is that the deterioration in efficiencies sharpens with the increase in sample size.

3.3 Test of Hypothesis

In Table 6 we present the simulated size of $t-$test (assumed size: 5%) for testing the null hypotheses $H_{0j}: b_j = -0.5 \quad (j = 1, 2, 3)$ and varying b_4 by taking the models of Lag Length 3 and Lag Length 4 (the true model) into consideration. While for the model of lag length 4 (true lag) the simulated test sizes stay quite close to the assumed size (5%), it is clear that for the model with lag order 3 it is not so. Rather, all simulated sizes are found to be increasing with b_4 systematically and start taking large values that make the test totally useless. Perhaps, the $t-$statistic may not be having the standard $t-$distribution in such cases.

Table 6:
Size (%) of Test

	$T=30$						$T=50$					
	LL = 4			LL = 3			LL = 4			LL = 3		
b_4	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3
0.00	3.39	4.05	3.00	3.12	3.66	2.70	4.90	5.05	4.25	4.80	4.05	4.10
0.04	3.87	4.41	3.30	3.96	4.68	3.21	3.95	4.25	4.30	4.25	5.65	4.55
0.08	3.99	4.20	3.78	4.50	6.12	3.81	3.95	4.35	4.30	4.85	8.05	4.30
0.12	3.45	3.84	3.18	4.86	7.80	4.56	4.30	5.00	4.00	6.85	10.80	7.20
0.16	4.02	4.14	3.60	6.87	11.85	6.81	4.75	5.05	4.70	10.26	16.41	8.85
0.20	3.57	4.68	3.15	8.52	15.21	7.53	5.15	5.00	4.80	15.71	22.86	14.11
0.24	3.57	4.23	3.45	11.25	20.56	10.32	4.60	5.10	4.50	21.11	30.06	18.76
0.28	3.96	4.98	3.33	14.97	26.71	14.52	4.45	4.95	3.50	29.26	41.57	27.01
0.32	3.51	4.35	3.39	21.55	33.55	21.07	4.15	5.25	3.80	39.22	48.57	37.17
0.36	3.48	4.44	3.54	28.33	39.46	25.51	3.95	4.40	3.40	50.47	58.48	47.67
0.40	3.93	4.98	3.99	34.78	49.00	32.95	4.70	5.45	4.35	60.33	69.08	57.53
0.44	4.35	4.08	3.51	41.47	56.78	40.99	4.55	5.60	4.85	69.58	76.64	67.58
0.48	4.56	4.95	3.63	50.32	65.00	49.90	4.90	5.65	5.05	77.64	85.24	77.24

	$T=100$					
	LL = 4			LL = 3		
b_4	1	2	3	1	2	3
0.00	4.10	4.10	3.70	4.80	3.70	5.20
0.04	4.01	5.61	4.10	4.91	7.81	5.80
0.08	5.40	6.10	4.10	8.11	11.51	6.61
0.12	4.40	5.91	4.90	11.61	15.51	11.21
0.16	4.80	4.61	5.81	19.42	26.22	17.92
0.20	6.31	6.10	5.60	30.32	34.93	27.22
0.24	4.51	5.20	5.21	42.94	48.54	41.14
0.28	4.21	4.91	4.00	56.15	62.65	52.85
0.32	4.60	5.50	4.20	72.36	74.77	69.96
0.36	4.11	4.00	3.41	82.87	81.38	82.17
0.40	4.80	6.41	4.51	89.29	90.49	88.89
0.44	4.91	5.31	4.71	95.29	95.39	94.79
0.48	4.81	4.51	4.31	97.70	98.40	97.39

4. CONCLUSION

In this paper we explore the problem of model specification in AR setup with commonly used lag length selection producers. The most interesting finding is that all such procedures are inherently flawed, as they cannot produce consistent results uniformly over the whole range of the true parameters values. The success of such procedures in detecting the true model depends heavily upon the true values of the regression parameters in the model. The errors in detecting the true lag order will be substantial for certain values of these parameters. Furthermore, ignoring the model selection uncertainties will produce serious problems in statistical inference to follow. The post-model-selection estimates are usually biased and inefficient and the tests of hypothesis based upon them are distorted.

REFERENCES

1. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Petrov, B.N., Csakim F., Akademiai Kiado (eds.). *Second International Symposium on Information Theory*, Budapest.
2. Basci, S. and Zaman, A. (1998). Effects of skewness and kurtosis on model selection criteria. *Economics Letters*, 59, 17-22.
3. Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997). Model Selection: An Integral Part of inference. *Biometrics*, 53, 603-618.
4. Claeskens, G. and Hjort, N.L. (2009). *Model Selection and Model Averaging*. Cambridge University Press.
5. Granger, C.W.J. and Jeon, Y. (2004). Forecasting performance of information criteria with many macro series. *Journal of Applied Statistics*, 31, 1227-1240.
6. Hacker, R.S. amd Hatemi-J, A. (2008). Optimal lag-length choice in stable and unstable VAR models under situations of homoscedasticity and ARCH. *Journal of Applied Statistics*, 35, 601-615.
7. Hafer, R. and Sheehan, R.G. (1989). The sensitivity of VAR forecast to alternative lag structures. *International Journal of Forecasting*, 5, 399-408.
8. Hannan, E.J. and Quinn, B.G. (1979). The determination of the order of an autoregression. *J. Roy. Statist. Soc.*, B, 41, 190-195.
9. Hatemi-J, A. and Hacker, R.S. (2007). Can the LR test be helpful in choosing the optimal lag order in the VAR model when information criteria suggest different lag orders? *Applied Economics* iFirst, 1-5.
10. Koreisha, S.G. and Pukkila, T. (1993). Determining the order of a vector autoregression when the number of component series is large. *Journal of Time Series Analysis*, 14, 47-69.
11. Kose, N. and Ucar, N. (2006). Effect of cross correlations in error terms on the model selection criteria for stationary VAR process. *Applied Economics Letters*, 13, 223-228.
12. Leeb, H. and Potscher, B.M. (2005). Model selection and inference: Facts and fictions. *Econometric Theory*, 21, 21-59.
13. Leeb, H. and Potscher, B.M. (2006). Can one estimate the conditional distribution of post-model-selection-estimators? *Annals of Statistics*, 34, 2554-2591.
14. Leeb, H. and Potscher, B.M. (2008). Can one estimate the unconditional distribution of post-model-selection-estimators? *Econometric Theory*, 24, 338-376.
15. Liew, V.K. and Chong, T.T. (2005). Autoregressive lag length selection criteria in the presence of ARCH errors. *Economics Bulletin*, 3, 1-5.
16. Lutkephol, H. (1993). *Introduction to Multiple Time Series Analysis*. Berlin, Heidelberg: Springer-Verlag.
17. Ng, S. and Perron, P. (2005). A note on the selection of time series models. *Oxford Bulletin of Economics and Statistics*, 67, 115-134.
18. Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 9-18

DETECTION OF OUTLIERS IN BIOEQUIVALENCE STUDIES DATA ANALYSIS WITH WILLIAMS DESIGN

A. Rasheed[1], S. Junaid[2] and T. Ahmad[3]

[1] Department of Research, Dow University of Health Sciences, Karachi, Pakistan. Email: rasheed_2526@yahoo.com

[2] Department of Statistics, University of Karachi, Karachi, Pakistan.

[3] Center for Bioequivalence Studies at ICCBS, University of Karachi, Karachi, Pakistan.

BACKGROUND

Drug Regulatory agencies all over the world generally discourage exclusion of outliers in a BE (BE) study; on the other hand Good Statistical Practices requires it. If the decision rules for identifying the outliers are clearly mentioned before the start of the study and laid down in protocol by the responsible biostatistician in collaboration with clinicians, the problem of outliers can be dealt smartly without jeopardizing the whole study for redoing. The purpose of this article is to introduce procedure for reliably detecting outlier subject(s) with Williams design.

EXPERIMENTAL

Literature review reveals many different methods for the detection of outlier values in BE studies; most of them are for BE of two treatments. For BE studies with more than two treatments use of Williams design seems imperative; but inclusion and deletion of outlying subjects may lead to profound effect on conclusion of BE which in turn may be dangerous for the health. The suggested method is an adjustment to a previously introduced method using exploratory data analysis technique such as principle component analysis and Andrews curves.

KEY WORDS

Bioequivalence; outliers; Williams design; Exploratory data analysis; Principal component; Andrews curves

1. INTRODUCTION

1.1 Back Ground and Rationale

Tukey (1977) proposed exploratory data analysis (EDA) a methodology for data analysis which generally employs a variety of techniques most of them are graphical. These methods are useful in detecting outliers, uncover underlying structure, maximize insight into a data set, extract important variables and determine optimal factor settings.

Enachescu and Enachescu (2009) discussed the EDA technique such as Andrews curves and Principal Component Analysis (PCA). Using PCA (Enachescu and Enachescu, 2009) mentioned that "first two principal axes in PCA span one such plane, providing a projection such that the variation in the projected data is maximized over all possible 2-D projections and applied these techniques to identify the outlying subjects in 2x2 crossover BE trial. In present work we extended these techniques to Williams design; a special variety of Crossover/Latin square designs. With the help of our proposed approach, gathering information regarding outliers subjects and their identification in a BE data with more than two formulations becomes easily possible.

1.2 Williams Design

In crossover designs if each formulation appears in same number of times within each sequence is called 'uniform within sequence' and if each formulation appears the same number of times within each period than is called 'uniform within period'. A crossover design is called uniform if it is uniform within sequences and within periods. A Latin square, in which every treatment is occurred once and only once in each row and each column yields uniform crossover designs. In a balanced design, each of the treatments occur the same number of times in each period and the number of subjects who receive treatment i in one period and treatment j in the next period is the same for all $i \neq j$. (Jones and Kenward , 2003)

Williams (1949) introduced a crossover design in which every treatment follows every other treatment the same number of times called Williams design that is the treatments are balanced over prefix and suffix. These Williams designs require fewer subjects than those based on the complete sets of orthogonal Latin squares (Wang et al., 2009). In Williams design when the number of formulations are even than balance can be achieved by a single Latin square design, but when the number of formulations are odd than two Latin square designs are needed.

2. EXPERIMENTAL TECHNIQUES

In this present work we used two EDA techniques, Andrews curve and PCA to ease the problem of detecting outliers in BE studies with more than two treatments. Modified z-scores method commonly used method for outlier detection also used here to insert the fictitious outliers in original data set. In Modified z-scores method subject having absolute z-scores greater than 3.5 are labeled as potential outliers.

EXPLORATORY DATA ANALYSIS TECHNIQUES

2.1 Andrews curves

Andrews (1972) introduced a method of visualization for multivariate data. Each multidimensional data point is mapped into a periodic function

$$f_X(t) = X_1 / \sqrt{2} + X_2 \sin t + X_3 \cos t + ... \quad (1)$$

This graphical approach displays a point in multidimensional space by a two-dimensional curve using the function $f_x(t)$ given above in the interval $-\prod < \tau < \prod$. The advantage of this method is that it allows the inclusion of many dimensions. A collection of multidimensional points, that is, a multivariate data set, is displayed as a group of curves. In this method numbers of variables are unlimited. These curves are dependent on the order of the variables. Lower frequency terms (i.e., those that are first in the sum given in the above equation) exert more influence on the shape of the curves we can get more information about data by re-ordering the variables and viewing the resulting plot. Observations showing quite apparent different curves are considered as outliers.

2.2 Principal component analysis

The objective of PCA is to discover or to reduce the dimensionality of the data set and identify new meaningful underlying variables.

In PCA number of (possibly) correlated variables are transformed into (smaller) number of variables which are uncorrelated called principal components. Large amount of variability is accounted by the first PC and each succeeding PC accounts for as much of the remaining variability as possible.

PCA can be performed either by using a covariance matrix or correlation matrix both matrices are calculated from the data matrix, if one is using correlation matrix so firs variables should be standardized.

2.2.1 Eigen analysis:

Eigen analysis is a mathematical technique used in the PCA, in this technique Eigen values and Eigen vectors of a square symmetric matrix with sums of squares and cross products are calculated. The eigen-vector associated with the largest Eigen-value has the same direction as the first principal component. The Eigen-vector associated with the second largest Eigen-value determines the direction of the second principal component.

X is $p \times n$ the data matrix where($p =$ number of variables and $n =$ number of observation) , Σ is covariance matrix obtained from the data matrix X , and Z is the standardized data matrix, R is correlation matrix obtained from the data matrix Z . λ_i is called Eigen value denotes the variance of the i-th PC (i.e., $\lambda_i = Var(ith\, PC)$ that can be calculated by setting $|R - \lambda I| = 0$ Where I is the identity matrix. $U_i = \beta^{(i)'} Z$ is called the i-th PC where $\beta^{(i)}$ is denote the i-th eigen vector that can be calculated by setting $(R - \lambda_i I)\beta^{(i)} = 0$ where $\beta = \left[\beta^{(1)}\ \beta^{(2)}\ \beta^{(3)} \ldots\ \beta^{(p)}\right]$ and each of $\beta^{(i)}$ is defined as $\beta^{(i)'} = \left[\beta_1^{(i)}\ \beta_2^{(i)}\ \beta_3^{(i)} \ldots\ \beta_p^{(i)}\right]$ with $\beta^{(\iota)'}\beta^{(\iota)} = \beta_1^{(\iota)^2} + \beta_2^{(\iota)^2} + \beta_3^{(\iota)^2} + \ldots + \beta_p^{(\iota)^2} = 1$. The sum of the Eigen-values equals the trace of the square matrix (i.e., $tr\,(\Sigma) = \sum_{i=1}^{p} \lambda_i$) and the maximum number of Eigen-vectors equals the number of rows (or columns) of this matrix.

Enachescu and Enachescu (2009) has mentioned that *"For normally distributed observation* U_i/λ_i *are independent* $\chi^2_{1,j}$ *variables. Consider* $\sum_{i=1}^{p} \lambda_i \chi^2_{1,j}$ *the weighted sum of square distance to zero of the projected data into principal factorial plane, with* $E\left(\sum_{i=1}^{p} \lambda_i \chi^2_{1,j}\right) = \sum_{i=1}^{p} \lambda_i = p$ *and* $Var\left(\sum_{i=1}^{p} \lambda_i \chi^2_{1,j}\right) = 2\sum_{i=1}^{p} \lambda_i^2$. *Now the Observations with a square distance greater than* m *(the rule of* 2σ *) may be considered as outliers where".*

3. PROPOSED NEW TECHNIQUE

EDA techniques, Andrews curve and PCA have been used to detect outliers in BE studies with two treatments, here we advocated Andrews curve and PCA for gathering information regarding outliers in Williams design with a reference (R) and two treatments formulations (T1 and T2).

Andrews curve function is defined as

$$f(t) = R/\sqrt{2} + T_1 \sin t + T_2 \cos t \qquad -\pi < t < \pi \qquad i = 1, 2, 3 \ldots 12 \tag{2}$$

Each observation is projected onto a set of orthogonal basis functions represented by sines and cosines and then plotted. Thus, each sample point is now represented by a curve. Observations with identical curves show the possible outlying subjects.

In PCA, X is $p \times n$ the data matrix where (p= 3, number of formulations and n=12, number of observation for each formulation) , $\sum$ is covariance matrix obtained from the data matrix X , and Z is the standardized data matrix, R is correlation matrix obtained from the data matrix Z and λ_1, λ_2 *and* λ_3 are Eigen values denoting the variances of first, second and third PC respectively. $\sum_{i=1}^{3} \lambda_i \chi^2_{1,j}$ the weighted sum of square distance to zero of the projected data into principal factorial plane, with mean $p = 3$ and variance $2\sum_{i=1}^{3} \lambda_i^2$. Now the observations with a square distance greater than m (the rule of 2σ) may be considered as outliers where $m = 3 + 2\sqrt{2\sum_{i=1}^{3} \lambda_i^2}$.

4. APPLICATION AND VALIDATION OF PROPOSED TECHNIQUE

In the present work we selected a data set of Areas Under the Curve; AUC from a bioequivalence study reported by (Purich, 1980). In the study twelve healthy volunteer were employed to investigate the bioequivalence of two test tablets formulations in comparison to a reference solution. (Chow and Liu, 2000) mentioned that no assignment of sequences and periods was given by (Purich, 1980). Thus for the purpose of illustration (Chow and Liu, 2000) assigned subject 1 and 2 to sequence 1; 3 and 4 to sequence 2; 5 and 6 to sequence 3; 7 and 8 to sequence 4; 9 and 10 to

Table 1: AUC data set with a reference and two test formulation (domestic and European tablets)

Sequence	Subject	R	T1	T2
(R, T2, T1)	1	5.68	4.21	6.83
	2	3.6	5.01	5.78
(T1,R, T2)	3	3.55	5.07	4.49
	4	7.31	7.42	7.86
(T2,T1,R)	5	6.59	7.72	7.26
	6	9.68	8.91	9.04
(T2,T1,R)	7	4.63	7.23	5.06
	8	8.75	7.59	4.82
(T2,R,T1)	9	7.25	7.88	9.02
	10	5	7.84	7.79
(R,T1, T2)	11	4.63	6.77	5.72
	12	3.87	7.62	6.74

sequence 5; 11 and 12 to sequence 6. Table 1 gives this AUC data set after rearrangement of reference and period according to Williams design for comparing the three formulations. The numerical results given bellow are obtained with popular software SAS.

To apply the above defined EDA techniques for determining the possible outlying subjects here AUC data set is presented formulations wise on the linear and logarithmic scale in Table 2.

Table 2: AUC(0-inf) data set for a reference and two test formulations on both linear and logarithmic scales

Subject	Linear Scale			Logarithmic Scale		
	R	T1	T2	R	T1	T2
1	5.68	6.83	4.21	1.737	1.921	1.437
2	3.6	5.78	5.01	1.281	1.754	1.611
3	5.07	3.55	4.49	1.623	1.267	1.502
4	7.42	7.31	7.86	2.004	1.989	2.062
5	7.26	7.72	6.59	1.982	2.044	1.886
6	9.04	8.91	9.68	2.202	2.187	2.27
7	5.06	4.63	7.23	1.621	1.533	1.978
8	4.82	8.75	7.59	1.573	2.169	2.027
9	7.88	9.02	7.25	2.064	2.199	1.981
10	7.84	7.79	5	2.059	2.053	1.609
11	4.63	6.77	5.72	1.533	1.913	1.744
12	3.87	7.62	6.74	1.353	2.031	1.908

4.1 Andrews curves

The Andrews curves for this data set are

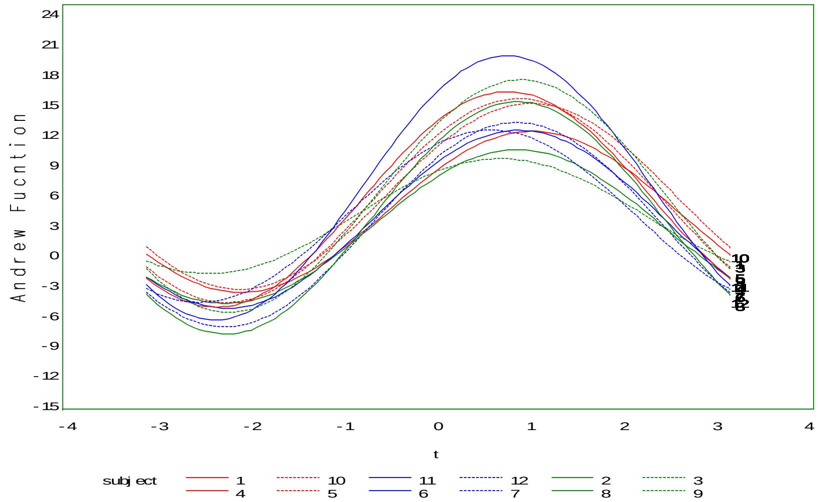
For linear scale

$$f(t) = R/\sqrt{2} + T_1 \sin t + T_2 \cos t \qquad -\pi < t < \pi \quad i = 1, 2, 3 \ldots 12 \tag{3}$$

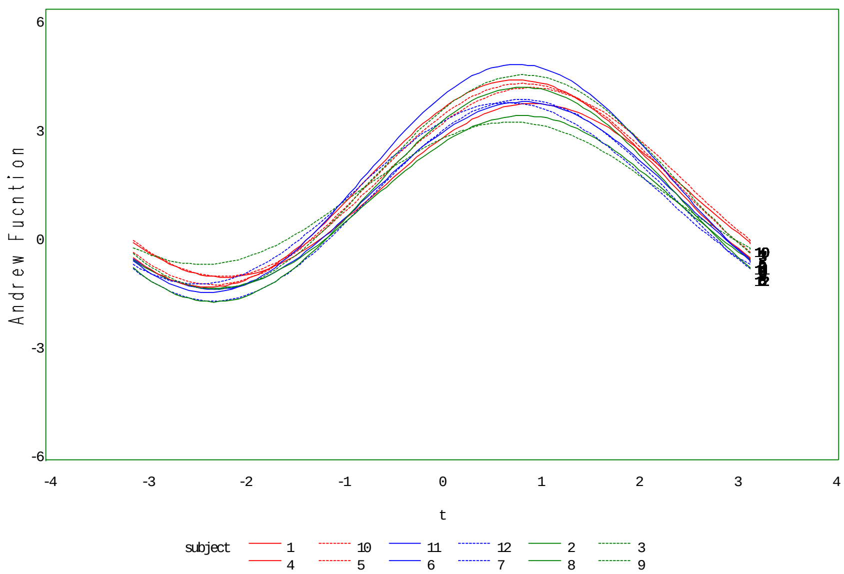
For logarithmic scale

$$f(t) = \ln(R)/\sqrt{2} + \ln(T_1)\sin t + \ln(T_2)\cos t \quad -\pi < t < \pi \quad i = 1, 2, 3 \ldots 12 \tag{4}$$

From linear and logarithmic Andrew Curves given in Figures-1a and 1b it is very evident that there is no curve reveals distinct or outlying behavior suggesting any subject a possible outlier.



1a



1b

Fig. 1: Andrews curves 1a: linear scale data; 1b: logarithmic scale data

4.2 Principal component analysis

We employed the same data set given in the tables 3 by using the principal components analysis we obtained the following results with correlation matrix R.

Table 3:
Eigen-values of the correlation matrix R for the linear and logarithmic scales

	Linear Scale				Logarithmic Scale			
	Eigen values	Differences	Proportion	Cumulative	Eigen values	Differences	Proportion	Cumulative
1	2.0421	1.5176	0.6807	0.6807	1.9098	1.2942	0.6366	0.6366
2	0.5245	0.091	0.1748	0.8555	0.6156	0.141	0.2052	0.8418
3	0.4335		0.1445	1	0.4746		0.1582	1

The threshold value and the square distance to the zero for each observation are given in table 4 for the linear and logarithmic scale

Table 4: Threshold value and the squared distance on linear and logarithmic scales

Subject	Linear Scale		Logarithmic Scale	
	Squared distance	Threshold	Squared distance	Threshold
1	1.9756731	9.088	2.4626745	8.832
2	3.1857674		3.5669672	
3	6.0492623		7.3577339	
4	1.4066709		1.5539635	
5	0.6469274		0.8066935	
6	8.0766893		6.05619	
7	2.5784951		2.4345028	
8	1.9496287		1.7079329	
9	2.6834404		2.3762979	
10	2.0350777		2.0368382	
11	0.8310584		0.6595056	
12	1.5813093		1.9807000	

Any observation with square distance greater than corresponding threshold value may be considered as outlier. As we can see that on both scales there is no observation with square distance greater threshold.

AS evident from the above analysis no observation is found as an outlier in the above data set. In order to verify the proposed extended EDA techniques (Andrews curves and PCA) it was felt imperative to introduce intentionally some outlying values in the original data. Accordingly we made some changes in the original data set by replacing few values with fictitious (obvious outlier) values.

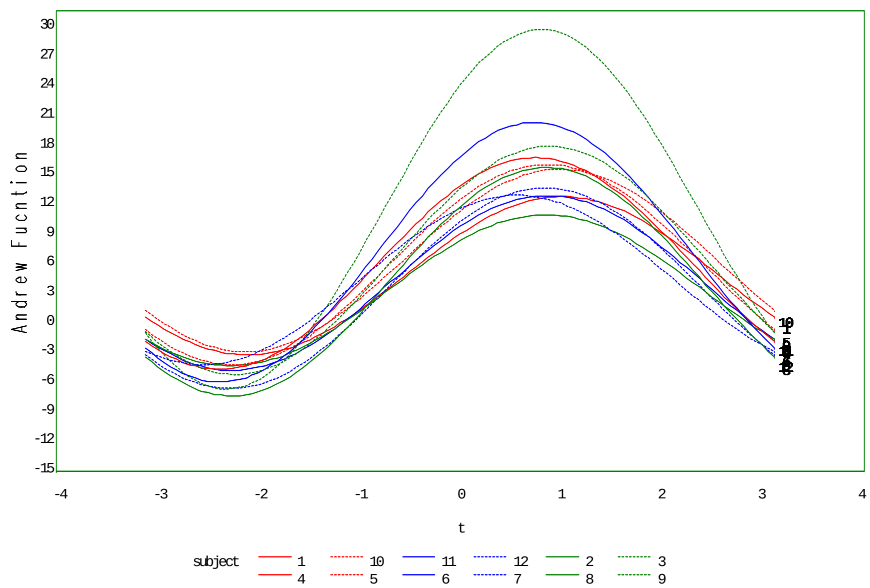
We replaced some original values with few fictitious extreme (very high and very low) values for each treatment (i.e., R, T1, and T2) which were identified as outliers by certainity by confirming them as outliers through modified z-scores method. We carried this exercise two times.

In first instance we randomly selected a subject 3 from original data set and replaced its values for all three treatments (5.07, 3.55 and 4.49) by fictitious values (15.2, 13.2 and 12.56) previously identified as outliers.

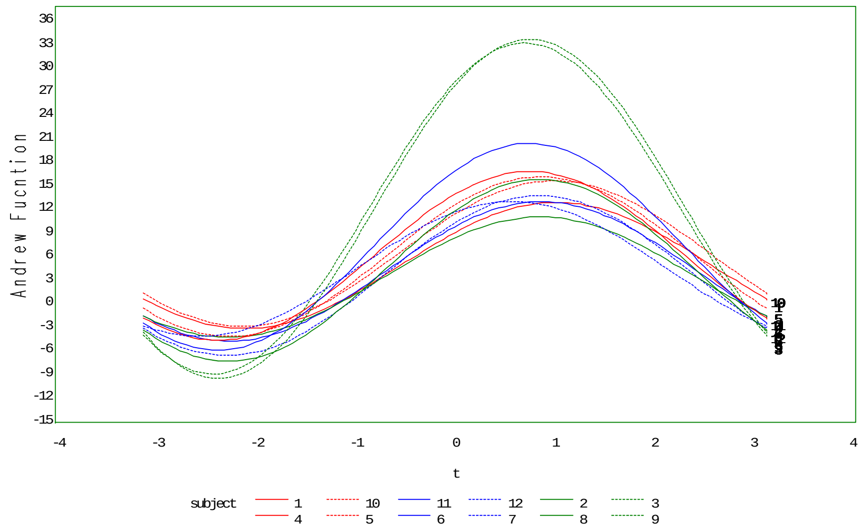
In second instance we selected two random subjects 3 and 9 from original data set and replaced their values by fictitious values previously identified as outliers. Subject 3 values (5.07, 3.55 and 4.49) replaced by (15.95, 13.56 and 16.12) and subject 9 values (7.88, 9.02 and 7.25) replaced by (15.98, 14.80 and 15.7).

On these two change data sets we applied the proposed EDA techniques to confirm the validity of these techniques that whether these techniques identify the outliers in these two data sets.

We are glad to report that both proposed EDA techniques Andrews curves and PCA correctly identified subject 3 for fictitious data set 1 and subject 3 and 9 for data set 2 on linear scale as well as on logarithmic scale. In Figure 2 (2a and 2b) the outlying subjects can be visualized very clearly and similarly in Table-V the Threshold values and the squared distance for fictitious data sets 1 and 2 can be seen. In Table-V Subject 3 in data set 1 and subjects 3 and 9 in data set 2 reveal squared distances are greater than the threshold values.



2a



2b

Fig. 2: (a) Andrews curves for the linear scale fictitious data set 1;
(b) Andrews curves for the linear scale fictitious data set 2

Table 5: Threshold value and the squared distance for fictitious data sets 1 and 2 from method of principal component analysis

Subject	Data Set # 1		Data Set # 2	
	Squared distance	Threshold	Squared distance	Threshold
1	2.0180746	10.431	1.4646404	11.072
2	2.8926605		2.277750	
3	19.067802		11.10280	
4	0.2055743		0.1342879	
5	0.0750083		0.2062721	
6	1.9922433		0.3078588	
7	2.6616014		1.9995872	
8	0.6348092		0.4616749	
9	0.4082725		12.259875	
10	0.9686317		0.6733027	
11	1.1422837		1.1532807	
12	0.9330383		0.9586664	

CONCLUSION

Through this work we report and recommend an extended exploratory data analysis techniques for identification of outliers in a Williams design data set generated during bioequivalence evaluation. In present research for identification of outliers we successfully applied the EDA techniques, Andrews curves and principal component analysis for the bioequivalence data set with more than two treatments.

REFERENCES

1. Andrews D. (1972). Plots of high-dimensional data. *Biometrics*, 28, 125-136.
2. Chow S-C. and Liu J-P. (2000). *Design and analysis of bioavailability and bioequivalence studies*. 2nd Edn. Dekker: New York.
3. Enachescu D. and Enachescu C. (2009). A new approach for outlying records in bioequivalence trials. Proceedings of the *13th International Conference on Applied Stochastic Models and Data Analysis*, Vilnius, Lithuania; 250-257.
4. Center for Drug Evaluation and Research (2001). *Statistical approaches to establishing bioequivalence*. Guidelines for Industry, U.S. Department of Health and Human Services Food and Drug Administration.
5. Jones B. and Kenward M.G. (2003). *Design and Analysis of Cross-Over Trials*. 2nd Edn. Chapman & Hall: London.
6. Purich E. (1980). *Bioavailability/Bioequivalence Regulations: An FDA Perspective in Drug Absorption and Disposition: Statistical considerations*, K.S Albert, Edn. American Pharmaceutical Association, Academy of Pharmaceutical Sciences, Washington, DC, 15-137.
7. Tukey J-W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.
8. Wang B-S., Wang X-J. and Gong L-K. (2009). The construction of a Williams design and randomization in crossover clinical trials in using SAS. *Journal of statistical software*, 29.
9. Williams E.J. (1949). Experimental Designs Balanced for the Estimation of Residual Effects of Treatments. *Aust. J. Sci. Res.*, 2(3), 149-168.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 19-24

A RETROSPECTIVE APPROACH IN SURVIVAL ANALYSIS FOR ESTIMATING DURATION FROM CHEST PAIN TO CORONARY ARTERY DISEASE (CAD)

Mehwish Hussain[1] and **Nazeer Khan**[2]

[1] Department of Statistics, University of Karachi,
Karachi, Pakistan; Email: mehvish.hussain@yahoo.com

[2] Department of Research, Dow University of Health Sciences,
Karachi, Pakistan; Email: n.khan@duhs.edu.pk

ABSTRACT

Coronary Artery Disease (CAD) is one of the most occurring and fatal disease in any region of the world. One of the perceptible symptoms of CAD is chest pain. But, every patient with chest pain not necessarily suffers with CAD as well. Also, all CAD patients may not bear chest pain in past. The duration from chest pain to CAD diagnosis is sill unevaluated. Survival analysis is one of the statistical technique which deals with time dependent event from any onset of study using prospective approach. In this paper, a retrospective approach was used for scrutinizing duration from chest pain in past to the diagnosis of CAD by the examiner at the time of study. Non-parametric techniques of survival analysis were used to estimate the particular duration.

KEYWORDS

Cross sectional study; duration analysis; CAD.

1. INTRODUCTION

The time it takes for a patient having cardiac chest pain to the diagnosis of CAD is getting significance. The two study designs can be referred for estimating such duration are prospective cohort and retrospective cohort (Kieding, Kvist, Hartvig, Tvede, & Juul, 2002). In prospective cohort, the patient with chest pain referred to cardiologist, who with time by time certain diagnosis tools signifies the patient having CAD. Then certain parameters would be examined for the treatment of CAD. In retrospective cohort design, the CAD patients are observed and their historical symptoms are scrutinized.

The dilemma with prospective cohort study is that it requires long lag of time, demanding recruitment and self – selection effort and keen concerns by the sample with accurate follow – up (Bonde, et al. 1998). Whereas, retrospective cohort study puts up with recall bias, mixes experiences with over a long period time calendar, and also usually suffers with missed data.

It would be therefore prudent to follow a third study design i.e. a cross-sectional study, in which patient currently suffering with CAD, have chest pain in past with observed duration to CAD diagnosed. This duration of chest pain to CAD diagnosis will be estimating for the whole population.

A new design has been developed for estimating time to pregnancy from current durations from a cross-sectional sample by Kieding (2002). He derived distribution of current duration and then parametric and non-parametric estimation were also studied (Kieding, Kvist, Hartvig, Tvede, & Juul, 2002).

2. DISTRIBUTION OF DURATION FROM CHEST PAIN TO CAD

For defining distribution of duration from chest pain to CAD, Kieding (2002) will be followed.

Let 'T' will be defined as *duration of chest pain to CAD*, and 'U' be the time for the patients suffered from chest pain but currently they are controlled with respect to CAD. Also let 'V' will be the discontinuation of follow-up since the start of chest pain (missing values in CAD). We are interesting in the distribution of T. Let $X = T \wedge U$ be the duration time of chest pain to whether having CAD or not, with probability density function $f(x)$ and survival function $S(x) = \int_x^\infty f(a)da$ and expectation $\mu_x = \int_x^\infty xf(x)dx = \int_x^\infty S(x)dx$, which we shall assume finite. Assuming the initiations happen according to Poisson process in a calendar time t with intensity $\beta(t)$, the observed experienced duration at t_0, $Y = X \wedge V = T \wedge U \wedge V$, has density proportional to $\beta(t_0 - y)\ S(y)$.

Since, in cross sectional sample, short calendar interval is considered. So, Cox (1969) showed that Y will be distributed as a backward recurrence time in a renewal process in equilibrium with renewal distribution $f(x)$ (Cox, 1969), that is, density of Y is

$$g(Y) = S(y)/\mu_X\ ; \quad 0 < g(0) < \infty$$

Thus, Y has decreasing density proportional to survival function of X .

An estimate of survival function of X can be obtained by setting

$$\hat{S}(x) = \hat{g}(x)/\hat{g}(0)$$

3. NON PARAMETRIC ESTIMATION

The non-parametric maximum likelihood estimate (NPMLE) for a decreasing density was derived by Grenander (1956) and generalized to right censored by Danby and Vardi (1986), Woodroof and Sun (1993) demonstrated that in an uncensored case $\hat{g}(0)$ is inconsistent and suggested penalized NPMLE of $g(0)$ (Grenander, 1956 Danby, et al. 1986, Woodroofe, et al. 1993). Kieding (2002) worked over this estimation with short interval of time, i.e., censoring the duration at 3 years and more. The NPMLE for the left-continuous step function by Kieding (2002) was then,

$$\widehat{g_n}(y) \min_{0 \le r \le k-1} \max_{k \le s \le n} \frac{s-r}{n(y_s - y_r)}, \quad y_{k-1} < k \le y_k,$$

$$\widehat{g_n}(y) = 0 \ \text{ for } y > y_n .$$

4. DATA DESCRIPTION

A community based study was conducted in Kingdom of Saudi Arabia during 1995 – 2000 to determine the prevalence of CAD among Saudi population. Subjects were included between 30 – 70 years of ages. The number of patients who were diagnosed to have CAD was established in finding of one or more of the following criteria: either physician's clinical assessment of the chest pain as anginal, previous MI, or findings of evidence of previous MI by ECG. For more details of the methodology, see *Al-Nozha, et al. 2004*.

Out of 17232 subjects, 1036 were diagnosed with CAD. Among those 884 (85.3%) had chest pain in past. Age, sex, body mass index (BMI), waist circumference, systolic & diastolic blood pressures (SBP & DBP), smoking status, fasting blood sugar (FBS), serum cholesterol, serum triglycerides and high density lipoprotein (HDL) were the significant factors associated with CAD (Al-Nozha, et al. 2004).

5. ESTIMATES FOR DURATION FROM CHEST PAIN TO CAD

The analyses were done in Stata v. 10.0. The data were converted from snapshots (cross sectional) to time dependent. The variable, time of chest pain, was put in time variable and CAD presence was set as event variable.

5.1 Descriptive Duration Analysis

The minimum duration from chest pain to the diagnosis of CAD was 1 year and maximum was 40 years. The median duration was 5 years with 2 and 10 were 1st and 3rd quartiles of the same respectively. The incidence rate was 0.1444.

5.2 Kaplan-Meier Estimates

Kaplan-Meier estimator reveals that 88% of the patients with chest pain did not experience CAD (95% CI: 0.8649 – 0.8951) (Fig. 1). This endurance reached about half of the above proportion (43%) within 5 years of the duration. All the patients with 17 years of chest pain were diagnosed as CAD patients with survival function of 8.26% (95% CI: 0.0656 – 0.1020). A similar panorama is suffered by the CAD patients with duration of at least 25 years of chest pain with very low survival function.

5.3 Cox Regression Model

Assumption Fulfillment

Fig. 2 portrays the fulfillment of proportional hazard (PH) function required for fitting Cox regression model. The graph clearly shows the parallel curves of –ln(-ln) survivor curves comparing male and females, indicating PH assumption is satisfied.

The Model

Duration Model with Natural Factors as Covariates:

$$\text{Duration} = D_0\,(t)\,\exp\,(0.870\ \text{Gender} + 0.999\ \text{Age})$$

OR

$$\text{Log Duration} = \underset{(0.014)}{-0.133\ \text{Gender}} - \underset{(0.009)}{0.001\ \text{Age}}$$

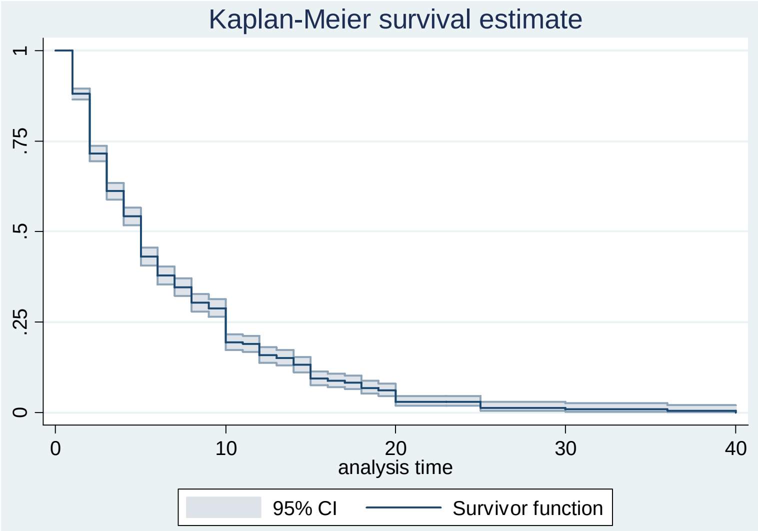


Fig. 1: Kaplan-Meier Curves portraying duration from chest pain to CAD in terms of Survivor function with 95% confidence interval.

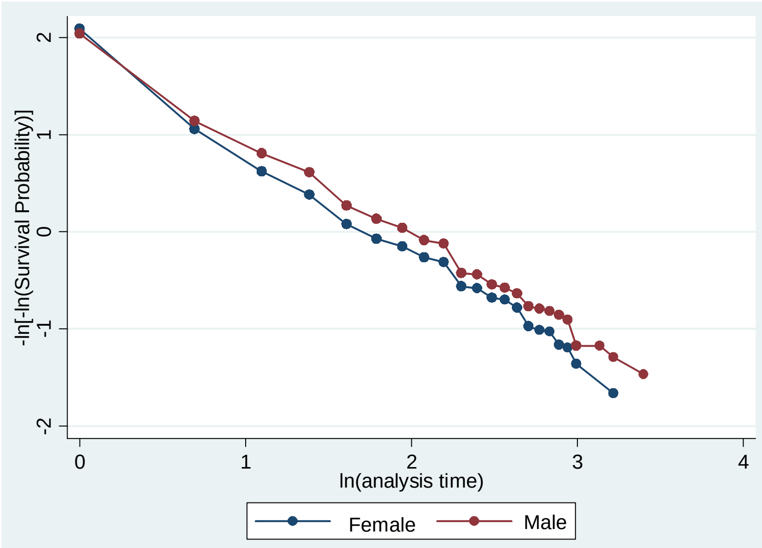


Fig. 2: Assumption of Cox-regression model in line with respect to gender.

Cox regression model portrays that males with chest pain in past are 0.87 times hazardous than females for having CAD. Also, with the increment of one year of age, subjects with chest pain are 2.72 fold higher to have CAD. The clinical factors i.e. triglyceride level, fasting blood sugar, high density level, diabetic status were not significantly hazardous for the CAD patient with chest pain in past.

Duration Model with Natural and Clinical Factors as Covariates

The effect of natural factors with clinical factors on duration from chest pain to CAD was also evaluated. Table 1 (a & b) shows the Cox regression model with hazard ration of duration and log duration as regressands respectively. Gender was the only variables which described the duration from chest pain to CAD significantly.

Table 1:
a) Cox regression model statistics for estimating duration from natural and clinical factors

Variable	Hazard Ratio	S.E (H.R)	Test Statistic	P Value	[95% Conf. Interval]	
Age	0.999	0.003	-0.19	0.850	0.994	1.005
Gender	1.218	0.078	3.09	0.002	1.075	1.381
Marital Status	1.067	0.070	0.99	0.322	0.939	1.212
Hypertension	0.935	0.083	-0.76	0.448	0.786	1.112
Diabetic Status	0.932	0.065	-1.01	0.311	0.813	1.068
High Density Level	0.948	0.031	-1.62	0.105	0.888	1.011
Triglyceride Nitrate	0.556	0.214	-1.52	0.128	0.265	1.182
Fasting Cholesterol	1.007	0.040	0.17	0.863	0.932	1.088
Fasting Glucose	1.030	0.038	0.80	0.422	0.958	1.107
Body Mass Index	0.968	0.036	-0.88	0.379	0.900	1.041

b) Cox regression model statistics for estimating log (duration) from natural and clinical factors

Variable	B	S.E (B)	Test Statistic	P Value	[95% Conf. Interval]	
Age	-0.001	0.003	-0.19	0.850	-0.006	0.005
Gender	0.197	0.064	3.09	0.002	0.072	0.323
Marital Status	0.065	0.065	0.99	0.322	-0.063	0.192
Hypertension	-0.067	0.088	-0.76	0.448	-0.241	0.106
Diabetic Status	-0.071	0.070	-1.01	0.311	-0.207	0.066
High Density Level	-0.054	0.033	-1.62	0.105	-0.118	0.011
Triglyceride Nitrate	-0.581	0.382	-1.52	0.128	-1.328	0.167
Fasting Cholesterol	0.007	0.040	0.17	0.863	-0.071	0.085
Fasting Glucose	0.029	0.037	0.80	0.422	-0.043	0.102
Body Mass Index	-0.032	0.037	-0.88	0.379	-0.105	0.040

6. CONCLUSION

Survival analysis, a prospective time dependent technique, can be applied to cross sectional data with retrospective approach. As seen in the above data, the estimated time from chest pain to CAD is 5 years. Males are less hazardous than females.

7. DISCUSSION

In the analysis of longitudinal data, two models are usually used, one is Cox proportional hazard model and another is accelerated failure time model. The observation of this time classifies it into either current duration or length biased sampled data. In medical setting, length biased data requires time need to observe the full duration of the disease of the sampled patients whereas current duration requires knowledge of the duration of disease of group of patients up to the present (Oluyede, 2007). Much work has been done on follow-up study of survival time of CAD patients. A follow up study of two years indicated the average survival time of CAD patients from day of diagnosis to the day of discharge equals 6.81 months. Only family history was the significant factor whereas age, sex and presence and absence of risk factors like hyperlipidemia, diabetes, hypertension, smoking etc. were the insignificant factors to survival times of CAD (Mingala & Estolano, 2007). Current duration survival analysis was introduced by Kieding (2002) while estimating time to pregnancy from current durations in a cross-sectional sample. Though, Ali M.M. analyzed incomplete durations with application to contraceptive use (Ali, Marshall, & Babiker, 2001). An attempt on analysis duration of CAD from chest pain from a cross sectional sample is done in this paper. The analysis present in this paper for estimating duration of CAD was not much different than Mingala (2007). Nevertheless, the parametric estimates of duration from chest pain to CAD are still not known. Some authors derived accelerated failure time models for cross sectional study. The next loom will be for estimating duration model for CAD.

REFERENCES

1. Ali, M.M., Marshall, T. and Babiker, A.G. (2001). Analysis of Incomplete Durations with Application to Contraceptive Use. *J. Roy. Statist. Soc.*, 164(3), 549-563.
2. Al-Nozha, M.M., Arafah, M.R., Al-Mazrou, Y.Y., Al-Maatouq, M.A., Khan, N.B., Khalil, Z.M. et al. (2004). Coronary Artery Disease in Saudi Arabia. *Saudi Med. J.,* 25(9), 1165-71.
3. Bonde, J.P., Ernst, E., Jensen, T.K., Hjollund, N.H., Kolstad, H., Henriksen, T.B. et al. (1998). Relation between semen quality and fertility: a population-based study of 430 first-pregnancy planners. *The Lancet*, 352(9135), 1172-77.
4. Cox, D.R. (1969). Some sampling problems in technology. In N.L. Johnson, *New Developments in Survey Sampling* (506-527). New York: Wiley.
5. Danby, L. and Verdy, Y. (1986). The survival curve with decreasing density. *Technometrics* , 28, 359-367.
6. Grenander, U. (1956). On the theory of mortality measurement, part II. *Skandinavisk Aktuarietidskrift , 39*, 125-153.
7. Kieding, N., Kvist, K., Hartvig, H., Tvede, M. and Juul, S. (2002). Estimating time to pregnancy from current durations in a cross-sectional sample. *Biostatistics, 3*(4), 565-78.
8. Mingala, M.C. and Estolano, M.P. (2007). Parametric and Nonparametric Survival Analysis of Myocardial Infarction Patients Admitted in Dr. Paulino J. Garcia Memorial Research and Medical Center in Cabanatuan City. *10th National Convention on Statistics (NCS).* EDSA - Shangrila Hotel.
9. Oluyede, B. (2007). *Length-Biased and Current Duration Sampling.* Georgia Southern University, Department of Mathematics.
10. Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of f(0+) when f is nonincreasing. *Statistica Sinica , 3*, 501-515.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 25-36

ON THE BAYESIAN ANALYSIS OF MAXWELL DISTRIBUTION UNDER DIFFERENT LOSS FUNCTIONS

Syed Mohsin Ali Kazmi, Muhammad Aslam and **Sajid Ali**
Department of Statistics, Quaid-i-Azam University Islamabad, Pakistan
Email: mohsinstats@yahoo.com

ABSTRACT

In the present paper, Maxwell distribution is considered for Bayesian analysis. The properties of Bayes estimators of the parameter are studied under different loss functions via simulated and real life data. A comprehensive simulation scheme is used under Noninformative priors. The loss functions are compared through posterior risk.

KEYWORDS

Bayesian Estimation; Squared Error Loss Function; Modified LINEX Loss Function; Entropy Loss Function; Square Logarithmic Loss Function; Credible interval (CI); Highest Posterior Density (HPD); Noninformative prior; Posterior risk.

1. INTRODUCTION

The Maxwell distribution is a probability distribution with application in physics and chemistry. The most frequent application is in the field of statistical mechanics. The temperature of any (massive) physical system is the result of the motions of the molecules and atoms which make up the system. These particles have a range of different velocities, and the velocity of any single particle constantly changes due to collisions with other particles. However, the fraction of a large number of particles within a particular velocity range is nearly constant. Then Maxwell distribution of velocities specifies this fraction, for any velocity range as a function of the temperature of the system.

Tyagi and Bhattacharya (1989a, b) considered Maxwell distribution as a lifetime model for the first time. They obtained Bayes estimates and minimum variance unbiased estimators of the parameter ad reliability function for the Maxwell distribution. Chaturvedi and Rani (1998) generalized Maxwell distribution by introducing one more parameter. They obtained Classical and Bayesian estimation procedures for this generalized distribution. Bekker and Roux (2005) studied Empirical Bayes estimation for Maxwell distribution. These studies give mathematical handling to Maxwell distribution but ignore the application aspect of the Maxwell distribution.

In this paper, random observations taken from Inverse Transform method of simulation, and the computations involved are conducted using Minitab, Mathematica and Excel. The Maxwell model and its likelihood are defined in section 2. Sections 3 and 4 elucidate the posterior distribution using uniform prior and Jeffreys prior respectively. Section 5 describe real life example with graphs of the posterior distribution. Credible intervals and Highest Posterior Density are elaborated in section 6. Section 7 presents the Bayes factors for different hypotheses. The different loss functions with real life examples as well as simulation study are performed in Sections 8, 9 and 10. Some concluding remarks are given in section 11.

2. LIKELIHOOD FUNCTION FOR THE MAXWELL MODEL

Let $X_1, X_2, ..., X_n$ be a random sample taken from the Maxwell distribution with unknown parameter θ, i.e.

$$f(x) = \frac{4}{\sqrt{\pi}} \frac{1}{\theta^{\frac{3}{2}}} x^2 \exp\left(-\frac{x^2}{\theta}\right) \qquad \theta > 0, 0 \le x \le \infty \tag{1}$$

The likelihood function of the Maxwell distribution with unknown parameter θ is:

$$L(\theta, \mathrm{x}) = \left(\frac{4}{\sqrt{\pi}}\right)^n \frac{1}{\theta^{\frac{3n}{2}}} \prod_{i=1}^{n} x_i^2 \exp\left(-\sum_{i=1}^{n} x_i^2 / \theta\right) \tag{2}$$

2.1 Objective Bayesian Analysis of Maxwell Model

Much effort has been expended by the Bayesians in the research for the so-called Noninformative priors or Objective Bayesian which represent lack of information about the parameters of a model. Here we use only two Noninformative priors, the Uniform and the Jeffreys priors.

3. USING THE UNIFORM PRIOR

The standard Uniform distribution is assumed as Non-informative prior for the parameter θ. The uniform prior for θ is defined as:

$$p(\theta) \propto 1, \quad 0 < \theta < \infty. \tag{3}$$

Using the likelihood (2) and the prior (3), the posterior distribution for the parameter θ takes the form:

$$p(\theta \mid \mathrm{x}) \propto p(\theta) L(\theta, x)$$

$$p(\theta \mid \mathrm{x}) = \frac{\left(\sum_{i=1}^{n} x_i^2\right)^{(3n-2)/2} \exp\left(-\sum_{i=1}^{n} x_i^2 / \theta\right)}{\Gamma\left((3n-2)/2\right) \theta^{(3n-2)/2+1}}, \quad 0 < \theta < \infty,$$

$$p(\theta \mid \mathrm{x}) = \frac{\beta_U{}^{\alpha_U}}{\Gamma \alpha_U} \frac{1}{\theta^{\alpha_U + 1}} e^{-\frac{\beta_U}{\theta}}, \qquad 0 < \theta < \infty \tag{4}$$

where $\alpha_U = \left(3n - 2/2\right)$ and $\beta_U = \sum_{i=1}^{n} x_i^2$ .are the parameters of the posterior distribution similar to the Inverted Gamma distribution.

4. THE JEFFREYS PRIOR

Since the parametric range for the Maxwell distribution is 0 to ∞, therefore, according to the Jeffreys rule of thumb, the Jeffreys prior becomes:

$$p_J\left(\theta\right) \propto \frac{1}{\theta}. \tag{5}$$

Using the likelihood (2) and the prior (5), the posterior distribution for the parameter θ takes the form:

$$p\left(\theta \mid \mathrm{x}\right) \propto \frac{1}{\theta^{\frac{3n}{2}+1}} \exp\left(-\sum_{i=1}^{n} x_i^2 / \theta\right)$$

or

$$p\left(\theta \mid \mathrm{x}\right) \propto \frac{1}{\theta^{\alpha_J+1}} \exp\left(-\beta_J / \theta\right), \qquad 0 < \theta < \infty \tag{6}$$

which is density kernel of the inverted gamma distribution with parameters α_J and β_J where $\alpha_J = \left(3n/2\right)$ and $\beta_J = \sum_{i=1}^{n} x_i^2$ are the parameters of the posterior distribution.

5. DATA SET USED FOR ANALYSIS

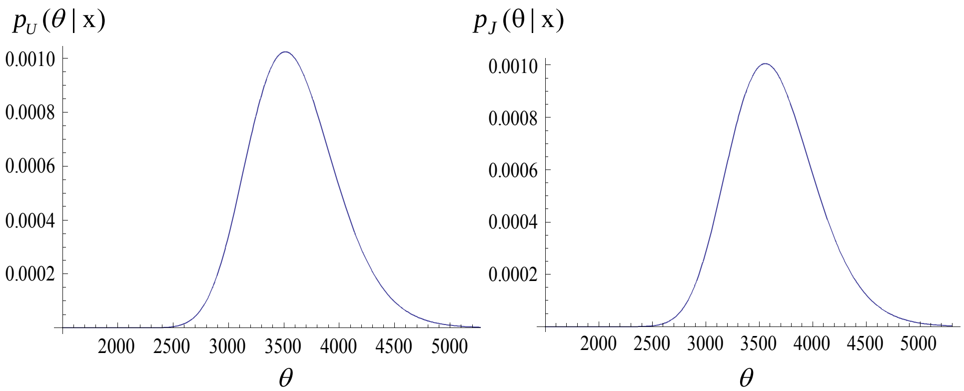
This section comprises the analysis of the posterior distribution of θ using Non-informative priors. For illustration, we take 55 observations related to the burning velocity of different chemical materials. As we know that the burning velocity is the velocity of a laminar flame under stated conditions of composition, temperature, and pressure. It decreases with increasing inhibitor concentration and can be determined by analyzing the pressure–time profiles in the spherical vessel and was checked by direct observation of flame propagation. The data related to the burning velocity (cm/sec) of different chemical materials is given below

68, 61, 64, 55, 51, 68, 44, 82, 60, 89, 61, 54, 166, 66, 50, 87, 48, 42, 58, 46,
67, 46, 46, 44, 48, 56, 47, 54, 47, 80, 38, 108, 46, 40, 44, 312, 41, 31, 40, 41,
40, 56, 45, 43, 46, 46, 46, 46, 52, 58, 82, 71, 48, 39, 41.

The source of the above explained data related to the burning velocity of different chemical materials for the year 2005 is the website (http://www.cheresources com/mists.pdf.).

Graphs of the Posterior Distributions using Non-informative Priors

By using above data set graphs of posterior distributions (4) and (6) using Non-informative priors i.e. Uniform and Jeffreys of parameter θ are provided as



The above graphs show that the posterior distributions using Uniform (UP) and Jeffreys' prior (JP) are positively skewed with miner difference.

Table 1: Properties of Posterior distribution using Real Data Set

Prior	**Mean**	**Mode**	**Variance**	**Skewness**	**Excess Kurtosis**
UP	3643.590062	3555.2606	166990.54723	0.45433259	3.394962
JP	3598.883435	3512.6826	160893.93796	0.45143040	3.386011

According to the above results the Jeffreys' prior is more preferable than uniform prior due to its better performance based on different statistical measures as stated in table 1.

6. BAYESIAN CREDIBLE AND HPD INTERVALS

The Bayesian credible intervals and highest posterior density intervals are derived below

6.1 Credible Intervals

If $p(\theta \mid x)$ is the posterior distribution given the sample, we may be interested in finding an interval (θ_1, θ_2) such that

$$P(\theta \in (\theta_1, \theta_2) \mid x) = \int_{\theta_1}^{\theta_2} p(\theta \mid x) d\theta = 1 - \alpha . \tag{7}$$

In Bayesian analysis, credible interval becomes the counterpart of the classical confidence interval, also credible interval may be unique for all models. The Bayesian credible interval, on the other hand, has a direct probability interpretation $P(\theta \in (a,b) \mid x) \geq 1 - \alpha$ and is completely determined from the current observed data **x** and the prior distribution.

The $(1-\alpha)100\%$ Bayesian credible intervals are obtained by using the posterior distribution of the parameter of interest. (See Ahmed, Abu-Taleb et al. 2007 and Saleem and Aslam 2009).

Credible Interval assuming Uniform Prior

$(1-\alpha)100\%$ Credible interval of posterior distribution of $(\theta \mid \mathrm{x})$ using Uniform prior is

$$\left[\frac{2\sum_{i=1}^{n} x_i^2}{\chi^2_{2(\frac{3n}{2}-1),1-\frac{\alpha}{2}}} \le \theta \le \frac{2\sum_{i=1}^{n} x_i^2}{\chi^2_{2(\frac{3n}{2}-1),\frac{\alpha}{2}}}\right].$$

Credible Interval assuming Jeffrey Prior

$(1-\alpha)100\%$ Credible interval for the posterior distribution of $(\theta \mid \mathrm{x})$ using Jeffreys Prior is

$$\left[\frac{2\sum_{i=1}^{n} x_i^2}{\chi^2_{2(\frac{3n}{2}),1-\frac{\alpha}{2}}} \le \theta \le \frac{2\sum_{i=1}^{n} x_i^2}{\chi^2_{2(\frac{3n}{2}),\frac{\alpha}{2}}}\right].$$

6.2 The Highest Posterior Density (HPD) Intervals:

The Highest Posterior Density (HPD) Interval may be declared as an advanced version of the Credible Interval. Actually, in case of a unimodal density, an additional condition can be imposed on a Credible Interval to make it unique by tilting it towards the most concentrated part of the posterior density. This unique Credible Interval is called an HPD Interval.

An interval (θ_1, θ_2) would be a $(1-\alpha)100\%$ HPD interval for θ if it satisfies the following two conditions simultaneously as given in Sinha (1998).

$$\int_{\theta_1}^{\theta_2} p(\theta \mid \mathrm{x})\, d\theta = 1-\alpha \text{ and} \tag{8}$$

$$p(\theta_1 \mid \mathrm{x}) = p(\theta_2 \mid \mathrm{x}) \tag{9}$$

HPD Interval assuming the Uniform Prior:

The posterior distribution of parameter θ assuming uniform prior (4), on simplification, two conditions which are given above are reduce to

$$\Gamma\left(\frac{3n}{2}-1, \frac{\sum_{i=1}^{n} x_i^2}{\theta_2}\right) - \Gamma\left(\frac{3n}{2}-1, \frac{\sum_{i=1}^{n} x_i^2}{\theta_1}\right) - (\alpha-1)\Gamma\left(\frac{3n}{2}-1\right) = 0$$

and

$$\left(\frac{3n}{2}\right)\ln\left(\frac{\theta_2}{\theta_1}\right)-\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)\left(\sum_{i=1}^{n}x_i^2\right)=0 .$$

Similarly for Jeffreys Prior, now solving these two equations simultaneously in order to get the $(1-\alpha)100\%$ HPD interval (θ_1,θ_2) for θ.

From Table 2, it is clear that the credible intervals and HPD intervals of the posterior distribution assuming Jeffreys' prior has less credible interval and HPD interval as compare to the posterior distribution using uniform prior so on the basis of this difference Jeffreys prior is more compatible as compare to Uniform prior also this is supported by the fact that the posterior distributions are positively skewed.

Table 2: 95% and 99% Bayesian Credible Intervals, HPD Intervals

Prior	95% CI	99% CI	95% HPD	99% HPD
UP	(2929.53,4528.37)	(2750.79,4878.34)	(2929.63, 4528.03)	(2750.84, 4878.03)
JP	(2897.46,4466.89)	(2721.64,4809.75)	(2896.52, 4466.10)	(2721.72, 4809.01)

7. BAYES FACTOR FOR DIFFERENT HYPOTHESES

In a Bayesian analysis to make the decision between two hypotheses is conceptually simpler than classical hypothesis testing. In the former case, the posterior probabilities of the hypotheses are calculated and the decisions about the hypotheses are made according to these posterior probabilities. An extensive literature on the subject of Bayesian hypothesis testing and comparison with classical method for hypothesis testing is available. Some of the many references are Lindley (1957, 1965); Zellner (1971); Zellner and Siow (1980); Shafer (1982); Berger and Delampady (1987); Berger and Pericchi (1993) and references cited therein. which are

$$P\left(H_1\right)=P(\theta\geq\theta_1)=\int_{\theta_1}^{\infty}p\left(\theta\,|\,\mathrm{x}\right)d\theta \tag{10}$$

where $p(\theta\,|\,\mathrm{x})$ is the posterior distribution of θ given that x

$$P(H_2)=1-P(H_1)$$

While Bayes factor is denoted by 'B' and the most general form of the Bayes factor can be described as follows.

$$B=\frac{Posterior\ odds\ ratio}{Prior\ odds\ ratio} \tag{11}$$

The Bayes factor can be interpreted as the 'odds for H_1 to H_2 that are given by the data'. While the Bayesian approach typically eschews arbitrary decision thresholds, Jeffreys (1961) gives the following typology for comparing H_1 vs. H_2.

Table 3: Posterior Probabilities and Bayes Factor for Different Hypotheses

Null Hypotheses	Alternative Hypotheses	Priors	Posterior Probability		Bayes Factor
H_1	H_2		$P(H_1)$	$P(H_2)$	B
$\theta \le 2700$	$\theta > 2700$	Uniform	0.0417473	0.9582527	0.0435661
		Jeffreys	0.0522460	0.9477540	0.0551261
$\theta \le 3500$	$\theta > 3500$	Uniform	0.3860966	0.6139034	0.6289208
		Jeffreys	0.4288498	0.5711502	0.7508529
$\theta \le 4500$	$\theta > 4500$	Uniform	0.9717702	0.0282298	34.423560
		Jeffreys	0.9784258	0.0215742	45.351661
$\theta \le 5000$	$\theta > 5000$	Uniform	0.9972645	0.0027355	364.56255
		Jeffreys	0.9981007	0.0018993	525.50977

Since from above table, 99% credible interval θ lies between 2750 and 4743 approximately. So it is reasonable to test that the hypotheses form $H_1 : \theta \le 2700$ verses $H_2 : \theta > 2700$ to $H_1 : \theta \le 5000$ verses. $H_2 : \theta > 5000$. The arbitrary decision thresholds for these hypotheses are based on Jeffreys (1961) typology for comparing model H_1 and H_2 mention above shows that: For $H_1 : \theta \le 2700$ verses $H_2 : \theta > 2700$, Bayes factor (B) lies between $10^{-\frac{1}{2}}$ and 1 using Non-informative priors i.e. (Uniform, Jeffreys) so according to the criterion defined by Jeffreys (1961) there is minimal evidence against the posterior distributions under null hypotheses (H_1).

Considering the next hypothesis $H_1 : \theta \le 3500$ verses $H_2 : \theta > 3500$, using Non-informative $10^{-\frac{1}{2}} \le B \le 1$, there is minimal evidence against the posterior distributions under H_1. For $H_1 : \theta \le 4500$ verses $H_2 : \theta > 4500$, and $H_1 : \theta \le 5000$ verses $H_2 : \theta > 5000$ H_1 has strongly evidence against H_2 specially for the Jeffreys prior.

8. BAYES ESTIMATION UNDER DIFFERENT LOSS FUNCTION:

A loss function should be an appropriate for the decision of problem under consideration. The selection of a loss function can be difficult and its choice is often made for reason of mathematical convenience, without any particular decision problem of current interest. This section presents the derivation of different loss functions for the posterior distributions derived under the Non-informative, i.e., Uniform, the Jeffreys. In order to estimate Byes estimates and posterior risk, four loss functions i.e. Square error loss function, Modified LINEX loss function, Entropy loss function and Square logarithmic loss function are used here.

8.1 Square Error Loss Function (SELF)

The square error loss function (SELF) was proposed by Legendre (1805) and Gauss (1810) to develop least squares theory. It is defined as

$$L_3 = L(\theta, a_3) = (\theta - a_3)^2, \tag{12}$$

The Bayes estimates and Posterior Risk under Square Error loss function are derived as:

$$a_3^* = E(\theta) = \frac{\beta}{\alpha - 1}, \ \rho(a_3^*) = E(\theta - a_3)^2 = \left[E(\theta^2) - (E(\theta))^2\right],$$

$$\rho(a_3^*) = \left(\beta^2 / (\alpha - 1)^2 (\alpha - 2)\right)$$

8.2 Modified LINEX Loss Function:

The LINEX loss function is an asymmetric loss function, which was introduced by Klebanov (1972) and used by Varian (1979) in the context of real life assessment. The linex loss function is almost similar to the square error loss function. The modified lines loss function is

$$L(\theta, a_4) = \left[\exp\left(\frac{a_4}{\theta} - 1\right) - \left(\frac{a_4}{\theta} - 1\right) - 1\right]. \tag{13}$$

The Bayes estimate of θ under modified LINEX loss is $a_4^* = \beta\left[1 - \exp\left(-\frac{1}{(\alpha + 1)}\right)\right]$.

By using the modified linex loss function the Bayes posterior risk is defined as $\rho(a_4^*) = e^{-1}E(\exp(a_4 / \theta)) - a_4(\alpha / \beta)$.

8.3 Entropy Loss Function (ELF)

Entropy Loss provides explicit estimators for the estimation of the natural parameter θ is the canonical form of the exponential family. The Calabria and Pulcini (1996) defined entropy loss function

$$L(\theta, a_5) = b\left(\frac{a_5}{\theta} - \log\frac{a_5}{\theta} - 1\right), \tag{14}$$

The Bayes estimate of the parameter θ is $a_5^* = \left(E\left(\theta^{-1}\right)\right)^{-1} = \frac{\beta}{\alpha}$.

Since the Bayes estimate under the entropy loss, is equal to the posterior harmonic mean. The Bayes posterior risk under the entropy loss function is provided as

$$\rho(a_5^*) = E(\log\theta) - \log(\alpha / \beta), \quad \theta \geq 1$$

which is the Bayes posterior risk but $E(\log\theta)$ is evaluated numerically through the computer package Mathematica and α and β are the posterior parameters.

8.4 Square Logarithmic Loss Function (SLLF)

The square logarithmic loss is defined as

$$L(\theta, a_6) = (\log a_6 - \log\theta)^2. \tag{15}$$

The Bayes estimate of θ relative to the square logarithmic loss is obtained as $a_6^* = \exp(E(\log\theta))$. The Bayes posterior risk under the Square Logarithmic Loss function is $\rho(a_6^*) = E(\log\theta)^2 - \{E(\log\theta)\}^2$. Here both terms $E(\log\theta)$ and $E(\log\theta)^2$ are numerically integrated by using computer software Mathematica.

9. BAYES ESTIMATION AND POSTERIOR RISK USING REAL LIFE DATA

By using different Loss functions i.e. Square Error loss function, Modified LINEX loss function, Entropy loss function and Square Logarithmic loss function, the Bayes estimates and Posterior Risk of the posterior distribution through Non-informative priors i.e. Uniform and Jeffreys priors are as follow where posterior risk are in parentheses.

Table 4: Bayes estimates and Posterior Risk under different Loss Function

Prior	SELF	MLLF	ELF	SLLF
UP	3643.59006 (166990.54771)	3533.80037 (0.0060361920)	3598.88344 (16.382905353)	3576.91686 (0.0123455222)
JP	3598.88344 (160893.93772)	3491.73232 (0.0059641911)	3555.26061 (16.358440141)	3621.07576 (0.0121949711)

*where SELF (Square Error loss function), MLLF (Modified LINEX loss function), ELF (Entropy loss function) Square Logarithmic loss function.

On comparing the Bayes posterior risk of different loss functions, we observe that the loss function MLLF (Modified LINEX loss function) has less Bayes posterior risk than other loss function. According to the decision rule of less Bayes posterior risk we conclude that MLLF (Modified LINEX loss function) is more preferable loss function.

10. SIMULATION STUDY

Simulation is a tractable practice; we can utilize it to examine the performance of a different field of study. In simulation one generates a sample of random data in such a way that mimics a real problem and recapitulates that sample in the similar approach. It is one of the most widely used quantitative schemes because it is so elastic and can yield so many constructive results. There are different method such as Monto Carlo simulation and Boot strap to simulate the data. Here simulation criteria (see Krishna 2009) is used, in order to get the Bayes estimates and the Bayes posterior risk under different loss functions along with Non-informative priors i.e. Uniform, Jeffreys.

It is clear from appendix Table 5-8, the comparison of Bayes posterior risk under different loss function (Square Error, Modified LINEX, Entropy and Square logarithmic) using Non-informative (Uniform and Jeffreys) priors has been made through which we conclude that with in each loss function Non-informative Jeffreys prior provides less Bayes posterior risk so it is more suitable for the parameter of the Maxwell distribution and between loss function Modified LINEX loss function is more preferable than all other loss function because under this loss function Bayes posterior risk is small for each and every value of parameter θ.

11. CONCLUSION AND SUGGESTIONS

We consider the Bayesian analysis of the Maxwell life time model under Non-informative prior. After analysis we conclude that the Non-informative priors the Jeffreys is suitable for the unknown parameter θ of the Maxwell distribution and preferable over the uniform prior because of having less posterior variance along with less Skewness combined with less kurtosis. On the comparison of different loss functions i.e. Square Error loss function, Modified LINEX loss function, Entropy loss function, Square logarithmic loss function with respect to posterior distributions under Non-informative priors.

The choice of loss function as concerned, one can easily observed based on evidence (different properties as discussed above) that modified LINEX loss function has smaller posterior risk. One thing is common as we increase sample size posterior risk comes down. Also note that we cannot compared results of complete data with censored data because in censored data we are using less information than the complete data set. In future, this work can be extended using informative priors and considering location parameter. The study is useful for the researchers and practitioners also for scientist in physics and chemistry, where Maxwell distribution is extensively sued.

REFERENCES

1. Abu-Taleb, A.A., Smadi, M.M. and Alawneh, J.A. (2007). Bayes Estimation of the Lifetime Parameters for the Exponential Distribution. J. Math. and Statist., 3(3), 106-108. ISSN 1549-3644 © 2007 Science Publications.
2. Bekker, A. and Roux, J.J. (2005). Reliability characteristics of the Maxwell distribution: A Bayes estimation study. *Comm. Statist. Theo. and Meth.*, 34(11), 2169-78
3. Berger, J.O. and Delampady, M. (1987). Testing precise Hypotheses. *Statistical Science*, 2, 317-352.
4. Berger, J.O. and Pericchi, L.R. (1993). The Intrinsic Bayes Factor for Model Selection and Prediction. *Technical Report*, 93-43C, Department of Statistics, Purdue University, West Lafayette.
5. Calabria, R. and Pulcini G. (1996). Point Estimation Under Asymmetric Loss Functions for Left Truncated Exponential Samples, *Comm. Statist. Theo. and Meth.*, 25(3), 585-600.
6. Chaturvedi, A. and Rani, U. (1998). Classical and Bayesian reliability estimation of the generalized Maxwell failure distribution. *J. Statist. Res.*, 32, 113-120.
7. Gauss, C.F. (1810). *Methode des Moindres Carres Morrire sur la Combinations des Observations*. Translated by J. Bertrand (1955). Mallet-Bachelier, Paris.
8. Klebanov, L.B. (1972). Universal Loss Function and Unbiased Estimation. *Dokl. Akad. Nank SSSR Soviet-Maths. Dokl. T.*, 203, N6, 1249-1251.
9. Krishna, H and Malik, M. (2009): Reliability estimation in Maxwell distribution with Type-II censored data. Int. J. of Quality & Reliability Management, 26(2), 184-195.

10. Legendre, A. (1805). *Nouvelles Methods Pour La Determination des Orbites des Cometes Courcier*, Paris.
11. Lindley, D.V. (1957). A Statistical Paradox. *Biometrics*, 44, 187-192.
12. Lindley, D.V. (1965). *An Introduction to Probability and Statistics form a Bayesian Viewpoint* Part 1 & 2, Cambridge University Press.
13. Saleem, M. and Aslam, M. (2009), On Bayesian analysis of the Rayleigh Survival Time assuming the Random censor Time. *Pak. J. Statist*., 25(2), 71-82.
14. Shafer, G. (1982), Lindley's Paradox. J. Amer. Statist. Assoc., 77, 325-351.
15. Tyagi, R.K. and Bhattacharya, S.K. (1989a). Bayes estimation of the Maxwell's velocity distribution function. *Stistica*, 29(4), 563-7.
16. Tyagi, R.K. and Bhattacharya, S.K. (1989b). A note on the MVU estimation of reliability for the Maxwell failure distribution. *Estadistica*, 41(137).
17. Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, Inc. New York.
18. Zellner, A. and Siow, A. (1980). Posterior Odds Ratios for Selected Regression Hypotheses, *Bayesian Statistics*, eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.M.F. Smith, Valencia, Spain: University Press, 585-603.

APPENDIX

Bayes Estimates with posterior risks under Different Loss Functions

Table 5: Bayes Estimates using UP and JP Priors under SELF

	UP			JP		
n	θ = 0.5	θ = 2	θ = 3	θ = 0.5	θ = 2	θ = 3
25	0.585411 (0.3651321)	2.916995 (3.4324650)	3.818662 (7.7354470)	0.5807501 (0.3250111)	2.812430 (3.3273980)	3.98615 (7.6651331)
50	0.551554 (0.3566970)	2.570916 (3.3939911)	3.656013 (7.6972050)	0.5498412 (0.3063490)	2.677520 (3.3150941)	3.62199 (7.6555041)
100	0.531931 (0.3226880)	2.361462 (3.3754621)	3.328069 (7.5169960)	0.5203300 (0.3042240)	2.229821 (3.2968690)	3.39378 (7.4577521)
200	0.525731 (0.2796570)	2.224965 (3.3677150)	3.317285 (7.4088231)	0.5166421 (0.2648650)	2.193470 (3.2753460)	3.22184 (7.3582290)
300	0.522297 (0.2502350)	2.118107 (3.3181831)	3.149792 (7.3893111)	0.5109303 (0.2357480)	2.112872 (3.2678221)	3.15372 (7.2712131)
500	0.507581 (0.2348131)	2.098865 (3.2181760)	3.101735 (7.3486841)	0.5052614 (0.2249790)	2.100531 (3.2263531)	3.09374 (7.2710160)

Table 6: Bayes Estimates using UP and JP Prior under MLLF

	UP			JP		
n	θ = 0.5	θ = 2	θ = 3	θ = 0.5	θ = 2	θ = 3
25	0.599493 (0.0132156)	2.710436 (0.0132156)	3.709585 (0.0132156)	0.835845 (0.0128753)	2.754276 (0.0128753)	3.814532 (0.0128753)
50	0.570168 (0.0066374)	2.622865 (0.0066374)	3.648372 (0.0066374)	0.560643 (0.0065502)	2.678264 (0.0065502)	3.560611 (0.0065502)
100	0.560988 (0.0033259)	2.425247 (0.0033259)	3.514312 (0.0033259)	0.541498 (0.0033039)	2.477036 (0.0033039)	3.347467 (0.0033039)
200	0.531286 (0.0016648)	2.243753 (0.0016648)	3.298186 (0.0016648)	0.533489 (0.0016593)	2.365087 (0.0016593)	3.228097 (0.0016593)
300	0.515806 (0.0011103)	2.120953 (0.0011103)	3.145297 (0.0011103)	0.513971 (0.0011078)	2.172904 (0.0011078)	3.136983 (0.0011078)
500	0.508533 (0.0006664)	2.051781 (0.0006664)	3.043019 (0.0006664)	0.505902 (0.0006655)	2.070151 (0.0006655)	3.023004 (0.0006655)

Table 7: Bayes Estimates using UP and JP Prior under ELF

	UP		JP	
n	θ = 2	θ = 3	θ = 2	θ = 3
25	2.678348 (1.3440732)	3.887486 (1.9340221)	2.731244 (1.3371321)	3.773743 (1.8873515)
50	2.445893 (1.1429511)	3.576926 (1.7862531)	2.569332 (1.1294802)	3.582891 (1.7285904)
100	2.362791 (0.9963072)	3.407866 (1.7324313)	2.252169 (0.9810290)	3.283017 (1.7018232)
200	2.186552 (0.9601375)	3.215507 (1.6462634)	2.164989 (0.9622241)	3.208927 (1.6394825)
300	2,100895 (0.9530471)	3.139385 (1.4421172)	2.110032 (0.9466091)	3.112841 (1.4052911)
500	2.008216 (0.7963182)	3.046711 (1.3752735)	2.048951 (0.7001782)	3.010771 (1.3446524)

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 37-42

SUSTAINABLE DEVELOPMENT AND AGRICULTURE SECTOR A CASE STUDY OF SINDH

Faiz M. Shaikh[1], **Anwar Ali Shah**[2] and **Muhammad Bachal Jamali**[3]

[1] Deptt: Agri: Economics, SZABAC-Dokri, Pakistan.
Email: faizmuhammed_2000@yahoo.com

[2] Faculty of Commerce and Business Administration,
University of Sindh, Jamshoro, Pakistan.

[3] Department of Commerce, Shah Abdul Latif University,
Khairpur, Pakistan.

ABSTRACT

This research investigates the Sustainable Development and Agriculture Sector A Case Study of Sindh. Data were collected from 900 respondents from nine districts by using simple random technique, A Structural questionnaire was the basic tool for the measurement the sustainability in agriculture sector. It was revealed that diverting a sizeable area from the existing cropping sequence to other crops and enterprises to meet the ever-increasing demand for food, fibre, fodder, fuel while taking care of soil health and agro-ecosystem. The cost benefit analysis shows that they promise good returns to the farmers, though the returns on maize are not so promising. Natural conditions are particularly suitable to the districts of Nawabshah and Halla to the cultivation of maize. Similarly cotton may be a natural choice in Ghotki and Sukkur to reduce the cropped area under rice in the districts.

KEY WORDS

Sustainable; Development; Agriculture.

INTRODUCTION

Achieving self-sufficiency in food grains and food security for the people has underlined Pakistan's Agricultural development programs. The seed-fertilizer revolution that began in the late sixties, and supported by the appropriate agricultural policies, led to a cropping system that helped the country in attaining the said goal of self-sufficiency in food grain and food security. Inspite of structural shift towards industrialization, Agriculture is still the largest sector of Pakistan's Economy, with deep impact on Socio-economic setup. It is the source of livelihood of around 47.8 percent of the people employed in this sector and contributed 21.7 percent in the economy of Pakistan. However declining soil and land fertility, falling water table, growing dependence of chemical fertilizers and the excessive use of pesticides, all of which aggravate the problem of environmental degradation, have raised questions about the sustainability of the very technologies and the policies which from the basis of the present cropping system. Thus, there are emerges a growing dominance of the food crops over the non-food crops. Even with in the food crops, Rice and Wheat emerges the focal points of the cropping pattern. Different considerations, like meeting the food security, profit maximization and output maximization all found shelter under the rice-wheat cultivation. The increase in the area of Rice and Wheat reduced under the areas other crops like gram jowar, bajra, barley, maize, etc. trends in Kharif crops signify change in the area of various crops like rice, maize, bajra, jowar sugarcane and cotton. The rice and cotton shows an impressive increase in area while maize, jowar and bajra registers declines. Sugarcane too record a decline but the decline is not very significant. Likewise the trend in Rabi crops is in favour of wheat and oil seeds and is against the crops like barley and gram. On the whole, it is rice and cotton under Kharif crops and wheat and oil seeds under the Rabi crops, which have exhibited a raising trend. Two important factors, namely the government policies and modern technology in agriculture in the form of HYY of seeds, chemical fertilizers, pesticides, machinery and equipments, etc. contributed a lot to the emergence of the dominating wheat rice combination in the cropping pattern in Sindh. Profit/output maximization at the micro level, considerations for food security at the macro level and the subsequent technological development, all worked in tandom to produce the wheat-rice combination.

Data Collection Methodology

Data were collected from 900 respondents from nine districts by using simple random technique, A Structural questionnaire was the basic tool for the measurement the sustainability in agriculture sector.

TOWARDS SUSTAINABLE AGRICULTURE

In the light of our discussion, at least, two types of measures related to sustainable agriculture may be suggested as damage control, exercise consists in incorporation certain short term measure within the existing cropping system so as to make it ecologically sustainable. The second exercise focuses on bringing about certain long-term changes in the present agricultural practices, including a shift from the chemical technology to a more eco-friendly biotechnology.

SHORT TERM MEASURE

The emphasis, here, is to incorporate certain short term measures within the present cropping system so as to make it ecologically sustainable or at least to check the damages it is causing to the natural resources. The following measures can be suggested to induce sustainability in the state agriculture.

1. Diversification of crops

The existing cropping system has led to the marginalization of other crops and the main trend that appears from such practices is the crop specialization than to the crop diversification. The emphasis here is to restore the area earlier held by crops other than the rice and wheat. Crop diversification, thus means diverting a sizeable area from the existing cropping sequence to other crops and enterprises to meet the ever-increasing demand for food, fiber, fodder, fuel while taking care of soil health and agro-ecosystem

2. Agro-forestry

Trough agro-forestry, the mix of agricultural and forestry systems. We make use of number of trees and bush which enhance the nutrient values of cretin crops and plants. The bushes and trees with nitrogen fixing capacity have beneficial effects on plants growing with or after them. Trees also improve the microclimate by acting as wind-breaks .by improving the water holding capacity of the soil, and by acting as shade trees for the livestock-so focusing the deposition of manure. Dlal (1999), while discussing the agricultural trends in Rural Sindh, has suggested new techniques for planting poplar tree to supplement the income for agriculture without adversely affecting the benefits form agricultural crops. The author has recommended and justified the cultivation of sugarcane also with the poplar plantation, which will contribute immensely in creating more income to the farmers and generating more employment and healthy environment for the society.

Table-1: Agriculture Growth Percent

Year	Agriculture	Major Crops	Minor Crops
2002-03	**4.1**	**6.8**	**1.9**
2003-04	**2.4**	**1.7**	**3.9**
2004-05	**6.5**	**17.5**	**1.5**
2005-06	**6.3**	**-3.9**	**0.4**
2006-07	**4.4**	**7.7**	**-1.5**
2007-08	**1.1**	**-6.4**	**10.9**
2008-09 (P)	**4.7**	**7.7**	**3.6**

Source: Federal Bureau of Statistics-2009

Table-2: Production of Major Crops (000 Tons)

Year	Cotton	Sugar cane	Rice	Maize	Wheat
2003-04	10048 (-1.6)	53419 (2.6)	4848 (8.3)	1897 (9.2)	19500 (1.6)
2004-05	14265 (42.0)	47244 (-11.6)	5025 (3.6)	2797 (47.4)	21612 (10.8)
2005-06	13019 (-8.7)	44666 (-5.5)	5547 (10.4)	3110 (11.2)	21277 (-1.6)
2006-07	12856 (-1.2)	54742 (22.6)	5438 (-2.0)	3088 (-0.7)	23295 (9.5)
2007-08	11655 (-9.3)	63920 (16.8)	5563 (2.3)	3605 (16.7)	20959 (-10.0)
2008-09 (P)	11819 (1.4)	50045 (-21.7)	6952 (24.9)	4036 (11.9)	23421 (11.7)

Economic Survey of Pakistan-2009-10

3. Reeducation of Area under Wheat and Rice

It has been suggested that at least 20 percent of the area presently under wheat and paddy need to be shifted to some other crops {see e.g. johl committee Report (1986) and Choudhry (1998). Of course, the alternative crops to wheat and rice must be less intensive in respect of the use of modern impact including water NAD at the same time be economically viable.

Alternative Crops to Wheat

The main complaint against wheat has been the tremendous expansion of area under its cultivation. The area under wheat, which was 31 percent in 1971-67 increased t o 41.8 percent in 2001-2 .The main impact of the expansion in area under wheat has been on the cultivation of gram. Area under gram cultivation, which was 20 percent in 2000-02 has been reduced to mere 9.96 percent in 2004 .The ecological had been reduced to mere 10.96 percent in 2004. The ecological significance of gram cultivation follows from the fact that it increases the level of nitrogen in the soil, which is facing depletion due to wheat cultivation. So, gram may be suggested as an effective alternative to what both on economic and ecological grounds. Especially in districts of Jacobabad, Shikarpur, Sukkur, Ghotki and Khairpur where climate conditions are conductive to its growth. Similarly, cultivation of oilseeds in the districts of Thatta and Badin. may be suggested as the alternative to wheat cultivation on economic and ecological grounds

Alternative crops to Rice

The main complaint against the cultivation of rice that it has adversely affected the ground water table particularly in Zones I and II and parts of Zone III. Here tow crops, namely maize and jowar may be suggested as alternative to rice. Both these crops require comparatively suggested as alternative to rice. Both these crops require comparatively less water than rice. Moreover a cost benefit analysis shows that they promise good returns to the farmers, though the returns on maize are not so promising. Natural conditions are particularly suitable to the districts of Nawabshah and Halla to the cultivation of maize. Similarly cotton may be a natural choice in Ghotki and Sukkur to reduce the cropped area under rice in the districts.

Damage Control Measures

As has already been brought out, tow most serious ecological impacts of the present cropping system are its adverse impact on the ground water table and soil nutrition in the state, Accordingly we suggest below toe sets of measure one to deal with the adverse impact of the copping system on the ground water table and other for dealing of the cropping system on the ground weather table and other for dealing with its adverse impact of the natural healthy of soil in the Nausheroferoze, and Nawabshah districts.

a) Measure of Deal with the Groundwater problem.

Make use of Rainwater

It is suggested that Thar State should make use of rainwater, which at present is being allowed to go water via a network of drains. This can be down by setting up recharging station all along the length of drains where water can be round the clock in a full proof manner before injecitng it into the specially designed bore holes. It is also cautioned that if into the specially designed bore holes. It is also cautioned that if the present decline in the ground water is not arrested brackish water form adjoining areas can intrude into the set water zones. It may also be suggested that the bunds of paddy filed should be raised so as the store the maximums amount of rainfall to the extent the rainfall storage does not decrease the crop yield .It can further be argued that seepage though canal system can be effective source of water replenishment of retable reserves but this system should not be applied in areas which have poor quality ground water otherwise this seepage will be a net loss of water.

Reeducation of Subsidies on Electricity and Installation of Pump set

There is a need of reviewing the policy on subsides specific relating to electricity and installation of the pump sets and tube wells has encouraged the farmers to over- irrigate their fields. A Punjab that the farmers on an average apply thirty-tow watering to paddy, well that the farmer on an average, apply thirty-two watering to paddy growing farmers have revealed above the recommended number of twenty-five watering. Second the flat tariff; irrespective of its pitch, induce the cultivation of heavily irrigation crops.

Reduction in Ground Water Draft

A reduction in ground water withdrawal will go a long way in checking the decline in water table in the Rural Sindh. A reduction in ground accomplished by making available to the farmers additional surface water for irrigation to substitute for ground water and its use for irrigation.

Water Conservation

Another option is to conserve or save water through adoption of more efficient irrigation management practices. This option assumes special significance in the cultivation of rice where farmers are known to use water to a level equivalent to the recommended one. There is an urgent need to educate farmers to improve their allocate efficiency in use of irrigation so as to promote more judicious and economic use of water especially in the cultivation of other crops option like sprinkler and drip irrigation need to be promoted.

b) Damage to soil health is mainly caused by the over use of chemical fertilizers. The supply of nutrients by chemical fertilizers is very much essential to achieve higher balance between different constituents (components) so as to promote their efficient use. Expert’s guidance to farmer in the use of chemical fertilizers assumes significant role here.

Livestock Measure and Composts

These sources are often chapter, more efficient than inorganic compounds and focus on recycling of nutrients. Livestock manures form cattle, and chickens are important as they positively affect soil integrated farms have more earthworms than those under conventional management do. Composting is a technique of long standing that combines that use of animal manures green material and household.

CONCLUSION

We have considered two alternative cropping systems with an eye primarily on an ecological sustainable organization of agricultural practices in Rural Sindh. One short term cropping system which seeks modification within the existing cropping system and the other, a long term cropping systems which is aim at bringing about significant changes in the goals and method/techniques employed under the present cropping system. These two alternatives cropping systems may well look at as the two phases of a single integrated programs to be adopted in a step-wise/phased manner.

REFERENCES

1. Baum II, C.L. (2002). A Dynamic Analysis of the Effect of Child Care Costs on the Work Decisions of Low-income Mothers with Infants. *Demography*, 39(1), 139-64.
2. Berger, M.C. and Black, D.A. (1991). Child Care Subsidies, Quality of Care, and the Labor Supply of Low Income, Single Mothers. *The Review of Economics and Statistics*, 70, 635-41.
3. Blau, D. and Robins, P. (1988). Child-care Costs and Family Labor Supply. *The Review of Economics and Statistics*, 70(3), 374-81.
4. Bormann, M.K, Quarm, D. and Gideonse, S. (1984). *Women in the workplace: Effects on families*. Norwood, New Jersey: Ablex Publishing Corporation.
5. Connelly, R. (1992). The Effect of Child Care Costs on Married Women’s Labor Force Participation. *The Review of Economics and Statistics*, 74(1), 83-90.
6. Doiron, D. and Kalb, G. (2005). Demands for Child Care and Household Labor Supply in Australia. *The Economic Record*, 81(254), 215-236.
7. Bal Gulshan Singh (1999). *How to sustain Soil Health the Turbine*. Chandigarh.
8. FAO (1989). *Sustainable Agricultural Production: Implications for international Agriculture Research*. Food and Agriculture Organization of the United Nations, Rome. ISBN: 92-510-2773-0.
9. Gips T. (1987). Breaking the pesticides Habit: Alternatives to Twelve Hazardous Pesticides Minnesota: International Services.
10. Singhal, D.C. (2001). Environmental Impact of Ground Water Development. *Journal of Resource Persons Aid in Need*, I(2), 12-24.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 43-50

SUSTAINABLE RURAL DEVELOPMENT THROUGH WOMEN PARTICIPATION IN SMEs BUSINESS GROWTH IN SINDH

Syed Maqsood Zia Ahmed[1] **Nazir Ahmed Gopang**[2] and **Faiz M. Shaikh**[3]
[1] Department of Statistics, Shah Abdul Latif University, Khairpur, Pakistan
[2] IBA, University of Sindh, Jamshoro, Pakistan.
[3] Deptt: Agri: Economics, SZABAC-Dokri, Pakistan.
Email: faizmuhammed_2000@yahoo.com

ABSTRACT

The present research paper is focused is focused on the sustainable development through women participation in SMEs business growth in Sindh. The main objective of current research is to examine the sustainable development through women participation in Small and medium enterprises in upper Sindh. The data were collected from 300 respondents from five Districts Dadu, Nawabshah, Shikarpur, Jacobabad and Kashmore district by using simple random technique. It was further revealed that the rural women is less confident and their husbands were always given them hard time once they are exposing themselves to outside the boundaries of the house. It was revealed the rural women is innovate designs of toppi (Sindhi Caps) as well as other SMEs products which are the only source of earning. They are also paid 60% less value of their products because of lack of marketing and other facilities. The biggest challenges which they were facing they were doing all business in house, lack of marketing facilities, Karo Kari criminal activities and they were deprived from the basic rights. This study contributes and explores the Rural Women challenges in SMEs business and how these critical unethical problems we can overcome like KARO KARI, and other various social issues growth.

INTRODUCTION

The present research highlight the sustainable growth of SMEs business opportunities in Pakistan. The implementations of different government schemes started from community development in 1950 has its limited impact on the rural masses in countryside. The green revolution in the north west regions has not impact on the rural livelihood in Sindh province. The developmental approaches in the dynamic development in Pakistan.

Several studies were conducted and critically analyzed the development process and listed various gaps like bureaucratic top down approach, grabbing the opportunities by rural elities, marginalization of poorer sections of society. (Dube, S.C. 1969, Desai A.R. 1969, Ratan V.W., 1984, Shastri, et al. 1988; Kurian 1989 etc. Women participation in Small and medium enterprises is the most neglected part of the development. The conceptual framework of the present study in the context of Small and medium Enterprises development in Sindh. Pakistan is recently facing lot of problems like unemployment, and slow growth in Agriculture we imported different agriculture related product from the world. Therefore, we have to look at sectors of the economy that have the potential to provide this employment, and simultaneously we have to ensure that the young are provided quality education and training for these more productive sectors of the economy. The greater growth potential lies in the modern high technology industries but it is also present in the labor-intensive industries of the traditional Small Enterprise sector and in the services that support it. "Young people in the future are more likely to end (therefore) should be working in organizations closer to the entrepreneurial mode" (Allan Gibbs). It is a worldwide phenomenon that Small Enterprises are an important part of a nation's economic and social structure. "Enterprise is the antithesis of command and control". On a global perspective Small Enterprises have acquired a significant stature in the economic development of a country. Globalization has put Small Enterprises directly in the limelight. Small Enterprises are increasingly a major force for national economic growth. The entrepreneurs who drive them are receiving serious attention from economists, planners, multilateral agencies and governments all over the world (Carter, Camille. 1999).

Human resource is one of the main contributing factors for economic growth and for social, political, and technological development. In the era of globalization, economic liberation and fast growing ICT, strengthening the national competitive advantage is the concern of the policy makers. Subsequently, various human resource development planning and strategies are formulated and action plan are outlined at various phases of development. Human resource development continued to be given priority in support the implementation of a productivity-driven growth, which required highly skilled, trainable and knowledge manpower. Emphasis continued to be given to increase accessibility to education at all levels in line with the democratization of the education policy. The high growth rate of the economy was achieved with the price stability and since 1995, with virtually full employment before the currency crisis erupted in the late 1997. The rapid expansion in manufacturing increased employment in the sector sharply during the Sixth Pakistan is the seventh largest population in the world and one of the dense populated countries in the world.

Developing Countries Policies for, Women as Entrepreneurship Business

Entrepreneur business Development in Turkey and Modalities of Intervention as was the practice of most countries; Turkey also practiced a state directed economy from its inception 1923. Hoping to achieve rapid industrialization it followed import substitution policy and relied on State Economic Enterprises, because it was the common belief that economic growth depended on heavy investment in large, capital-intensive industrial projects. Notwithstanding the imbalance of such a policy there was considerable growth of approximately 7% per annum. However true to the world pattern the growth declined in the 1970s and a new trend emerged in the 1980s. It was the trend based on free market economy and hence structural reforms were undertaken to give the economy a new shape.

2. WOMEN'S LABOR FORCE PARTICIPATION RATES IN PAKISTAN

Women are highly important contributors to the country's economic and social development. Over the years women participation in the economy has increased rapidly and they constitute almost half of the total population. Since 1990, women's participation in the SMEs business has increased enormously. Even though Pakistan is a newly industrializing country, its female labor force participation rate compares favorably with those of the industrialized countries of the Asia and Pacific region. Women's labor force participation rate has increased over the years, but is still significantly lower than that of men in early decades.

Table 1:
Participation Rates by Gender in SMEs Business 2001-2009

Gender	2001	2002	2003	2004	2005	2006	2007	2008	2009
Male	35	30	29	30	32	30	29	28	30
Female	20	35	31	33	35	33	33	35	36
Total	55	65	60	63	67	63	62	62	66

Survey-2009

Obviously, it can be seen that employment rates between the three ethnic groups are different: in 2001, 55 % of the total work force employed in SME business the share of women is 20% in sample area. In 2002, 65 % of the total work force employed in SME business the share of women is 35% in sample area, the women of Rural Sindh are so innovative that they can design the different kind of flowers without using the machines. In 2003, 6 % of the total work force involves in SME business and the share of women is 31%. In 2004, 63 % of the total work force involves in SME business and the share of women is 33%. In 2005, 67 % of the total work force involves in SME business and the share of women is 35%. In 2006, 63 % of the total work force involves in SME business and the share of women is 33%. In 2007, 62 % of the total work force involves in SME business and the share of women is 33%. In 2008, 62 % of the total work

3. DATA COLLECTION METHODOLOGY

The data were collected from 300 respondents from five Districts Dadu, Nawabshah, Shikarpur, Jacobabad and Kashmore district by using simple random technique. Structural questionnaire were design as a measurement tool.

Table 2: Summary Statistics

Variables n=300		Frequencies	Percent
Location	Jacobabad	100	33.33
	Kashmore	100	33.33
	Shikarpur	40	13.33
	Larkana	60	20
Age-Minimum-18	19-30	150	50
Maximum-50	31-50	150	50
Race	Sindhi	200	66.66
	Balouch	50	16.66
	Punjabi	30	10
	Urdu	20	6.66
Education		Male	Female
Primary	Jacobabad	17%	1%
	Kashmore	20%	2%
	Shikarpur	33%	17%
	Larkana	37%	20%
Secondary	Jacobabad	12%	0.5%
	Kashmore	11%	1%
	Shikarpur	22%	14%
	Larkana	28%	16%
University/Graduate	Jacobabad	6%	.33
	Kashmore	5%	.4
	Shikarpur	10%	11%
	Larkana	16%	13%
Health Condition		Satisfactory	Non-Satisfactory
	Jacobabad	12%	88%
	Kashmore	11	89%
	Shikarpur	40	60
	Larkana	55	45

Table-3: Housing Status

Housing		Own	Rental
	Jacobabad	90%	10%
	Kashmore	93%	7%
	Shikarpur	80%	20%
	Larkana	75%	25%

Survey-2009

Table 4: Working Status

Working Status	Government Job	SMEs Business	Unemployed
Jacobabad	30%	20%	50%
Kashmore	10%	30%	60%
Shikarpur	24%	10%	66%
Larkana	30%	20%	50%

Survey-2009

Table 5: Monthly Income from SMEs Business

Monthly Income		SMEs Business
	Jacobabad	Rs.20,000 / per family
	Kashmore	Rs.20,000 / per family
	Shikarpur	Rs.30,000 / per family
	Larkana	Rs.30,000 / per family

Survey-2009

Table 6: Probit Model for the Women Participate in SMEs Business

Variable	Coefficient	P>│t│	Marginal effect
- Intercept	**-.56677**	**0.78**	
Personal Characteristics			
Age	.513879	0.01	.344465
Years of Education	.00987765	0.909	.00098877
Status	-0.99876	0.878	.098866
Health	.4123467	0.0067	-0.4322
Birth of Origin	-.87654	0.177	-76543
Experience	.1988766	0.7654	-0.44556
Income	.98766	.66689	.097655
Marketing	-0.8765	-0.8876	-86544
Women Wages in SMEs Business	-0.877665	0.0012	-7654
Cost on SMEs products	-76554	0.2345	-765433
Log likelihood		**-865433**	
Pseudo		**0.7654**	

As the theory predicts, years of education and working experience are significantly positive indicating that higher education or having more experience raise the probability of participation. As a proxy to wages, having more education means greater possibility of getting better jobs, hence have higher wages. But the situation in SMEs business is entirely different in case f education and wages. Women are getting low wages and due to non availability of market facility they are paying low wages. The negative impact of husbands' income and having children at the age of 0-6 years old are also significant. High income of other household member, raises the reservation wage of mothers, thus lowers the probability of participation when the objective of working is to help family's financial need. The impact of childcare on mothers' labor force participation is significant but not as the theory predicted. The probability of participation in labor force is significantly higher for women who lived in the urban areas where jobs opportunities are greater compared to the rural.

4. FACTORS INFLUENCING THE INCREASE OF EMPLOYED WOMEN

The increase in the participation rate of women in SMEs business in Pakistan could to a certain extent be explained by the above analysis because due to factors that women has no access to market and other amentias facilities. The increase in the female labor force participation may be attributable to improving economic incentives in employment and policies favoring the employment of women. In addition, the combined effects of improved maternal and child health care, access to family planning services, increased years of schooling, leading to arise in the average age at marriage, have allowed women to take advantage of the increased employment opportunities

Despite their significant role of women in SMEs Business, they have been largely ignored in the government's programs until recently, and the effects of the current programs focusing on income-generating activities such as food processing and handicrafts remain to be seen, iii) Women's low earning can be attributed to lifetime choices between work and family formation (from the viewpoint of labor supply) and to employment discrimination (from the viewpoint of labor demand). Since women usually have a greater role than men in caring for the family, they may invest less in their own education and may work for shorter periods and in occupations that require fewer hours or less effort than men. This combined with interruptions in labor-force participation limits women's access to better jobs and promotions. Furthermore, employers, in turn, may invest less in nurturing women's skills through training or education because women are expected to drop out of the labor force while they are raising young children or, in many circumstances, to stop all work outside the home once they are married.

5. CONCLUSION

Much progress has been achieved in the past few decades in narrowing the gender gap in Pakistan. It can be shown in the developments in women's roles, both in absolute and relative terms, in the major socioeconomic aspects of the country's development: increasing rates of female labor force participation, gains in productive activities of women and their strengthened economic standing and their increased participation in education. Pakistan's impressive economic growth has been accompanied by the greater

Participation of women in the formal workforce and in a range of other activities. There are equal opportunities for employment for both men and women workers in Pakistan. Based on the Labor Force Survey, in the first quarter of year 2002, women a made up 35.5 per cent of the labor force. Policy statements in the Government's also provide opportunities for women in SMEs business and economic participation as well as participation in education and training. Viewed from this perspective, women as active actors, in both the private and public spheres-should be trained with their male counterparts focusing not only on their domestic role but also on their productive role.

REFERENCES

1. Anderson, P.M. and Levine, P.B. (1999). *Child Care and Mother's Employment Decisions*. Working Paper for National Bureau of Economic Research http://www.nber.org/papers/w7058.
2. Averett, S.L., Peter, H.E. and Waldman, D.M. (1997). Tax Credits, Labor Supply and Child Care. *The Review of Economics and Statistics*, 79(1), 125-36.
3. Baum II, C.L. (2002). A Dynamic Analysis of the Effect of Child Care Costs on the Work Decisions of Low-income Mothers with Infants. *Demography*. 39(1), 139-64
4. Berger, M.C. and Black, D.A. (1991). Child Care Subsidies, Quality of Care, and the Labor Supply of Low Income, Single Mothers. *The Review of Economics and Statistics*, 70, 635-41.
5. Blau, D. and Robins, P. (1988). Child-care Costs and Family Labor Supply. *The Review of Economics and Statistics*, 70(3), 374-81.
6. Bormann, M.K, Quarm, D. and Gideonse, S. (1984). *Women in the workplace: Effects on families*. Norwood, New Jersey: Ablex Publishing Corporation.
7. Connelly, R. (1992). The Effect of Child Care Costs on Married Women's Labor Force Participation. *The Review of Economics and Statistics*, 74(1), 83-90.
8. Doiron, D. and Kalb, G. (2005). Demands for Child Care and Household Labor Supply in Australia. *The Economic Record*, 81(254), 215-236.
9. Han, Wenjui and Waldfogel, J. (2001). Child Care Costs and Women's Employment: A Comparison of Single and Married Mothers with Pre-School-Aged Children. Social *Science Quarterly*, 82(3), 552-68.
10. Heckman, J. (1974). Effects of Child Care Programs on Women's Work Effort. *Journal of Political Economy*, 82, s136-s163.
11. Hofferth, S.L. and Wissoker, D.A. (1991). Price and Quality in Child Care Choice. *Journal of Human Resources*, 27(1), 70-111.
12. Kaufman, E.B. (1994). *The economics of labor markets*. 4th Edition. Georgia State Universities: The Dryden Press.
13. Kimmel, J. (1993). Child Care Costs As a Barrier to Employment for Single and Married Mothers. *The Review of Economics and Statistics*, 287-299.
14. Kimmel, J. and Powell, L.M. (2006). Nonstandard Work and Child Care Choices of Married Mothers. *Eastern Economic Journal*, 32(3), 397-419.
15. Leibowitz, A; Klerman, J.A. and Waite, L.J. (1992). Employment of New Mothers and Child Care Choice: Differences by Children's Age. *Journal of Human Resources*, 27(1) 112-133.
16. Leibowitz, A; Waite L.J. and Wittsberger, C. (1988). Child care for Preschoolers: Differences by Child's Age. *Demography*, 205-220.
17. Lokshin, M. and Fong, M. (2006). Women's labor Force Participation and Child Care in Romania. *Journal of Development Studies*, 42(1), 90-109.
18. Michalopoulos, C.; Robins, P. and Garfinkel, I. (1992). A Structural Model of Labor Supply and Child Care Demand. *Journal of Human Resources*, 27, 166-203.
19. Oishi, A.S. (2001). The Effect of Childcare Costs on Mothers' Labor Force Participation. *Paper for the Distribution of Income Project 1999-2001*, s51-s65.
20. Powell, L.M. (1998). Part-time versus Full-time Work and Child Care Costs: Evidence for Married Mothers. *Applied Economics*, 30(4), 503-11.

21. Ribar, D. (1992). Child Care and the Labor Supply of Married Women. *Journal of Human Resources*, 27(1), 134-65.
22. Robins, P.K. and R.G. Spiegelman. (1978). An Econometric Model of the Demand for Child Care. *Economic Inquiry*, 16, 83-94.
23. Stolzenberg, R.M. and Waite, L.J. (1988). Local Labor Market, Children and Labor Force Participation of Wives. *Demography*, 21(2), 157-68.
24. Stromquist, P.N. and Monkmen, K. (1998). *Women in the third world: An encyclopedia of the contemporary issues*. New York & London: Garland Publishing Incorporation.
25. Viitanen, T.K. (2005). Cost of Childcare and Female Employment in the UK. *Labor*, 19 (Special Issue), 149-170.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 51-60

PROPORTIONAL ODDS MODELS WITH L2 PENALIZATION

Faisal M. Zahid[1] and **Shahla Ramzan**[2]

[1] Department of Statistics, Ludwig-Maximilians-University Munich, Germany. Email: faisal-maqbool.zahid@stat.uni-muenchen.de

[2] Department of Statistics, Govt. College University Faisalabad, Pakistan. Email: shahla_ramzan@yahoo.com

ABSTRACT

In regression analyses, proportional odds models are commonly used when response variable has ordered categories. In case of proportional odds models, the usual likelihood approach becomes unstable with multicollinearity. The likelihood estimates do not exist when number of observations is less than the number of parameters to be estimated. Penalized likelihood with L2 penalty is used to address these issues. For categorical predictors, the considered penalized likelihood function penalizes the parameters associated with ordinal predictors differently than nominal predictors by preserving the order of categories. Prediction Error, mean squared error of parameter estimates and fitted probabilities for penalized estimates are compared with MLE in a simulation study and an application.

KEYWORDS

Likelihood estimation; Logistic regression; Penalization; Proportional odds model; Ridge regression.

1. INTRODUCTION

In regression analysis, maximum likelihood estimation is a common approach to compute the parameter estimates in categorical response models. But this approach may fail with large number of parameters relative to the sample size. For $p > n$, the maximum likelihood estimates do not exist at all. The same problem arises for the data with high correlation among the covariates. An alternative to the usual likelihood approach is to use penalized likelihood function. Penalization techniques combine log-likelihood function with a penalty term which measures the smoothness of the fit. In recent years several penalization techniques with different types of penalties have been proposed. The main objective of using penalized log-likelihood is to obtain unique estimates of the parameters, better prediction with a good compromise between bias and variance. Ridge regression is the most familiar penalization approach in the literature. In the context of linear models much literature is available for ridge regression. Schaefer et al. (1984) and Schaefer (1986) discussed the ridge penalty for logistic regression with binary response. An extension of ridge regression for GLM type models is considered by Nyquist (1991). LeCessie and Houwelingen (1992) discussed different ways to select the ridge penalty and also for computing prediction error in case of logistic ridge regression. In the literature univariate GLM's are more focused than the multivariate GLM. Zhu and Hastie

(2004) used penalized logistic regression with quadratic penalty as an alternative to the support vector machine (SVM) for microarray cancer diagnostic problems. Zahid and Tutz (2009) used ridge penalty to get penalized estimates for logistic regression with multi- category (unordered) responses, which are independent of choice of the reference category. Ridge regression shrinks the parameter estimates to zero but none of them is exactly zero. As a result we do not have a parsimonious model but a model with all predictors. Another penalization approach called Lasso was proposed by Tibshirani (1996). Lasso technique not only shrinks the parameter estimates to zero but also serves as subset selection by setting some of the estimates exactly to zero. The lasso approach for multinomial logit models was considered by Friedman et al. (2010). In many applications multi-category responses are ordered. According to our knowledge penalization has not been addressed for ordered category response models. For ordinal responses there were several models discussed in the literature (see McCullagh (1980), Ananth and Kleinbaum (1997) and Agresti (1999)). However the proportional odds model (also known as cumulative logit models) is the most popular among all other models for ordered category responses. Unlike multinomial logit models the proportional odds model (POM) has simple form in the sense that it has so-called global parameter estimates which are not category specific. But still in the case of large number of covariates maximum likelihood estimates may not exist. To resolve this problem we are using penalized log-likelihood with L2 penalty to compute the estimates in proportional odds models. If the response variable Y has k ordered categories as *1, ..., k*, the proportional odds model is given by

$$\log\left[\frac{\mathrm{P}(Y \le r|\mathrm{x})}{\mathrm{P}(Y > r|\mathrm{x})}\right] = \gamma_{0r} - \mathrm{x}^T\gamma, \quad r = 1,\dots, q = k-1, \tag{1}$$

The parameters vector $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)$ does not depend on the response categories. The ordered response Y can be viewed as a categorized version of an unobservable continuous latent variable Z as $Y = r \Leftrightarrow \gamma_{0,\,r-1} < Z \le \gamma_{0r}$ for $r = 1, \dots, k$, where $-\infty = \gamma_{00} < \gamma_{01} < \dots < \gamma_{0k} = \infty$ define the category boundaries on the unobservable latent continuum. The intercepts $\{\gamma_{0r}\}$ are different for each cumulative logit and must satisfy the stochastic ordering $\gamma_{01} < \dots < \gamma_{0q}$ to have positive probabilities.

In the following text penalized estimates are computed using the penalized log-likelihood based on L2 penalty for the proportional odds models given in (1). In Section 2 penalized likelihood is discussed with some computational issues. Empirical results of ordinal ridge regression are compared with usual ML estimates in Section 3. The ordinal ridge regression is fitted and compared with MLE using a real data in Section 4. Some final and concluding remarks are given in Section 5.

2. PENALIZED LIKELIHOOD

For the proportional odds model given in (1), let $\varphi_{ir}(\mathbf{x})$ denotes the cumulative probability for the occurrence of response levels up to and including rth level with a given covariate vector x_i given as

$$\varphi_{ir}(\mathbf{x}) = P(Y_i \le r|\mathbf{x}_i) = F(\eta_{ir}) \quad r = 1,\ldots,q = k-1,$$

where F is a strictly monotone distribution function. The model with homogeneous effects has the predictor

$$\eta_{ir} = \gamma_{0r} - x_i^T\gamma \tag{2}$$

In the context of multivariate generalized linear models, the proportional odds model can be given as

$$\boldsymbol{\pi}_i = h(\mathbf{X}_i\boldsymbol{\beta}) \quad \text{or} \quad g(\boldsymbol{\pi}_i) = \mathbf{X}_i\boldsymbol{\beta},$$

where $\pi_i^T = \left(\pi_{i1},\ldots,\pi_{iq}\right)$ with components $\pi_{ir} = \mathbf{P}(Y_i = r|\mathbf{x}_i)$ and $g = (g_1,\ldots, g_{k-1})$ is a logit link type function given by

$$g_r\left(\pi_i\right) = \log\left[\log\left(\frac{\varphi_{ir}}{1-\varphi_{ir}}\right) - \log\left(\frac{\varphi_{i,r-1}}{1-\varphi_{i,r-1}}\right)\right]. \tag{3}$$

For $p^* = p + q$, $\mathbf{X}_i = [\mathbf{I}_{q\times q}, \mathbf{I}_{q\times 1} \otimes x_i^T]$ is a $q \times p^*$ matrix and $\boldsymbol{\beta}^{\mathbf{T}} = (\gamma_0^T, \gamma^T) = (\gamma_{01}, \ldots, \gamma_{0q}, \gamma_1, \ldots, \gamma_p)$ is a $p^* \times 1$ vector. The complete design matrix of order $nq \times p^*$ is given as $\mathbf{X}^{\mathbf{T}} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]$. For further details see McCullagh and Nelder (1989) and Fahrmeir and Tutz (2001).

The predictor space may contain some categorical predictors with more than one parameters associated with it. Let we have K_j parameters associated with predictor X_j. So for a binary or continuous covariate we have $K_j = 1$ and if covariate is categorical then $K_j > 1$ depending upon the number of categories of the predictor X_j. The penalized log-likelihood with L2 penalty is given as

$$l_p(\beta) = l(\beta) - \frac{\lambda}{2}J(\beta) \tag{4}$$

where $l(\boldsymbol{\beta})$ is the usual log-likelihood function given by

$$l(\beta) = \sum_{i=1}^{n}\sum_{r=1}^{k} y_{ir}\log(\pi_{ir}), \tag{5}$$

and λ is a tuning parameter. The penalty term $J(\boldsymbol{\beta})$ can be given as

$$J(\boldsymbol{\beta}) = \boldsymbol{\beta}^T\mathbf{P}\,\boldsymbol{\beta}. \tag{6}$$

The definition of penalty matrix **P** depends on how we perform penalization with different types of predictors. In penalization, it is common practice to penalize the parameter estimates but with ordinal predictors rather than estimates, difference between the parameters estimates of adjacent categories should be penalized (see Gertheiss and Tutz (2009)). Penalizing such differences will cause a smoothed version of the parameter estimates by avoiding the large difference among the estimates associated with the dummies of the ordinal predictors. Let the predictor X_j is ordinal with K_j+1 categories and first category is treated as reference category. If $\boldsymbol{\gamma}_j$ denotes the parameter vector for K_j parameters/dummies associated with ordinal predictor X_j , the penalty term for penalizing the differences between the parameters of adjacent categories takes the form as

$$J(\gamma_j) = \sum_{l=2}^{K+1} (\gamma_{jl} - \gamma_{j-1,l})^2 = \gamma_j^T \Omega_j \gamma_j$$

With $\Omega_j = U_j^T U_j$, for a $K_j \times K_j$ matrix $\mathbf{U}_j$ given by

$$U_j = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & \ddots & & \vdots \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}$$

If the predictor is not ordinal, the penalty term for predictor X_j is given by

$$J(\gamma_j) = \sum_{l=2}^{K+1} \gamma_{jl}^2 = \gamma_j^T I_j \gamma_j$$

where $\mathbf{I}_j$ is a $K_j \times K_j$ identity matrix. For the predictor space with all types of predictors (i.e., binary/continuous, nominal and ordinal predictors), the penalty matrix $\mathbf{P}$ given in (6) is given by

$$\mathbf{P} = \mathrm{diag}(\mathbf{0}_{q \times q}, \mathbf{P}_1, \ldots, \mathbf{P}_p). \tag{7}$$

Here $\mathbf{0}_{q\times q}$ is a zero matrix with zeros for the category specific intercept terms which are not penalized. If the jth predictor is ordinal then $(K_j \times K_j)$−submatrix $\mathbf{P}_j$ assumes the value $\boldsymbol{\Omega}_j$ otherwise $\mathbf{P}_j = \mathbf{I}_j$.

The penalized log-likelihood function given in (4) can be written as

$$l_p(\beta) = \sum_{i=1}^{n} l_i(\beta) - \frac{\lambda}{2} \sum_{j=1}^{p} \beta_j^T P \beta_j$$

Score function $s_p(\boldsymbol{\beta})$ for the penalized log-likelihood is given by

$$\begin{aligned} s_p(\beta) &= \sum_{i=1}^{n} X_i^T D_i(\beta) \Sigma_i^{-1}(\beta)[y_i - h(\eta_i)] - \lambda P\beta \\ &= X^T D(\beta) \Sigma^{-1}(\beta)[y - h(\eta)] - \lambda P\beta \end{aligned} \tag{8}$$

where $D_i(\beta) = \partial h(\eta_i)/\partial \eta$ is derivative of $h(\boldsymbol{\eta})$ evaluated at $\boldsymbol{\eta}_i = \mathbf{X}_i \boldsymbol{\beta}$, $\boldsymbol{\Sigma}_i(\boldsymbol{\beta}) = \mathrm{cov}(\mathbf{y}_i)$ is the covariance matrix of ith observation of $\mathbf{y}$ given parameter vector $\boldsymbol{\beta}$ and $\mathbf{W}_i(\boldsymbol{\beta}) = \mathbf{D}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1} \mathbf{D}^T(\boldsymbol{\beta})$. $\mathbf{y}$ and $h(\boldsymbol{\eta})$ are given by $y^T = (y_1^T, \ldots, y_n^T)$ and $h(\boldsymbol{\eta})^T = (h(\boldsymbol{\eta}_1)^T, \ldots,$

$h(\boldsymbol{\eta}_n)^T$). The matrices have block diagonal form $\boldsymbol{\Sigma}(\boldsymbol{\beta}) = \mathrm{diag}(\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}))$, $\mathbf{D}(\boldsymbol{\beta}) = \mathrm{diag}(\mathbf{D}_i(\boldsymbol{\beta}))$ and $\mathbf{W}(\boldsymbol{\beta}) = \mathrm{diag}(\mathbf{W}_i(\boldsymbol{\beta}))$. By equating the score function to zero we obtain the estimation equations as

$$X^T D(\beta)\Sigma^{-1}(\beta)[y - h(\eta)] - \lambda P\beta = 0$$

where $\boldsymbol{\beta}$ is $p^* \times 1$ parameter vector for $p^* = q + \sum_{j=1}^{p} K_j$, and $\mathbf{P}$ is a $p^* \times p^*$ matrix given in (7). Fisher scoring iteration yields

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + \left(X^T W(\hat{\beta}^{(k)})X + \lambda P\right)^{-1} s_p(\hat{\beta}^{(k)}) .$$

If $\hat{\boldsymbol{\beta}}$ are penalized estimates for the true parameter $\boldsymbol{\beta}$, the covariance matrix can be approximated by

$$\mathrm{cov}(\hat{\beta}) = \left(X^T W(\hat{\beta})X + \lambda P\right)^{-1} \left(X^T W(\hat{\beta})X\right) \left(X^T W(\hat{\beta})X + \lambda P\right)^{-1}$$

Generalized cross-validation (GCV) criterion is used to find the optimal value of ridge penalty λ. In the generalized linear models (GLM) environment we are using likelihood-based criterion deviance for GCV instead of squared distances of $y_{i.}$ and $\boldsymbol{\pi}_i$. Deviance based generalized cross-validation is given by

$$\mathrm{GCV} = \frac{-2.\sum_{i=1}^{n}(l_\lambda(\hat{\pi}_i) - l_i(y_i))}{(1 - tr(H(\lambda))/n)^2}, \tag{9}$$

where the hat matrix $\mathbf{H}$ is given as

$$H = W^{T/2}X\left(X^T W(\hat{\beta})X + \lambda P\right)^{-1} X^T W^{1/2} .$$

3. SIMULATION STUDY

The effectiveness of ridge regression is discussed in this section using simulated data. For a sample of size n, predictor space contains continuous covariates (denoted by C), binary covariates (denoted by B), nominal covariates (denoted by N_{K+1}) and/or ordinal covariates (denoted by O_{K+1}) with K+1 categories. The continuous covariates are drawn from a p-dimensional centered multivariate normal distribution with covariance between two covariates X_j and X_k being $\rho^{|j-k|}$. Four values of ρ = 0.0, 0.3, 0.7 and 0.9 are used. To study the problems with existence of usual MLE, we consider proportional odds models with k = 3 response categories for each setting of predictor space given in Table 1. The true values used for the intercept terms are (-0.3, 0.8). The true values for global parameters are obtained as $(-1)^j \exp(-2(j-1)/20)$ for j = 1,..., $\sum_{j=1}^{p} K_j$. A multiplicative factor c_{snr} for $\boldsymbol{\beta}^T = (\gamma_0^T, \gamma^T)$ is chosen so that the signal-to-noise ratio is 1.

For each setting mentioned in Table 1 with different number and type of covariates, S = 50 samples of size n are used in the study. To compare the results of ridge estimates with likelihood estimates, we consider only those samples for which ML estimates exist. In order to obtain S = 50 samples, S' samples are ignored because ML estimates with their standard errors are not existing for these samples using *polr* function of library *MASS* in statistical language R 2.10.0. In Table 1, the columns with title S' showing the number of samples for which ML estimates did not exist, highlights the need of a penalization technique. The results of Table 1 show that value of S' is increasing with increasing number of predictors drawn from different distributions. For sample size n = 30, we cannot generate 50 samples for which MLE is existing in setting 2, and in setting 1 although estimates are presented but they have quite large standard errors (not shown here). Although ML estimates are stable for independent predictors or in case of moderate correlation among predictors with increasing sample size, they are deteriorated in case of high multicollinearity. The ridge regression may be the best choice to obtain stable estimates of parameters for all predictors in the model when usual MLE is not existing or deteriorated because of ill-conditioned predictor space. In Table 1, ML estimates are computed using *polr* function of statistical environment/language R 2.10.0. Ridge estimates are compared with likelihood estimates in terms of mean squared error (MSE) of $\hat{\boldsymbol{\beta}}$, deviance of the fitted probabilities and mean prediction error (MPE). In all settings we use samples of size n for the training data and then a sample of size n_{test} = 1000 is generated for comparing the prediction performance of ridge estimates with likelihood estimates.

The results of Table 1 show that ridge is performing better than MLE as is expected. In case of ten normally distributed covariates, ML estimates have very large standard errors for n = 30. In the simulation setting 2 considered with categorical predictors, there are twenty parameters to be estimated (other than intercept terms). Here the situation becomes more critical with MLE and we have to leave a large number of samples because of problems with existence of estimates. Even we could not get enough samples with n = 30 for which likelihood estimates with their standard errors exist. So for setting 2, with n = 30, only the results of ridge regression are given in Table 1. The ridge estimates do exist in all considered situations even when the likelihood estimates do not exist. For comparing ridge estimates with ML estimates, mean squared error is computed as $\frac{1}{S}\Sigma_S\left(\hat{\beta}_S^{method}-\beta^{true}\right)\left(\hat{\beta}_S^{method}-\beta^{true}\right)$ and the formula used to compute deviance of fit i.e., Dev($\hat{\pi}$) is given by $D=2\sum_{i=1}^{n}\sum_{j=1}^{k}y_{ij}\log\left(\frac{y_{ij}}{\hat{\pi}_{ij}}\right)$ with $y_{ij}\log\left(\frac{y_{ij}}{\hat{\pi}_{ij}}\right)=0$ for y_{ij} = 0. Mean prediction error based on 1000 test observations is computed as

$$\text{MPE}=\frac{1}{S}\Sigma_S D_S=\frac{1}{S}\Sigma_S 2.\left[\sum_{i=1}^{n}\sum_{j=1}^{k}\pi_{ijs}^{test}\log\left(\frac{\pi_{ijs}^{test}}{\hat{\pi}_{ijs}^{test}}\right)\right].$$

Table 1:
Comparison of ridge estimates and maximum likelihood estimates in terms of $MSE(\hat{\beta})$, deviance($\hat{\pi}$) and mean prediction error (MPE). Results of simulation study for setting 1 (C = 10), and setting 2 (C=B =5, N_3 =N_4 =O_3 =O_4 =1).

Predictors	ρ	N	MLE			Ridge			S'_{S^0}
			MSE $(\hat{\beta})$	dev $(\hat{\pi})$	MPE	MSE $(\hat{\beta})$	dev $(\hat{\pi})$	MPE	
Setting 1	0.0	30	180.1517	13.6873	812.7266	7.0702	14.3484	311.5942	581
	0.3	30	291.1177	12.0787	1798.7247	9.2962	13.3363	370.9956	500
	0.7	30	343.8420	13.6453	941.6172	11.4605	14.8170	292.5661	535
	0.9	30	421.3659	13.1603	883.5063	15.3219	14.9178	179.8458	366
	0.0	50	14.7393	21.3033	272.5190	4.4841	21.6354	177.3077	34
	0.3	50	14.2448	24.1852	257.8009	3.9573	24.6463	168.2815	191
	0.7	50	19.2019	22.5475	206.5454	5.9453	23.0144	122.8994	55
	0.9	50	43.6299	22.4499	349.7897	11.8912	23.4773	171.2196	28
	0.0	100	3.7835	47.4030	97.2800	2.3563	47.5027	82.8500	0
	0.3	100	2.7517	48.1384	96.9192	2.0136	48.2122	86.0813	0
	0.7	100	7.1912	46.4471	111.7852	4.2880	46.6742	89.7360	0
	0.9	100	26.2696	41.8972	74.2318	10.9573	44.0150	51.0959	0
Setting 2	0.0	30	−	−	−	43.7581	6.3059	843.4785	−
	0.3	30	−	−	−	66.1659	5.2765	833.6123	−
	0.7	30	−	−	−	83.3968	6.3192	618.4994	−
	0.9	30	−	−	−	80.9918	4.9503	529.5153	−
	0.0	50	1115.9632	21.1014	1579.6180	24.0194	22.9401	372.7722	3729
	0.3	50	606.7988	21.6298	977.6950	35.5278	22.5852	336.4792	3202
	0.7	50	115.9206	23.6653	786.0302	24.6058	25.3760	330.2098	11037
	0.9	50	276.7447	23.0867	980.9997	34.9713	25.3711	292.1317	13631
	0.0	100	48.8578	41.6153	265.6948	22.0897	42.4914	175.4264	121
	0.3	100	36.9882	43.8864	267.5244	19.2712	44.4978	184.9282	46
	0.7	100	47.7568	43.4548	281.4263	17.4481	44.6789	168.8737	24
	0.9	100	41.9465	47.0938	242.1356	17.8432	48.1940	151.0515	44

4. APPLICATION

In this section, for computing and comparing the ridge estimates with ML estimates, we are considering the housing data set from UCI repository (http://archive.ics.uci.edu/ml/datasets/Housing). The response variable (MEDV) is about the median values of the owner-occupied houses in suburbs of Boston. For the analysis purpose, response variable which is measured in \$1000's is divided into four price categories as lower , lower middle, upper middle and high price category using the thresholds as MEDV< 10, 10≤MEDV< 25, 25≤MEDV < 40 and MEDV≥40. The predictors used to predict the price range of a house are: per capita crime rate by town (CRIM); proportion of residential land zoned for lots over 25,000 sq.ft. (ZN); proportion of non-retail business acres per town (INDUS); Charles River dummy variable (CHAS=1 if tract bounds river; CHAS=0 otherwise); nitric oxides concentration (NOX)(parts per 10 million); average number of rooms per dwelling (RM); proportion of owner-occupied units built prior to 1940 (AGE); weighted distances to five Boston employment centres (DIS); index of accessibility to radial highways (RAD); full-value property-tax rate per \$10, 000 (TAX); pupil-teacher ratio by town (PTRATIO); $1000(\mathrm{Bk} - 0.63)^2$ where Bk is the proportion of blacks by town (B) and % lower status of the population (LSTAT). Although the *polr* function of R provides the likelihood estimates of parameters but fails to produce standard errors of these estimates. So the usual ML estimates and the corresponding standard errors are not computed with *polr* as in simulation study but ML estimates are computed with ridge function for $\lambda = 0$ (i.e., MLE). For computing the ridge estimates, ridge penalty is decided on the basis of deviance based generalized cross-validation (GCV). The parameter estimates with their standard errors are given in Table 2 for the complete data set. The results of ridge regression are based on the optimal value of ridge penalty $\lambda = 0.1$. In order to check the existence of MLE in case of small samples, different random samples of size $n = 30$ are drawn from the complete data set but likelihood estimates are not existing for any of the these samples. Similarly ML estimates are not existing for most of the random samples of size $n = 50$ drawn from total sample of size $n = 506$. However ridge estimates exist for all such small samples.

Table 2:
MLE and ridge estimates with corresponding standard errors for housing data set with n = 506.

	MLE (n = 506)		Ridge (n = 506)	
	Estimate	S.E.	Estimate	S.E.
CRIM	0.0816	0.0210	0.0804	0.0209
ZN	−0.0090	0.0075	−0.0093	0.0075
INDUS	−0.0344	0.0393	−0.0253	0.0384
CHAS	−0.7756	0.5078	−0.7710	0.4909
NOX	5.6532	2.6104	3.3769	1.5477
RM	−1.4677	0.2715	−1.4615	0.2678
AGE	−0.0028	0.0077	−0.0010	0.0075
DIS	0.4321	0.1243	0.3970	0.1186
RAD	−0.2029	0.0475	−0.1952	0.0465
TAX	0.0077	0.0025	0.0077	0.0025
PTRATIO	0.3756	0.0844	0.3501	0.0804
B	−0.0077	0.0022	−0.0077	0.0022
LSTAT	0.3082	0.0432	0.3119	0.0430

5. CONCLUDING REMARKS

Ridge regression provides stable estimates in logistic regression when maximum likelihood estimates are deteriorated because of multicollinearity in the design space. Also when sample size is small relative to the number of parameters, existing softwares for fitting proportional odds models may fail to provide the parameter estimates. The maximum likelihood estimates will not exist if the design matrix is not of full rank (with $p > n$). To address these issues, ridge regression is used in this paper for ordinal response models with a focus on proportional odds models. Since ridge penalty shrinks the parameters estimates to zero but does not perform variable selection, it is useful in case of limited number of predictors where the analysts are interested in fitting a model by keeping all the predictors in the model. For ordinal predictors, natural ordering between the successive categories is considered and differences between parameters estimates associated with the dummies of adjacent categories are penalized.

REFERENCES

1. Agresti, A. (1999). Modelling ordered categorical data: Recent advances and future challenges. *Statistics in Medicine*, 18, 2191-2207.
2. Ananth, C.V. and Kleinbaum, D.G. (1997). Regression models for ordinal responses: A review of methods and applications. *International Journal of Epidemiology*, 26(6), 1323-1333.

3. Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling based on Generalized Linear Models*. Second Ed. Springer-Verlag New York, Inc.
4. Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33(1).
5. Gertheiss, J. and Tutz, G. (2009). Penalized regression with ordinal predictors. *International Statistical Review*, 77, 345-365.
6. LeCessie, S. and Houwelingen, V. (1992). Ridge estimators in logistic regression. *Applied Statistics*, 41, 191-201.
7. McCullagh, P. (1980). Regression models for ordinal data. *J. Roy. Statist. Soc.*, B 42, 109-142.
8. McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Second Ed. Chapman & Hall, New York.
9. Nyquist, H. (1991). Restricted estimation of generalized linear models. *J. App. Statist.*, 40, 133-141.
10. Schaefer, R. (1986). Alternative estimators in logistic regression when the data are collinear. *J. Statist. Compu. and Simul.*, 25, 75-91.
11. Schaefer, R., Roi, L. and Wolfe, R. (1984). A ridge logistic estimator. *Commun. in Statist.: Theo. and Meth.*, 13, 99-113.
12. Tibshirani, R. (1996). Regression shrinkage and selection via lasso. *J. Roy. Statist. Soc.*, B 58, 267-288.
13. Zahid, F.M. and Tutz, G. (2009). *Ridge estimation for multinomial logit models with symmetric side constraints.* Technical Report No. 67. Institute of Statistics, Ludwig-Maximilians-University Munich, Germany.
14. Zhu, J. and Hastie, T. (2004). Classification of gene microarrays by penalized logistic regression. *Biostatistics*, 5, 427-443.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 61-70

DEVELOPED HIDDEN MARKOV MODEL IN DNA SEQUENCE

Naila Rozi[1] and **Nasir Uddin Khan**[2]

[1] Sir Syed University of Engineering & Technology, Karachi, Pakistan. Email: nrozi@yahoo.com

[2] University of Karachi, Karachi Pakistan. Email: nasir.prof@yahoo.com

ABSTRACT

We have developed a method to extract the HMM in DNA sequences. In this method, the Genetic Algorithm (GA) and Baum- Welch algorithm are used to obtain the best Hidden Markov Model (HMM) representations of the geometric patterns in DNA sequences. The GA is used to search the best network shapes and the initial parameters of the HMMs. Baum-Welch algorithm is used to optimize the HMM parameters for the given network shapes. Akaike Information Criterion (AIC), which gives a criterion for the balance of adaptation and complexity of a model, is applied in the HMM evaluation. We have applied the method to the extraction of the signal patterns in human promoters and 5' ends of yeast introns. As a result, we obtained HMM representations of characteristic features in these sequences. To validate the efficiency of the method, we have performed promoter recognition using obtained HMMs. Two entries including nine promoters are selected from GenBank 76.0.

KEYWORDS

DNA sequence; signal pattern extraction and recognition; Hidden Markov Model; Genetic Algorithm; Akaike Information Criterion

1. INTRODUCTION

The method to the extraction of the signal patterns in human promoters and 5` end of yeast introns. As a result, we obtained HMM representation of characteristic features in these sequences. To validate the efficiency of the method, we have performed promoter recognition using obtained HMMs. Two entries including nine promoter recognition using obtained HMMs. Two entries including nine promoters are selected from GenBank 76.0, and it is observed that the HMM can predicts eight promoters correctly. These results imply that the method is efficient to design preferable HMM networks, and provides reliable models for the recognition of the geometric patterns. DNA sequencing reactions the PCR reactions for replicating DNA (refer to the previous page DNA Denaturation, Annealing and replication. The reaction mix includes the template DNA, free nucleotides, an enzyme (usually a variant of Taq polymerase) band a ‘primer’ – a small piece of single-stranded DNA about 20-30 nt long that can hybridize to one strand of the template DNA.

The reaction is initiated by heating until the two strands of DNA separate, then the primer sticks to its intended location and DNA polymerase starts elongating the primer. If allowed to go to completion, a new strand of DNA would be the result. If we start with a billion identical pieces of template DNA, we 71 get a billion new copies of one of its strands.

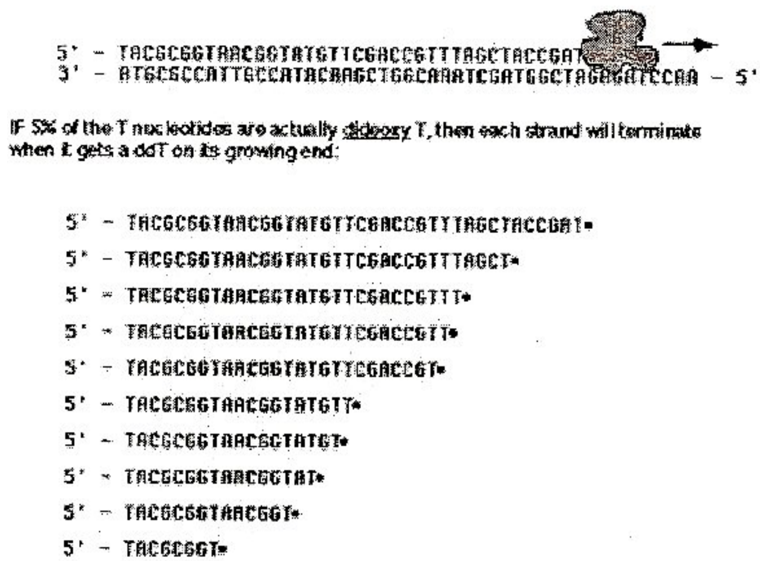
Dideoxynucleotides:

We run the reactions, however, in the presence of didcoxyribonucleotide. This is just like regular DNA, except it has no 3` hydroxyl group – once it's added to the end of a DNa strand, there's nom way to continue elongating it. Now they key to this is that MOST of the nucleotides are regular ones, and just a fraction of them are dideoxy nucleotides.

Replicating a DNA strand it the presence of dideoxy-T:

Most of the time when a "T" is required to make the new strand, the enzyme will get a good one and there's no problem. Most of the time after adding a T, the enzyme will go ahead and add more nucleotides. However 5% of the time, the enzyme will get a dideoxy.-T, and that strand can never again be elongated. It eventually breaks away from the enzyme, a dead end product.

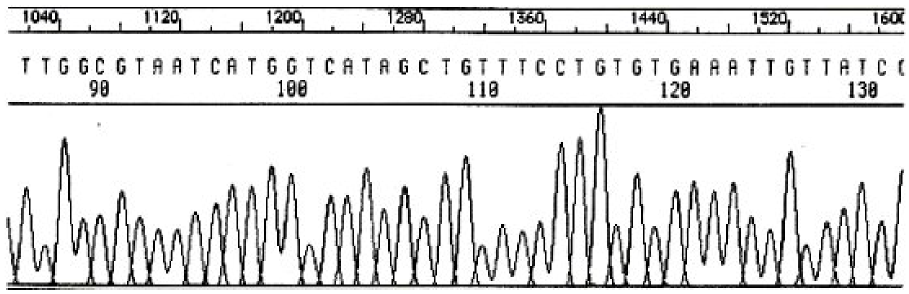
Sooner or later ALL of the copies will get terminated by a T, but each time the enzyme makes a new strand, the place it gets stopped will be random. In million of starts, there will be strands stopping at every possible T along the way. ALL of the strands we make started at one exact position. ALL of them end with a T. there arillions of them



…many millions at each possible T position b
Here's how we find out those fragment size.

Well, OK, it’s not so easy reading just C’s as you perhaps saw in the last figure. The spacing between the bands isn’t all that easy to figure out. Imagine, though, that we ran the reaction with *all four* of the didcoxy.

The sequencer also gives the operator a text file containing just the nucleotide sequence, without the color traces.



As you have seen, we can get the sequence of a fragment of DNA as long as 900 or so nucleotides. Great! But what about longer pieces? The human genome is 3 *bases* long, arranged on 23 pairs of chromosomes. Out sequencing machines reads just a drop in

We’d get (at left). The sequence of the DNA is rather obvious if you know the color codes… just read the colors from bottom top: TGG, GTCCA-(etc).

That’s exactly what we do to sequence DNA, then – we run DNA replication reactions in a test tube, but in the presence of trace amounts of all four of the dideoxy terminator nucleotides. Electrophoresis is used to separate the resulting fragments by size and we can ‘read’ the sequence from it, as the colors march past in order.

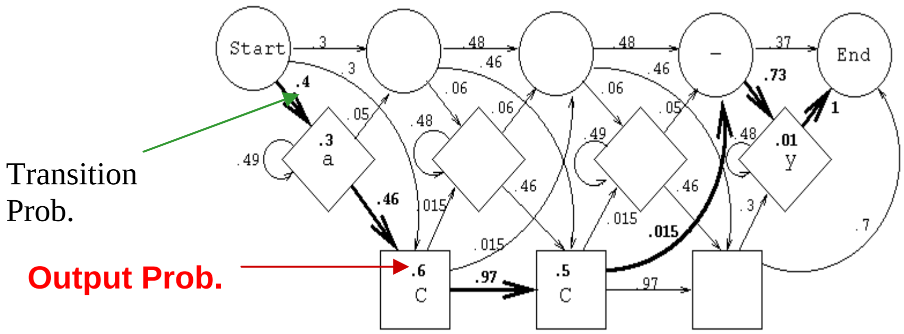
In a large-scale sequencing lab, we use a machine to run the electrophoresis step and to monitor the different colors as they come out. Since about 2001, these machines – not surprisingly called automated DNA sequencers – have used ‘capillary electrophoresis’, where the fragments are piped through a tiny glass-fiber capillary during the electrophoresis step, and they come out the far end in size-order. There’s an ultraviolet laser built into the machine that shoots through the liquid emerging from the end of the capillaries, checking for pulses of fluorescent color to emerge. There might be as many as 96 samples moving through as many capillaries (‘lanes’) in the most common type of sequencer. At left is a screen shot of a real fragment of sequencing gel (this one from an older model of sequencer, but the concept are identical). The four colors red, green, blue and yellow each represent one of the four nucleotides.

The actual gel image, if you could get a monitor large enough to see it all at this magnification, would be perhaps 3 or 4 meters long and 30 to 40 cm wide. We don’t even have to ‘read’ the sequence from the gel – the computer does that for us! Below is an example of what the sequencer’s computer shows us for one sample. This is a plot of the colors detected in one ‘lane’ of a gel (one sample), scanned from smallest fragments to largest. The computer even interprets the colors by printing the nucleotide sequence across the top of the plot. This is just a fragment of the entire file, which would span around 900 or so nucleotides of accurate sequence.

2. HIDDEN MARKOV MODEL (HMM) ARCHITECTURE

- **state transition matrix** : holding the probability of a hidden state given the previous hidden state.
- **output matrix** : containing the probability of observing a particular observable state given that the hidden model is in a particular hidden state.

Thus a hidden Markov model is a standard Markov process augmented by a set of observable states, and some probabilistic relations between them and the hidden states.
An example of a HMM for Protein Sequences



This is a possible hidden Markov model for the protein ACCY. The protein is represented as a sequence of probabilities. The numbers in the boxes show the probability that an amino acid occurs in a particular state, and the numbers next to the directed arcs show probabilities, which connect the states. The probability of ACCY is shown as a highlighted path through the model. There are three kinds of states represented by three different shapes. The squares are called **match states**, and the amino acids emitted from them form the conserved primary structure of a protein. These amino acids are the same as those in the common ancestor or, if not, are the result of substitutions. The diamond shapes are **insert states** and emit amino acids that result from insertions. The circles are special, silent states known as **delete states** and model deletions. These type of HMMs are called Protein Profile-HMMs and will be covered in more depth in the later sections.

Scoring a Sequence with an HMM

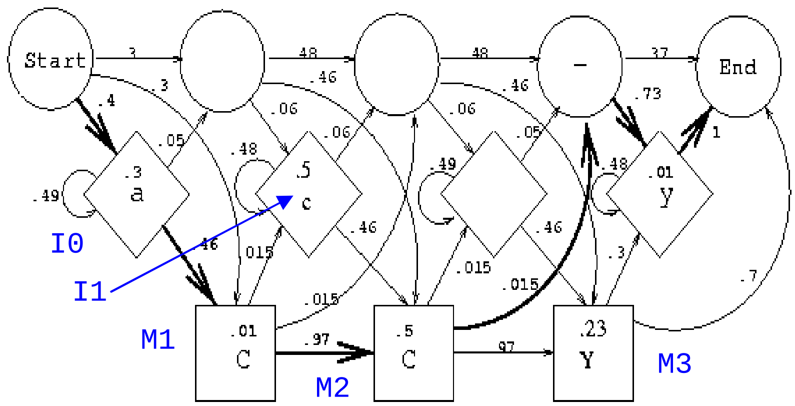
Any sequence can be represented by a path through the model. The probability of any sequence, given the model, is computed by multiplying the **emission** and **transition** probabilities along the path. A path through the model represented by ACCY is highlighted. For example, the probability of A being emitted in position 1 is 0.3, and the probability of C being emitted in position 2 is 0.6. The probability of ACCY along this path is

$$.4*.3*.46*.6*.97*.5*.015*.73*.01*1 = 1.76 \times 10^{-6}.$$

3. THREE PROBLEMS OF HIDDEN MARKOV MODELS

3.1 Scoring Problem

We want to find the probability of an observed sequence given an HMM. It can be seen that one method of calculating the probability of the observed sequence would be to find each possible sequence of the hidden states, and sum these probabilities. We use the Forward Algorithm for this.



Consider the HMM shown above. In this figure several paths exist for the protein sequence ACCY.

The **Forward algorithm** employs a matrix, shown below. The columns of the matrix are indexed by the states in the model, and the rows are indexed by the sequence. The elements of the matrix are initialized to zero and then computed with these steps:

i) The probability that the amino acid A was generated by state I0 is computed and entered as the first element of the matrix. This is .4*.3 = .12
ii) The probabilities that C is emitted in state M1 (multiplied by the probability of the most likely transition to state M1 from state I0) and in state I1 (multiplied by the most likely transition to state I1 from state I0) are entered into the matrix element indexed by C and I1/M1.
iii) The sum of the two probabilities, sum(I1, M1), is calculated.
iv) A pointer is set from the winner back to state I0.
v) Steps 2-4 are repeated until the matrix is filled.

The probability of the sequence is found by summing the probabilities in the last column.

	I0	I1	M1	I2	M2	I3	M3
A	.12	0	0	0	0	0	0
C	0	.015	.005	0	0	0	0
C	0	0	0	0	.012	0	0
Y	0	0	0	0	0	.0000001	.002

Matrix for the Forward algorithm

3.2 Alignment Problem

We often wish to take a particular HMM, and determine from an observation sequence the most likely sequence of underlying hidden states that might have generated it. This is the alignment problem and the **Viterbi Algorithm** is used to solve this problem.

The Viterbi algorithm is similar to the forward algorithm. However in step 3, maximum rather than a sum is calculated. The most likely path through the model can now be found by following the back-pointers.

	I0	I1	M1	I2	M2	I3	M3
A	**.120**	0	0	0	0	0	0
C	0	**.015**	.005	0	0	0	0
C	0	0	0	0	**.23**	0	0
Y	0	0	0	0	0	.0001	**.22**

Matrix for the Viterbi algorithm

Once the most probable path through the model is known, the probability of a sequence given the model can be computed by multiplying all probabilities along the path.

3.3 Training Problem

Another tricky problem is how to create an HMM in the first place, given a particular set of related training sequences. It is necessary to estimate the amino acid emission distributions in each state and all state-to-state transition probabilities from a set of related training sequences. This is done by using the Baum-Welch Algorithm or the Forward Backward Algorithm.

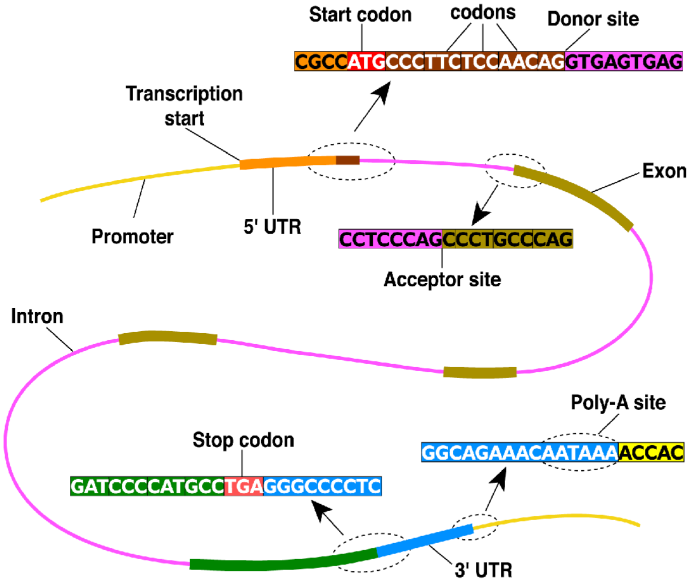
The algorithm proceeds by making an initial guess of the parameters (which may well be entirely wrong) and then refining it by assessing its worth, and attempting to reduce the errors it provokes when fitted to the given data. In this sense, it is performing a form of gradient descent, looking for a minimum of an error measure.

4. GENE FINDING AND PREDICTION

We introduce here the gene-prediction HMMs that can be used to predict the structure of the gene. Our objective is to find the coding and non-coding regions of an unlabeled string of DNA nucleotides.

The motivation behind this is to

- assist in the annotation of genomic data produced by genome sequencing methods
- gain insight into the mechanisms involved in transcription, splicing and other processes



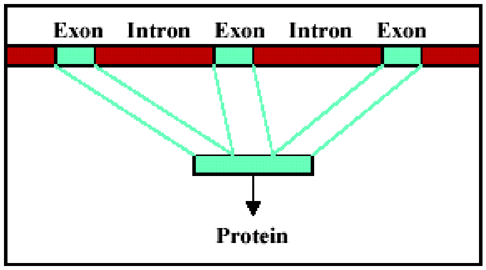
As shown in the diagram above, a string of DNA nucleotides containing a gene will have separate regions

- Introns – non-coding regions within a gene
- Exons – coding regions

These regions are separated by functional sites

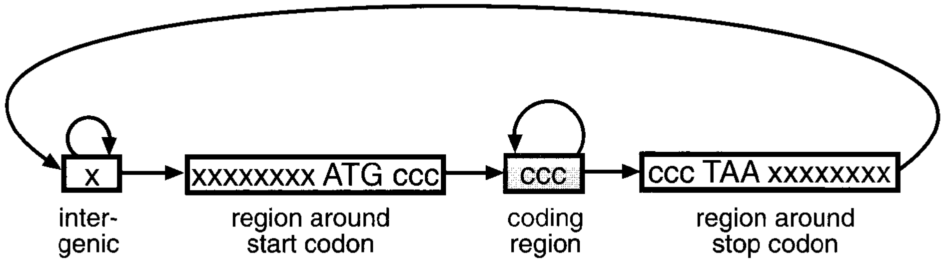
- Start and stop codons
- Splice sites – acceptors and donors

In the process of transcription, only the exons are left to form the protein sequence as depicted below.



Many problems in biological sequence analysis have a grammatical structure . HMMs are very useful in modeling grammar. The input to such a HMM is the genomic DNA sequence and the output, in the simplest case is a parse tree of exons and introns on the DNA sequence.

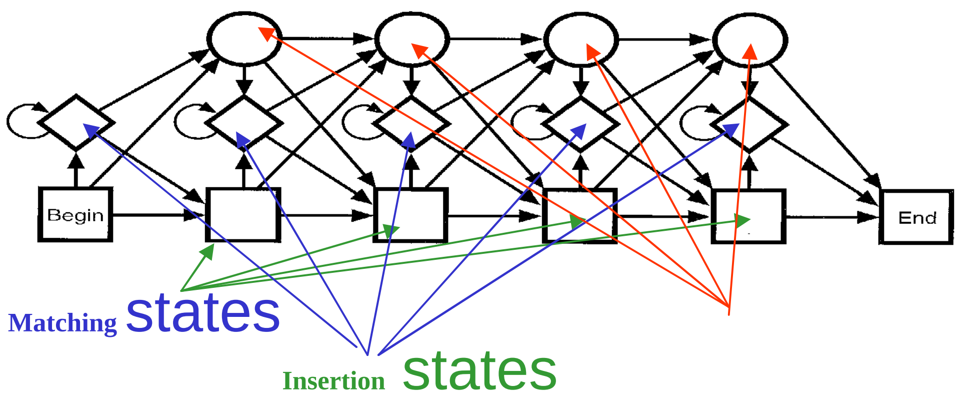
Shown below is a simple model for unspliced genes that recognizes the start codon, stop codon (only one of the three possible stop codons are shown) and the coding/non-coding regions. This model has been trained with a test set of gene data.



Having such a model, how can we predict genes in a sequence of anonymous DNA? We simply use the Viterbi algorithm to find the most probable path through the model

Protein- Profile HMMs

As we have seen earlier, protein structural similarities make it possible to create a statistical model of a protein family which is called a **profile**. The idea is, given a single amino acid target sequence of unknown structure, we want to infer the structure of the resulting protein. The profile HMM is built by analyzing the distribution of amino-acids in a training set of related proteins. This HMM in a natural way can model positional dependant gap penalties.



The basic topology of a profile HMM is shown above. Each position, or module, in the model has three states. A state shown as a rectangular box is a **match** state that models the distribution of letters in the corresponding column of an alignment.

5. PREDICTION OF PROTEIN SECONDARY STRUCTURE USING HMM'S

Prediction of secondary structures is need for the prediction of protein function. As an alternative method to direct X-ray analysis, a HMM is used to

- Analyze the amino-acid sequences of proteins
- Learn secondary structures such as helix, sheet and turn
- Predict the secondary structures of sequences

The method is to train the four HMMs of secondary structure – helix, sheet, turn and other – by training sequences. The Baum-Welch method is used to train the HMMs. So, the HMM of helix is able to produce helix-like sequences with high probabilities. Now, these HMMs can be used to predict the secondary structure of the test sequence. The forward-backward algorithm is used to compute the probabilities of these HMMs outputting the test sequence. The sequence has the secondary structure whose HMM showed the highest probability to output the sequence.

6. HMM IMPLEMENTATION

These are the two publicly available HMM implementation software.

HMMER - http://hmmer.wustl.edu/
SAM system - http://www.cse.ucsc.edu/research/compbio/sam.html

7. CONCLUSION

- **Integration of structural information into profile HMMs.**
  Despite the almost obvious application of using structural information on a member protein family when one exists to better the parameterization of the HMM, this has been extremely hard to achieve in practice.

- **Model architecture**
  The architectures of HMMs have largely been chosen to be the simplest architectures that can fit the observed data. We can use protein structure knowledge to make better architecture decisions, or, in limited regions, to learn the architecture directly from the data. These will implied architectures have implications for our structural understanding.

- **Biological mechanism**
  In gene prediction, the HMM's may be getting close to replicating the same sort of accuracy as the biological machine (the HMM's have the additional task of finding the gene in the genomic DNA context, which is not handled by the biological machine that processes the RNA). What constraints does our statistical model place on the biological mechanism— in particular, we can consider a biological mechanism that could use the same information as the HMM.

REFERENCES

1. Rabiner, L.R. and Juang, B.H. (1986). An Introduction to Hidden Markov Models. *IEEE ASSP Magazine*, January, 1-16.
2. Asai, K.; Hayamizu, S. and Handa, H. (1993). Prediction of protein secondary structures by hidden Markon models. *Computer Application in the Biosciences* (CABIOS), 9(2), 141-146.
3. Krogh, A.; Brown, M.; Mian, I.S. Sjolander, K. and Haussler, D. (1994). Hidden Markov Models in Computational Biology: Applications to Protein Modeling, *J. Mol. Biol.*, 235, 1501-1531.
4. Eddy, S. (1998). Profile hidden Markov models. *Bioinformatics*, 14, 755-763.
5. Rabiner, L.R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2), 257-285.
6. Baldi, P., Chauvin, Y., Hunkapiller, T. and McClure, M.A. (1993). *Hidden Markov Models in Molecular Biology: New Algorithms and Applications*, *In Advances in Neural Information Processing Systems* 5, Eds. S.J. Hanson, J.D. Cowan and C. Lee Giles, (Morgan Kaufmann) pp 747-754.
7. Baldi, P.; Chauvin, Y.; Hunkapiller, T. and McClure, M.A. (1994). Hidden Markov Models of Biological Primary Sequence Information. *Proceedings of the National Academy of Science*, USA 91, 1059-1063.
8. David Kulp, David Haussler, Martin G. Reese, and Frank H. Eeckman, (1996). A generalized hidden markov model for the recognition of human genes in DNA, Proceedings of the *Fourth International Conference on Intelligent Systems for Molecular Biology* (Menlo Park, CA), AAAI Press.
9. Hughey, R. and Krogh, A. (1996). Hidden Markov models for sequence analysis: extension and analysis of the basic method. Computer Applications in the Biosciences, 12:95-107. http://www.cse.ucsc.edu/research/compbio/html_format_papers/hughkrogh96/cabios.html

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 71-80

TIME OF ERUPTION OF PERMANENT TEETH IN PAKISTANI CHILDREN

Nazeer Khan
Dow University of Health Sciences, Karachi, Pakistan
Email: n.khan@duhs.edu.pk

ABSTRACT

Objectives: To establish a norm for mean eruption time of permanent teeth, except the third molars, of Pakistani children. Furthermore to evaluate the effect of gender, type of schools (private/public), height, weight and body mass index on the eruption time.

Subjects and Methods: 4370 children of 'just erupted' teeth (cases) were obtained from 102 randomly selected schools from 18 towns of Karachi, using systematic random sampling procedure. A team of 2 dentists and 2 assistants visited the school to collect the data. The dental examination was carried out for the selected child. Height and weight were also measured. The date of birth was obtained from the school record.

Results: Fifty five percent of the children were male and the mean age was 9.31±2.27 years. The right first molars (# 16 & #46) showed the minimum eruption time in upper and lower jaws. Second molars were the last tooth to emerge in this survey. There was no significant difference of eruption time between gender, except tooth #15, #25 and #43. None of contra lateral (left and right side) teeth showed any statistical significance difference. All the mandible teeth, except the premolars, erupted earlier than maxillary teeth. Private schools children showed early eruption than the public schools children. The Pearson and partial correlation were significant positively correlated with height for most of the teeth. Eruption time of all the teeth, except one, showed positive correlation with weight. Eruption time of only few teeth showed significant partial correlation with weight, and significant Pearson correlation with BMI. Conclusion: In general, the study showed that the eruption time of Pakistani children are different in few respects and agreed in others, when comparing with the information reported in the literature of other countries.

KEYWORDS

Eruption time; permanent teeth; Pakistani children.

INTRODUCTION

Parents consider tooth eruption as an important event in the child's development, and they have often showed their concern about the timing of eruption of teeth. Mostly the information on the age of permanent teeth emergence used in clinical and academic situations in Pakistan is based on American and European standards[1,2]. However, it has been suggested in the literature that standards for tooth emergence should be derived from the population in which they are to be applied because factors related to emergence may vary considerably in both dentitions[3]. Similarly, adequate knowledge of timing of permanent tooth emergence is essential for diagnosis and treatment planning in Pediatric Dentistry and Orthodontics[4]. Furthermore, information on tooth emergence is also used to supplement other maturity indicators in the diagnosis of certain growth disturbances, and in forensic dentistry to estimate the chronological age of children with unknown birth records[4-6]. Therefore, the specific standards of the time of emergence of teeth characterize an important resource for general dental practitioners, orthodontists and pedodontists.

Many studies are conducted in different population and among different ethnic groups all over the world[4,5,7-21]. However, no study is reported from Pakistan, except one for primary teeth[19] and one for permanent teeth, conducted only for boys in pre-partition time[20]. Therefore, there was a need to conduct such a study to establish proper norm for time of eruption of Pakistani children. The objective of the study was **t**o establish a norm for the mean eruption time of permanent teeth, except the third molars, of Pakistani children. Furthermore, to find out the effect of gender, type of schools (private/public), height, weight and body mass index on the eruption time:

SUBJECTS AND METHODS

Karachi city is divided into 18 administrative towns. In year 2007, 3948 public and 2560 private schools were registered in the city schooling system. However, for the better administration a public school is divided into primary, secondary and high schools sections separated for each gender, and are registered as different schools. Therefore, if we pick up a high school, the attached primary and secondary school will automatically be included in the sample. In private schools listing, the system is just opposite. Not only one registered school covers all the primary, secondary and high school sections for both the gender, it also covers many campuses, located at different areas. Therefore, by looking the school lists with this complex setting, it was determined that the number of cases in private and public schools should be divided in the ratio of 3:1. Four thousand cases were planned to obtain for the study. Ten percent more is added into this sum to make sure that minimum committed number is obtained. Literature indicates that about 15-20% of the total children have at least one tooth just erupted. Therefore, we were expecting to examine about 25,000 children to obtained 4,400 cases of just erupted teeth. Assuming that each school enrolls on average 250 children, we were needed to visit about 100 schools to obtain the required number of cases. Furthermore, it was also decided to collect at least 150 cases for every tooth to make sure that mean eruption time for each tooth will be calculate on the basis of a reasonably good number of cases. The expected total was divided into different towns according to the proportion of schools. Schools were randomly selected from the list of schools, using systematic random sampling procedure.

Letters were posted at the addresses mentioned in the list of schools to obtain the permission from the administration. Since the response rate was very poor, the author has visited by himself to the selected schools to explain the purpose of the projects and get the permission from the administration. Time and dates were arranged with administration. A schedule calendar was prepared for the investigation team. A team of 2 dentists (1 male & 1 female) and 2 assistants (1 male & 1 female) visited each school on the assigned day and time. The objective and the benefits of the project were explained to all the students of the class and informed consent was taken. Every present student of the class, who agreed to be in the project, was examined for general checkup. If a child has have just erupted tooth, then that child was taken away from the class room. The criterion of the just erupted teeth was defined as: a tooth deemed to have emerged if any part of it was visible in the mouth. The dental examination was carried out by field examiners using the dental examination kit under fluorescent light. The height was measured in centimeter, using wall-mounted ruler on the child's head with their back and knees completely straight, and their feet together. The weight was measured in kilogram using a commercial digital scale after removal of the shoes only. The date of birth was obtained from the school records. The clinicians were trained and calibrated by showing many clinical pictures of just erupted, unerupted or erupted teeth. No casts or subjects were used for inter or intra examiner calibration. Because it s very easy to distinguish among the above three different conditions for a tooth. Virtanen[21] indicated that the criteria for tooth emergence are so clear that evaluation of the error of the method is not necessary. Kochhar[22] also mentioned that it is relatively simple to decide whether a tooth has emerged or not. The data were analyzed using SPSS statistical software.

RESULTS

Descriptive Statistics: The total number of private and public schools included in this study were 76 and 26, respectively. About 25,000 children were screened from these 102 schools. Out of those, 4394 children of 'just erupted' teeth (cases) were obtained for the study. Twenty four of them were non-Pakistani. Therefore, the analysis was carried out only for 4370 Pakistani children. Fifty five percent of the sample was male and 80.3% of children were studying in grade 1 to grade 5. The mean age of the children was 9.31±2.27 years (R: 2.5 – 17.7 years).

Table 1 shows the descriptive statistics (number of cases, mean, median, standard deviation, and 95% confidence interval of mean) of eruption time of both the jaws and p-value for antagonist teeth (upper and lower corresponding teeth). The minimum mean eruption time was 6.5±1.1 years of right first molar of mandibular jaw (# 46). The maximum eruption time of 11.8 years was for the 2nd molars of maxillary jaw. All the mandible teeth, except the premolars, erupted earlier than maxillary teeth. The difference of mean eruption time all the contra-lateral (left and right) teeth did not show any statistical significance. Table 2 discusses the mean eruption time for male and female children. There was no significant difference of eruption time between gender in all the studied teeth, except tooth #15, # 25 and # 43. Females showed significantly late eruption in mandible second premolars (#15 and #25) and early eruption in right mandible canine (#43). Table 3 depicts the mean eruption time among private and public schoolchildren. Twenty five teeth, out of 28, of the of private schools children showed early eruption than the public schools and 17 of them: tooth type (#17, #15, #13, #21, #22, #23, #25, #26, #27, #47, #44, #42 #31, #32, #34, #35 and #37), were statistically significant. Table 4 shows the Pearson and partial correlation of eruption age with height of the patients. The Pearson correlation was significant positively correlated with height for all the teeth ($p<0.0001$), except tooth #31 and #35 ($p=0.057$ and 0,076, respectively). The partial correlation of eruption time with height, controlled for weight, for all the teeth was also significant positively correlated ($p<0.0001$), except tooth #31 and #35. Table 5 shows the Pearson and partial correlations of eruption time with weight. Eruption time of all the teeth, except tooth # 42, showed positive correlation with weight. However, only 15 of them showed statistically significant for non-zero correlation. Only five teeth (#17, #26, #42, #41, #32) showed significantly non-zero partial correlation of weight, controlling with height. All of them showed negative sign. Table 6 shows the Pearson correlation between the eruption time and Body Mass Index (BMI). Only 7 teeth (#16, #26, #45, #44, #43, #41, and #32) showed significant correlation.

DISCUSSION

In the literature different population groups are targeted to determine the mean eruption time of permanent teeth. However, no reported data are available for Pakistani children, except an article published in pre-partition time for the mean eruption time of Lahori boys[17]. Due to unavailability of local data, the standards for eruption time being taught in dental colleges of Pakistan, are based on non-Pakistani population, especially American and European standards[1,2]. It is documented in the literature that significant variation exists in time of eruption and emergence sequence in different population[3]. Therefore, it was necessary to make an investigation of the standard values of eruption time of Pakistani children. This report presents baseline information for time and sequence of eruption of permanents teeth of Pakistani children. Furthermore, except the Iranian study[12], all the previous studies established the standard of eruption time on moderate or small sample sizes. This study is quit a comprehensive one and covered all towns of Karachi, the largest metropolitan city of Pakistan.

About thirty three million children were enrolled up to the secondary schools (grade 1 to grade 10) in Pakistan in 2006[23]. Fifty seven percent of them were male children. In this study the percentage of male children was 55%, which was not very far from the national data. In many parts of Pakistan, especially in rural areas, families do not send their daughters to school. Karachi is mostly urbanized and this type of negative attitude against females does not exist in this town, therefore, we were expecting higher percentage of females' enrolment.

The study was not a hospital based-study. Therefore no radiograph was available. Consequently, it was not possible to determine the congenitally missing teeth. Holman et al[24] discussed the impact of congenitally missing teeth on the mean eruption time. They concluded that estimates of eruption time without considering congenitally teeth were biased upward (always less than 1%), and the standard deviations were consistently overestimated by 3-5%. However, for adequate sample sizes agenesis does not lead to substantially biased estimates. Since in this study the sample size was quit large enough, therefore the effect of congenitally missing teeth would not be of significant.

This study did not show any trend of difference of eruption time between male and female children. None the study teeth, except maxillary second premolars (#15 & #25) and mandibular right canine (#43), showed any statistical significant difference among gender. Maxillary and mandibular significant values were in opposite directions for time of eruption in males and females. These results of mostly insignificant differences and no clear-cut trends in males and females eruption time were not agreed with almost all the other studies, where they have shown that the girls have advanced eruption time than male children[4,6-8,14,22, 25-28]. However, these results agreed with few other studies[29-31]. This result of no significant eruption time between male and female children is most important finding of this study. Most of the studies, which have indicated no significant difference, were small scale studies, except the Nigerian study of Ajamni and Jain[29]. Therefore, the eruption pattern among Pakistani children is very much distinct with other nationalities in this regard.

The mean eruption time of none of the contralateral (right and left) teeth were statistically significant. Therefore, the eruption time of contralateral teeth was symmetrical. This finding was agreed by almost all the studies mentioned in the literature. Mean eruption time of all the mandible teeth, except the premolars and first molars, showed statistically significant early eruption than maxillary teeth. The largest difference between antagonist teeth was observed in canines and incisors. This trend of early eruption with significant differences of mandible teeth, except the premolars and first molars, is agreed to other studies[7,8,12,17,22,32]. However, the study of Nanda[33] did not agree with this result.

The children from private schools showed early eruption than public schools. Seventeen teeth showed significantly earlier eruption to the private schoolchildren. In Pakistan, usually children of low socio-economic class enroll in the public schools. Therefore, malnutrition could play a major role in the delayed eruption. Triratana et al[34] and Psoter et al.[35] showed that the malnutrition children usually have delayed eruption than the children grow with normal healthy diet.

In our study it was observed that eruption of teeth was positively related to somatic growth (height and weight) of the children and the results are in agreement with Billewicz[8] and Agarwal[15]. All partial correlations between eruption times with heights (controlling the weights) were positively correlated with statistical significant ($p<0.05$), except tooth # 31 and #35 and these partial correlations were not very much different to the Pearson linear correlations (without controlling the weight). Pearson linear correlations of eruption times with weight (without controlling height) were all positive and statistically significant. However, when partial correlations with weights were computed (controlling the height), all the values became negative except one tooth in maxillary jaw and 5 teeth in mandible jaw, and only five of those negative correlations were statistically significant. It infers that the children who are tall, it did not matter whether they are heavy weight or not, going to have delayed eruption. However, if they are heavy, it would be early eruption if they are not tall and delayed eruptions if they are tall. Due to these conflicting outcomes of the linear and partial correlation height and weight with eruption time, the BMI showed 50% positive correlation and 50% negative correlations. Therefore, no prediction for the eruption time can be made on the basis of BMI.

The difference of eruption of time in different populations are linked to the many attributes, like genetics factors[36,37], environmental factors including the socioeconomic status and nutrition[6,8,10,11,38,39], and climate[40]. It was also found that eruption times were also varied among the same ethnic groups[3,9,409] and also attributed to genetic variations[9]. In reading of reported differences in the eruption time, the readers should be careful for the definition of emergence and eruption, sampling methods, sample sizes, age groups, number of cases in each group, eating habits and socio-economic effects.

CONCLUSIONS

The following conclusions were obtained from this study:

- On average the first erupted tooth was the right maxillary first molar, emerged at the age of 6.5 years and the last tooth was mandible 2nd molar, emerged at the age of 11.8 years,
- There is no significant difference of eruption time between gender in all the studied teeth, except tooth #15, # 25 and #43.
- There was no significant difference between the contralateral teeth of right and left side,
- All the mandible teeth, except the premolars, erupted earlier than maxillary teeth.

- Ninety percent of the teeth of private schools children showed early eruption than the public schools.
- The children who are tall, it did not matter whether they are heavy weight or not, showed delayed eruption. However, if they are heavy, it would be early eruption if they are not tall and delayed eruptions if they are tall.
- There was no systematic relationship between the eruption time and Body Mass Index (BMI),

In general, the study showed that the eruption time of Pakistani children are different in few respects and agreed in others, when comparing with the information reported in the literature of other countries.

ACKNOWLEDGEMENTS

First of all, I am thankful the Higher Education Commission for providing me the fund to conduct this study. I am extremely grateful to my team members: the coordinator, Ms Maria Khan; the field investigators: Drs. Rafia Amber, Mohammad Amir, Danish Younus and Syed Wajahat AA Zaidi; and the data collectors: Ms. Tahseen Bashir, and Mr. Syed Qamarul Hasan, for completing this project with full interest and energy. I am also indebted to the principals and administrative staff of all 102 schools for their cooperation in screening their school children.

REFERENCES

1. Profitt, W.R., Fields, H.W., Ackerman, H.L., Thomas, P.M. and Tulloch, J.F.C. (2000). *Contemporary Orthodontics*. St. Louis: The CV Mosby Company, 69-80.
2. Thilander, B. and Ronning, O. (1995). *Introduction to Orthodontics*. Stockholm: Gothia, 43-49.
3. Demirjian, A. Dentition (1986). In: Falkner F, Tanner JM, editors. *Human growth*. London: Baillere Tindall, 413-444.
4. Mugonzibwa, E.A., Kuijpers-Jagtman, A.M., Laine-Alava, M.T., van't Hof M.A. (2002). Emergence of permanent teeth in Tanzanian children. *Community Dent Oral Epidemiol*, 30, 455-462.
5. Garn, S.M., Sandusky, S.T., Nagy, J.M. and Trowbridge, F.L. (1973). Negre-Caucasoid difference in permanent tooth emergence at a constant income level. *Arch Oral Biol*, 18, 606-615.
6. Nyström, M., Peek, L., Kleemola-Kujala, E., Evälahti, M. and Kataja, M. (2000). Age estimation in small children: reference values based on counts of deciduous teeth in Finns. *Forensic Sci. Int*., 110(3), 179-88.
7. Diamanti, J. and Townsend, G.C. (2003). New standards for permanent tooth emergence in Australian children. *Aust. Dent. J.*; 48, 39-42.
8. Billewicz, W.Z. and McGregor, I.A. (1975). Eruption of permanent teeth in West African (Gambian) children in relation to age, sex and physique. *Ann of Hum Biol*, 2, 17-28.
9. Pahkala, R., Pahkala, A. and Laine T. (1991). Eruption pattern of permanent teeth in a rural community in North Eastern Finland. *Acta Odontol Scand*; 49, 341-349.
10. Lee, M.M.C., Low, W.D. and Chang, S.F.S. (1965). Eruption of the permanent dentition of Southern Chinese children in Hong Kong. *Arch Oral Biol*.; 10, 849-861.

11. Nonaka, K., Ichiki, A. and Miura, T. (1990). Changes in the eruption order of the first permanent tooth and their relation to season of birth in Japan. *Am. J. Phys. Anthropol.*, 82, 191-198.
12. Moslemi, M. (2004). An epidemiological survey of the time and sequence of eruption of permanent teeth in 4-15 –year-olds in Tehran. *Iran Int. J. Paediatr Dent.*, 14, 432-438.
13. Wedl, J.S., Schoder, V. and Friedrich, R.E. (2004). Tooth eruption times of permanent teeth in male and female adolescents of a country district in Lower Saxony. *Arch Kriminol*; 213, 84-91.
14. Wedl, J.S., Schoder, V., Blake, F.A., Schmelzle, R. and Friedrich, R.E. (2004). Eruption times of permanent teeth in teenage boys and girls in Izmir (Turkey). *J. Clin. Forensic Med.*, 11, 299-302.
15. Agarwal, K.N., Gupta, R., Faridi, M.M. and Arora, N.K. (2004). Permanent dentition in Delhi boys of age 5-14 years. *Indian Pediatr*; 41, 1031-1035.
16. Chohan, A.N., Khan N.B., Al Nahedh L., Bin Hassan, M., and Al Sufyani N. (2007). Eruption time of permanent first molars and incisors among female primary school children of Riyadh. *J. Dow Univ. Health Sci.*, 1(2), 53-58.
17. Khan. N., Chohan, A.N., Al-Nasser, F., Al Sharani, D., Al-Jorais, R., Al-Salehi, R. and Al-Yousef, T. (2008). Statistical presentation of eruption age of permanent second molars, premolars and canines in female school children living in Riyadh, Saudi Arabia. *Saudi Dent. J.*, 20(3), 140-9.
18. Khan, N.B., Chohan, A.N., l-Magrabi, B., Al-Deyab, S., Zahid, T. and Al-Moutairi, M. (2006). Eruption time of permanent first molars and incisors among a sample of Saudi male schoolchildren. *Saudi Dent. J.*, 18(1), 18-24.
19. Saleemi, M.A., Hägg, U., Jalil, F. and Zaman, S. (1994). Timing of emergence of individual primary teeth. A prospective longitudinal study of Pakistani children. *Swed. Dent. J.*, 18, 107-12.
20. Shourie, K.L. (1946). Eruption age of teeth in India. *Ind. J. Med. Res.*; 34(1), 105-18.
21. Virtanen, J., Bloigu, R.S. and Larmas, M.A. (1994). Timing of eruption of permanent teeth: standard Finnish patient documents. *Community Dent Oral Epidemiol.*, 22, 286-8.
22. Kochhar, R. and Richardson, A. (1998). The chronology and sequence of eruption of human permanent teeth in Northern Ireland. *Int. J. Paediatr Dent.*; 8, 243-52.
23. Government of Pakistan (2007). *Monthly Bulletin of Statistics*, Statistics Division, Federal Bureau of Statistics, 55(2), 12.
24. Holman, D.J. and Jones, R.E. (1998). Longitudinal analysis of deciduous tooth emergence, Part II. Parametric survival analysis in Bangladeshi, Guatemalan, Japanese, and Javanese children. *Am. J. Phys. Anthropol*, 105, 209-30.
25. Nizam, A., Naing, L. and Mokhtar, N. (2003). Age and sequence of eruption of permanent teeth in Kelantan, North Eastern Malaysia. *Clin. Oral Investiq*, 7, 222-25.
26. Nonaka, K., Ichiki, A. and Miura, T. (1990). Changes in the eruption order of the first permanent tooth and their relation to season of birth in Japan. *Am. J. Phys. Anthropol*; 82(2), 191-8.
27. Friedrich, R.E., Katerji, H., Wedl, J.S. and Scheuer, H.A. (2006). Eruption times of permanent teeth in children and adolescents of Federborn, Westphalia, Germany. Arch Krimnol; 217(1-2), 20-35.
28. Wedl, J.S., Schoder, V. and Friedrich, R.E. (2004). Tooth eruption times of permanent teeth in male and female adolescents in Niedersachsen. *Arch Kriminol*; 213(3-4), 84-91.

29. Ajamani, M.L. and Jain, S.P. (1984). Eruption age of teeth in Nigeria. Anat. Anz.; 157, 245-52.
30. Friedrich, R.E., Leist, A. and Scheuer, H.A. (2008). Eruption times of permanent teeth in children and adolescents in the German state of Saarland. *Arch Kriminol*, 222(3-4), 73-104.
31. Khatskevich, G.A. and Bogomolova, I.A. (2004). Time of permanent teeth eruption in schoolchildren of Saint-Petersburg. *Stomatologiia* (Mosk), 83(3), 53-7.
32. Leroy, R., Bogaerts, K., Lesaffre, E. and Declerck, D. (2003). The emergence of permanent teeth in Flemish children. *Comm. Dent. Oral Epidemiol*, 31, 30-39.
33. Nanda, R.S. (1960). Eruption of Human Teeth. *Am. J. Orthod*; 46(5), 363-78.
34. Triratana, T., Hemindra, Kiatiparjuk, C. (1990). Eruption of permanent teeth in malnourished children. *J. Dent. Assoc. Thai*; 40, 100-108.
35. Psoter, W., Gebrian, B., Prophete, S., Reid, B. and Katz, R. *Effect of early childhood malnutrition on tooth eruption in Haitian adolescents*.
36. Arvystals, M.G. (1976). Familial generalized delayed eruption of the dentition with short stature. *Oral Surg. Oral Med. Oral Pathol.*; 41, 235-243.
37. Hatton, M.E. (1955). Measure of the effects of heredity and environment on eruption of the deciduous teeth. *J. Dent. Res.*, 34, 397-401.
38. Adler P. (1963). Effect of some environmental factors on sequence of permanent tooth eruption. *J. Dent. Res.*, 42, 605-616.
39. Johannessen, A.C., Bjorvatn, K., Myklebust, D. (1989). Oral health in children attending the church-afflicted schools in Northern Cameroon: tooth development and dental caries. *Odontostomatol Trop.*, 12, 67-8.
40. Friedlaender, J.S., Bailit, H.L. (1969). Eruption times of the deciduous and permanent teeth of natives on Bougainville island, territory of New Guinea: a study of racial variations. *Hum. Biol.*; 41, 51-65.

Table 1:
Descriptive statistics of eruption time of all the teeth, except third molars, in maxillary jaw

Tooth No	No of Cases	Mean	Median	SD	95% CI of mean	Tooth No	No of Cases	Mean	Median	SD	95% CI of mean	p-value*
17	228	11.8	11.8	1.6	(11.6,12.0)	47	428	11.3	11.2	1.6	(11.2, 11.5)	< 0.0001
16	157	6.6	6.4	1.2	(6.4,6.8)	46	239	6.5	6.4	1.1	(6.4, 6.6)	0.278
15	191	10.4	10.3	1.5	(10.2,10.6)	45	166	10.7	10.7	1.7	(10.5, 11.0)	0.067
14	286	10.1	10.0	1.4	(9.9,10.2)	44	287	10.5	10.3	1.6	(10.3, 10.7)	0.001
13	597	10.9	10.8	1.5	(10.8,11.0)	43	359	10.2	10.2	1.7	(10.1, 10.4)	< 0.0001
12	337	8.4	8.3	1.3	(8.2,8.5)	42	308	7.8	7.7	1.2	(7.8, 7.9)	< 0.0001
11	326	7.5	7.4	1.5	(7.4,7.7)	41	216	6.9	6.8	1.1	(6.7, 7.0)	< 0.0001
21	324	7.5	7.3	1.4	(7.3,7.6)	31	193	7.0	7.0	1.2	(6.9, 7.2)	< 0.0001
22	323	8.4	8.3	1.3	(8.3,8.5)	32	293	7.9	7.8	1.3	(7.8, 8.1)	< 0.0001
23	570	10.9	10.9	1.4	(10.8,11.0)	33	353	10.1	9.9	1.6	(9.9, 10.3)	< 0.0001
24	305	10.1	10.1	1.5	(10.0,10.3)	34	279	10.3	10.3	1.4	(10.1, 10.4)	0.222
25	229	10.3	10.2	1.4	(10.1,10.5)	35	162	10.6	10.6	1.5	(10.3, 10.8)	0.054
26	159	6.7	6.6	1.0	(6.6,6.9)	36	220	6.6	6.3	1.1	(6.4, 6.7)	0.157
27	230	11.8	11.7	1.5	(11.6,12.0)	37	415	11.4	11.3	1.6	(11.2, 11.5)	0.001

Table 2:
Comparison of eruption time among gender

Tooth type	Male N	Male $\bar{X} \pm SD$	Female N	Female $\bar{X} \pm SD$	P-value	Tooth type	Male N	Male $\bar{X} \pm SD$	Female N	Female $\bar{X} \pm SD$	P-value
17	137	11.6 ± 1.6	91	12.0 ± 1.5	0.865	47	248	11.4 ± 1.5	180	11.2 ± 1.6	0.393
16	85	6.6 ± 1.2	72	6.6 ± 1.2	0.952	46	130	6.6 ± 1.3	109	6.4 ± 1.0	0.189
15	118	10.2 ± 15.	73	10.8 ± 1.5	0.008	45	88	10.7 ± 1.9	78	10.8 ± 1.4	0.619
14	158	10.1 ± 1.3	128	10.1 ± 1.5	0.940	44	155	10.5 ± 1.6	132	10.4 ± 1.5	0.521
13	396	11.0 ± 1.5	201	10.7 ± 1.5	0.065	43	203	10.4 ± 1.8	155	10.0 ± 1.5	0.009
12	177	8.4 ± 1.1	160	8.4 ± 1.5	0.625	42	188	7.8 ± 1.2	120	7.7 ± 1.1	0.488
11	192	7.5 ± 1.3	134	7.5 ± 1.7	0.985	41	121	6.8 ± 1.1	95	7.0 ± 1.2	0.316
21	187	7.5 ± 1.1	137	7.5 ± 1.6	0.919	31	122	7.0 ± 1.2	71	7.1 ± 1.4	0.558
22	180	8.5 ± 1.2	143	8.3 ± 1.4	0.398	32	172	7.9 ± 1.1	121	8.0 ± 1.5	0.287
23	385	10.9 ± 1.4	185	10.9 ± 1.4	0.959	33	201	10.2 ± 1.5	151	10.0 ± 1.6	0.074
24	165	10.1 ± 1.4	140	10.1 ± 1.6	0.990	34	154	10.3 ± 1.4	125	10.3 ± 1.5	0.978
25	137	10.0 ± 1.3	92	10.7 ± 1.5	0.001	35	97	10.5 ± 1.4	65	10.7 ± 1.7	0.339
26	85	6.7 ± 1.1	74	6.7 ± 1.0	0.972	36	124	6.6 ± 1.2	96	6.5 ± 1.0	0.259
27	141	11.7 ± 1.5	89	12.0 ± 1.5	0.170	37	238	11.3 ± 1.7	177	11.5 ± 1.6	0.093

Table 4:
Pearson and partial correlation of eruption time with height of the children

Tooth Type	No. of Cases	Pearson Correlation r	Pearson Correlation p-value	Partial Correlation r	Partial Correlation p-value	Tooth Type	No. of Cases	Pearson Correlation r	Pearson Correlation p-value	Partial Correlation r	Partial Correlation p-value
17	229	0.319	<0.0001	0.342	<0.0001	47	429	0.365	<0.0001	0.318	<0.0001
16	157	0.499	<0.0001	0.254	<0.0001	46	239	0.333	<0.0001	0.263	<0.0001
15	194	0.249	0.001	0.227	0.002	45	168	0.336	<0.0001	0.206	0.008
14	289	0.385	<0.0001	0.306	<0.0001	44	297	0.304	<0.0001	0.187	0.002
13	559	0.210	<0.0001	0.209	<0.0001	43	408	0.459	<0.0001	0.336	<0.0001
12	337	0.245	<0.0001	0.202	<0.0001	42	308	0.149	0.009	0.188	0.001
11	326	0.207	<0.0001	0.222	<0.0001	41	215	0.266	<0.0001	0.316	<0.0001
21	324	0.205	<0.0001	0.217	<0.0001	31	193	0.137	0.057	0.109	0.134
22	324	0.170	0.002	0.182	0.001	32	294	0.271	<0.0001	0.324	<0.0001
23	529	0.173	<0.0001	0.150	<0.0001	33	393	0.368	<0.0001	0.268	<0.0001
24	313	0.357	<0.0001	0.292	<0.0001	34	282	0.270	<0.0001	0.250	<0.0001
25	234	0.275	<0.0001	0.224	0.001	35	162	0.140	0.076	0.072	0.363
26	158	0.285	<0.0001	0.365	<0.0001	36	222	0.310	<0.0001	0.226	0.001
27	230	0.234	<0.0001	0.256	<0.0001	37	417	0.314	<0.0001	0.266	<0.0001

Table 5:
Pearson and partial correlation of eruption time with weight of the children

Tooth Type	No. of Cases	Pearson Correlation		Partial Correlation		Tooth Type	No. of Cases	Pearson Correlation		Partial Correlation	
		r	p-value	r	p-value			r	p-value	r	p-value
17	229	0.080	0.230	-0.154	0.022	47	429	0.201	<0.0001	-0.070	0.152
16	157	0.452	<0.0001	0.101	0.213	46	239	0.212	0.001	0.021	0.745
15	194	0.118	0.105	-0.056	0.443	45	168	0.276	<0.0001	0.052	0.510
14	289	0.250	<0.0001	-0.501	0.400	44	297	0.246	<0.0001	0.034	0.574
13	559	0.076	0.063	-0.075	0.070	43	408	0.333	<0.0001	-0.019	0.718
12	337	0.143	0.009	-0.011	0.084	42	308	-0.004	0.949	-0.116	0.042
11	326	0.050	0.368	-0.096	0.085	41	215	0.034	0.619	-0.180	0.008
21	324	0.053	0.345	-0.089	0.111	31	193	0.087	0.229	-0.021	0.771
22	324	0.049	0.382	-0.078	0.162	32	294	0.048	0.411	-0.190	0.001
23	529	0.094	0.026	-0.033	0.431	33	393	0.263	<0.0001	-0.013	0.805
24	313	0.224	<0.0001	-0.064	0.267	34	282	0.137	0.023	-0.086	0.154
25	234	0.168	0.011	-0.033	0.619	35	162	0.126	0.109	0.037	0.644
26	158	0.046	0.567	-0.242	0.002	36	222	0.225	0.001	0.060	0.380
27	230	0.053	0.425	-0.120	0.071	37	417	0.178	<0.0001	-0.034	0.494

Table 6:
Pearson and partial correlation of eruption time with BMI of the children

Tooth Type	No. of Cases	Pearson Correlation		Tooth Type	No. of Cases	Pearson Correlation	
		r	p-value			r	p-value
17	229	-0.060	0.370	47	429	0.026	0.595
16	157	0.190	0.018	46	239	-0.085	0.188
15	194	-0.016	0.824	45	168	0.182	0.019
14	289	0.050	0.400	44	297	0.121	0.042
13	559	-0.036	0.383	43	408	0.127	0.016
12	337	-0.004	0.946	42	308	-0.097	0.089
11	326	-0.082	0.139	41	215	-0.206	0.002
21	324	-0.088	0.113	31	193	-0.005	0.940
22	324	-0.071	0.204	32	294	-0.142	0.015
23	529	0.008	0.845	33	393	0.100	0.060
24	313	0.030	0.597	34	282	-0.010	0.863
25	234	0.018	0.784	35	162	0.072	0.363
26	158	-0.212	0.007	36	222	-0.045	0.505
27	230	-0.060	0.365	37	417	0.016	0.749

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 81-86

ENGEL EQUIVALENCE SCALE AND THE COST OF CHILDREN

Hazrat Yousaf[1] and **Khalid Khan**[2]
Department of Economics, Lasbela University of Agriculture,
Water and Marine Sciences, Uthal, Balochistan, Pakistan
Email: [1]hazrataerc@yahoo.com; [2]khlidkk82@yahoo.com

ABSTRACT

This study presents and discusses the Engel equivalence scales and the cost of children for twenty household types by applying the Semi-logarithm specification of Engel's model and the Household Integrated Expenditure Survey (HIES) of Pakistan for the year 2005-06. The coefficients have expected sings, reasonable magnitudes, individually and jointly significant. The value of Adj. R^2 is plausibly high for cross-section data with larger number of observations. The values of Engel scales are setting in the range of between 1.00 and 3.2. The cost of a child aged 0 -10 to a childless couple is equal to 16.50%; whereas that a child aged 10 -15 is 18.95% larger than that of former. The study also concludes that the households' expenditure and demographic compositions by age are the main determinants that effect the households' consumption of food items significantly.

1. INTRODUCTION

Equivalence scale is an index number and "is usually expressed as a set of numbers where the value corresponding to an arbitrary chosen reference household is set equal to unity. The values that correspond to other household types are then expressed relative to this base." (Tran and Peter, 1991, p.221). The existent literature regarding equivalence scales has mentioned three main approaches on the construction of equivalence scales. 'The first approach uses nutritional needs by household demographic compositions by age and gender. The second is based on the use of survey questionnaires directly collected from households. The third approach is included observed expenditure patterns of households' (Panos Tsakloglou, March 1991, p. 343). Equivalence scales are the economist's way of answering to the questions of cost of living, welfare, poverty, inequality and the cost of children of different household types.

The first approach application was by Rowntree (1901) in his pioneering study of poverty and the bundle of goods employed was just a minimum acceptable quantity of food, rent and a small allowance.

Two adult household with no children was assigned the index of 100 and the measuring costs of all other household compositions relative to this. The results suggest that the minimum need of a couple with one child cost 24 percent is larger than of childless couple's cost.

Beveridge (1942) extended the study of Rowntee just by add fuel, light and a margin for "inefficiency" in purchasing. The findings suggest that minimum need of a couple with one child cost 22 percent to childless couple.

Rueben C. Buse and Larry E. Salathe (1978) estimates the adult equivalence scales for five food groups and total food consumed at home by using the USDA 1955 and 1965 household food survey data, respectively. The study concludes that children consume less total food—beef, pork and vegetables that adults, while the middle age spend less on fruit than the elderly.

Julie A. Nelson (1993) presents household equivalence scales in use and historical perspective. The paper is completely in descriptive form and the objective is to add clarity to equivalence scales literature in the light of its historical and philosophical foundation.

The study of Gianni Betti (2000) estimates the equivalence scales based on the Quadratic Almost Ideal Demand System (QAIDS). The Italian households' budget survey 1985-1994 pooled data are used by the author.

Tran Binh and Peter Whiteford (1991) estimate the household equivalence scales by using the 1984 Household Expenditure Survey (HES) of Australia. They apply the Engel and ELES. They suggest that it is sensible to adopt a range of values instead of a single value for the equivalence scales of a particular household type.

White and Masset (2007) estimate child poverty in Vietnam by using adult equivalence scales. They use the Vietnam Living Standard Surveys (VLSS) of the years 1992-93 and 1997-98, which was conducted by the Vietnam's Central Statistical Office. The Engel and Rorbarth models are estimated for the equivalence scales computations and suggest that the better-off families (which are largely urban, more educated, female-headed and from the Khin ethnic majority) spend a lower share of their income on children than the other household groups. They also reveal that the children from better-off family groups have experienced a more rapid increase in welfare level than the children from other sectors of Vietnam society by comparison of the data 1992-93 and 1997-98.

The study of Aly Ercelawn (1991) includes the HIES 1984/85 data set of Pakistan in order to compute adult equivalent household. The author adopts nutritional needs by household demographic compositions and daily calorie intake of 2550 calorie per adult equivalent as a reference.

Panos Tsakloglou (1991) covers estimation and comparison of two simple models (Engel and Rothbarth) of equivalence scales for the cost of children. The author uses Greek Household Expenditure Survey (GHES) which was conducted in 1981/82. He observes that the estimated scales seem to differ considerably across models. The Engel scales are substantially higher than the Rothbarth scales. The study also identifies that the Engel model overestimates whilst the Rothbarth model underestimates the cost of children.

The current piece of work is organized as follows: Section 2 discuss the methodology and data applied for estimations. Section 3 presents results and discussion. Finally, section 4 consists of conclusion and policy implication.

2. METHODOLOGY AND DATA

The current study is based on the third approach of equivalence scale (*Engel scale*) estimation and the following *Panos Tsakloglou (March 1991)* methodology is adopted. The direct utility function is given by

$$u = u(q, z) \qquad (1)$$

where u is utility, q and z are the vector of commodities consumed and the vector of household demographic compositions, respectively. The relation in *(1)* must be hold if at the minimum money cost it is possible to achieve a given utility at given demographic compositions as

$$c\ (u, p, z) = m \tag{2}$$

Since the equivalence scale must be calculated with respect to a reference (the childless two adult) household and then divide the cost function of any household say (*h)* by the cost function of the reference household r as

$$\lambda^h = c(u^0,\ p^0, z^h)\ /c(u^0, p^0, z^r) \tag{3}$$

u^0 *and* p^0 indicate selected a reference utility level and price vector, respectively.

Using Shepherd's lemma due to the unobservable element of equations *(1), (2) and (3)* we can derive the Hicksian (Compensated) demand function

$$p_i\, q_i\ \partial c(u,p,z)\ /\ \partial\ lnp_i \tag{4}$$

Using the Roy's Identity and derive the Marshallian (Ordinary) demand function (see layard and walter, p.145)

$$p_i q_i\ = f\ (m, p, z) \tag{5}$$

The components of equation *(5)* are observable and these relations are usually estimated by appropriate econometric methods. A simple way to estimate the relations of dependent variable and explanatory variables, to compute equivalence scale and determine the cost of children is to use a functional form of the Engel curve in the following matrix notation

$$Y = X\beta + u_i \tag{6}$$

In general, "$\boldsymbol{Y}$" is an (n * 1) vector of observations, "$\boldsymbol{X}$" is an (n *k) matrix of the level of explanatory variables, $\boldsymbol{\beta}$ is a (k * 1) vector of the regression coefficients and *ui* is (n * 1) vector of error term.

For empirical analysis the following model is estimated

$$w_f = \beta_0 +\ \beta_1\ lnm + \beta_2 n_1 + \beta_3 n_2 + u_i \tag{7}$$

where

w_f = Budget share for *food* commodity group.
m = total household expenditure (income).
n_1 = children in the age bracket 0-10 year old.
n_2 = children in the age bracket 10-15 year old.
β_s are parameters to be estimated.

The basic Engel equivalence scale for the *h type* household is given by

$$\lambda^h = m^h\ /m^r\ = \exp\ \{ -\ (\ \beta_2 n_1 + \beta_3 n_2) /\ \beta_1 \} \tag{8}$$

DATA

The estimations are based upon the data available in the Household Income and Expenditure Survey (HIES)[1] of Pakistan. This survey is conducted by the Federal Bureau of Statistics on annual basis[2]. We use the survey for 2005-06. The data tapes were being the latest available at the start of this study. The gross sample consists of over 15453 households. The universe of HIES survey consists of all urban and rural areas. The entire sample of households has been drawn from 1109 primary sampling units out of which 531 are urban and 578 are rural that comprises 50,588 mouzas/village/dehs and 28,000 enumeration blocks in all rural and urban areas of the country in the survey. The raising factors given in the survey are used to give household weights[3] to correct for over sampling of some provinces or locations. The sampling universe does, however, included AJK, Northern and FATA areas. At the first stage we obtain the monthly quantities consumed (for sixty five food commodities) and their respective consumption expenditures through SPSS package. The food expenditures include as the sum value of paid and consumed, wages and salaries in kinds, own produced and consumed receipts from assistance gift, dowry etc. The final data set consist of monthly food share, total household expenditure (income), demographic composition by age and the net sample includes of 4285 households.

3. RESULTS AND DISCUSSION

The parameter estimates of the preferred Engel model are presented in *table I* alongwith their significances as shown by the *t-ratios* and the last two columns assign the *Adj.* R^2 and *F-test* of the regression. The coefficients of demographic variables are positive and statistically significance at 5 per cent level. It implies that the proportion of household budget is devoted to food due to the presence of children. Furthermore, the

Table I: Parameter Estimates of Engel Equation for Pakistan: 2005-06

$$w_f = \beta_0 + \beta_1 \ln m + \beta_2 n_1 + \beta_3 n_2 + u_i$$

Parameters	*Estimates*	*t-ratio*	*Adj.R²*	*F-test*
β_0	1.650	69.32	0.42	1027.08
β_1	-0.144	-52.32		
β_2	0.022	24.66		
β_3	0.025	12.86		

Source: Calculated by authors using the HIES 2005-06 data set of Pakistan.

---

1 HIES was started in 1963 and continued to be carried out with some breaks. In 1990 the questionnaire was revised in order to address the requirements of new system of National Accounts. The surveys were constructed during 1990, 1992-93, 1993-94 and 1996-97 using the revised questionnaire. The HIES was merged with Pakistan Integrated Household Survey (PIHS) in 1998-99 and the questionnaire was improved further and was split into two modules in order to obtain better quality of information separately from male and female respondents by the male and female enumerators respectively. The last round of HIES was conducted in 2004-05 as a sub sample of Pakistan Social Living Measurement (PSLM) 2004-05 district level survey, covering 14708 households taken as sub sample of the 77,000 households of PSLM survey.

2 The survey is spread over a year commencing in the first quarter from July and completed by next June.

3 The weights assigned from lowest 13.75 to the largest 8707.44.

coefficients of n_2 is larger than the coefficient of n_1 involving that the children of the age bracket *10-15* years have heavier food requirement than the children age bracket *0-10* years, as we would expect. The model also yields an *F-test* of 1027.08, which is significant at *5* per cent, indicating that the explanatory variables are jointly significant alongwith *t-ratios*.

The basic Engel scales are computed for twenty household types, corresponding to $n_1 = 0, 1,2,3,4$ and $n_2 = 0,1,2,3$ with reference to food basket (see *Appendix-A*). The detail Engel equivalence scales and the costs of children for the two adult household types are included by *table II.* It appears that the values of Engel scales are setting in the range of 1.00 and 3.20. The Engel equivalence scales of two adult and one child, two, three and four children of the age brackets $n_{(0\text{-}10)}$ to the childless couples are 1.165, 1.357, 1.581 and 1.842; while the costs of the household types are 16.50%, 35.70%, 58.10% and 84.20%, respectively.

Table II: Engel Equivalence Scales and Costs of Children for Pakistan: 2005-06
Reference Household: Childless two-adult household = 1.00

Household types (n1, n2)	***Engel Equivalence Scales*** $exp\{-(\beta 2n1+\beta 3n2)/\beta 1\}$	***Costs of Children (In %)***
0,0	1.00	0.00%
0,1	1.1895	18.95%
1,1	1.385	38.50%
2,1	1.614	61.40%
3,1	1.881	88.10%
4,1	2.191	119.1%
0,2	1.415	41.50%
1,2	1.648	64.80%
2,2	1.9208	92.08%
3,2	2.237	123.7%
4,2	2.607	160.7%
0,3	1.683	68.30%
1,3	1.961	96.10%
2,3	2.285	128.5%
3,3	2.662	166.2%
4,3	3.101	210.16%
1,0	1.165	16.50%
2,0	1.357	35.70%
3,0	1.581	58.10%
4,0	1.842	84.20%

Source: Calculated by authors using the HIES 2005-06 data set of Pakistan.

The Scales of two adult and one child, two and three children of the age brackets $n_{(10\text{-}15)}$ are 1.1895, 1.415 and 1.683; while the costs are 18.95%, 41.50% and 68.30% to the childless couples , respectively. The cost of a little child $n_{(0\text{-}10)}$ plus one big child $n_{(10\text{-}15)}$, two little plus two big child, three little plus one big child and one little plus three big child is equal to 38.50%, 92.08%, 88.10% and 96.10% of a childless couple's cost. The table of equivalence scales and cost of children also implies that the cost of big children is larger than that of little children household types. The household expenditure on food items increases alongwith adding a child aged 0 -10 or a child aged 10 -15 considerably.

4. CONCLUSION AND POLICY IMPLICATION

The empirical analysis of Engel Equivalence scales and the cost of children reported here have yielded broadly satisfactory results both in terms of economic theory and statistical fit. The result of the regression implies that household demographic compositions bring significant changes in the consumption of the food items and the proportion of household expenditure on food item is increased alongwith the household demographic compositions by age. The Engel model of food share is also suggested an inverse relation between income (total expenditure) of household and food share, *ceteris paribus.* In addition, the coefficients of children of the age (10-15) is greater than the coefficient of children of the age (0-10) including that the former is heavier food requirements than the second category of children of two adult household types. In addition, the welfare of the household type of the children aged 10-15 is also significantly affected.

REFERENCES

1. Aly Ercelawn (1991). *Undernourishment as poverty in Pakistan: Are there regional differences in the risk of hunger*? Applied Economics Research Centre, University of Karachi. Discussion Paper Number 148, 01-37.
2. Beveridge (1942). *Social Insurance and Allied Services*. (London: HMSO).
3. Gianni A. Betti (2000). *Quadratic Engel curves and household equivalence scales: the case of Italy 198501994*. Discussion Paper. Department of statistics. Landon School of Economics, September 15, 2000.
4. Julie A. Nelson (1993). Household equivalence scales: theory versus policy. *Journal of Labor Economics*, 11(3), 471-493.
5. Rowntree (1901). Poverty: A study of town life. New York: Macmillian; W.H.
6. Rueben C. Buse; Larry E. Salathe (1978). Adult Equivalent Scales: An Alternative Approach. *Amer. J. of Agri. Econ.*, 60(3), 460-468.
7. Tran Nambina and Peter Withford (1991). Household equivalence scales: New Australian estimates from the 1984 household expenditure survey. *The Economic Record*, 221-233.
8. Tsakloglou, P. (1991). Estimation and comparison of two simple models of equivalence scales for the cost of children. *Economic Journal*. 101, 343-57.
9. Van Hoa, T. (1986). Measuring Equivalence Scales: A New System – wide Method. *Economic Letters*, 20, 95-99.
10. White, H. and Masset, E. (2007). Child Poverty in Vietnam: Using Adult Equivalence Scales to Estimate Income - Poverty for Different Age Groups. *MPRA Paper* No. 777, 01-27.

APPENDIX-A

Table A: ***nc_{010} * nc_{1015} Cross Tabulation***

		nc_{1015}				Total
		.00	1.00	2.00	3.00	
nc_{010}	.00	555	108	68	17	748
	1.00	380	96	108	40	624
	2.00	524	154	153	45	876
	3.00	502	203	134	38	877
	4.00	348	196	102	31	677
	Total	2309	757	565	171	3802

Source: Computed by authors using the HIES 2005-06 data set of Pakistan.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 87-90

COMPUTER IN DEFENSE: A SURVEY

Irshad Ullah
Government Higher Secondary School, OUCH Dir (Lower), Pakistan
Email: irshadullah79@gmail.com

ABSTRACT

Computers have given the modern military a degree of efficiency and effectiveness that has completely changed the way the military operates at sea, in the air, and on the ground. The fast computational abilities and the accuracy of digital computing enable the military to perform more quickly with less risk to themselves and with less cost to their government.

In this work, it will be studied that how useful is the use of computer in the field of defense for the environment as well as for human.

KEY WORDS

Sea; Computational; Military; Defense; Human.

1. INTRODUCTION

The military funding of science has had a powerful transformative effect on the practice and products of scientific research since the early 20th century. Particularly since World War I, advanced science-based technologies have been viewed as essential elements of a successful military. While there were numerous instances of military support for scientific work before the 20th century, these were typically isolated instances; knowledge gained from technology was generally far more important for the development of science than scientific knowledge was to technological innovation (Hacker, Barton C. (2005)).

World War I is often called "the chemists' war", both for the extensive use of poison gas and the importance of nitrates and advanced high explosives. Poison gas, beginning in 1915 with chlorine from the powerful German dye industry, was used extensively by the Germans and the British; over the course of the war, scientists on both sides raced to develop more and more potent chemicals and devise countermeasures against the newest enemy gases. Harris, Robert and Jeremy Paxman (2002)).

Physicists also contributed to the war effort, developing wireless communication technologies and sound-based methods of detecting U-boats, resulting in the first tenuous long-term connections between academic science and the military. (Kevles, Daniel J (1971)).

World War II marked a massive increase in the military funding of science, particularly physics. In addition to the Manhattan Project and the resulting atomic bomb, British and American work on radar was widespread and ultimately highly influential in the course of the war; radar enabled detection of enemy ships and aircraft, as well as the radar-based proximity fuze. Mathematical cryptography, meteorology, and rocket science were also central to the war effort, with military-funded wartime advances having a significant long-term effect on each discipline. The technologies employed at the end—jet aircraft, radar and proximity fuzes, and the atomic bomb—were radically different from pre-war technology; military leaders came to view continued advances in technology as the critical element for success in future wars. The advent of the Cold War solidified the links between military institutions and academic science, particularly in the United States and the Soviet Union, so that even during a period of nominal peace military funding continued to expand. Funding spread to the social sciences as well as the natural sciences, and whole new fields, such as digital computing, were born of military patronage. Following the end of the Cold War and the collapse of the Soviet Union, military funding of science has decreased substantially, but much of the American military-scientific complex remains in place.

The sheer scale of military funding for science since World War II has instigated a large body of historical literature analyzing the effects of that funding, especially for American science. Since Paul Forman's 1987 article "Behind quantum electronics: National security as a basis for physical research in the United State, 1940-1960," there has been an ongoing historical debate over precisely how and to what extent military funding affected the course of scientific research and discovery. (Forman, Paul (1940-1960)) Forman and others have argued that military funding fundamentally redirected science—particularly physics—toward applied research, and that military technologies predominantly formed the basis for subsequent research even in areas of basic science; ultimately the very culture and ideals of science were colored by extensive collaboration between scientists and military planners. An alternate view has been presented by Daniel Kevles, that while military funding provided many new opportunities for scientists and dramatically expanded the scope of physical research, scientists by-and-large retained their intellectual autonomy.

World War I marked the first large-scale mobilization of science for military purposes. Prior to the war, the American military ran a few small laboratories as well as the Bureau of Standards, but independent inventors and industrial firms predominated. (Ret^ Kevles, URL (2010)) Similarly in Europe, Harris and Paxman URL (2010)) military-directed scientific research and development was minimal. The powerful new technologies that led to trench warfare, however, reversed the traditional advantage of fast-moving offensive tactics; fortified positions supported by machine guns and artillery resulted in high attrition but strategic stalemate. Militaries turned to scientists and engineers for even newer technologies, but the introduction of tanks and aircraft had only a marginal impact; the use of poison gas made a tremendous psychological impact, but decisively favored neither side. The war ultimately turned on maintaining adequate supplies of materials, a problem also addressed by military-funded science—and, through the international chemical industry, closely related to the advent of chemical warfare.

War I was the chemists' war, World War II was the physicists' war. As with other total wars, it is difficult to draw a line between military funding and more spontaneous military-scientific collaboration during World War II. Well before the Invasion of Poland, nationalism was a powerful force in the German physics community; the military mobilization of physicists was all but irresistible after the rise of National Socialism. German and Allied investigations of the possibility of a nuclear bomb began in 1939 at the initiative of civilian scientists, but by 1942 the respective militaries were heavily involved. The German nuclear energy project had two independent teams, a civilian-controlled team under Werner Heisenberg and a military-controlled led by Kurt Diebner; the latter was more explicitly aimed at producing a bomb (as opposed to a power reactor) and received much more funding from the Nazis, though neither was ultimately successful. (A Study in German Culture. URL (2010)).

Wars fought prior to the advent of digital computing could take years to complete and resulted in tremendous loss of property on both sides. Weapons of mass destruction used to be the fastest way to achieve the objective. A discussion with Dan Carroll, Vietnam War veteran and current contract engineer for major aerospace companies revealed that in World War II and in Vietnam, American forces employed a technique known as carpet bombing to wipe out enemy troops and resources. Many large bombers would over fly an area and drop thousands of bombs in an effort to destroy the enemy. This proved to be very costly for both sides in terms of property and human life.

Assuming that war is inevitable and that enemies must sometimes be destroyed, small, powerful computers are useful. Weapons can now be made intelligent enough to know precisely where and what the target is. It is now possible to launch a projectile from hundreds of miles away and destroy one particular building in an enemy installation. This sort of precision warfare is much less expensive in the long run and makes much more efficient use of resources. When human beings must be killed, we have the ability to spare the vast majority of people in a city who are innocent bystanders and target only those directly involved in the war effort.

CONCLUSION AND FUTURE WORK

Assuming that war is inevitable and that enemies must sometimes be destroyed, small, influential computers are useful. Weapons can now be made intelligent sufficient to know precisely where and what the target is. But steps must be taken to prevent the world from wars. Because it's not only destroy the humans but also it's destroy the environment. From fast we have an example when the last prophet of Allah conquer Makkah n casualty is reputed in that war even the who are responsible for the killing of his nearest family members are forgave. Because war is not for power it's for peace and for the stability and maintenance of peace.

REFERENCES

1. Forman, Paul (1987). Behind quantum electronics: National Security as basis for physical research in the United States, 1940-1960. *Historical Studies in the Physical and Biological Sciences*, 18(1), 149-229.
2. Hacker, Barton C. (2005). The Machines of War: Western Military Technology 1850-2000. *History and Technology*, 21(3), 255-300.
3. Harris and Paxman, *A Higher Form of Killing*, pp 11-12 http://en.wikipedia.org/wiki/History_of_military_technology
4. Harris, Robert and Jeremy Paxman (2002). *A Higher Form of Killing: The Secret History of Chemical and Biological Warfare*. Chapter 1.
5. Kevles, Daniel J. (1971). *The Physicists: The History of a Scientific Community in Modern America*. New York: Alfred K. Knopf, 137-138.
6. Ret^ Kevles, the Physicists, pp 103-104 Retrieved (29.12.2010)
7. ^The extent to which Heisenberg's team was devoted to aiding the Nazi's by producing an atomic bomb is a matter of some historical dispute. However, most recent scholarship suggests that the stagnation of the German project stemmed from Heisenberg's doubts about the feasibility rather than the desirability of a Nazi bomb. See: Rose, Paul Lawrence. Heisenberg and the Nazi Atomic Bomb Project: A Study in German Culture. Berkeley: University of California Press, 1998. Retrieved (29.12.2010) http://en.wikipedia.org/wiki/History_of_military_technology

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 91-94

DATA ANALYSIS IN THE FIELD OF ENGINEERING

Irshad Ullah
Government Higher Secondary School, OUCH Dir (Lower), Pakistan
Email: irshadullah79@gmail.com

ABSTRACT

Use of computer in the field of engineering is very vast. Even there are many branches of computer Software tools that have been developed to support these activities are considered CAE tools. CAE tools are being used, for example, to analyze the strength and performance of components and assemblies. The term encompasses simulation, justification, and optimization of products and manufacturing tools. In the future, CAE systems will be major providers of information to help support design teams in decision making. In regard to information networks, CAE systems are individually considered a single node on a total information network and each node may interrelate with other nodes on the network. CAE systems can provide support to businesses. This is achieved by the use of reference architectures and their ability to place information views on the business process. Reference architecture is the basis from which information model, especially product and manufacturing models.

In this work, Data Mining techniques are used for the data analysis, and knowledge discovery procedure to show experimentally that how consistent, able and fast are these techniques for the study in the particular field? A solid mathematical threshold (0 to 1) is set to analyze the data.

KEY WORDS

CAE; Branch; Network; Consistent.

1. INTRODUCTION

Data Analysis

Data analysis is a process in which raw data is prepared and structured so that valuable information can be extracted from it. The process of organizing and thinking about data is way to accepting what the data does and does not contain (Irshad Ullah (2010)). Data Mining (Jiawe Han and Micheline Kamber (2001)) is a process of extracting knowledge from databases or data warehouses. Data mining, also known as knowledge discovery in databases, is a non-trivial extraction of implicit, previously unidentified and potentially useful information from data. ‘Data Mining’ and ‘Knowledge Discovery’ have been used interchangeably.

Data mining presents new perspectives for data analysis. The purpose of data mining is to mine and discover new knowledge from data. Over the past few decades, new methods have been developed about the capabilities of data collection and data creation. Data collection tools have provided us with a huge amount of data. Data mining processes have integrated techniques from multiple disciplines such as, statistics, machine learning, database technology, pattern recognition, neural networks, information retrieval and spatial data analysis. Data mining techniques have been used in many different fields such as, business management, science, engineering, banking, data management, administration and many other applications. (Jiawe Han and Micheline Kamber (2001)

The AIS algorithm was the first published algorithm developed to generate all large itemsets in a transaction database. It focused on the enhancement of databases with necessary functionality to process decision support queries. This algorithm was targeted to discover qualitative rules. This technique is limited to only one item in the consequent. The AIS algorithm makes multiple passes over the entire database. During each pass, it scans all transactions. In the first pass, it counts the support of individual items and determines which of them are large or frequent in the database. Large itemsets of each pass are extended to generate candidate itemsets (Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami, (1993)).

Similar to the AIS algorithm, the SETM algorithm makes multiple passes over the database. In the first pass, it counts the support of individual items and determines which of them are large or frequent in the database. Then, it generates the candidate itemsets by extending large itemsets of the previous pass. In addition, the SETM remembers the TIDs of the generating transactions with the candidate itemsets. The relational merge-join operation can be used to generate candidate itemsets (M. Houtsma and A. Swami (1995)).

Apriori scans the entire database in each pass to count support. Scanning of the entire database may not be needed in all passes. Based on this conjecture, (M. Houtsma and A. Swami (1995)) proposed another algorithm called Apriori-TID. Similar to Apriori, Apriori-TID uses the

Apriori's candidate generating function to determine candidate itemsets before the beginning of a pass. The main difference from Apriori is that it does not use the database for counting support after the first pass. Rather, it uses an encoding of the candidate itemsets used in the previous pass denoted by k C. The advantage of using this encoding function is that in later passes the size of the encoding function becomes smaller than the database, thus saving much reading effort.

2. ASSOCIATION RULE PROBLEM

Association rule mining helps in finding interesting association relationships among large set of data items. The discovery of such associations can help develop strategies to predict. Association rule mining explores for interesting relationships among items in a given data set. Association analysis is used for market basket or transaction data analysis. Market basket process finds associations between the different items. An objective of association rule mining is to develop a systematic method using the given data set and finds relationships between the different items. Association is a rule, which implies certain association relationships among set of objects such as occur together or one implies the other (Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami, (1993))

2.1 Formal Problem Description

The following is the formal problem definition. Let $I = \{i_1,\ i_2, \ldots, i_n\}$ be a set of literals called items. Let D be a set of transaction, where each transaction is a set of item such that $T \subseteq I$. Associated with each transaction is a unique identifier called TID. We say that a transaction contain T contain X, a set of some item in I, if $X \subseteq T$. An association rule is an implication of the form $X \Rightarrow Y$ where $X \subset I$ and $Y \subset I$, and $X \cap Y = \varnothing$. The rule $X \Rightarrow Y$ hold in the set D with confidence c if c% of transaction in D that contain X also contain Y. The rule $X \Rightarrow Y$ has confidence S in the transaction set D if s% of transaction in D contain XUY (Rakesh Agrawal and Ramakrishnan Srikant (1994)).

3. RESULTS

The algorithm requires the data in the database to be in binary format. So the dataset transa has been downloaded from the net. The data was stored in a text file on the site (URL(200)). The data was in a format:

0 1 0 0 1 0 0 1

first convert the data into a format that the item now is separated by commas instead of spaces. Now the data is loaded to the table with the help of sql loader and look like 0,1,0,0,1,0,0,1

After loading the data into table the algorithms are implemented on the database having ten thousand records initially.

Figure 1

SI algorithm Results (Salam, Irshad)		
Dissimilarity	**Value**	**Frequent List**
.71	.80	I1,I2
.38	.80	I1,I3
.33	.80	I1,I4
.73	.80	I1,I5
.64	.80	I2,I3
71	.80	I2,I4
79	.80	I2,I5
.5	.80	I3,I4
.75	.80	I1,I2,I3
.53	.80	I1,I3,I4

Figure 2

Appriori Algorithm Results		
Min Support	**Support count**	**Frequent item List**
20	4838	I1
20	3868	I2
20	5332	I3
20	4848	I4
20	4344	I5
20	1938	I1,I2
20	3878	I1,I3
20	3878	I1,I4
20	2423	I2,I3
20	1938	I2,I4
20	3393	I3,I4
20	2424	I4,I5
20	1938	I1,I2,I3
20	3393	I1,I3,I4

The largest frequent list generated by the algorithms are same I1, I2, I3 and I1, I3, I4. The experiments are performed by changing size of the database. But the results produced were the same (The largest frequent list). It means that to apply such techniques to the field of engineering a useful information and hidden pattern be recognized which may be used in future decision making for different purpose e.g. Course analysis, Subject analysis etc.

4. CONCLUSION AND FUTURE WORK

In this research an attempt has been made to apply data mining algorithm to the field of engineering for the generation of meaningful information to be used by the experts of the field for different purposes. In future these algorithms may be applied to other data storages and may be for the clustering purpose in the said field.

REFERENCES

1. http://www2.cs.uregina.ca/~dbd/cs831/notes/itemsets/itemset_prog1.html.
2. Irshad Ullah (2010). Data Analysis by Data Mining Algorithms A Practical Approach. Proceeding of *Two Day International Conference on Word Statistics Day: Statistics for Society* organized by ISOSS, Lahore, Pakistan.
3. Jiawe Han and Micheline Kamber (2001). *Data mining concepts and techniques*, San Franciso Moraga Kaufman.
4. Houtsma, M. and Swami, A. (1995). Set-Oriented Mining for Association Rules in Relational Databases, Proceedings of the *11th IEEE International Conference on Data Engineering*, 25-34, Taipei, Taiwan.
5. Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami (1993). Mining Association Rules Between Sets of Items in Large Databases, Proceedings of the *1993 ACM SIGMOD I international Conference on Management of Data*, 207- 216, Washington DC.
6. Rakesh Agrawal and Ramakrishnan Srikant (1994). Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the *Twentieth International Conference on Very Large Databases*, 487-499, Santiago, Chile.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 95-98

COMPUTER SCIENCE AND POPULATION STUDY

Irshad Ullah
Government Higher Secondary School, OUCH Dir (Lower), Pakistan.
Email: irshadullah79@gmail.com

ABSTRACT

Computer is a programmable machine. This means that it can execute a programmed list of instructions and respond to new instructions those are given. A computer is a machine that receives input, stores and manipulates data and provides output in a useful format.

Here its will be studied that's how computer helps in population study. Some practical will be performed for the efficiency purpose. Experiments to be performed will base on some solid values. The results produces may be used for the analysis and decision purpose.

KEY WORDS

Population; Response; List; Program.

1. INTRODUCTION

Data mining presents new perspectives for data analysis. The purpose of data mining is to extract and discover new knowledge from data. Over the past few decades, new methods have been developed about the capabilities of data collection and data generation. Data collection tools have provided us with a huge amount of data. Data mining processes have integrated techniques from multiple disciplines such as, statistics, machine learning, database technology, pattern recognition, neural networks, information retrieval and spatial data analysis. Data mining techniques have been used in many different fields such as, business management, science, engineering, banking, data management, administration and many other applications. (Jiawe Han and Micheline Kamber (2001)) Data Mining techniques may be used in medical Science.

Understanding and definition of the problem is the first step in data mining process. Once a problem has been defined, relevant data must be collected. The relevant data is extracted from an existing database or data warehouse. (Jiawe Han and Micheline Kamber (2001)).

2. FORMAL PROBLEM DESCRIPTION

The following is the formal problem definition. Let $I = \{i_1,\ i_2, \ldots, i_n\}$ be a set of literals called items. Let D be a set of transaction, where each transaction is a set of item such that $T \subseteq I$. Associated with each transaction is a unique identifier called TID. We say that a transaction contain T contain X, a set of some item in I, if $X \subseteq T$. An association rule is an implication of the form $X \Rightarrow Y$ where $X \subset I$ and $Y \subset I$, and $X \cap Y = \varnothing$. The rule $X \Rightarrow Y$ hold in the set D with confidence c if c% of transaction in D that contain X also contain Y. The rule $X \Rightarrow Y$ has confidence S in the transaction set D if s% of transaction in D contain XUY ([3]Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami, (1993)).

Association rules can be classified based on the *type of vales*, *dimensions of data*, and *levels of abstractions* involved in the rule. If a rule concerns associations between the presence or absence of items, it is called Boolean association rule. And the dataset consisting of attributes which can assume only binary (0-absent, 1-present) values is called Boolean database.

3. RESULTS

To perform experiments to check the results and accuracy of the algorithm. The algorithm requires the data in the database to be in binary format. The dataset is download from the net. The data was stored in a text file on the site (URL (2008)). The data was in a format:

0 1 0 0 1 0 0 1

1 1 0 0 1 1 0 0

We first convert the data into a format that the item now is separated by commas instead of spaces. Now the data is loaded to the table with the help of sql loader and look like

0,1,0,0,1,0,0,1

1,1,0,0,1,1,0,0

After loading the data into table the algorithms are implemented on the database having ten thousand records.

Figure 1

SI algorithm Results (Salam, Irshad)		
Dissimilarity	**Value**	**Frequent List**
.71	.80	I1,I2
.38	.80	I1,I3
.33	.80	I1,I4
.73	.80	I1,I5
.64	.80	I2,I3
71	.80	I2,I4
79	.80	I2,I5
.5	.80	I3,I4
.75	.80	I1,I2,I3
.53	.80	I1,I3,I4

Figure 2

Appriori Algorithm Results		
Min Support	**Support count**	**Frequent item List**
20	4838	I1
20	3868	I2
20	5332	I3
20	4848	I4
20	4344	I5
20	1938	I1,I2
20	3878	I1,I3
20	3878	I1,I4
20	2423	I2,I3
20	1938	I2,I4
20	3393	I3,I4
20	2424	I4,I5
20	1938	I1,I2,I3
20	3393	I1,I3,I4

From the results its clear that the algorithms generate the same results with largest frequent list I1,I2,I3 and I1,I3,I4. So these algorithms may be used for the analysis of population to find different pattern in different areas. The knowledge may be use for the welfare of the population of either kingdom.

4. CONCLUSION AND FUTURE WORK

In this research an experiment is performed to collect meaning full data about the population and then on the basis of the results the experts may suggest and take some steps for its welfare. In future experiments may be performed for large data volumes

REFERENCES

1. http://www2.cs.uregina.ca/~dbd/cs831/notes/itemsets/itemset_prog1.html.
2. Jiawe Han and Micheline Kamber (2001). *Data mining concepts and techniques San Francisco Moraga Kaufman*.
3. Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami, (1993). Mining Association Rules Between Sets of Items in Large Databases, Proceedings of the *1993 ACM SIGMOD I international Conference on Management of Data*, 207-216, Washington, DC.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 99-102

DATA MINING AND STUDENTS ANALYSIS

Irshad Ullah
Government Higher Secondary School, OUCH Dir (Lower), Pakistan
Email: irshadullah79@gmail.com

ABSTRACT

Data Mining is a process of extracting information from databases or data warehouses. Data mining, also known as knowledge discovery in databases, is a non-trivial extraction of hidden, previously unknown and potentially useful information from data. 'Data Mining' and 'Knowledge Discovery' have been used interchangeably.

In this paper data Mining Algorithms will be use for the student's databases. To analyze students behavior and characteristics practical implementations will be made for the future analysis purpose.

KEY WORDS

Students; Discovery; Database; Non-Trivial.

1. INTRODUCTION

Data mining is a repetitive process consisting of several steps. Starting with the understanding and definition of a problem and ending with the analysis of results and determine a strategy with using the result (Jiawe Han and Micheline Kamber(2001)) Association rule mining is one of the most researched areas of data mining and has recently received more attention from the database side. It has proven to be quite useful in the marketing and retail communities as well as other more diverse fields.

AIS Algorithm, first introduced in 1993 (Ramakrishnan Srikant (1996)) is used to identify relationships among a set of items in a database. These relationships are not based on inherent properties of the data themselves (as with functional dependencies), but rather based on co-occurrence of the data items.

This algorithm was targeted to discover qualitative rules. This technique is limited to only one item in the consequent. The AIS algorithm makes multiple passes over the entire database. During each pass, it scans all transactions. In the first pass, it counts the support of individual items and determines which of them are large or frequent in the database.

The SETM algorithm was proposed in (Ashoka Savasere, Edward Omiecinski, and Shamkant B. Navathe (1995)) and was motivated by the desire to use SQL to calculate large itemsets (Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami (1993)).

Similar to the AIS algorithm, the SETM algorithm makes multiple passes over the database. However, (David Wai-Lok Cheung, Vincent T. Ng, Ada Wai-Chee Fu, and Yongjian Fu (1996)) mentioned that SETM is not efficient and there are no results reported on running it against a relational DBMS. The Apriori algorithm developed by rules (Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami (1993)) is a great achievement in the history of mining association. It is by far the most well-known association rule algorithm. This technique uses the property that any subset of a large itemset must be a large itemset. Also, it is assumed that items within an itemset are kept in lexicographic order. The fundamental differences of this algorithm from the AIS and SETM algorithm are the way of generating candidate itemsets and the selection of candidate itemsets for counting.

Partition (Ashoka Savasere, Edward Omiecinski, and Shamkant B. Navathe (1995)) reduces the number of database scans to 2. It divides the database into small partitions such that each partition can be handled in the main memory.

Goal of association rule is finding associations among items from a set of transactions which contain a set of items. Main problem of Association Rule inductions is that there are so many possible rules. A typical association rule mining algorithm has the following components:

- Structure (used for implementation of the pattern)
- Score function (simply binary function, support and confidence)
- Search method (systematic search methods)
- Data management technique (number of linear scan through the database)

2. FORMAL PROBLEM DESCRIPTION

The following is the formal problem definition. Let $I = \{i_1,\ i_2, \ldots, i_n\}$ be a set of literals called items. Let D be a set of transaction, where each transaction is a set of item such that $T \subseteq I$. Associated with each transaction is a unique identifier called TID. We say that a transaction contain T contain X, a set of some item in I, if $X \subseteq T$. An association rule is an implication of the form $X \Rightarrow Y$ where $X \subset I$ and $Y \subset I$, and $X \cap Y = \varnothing$. The rule $X \Rightarrow Y$ hold in the set D with confidence c if c% of transaction in D that contain X also contain Y. The rule $X \Rightarrow Y$ has confidence S in the transaction set D if s% of transaction in D contain XUY Rakesh Agrawal and Ramakrishnan Srikant (1994) Association rules can be classified based on the *type of vales*, *dimensions of data*, and *levels of abstractions* involved in the rule. If a rule concerns associations between the presence or absence of items, it is called Boolean association rule. And the dataset consisting of attributes which can assume only binary (0-absent, 1-present) values is called Boolean database.

3. RESULTS

To perform experiments to check the results and accuracy of the algorithm. The algorithm requires the data in the database to be in binary format. The dataset is downloaded from the net. The data was stored in a text file on the site (URL (2010)). The data was in a format:

0 1 0 0 1 0 0 1

First the data is converted into a format that the item now is separated by commas instead of spaces. Now the data is loaded to the table with the help of sql loader and look like

0,1,0,0,1,0,0,1

After loading the data into table the algorithms are implemented on the database having twenty thousand records.

Figure 1

SI algorithm Results (Salam, Irshad)		
Dissimilarity	**Value**	**Frequent List**
.71	.80	I1,I2
.38	.80	I1,I3
.33	.80	I1,I4
.73	.80	I1,I5
.64	.80	I2,I3
71	.80	I2,I4
79	.80	I2,I5
.5	.80	I3,I4
.75	.80	I1,I2,I3
.53	.80	I1,I3,I4

Figure 2

Appriori Algorithm Results		
Min Support	**Support count**	**Frequent item List**
20	4838	I1
20	3868	I2
20	5332	I3
20	4848	I4
20	4344	I5
20	1938	I1,I2
20	3878	I1,I3
20	3878	I1,I4
20	2423	I2,I3
20	1938	I2,I4
20	3393	I3,I4
20	2424	I4,I5
20	1938	I1,I2,I3
20	3393	I1,I3,I4

By changing size of the database both algorithms generate same results. With largest frequent list I1,I2,I3 and I1,I3,I4.

Hence this is clear that results are reliable and accurate. By this meaning full data about the students of any organization can be collected and evaluated.

4. CONCLUSION AND FUTURE WORK

This research provide a methodology to collect and evaluate students data and converted it into information for the analysis purpose. The outcome may be helpful an enhancing students performance. In future this research may be extended to perform more experiments and arrange the results from different point of view.

REFERENCE

1. Ashoka Savasere, Edward Omiecinski, and Shamkant B. Navathe (1995). An Efficient Algorithm for Mining Association Rules in Large Databases. Proceedings of the *21st International Conference on Very Large Databases*, 432-444, Zurich, Switzerland.
2. David Wai-Lok Cheung, Vincent T. Ng, Ada Wai-Chee Fu, and Yongjian Fu (1996). Efficient Mining of Association Rules in Distributed Databases. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 911-922.
3. http://www2.cs.uregina.ca/~dbd/cs831/notes/itemsets/itemse t_prog1.html.
4. Jiawe Han and Micheline Kamber (2001). *Data mining concepts and techniques San Francisco Moraga Kaufman*.
5. M. Houtsma and A. Swami (1995). Set-Oriented Mining for Association Rules in Relational Databases. Proceedings of the *11th IEEE International Conference on Data Engineering*, 25-34, Taipei, Taiwan.
6. Rakesh Agrawal and Ramakrishnan Srikant (1994). Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of the *Twentieth International Conference on Very Large Databases*, 487-499, Santiago, Chile.
7. Ramakrishnan Srikant (1996). Fast Algorithms for Mining Association Rules and Sequential Patterns, Ph.D. Dissertation, University of Wisconsin, Madison.
8. Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami (1993). Mining Association Rules between Sets of Items in Large Databases. Proceedings of the *1993 ACM SIGMOD I international Conference on Management of Data*, 207-216, Washington DC.
9. Sarawagi Sunita, Thomas Shiby, and Agrawal Rakesh (1998). Integrating Mining with Relational Database Systems: Alternatives and Implications. Proceedings *ACM SIGMOD International Conference on Management of Data*, SIGMOD, Seattle, Washington.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 103-105

LOGICAL REASONING AND DATA MINING ALGORITHMS

Irshad Ullah
Government Higher Secondary School, OUCH Dir (Lower), Pakistan
Email: irshadullah79@gmail.com

ABSTRACT

The term 'logic' is used quite a lot, but not at all times in its technical sense. Logic, strictly speaking, is the science or study of how to assess arguments and reasoning. Logic is what allows us to distinguish correct reasoning from poor reasoning. Logic is important because it helps us reason correctly - without correct reasoning, we don't have a viable means for knowing the truth or arriving at sound beliefs.

Here its will be studied that how algorithms from the field of computer science is helpful in making the correct reasoning or decision. It will be proved practically that such algorithms help in making and assessing the reasoning and decision.

KEY WORDS

Logic; Reason; Technical; Decision; Argument.

1. INTRODUCTION

What is an argument? What are logical fallacies? In order to debate well and think critically, we need the proper skills. Whether reading advertising, listening to politicians, buying a used car, or considering arguments for religion, we need to know how to carefully examine what is being said and consider the validity of the content or arrangement. Logical analysis is a system that we use to make an argument from observation and known facts.

We can use logical reasoning to make assumptions, through known data. A very popular phrase that depicts logic is this: Every man is mortal, Socrates is a man, and therefore Socrates is mortal. Logic is obviously used everywhere, at work, when we are learning a new language, etc. etc. Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information–information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships recognized. Technically, data mining is the process of finding correlations or patterns among thousands of fields in large relational databases.

Data mining (Jiawe Han and Micheline Kamber (2001)) presents new perspectives for data analysis. The purpose of data mining is to extract and discover new knowledge from data. Over the past few decades, new methods have been developed about the capabilities of data collection and data generation.

AIS Algorithm, first introduced in 1993 (Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami (1993)) is used to identify relationships among a set of items in a database.

The SETM algorithm was proposed in (M. Houtsma and A. Swami (1995)) and was motivated by the desire to use SQL to calculate large itemsets (Ramakrishnan Srikant (1996)). Apriori scans the entire database in each pass to count support. Scanning of the entire database may not be needed in all passes. Based on this conjecture, (Rakesh Agrawal and Ramakrishnan Srikant (1994)) proposed another algorithm called Apriori-TID. Similar to Apriori, Apriori-TID uses the Apriori's candidate generating function to determine candidate itemsets before the beginning of a pass.

2. PROBLEM DESCRIPTION

The following is the formal problem definition. Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of literals called items. Let D be a set of transaction, where each transaction is a set of item such that. Associated with each transaction is a unique identifier called TID. We say that a transaction contain T contain X, a set of some item in I, if. An association rule is an implication of the form where and $X \cap Y =$. The rule hold in the set D with confidence c if c% of transaction in D that contain X also contain Y. The rule has confidence S in the transaction set D if s% of transaction in D contain XUY (Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami (1993)).

3. RESULTS

To perform experiments dataset is downloaded from the net. The data is converted into the required format. Now the data is loaded into a table with the help of SQL loader. Two algorithms were used for the purpose. Initially ten thousands records were loaded in table. After performing different experiments the largest frequent list generated by the algorithms in all cases.

Frequency	List
1938	I1, I2, I3
3393	I1, I3, I4

On the basis of these results the experts may make decisions. Because the database on which the experiments performed were in binary format. And a well known jaccqard dissimilarity quotient for the non-invariant similarities was used. Now on the basis of such inferences data may be judged and also reasoning from different experts may be generated. And the performance may be more enhanced and thinking capability may be improved.

4. CONCLUSION AND FUTURE WORK

Here logical data have been analyzed to improve thinking and reasoning capability. In future more experiments may be performed with some other formula for the efficiency purpose.

REFERENCES

1. Jiawe Han and Micheline Kamber (2001). *Data mining concepts and techniques San Francisco Moraga Kaufman*.
2. M. Houtsma and A. Swami (1995). Set-Oriented Mining for Association Rules in Relational Databases. Proceedings of the *11th IEEE International Conference on Data Engineering*, 25-34, Taipei, Taiwan.
3. Rakesh Agrawal and Ramakrishnan Srikant (1994). Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of the *Twentieth International Conference on Very Large Databases*, 487-499, Santiago, Chile.
4. Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami (1993). Mining Association Rules between Sets of Items in Large Databases. Proceedings of the *1993 ACM SIGMOD I international Conference on Management of Data*, 207-216, Washington DC.
5. Ramakrishnan Srikant (1996). *Fast Algorithms for Mining Association Rules and Sequential Patterns*, Ph.D. Dissertation, University of Wisconsin, Madison.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 107-112

PERMANENT INCOME HYPOTHESIS, MYOPIA AND LIQUIDITY CONSTRAINTS: A CASE STUDY OF PAKISTAN

Khalid Khan[1], **Hazrat Yousaf**[1] and **Mohammed Nishat**[2]

[1,2] Department of Economics, Lasbela University of Agriculture, Water & Marine Sciences, Uthal, Balochistan, Pakistan.
Email: khalidkk82@yahoo.com; hazrataerc@yahoo.com

[3] Department of Finance and Economics, Institute of Business Administration, Karachi, Pakistan. Email: mnishat@iba.edu.pk

ABSTRACT

In this paper, it is attempted to test the permanent income hypothesis (PIH) for Pakistan. The results indicate the strong validity of absolute income hypothesis (AIH) rather than PIH. Besides this, the study attempts to find out the reasons of the rejection of the PIH in Pakistan, for this purpose the study used the Shea (1995) model. The results of the Shea (1995) model rejected the symmetric relationship between consumption and expected income and provide a little evidence of existence of liquidity constraints.

KEYWORDS

Permanent Income Hypothesis; Absolute Income Hypothesis; Symmetric Relationship; Liquidity Constraints.

1. INTRODUCTION

Campbell and Mankiw (1990) showed that in an economy consumers can satisfy both absolute income hypothesis (AIH) and permanent income hypothesis (PIH) simultaneously. Campbell and Mankiw divided consumers in two different segments i.e. 'λ' and $(1-\lambda)$, λ shows the proportion of backward looking consumers, while $(1-\lambda)$ shows the proportion of forward looking consumers. Khalid (1994) tested the PIH for Pakistan, using distributive lag model. His results show that PIH is not valid in Pakistan but he did not show any reason for the failure of PIH. Since in consumption literature, there are two reasons for the rejection of PIH, one is liquidity constraints (Zeldes (1989)), and another is myopia (Runkle (1991)). In case of liquidity constraints consumers are unable to borrow against their future income but they can save freely when their current income is increases. Therefore, liquidity constraints show an asymmetric relationship between consumption and expected income. But myopia consumption follow current income, because in myopia, consumers response equally to predictable income increases and decreases. Therefore, myopia causes symmetric relationship between consumption and expected income.

The current study attempts are to test the PIH for Pakistan and interested to find out the reasons of the rejection of PIH in Pakistan. The rest of the paper is organized as follows. Section two offers methodology and data description, section three presents the results and discussion and section four summarizes conclusion.

2. METHODOLOGY AND DATA DESCRIPTION

Adopting the Hall (1978) that the typical forward looking consumer maximizes:

$$E_t \sum_{s=0}^{\infty} (1+\delta)^{-s} u(C_{t+s}) \qquad u' > 0, u'' < 0 \tag{1.1}$$

E_t = expectation at period t

C_t = private consumption at period t

u = utility function of individual

δ = rate of subjective time preference

If the representative individual can land and borrow at a real interest rate r, then the first order condition from the above maximization is:

$$E_t u'(C_{t+1}) = \left(\frac{1+\delta}{1+r}\right) u'(C_t) \tag{1.2}$$

If we assume that $r = \delta$ and the marginal utility (u') is linear or log linear then we finds that the current consumption is the best forecast of future consumption in the next period, i.e.:

$$C_{t+1} = C_t + \varepsilon_t \tag{1.3}$$

or

$$E_t C_{t+1} = C_t \tag{1.4}$$

Therefore,

$$\Delta C_t = \varepsilon_t \tag{1.5}$$

where ε_t is random error, the above equation (1.5) implies that all the available information is used in current period to forecast the future consumption.

2.1 Econometric Model

To test the PIH, for this purpose the study uses following model:

$$\Delta C_t = \alpha + \lambda \Delta Y_t + \varepsilon_t \quad (1.6)$$

where ΔC_t consumption growth, ΔY_t is expected income growth. When $\lambda = 0$ then PIH will be satisfied.

The Campbell and Mankiw (1990) allow the interest rate in the above equation (1.6) for intertemporal substitution for the forward looking consumer. Then the augmented model as become:

$$\Delta C_t = \mu + \lambda \Delta Y_t + \sigma . r_t + \varepsilon_t \tag{1.7}$$

where r_t real expected interest rate.

The PIH postulated that predictable changes in income have no affect on consumption but consumption is only can affected from interest rate. Therefore, if (λ) equals to zero it shows the validity of PIH. Under myopia consumers consume a fixed share of current income. Therefore, consumption should response symmetrically the increases and decreases of expected income. While, under liquidity constraints individual consumption should response more strongly to increases of expected income then decreases of expected income. This asymmetrical behavior is due to this: that under the liquidity constraints individuals cannot borrow but save freely.

Following Shea (1995) model, to test the presence of liquidity constraint and myopia, by running the following regression:

$$\Delta C_t = \mu + \lambda_1 (POS_t)\Delta Y_t + \lambda_2 (NEG)\Delta Y_t + \sigma . r_t + \varepsilon_t \qquad (1.8)$$

where POS and NEG are dummies variables for periods in which $\Delta Y_t > 0$ and $\Delta Y_t < 0$ respectively. The PIH implies that both $\lambda_1 = \lambda_2 = 0$, under myopia the $\lambda' s$ should be significant, equal and greater than zero, while in liquidity constraints implies that: λ_1 is positive, significant and greater than λ_2.

2.2 Data

The data is collected on annual basis from 1971 to 2009 at the International Financial Statistics (IFS) CD-ROM. The data consists of real disposal income, real consumption, real interest rate and consumer price index (CPI).Real consumption is calculated by adjusting nominal consumption with CPI, real disposable income is find out by deflating GDP with GDP deflator, while the real interest rate is computed as the discount rate minus the change of consumers price index (CPI).

3. EMPIRICAL RESULTS

Firstly, the study tests the PIH for Pakistan, by estimating the equations (1.6) and (1.7). Table 2 presents the empirical results of equation (1.6) and equation (1.7). These equations are estimated by using OLS and instrumental variables (IV) method. Equation (1.6) is estimated by applying the OLS method; the results of equation (1.6) show the strong validity of AIH for Pakistan. Therefore, consumption follows current income rather than permanent income. Both OLS and IV methods are used to estimate equation (1.7). Model 1, 2, 3, 4, and 5 are estimated by using the IV lists; first, second, third, fourth and fifth, offers by table 1.

Table 1: Lists of Instrumental Variables

Instrumentals variables lists which are used in Empirical estimations	
List	**Instrumentals variables**
First list	$\Delta y_{t-2},...,\Delta y_{t-6}; r_{t-2},...,r_{t-6}$
Second list	$\Delta c_{t-2},...,\Delta c_{t-6}; r_{t-2},...,r_{t-6}$
Third list	$\Delta y_{t-2},...,\Delta y_{t-6}; \Delta c_{t-2},...\Delta c_{t-6}; cy_{t-2}; r_{t-2},...,r_{t-6}$
Fourth list	$\Delta i_{t-2},...,\Delta i_{t-6}; r_{t-2},...,r_{t-6}$
Fifth list	$\Delta i_{t-2},...\Delta i_{t-6}; \Delta y_{t-2},...,\Delta y_{t-6}; \Delta c_{t-2},...\Delta c_{t-6}; cy_{t-2}; r_{t-2},...,r_{t-6}$

Note: Δy growth rate of real GDP, r discount rate minus change in CPI, Δc growth rate of real household consumption, cy log of MPC, and Δi change in nominal interest rate call money rate.

The results of the equation (1.7) also support the AIH for Pakistan. The high and significant values of marginal propensity to consume (MPC) ensure the invalidity of PIH and indicate that the predictable changes in income revise consumption decision of consumers in Pakistan.

Table 2: Estimated Results of Equation (1.6) And (1.7)

Estimators of equations		$\Delta C_t = \alpha + \lambda \Delta Y_t + \varepsilon_t$				$\Delta C_t = \mu + \lambda \Delta Y_t + \sigma . r_t + \varepsilon_t$			R^2	DW
Models	Instr: List	α	λ	R^2	DW	μ	λ	σ		
Model: 1(OLS)	------	1558.95	0.7401***	0.97	1.5386	2045	0.783***	-0.049	0.96	1.50
Model: 2 (IV)	1					2045	0.888***	-0.171	0.85	1.47
Model: 3 (IV)	2					840	0.748***	-0.009	0.81	1.53
Model: 4 (IV)	3					1048	0.777***	-0.042	0.92	1.51
Model: 5 (IV)	4					1294	0.793***	-0.060	0.95	1.49
Model: 6 (IV)	5					1169	0.785***	-0.051	0.89	1.50

Note: model one is estimated with the help of OLS method and model two, three, four, five and six are estimated through 2SLS method, by using the instrumental lists one, two three, four and five respectively . The coefficient on (**, ***) are statistically significant at the level of (5, 1) percent respectively

Secondly, the study estimates equation (1.8) to highlights the reasons of the rejection of the neoclassical consumption hypothesis i.e. myopia and liquidity constraints. Myopia violates the PIH and implies that consumers consume a constant fraction of their current income. While in liquidity constraints consumers can smooth rising income through saving but cannot able to smooth falling income. The equation (1.8) clarifies the source of the rejection of PIH. The estimates of the equation (1.8) are reported in table 3.

Equation (1.8) is estimated with the help of OLS and 2SLS by using the different instrumental variables, mentioned in table 1. There are three hypotheses regarding the equation (1.8). First, if the permanent income hypothesis is valid if: $\lambda_1 = \lambda_2 = 0$, second, if the PIH is not valid due to myopia then, the significant value of λ_1 and λ_2 will be: $\lambda_1 > 0, \lambda_2 > 0$ and $\lambda_1 = \lambda_2$, and third, if the PIH is not valid due to liquidity constraints then the significant value of λ_1 and λ_2 will be : $\lambda_1 > 0$, and $\lambda_1 > \lambda_2$.

Table 3: Estimated Results of Equation (1.8)

Estimators of equations		$\Delta C_t = \mu + \lambda_1 (POS_t)\Delta Y_t + \lambda_2 (NEG)\Delta Y_t + \sigma . r_t + \varepsilon_t$						
Models	Instrument List	μ	λ_1	λ_2	σ	R^2	F.statistics $H_0 : \lambda_1 = \lambda_2 = 0$ $H_1 : \lambda_1 \neq \lambda_2 \neq 0$	DW
Model:1 (OLS)	-------	-6.30E-10	-0.00269***	1.13	-1.52E-5	0.95	-----------	1.52
Model: 2 (IV)	1	3.40E-09	0.00165***	0.78	-2.52E-7	0.85	20.10***	1.47
Model:3 (IV)	2	-1.2E-08	0.00321***	1.10	4.92E-6	0.90	34.05***	1.47
Model: 4 (IV)	3	1.36E-08	4.92E-3***	0.85	-1.46E-4	0.95	15.23***	1.51
Model: 5 (IV)	4	-3.40E-09	1.95E-3***	1.20	1.95E-4	0.82	22.01***	1.47
Model: 6 (IV)	5	6.12E-09	4.45E-3***	0.98	-5.64E-5	0.92	26.12***	1.76

Note: model one is estimated with the help of OLS method and model two, three, four, five and six are estimated through 2SLS method, by using the instrumental lists one, two three, four and five respectively . The coefficient on (**, ***) are statistically significant at the level of (5, 1) percent respectively

Thus, the empirical results of table 3 shows that the value of λ_1 is negative in model 1 and positive and significant in all IV models. While the value of λ_2 is found insignificant in OLS and all IV models.
These results of equation (1.8) support the results of equation (1.6) and (1.7) and reject validity of PIH because the possible condition for the validity of PIH is not found i.e. the expected negative and positive income do not affect consumption ($\lambda_1 = \lambda_2 = 0$).The condition which ensures myopia is also not observed in all models from 1 to 6. Because the value of λ_2 is insignificant in OLS model and through out the IV models. Furthermore, the results are somehow fulfill the condition of the liquidity constraints i.e. λ_1 is positive, significant and greater than zero. Furthermore, the results of equation (1.8) indicate that consumption is more sensitive to increases than decreases predictable income, due to the significant values λ_1 and the insignificant values of λ_2 . As a result of this evidence the presence of liquidity constraints is ensure rather than myopia or perverse asymmetry. Therefore, we can say that the rejection of PIH in Pakistan is not subject to myopia but liquidity constraints.

4. CONCLUSION

The study investigates validity of the PIH in Pakistan. The results indicate invalidity of PIH in Pakistan. Because the predictable changes in income can affect consumption. This is the clear violation of the neoclassical consumption hypothesis. Furthermore, to find out the reasons of the rejection of the PIH, the study applies the Shea (1995) model to investigate the reasons of rejection of the PIH i.e. Myopia and liquidity constraint. The results are supported the presence of liquidity constraints rather than myopia and perverse asymmetry.

REFERENCES

1. Campbell, J. and Mankiw, N. (1990). Permanent Income, Current Income, and Consumption. *Journal of Business and Economic Statistics*, 8, 265-79.
2. Friedman, M. (1957). *A theory of the consumption function*. Princeton University Press.
3. Hall, E.E. (1978). Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence. *Journal of Political Economy*, 86, 971-87.
4. Khalid, A.M. (1994). Empirical Test of the Rational Expectation and Permanent Income Hypothesis Evidence from Pakistan. *The Pakistan Development Review*, 33, 1043-1053.
5. Runkle, D. (1991). Liquidity Constraints and the Permanent Income Hypothesis. *Journal of Monetary Economics*, 27, 73-98.
6. Shea, J. (1995). Myopia, Liquidity Constraints, and Aggregate Consumption: A simple Test. *Journal of Money, Credit and Banking*, 27, 798-805
7. Zeldes, S.P. (1989). Consumption and Liquidity Constraints: An Empirical Investigation. *Journal of Political Economy*, 97, 305-346.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 113-119

AGE AND GENDER AS A RISK FACTOR OF MYOCARDIAL INFARCTION

Shazia Azeem[1] and **Nazeer Khan**[2]

[1] Department of Statistics, Federal Urdu University, Karachi, Pakistan. Email: shazki22@yahoo.com

[2] Dow University of Health Sciences, Karachi, Pakistan. Email: n.khan@duhs.edu.pk

ABSTRACT

Objective: To asses the effects of age and gender on different risk factors of Myocardial Infarction (MI).

Method: The data were collected from four different hospitals (NICVD, KIHD, and LNH, CIVIL) of Karachi. Three hundred fifty patients and control each were included in this study.). The inclusion criteria of case were their positive ECG report. The inclusion criteria for control were no previous personal or family (no first degree relative) history of coronary disease or chest pain. The exclusion criteria for both case and control were disagreement with informed consent, pregnancy, age more than 70 years and less than 30 years. The subjects were interviewed (through already prepared questionnaire) and investigated from the cardiac ward and out patients clinics of the hospital. The questionnaire included basic demographic and socio-economic information in detail and the risk factors attached with MI patients.

Results: Eighty two percent of the MI patients were male ($P<0.0001$) with average age of 52.7 years. Forty eight percent ($P<0.0001$) of the MI patients were from age group (45-59.99) years. DM was significant with age groups ($P=0.012$) among MI patients. Highest proportion of MI diabetic patients was observed in age group (≥60) years. Highest proportion of MI hypertensive patients was also observed in age group (≥60) years. The proportion of smoker MI patients was higher in age group (30-44.99) years.HDL was insignificant ($P=0.095$) among MI patients while the highest proportion of MI patients was observed in age group (30-45.9) years. Seventy four percent female MI patients ($P<0.0001$) were hypertensive while the overall prevalence of hypertension among MI patients in the study sample was 40.2%. The proportion of male diabetic MI patients was higher than female diabetic MI patients. Sixty percent male MI patients were smoker ($P<0.0001$) while overall prevalence of smoking among MI patients was 54%. The proportion of male (TG: ≥150 mg/dl) was higher than female ($P=0.027$) in control group.

Conclusions: The finding of the study revealed that male gender and increasing age are the two important risk factors of MI. Hypertension was more common in female MI patients while DM was more common in male MI patients.

INTRODUCTION

In Pakistan Ischemia Heart Disease (IHD) is the 2nd leading cause of death at all ages contributing to 11% of all deaths (Andrieu N., Prevost T., Rohan T.E., et al.: 2000). In

2002 about 100,000 (Acute Myocardial Infarction) AMI patients were reported in Pakistan (Iqbal M.P., Ishaq M. et al. 2005). In many epidemiologic surveys, age remains one of the strongest predictors of disease. More than half of those who have heart attacks are 65 or older, and about four out of five who expired of such attacks are 65 and above. Of course, nothing can be done to reduce age. However, careful attention to diet and maintaining fitness may delay the degenerative changes associated with aging. The incidence of Coronary heart disease (CHD) is lower in women than in men. In general, the incidence of CHD in women begins to increase after menopause. But once past the menopause, a woman's risk is similar to a man's (Heart Book, 1992).

METHOD

The data were collected from four different hospitals (NICVD, KIHD, and LNH, CIVIL) of Karachi. Three hundred fifty patients and control each were included in this study.). The inclusion criteria of case were their positive ECG report. The inclusion criteria for control were no previous personal or family (no first degree relative) history of coronary disease or chest pain. The exclusion criteria for both case and control were disagreement with informed consent, pregnancy, age more than 70 years and less than 30 years. The subjects were interviewed (through already prepared questionnaire) and investigated from the cardiac ward and out patients clinics of the hospital. The questionnaire included basic demographic and socio-economic information in detail and the risk factors attached with MI patients. The study was approved by the Ethical Review Board (ERB) of Dow University of Health Sciences (DUHS). All the study hospitals accepted the ERB approval of DUHS and gave the permission for the study. The study was explained to the patients and the consent was taken verbally.

RESULTS

Eighty two percent of the MI patients were male ($P<0.0001$) with average age of 52.7 $\pm$ 9.97 years. Fifty one percent ($P<0.0001$) of the MI patients were from age group (45-59.99) years. Table-1 shows the risk factors categorized by gender. Female gender was significant ($P<0.0001$) for both case and control among hypertensive subjects. Among control, female diabetic patients were significant higher ($P=0.037$). Male gender was significant ($P<0.0001$) among smokers for both case and control. The proportion of female with Triglycerides (TG: $\geq$150 mg/dl) was higher among MI patients. Male gender was significant with HDL for both case and control. Male gender with TG ($\geq$150 mg/dl) was significant ($P=0.027$) among control subjects. The prevalence of overweight and obese were highly significant ($P=0.019$) among male MI patients. In dietary habits use of vegetables and fruits and chicken was significantly higher in male MI patients ($P<0.0001$).

Table 1: Risk factors categorized by gender

Risk factors	Status	Male (%)	Female (%)	Total (%)	P-value
Hypertension yes	case	96(33.0)	46(74.2)	142(40.2)	<0.0001
	control	20(23.0)	96(36.4)	116(33.0)	0.014
DM yes	case	135(46.4)	28(45.9)	163(46.3)	0.529
	control	19(21.8)	86(32.6)	105(29.9)	0.037
Smoke yes	case	175(60.1)	15(25.0)	190(54.1)	<0.0001
	control	25(28.7)	28(10.6)	53(15.1)	<0.0001
HDL	case	232(79.7)	59(95.2)	291(82.4)	0.001
	control	53(60.9)	226(85.6)	279(79.5)	<0.0001
DM(father) yes	case	61(21.0)	22(36.1)	83(23.6)	0.011
	control	19(21.8)	70(26.5)	89(25.4)	0.235
TG(≥150 mg/dl)	case	142(48.8)	37(59.7)	179(50.7)	0.078
	control	40(46.5)	90(34.1)	130(37.1)	0.027
Glucose(≥126 mg/dl)	case	158(54.5)	30(48.4)	188(53.4)	0.232
Glucose(≥126 mg/dl)	control	17(19.5)	85(32.3)	102(29.1)	0.015
BMI (<23) kg/m^2	case	66(22.7)	13(21)	79(22.4)	0.517
(23-24.9) kg/m^2		83(28.5)	14(22.6)	97(27.5)	
(≥25) kg/m^2		142(48.8)	35(56.5)	177(50.1)	
BMI (<23) kg/m^2	control	30(34.5)	42(15.9)	72(20.5)	0.001
(23-24.9) kg/m^2		23(26.4)	107(40.5)	130(37.0)	
(≥25) kg/m^2		34(39.1)	115(43.6)	149(42.5)	
Veg_fruit no	Case	11(3.8)	8(12.9)	19(5.4)	<0.0001
once		14(4.8)	10(16.1)	24(6.8)	
≥2 times		266(91.4)	44(31.0)	310(87.8)	
Veg_fruit no	control	33(37.9)	71(26.9)	104(29.6)	0.020
Once		25(28.7)	60(22.7)	85(24.2)	
≥2 times		29(33.3)	133(50.4)	162(46.2)	
Chicken_week no	case	32(11.0)	19(30.6)	51(14.4)	<0.0001
Once		93(32.0)	15(24.2)	108(30.6)	
≥2 times		166(57.0)	28(45.2)	194(55.0)	
Chicken_week no	control	68(78.2)	149(56.4)	217(61.8)	<0.0001
Once		15(17.2)	49(18.6)	64(18.2)	
≥2 times		4(4.6)	66(25.0)	70(19.9)	

Table 2 shows the risk factors categorized by different age groups. The proportion of diabetic patients was significantly higher in age group (≥60) years for both case (P=0.012) and control (P<0.0001). Hypertension was not significant with different age groups; however the prevalence of hypertensive patients was higher in age group (≥45) years. The highest proportion of smoker was observed in age group (≥60) years among control subjects (p < 0.001). The family history of DM (father) and DM (mother) was significant with different age groups among control subjects. The proportion of patients with glucose (≥126 mg/dl) was significantly higher in age group (≥60) years for both case (P=0.012) and control (P<0.0001). The highest proportion of MI patients with HDL (<40 mg/dl,<50 mg/dl) was observed in age group of 30-44.99) years. Obesity was observed in age group of 30-45 years for both case and control. The consumption of fresh fruits and vegetables were significantly higher among MI patients. Use of red meat was significantly higher in age group of 30-45 years among MI patients.

Table 2: Risk factors categorized by Age:

Risk factors	**Status**	**(30-45) yr**	**(45-60)yr**	**(≥ 60)yr**	**Total**	**P-value**
CHD death sibling	case	16(32.0)	60(33.1)	31(25.6)	107(30.4)	0.365
	control	50(33.8)	42(32.3)	11(15.1)	103(29.3)	0.010
DM	case	14(28.0)	85(47.0)	64(52.9)	163(46.3)	0.012
	control	23(15.5)	45(34.6)	37(50.7)	105(29.9)	<0.0001
Hypertension	case	17(33.3)	69(40.8)	56(42.1)	142(40.2)	0.541
	control	42(28.4)	43(34.7)	31(39.2)	116(33.0)	0.226
waist	case	42(82.4)	111(65.7)	88(66.2)	241(68.3)	0.065
	control	92(62.2)	60(48.4)	48(60.8)	200(57.0)	0.055
Smoking	case	31(62.0)	101(56.1)	58(47.9)	190(54.1)	0.182
	control	7(4.7)	28(21.5)	18(24.71)	53(15.1)	<0.0001
DM (father)	case	12(24.0)	43(23.9)	28(23.1)	83(23.6)	0.987
	control	51(34.5)	25(19.2)	13(17.8)	89(25.4)	0.004
DM (Mother)	case	9(18.0)	55(30.4)	37(30.3)	101(28.6)	0.201
	control	40(27.0)	17(13.1)	16(21.9)	73(20.8)	0.016
Glucose (≥126 mg/dl)	case	17(34.0)	101(56.0)	70(57.4)	188(53.4)	0.012
	control	34(23.0)	30(23.3)	38(52.1)	102(29.1)	<0.0001
HDL (<40 mg/dl, <50 mg/dl)	case	45(90.0)	152(84.0)	94(77.0)	291(82.4)	0.095
	control	124(83.8)	92(70.8)	63(86.3)	279(79.5)	0.007
BMI (<23) kg/m^2	case	7(14.0)	44(24.3)	28(23.0)	79(22.4)	0.503
(23-24.9) kg/m^2		13(26.0)	48(26.5)	36(29.5)	97(27.5)	
(≥25) kg/m^2		30(60.0)	89(49.2)	58(47.5)	177(50.1)	
BMI (<23) kg/m^2	control	18(12.2)	26(20.0)	28(38.4)	72(20.5)	<0.0001
(23-24.9) kg/m^2		52(35.1)	57(43.8)	21(28.8)	130(37.0)	
(≥25) kg/m^2		78(52.7)	47(36.2)	24(32.9)	149(42.5)	
Waist **(≥80 cm, ≥90 cm)**	Case	41(82.0)	123(68.0)	77(63.1)	241(68.3)	0.053
	control	92(62.2)	66(50.8)	42(57.5)	200(57.0)	0.159
Veg_fruit no	Case	2(4.0)	10(5.5)	7(5.7)	19(5.4)	0.888
once		5(10.0)	11(6.1)	8(6.6)	24(6.8)	
≥2 times		43(86.0)	160(88.4)	107(87.7)	310(87.8)	
Veg_fruit no	control	51(34.5)	38(29.2)	15(20.5)	104(29.6)	0.001
Once		38(25.7)	38(29.2)	9(12.3)	85(24.2)	
≥2 times		59(39.9)	54(41.5)	49(67.1)	162(46.2)	
Meat no	Case	15(30.0)	52(28.7)	47(38.5)	114(32.3)	0.037
once		5(10.0)	48(26.5)	26(21.3)	79(22.4)	
≥2 times		30(60.0)	81(44.8)	49(40.2)	160(45.3)	
meat no	control	27(18.2)	27(20.8)	17(23.3)	71(20.2)	0.913
once		26(17.6)	21(16.2)	13(17.8)	60(17.1)	
≥2 times		95(64.2)	82(63.1)	43(58.9)	220(62.7)	

Table 3 shows descriptive statistics categorized by gender. Among MI patients, average TG of females was significantly (P=0.044) higher than male. Among control group average glucose level of females was significantly (P=0.030) higher than males.

Table-3: Descriptive statistics categorized by gender:

Status	Risk factors	gender	Mean	SD	P-value
MI	TG	male	153.29	75.515	
		female	177.55	86.605	.044
control	glucose	male	110.05	52.563	.030
		female	129.13	75.706	
	TG	male	159.20	81.082	
		female	142.30	66.145	.081

Table-4 shows descriptive statistics categorized by different age groups. Among MI patients average glucose level was significantly (P=0.030) higher in age group of 45-60 years as compare to age group (30-44.9) years. Average value of BMI and TG were significantly higher, while HDL was significantly lower in age group (45-59.9) years as compare to age group (≥60) years among MI patients. In control subjects' average value of BMI and cholesterol were significantly higher ($p < 0.05$) in age group (45-59.9) years as compare to age group (≥60) years. In control subjects average value of glucose (P=0.019) was significantly higher in age group (≥60) years as compare to age group (45-59.9) years.

Table-4: Descriptive statistics categorized by age:

Status	Risk factors	age	Mean	SD	P-value
MI	glucose	30-44.99	120.08	44.191	0.030
		45-55.99	142.25	67.713	
Control	BMI	30-44.99	27.74	4.790	0.004
		45-55.99	26.06	4.784	
MI	BMI	≥60	24.81	3.858	0.018
		45-55.99	26.06	4.853	
	TG	≥60	144.10	67.919	0.036
		45-55.99	163.44	84.551	
	HDL	≥60	37.55	13.819	0.045
		45-55.99	35.25	5.542	
Control	BMI	≥60	23.65	3.523	<0.0001
		45-55.99	26.06	4.784	
	Cholesterol	≥60	169.33	39.842	0.026
		45-55.99	182.91	42.170	
	Glucose	≥60	152.42	87.471	0.019
		45-55.99	124.05	79.310	

DISCUSSION

This study aimed to identify the effect of age and gender on different risk factors of MI. Male gender was found one of the risk factor of MI and this result is consistent with a study which showed that in both young and old age groups, prevalence of AMI is higher in male (Dang A., Dias A. 2008). The prevalence of hypertension was higher among females both case and control. This result is consistent with another study according to that the proportion of hypertension (20.2% vs. 17.4%, P=0.003), hyperlipidemia (14.6%, vs. 10.1%, P<0.001), and central obesity (42.4% vs. 14.7%, P<0.001) were significantly greater in women than in men (Davis M.A. et al, 1994).The prevalence of DM was significantly higher in females among control subjects. This result is in contrast with another study that showed the prevalence of CHD was higher in diabetic women (Hu G., Jousilahti P., Qiao Q., Katoh S., Tuomilehto J. 2005).The proportion of female MI patients with TG(≥150 mg/dl) was significantly higher than male MI patients. The main reason of this high TG among females was trend of keeping maids for household jobs in Karachi. This result does not agree with another Pakistani study that reported that the mean triglyceride levels among males were significantly higher than among females (Iqbal M.P., Shafiq M., Mehboobali N., et al. 2004). Smoking, obesity, HDL (<40 mg/dl) and use of vegetables, fruits and chicken were significant higher among with male subjects. This finding is consistent with another study according to that smoking was less common in women, and diabetes and high intake of meat were similar in both male and females (Davis MA et al. 1994). Among MI patients highest proportion of hypertension and diabetic was observed in old patients. This result is consistent with a study done in Saudi Arabia that demonstrate an increasing prevalence of diabetes with advancing age (Al-Nozah M, Khan NB, et al, 2004). HDL, obesity and high consumption of red meat was observed in young MI patients that reflect the sedentary life style and unhealthy eating habits are the main cause of MI in young generation of Pakistan. The prevalence of MI was higher among 45-60 years. The glucose level, BMI, TG and HDL were higher in this age group among MI patients. This result is consistent with a study conducted in Pakistan. (Jafary M., Samad A., Ishaq M., Shaukat Jawaid S., Ahmad M., Vohra E, 2007).

CONCLUSION

The finding of the study revealed that male gender and increasing age are the two important risk factors of MI. In Karachi, Pakistan the main risk factors of Myocardial infarction among females were Hypertension and TG, while smoking, obesity and HDL among males. Age group (45-60) years were the high risk age group for MI patients.

REFERENCES

1. Al-Nozha M.M., Al-Maatouq M.A., Al-Mazrou Y.Y., Al-Harthi S.S., Arafah M.R., Khalil M.Z., Khan N.B., Al-Khadra A., Al-Marzouki K., Nouh M.S., Abdullah M., Attas O., AlShahid M.S. and Al-Mobeireek A. (2004). Diabetes mellitus in Saudi Arabia. *Saudi Med. J.*, 25(11), 1603-1610.
2. Andrieu, N., Prevost, T., Rohan, T.E., et al. (2000). Variation in the interaction between familial and reproductive factors on the risk of breast cancer according to age, menopausal status, and degree of familiarity. *Int. J. Epidemiol.*, 29(2), 214-23.

3. Dang, A. and Dias, A. (2008). Differences in risk factors and mortality, in young and old individuals with Acute Myocardial Infarction in INGOA. *J. Clin. and Diagn. Res*., [serial online] (cited: 2009, Dec., 2) 2, 715-719 http://www.jcdr.net/back_issues.asp?issn=0973709x&year=2008&month=April&volume=2&issue=2&page=715-719&id=223.
4. Davis M.A., Neuhaus, J.M., Moritz, D.J., Lein,D., Barclay, J.D. and Murphy, S.P. (1994). Health behaviors and survival among middle-aged and older men and women in the NHANES I Epidemiologic Follow-up Study. *Preventive Medicine*, 23(3), 369-76.
5. Hu, G., Jousilahti, P., Qiao, Q., Katoh, S, Tuomilehto J. (2005). Sex differences in cardiovascular and total mortality among diabetic and non-diabetic individuals with or without history of myocardial infarction. *Diabetologia,* 48, 856-861.
6. Iqbal M.P., Shafiq M., Mehboobali N., Iqbal S.P. and Abbasi K. (2004). Hypertriglyceridemia and low HDL-cholesterol are the major Lipid abnormalities in Normal Pakistani adults: Are these contributing to the high rates of coronary artery disease in this population? *Pak. J. Med. Sci*., 20(4), 385-91.
7. Iqbal, M.P., Ishaq, M., Kazmi, K.A., Yousuf, F.A., Mehboobail, N., Ali, S.A., Khan A.H. and Waqar, M.A. (2005). Role of vitamin B6, B12 and folic acid on hyperhomocysteinemia in a Pakistani population of patient with acute myocardial infarction. *Nutr. Metab. Cardiovasc. Dis.*, 15, 100-8.
8. Jafary, M., Samad, A., Ishaq, M., Shaukat, Jawaid S., Ahmad, M., Vohra, E. (2007). Profile of Acute Myocardial infarction (AMI) in Pakistan. *Pak. J. Med. Sci.,* 23.
9. Zaret, B.L., Moser, M. (1992). *Yale University School of Medicine Heart book*. Cohen LS, Medical editors. Subak-sharpe GJ, Editorial director. New York: Hearst Books, ISBN 0-688-09719-7.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 121-124

COMPUTER SCIENCE IN EDUCATION: A SURVEY

Irshad Ullah
Government Higher Secondary School, OUCH Dir (Lower), Pakistan
Email: irshadullah79@gmail.com

ABSTRACT

Computer is used in the field of education for different function. Computer may be used to practice software to make the office work consistent and speedy in a fresh and soft environment. Computer may be used to provide education called by the society special peoples. Because if they cannot see so they may use micro phone to hear the lectures and use the fingers for typing purpose. Peoples may record the lectures for future use and may to be keeping in library. Its may be used to broad cast the lectures. Video conferencing facility may be used to give the lecture at the same time at multiple Places. Its may work just like a teacher in special situation. Educational institutions may be networked with each other and with their corresponding offices.

KEY WORDS

Software; Speedy; Environment; Lecture; Video.

1. INTRODUCTION

The Use of Computers in Education The typical school has 1 computer per 20 students, a ratio that computer educators feel is still not high enough to affect classroom learning as much as books and classroom conversation. Some critics see computer education as merely the latest in a series of unsuccessful attempts to revolutionize education through the use of audio- and visually-oriented non print media. For example, motion pictures, broadcast television, filmstrips, audio recorders, and videotapes were all initially heralded for their instructional potential, but each of these ultimately became minor classroom tools alongside conventional methods.

1.1 Communications Satellite: A communications satellite is an artificial Satellite placed into orbit around the Earth to facilitate communications on Earth. Most long-distance radio communication across land is sent via MICROWAVE relay towers. In effect, a satellite serves as a tall microwave tower to permit direct transmission between stations, but it can interconnect any number of stations that are included within the antenna beams of the satellite rather than simply the two ends of the microwave link. Computer Crime Computer crime is defined as any crime involving a computer accomplished through the use or knowledge of computer technology. Computers are objects of crime when they or their contents are damaged, as when terrorists attack computer centers with explosives or gasoline, or when a "computer virus" a program capable of altering or erasing computer memory is introduced into a computer system.

1.2 Personal Computer: A personal computer is a computer that is based on a microprocessor, a small semiconductor chip that performs the operations of a c.p.u. Personal computers are single-user machines, whereas larger computers generally have multiple users. Personal computers have many uses such as: Word processing, communicating to other computers over a phone line using a modem, databases, and leisure games are just some of the uses of a Personal Computer. Computers for Leisure Games As they proliferated, video games gained color and complexity and adopted the basic theme that most of them still exhibit: the violent annihilation of an enemy by means of one's skill at moving a lever or pushing a button. Many of the games played on home computers are more or less identical with those in video arcades. Increasingly, however, computer games are becoming more sophisticated, more difficult, and no longer dependent on elapsed time a few computer games go on for many hours. Graphics have improved to the point where they almost resemble movies rather than rough, jagged video screens of past games. Some of the newest arcade games generate their graphics through C.D R.O.M. Many include complicated sounds; some even have music and real actors. Given an imaginative programmer, a sophisticated video game has the potential for offering an almost limitless array of exotic worlds and fantastic situations. In the early 90s parents and government were becoming increasingly aware of violence in video games so they introduced warnings on the box like in the movies. (essaysample.com/essay/003335.html 2011).

A computer helps to manipulate data according to a list of instructions, called a program. It can save a lot of data and produce the same, instantly. It is also called a universal information-processing machine. It is a common operating system used in corporate businesses, educational organizations and many research programs. The computer technology has been used proficiently in various educational fields. There are many professional courses that program their curriculum on computers. It enables the students and teachers to methodically study or conduct classes. Most schools highlight the importance of computer education. They provide computer education to children, at a very young age. It helps them to learn and develop interest in the basics of computers. The main purpose is to make them comfortable using the system, as the future holds a bright promise for the technology. Many students become proficient and plan to pursue careers in the world of computers.

Computers are used in running school and college administrations, during the admission procedures, storing of official and student records. They are also used in syllabus planning and decision-making, controlling, assisting instructions and simulation. Computers are helpful in directing aptitude tests and achievement tests, at the time of entrance exams. There is computer software designed to process performances related to teachers and employees promotion avenues. They also process records of salaries, examinations, schemes of examination, printing of papers and question papers, evaluation of answer sheets, mark sheets and certificates.

Computers are used in colleges, by the professors, to conduct special classes and enable their students to adopt a methodical way of study. Students take more interest in the documented programs, designed on different topics. Internet has provided a favorable means of pursuing courses from renowned universities, across the world. These facilities become available at a click of the mouse.

Computer Education provides detailed information on Computer Education, Computer Education Online, Computer Science Education, Computer Education Institute and more. Computer Education is affiliated with Computer Graphic Training. (Article Source: http://ezinearticles.com/?expert=Max_Bellamy2011).

Computers have changed the way we work, be it any profession. Therefore, it is only but natural the role of computers in education has been given a lot of prominence in the recent years. Computers play a vital role in every field. They aid industrial processes; they find applications in medicine; they are the heart of the software industry; they play a vital role in education. The uses of computers in education are manifold. Here, we shall discuss the important facets of the role of computers in education.

The computer technology has a deep impact on education. Computer education forms a part of the school and college curricula, as it is important for every individual today, to have the basic knowledge of computers. The advantages of computers in education include an efficient storage and rendition of information, quick information processing and very importantly the saving of paper. Know more about the importance of computer education.

Computer teaching plays a key role in the modern systems of education. Students find it easier to refer to the Internet than searching for information in heavy reference books. The process of learning has gone beyond learning from prescribed text books. Today, aspirers can satiate their thirst for knowledge by means of the Internet. It is easier to store information on computers than maintaining hand-written notes. To know more on the subject, read about textbooks versus computer teaching.

Online education has revolutionized the education industry. The computer technology has made the dream of distance learning, a reality. Education is no more limited to classrooms. It has reached far and wide thanks to the computer technology. Physically distant locations have come close to each other only due to computer networking.

Computers facilitate an efficient storage and effective presentation of information. Presentation software like PowerPoint and animation software like Flash and others can be of great help to the teachers while delivering information. Computers can turn out being a brilliant aid in teaching. Computers facilitate an audio-visual representation of information, thus making the process of learning interactive and interesting. Computer-aided teaching adds a fun element to education.

Internet can play an important role in education. As it is an enormous information base, it can be harnessed for the retrieval of information on a wide variety of subjects. The Internet can be used to refer to information on various subjects to be taught to the students.

Moreover, computers facilitate an electronic format for storage of information, thereby saving paper. Homework and test assignments submitted as soft copies save paper. Electronically erasable memory devices can be used repeatedly. They offer a robust storage of data and reliable data retrieval. The computer technology thus eases the process of learning.

A life without computers would seem almost unimaginable for many. The importance of computers is evident today and having the perfect know-how of computers can only propel one’s career in the right direction. Today, computers are a part of almost every industry. They are no more limited to the software industry. They are widely used in networking, information access, data storage and the processing of information. So why not introduce computers early in education? Introducing computers early in education lays the foundation of most of the major competitive careers. Computers play a significant role in one’s personal and professional life. (buzzle.com/articles/role-of-computers-in-education.htm 2011).

Computers have become the life line of young generation. The present generation students like to embrace all the things in the ambit of this modern technology. The educators are feeling shy and reluctant to accept technology as their work partner. Will our present and the immediate next generation educators like to view the gap widening between them and further generation students in terms of usage of technology for teaching learning purpose or they would like to modernize themselves in the struggle to keep intact their responsible professional position intact? Are universities and institutions in the field of teacher education focusing upon the courses for using computers in education or they simply feel delighted to see the state boards at school level and National/State councils for Educational Research and Training are advancing at great speed to plan out integration of technology with education with teachers struggling to complete even the prescribed syllabus?

To ‘teach’ is one of the simplest tasks! Just you know the content and “make available” a student willing to learn; start teaching. You may pursue a diploma/degree in education for teaching in more formalized structure. After completing a formal teacher training course, a teacher gets a tag of “A Qualified Teacher”. But does this really provide courage and confidence to face a modern technological high tech class of young generation? Today students of this modern generation are much advanced and smarter than their teachers in terms of usage of modern technical gadgets. It looks as if technology had been with the children since birth, especially computers been an integral part of child’s life and personality. He/she would once forget what his/her parents had instructed, but would never forget to be “Online” at a specified time to join his/her friends at chat room.

But they even want that computers should be there for their routine ‘boring education’, although not at the cost of their dearly loved teachers. Students are ready to accept the changes in their learning styles, but their teachers are feeling feared that computers are complicated to handle, if they are used in education for teaching learning purpose and computers might substitute the traditional teachers. Neither the pre-service teacher training courses are focusing to include computers in integration with education, nor are the in-service trainings focusing upon using computers in education. Schools all over the country are demanding teachers who are ready to initiate the process of Computer Assisted Instructions and initialize integration of technology with education, but the teachers we are producing are just like the conventional and that too overlapped with traditional methods of teaching. When the time will come, that universities and institutions in the field of teacher education, feel an urge to revise their curriculum keeping in view the modern technological changed society?

In order to meet the challenges provided by present society, many universities in the field of Teacher Education had introduced an optional course in computers. This only brought awareness about computers in the field of Teacher Education. The student-teachers were made to know about basic computer hardware and some software application packages. But keeping in mind the interest and psychology of modern learner, many universities tried to ‘Integrate Technology with Education’. They primarily focused upon use of computers in education hence; they initiated to teach a core paper “Computers in Education” with other foundation papers in education, so as to initiate a process of making a relation between teacher-computer-student. (wikieducator.org/Computers_in_Education 2011).

REFERENCES

1. Article Source: http://ezinearticles.com/?expert=Max_Bellamy.
2. http://www.buzzle.com/articles/role-of-computers-in- education.html.
3. http://wikieducator.org/Computers_in_Education.
4. http://www.essaysample.com/essay/003335.html 2011.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 125-128

COMPUTER SCIENCES AND ENVIRONMENTAL SCIENCES: A SURVEY

Irshad Ullah
Government Higher Secondary School, OUCH Dir (Lower), Pakistan
Email: irshadullah79@gmail.com

ABSTRACT

The term “green computing” is one that is coming out in many different markets and areas all over the world in today's universal technology. Yet, many computer users out there aren't sure what it means. Green computing is actually attractive easy to explain and to do – it is basically learning to use computer resources more efficiently to help the environment as well as energy savings. The main goal of a green computing program is to help the triple bottom line, which is an expanded spectrum of settings for measuring organizational achievement, and is extremely similar to green chemistry, which reduces the use of energy and resources that harm the environment. Green computing helps to promote the usage of biodegradable products and recycle computer components whenever possible.

KEY WORDS

Green computing, User, Achievement, Chemistry, Environment.

1. INTRODUCTION

Environmental science is an interdisciplinary academic field that integrates physical and biological sciences, (including physics, chemistry, biology, soil science, geology, and geography) to the study of the environment, and the solution of environmental problems. Environmental science provides an integrated, quantitative, and interdisciplinary approach to the study of environmental systems (**Iowa** State University, URL2010).

Related areas of study include environmental studies and environmental engineering. Environmental studies incorporate more of the social sciences for understanding human relationships, perceptions and policies towards the environment. Environmental engineering focuses on design and technology for improving environmental quality.

Environmental scientists work on subjects like the understanding of earth processes, evaluating alternative energy systems, pollution control and mitigation, natural resource management, and the effects of global climate change. Environmental issues almost always include an interaction of physical, chemical, and biological processes. Environmental scientists bring a systems approach to the analysis of environmental problems. Key elements of an effective environmental scientist include the ability to relate space, and time relationships as well as quantitative analysis.

Components Atmospheric sciences focus on the Earth's atmosphere, with an emphasis upon its interrelation to other systems. Atmospheric sciences can include studies of meteorology, greenhouse gas phenomena, atmospheric dispersion modeling of airborne contaminants, (Beychok, M.R. (2005) (Turner, D.B. (1994)) sound propagation phenomena related to noise pollution, and even light pollution. Taking the example of the global warming phenomena, physicists create computer models of atmospheric circulation and infra-red radiation transmission, chemists examine the inventory of atmospheric chemicals and their reactions, biologists analyze the plant and animal contributions to carbon dioxide fluxes, and specialists such as meteorologists and oceanographers add additional breadth in understanding the atmospheric dynamics.

Ecology an interdisciplinary analysis of an ecological system which is being impacted by one or more stressors might include several related environmental science fields. For example, one might examine an estuarine setting where a proposed industrial development could impact certain species by water and air pollution. For this study, biologists would describe the flora and fauna, chemists would analyze the transport of water pollutants to the marsh, physicists would calculate air pollution emissions and geologists would assist in understanding the marsh soils and bay muds.

Environmental chemistry is the study of chemical alterations in the environment. Principal areas of study include soil contamination and water pollution. The topics of analysis include chemical degradation in the environment, multi-phase transport of chemicals (for example, evaporation of a solvent containing lake to yield solvent as an air pollutant), and chemical effects upon biota.

As an example study, consider the case of a leaking solvent tank which has entered the habitat soil of an endangered species of amphibian. As a method to resolve or understand the extent of soil contamination and subsurface transport of solvent, a computer model would be implemented. Chemists would then characterize the molecular bonding of the solvent to the specific soil type, and biologists would study the impacts upon soil arthropods, plants, and ultimately pond-dwelling organisms that are the food of the endangered amphibian.

Geosciences include environmental geology, environmental soil science, volcanic phenomena and evolution of the Earth's crust. In some classification systems this can also include hydrology, including oceanography.

As an example study of soils erosion, calculations would be made of surface runoff by soil scientists. Fluvial geomorphologists would assist in examining sediment transport in overland flow. Physicists would contribute by assessing the changes in light transmission in the receiving waters. Biologists would analyze subsequent impacts to aquatic flora and fauna from increases in water turbidity**.**(Beychok, M.R. (2005) (Turner, D.B. (1994).

Encouraging the use of computer algorithms by developing new algorithms and introducing uncommonly known algorithms for use on environmental science problems is a significant contribution, as it provides knowledge discovery tools to extract new aspects of results and draw new insights, additional to those from general statistical methods. Conducting analysis with appropriately chosen methods, in terms of quality of performance and results, computation time, flexibility and applicability to data of various natures, will help decision making in the policy development and management process for environmental studies. (Kyoko Fukuda Thesis 2010).

Everywhere in the world, people depend upon a healthy environment. Monitoring our oceans, forests, farms, rivers, and even our cities is critical to understanding how to preserve our world and make the most of our natural resources. Additionally, understanding the behavior and health of the organisms that inhabit these spaces is crucial in our role as caretakers of the planet.

Computer scientists have been playing an active role in these tasks, working with environmental scientists, oceanographers, hydrologists, and urban planners. For example, our undergraduates have been deploying sensors along the river in Amherst to monitor contaminants in the water. To the right, undergraduates Antony Partensky and Jeffrey Cleveland deploy a buoy that uses acoustics to communicate with underwater sensors on the Quabbin Reservoir.

In another project, wood and snapper turtles have tagged with small computers, radios, and GPS devices. Together with environmental scientists from UMass, Prof. Mark Corner deploys sensors on the backs of wood turtles (Clemmys insculpta). The goal is to track the turtle’s movements to ensure the land set aside for protecting their habitat is sufficiently large.

The turtles are found throughout the northeast, living along streams and woodlands. Its numbers are dwindling through loss of habitat and highway mortality, but conservation efforts to study the creatures in their natural habitat have been hindered by a lack of tracking data. Researchers currently track turtles manually using radio telemetry and re practically limited to recording a location every two to three days per animal. In order to more accurately understand how these turtles behave and use their habitat, Mark is developing a new tracking system to collect more frequent and detailed data.

Engineering a sensor suitable for the back of a turtle highlights many challenges. The mounted sensor contains a GPS receiver and processor and must be small and lightweight. Power consumption is a key concern, since a typical GPS receiver will completely drain a small battery in two hours. Luckily, the cold-blooded turtles need to sun themselves, so a small solar panel can recharge the battery. Still, programming the devices requires careful consideration of varying energy availability and demand. (URL 2011).

REFERENCES

1. Beychok, M.R. (2005). *Fundamentals of Stack Gas Dispersion*. 4th Ed. Author-Published. ISBN 0-9644588-0-2.
2. *Computer Science Building 140 Governors Drive University of Massachusetts Amherst*, MA 01003-9264) http://www.cs.umass.edu/admissions/environmental-science].
3. Environmental Science: Iowa State University, http://www.ensci.iastate.edu (Accessed 17 February 2010).
4. Kyoko Fukuda (2010). *Computer-Enhanced Knowledge Discovery in Environmental Science*. Thesis Dissertation.
5. Turner, D.B. (1994). *Workbook of atmospheric dispersion estimates: an introduction to dispersion modeling*. (2nd Ed.). CRC Press. ISBN 1-56670-023-X.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 129-134

APPLICATION OF REGRESSION ERROR SPECIFICATION TEST (RESET) FOR GENERALIZED LINEAR MODELS: OPPORTUNITIES AND OUTCOMES

Mudassir Ahmed[1] and **Muhammad Usman Sikander**[2]
[1] Directorate General Monitoring & Evaluation, P&D Department,
Govt. of the Punjab, Lahore, Pakistan. Email: mudassir_a1@yahoo.com
[2] Center for Research in Economics & Business, Lahore School
of Economics, Lahore, Pakistan. Email: usman.lse@gmail.com

ABSTRACT

Generalized Linear Models are the extension of the popular General Linear Models. The generalization process includes the transformation of non-linear functions into linear functions using appropriate link functions e.g. Identity, *LOGIT*, *PROBIT* and *C-LOG-LOG*. Using generalized linear models for modeling binary response or count data require the selection of the appropriate link function that still remains a challenge. The paper provides an application of the Regression Error Specification Test *(RESET)* for generalized linear models using the data on Maternal Health from *MICS 2007-08* in order to provide a framework for the selection of appropriate link function.

KEY WORDS

Generalized Linear Models; Tolerance Distribution; *RESET*; Link Function; *LOGIT*; *PROBIT*; *C-LOG-LOG* and *MICS 2007-08.*

1. INTRODUCTION

Normality is the most important assumption within the domain of inferential statistics. General linear regression models therefore, assume the normality of the response variables for all such inferential procedures and statistics. However in most cases, this assumption of normality of the dependent variable does not hold true. Several statistical methods and tests are available in the literature that can be used for testing the normality of the response variable, functional form of the model and, selection of the explanatory variables. But still these tests are not frequently used for generalized linear models. Some tests like over-dispersion test for count data and specification test for qualitative dependent variable by *(Davidson & G. MacKinnon, 1982)* are used frequently in research studies but these tests does not provide help for the appropriateness of the 'link function'.

We are therefore motivated by the fact that the selection of inappropriate link function not only ignores the shape of the underlying tolerance distribution but also provides the regression estimates that may be inconsistent and vague. We use the data on Maternal Health from *MICS 2007-08* and test the newly modified version of the *RESET* test[1] for generalized linear models. The paper has been organized in the following way.

[1] (Sapra, 2005)

Section two of the paper provides some of the important determinants of the maternal health that we use for empirical investigation and extracts evidence from the literature on the *RESET* test and its application. In section three, we provide some description on our data, the methodology of the *RESET* test is also discussed in this section. Section four outlines the discussion on our empirical results and provides interpretation of the *RESET* test to select the appropriate link function. Finally, section five concludes the paper.

2. LITERATURE REVIEW

Chowdhury, Islam, & Hossain, June-2010 using multi-stage sampling, investigate the socio-economic factors that affect childhood mortality. Results suggest that mother's education, father's occupation, monthly income, types of latrine and, electricity are the major factors for postnatal and neonatal care and, infant and child mortality.

Dhakal, Chapman, Simkhada, Teijlingen, & Stephans, 2007 concludes that woman's own occupation, husband's socio-economic status (occupation and education), the number of pregnancies and number of children are the prominent factors associated with the utilization of postnatal care.

Salama and Ismail (2007) explore the factors affecting mother health services. The affect of age, duration of marriage, education, occupation, income, parity, husband's education and, distance from the health service have been examined on mother health services. The discriminate analysis supports that income, mother education and distance from the health service are the major factors responsible for choice of mother health services.

Broeck, Eleckels, & Massa, 1996 found factors associated with excess mortality in analysis found to be mother's parity, distance from the health centre, and invaliding maternal diseases. Maternal school education was significant in the multiple logistic regression only. In contrast to the other risk factors, mother-child separation or problems with breastfeeding are rare and do not significantly increase mortality.

Sunil Sapra (2005) extended the popular regression specification error test *(RESET)* by *Ramsey (1969)* for general linear regression models to *GLMs* in his paper titled "A regression error specification test *(RESET)* for generalized linear models". He applied this *RESET* test to three different economic data sets and studies the power properties of the test via a Monte Carlo experiment.

The selection of the appropriate link function is important step in fitting the *GLM* when inference and predictions are the objectives. *Huettmann & Linke, 2003* studied the effect of different link functions for binary response models. *Breslow (1996)* checked the different assumptions of GLMs in his paper and used the Pregibon's goodness of Link test for checking the appropriateness of the link function.

After fitting the *GLM* with standard link function and obtaining the fitting values, one can test the significance of the constructed variable using chi square with one degree of freedom. The test takes the following form.

$$z = -\left[\frac{\partial g\left(\hat{p};\lambda\right)}{\partial\lambda}\right] \; at \;\; \lambda = \lambda o^2 .$$

3. DATA AND METHODOLOGY

The data used for the present study was collected by the Bureau of Statistics, Planning and Development Department, Government of the Punjab. This survey *MICS 2007-08* was conducted from December *2007* to February *2008* and surveyed a total of almost *91,280* households from urban and rural locations of Punjab. This survey is a reliable source for the estimation and assessment of various socio-economic and demographic indicators in order to measure the level of achievement on *MDGs*. We use just the data for the *Gujrat* district only as the proportion of women reporting taking Antenatal care from improved sources i.e. from doctors, nurses, lady health visitors and lady health workers is highest in this district.

Regression Error Specification Test (*RESET*)[2]

The *RESET* test makes a comparison between a preliminary *GLM* to another *GLM* including higher order terms of predicted preliminary link function. Specifically, *RESET1* compares the preliminary *GLM* to the *GLM* with the second power of the predicted link function of preliminary *GLM*. Similarly, the *RESET2* compares the former with a *GLM* with the third power of the predicted link function of preliminary *GLM*. Following three *GLMs* with the same stochastic component but different systematic components below can provide a better illustration of the method.

$$GLM1 = \eta_{1i} = \beta' x_i$$
$$GLM2 = \eta_{2i} = \beta' x_i + \gamma_1 \ \hat{\ }_{1i}^{2}$$
$$GLM3 = \eta_{3i} = \beta' xi + \gamma_1 \ \hat{\ }_{1i}^{2} + \gamma_2 \ \hat{\ }_{1i}^{3}$$

Now we consider the following *RESET* tests.

RESET1 Compares *GLM1* and *GLM2* and tests $Ho: \gamma_1 = 0$ *against* $Hi: \gamma_1 \neq 0$

RESET2 Compares *GLM1* and *GLM3* and tests $Ho: \gamma_1 = \gamma_2 = 0$ *against* at least one of the coefficients γ_1 and γ_2 is not equal to zero.

Let L_1, L_2 and L_3 denote the log likelihood functions evaluated at the *MLEs* under *GLM1*, *GLM2* and *GLM3* respectively. Then the test statistics for *RESET1* and *RESET2* are as under

$\lambda_1 = -2\ (L_1 - L_2) \sim \chi^2(1)$ under *Ho and*
$\lambda_2 = -2\ (L_1 - L_3) \sim \chi^2(2)$ under *Ho.*

4. RESULTS

We are going to model the probability for women taking Antenatal Care *(ANC)* from the improved sources. Certain area specific variables e.g. the locality *(AREA)* and distance to available health facility *(DHF)* and others specific to the household e.g. ownership of the household *(OH)*, income category *(IC)*, sex of the head of the household *(HHS)* and, whether the household is in a combined family setup *(CF)* used as possible determinants that explain the probability for a women to take Antenatal Care *(ANC)* from improved sources. Mother specific indicators including the current pregnancy *(CP)*, education level of mother *(ME)*, age of the mother *(MAGE)*, square of mother age

---

[2] The detailed discussion on the *RESET* test has been taken from Sunil Sapra (2005).

(MAGESQ) and, total child ever born by mother *(CEB)* provide valuable information in relation to the topic.

Three models *LOGIT*, *PROBIT* and *C-LOG-LOG* were fitted on the data. Once the models were finalized different methods were adopted to make comparison between the three models. This comparison includes goodness of fit statistics i.g. *Deviance, Pearson Chi Square, Log Likelihood, AIC and BIC.* Minimum values of these diagnostic tools suggested that *PROBIT* is the first best, *LOGIT* as second and *C-LOG-LOG* as the third best model. Once we employ *RESET* test the *LOGIT* turns out to be the best link function for the data. *RESET* Statistics λ_1 *and* λ_2 are used to test for the adequacy of the link function. An insignificant value of the *RESET* Statistic λ presents the notion that the link function is appropriate whereas a significant value does not.

Our results suggest that the *RESET* statistic λ_1 is insignificant only for *LOGIT* model at 1% level of significance and significant at 1% for all two models and λ_2. Moreover, the correct predictions made by the three models clearly distinguish between the adequate and in-adequate models. The predictability of the *LOGIT* model is the best amongst the three models providing evidence that the *LOGIT* best describe the variation in the dependent variable.

Prediction

Prediction	**Fitted Category**	**ANC**	
		No	**Yes**
LOGIT	No	1292(90%)	313(52%)
	Yes	144(10%)	285(48%)
PROBIT	No	1293(90%)	320 (54%)
	Yes	143(10%)	278(46%)
C-LOG-LOG	No	1320(92%)	344(58%)
	Yes	116(8%)	254(42%)

Goodness of Fit Tests

Statistics	***LOGIT***	***PROBIT***	***C-LOG-LOG***
Deviance	1590.913	1583.237	1618.781
P Chi-Square	1850.311	1834.277	1895.332
Log Likelihood	-950.966	-947.127	-964.899
AIC	1937.931	1930.255	1965.799
BIC	2035.383	2027.706	2063.250

RESET Statistics

RESET	***LOGIT***		***PROBIT***		***C-LOG-LOG***	
	Values	χ^2	**Values**	χ^2	**Values**	χ^2
λ_1	4.611	6.635	45.948	6.635	51.03	6.635
λ_2	25.666	9.210	46.352	9.210	40.504	9.210

5. CONCLUSION

The application of the *RESET* test helps selecting the appropriate link function and help determining the accurate magnitude of regression coefficients for independent variable. Although, our diagnostic tools suggest that *PROBIT* link function is the best because it has minimum Log likelihood value whereas predictability of the model favors *LOGIT* Model. *RESET* test also suggest *LOGIT* as the best link function amongst all three link functions included in our analysis. The problem remains with the objective of the study. If one is willing to focus more on statistical inference, prediction or both together, the selection of the best link function even after utilizing the available diagnostic tools remains crucial. As *Huettmann and Linke (2003)* suggest that there exists a vacuum due to unavailability of widely accepted fitting diagnostic tools. Therefore, our work is just an illustration of our results taking in to account various link functions and applying different diagnostic tools for reliable inferences and correct predictions in case of binary response models or *GLMs*. The results highlight a strong need to further deepen the analysis in order to devise a strategy utilizing all available diagnostic tools and making a rationalized decision regarding the choice of the link function. Future research can be devoted to assigning some arbitrary weights to different diagnostic tools in order to arrive at a collective score for each link function and then selecting the one with the highest score.

REFERENCES

1. Breslow, N.E. (1995). *Generalized Linear Models: Checking Assumptions and Strengthening Conclusions*. Congresso Nazionale Societa' Italiana di Biometrica Centro Convegni S. Agostino. Cortona.
2. Broeck, J.V., Eleckels, R. and Massa, G. (1996). Maternal Determinants of Child Survival in a Rural African Community. *Internaational Journal of Epidemiology*.
3. Chowdhury, Q.H., Islam, R. and Hossain, K. (2010). Socio-economic Determinants of Neonatal, Post Noental, Infant and Child Mortality. *International Journal of Sociology and Anthropology*, 2(6), 118-125.
4. Davidson, R. and G. MacKinnon, J. (1982). Convenient Specification Tests for Logit and Probit Models. *Queen's Economics Department Working Paper* No. 514.
5. Dhakal, S., Chapman, G.N., Simkhada, P.P., Teijlingen, E.R. and Stephans, J. (2007). *Utilisation of postnatal care among rural women in Nepal*. London, United Kingdom.
6. Huettmann, F. and Linke, J. (2003). Assessment of Different Link Functions for Modeling Binray Data to derive sound Inference and Prediction. *ICCSA 2003, LNCS 2669* (pp. 43-48). Springer Verlag Berlin Heidelberg.
7. Prebigon, D. (1980). Goodness-of-link Test for Generalized Linear Models. *Appl. Statist., 29*, 15-24.
8. Sapra, S. (2005). Regression Error Specification Test (RESET) for Generalized Linear Models. *Economic Bulletin*, 3, 1-6.

APPENDIX

Variables	Logit Model		Probit Model		C-LOG-LOG Model	
	Coefficient	**P-Values**	**Coefficient**	**P-Values**	**Coefficient**	**P-Values**
Intercept	-1.186	0.252	-0.904	0.119	-0.856	0.317
AREA	0.286	0.039	0.700	0.038	0.164	0.119
DH1	1.469	0.075	0.786	0.081	1.108	0.120
DH2	1.002	0.248	0.517	0.278	0.731	0.325
OH0	0.433	0.023	0.243	0.029	0.322	0.018
CP1	-0.356	0.073	-0.204	0.082	-0.283	0.059
ME1	-0.487	0.002	-0.291	0.002	-0.457	0.000
ME2	-0.161	0.301	-0.110	0.235	-0.175	0.131
ME3	-0.155	0.399	-0.104	0.344	-0.116	0.383
IC1	0.723	0.070	0.438	0.056	0.498	0.098
IC2	1.093	0.013	0.648	0.011	0.803	0.015
IC3	0.717	0.195	0.438	0.170	0.517	0.228
HHS1	-0.275	0.064	-0.167	0.053	-0.206	0.058
CF0	-0.393	0.004	-0.233	0.003	-0.255	0.015
MAGE	-0.078	0.000	-0.038	0.000	-0.067	0.000
CEB	5.815	0.000	2.552	0.000	1.633	0.000
MAGE*CEB	-0.106	0.000	-0.114	0.000	-0.059	0.001
MAGESQ*CEB	0.001	0.000	.001	0.000	.0001	0.053

<u>RESET Test Calculation</u>

Log likelihood Functions

Like Functions	LOGIT Model	PROBIT Model	C-LOG-LOG Model
L_1	-950.966	-947.127	-964.899
L_2	-948.661	-927.153	-939.384
L_3	-938.133	-923.951	-944.647

Coefficients of Link Function Powers in Model

Coefficients	Logit Model		Probit Model		C-LOG-LOG Model	
	Coefficient	**P-Values**	**Coefficient**	**P-Values**	**Coefficient**	**P-Values**
γ_1	-0.075	0.031	-0.422	0.000	-0.508	0.000
γ_2	-0.053	0.000	-0.024	0.522	-0.114	0.000

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 135-140

TERMINAL EVALUATION OF PUBLIC SECTOR DEVELOPMENT PROJECTS: ANALYSIS & REMODELING

Sajjad Mubin[1], **Jamshed Hussain Sial**[2] and **Mudassir Ahmed**[3]
Directorate General Monitoring & Evaluation, Planning & Development Department,
Government of the Punjab, Lahore.
Email: [1]sajjadmubin@yahoo.com
[2]Jamshed369@hotmail.com
[3]mudassir_a1@yahoo.com

ABSTRACT

Development means change in quality of people's life. Evaluation of development projects ought to assess this change, White (2009). Since July 2008, DGM&E has conducted terminal evaluation of 85 projects. In this paper, a descriptive analysis of the projects has been made. It is observed that in most cases projects failed to achieve the objectives and they were revised for time, cost and scope. This underscores lack of proper pre-execution project planning and learning from lessons learnt of evaluation. In fact, evaluation recommendations are also set aside. On the contrary, planning is the most critical stage because here the risks are at the highest. Feedback from evaluation results must be incorporated in planning, executing and implementing development schemes. Therefore, institutionalization of feedback, throughout a project life cycle, is immensely important for increasing efficiency and effectiveness of public development interventions.

KEY WORDS:

Project Terminal Evaluation; DGM&E; Project Life Cycle; Good Governance and Development Interventions.

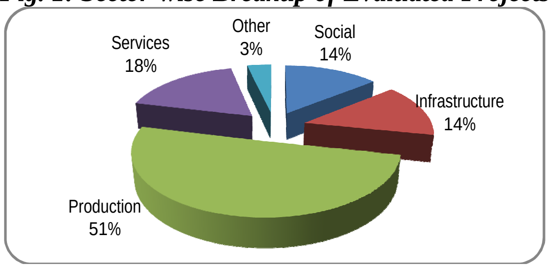
1. INTRODUCTION

The word development may be used in numerous contexts. Dimensions of development programs envelop various aspects ranging from economic and social betterment of the masses, institutional development and infrastructure improvement. In all these contexts, it denotes some kind of change. It can be viewed as a process of societal change that generates some perceived benefits for people, or as a state of perceived quality of life attained through such a process, Dale (2004). Fiscal strains, ever-rising expectations from ordinary citizens, accountability pressures from civil society and parliaments, desire for curbing and controlling corruption and other leakages, and burgeoning requirements from international donors provide a continuous impetus to governments for enhancing the quantity and quality of government services. However, constrained development budget leaves quality improvement as the most viable option. M&E is a bridge to reach this frontier and is an essential element of Good Governance. It results in improvement in physical effectiveness of the projects, Thomas (2010). Therefore, it is pertinent that M&E systems are developed, overhauled and strengthened.

2. ANALYSIS: A CASE STUDY OF 85 EVALUATED PROJECTS

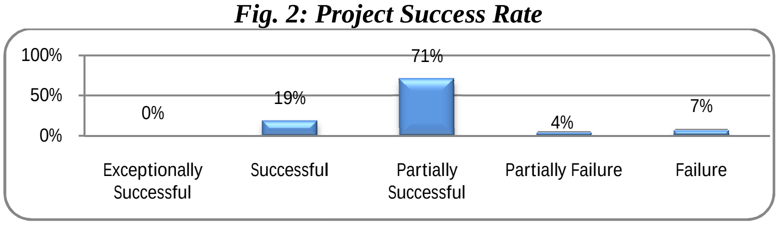
In 2010-2011, a target of 3128 development projects having a total investment volume of Rs. 637.5 billion (current year allocation Rs. 182 billion) has been set in Punjab. Social and Infrastructure sector constitute 80% development funds. However, **Fig. 1** shows that only 28% of the total evaluated projects belong to these sectors. This shows that uniformity in selection of projects for evaluation is not being maintained.

Fig. 1: Sector-wise Breakup of Evaluated Projects



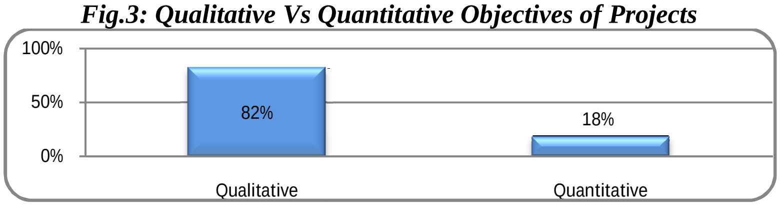
Development projects are initiated to achieve specific objectives, which if not attained, constitutes a waste of public money. However, it can be seen in **Figure 2** that in most of the cases the projects failed to achieve the desired aims.

Fig. 2: Project Success Rate



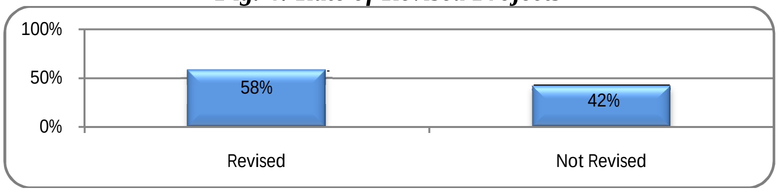
An important finding that throws into doubt even this meager achievement is that mostly the objectives of the projects are intangible and therefore immeasurable. This aspect is presented in **Figure 3**. In fact, in 70 out of total 85 evaluated projects more than 50% of the objectives were qualitative.

Fig.3: Qualitative Vs Quantitative Objectives of Projects



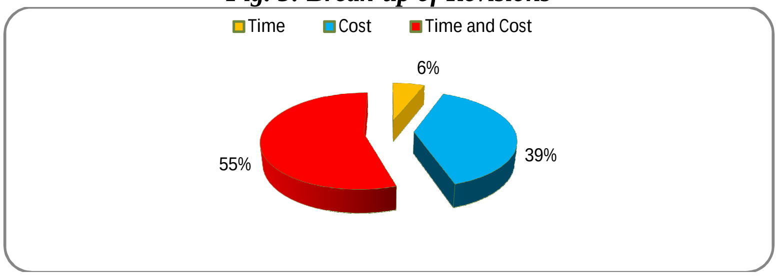
D.P. Gregg, former US Ambassador to South Korea, said that “Timing is everything in politics, war and peace”. The same is true for development projects. The old adage ‘it’s better to be late than never' cannot be applied to projects because in most cases the whole purpose of a project would be lost if it is not completed in time. **Figure 4** highlights this issue.

Fig. 4: Rate of Revised Projects



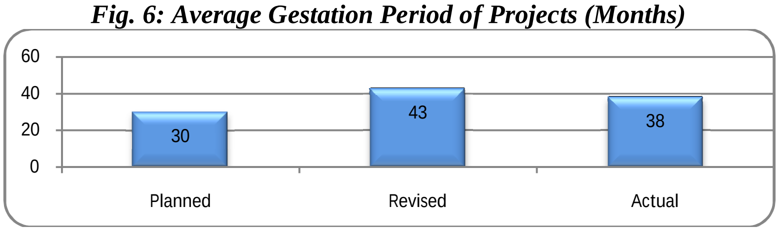
On the other hand, **Figure 5** gives further breakup of the revised projects into time revised, cost revised, and both time and cost revised.

Fig. 5: Break-up of Revisions



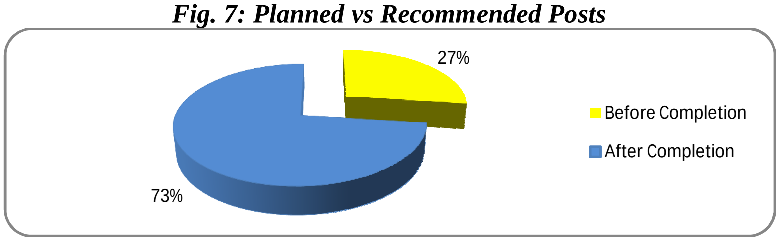
In fact, there was on average an increase of 8 months in gestation period of the projects as evident in **Figure 6**.

Fig. 6: Average Gestation Period of Projects (Months)



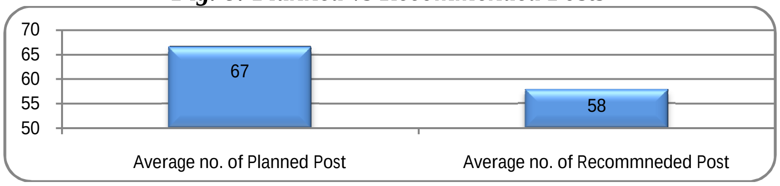
It was also observed almost three fourth of the evaluation requests were made after the completion of the Projects (**Figure 7)**

Fig. 7: Planned vs Recommended Posts



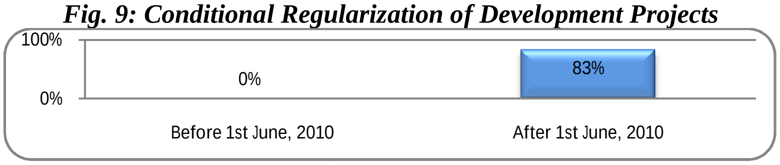
Efficiency is essential for achieving the best out of scarce resources. However, it was observed that extra resources are used, especially human resource, as in **Figure 8** which shows that recommended posts were always less than the planned project posts.

Fig. 8: Planned vs Recommended Posts



No 'Impact Evaluation' exercise or lack of scrutiny of PC-V proves to be the last nail in the coffin of M&E information importance. DGM&E has started conditional regularization of those development projects which have failed to achieve the desired objectives i.e. the projects are to be re-evaluated after 1 year to 3 years with respect to the problems indentified. As shown in **Figure 9,** out of 21 evaluated projects since June 2010, 17 projects have been regularized on conditional basis because they desired objectives have not been achieved. However, this is a short term measure.

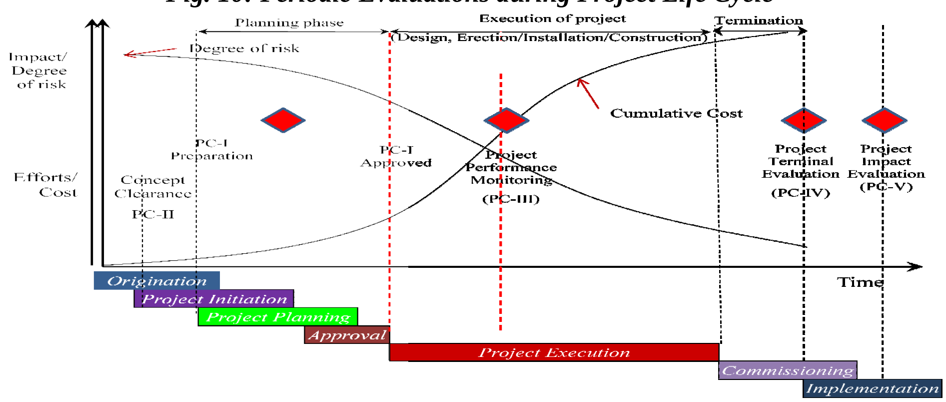
Fig. 9: Conditional Regularization of Development Projects



3. REMODELING EVALUATION: INSTITUTIONALIZATION OF FEEDBACK

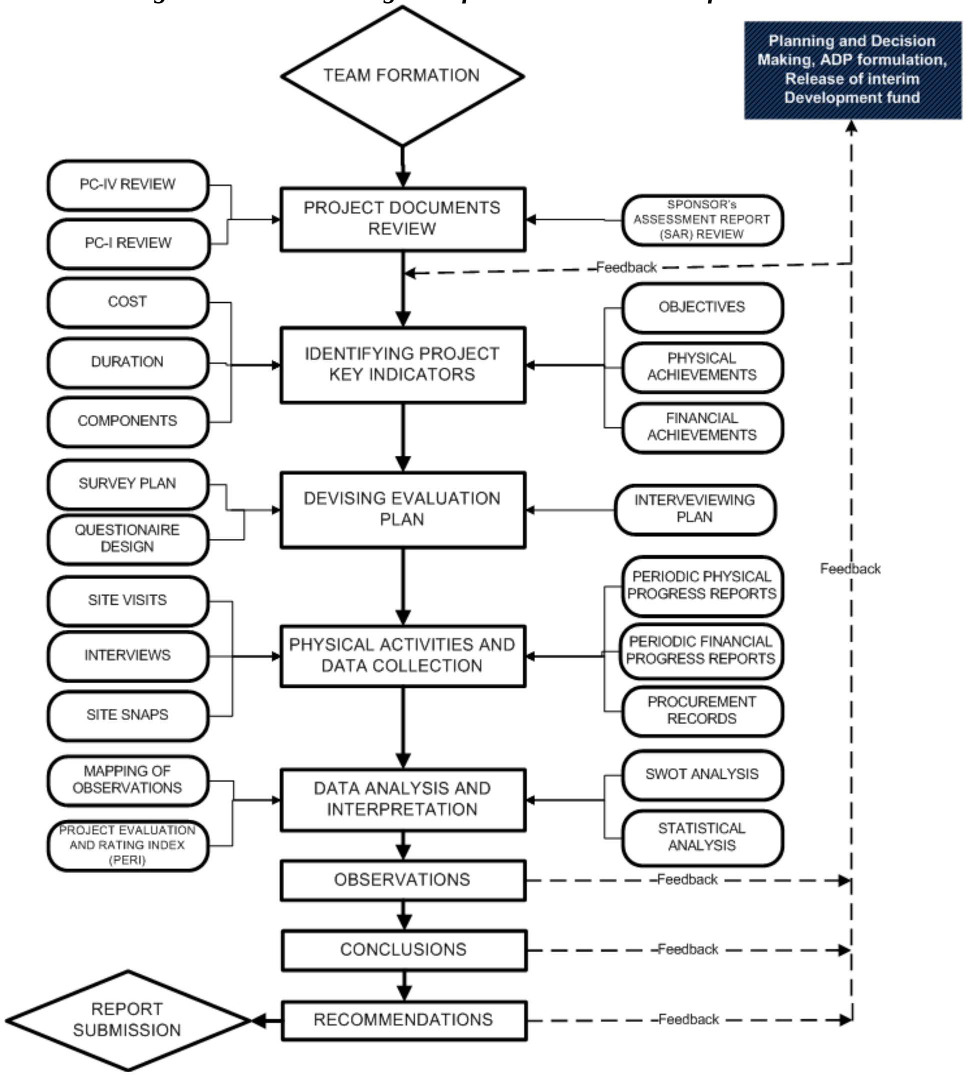
As a long term strategy, the evaluation mechanism ought to be remodeled. Periodic and comprehensive evaluations should be started so that the iron is stricken when it is hot. In **Figure 10**, proposed periodic evaluations are shown by red diamonds. Impact evaluation should be initiated, Chambers et al. (2009).

Fig. 10: Periodic Evaluations during Project Life Cycle



Another important aspect is that those periodic feedbacks ought to be incorporated both in planning and evaluation of projects (**Figure 11**).

Fig. 11: Schematic Diagram of Institutionalization of Feedback



4. CONCLUSIONS AND RECOMMENDATIONS

The significance of this study can be ascertained from the fact that data maintenance at such scale and data analysis on the basis of that data has never been carried out in Pakistan, let alone Punjab. The study has removed the dust from the reality of public sector development project management in Punjab. On the basis of this study the remodeling of evaluation has been proposed by institutionalization of feedbacks periodically. In additions, the following recommendations are put forth:

1. Projects for evaluation should be selected by taking into account sector wise allocation of development fund so that.
2. During appraisal and planning of projects, it should be made mandatory that the projects are quantifiable and measureable so that proper M&E can be conducted.
3. Timely completion of projects ought to be disciplined.
4. The time for evaluation should be included in the project gestation period.
5. Proper human resource management techniques should be carried out during the appraisal stage so that appropriate level of HR requirement is ascertained.

BIBLIOGRAPHY

1. Francois Bourguignon and Francisco H.G. Ferreira (2002). Ex Ante Evaluation of Policy Reforms Using Behavioral Models. The Impact of Economic Policies on Poverty and Income Distribution: *Evaluation Techniques and Tools*, 123-142.
2. Howard White (2009). Theory-Based Impact Evaluation: Principles and Practice. *International Initiative for Impact Evaluation Working Paper* 4. International Initiative for Impact Evaluation (3ie).
3. Robert Chambers, Dean Karlan, Martin Ravallion, and Patrica Rogers (2009). Designing Impact Evaluations: Different Perspectives. *International Initiative for Impact Evaluation Working Paper* 3. International Initiative for Impact Evaluation (3ie).
4. Reidar Dale (2004). *Evaluating Development Programmes and Projects*. Second Edition Thailand: Sage.
5. Vinod Thomas (2010). *Evaluation for Greater Development Effectiveness*. Korea Development Institute Conference on Economic Development and Impact Evaluation, Seoul, South Korea.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 141-144

MAJOR DETERMINANTS OF POVERTY IN PUNJAB: LOGIT & PROBIT REGRESSION ANALYSIS

Majida Jawad[1], **Mudassir Ahmed**[1] and **Muhammad Usman Sikander**[2]
[1] Directorate General Monitoring & Evaluation, P&D Department, Govt. of the Punjab, Lahore, Pakistan.
Email: majidajee@yahoo.com; mudassir_a1@yahoo.com
[2] Center for Research in Economics & Business, Lahore School of Economics, Lahore, Pakistan. Email: usman.lse@gmail.com

ABSTRACT

In this research paper the two regression models, logit and probit have been developed to find out the real factors of poverty in rural and urban areas of *Punjab*. Both the models have been verified on the basis of diagnostic tools for binary regression. The analysis reveals that the Logit model is the best fitted model when compared to the probit model for determining the factors of poverty in Punjab. Household Head's profession and possessions of agriculture and non-residential land are more likely factors to help exiting the poverty trap. The household composition of proportion of below 15 years of aged persons positively affects the possibility of entering the poor household group. The empirical results suggest considerations for the policy makers and provide poverty dynamics.

KEY WORDS

HIES (2004-05); Household; Poverty; Logit; Probit and Punjab.

1. INTRODUCTION

Pakistan being a developing country is the sixth most populous economy in the world. The depth and nature of poverty in post-conflict Pakistan is not known as the civil unrest prevented data collection and analysis in that period. In Pakistan, the process of writing a Poverty Reduction Strategy Paper (PRSP) had the main goal of understanding its causes and identifying policies able to successfully achieve reduction in poverty. After the September 2000, when 189 member countries of United Nations signed the Millennium Development Goals (MDGs), Pakistan also agreed to reduce the population in poverty by half till 2015. The purpose of this paper is to examine the cause and effect relationship between the poverty and other socio-economic indicators in Punjab, Pakistan by using HIES 2004-05 data set. This objective is achieved by means of both descriptive and regression analysis. The second part of the paper contain the relevant literature on poverty, third part discuss the results of data analysis. While fourth and last part concludes the paper and provides some policy recommendations respectively.

2. LITERATURE REVIEW

Poverty affects many aspects of human life like economic, social, physical, moral, psychological conditions etc. As a result, there are different approaches to conceptualize the poverty. Many researchers have proposed several theories on the poverty and utilized different socio-economic indicators as key determinants of poverty un urban and rural settings for different countries. Sonja and Lindsay (2007), Mok et al. (2007), Khalid et al. (2005), Mulat et al. (2003), Bruck (2001), Geda et al. (2001), Oyugi (2000), Bigsten et al. (1999), Daff and Jollifee (1999), Malik (1996), Jalan and Ravallion (1998) have applied different statistical methods/techniques (e.g. Multiple Linear Regression Model, Binomial Logistic Regression Model, Multinomial Logit Regression Models, Ordered Logit Models, Log-Linear Regression Model) to determine the effects of explanatory variables for measuring likelihood of poverty. In all these research papers household composition, education, work, assets, remittances, total adult equivalent consumption of the household on food and non-food items, female and male adults in a household, higher proportion of children under 15 years of age, education of the members of the household and gender of the head of household, household welfare, the refugee status and maternal education, consumption expenditure and income, landholding area, livestock unit, the proportion of household members who are able to read and write, household size, source of water for household use and off farm employment to be important variables in determining the household's welfare, female headship, average year of schooling, parent having completed primary schooling and unemployment, owned land, value of livestock and sector of employment, income per capita, education and dependency ratio, remote mountainous areas, limited transport, education levels and, health status of household members were found the major source of poverty.

3. ANALYSIS AND MODEL BUILDING

For the present study, data were taken from HIES 2004-05 survey collected by Federal Bureau of Statistics, Pakistan. Since the research question is specific to the province of Punjab, we just focused on the Punjab data and included some 5875 households from survey. . Initially, thirty four independent variables were included in the regression model to find the best sub set of the independent variables. Using backward elimination procedure, we were able to short list the key determinants of poverty in Punjab on a binary response dependent variable for Logit and Probit Regression Model. The diagnostic tools for accessing the best fitted model between the two are Hosmer Lemeshow Goodness of Fit Statistics, Deviance and Sum of Square of Residual, graphical view of Logit link function which suggest that the logit model fits the data well as compare to Probit models however results of regression coefficients show similar results.

The results from this study shows that the coefficients of age of head of household (ageHHhead) and square of age of head of household (SqageHHhead), OC and SC does not have a significant effect on the household to be a poor in all the models. The coefficient for female headship (female HHH) and the coefficient of number of persons whose age is below 15 years (below15s) are significantly different from zero at 5% level of significance only in Punjab and Rural model. Similarly educational variables emerge as a strong determinant of a household to be poor, results indicates that the all variables related to the level of education are negatively associated with the household to being a poor and by improving the level of education of the members of the household their marginal effects of household fall into poverty become low in all the models. The other variables that is found to be correlate with poverty in overall Punjab model, is the EMP, SE, UFH, agriland and nonresident. The coefficients of these variables are negatively significant different from zero at 5 % level of significant that shows an inverse relationship with the household to being a poor in all the models. Participation of household's members working in government sector is negatively and coefficient of number of goats (Goat) is positive associated with poverty at 5 % level of significant in Overall Punjab and Urban model. AFF shows a positive association while coefficient of number of poultry (poultry) was negatively associated with poverty in Punjab and rural model.

4. CONCLUSIONS

All the variables selected for the research work are significant to verify the key determinants of poverty in the area. The analysis on poverty dynamics highlights that the poverty predictors may revolve around the demographic characteristics of the household and the household head. To lend scientific and empirical basic to the determination of priorities for the improvement in welfare and alleviate poverty in *Punjab*, we present some policy implementations in the following lines:

- The structural reform policies that emerge as crucially important relating to education, employment sector, occupational type, ownership of livestock and household assets.
- The breakdown according to household size confirms that poverty in Punjab is relatively widespread in larger households which results the sluggish improvement in living standards of the people along with their worse livelihood conditions. Rapid growth of the population relative to resource generation is the main cause of this dilemma because high population growth adversely affects the per capita income. Moreover, government should make efforts to control the increasing population.
- A very remarkable finding of this study is the negative relationship between the female headship of household and the poverty status. Results exhibit that a household headed by female have less probability to fall into poverty. According to the disaggregated data, ratio of women is 48.17 percent of the total population of Punjab in 2007; therefore, government should take initiative to create women friendly environment and encourage them as a diligent workforce.
- A significant finding of the present study is that a large proportion of households have low education level. Education at the higher secondary level emerges as a powerful explanatory variable in the analysis of poverty dynamics. Primary education would, thus, appear to in-sufficient medium term national goal for poverty reduction. Functional literacy for the household's head and higher secondary education may be the considerable part of Pakistan's national goals if a serious dent is to be made in incidence of poverty.
- Government may conduct authentic research to locate the deprived areas along with the level of education required for household' members of those areas.
- The employment status of the household's member is critical for reducing the incidence of poverty in Punjab. It is found that the members of poor household are generally being employed in low paid informal sector; therefore, government should take concrete steps to increase the wage level especially in private sector organizations and informal employment sector.

- The deeper analysis of the employment sector reveals the fact that self business and self employment along with foreign investment should be encouraged as these may change the living standard of the people and directly impact the poverty.
- Government should make policies to control escalation in prices of eatables which would automatically increase the purchasing power of the people and make them possible to fulfill other necessities.
- Engagement of households in government sector is another important variable that determines households' welfare. Commencing of the new public sector projects will generate more employment opportunities that can help in improving the living standard of the people of the province. Moreover, there is a need to make the merit of selection more transparent to avoid anomalies in the recruitment system.
- Special importance is required towards the services like social and personal services along with working in the agriculture, forestry and fisheries. Agriculture financial statement for about 23 percent of GDP and employs about 47 percent of the labor force but still this sector in ignored by the policy makers. Therefore, if a serious dent is required for poverty alleviation then this sector should be strengthened.
- A large proportion of population of Punjab, particularly in the Southern region is involved in livestock related activities which is relatively ignored area in providing the employment opportunities by the government as well as other non government organization (NGOs). Therefore, government along with NGOs should make efforts jointly on war footing basis to get better with this sector so that the people engaged with this profession may improve their living standards.

5. REFERENCES

6. Sen, A. (1976). An Ordinal Approach to Measurement of Poverty, *Econometrica* 44.
7. Jalan, J. and Ravallion, M. (1998). Determinants of Transient and Chronic Poverty: Evidences from China. *World Bank Policy Research Working Paper*, 1936. The World Bank.
8. Sonja Fagernas and Lindsay Wallace, (2003). Determinant of poverty in Sierra Leone, *Economic and Statistics Analysis*, ESAU Unit, Working paper 19.
9. Khalid, U. Shahnaz, L. and Bibi, H. (2005). Determinants of Poverty in Pakistan: A Multinomial Logit Approach. *The Lahore Journal of Economics*, 10(1), 65-81.
10. Alemayehu Geda, N. de Jong, G. Mwabu and M.S. Kimenyi (2001). *Determinants of Poverty in Kenya: A Household Level Analysis*. The Institute of Social Studies and the Kenya Institute for Public Policy Research and Analysis (KIPPRA), Kenya, Nairobi.
11. Datt, G. and Jolliffe, D. (1999). Determinants of Poverty in Egypt: 1997. *Discussion Paper No. 75*, Food Consumption and Nutrition Division, IFPRI Washington DC.
12. Malik (1996). Determinants of Rural Poverty in Pakistan: A Micro Study. *The Pakistan Development Review*, 35 2, 171-187.
13. Ravallion (1992) Poverty Comparisons: A Guide to Concepts and Methods, *World Bank, LSMS, Working Paper* 88.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 145-150

RISK MANAGEMENT AND MITIGATION TECHNIQUES IN ISLAMIC FINANCE – A CRITICAL REVIEW

Muhammad Mazhar Manzoor[1] and **A.R. Zaki**[2]

[1] Department of Business Administration, FUUAST, Karachi, Pakistan. Email: mmazher@gmail.com

[2] Department of Business Administration, University of Karachi, Karachi, Pakistan.

ABSTRACT

After the decades movement Islamic banking has reached on its peak and going on day by day due to its religious foundation among Muslims all over the world specially during the last few years. The study relate to Islamic finance which in their objectives and operations, working on Qur'an principles. Islamic financing is thus set apart from 'conventional financing', which have no such religious preoccupations. The aim of the study is "to indicate risks facing by Islamic finance institutions and how can they manage these risks and how can they mitigate risks in the light of Shariah principles? This study grew up from secondary data and analyzes the importance of effective risk management in Islamic financing. Also study analytically confer that the Islamic banks become more efficient if they are managing risks effectively.

INTRODUCTION

Financing in an Islamic way for specific system creates real assets. This is true even in the case of Murabahah and leasing, despite the fact that they are not believed to be ideal modes of financing and are often criticize by for their being close to the interest-based financing in their net results. It is known, on the other hand, that interest – based financing does not necessarily create real assets. Therefore, the supply of money through the loans advanced by the financial institutions does not normally match with the real goods and services produced in the society, because the loans create artificial money through which the amount of money supply is increased, and sometimes multiplied without creating real assets in the same quantity. This gap between the supply of money and production of real assets creates or fuels inflation. Finally the financing in Islam is backed by physical assets; it is always matched with corresponding goods and services.

Risk management is a process of identify, analysis and elaborate the risk and determining how to best handle such exposure. Although, it involves identification, measurement, and monitoring, reporting and controlling risks to do this, firms made policies, procedures or/and practices to minimization or elimination of unacceptable risk. The focus of risk management is the identification and treatment of risk which prevent the financial institutions in our case example from devastating impact. Its goal is to add maximum and sustainable value to all the activities of the organization. It also organizes the understanding of the possible upside and downside of all those factors that can affect the chance of success and lessens both probability of failure and the uncertainty of achieving the organizations' objectives. Study shows that how banks are more effective by good risk management. IBS has high and rapid growth rate and many conventional banks just converted or are about to convert. Islamic banks are facing similar but not identical risk as conventional banks.

LITERATURE REVIEW

Shariah provides guidelines for aspects of Muslim life, including religion, politics, economics, banking, business, and law, Michael Silva (2006). Under shariah, money is not considered an asset class because it is not tangible and thus, may not earn a return, The National Bureau of Asian Research (2008). Therefore Islamic banking emerged in the Islamic economists' literature as a financial system based exclusively on profit-and-loss sharing, which was argued to be more equitable and stable, Chapra (1996) and Siddiqi (1983).

Now Islamic banking has developed phenomenally in recent years not only in the Muslim world, but also in the West (Wilson, 2007). And according to Stephen Timewell (2009), Islamic banking industry may be able to strengthen its position in the international market as investors and companies seek alternate sources of financing. The earnings of profits or returns from assets are permitted so long as the business risks are shared by the lender and borrower, Andreas Jobst (2007).

The survival and success of financial organization depends on the efficiency in with they can manage its risks, hence, risk management is one of the critical factors in providing better returns to the shareholders (Akkizidis and Khandelwal, 2008). Khan and Bhatti (2008) observed that Islamic banks face another crucial challenge to improving their risk management strategies and corporate governance. Risk management is increasingly recognized as being concerned with both positive and negative aspects of risk (IRM, AIRMIC and ALARM, 2002). Academic interest in the study of risk management in Islamic banking probably accelerated from scholars such as Khan (1997), Vogel and Hayes (1998), Obaidullah and Wilson (1999), and Khan and Ahmad (2001). Gallati (2003) provides the stage of shareholders value objectives that are expected to be met by improving on risk management sophistication. Effective risk management requires a reporting and review structure to ensure that risks are effectively identified and assessed and that appropriate controls and responses are in place (IRM, AIRMIC and ALARM; 2002). The high perception of risks may be an indication of the low degree of active risk management due to the absent of risk control through internal processes and control, especially in the case of operational risk (Iqbal and Mirarkhor, 2007). In relation to commercial banks' practice of risk management, Al-Tamimi (2002) found that the UAE commercial banks were mainly facing credit risk. The study also found that inspection by branch managers and financial statement analysis are the main methods used in risk identification. The main techniques used in risk management are establishing standards, credit score, credit worthiness analysis, risk rating and collateral. Khan and Ahmed 2001 conducted a survey of risk management practice and found that on average the lowest percentage is on the measuring mitigating and monitoring risk that is 69% score as compared to risk management policies and procedures that is 82.4% and internal control of Islamic banks that is 76.7%. A comprehensive explanation of risk management in Islamic Banking are made by Akkizidis and Khandelwal (2008) covering the aspect of risk management issues in Islamic financial contracts, Basel II and Islamic Financial Services Board (IFSB) for Islamic financial risk, and examining the credit, market and operational risk management for IBs. They also explain the unique mixes or risk for each financial contracts in IBs. BCBS (2001) defines financial risk management as a sequence of four (4) Processes: (1) the identification of events into one or more broad categories of market, credit, operational and other risks into specific sub-categories; (2) the assessment of risks using data and risk model; (3) the monitoring and reporting of the risk assessments on a timely basis; and (4) the control of these risks by senior management. In addition, as suggested by Al-Tamimi (2002), in managing risk commercial banks can follow comprehensive risk management process which includes eight (8) steps: exposure identification; data gathering and risk quantification; management objectives; product and control guidelines risk management evaluation; strategy development; implementation; and performance evaluation.

RISK MANAGEMENT PROCESS

Risk management process comprises of several steps including identification, measurement, monitoring, reporting, mitigation & control etc.

In support, State Bank of Pakistan gave 15 Guiding Principles on managing risk for Islamic banking institutions which divided into following pattern.

1. General (1 Principle)
2. Credit risk (4 Principles)
3. Equity investment risk (3 Principles)
4. Market risk (1 Principle)
5. Liquidity risk (2 Principles)
6. Rate of return risk (2 Principles)
7. Operational risk (2 Principles)

These principles are not radically different from those applicable to conventional banks. However, these are some fundamental differences:

- Emphasis on Shariah compliance
- 6 out of 15 principles make explicit reference to Shariah rules

VARIOUS RULES TO RISK MANAGEMENT

1. There is no return without risks, Rewards go to those who take risks.
2. Be transparent. Risk should be fully understood
3. Seek experience. Risk is measured and managed by people, not by mathematical models
4. Know what you don’t know. Question the assumptions made
5. Communicate. Risk should be discussed openly
6. Diversify-avoid concentration. Multiple risks will produce more consistent rewards
7. Show discipline. A consistent and rigorous approach will beat a constantly changing strategy

8. Use common sense. It is better to be approximately right, than to be precisely wrong
9. Return is only half of the equation. Decisions should be made only after considering the risks and returns of the possibilities
10. Oversight must be enterprise-wide. Risks cannot be managed in isolation

METHODOLOGY

This study grew up from secondary data and analyzes the importance of effective risk management in Islamic banking systems. The study is completely descriptive and based on secondary data grew from different sources. In this study we try to find out the importance of effective risk management in an Islamic banking system. To promote the study we used different statistical data to highlight the risk in Islamic banking system and grew up the data to explain the results of effective risk management.

To express the research hypothesis the data is grew from secondary sources that includes different books related to Islamic banking, articles from different magazines, research articles, research papers, web-browsing, annual reports and different reviews of literature. Common size analysis is used to propel the study and to express the desired statement. This type of analysis great value relative analysis in which financial statements converts into percentage or the values is compared. In this common size analysis data is extracted from balance sheet and income statement.

ANALYSIS AND DISCUSSION

Islamic banking has grown at a rapid pace in the last few years. Both the local and global Islamic Financial Services Industry (IFSI) shared and enjoy this prosperity. The main drivers or motivational factors of this tremendous growth are (a) faith based nature of Islamic Finance, (b) surplus liquidity in the market, and (c) the interest of world's influential financial centers. In Pakistan, this growth has resulted from the highly positive response of private sector to policies for promotion of Islamic banking and finance introduced after 2001 by the State Bank of Pakistan. Collectively these factors have contributed to evolution of new and innovative products by Islamic Financial Industry in tandem with a strong and Shariah Compliant regulatory framework. According to an estimate, total Shariah compliant assets worldwide have grown to about US$ 700 billion – with annual growth exceeding 10.0 percent during the past decade - and are projected to grow to US$ 1.6 trillion by 2012

By the review of literature it is quite clear that without risk management a bank cannot be efficient. To captures the market share and to attract the customers well and to provide highest profit to shareholders the effective implementation of risk management process is the first requirement.

To prove the statement "IBs that have good risk management are more efficient" we put common size analysis and analyze the balance sheet structure, profitability ratios, income statement and capital adequacy ratio. There are different banks running in the state of Pakistan. We will analyze two of them to show our research statement positive. The analysis of Meezan bank having (201 branches) and bank Islami having (102 branches) are as follows:

Rs. in million

Particulars	Bank Islami		Meezan Bank	
	2007	**2008**	**2007**	**2008**
Total Deposits	6,246	11,242	54528	70234
Total Assets	9,985	17,231	67179	85276
Total Financing	3963	6528	34576	39768
Total Investments	3,486	6827	10536	14287
Asset to Deposit Ratio	39.89%	52.31%	56.90%	51.62%
Return on Assets	-0.36%	0.04%	1.70%	.82%
Return on Equity	-1.14%	0.15%	18.39%	10.30%
Earnings Per share (Rs.)	-0.13	-0.12%	1.96%	1.26%
No. of Employees	1188	563	2205	3170
No. of Branches	36	102	100	166

From above it is clearly depicted that Meezan bank is efficiently increases its deposits and assets as compare to bank Islami. It means that the bank is providing attractive products to increase its capital. Because increase in capital decreases bankruptcy and efficient capital is the basic requirement of credit risk management. As we have discussed the types of risk, credit risk suffers 80% losses so to increase capitals introducing attractive products are necessary. After that there is a vast network of branches of Meezan bank as compared to bank Islami so that is why there are large no of customers who can access their accounts well.

If we focus the ratios then we conclude that the bank that have good risk management has good ratios according to industry average. Meezan bank's market value is greater than bank Islami it means that bank Islami needs to improve product development to attract customers. Earnings per share are also low as compare to Meezan bank.

After studying the risk management policies of both banks it is clear that Meezan bank has effective credit risk management committee and asset liability management committee(ALCO) which is responsible for approving monitoring and ensuring that financial transactions that are within the acceptable risk rating criteria, policies, procedures and manuals are in place and authorities have been appropriately delegated to ensure credit quality, proper risk –reward trade off, industry diversification and periodic credit reviews. ALCO is responsible for reviewing and recommending all market risk and liquidity risk policies and ensuring that sound risk measurement systems are established. The bank applies stress testing and values at risk (VAR) techniques as market risk management tools. Liquidity management is done through cash flows matching investments in commodity Murabaha, Sukuks and placements in foreign exchange. Operational risks are measured and managed timely and effective manner through enhanced operational risk awareness, dual checks and improving early warning signals. The bank has developed effective manuals and procedures necessary for mitigation of operational risk. There is also a bank internal audit department who identifies bank's weaknesses and implements regulatory standards.

CONCLUSION

Every social system has its own economic system. Islam being a comprehensive and distinct religion possesses a corresponding economic system of its own. Islamic economics is fast developing into a different and distinct paradigm of economics. Therefore, a number of Islamic financial institutions have emerged in various Muslim as well as some non-Muslim countries reaching finally at long last to our country. Hence the development of Islamic or interest free banking and Islamic finance has been gaining momentum on a global scale for the last few years.

Islamic banking system is more sensitive to risk therefore it needs efficient risk management policies and procedures. The risk management function has now taken its root within the financial institutions on worldwide basis. Every bank should perceive the risk management to become effective and to remain active. The review of policies is also required for effectiveness and for sound banking system. Our results concluded that the risk management process is the key requirement for efficient banking systems.

REFERENCES

1. Wilson, R. (2007). Islamic investment in the UK. *Business Islamic*, 1(12), 68-72.
2. Hassan, M.K. and Dicle, M.F. (2006). *Basel II and capital requirements for Islamic banks*. Working paper, Department of Economics and Finance, University of New Orleans, New Orleans, LA.
3. Khan, T. (1997). *An analysis of risk sharing in Islamic finance with special reference to Pakistan*. Ph.D. dissertation (unpublished), Loughborough University, Loughborough.
4. Vogel, F.E. and Hayes, S.L. (1998). *Islamic Law and Finance: Regulation, Risk and Return*, Kluwer Law International, The Hague.
5. Obaidullah, M. and Wilson, R.J.A. (1999). Risk sharing and management in infrastructure financing: an Islamic evaluation. Paper presented at the *International Conference on Islamic Economics in the 21st Century*, jointly organized by International Islamic University, Malaysia and IRTI of IDB, Kuala Lumpur, August 9-13.
6. Khan, T. and Ahmad, H. (2001). *Risk management: an analysis of issues in Islamic financial industry*. IRTI Occasional Paper, Jeddah.
7. Khan, M.M. and Bhatti, M.I. (2008). Development in Islamic banking: a financial risk allocation approach. *Journal of Risk Finance*, 9(1), 40-51.
8. Al-Tamimi, H. (2002). Risk management Practices: An Empirical Analysis of the UAE Commercial Banks., *Finance India*, XVI(3), 1045-1057.
9. Iqbal, Z. and Mirakhor, A. (2007). *An Introduction to Islamic Finance: Theory and Practice*. John Wiley & Son (Asia) Pte Ltd.
10. BCBS (2001). *Consultative Document: Principles for the management and Supervision of Interest Rate, Risk*, Bank for International Settlements.
11. Harker, P.T. and Zenios, S.A. (2000). *Performance of Financial Institutions: Efficiency, Innovation, Regulation*. Cambridge University Press.
12. IRM, AIRMIC and ALARM (2002). *A Risk management Standard*.
13. Akkizidis, I. and Khandelwal, S.K. (2008). *Financial risk management for Islamic Banking and Finance*. Palgrave Macmillan, First Edition.
14. Chapra, U. (1996). What Is Islamic Economics? *IDB Prize Winner's Lecture Series*, No. 9, Islamic Development Bank, Jeddah.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 151-157

POPULATION GROWTH, POVERTY AND SOCIAL EXCLUSION: CHALLENGES OF POOR URBAN LIVELIHOOD – A CASE OF KARACHI

Muhammad Mazhar Manzoor[1], Sanober Shaikh[2], Muhammad Ibrahim[1], Rabia Essa[1], Muhammad Mohiuddin[1] and **Syed Ahsan Zafar[1]**

[1] Department of Business Administration,
Federal Urdu University of Science and Technology,
Karachi, Pakistan. Email: mmazher@gmail.com

[2] Department of Business Administration,
University of Sindh, Jamshoro, Pakistan.

ABSTRACT

Objective:

In recent years, Urbanization has become continuous and ongoing phenomena for both developed and developing countries, more than half of the world population now living in urban areas. Pakistan has also experienced unprecedented urban growth. Due to declining reliance on agricultural sector, the rural population continued to shifting to cities and make dominant contribution in urban population. These migrated peoples usually add in low income population and proportion of these persons living in slums where conditions are more deplorable, with high rates of malnutrition, infectious diseases, and exposure to violence because slum residents are often systematically barred from opportunities, security, capacity and empowerment that would facilitate them to have control over their lives and health. The projected urban population growth of year 2010 is 39.89% and it will become 49.45% in 2030. If the issue of slums is not addressed immediately then the situation will have catastrophe results. This paper evaluates planning standards and administrative procedures to reducing unplanned development and making a positive impact on dealing with urban poverty. It is recommended that for future policy making Governments must work to draw up proper Urban Development Plans together with the incorporation of communities and the private sectors in land delivery, and the devolution and reformation of procedures, increasing the assets of the urban poor, and provide access to basic health care, water and proper sanitation at subsidize rate and promote return migration on large scale.

Data source:

This study is based on secondary sources of published data and may be jerk down by utilizing existing surveys, open and closed end questionnaires etc. While Primary data source for this basal study is library based,

Methodology and Policy Relevance:

For this study we restrict population size to few informal Settlements of Karachi while random and stratified sampling is use. This Study is helpful to recognize the factors of increasing unplanned development and livelihood challenges of poor urban peoples and meanwhile we opt such policies on national level which can overcome the livelihood challenges.

KEYWORDS

Urban Population Growth; Livelihood, Poverty; Unplanned Development; Social Exclusion.

1. INTRODUCTION

1.1 Background of the Study

The rapid growth of cities is a common and persisting demographic phenomenon in most of the developing countries including Pakistan. This growth has led to an increase in the degree of urbanization. In Pakistan, for example, the proportion of total population living in urban areas, commonly referred as the level of urbanization, has increased from only 32.5% in 1998 to about 39.89% in 2010. Although within the Asia-Pacific region, based on both the level of urbanization and urban growth, Pakistan is grouped with countries having moderate level of urbanization, it has the highest share of population living in urban areas among the South Asian countries. It has also been projected that about half of Pakistan's population will be living in cities by the year 2030.

Poverty in Pakistan is largely a rural phenomenon, and lack of a stable source of income, landlessness, skewed distribution of land, droughts and less economic opportunities are the main causes of rural poverty. More than half of the rural households in Pakistan are landless. These factors encourage rural peoples to migrate. Mostly these rural migrants add into low income population of urban areas and unable to afford proper housing. Allocating resources for these migrants is become a major challenge for urban management and slow industrialization also make critical the situation. That's the result existing resources of urban localities are also use by increasing proportion of population that put pressure on urban areas infrastructure and civic institutions and create serious problems of economic and social crumbling and environmental deprivation. Also generate problem of housing, which increase the cost of urban property and causes high rent that's the reason large number of peoples have to sleep under bridges and also large proportion of population living in slums. Nations Agency UN-HABITAT defines a slum as a heavily populated urban area characterized by extremely poor housing and squalor, usually inhabited by the very poor or socially disadvantaged. Unemployment and lack of job have given momentum of widespread of poverty in urban areas and increasing population growth rate put extra burden on it. Poverty has various dimensions in terms of income, health, education security and social inclusion. Miscellaneous factors such as inflation, huge defense expenditure, foreign debt-dependence, backward agricultural instruments, slow industrialization, lack of modern technologies, load shading, corruption, terrorism, insufficient government, illiteracy and lack of will and determination on the part of the public etc are also responsible for widespread poverty.

The urban poor live in often awful conditions, which in some instances are worse than those in the rural areas that so many of them left behind. Women's status, youth issues, and reproductive health concerns all take on a different complexion in the context of urban slums and informal settlements. Problems of social and economic exclusion, HIV/AIDS and violence are more pronounced in these areas than in rural or more prosperous urban areas.

1.2 Significance of the Study

This study is helpful to understand the causes of migration and recognize the factors of increasing unplanned development and livelihood challenges of poor urban peoples and meanwhile such policies on national level which can overcome the livelihood challenges.

1.3 Variables

There are several pull and push factors that discourage rural peoples to stay in their homeland and encourage migrate towards cities. Here highlight some push and pull factors that cause of cities growth, poverty and social exclusion in urban areas.

<table>
<tr><th>PUSH FACTORS</th><th>PULL FACTORS</th></tr>
<tr><td>

To earn higher income
Better social and welfare services
Improve access to courts for justices
Better Paid Jobs
Better economic opportunities (that attract rural peoples towards the act of migration)
Higher the rate of poverty reduction among the migrant family and individual attract others rural peoples to migrate
Psychological attraction
Personal development
Strengthening social fabrics and building personal relations
Better living standard
High return on investment
Improved access to services such as medical and health care services
Better Health care and medical facilities
Carrier building opportunities
Flexible work environment
Cultural and re creational activities
Number of Educational institutes and vocational training centers

</td><td>

Machination of Agriculture (Such as agricultural sector is not showing any sign of improvement)
Shortages of non agricultural occupation
Shortages of educational institutes (like school, collages and other vocational training centers)
General lack of economic opportunities
Difficulty in access to Medicare and welfare services.(less % of doctors work in villages)
Less of employment opportunities
absence of industries
Less of employment opportunities
Lack of personal development opportunities
Lack of re creational opportunities
Jagirdari System
Less income level
Poor social provision
Working pressure group
Unfavorable situation (such as warfare or political problem or religious persecution)
Natural disaster (such as flood, drought and recent life threatening epidemic of HIV/AIDS)
Inadequate infrastructure
High poverty rate
High rural population growth
Less privileged standard of living
Due to non fertile land
Resettlement due to dams

</td></tr>
</table>

2. LITERATURE REVIEW

In Pakistan 10 million people or 8% of the population of Pakistan consisted of internal and external migration. (Population and Housing Census of Pakistan, 1998); the factors leading to high and rapid urbanization in Pakistan, divided into two components. First, the factors that create hurdles, in making the local conditions conducive and for better and prosperous life, are less fertile area, less production, no availability of health facilities and sanitation problems, lack of quality education, no availability, on- existence of social system based on justice and equality. According to weighted-Matrix approach the Urban- Rural population projection indicate that the urban population of Pakistan in 1998 with 32.5% of the total is estimated to rise to 50% by the year 2030. Province-wise population projection reveals that N.W.F.P is the lowest urbanized province of Pakistan, having 16.87% urbanization according to 1998 census. However, the proportion urbanized is expected to rise and become 41.36% by the year 2030. Baluchistan is another low urbanized province, the census indicate that 23.89% of the people reside in the urban area. The projected urban percentage is 45.56 by 2030, showing a rise of 22 percent Punjab where 31.267% population is urbanized according to 1998 census is projected to 50.07% urbanized by 2030. Sindh province currently has the highest proportion of urban population, which is 48.75% according to 1998 census. This figure will raise by 12 percent, which means the projected level of urbanization for the year 2030 is 60.73%, the pace of urbanization in the low urbanized provinces like N.W.F.P and Baluchistan is considerably higher, while it is low in the highly urbanized province of Sindh. (J. Bahrawar et al. 2008).

In-migration play a key role in economic development of rural peoples and as well as In-migration can modify the conditions of supply of entrepreneurship in those areas where there is a strong and favorable demand side condition and In-migration can contribute to the process of economic development through entrepreneurship and increased rural urban interdependencies. This means that the role of new arrivals in the countryside is in the main cumulative, rather than transformational. (Kalantaridis Christos, 2010). At the same time, like the people of so many low- and middle-income nations, Pakistani citizens have sought work abroad, and in the 1970s large-scale labor migration to the Middle East began in earnest. Remittances have since become an important component of the national economy and of the livelihoods of many households. These complex and substantial movements have resulted in profound changes in settlement patterns, and also in deep socioeconomic and cultural transformations. Smaller urban centers, such as the ones described in this paper, reflect the growing discrepancy between changing values and widening economic opportunities on the one hand, and the perseverance of a feudal system of political power often supported by a highly contentious managerial and political decentralization plan, on the other hand. (Hasan A., Raza M., 2009); Migration can cause of various socio-economic problems and factors of human security, housing, and unemployment, food that put extra burden on existing infrastructure and increase unhealthy competition of bread and butter. Due to industrialized vision, Karachi face rapid process of urbanization and lack of proper planning and management it become a critical and challenging task for the administration of Karachi. Many peoples from all over the Pakistan and cross border migrate to Karachi through proper and illegal channels for better employment and livelihood. Government enables to register them properly, that the reason they cannot create human securities for them. It increases the burden on infrastructure and resources of Karachi because services and resources of urban sector are divided into increasing population and create various socio economic problems. When expectation of having good life and better employment are not accomplish, they become frustrate and deprivation and they involve in illegal or wrong activities such as mobile snatching and organized crime such as trafficking in narcotics and arms also increase their exposure of joining political or religious extremist groups. (Manzoor, M., 2010).

3. KEY FINDINGS

Economical	Security measures
• Inflation rate rises due to supply and demand factors • Government literally lost control of housing allocation and land taxes • Lack access to land, finances and adequate shelter • Difficult to estimate food, housing and clothing • Increase unemployment • Increase the prices of houses and house rent • Depletion of natural resources especially fossil oil	• Make cities ungovernable because government reduce control • Increase illegal activities • Security collapse • Corruption by agents that support illegal migration • Absence of rule and regulation • Evaluated crime rate due to drug cartels • Street crimes such as mobile snatching, robbery increases • Human trafficking • Unprecedented demographic and political problems • Increase thread of political violence • Increase clashes among parties on land • Regulation constraint such as household registration problem

Social	Environmental
• Deteriorating and poor Social services • Increasing number of Baggers • Unhygienic living conditions • Increase population • Difficult to manage and control population growth • Child labor increases (because when family migrants in cities mostly send their children's to work in order meet there • Increase desertification • Unhealthy competition of getting bread and butter increase • Decline quality and standard of living • Overburdening on existing resources and infrastructure create shortage of basic needs	• Environmental deprivation and insecurities • Irregular traffic jam and busy signals • Faces the problems of Pollution, air, sound etc • Deforest and loss eco- system that sustain global atmospheric oxygen and carbon dioxide balances • Shortages of water supply • Inadequate transportation • Crowded Housing • Environmental Securities • Inadequate fresh water, sewerage treatment and effluent discharge

Medical

- High infant and children mortality
- Scarcity of medical resources
- Polio vaccine program not successfully work
- Low birth weight due to lack of resources
- High infectious disease like HIV/AIDS, Dengue virus, Bird flu, Hepatitis etc

4. CONCLUSION AND RECOMMENDATIONS

Migration has played a key role in reshaping the size and distribution of the population. Urban population growth has also increased through the contribution of rapid rural-urban migration, especially in the industrialized provinces of Sind and Punjab. In rural areas people face numerous problems not only regarding their earning and jobs but also lack access to basic needs such as education, health and proper sanitation, many of them see urban areas as the solution of their problems also rural-urban income gap, better paid jobs and industrialization encourage rural peoples migrate to urban areas. On the other hand lack of capital, shortage of power and iron, narrowness of home and foreign markets for the manufactured goods, frequent breakdown of electricity, lack of proper planning, use of backward technology are major obstacle in the expansion and growth of industrial sector in Pakistan, that's the reason many educated and skilled urban peoples are jobless in their homeland.

Internal and international migration relative to socio economic status may create serious problems of unemployment and lack of jobs due to limited consumption capacity of both industries and urban social services. That ultimately increase the momentum of poverty in urban areas and consequently put pressure on civic institutions and also generate problem of housing, which increase the cost of urban property and causes high rent that's the reason large number of peoples have to sleep under bridges, increase density of occupation of Kachi Abadis, to meet there expenses many of migrated facilities send their children to workplaces rather than schools that cause of increment in child labor, overcrowding and social fragmentation.

Pakistani authorities may have two policy options one is to create jobs and opportunities for migrants in urban industries and social services. That must require proper registration of individual and family migrants for allocating human resources, such as food allocation, health and economic security. In formulation process for domestic and international migration and their related issues the involvement of ministries of education, labor, immigration bureau, foreign affairs and the bureau of statistics is necessary and also promote teamwork, patience, harmony and mutual respect among the people of both sending and receiving societies between which the migrant moves. Develop programs under the involvement of Federal, provincial and district governments, NGO's and civic institutions with respect to national and specific local concerns, that help to minimize the cost and risk of migration that's the result migration is a positively influenced.

The second option is to control migration. Majority of rural peoples directly or indirectly depend on agricultural for their livelihood. Unequal distribution of land and water shortages negatively effect on agricultural sector and encourage many rural peoples to diversify their source of revenue through seeking opportunities away from home. Government need to overcome the water problem in rural areas by making dams, facilitating farmers with land and other resources that increase their production and income level, that reduce the rural-urban income gap and also minimized the imbalances between economic opportunities of rural and urban area. Government also paid more attention in rural areas infrastructure and the innovation of institutions is necessary.

If opportunities, justice and equity are available in rural society then there would be no need to go away from home for searching work. Because it is not easy to leave their homeland and family.

BIBLIOGRAPHY

1. Adger, W.N., Kelly, M.P., Winkels, A., Loung, G.H. and Locke C. (2002). Migration, Remittance, livelihood trajectories and social resilience. *Ambio* 3/4, 358-366.
2. Coast, Ernestina (2006). Local understandings of, and responses to, HIV: Rural–urban migrants in Tanzania. *Social Science & Medicine*, 63, 1000-1010.
3. Faooq, M. and Cheema, M.A. (2005). Likelihood Impact of Internal Migration on Family Institution in Punjab, Pakistan. *J. Agri. and Social Sci.*, 1/2, 148-151
4. Hasan, A. and Raza, M. (2009). Migration and Small town in Pakistan. *Working paper series on Rural-Urban Interactions and Livelihood Strategies*, 15.
5. Jan, B., Iqbal, M. and Iftikharuddin (2008). Urbanization trend and urban population projections of Pakistan using weighted approach. *Sarhad J. Agric*., 42/6, 173-180.
6. Kalantaridis, C. (2010). In-migration, entrepreneurship and rural-urban interdependencies: The case of East Cleveland, North East England. *Journal of Rural Studies*, 1-10.
7. Lu, Qi. and Chunyue, Y. (2008). Characteristics of Rural Labor Emigration in Minority Areas of Honghe, Yunnan and its Impact on New Socialist Countryside Construction in the 21st Century. *China population resources and environment*, 18/4, 085-088.
8. Ahmed, M.I.; Manzoor M.M. and Agha S. (2009). Urban growth and environment – Contemporary challenges and need for integration. Proc. of *5th International Conference on Statistical Sciences*, Lahore, Pakistan.
9. *Microsoft Encarta Encyclopedia Deluxe*, 2002.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 159-166

MULTIDIMENSIONAL ANALYSIS OF MIGRATION AND URBANIZATION: CASES FROM KARACHI

Muhammad Mazhar Manzoor
Department of Business Administration
Federal Urdu University of Science and Technology
Karachi, Pakistan. Email: mmazher@gmail.com

ABSTRACT

Urbanization refers to a process in which an increasing proportion of an entire population lives in cities and the suburbs of cities. Historically, it has been closely connected with industrialization. When more and more inanimate sources of energy were used to enhance human productivity (industrialization), surpluses increased in both agriculture and industry. Larger and larger proportions of a population could live in cities. Because of security relevance, which is an emerging paradigm for understanding global vulnerabilities and it holds that a people-centered view of securities is necessary for Economic security, Food security, Health & Environmental security, Personal security, Community security, Political security and "Freedom from Fear". Economic forces were such that cities became the ideal places to locate factories and their workers.

This study focuses the case of Karachi, Pakistan as urbanizing city. Karachi, with over 10 million residents, is the largest city of Pakistan. It accounts for over 7 per cent of the population, and for over 22 per cent of all people living in urban areas in the country. The city has recorded high rates of population growth over the last six decades. Its population increased nearly three-fold in the 1940s to over a million people in 1951, and has increased ten-fold since then. Migration has been a key determining factor in the growth of the city throughout its history. This study focuses on the multidimensional analysis of the migration of people from all over Pakistan and cross border migration from Afghanistan to the metropolitan city of Karachi. It reveals the causes, methodology and resources used by such migrants. It has been extensively tried to keep the areas of Human Security disorders in the spotlight. The grave issues likes' illegal migration random increase in urban population and their effects on the socio-economic infrastructure are critically analyzed. As a harbor and an industrial and commercial hub, it is not surprising that the city has attracted migrants from within the country and beyond. It has also served as a transit point for people migrating onwards to destinations of greater economic opportunity abroad.

KEYWORDS

Migration; Urbanization; Population; Economics Forces; Human security.

INTRODUCTION

The population of the Karachi had also risen to about 105,000 inhabitants by the end of the 19th century and was a cosmopolitan mix of Sindhi's, European traders, Parsis, Iranians, Lebanese, and Goan merchants. Pakistan is one of the many Asian countries who are experiencing not only high rate of population growth but also rapid urban growth.

Table 1
Percentage growth of population in Pakistan

Serial No.	Census Years	Population Increasing Rate (PIR) %age
1	1951	1.8
2	1961	2.4
3	1972	3.1
4	1981	3.0
5	1998	2.6
6	2004	1.64
7	2008	2.04

Source: Census of Pakistan, Government of Pakistan & United Nation.

The growth of urban population is indicated by increasing number of urban centers and also by the concentration of population in big cities. At the eve of partition 82.5% population lived in rural areas while according to the census (2004) 67.5% population lived in rural areas. A large number of populations transferred from rural to urban areas due to the social and economic problems.

In 1980’s to date saw an influx of illegal Afghan refugees from the Afghan war into Karachi, and the city now also called, a "city of illegal refugees". August 14, 1947 when it became the capital city of a new Dominion of Pakistan, its population was about 450,000 inhabitants. However, the population rapidly grew with large influx of refugees from neighboring Union of India (after the partition. By 1951, the city population had crossed one million marks. The city's population continues to grow at about 5% per annum, largely thanks to its strong economic base. According to the census of 1981 average population was consists of 106 people per km. While due to the migration in 1988 it has increased about 164 people live in per km. Lahore is the most populated city of Pakistan where about 3566 people live in a km & second populated city of Pakistan is Karachi where about 2795 live in a km & in Baluchistan especially in Awaran & Chagi only 4 people live in a km.

Studies show that long-distance and single-step migrations from mountain and hilly districts of the NWFP and plains to the Punjab are predominant among the lower income rural migrants in Karachi. This study covers the consequences of migration for the receiving on the determinants of the flows, less on the consequences for the receiving and even less on the consequences for the receiving and even less on those for sending communities. Pakistan’s primate city, formation of katchi-abadis, internal migration from all over the country to Karachi, immigration from Afghanistan which settled in many area of Karachi and their effect on social and economical infrastructure, and also effect of migrants in human security issues.

Major Issues of Migration

As observed by Mangalam (1968); ‘Migration is a relatively permanent moving away of a collectively called migrant, from one geographical location to another preceded by decision-making on the part of the migrants on the basis of a hierarchically ordered set of values or valued ends and resulting in changes in the interactional system of migrant’. Major issues consider by Rauf (1984) of migration such as urbanization, industrialization, disguised unemployment, squatter settlements, pressure upon urban services, cultural adjustment and acculturation.

Table 2
Urban population growth yearly in Karachi

Year	**1875**	**1872**	**1881**	**1891**	**1901**
Urban Population (Karachi)	56,875	56,753	73,560	105,199	136,297
Year	**1911**	**1921**	**1931**	**1941**	**1951**
Urban Population (Karachi)	186,771	244,162	300,799	435,887	1,068,459
Year	**1961**	**1972**	**1981**	**1998**	**2009**
Urban Population (Karachi)	1,912,598	3,426,310	5,208,132	9,269,265	18,000,000 (estimated)

Source: Census of Pakistan, Government of Pakistan.

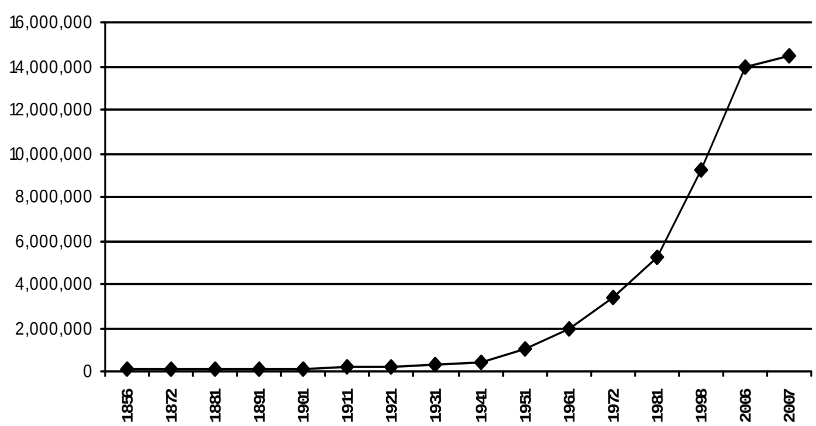


Fig. 1: Trend of population growth in Karachi.
Source: Official Website of City District Government Karachi

Migration is usually analyzed in terms of Push and Pull theory was extended by Herberle (1983). There are the several influence on the people to migrate from one place to another place, it also possible people move from one country domain to another comes in the category immigration. Mostly younger people moves from one place to another for improving the social status in their own communities and migrate for a short period of time or it may be settled permanently, if the living status continuous improves. Push factors build a force in mind to move from one place because of fulfilling their needs and wants i.e. better shelter, quality food, strong education, luxurious life style, and other necessities become the major cause for migration or immigration. Given diagram analyze the key factors and indicators directly influence on the people and become reason of migration. All given factors in the diagram having their own importance and value but loss of employment, working pressure group, catastrophe faces, division of agricultural land, less privileged standard of living and shortage of non-agricultural occupations create a negativity in the mind of people and generate a force to migrate or immigrate immediately.

Most probably that several attractive factors boost up the mind of the people for movement from one pacific to another. Pull factors most of the time highly rated in the mind of people and overcome the challenges in the mind of people which faces in future from new city or country. Key attraction in the mind of the people because of good living style, high return of investment, opportunity to earn maximum, employment opportunity, career building specially for the youngsters and cultural or re-creational activities. Given diagram highlight the indicators for the migration or immigration of people.

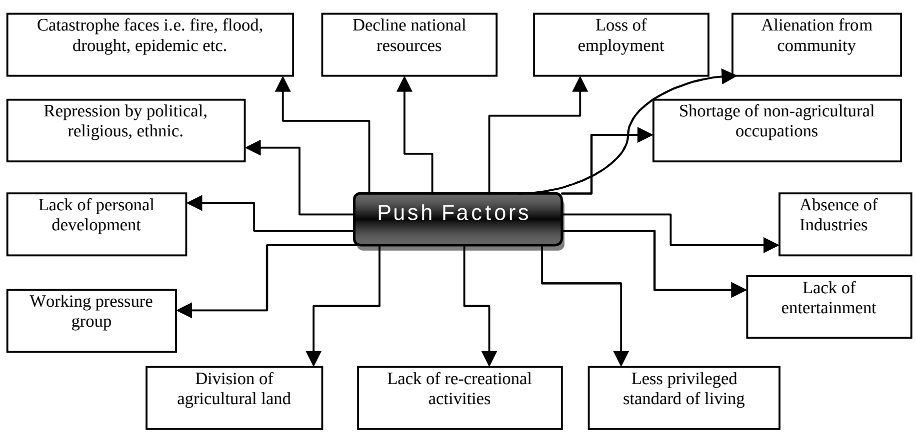


Fig. 2: Major Issues of Migration (Push Factors)

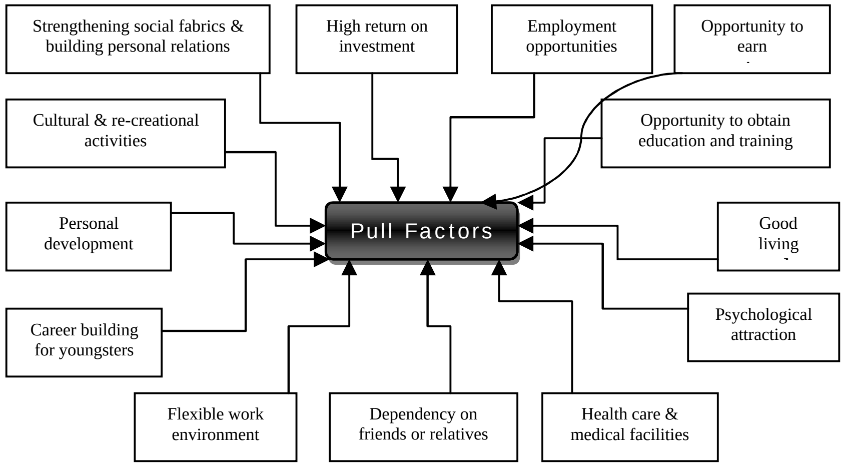


Fig. 3: Major Issues of Migration (Pull Factors)

Dimensions of Migration

Several types of migration may observed in the case of Karachi such as internal migration, Emigrant migration in which peoples from across the border such as Afghanistan, Iran, China, Bangladesh, Srilanka etc comes in Karachi according to their own needs and wants. Similarly another type of Migration is Pendular migration in which seasonal tenure may witnessed. Another dimension is Chain migration which come into existence due to several push and pull factors.

URBANIZATION

When the person move from one domain to another domain for fulfilling their needs and wants for improving their status in their own society or boost up their living style or giving all necessities to upcoming generation, decided to migrate/emigrate immediately and arrange all sources that initially supportive in big city Karachi. However, the implications of rapid urban growth raise the problems for the administration of Karachi.

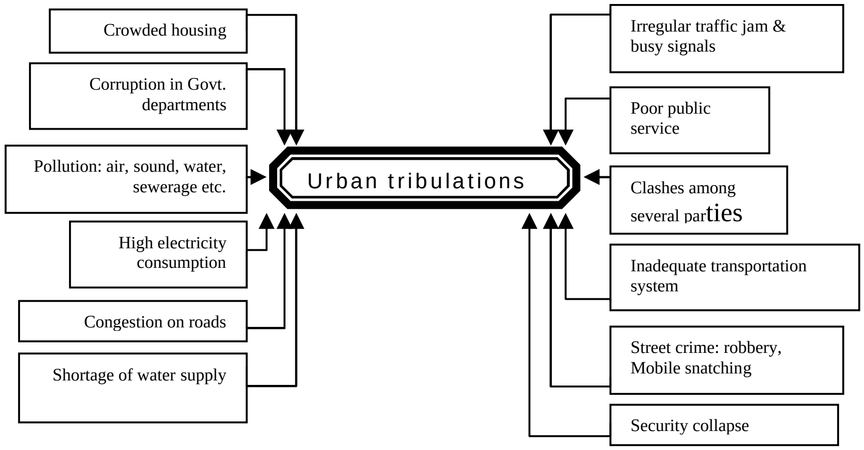


Fig. 4: Urbanization

ANALYSIS AND DISCUSSION

State Weakness

In Karachi, government has literally lost control over the management of housing allocations, land taxes, and policing. The urban land market in the city is unique. Approximately 80 percent of Karachi's land is owned by the government. These lands are often under long-term lease and managed by various governmental agencies. The main problem is corruption. According to John Adams, the violence, curfews, and plant closings cost the economy

Economic Impact

The economic impact of Karachi's civil strife extends far beyond the city itself. Businesses have suffered nationwide. Fan manufacturers in Gujrat anticipate a 25 percent drop in sales largely due to the market shutdowns caused by turmoil in Karachi in 1995. Textile manufacturers in Multan and Faisalabad have reported rising difficulties in recovering payments from Karachi traders due to falling sales in the city, and businessmen in the NWFP expect problems from a large influx of unemployed youth returning to the region in the wake of layoffs from Karachi factories.

Scarcity and Urban Violence

High natural population growth and a rapid influx of migrants have pushed diverse and contending societal groups into close contact. Inequalities among economic classes and ethnic groups are therefore more obvious than would be the case in rural areas, and these inequalities are exacerbated by competition for exceedingly limited urban resources. The city's high urban growth rate - about three to four hundred thousand persons per year - has accentuated these conflicts.

Human Trafficking

The coastal belt from Karachi to Balochistan is still a paradise for human traffickers as the route to Iran is being used for this illegal and heinous practice. Human trafficking from Pakistan has gained momentum since the 1950s, said a report prepared by various NGOs. Karachi is believed to be a hub for human trafficking towards Gulf and western countries. Recently, dozens of people are coming from Badakhshan, Kunduz and northern regions of Afghanistan every day to enter Iran. Al-Asif Square, Banaras and Lee Market are the main markets for transit points developed by human traffickers.

Health Contagion

As many people live in Katchi Abadi and slums the health problems are severe in those areas. As the case of polio continuously programs by Govt. of Pakistan are going on to eliminate this disease from the country but it is not eliminating due to Afghan mohajrins.

NEGATIVE EFFECT OF ILLEGAL MIGRANTS ON SOCIO-ECONOMIC STRUCTURE

The presence of migrants in Karachi has caused a number of problems in the socio-economic structure of the city. Some of them are discussed below.

- Since the illegal migrants are living in Karachi without any proper documentation, so it is very difficult to estimate their food, clothing and housing need. Katchi bastis are containing people above their capacities, hence finding it difficult to meet everyone’s need. The result is continuous hunger and greed of grabbing as much as possible.
- Karachi having four main industrial zones i.e. S.I.T.E., Korangi, New Karachi and West Wharf, attracts people from all over Pakistan to have employment in its vast agricultural potentials. The migrants, normally dwelling in katchi bastis in suburban areas of the city, have to travel long distances to reach their work places. It’s normally carried out by public transport. So, the trans-city traveling of this labour causes transport and traffic problems, specially during typical rush hours.
- As no proper procedure is followed while sale, purchase or hiring a residential place in such bastis, so it becomes very easy for the thieves, dacoits, robbers, gangsters and wrong-doers to accommodate themselves in such areas. On the other hand, it’s also practically impossible for law inforcement agencies to arrest and detain such culprits as no evidence of their living place and hide-outs is available.
- There’s no proportion in the increasing number of migrants and the number of industries. Rather huge number of industry has shifted from Karachi to the other parts of country due to the political unrest during the last two decades. So, less number of employment opportunities have to accommodate larger and even larger number of workers in them. It seems virtually impossible. As a result, the unhealthy competition of getting bread and butter is increasing day by day.

- The migrants after getting failed to receive their rights of better living from the administration, start getting involved in illegal activities. Group associated to the underworld pick them, train them and use them for their wrong purposes. The migrants are highly paid by such groups so they can trap more and more people towards them.

CONCLUSION

Due to industrialized vision of Karachi, peoples from all over the country and abroad (mostly from Afghanistan & Bangladesh) find it right place to meet their both ends. The migrants in Karachi are earning a better livelihood and shift from agriculture to urban wages in worthwhile. However, it is observed that migrants have not moved from the place of origin just to enjoy city life and comfort as it offered. Nevertheless, due to abundant migration and illegal migration many socio-economic problem occurred and factor of human security are high as illegal migrants can be a cause of human trafficking, crimes etc. they are also a cause of Katchi Abadis and slums in Karachi and due to unemployment they take bagging as their profession. Thus, the challenge is to manage and document migration effectively and efficiently such that states and international community may also combat with the challenges that come with the high mobility of workers. Given the diversity and volume of migration flows, it is important that states come up with a coherent migration policy so that the benefits of migration can be enjoyed fully as this is a vital process for the poor and for development.

REFERENCES

1. Baloch, S. (2008). Human trafficking on the rise in Pakistan, *The Nation*, Karachi, Pakistan, June 13, 2008.
2. Mazhar, M.M., Ahmed, I. and Agha, S. (2010). Urban growth & environment-contemporary challenges and need for integration. *5th International Conference on Statistical Sciences, Lahore, Pakistan.*
3. Mangi, N.A. (2009). Pakistanis Protest 42-Hour Power Outage in Karachi (Update2), [Online:http://www.bloomberg.com/apps/news?pid=20601091&sid=aoqgNMycnas0].
4. Selier, F.J. (1988). *Rural-Urban migration in Pakistan: The Case of Karachi*. Lahore, Pakistan: Vanguard Books, Inc., 166.
5. Selier, F. and Karim, S. (1986). *Migration in Pakistan theories and facts*. Lahore, Pakistan: Vanguard Books, Inc., 193.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 167-172

INCORPORATING ISLAMIC FINANCIAL FRAMEWORK FOR COMBATING GLOBAL FINANCIAL CRISIS

Muhammad Mazhar Manzoor[1], **Abdul Sattar**[2], **Muhammad Aqeel**[1], **Shagufta Rasheed**[3] and **Hassan Raza**[4]

[1] Department of Business Administration, Federal Urdu University of Science and Technology, Karachi, Pakistan
Email: mmazher@gmail.com; aqeel_xp@yahoo.com

[2] Department of Management Sciences, Balochishtan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan.

[3] Meezan Bank Limited, Karachi, Pakistan

[4] National Bank of Pakistan, Karachi, Pakistan

ABSTRACT

This paper attempts to analyze the certain causes of recent global crisis by investigating the key factors mainly interest based activities including subprime mortgage and credit default swaps and proposing a comprehensive Islamic financial framework as an alternative to conventional finance. The root causes of financial crunch has been investigated as interest based activates, mortgage, imprudent lending, credit defaults, lack of transparency and loss of confidence etc. But this study entails only interest as the major factor for financial crunch which leads to increase in mortgage and other sort of loans which aroused from deregulation. This study is conducted by exploring the key factors of financial crisis by theoretical as well as factual grounds. This in turn leads to formation of Islamic financial model based on features of Islamic finance such as asset backed financing, equity financing and profit and loss sharing, thus findings suggest specific approach to the solutions of crisis by modeling the Islamic finance on theoretical and conceptual bases.

KEYWORDS

Financial crisis; Subprime Mortgage; Credit Default Swaps Islamic financial framework; conventional finance; asset based financing; equity based financing; profit and loss sharing.

INTRODUCTION

It is not the first time that most developed countries have faced sever financial crisis. Financial institutions of the world had to countenance such crisis because of number of reasons which are believed to be intrinsic in interest bearing economic system. Many of the economic gurus anonymously agree on certain core causes such interest based activities, illusionary money and credit and other roots of these financial crises including subprime lending. The recent economic crunch had greater impact on world economy than depression of 1930s (Davies and Davies 1996). The advent of recent financial crisis is witnessed from America (Rarick and Han 2009). The major US banks are blamed of deregulated lending which resulted in a decrease in bonds' values (Hassan, 2009).

The origin of almost all the financial crises has been traced from disproportionate and unwise lending by banks (BIS 2008). The conventional financial system which is based on interest has proven to be insufficient to deal with banking system which resulted in excessive risk taking and subprime mortgage crisis in the United States (SESRIC 2009**).**

Financial derivatives occupy a main place in conventional financial system, these derivatives deal in the transactions where the exchange of real goods or services do not take place hence creating illusionary money, speculation and excessive credit. In the recent years there has been a huge increase in credit derivatives such as Credit Default Swaps (CDS) and Over -The-Counter (OTC) observed which was $592.0 trillion at the end of 2008 for OTC while CDS reached to $58 trillions at the end of December 2007 which representing 105% of worlds total GDP. It has been shown in the following figure (BIS 2009).

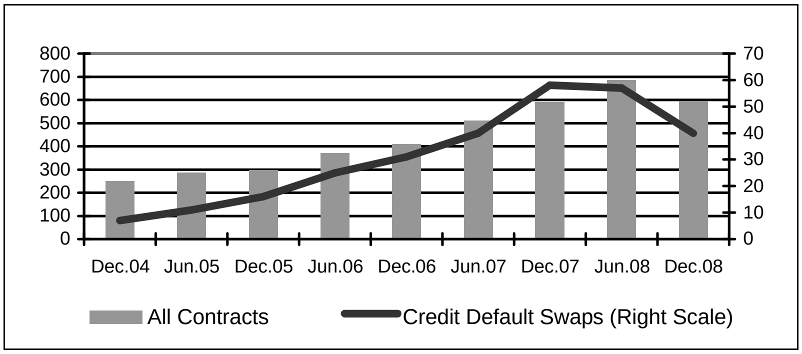


Fig. 1: National Amounts Outstanding (Trillion US Dollar)
Source: Bank for International Settlement, Statistics on Derivatives.

One of the prime causes of recent financial crunch is subprime mortgage which was the result of some other reasons for this economic mishap including poor risk modeling, deregulations lack, of transparency and governance lapses (Blundell-Wignall, et al. 2008, Goodhart, 2008).

The most striking observation in the recent years has been the stability of Islamic financial system. This stability make the Islamic economists believe that interest based mechanism of conventional finance along with excessive monetary expansion; large balance of payment deficits and inadequate regulations are the core basis for such crisis. Siddiqui, (2009) proposed that such crisis can be successfully dealt with the application of Islamic financial principles which suggest asset based financing, all-encompassing equity financing and socio economic applications.

In this paper, most common causes of recent financial crisis have been highlighted and a proposed model has been put forwarded as an alternative to conventional economic system.

LITERATURE REVIEW

Seidu (2009) pointed out some radical weaknesses in conventional financial system according to him strategy failure, financial regulatory, complacency high ratio of bad debt loans and flawed ideologies are the main contributors in financial crunch and Islamic financial system can be a best alternative.. A similar exploration made by Adrian and Shin (2008) according to which deregulation, lack of lucidity and legislation are the contributors in recent financial crisis.

Ivashina and Scharfstein (2009) blamed excessive and credit boom for the recent financial crisis which were at the epic during the summer of 2007. The followed by downfall of subprime mortgage and securitized products. All these factors led to the concerns regarding the solvency and liquidity of financial institutions.

SESRIC (2009) report highlighted similar causes of financial crisis. Baily, et al, (2008) and Stiglitz, (2007) deemed that structural changes in conventional financial systems are prerequisite in order to minimize the occurrence and magnitude of financial crisis.

Sakti (2009) expounds speculative and derivative market transaction that this non productive activity is a big threat for crisis in real economy situation because he observed that volume of transactions in money market amounted US $1.5 trillion in a day but the transaction volume in the world real sectors amounted only US$6 trillion in a year.

Islamic financial system does not only prohibit interest but regulated with shariah compliance and all the transaction have real value there is no illusionary money as described by Tayyebi (2009) that Assets backed transactions restrict in trading of debts thus reducing the possibilities of excessive lending. Islamic finance does involve home financing but greater initial deposits and assets involvement reduce the threat of subprime issue.

The need of a regulated and substantial financial system has increased in the recent past and Islamic financial system is proving to be the best alternative of orthodox interest based financial setup. There is rich literature regarding the causes and remedies is available. This paper analyzes the selected literature and proposes Islamic financial model to replace conventional interest based system.

METHODOLOGY

Much of the literature regarding causes and remedies of financial crisis have been written A bulk of secondary data have been taken under study in order to explore the root causes of the current financial crisis and propose an alternative to avoid such financial crunch. Secondary data are the principal sources of data collection for this study. The data explored many reasons of financial crisis including interest, deregulation and credit derivatives such as credit default swaps. In the lights of the collected data the interest based activities as the root cause of financial crisis have been explicitly explained in the discussion section and an interest-free mechanism as Islamic financial model has been proposed to prove that Islamic financial system can put up barricade in the way of financial crises.

DATA ANALYSIS AND DISCUSSION

Murphy (2008) enlightened that mispricing in the Credit Default Swap Market is one of the root causes of financial crisis. According to him the estimated principle amount of credit default swaps was to be $55 trillion by Security Exchange Commission in United States but it may actually exceed $60 trillion. This shows towering degree of deregulation. A similar analysis has been made by Simon (2008) that even the contracts were made on phone calls without having documentation this shows unregulated approach and contributed mainly in financial crisis.

According to Daly (2008) the financial assets have overgrown in comparison to the growth of real wealth which has given rise to too much liquidity consequently giving rise to financial crisis. On the other hand, in Islamic finance growth is directly relates to growth of real sectors. A study of Kallis et al (2009) about financial crisis shows that future based transactions are carried out in interest based economic system with the expectation that blurred economic growth will provide ways to repay the debt and interest. But this is not a certain case. In case economy does not grow it makes the debtor defaulter. Increasing debts force economy to grow but to some limits thus real GDP does not grow this scenario leads to credit crunch. Islamic finance on the other hand does not involve in interest based lending it rather participates in profit and loss sharing and equity financing thus having a direct contribution in economic growth.

Chapra (2009) also blames the excessive lending of banks in giving rise to financial crisis he profound that bank would have avoid that only if they were afraid of losses, reputation distortion and bankruptcy but they were sure to receive interest with repayment of loans and there was lack of profit and loss sharing practices. This leads the banks to undertake excessive leverage and huge volume of debts the consequent of this entire state of affairs is unsound rise in prices of assets a speculative investment. (G-20 2008) summit also agree that excessive lending is one of the prime causes of financial crisis.

While Islamic financial principles stresses on more financial discipline, as mentioned by Chapra (2009) in his book that Islamic finance is pillared on the Principe of “No risk, no gain” this intensifies greater discipline in financial market this stimulates financiers to closely monitor risks.

Nasiruddin Ahmed (2009) argued that equity based Shariah banking system works as an alternative to conventional finance and the recent financial crisis have increased the need of Islamic financial system.

A strong sense of alternative economic and financial systems has been felt. In the recent scenario Islamic financial system has been one of the key alternatives that can reduce the effect of such financial downfall. Because, after the analysis of recent financial down fall, it has come into the surface of study that the root causes of financial crises do not exist in Islamic financial system altogether. In Islamic finance profit maximization is not the core objective, Islamic finance is rather based on asset backed financing, profit and loss sharing and equity financing. Islamic finance involves house financing but there is no involvement of interest, hence no threat of subprime mortgage. Islamic finance is regulated under well-defined Shariah regulations, therefore regulatory issues are also counter with.

Thus under these discussion, we propose an Islamic financial model as an alternative to conventional finance.

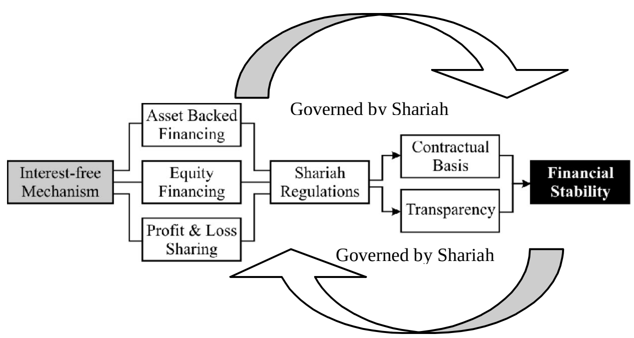


Fig. 2: Islamic financial model.

The model suggests that concept of Islamic finance come into view from interest free business practices. The mechanism of Islamic finance is based on asset backed financing Equity financing and profit and loss sharing. In conventional financial deregulation of financial activities is the give rise crisis there were no tangible contracts involved, mispricing credit default swaps and subprime mortgage were resulted. The regulatory framework shows that in Islamic finance, economic stability is resulted from Shariah tenets to avoid interest based activities and once the stability is achieved it is continuously carried out under the regulations of Islamic Shariah laws. The Shariah regulation brings transparency and stability in financial transactions. The growth in Islamic financial institutions has a direct relationship with the growth in real sector of economy and production. This mechanism can be better understood by analyzing pillars of Islamic finance i.e. profit and loss sharing asset backed financing and equity financing etc. Islamic financial system provides a comprehensive solution to current financial crunch. Because root causes of financial crisis do not prevail in Islamic financial practices. Interest based and excessive lending activities in orthodox financial system are replaced by equity financing, profit and loss sharing and asset backed financing. The core object of Islamic financial system is the promotion of human and social well being. Conventional finance have witnessed a mammoth deregulation which head towards subprime mortgage and credit default swaps losing the confidence of investors and resulting to financial crisis. More importantly the growth of Islamic financial institutions is tied with the growth of real sectors of economy thus a positive advancement in Islamic financial institutions results economic growth and stability.

Hence financial crises are hard to come under Islamic financial system.

The results of our study show that interest free mechanism gives rise to asset backed financing, equity financing and profit and loss sharing which are governed by Shariah regulations. Mathematically we can express our model as under

Thus:

$$\textbf{\textit{Financial stability}} = \beta_1 \textbf{\textit{ asset backed financing}} + \beta_2 \textbf{\textit{ equity financing}} \\ + \beta_3 \textbf{\textit{ Profit and Loss Sharing}} + e \ldots\ldots\ldots\ldots (i)$$

CONCLUSION AND RECOMMENDATIONS

CONCLUSION

The economies of the world are not facing financial crisis for the first time. Since the Great Depression of 1930s, world has been facing sever or low intensified financial shocks. The financial crunch which emerged from the mid of 2007 has received more attention and most of the economist agree that it has far more sever impact then older ones. This suggests some inherent flaws in the functioning of orthodox financial system. Mainly interest based activities, subprime mortgage and credit default swaps are considered more responsible which resulted deregulation and loss of confidence. In this regard reforms in the terms of regulations and alternative economic systems are being considered among which Islamic financial system provides a broad and all-encompassing solution to this problem. Islamic financial system does not involve inertest based activities it rather promotes profit and loss sharing thus encourages growth of real sector of economy and hindering financial crises.

This paper concludes that Islamic financial system regulated under well defined Shariah governance, reduces the threats of deregulation and promotes equity based financing; profit and loss sharing and asset backed financing which provides a best alternative to conventional financial system.

RECOMMENDATIONS

Economists of the world should agree on certain applicable and long term alternatives or remedies to the burning issue. This paper recommends application of financial system which discourages inertest based activities, contributes in economic growth by actively investing in real sectors of economy and regulates well defined rules for financial transactions. In present time Islamic financial system has certain attributes to meet the challenges of present crisis. Thus Islamic financial system should be considered as a remedial step to confront such financial crisis.

REFERENCES

1. Adrian, T. and Shin, H.S. (2008). *Liquidity, Monetary Policy, and Financial Cycles*. Federal Reserve Bank of New York, Current Issues in Economics and Finance,
2. Baily, M., Robert, L. and Johnson, M. (2008). *The Origins of the Financial Crisis*. The Initiative on Business and Public Policy at the Brookings Institution.
3. Blundell-Wignall, A., Atkinson, P. and Lee, S.H. (2008). *The Current Financial Crisis: Causes and Policy Issues*. OECD Financial Market Trends, ISSN 1995-2864.
4. Chapra, M.U. (2009). *The Global Financial Crisis Some Suggestions for Reform of the Global Financial Architecture in the light of Islamic Finance*. Joint Usage / Research Center for Islamic Area Studies Center for Islamic Area Studies at Kyoto University (KIAS).
5. Daly, H. (2008). *Credit crisis, financial assets and real wealth*. The Oil Drum, www.theoildrum.com
6. Davies, R. and Davies, G. (1996). *A History of Money from Ancient Times to the Present Day*. Cardiff: University of Wales Press. Economics and Economic Policy.
7. G-20 Summit. (2008). *Decelerate of the Summit on Financial Markets and the World Economy*. Washington DC: White House.
8. Goodhart, C.A.E. (2008). The background to the 2007 financial crisis. International Economics and Economic Policy, *Springer*, 4(4), 331-346.
9. Hassan A. (2009). *The Global Financial Crisis and Islamic Banking*. www.islamic-foundation.org.uk/IslamicEconomicsPDF/Hassan-financialcrisis-if.pdf [online]http://ssrn.com/abstract=1295344
10. Ivashina, V. and Scharfstein, D. (2009). *Bank Lending During the Financial Crisis of 2008*. [online] http://ssrn.com/abstract=1297337.
11. Kallis, G., Martinez-Alier, J. and Norgaard, R.B. (2009). Paper assets, real debts an ecological-economic exploration of the global economic crises. *Critical Perspectives on International Business*, 5(1/2), Emerald Group Publishing Limited.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 173-178

ELIMINATION OF RIBA FROM BANKING BY USING ISLAMIC FINANCING METHODS

Muhammad Mazhar Manzoor[1], **Muhammad Fahim**[1], **Sajid Ahmad Chauhan**[1]
Waseem Hasan Raja[2] and **Jamshan Siddiqui**[2]

[1] Department of Business Administration, Federal Urdu University of Science and Technology, Karachi, Pakistan Email: mmazher@gmail.com

[2] Department of Management Sciences, KASBIT, Karachi, PAKISTAN.

ABSTRACT

Islamic Economic principles promote equity and asset based investing and prohibits charging interest payment as in conventional banking. The purpose of the study is to find out the permissible Islamic financing methods that are free from riba and their implementation by the Islamic banks, and other financial institution, which are offering these modes of financing. Islamic banks an alternate to interest based banking is not banking in traditional sense of the word. It derives its inspiration and guidance from the religious edicts of Islam and has to conduct its operation strictly in accordance with the direction of shariah. Islamic finance base on concept of sharing. While the conventional banks have no concerns with the sharing principles. In this study the data was gathered from secondary source and no statistical tool was applied on the data as the data was descriptive not in numeric form and we have seen in the study that Islam has given many methods that are riba free and by implementing these methods we can run the banking system as well as the whole economy in an Islamic way.

INTRODUCTION

Elimination of Riba from economy is the key issue of the Islamic world. As Riba is strongly prohibited in Islam. And there are strict warning against this practice given in Quran and Ah Hadis. The purpose of this dissertation is to give others to focus on interest free mode of financing as the conventional banking system is based on interest and there are many people who want to finance in an Islamic way. Due to the frequent use of interest or we can say that Riba there is injustice to the people and the money is just hold by a fist of people. As a result of this unbalance economy is created where only accumulation occurs for the sake of accumulation. Islamic banking emerged as a response to both religious and economic needs. Islam calls for avoiding any transaction based on Riba while economic needs ask for diversity in the role of banking foe promoting investment/productive activities, facilitating economic justice and adding stability to the economy. Islamic banking can thus be perceived as an improved and better-equipped system in different dimensions. The banking system in an economy works like the blood circulation system of a body. As only an efficient blood circulation system can ensure a healthy body. Similarly an efficient and equitable banking system can dispense economic efficiency and justice. These basic concepts and objectives are common to any banking system whether it be conventional or Islamic. The difference lies in the methodology adopted to achieve these objectives through use of interest-based contracts while Islamic banking achieves these objectives through trade-based contracts.

LITERATURE REVIEW

Hussain, Dr. Ishrat (2004) Islam is not a new religion it is the same truth that God revealed through all His prophets. All religions are the same in essence, whether given, for example, to Noah, Abraham, Moses, or Jesus, or to the holy Prophet of Islam. For a fifth of the world's population, Islam is both a religion and a complete code of life. A unique feature of Islamic banking is its profit-and-loss sharing (PLS) paradigm, which is principally based on the Mudarabah (profit sharing) and Musharakah (joint venture) concepts of Islamic contracting. Under the PLS paradigm, the assets and liabilities of Islamic banks are integrated in the sense that borrowers share profits and losses with the banks, which in turn share profits and losses with the depositors. Advocates of Islamic banking, thus, argue that Islamic banks are theoretically better poised than conventional banks to absorb external shocks because the banks' financing losses are partially absorbed by the depositors Muhammad Fahim Khan (2003) We conveniently found Murabahah based financing as a permissible alternative close to interest based lending by simply utilizing Shariah permissibility of appointing banks own clients (seeking funds) as its agent to purchase the goods needed by the client. Tariq Ahmed Saeedi, (2008) Islamic financial market industry uses KIBOR that follows conventional financial mechanism as a benchmark to determine its profit. Islamic financial industry is not in position to develop its own interest mechanism nor has its sustainable derivative market. Not only Riba is prohibited in Islam but also uncertainty in profit or speculative investment is also prohibited. Islamic financial institutions as a whole need proper planning and support of aggressive awareness campaign. Muhammad El Oorchi (2005), The fact that Islamic laws prohibit paying and receiving interest does not imply that they frown on making money or encourage reverting to an all-cash or barter economy. They encourage all parties in a financial transaction to share the risk and profit or loss of the venture. Depositors in Islamic banking can be compared to investors or shareholders, who earn dividends when the bank makes a profit or lose part of their savings if the bank posts a loss. The rationale is to link the return in an Islamic contract to productivity and the quality of the project, thereby ensuring a more equitable distribution of wealth. Dr. S. Alam (2009) stated that, Riba means increase or any addition amount over the principle amount it includes all like usury and interest. Riba is of two kinds (a) Riba An-Nasiah i.e. interest on lent money. (b) Riba Al-fadl, i.e. taking a superior thing of the same kind of goods of inferior quality. Riba is strongly prohibited in the Holy Quran and Holy Prophet Hazrat Muhammad Peace be upon Him has also described on many occasions and strictly prohibited taking riba in any form, from any source. Allah does not love those sinners who, in spite of the divine favors of admonition and riches continue to receive usury and consider to legally valid. That reflects rank ingratitude plus contumacious sinfulness on their part. Usury and interest is totally banned item and earning it from banks or any other institution is an illegal money and useless for everyone and anyone who is doing usury transaction and is not good on his part. Attacking with several diseases and calamities are due to taking interest /Riba/usury. The important feature of Islamic financial system is the prohibition of paying and receiving of interest on capital, i.e. guaranteed fix predetermined rate tied to maturity called riba, while the profit on capital comes from sharing is encouraged this increase the wealth through entrepreneurship. Islamic principles related to Islamic financial system are; advocating risk sharing, promotion of entrepreneurship, discovering speculative behavior, preservation of property rights. The nature of capital as solely being a medium of exchange is central to the prohibition of interest.

HYPOTHESIS

Ho: {Elimination of riba is observed by using Islamic financing methods.}

H_1: {Elimination of riba is not possible by using Islamic financing methods.}

RESEARCH DESIGN

Our study is descriptive as we first describe the different Islamic financing methods and how the banks are using them and the data we are collected is from secondary sources.

The data is collected from secondary data resources, such as literature review, Journals Newspapers, web browsing, articles and books related to the topic/Islamic Banking. The data is collected from state bank of Pakistan's books. Our main focus is the Islamic banks working in Pakistan using different Islamic methods that are really interest free and not just merely change of name that the conventional banks are using for financing.

Sampling or data collection is always a difficult and the most important step to do any study as our study is based on the outcome of that data. Here we are using the data present in the SBP's Published books which are very reliable and authentic and we will see how many Islamic banks are working in Pakistan and how many permissible Islamic financing methods they are using.

ANALYSIS AND DISCUSSION

1-**Participatory Mode of Finance**

Various forms of partnership can be direct financing methods. In the early days of Islamic banking and finance, those forms were commonly grouped under the banner "profit and loss sharing", to be contrasted with the debt-based forms of financing. It was assumed by some that the profit and loss sharing methods were somehow more ideal from an Islamic point of view. The fact that most Islamic banking practice concentrated on credit sales and leases was thus often lamented as re-labeling of the forbidden interest.

Musharakah“*Musharakah*” is a term frequently referred to in the context of Islamic modes of financing. The connotation of this term is a little limited than the term “Shirkah” more commonly used in the Islamic jurisprudence. For the purpose of clarity in the basic concepts, it will be pertinent at the outset to explain the meaning of each term, as distinguished from the other. “Shirkah” means “Sharing” and in the terminology of Islamic Fiqh, it has been divided into two kinds:

1. *Shirkat-ul-milk:* It means joint ownership of two or more persons in a particular property. This kind of “Shirkah” may come into existence in two different ways: \ sometimes it comes into operation at the option of the parties. For example, if two or more persons purchase equipment, it will be owned jointly by both of them and the relationship between them with regard to that property is called “Shirkat-ulmilk”.
2. *Shirkat-ul-‘aqd*: This is the second type of Shirkah which means “a partnership effected by a mutual contract”. For the purpose of brevity it may also be translated as “joint commercial enterprise”.*Shirkat-ul-‘aqd* is further divided into three kinds:
   - *Shirkat-ul-amwal* where all the partners invest some capital into a commercial enterprise.
   - *Shirkat-ul-A’mal* where all the partners jointly undertake to render some services for their customers, and the fee charged from them is distributed among them according to an agreed ratio.

- *Shirkat-ul-'aqd* is *Shirkat-ul-wujooh*. Here the partners have no investment at all. All they do is that they purchase the commodities on a deferred price and sell them at spot. The profit so earned is distributed between them at an agreed ratio

MUDARABAH *"Mudarabah"* is a special kind of partnership where one partner gives money to another for investing it in a commercial enterprise. The investment comes from the first partner who is called "*rabb-ul-mal*", while the management and work is an exclusive responsibility of the other, who is called "*mudarib*".

Types of Mudarabah

1. Al Mudarabah Al Muqayyadah Rub-ul-maal may specify a particular business or a particular business or a particular place for the mudarib, in which case he shall invest the money in that particular business or place. This is called Al Mudarabah Al Muqayyadah (restricted Mudarabah).
2. Al Mudarabah Al Mutlaqah However if Rab-ul-maal gives full freedom to Mudarib to undertake whatever business he deems fit, this is called Al Mudarabah Al Mutlaqah (unrestricted Mudarabah). However mudarib cannot, without the consent of Rab-ul-Maal, lend money to anyone. Mudarib is authorized to do anything, which is normally done in the corse of business. However if they want to have an extraordinary work, which is beyond the normal routine of the traders, he cannot do so without express permission from Rab-ul-Maal. He is also not authorized to:
   a) Keep another Mudarib or a partner
   b) Mix his own investment in that particular Modarahab without the consent of Rab-ul- Maal.

Non-Participatory Mode of Finance

There are three major sale-based financing contracts that are discussed in the classical literature on shariah: deferred or installment payment sale, forward sale with immediate payment, and manufacturing sale

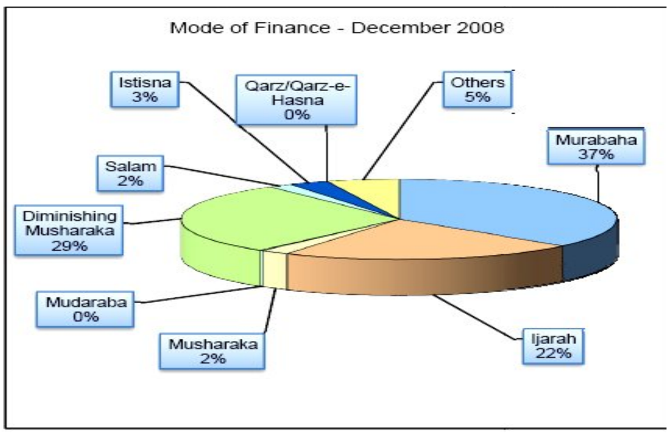
a) Murabaha In this sale, the buyer knows the price at which the seller obtained the object to be financed, and agrees to pay a premium over that initial price. Under this arrangement the bank discloses its cost and profit margin to the client. In other words rather than advancing money to a borrower, which is how the system would work in a conventional banking agreement, the bank will buy goods from a third party and sell those goods on to the customer for a pre-agreed price. Murahababahah is a mode of financing as old as Musharakah. Today in Islamic banks world-over 66% of all investment transactions are through Murabahah.
b) MUSAWAMAH Musawamah is a general and regular kind of sale in which price of the commodity to be traded is bargained between seller and the buyer without any reference to the price paid or cost incurred by the former. Thus, it is different from Murabaha in respect of pricing formula. Unlike Murabaha, seller in Musawamah is not obliged to reveal his cost. Both the parties negotiate on the price. All other conditions relevant to Murabaha are valid for Musawamah as well. Musawamah can be used where the seller is not in a position to ascertain precisely the costs of commodities that he is offering to sell.
c) Ijarah "*Ijarah*" is a term of Islamic fiqh. Lexically, it means 'to give something on rent'. In the Islamic jurisprudence, the term '*ijarah*' is used for two different situations. In the first place, it means 'to employ the services of a person on wages given to him as a consideration for his hired services.' The employer is called

musta'jir while the employee is called *ajir*.The rule of ijarah, in the sense of leasing, is very much analogous to the rules of sale, because in both cases something is transferred to another person for a valuable consideration. The only difference between ijarah and sale is that in the latter case the corpus of the property is transferred to the purchaser, while in the case of ijarah, the corpus of the property remains in the ownership of the transferor, but only its usufruct i.e. the right to use it, is transferred to the lessee.

d) Ijarah Wa Iqtina In Islamic Shariah, it is allowed that instead of sale, the lessor signs a separate promise to gift the leased asset to the lessee at the end of the lease period, subject to his payment of all amounts of rent. This arrangement is called Ijarah wa Iqtina. It has been allowed by a large number of contemporary scholars and is widely acted upon by the Islamic banks and financial institutions. The walidity of this arrangement is subjected to two basic conditions.(I)The agreement of Ijarah itself should not be subjected to signing this promise of sale or gift but the promise should be recorded in a separate document,(II)The promise should be unilateral and binding on the promisor only. It should not be a bilateral promise binding on both parties because in this case it will be a full contract affected to a future date, which is not allowed in the case of sale or gift.

CONCLUSION

Islam was the basis of creation of an independent state within the undivided Indo-Pak Sub-Continent. Since its creation, the people of Pakistan have held the demand for elimination of Riba from the financial system of Pakistan on the basis of Islamic precepts. All Constitutions of Pakistan have incorporated, within the principles of policy, the elimination of Riba as an important objective of the State policy. Since Pakistan started with an approach to convert the whole system into Islamic one, a number of amendments in relevant laws were introduced providing legal cover for Islamic financial products and services. Similarly, some new laws were introduced to allow new financial institutions or facilitate the existing ones. The legal and regulatory infrastructure developed during that era has proved to be invaluable asset as we keep on charting the present landscape of the industry today on the same. Islamic Banking Industry of Pakistan continued its progress during the year 2008. They have increased their share of assets in the overall banking system to 4.9% up to December 2008. The growth is also reflected in increased share of deposits and financing & investment that stood at 4.8% and 4.4% respectively at the end of Dec 2008.



Last but not least, Islamic finance can meet all the transaction needs of the market it does so more efficiently than conventional finance.The Islamic financing system replaces the concept of riba(Ho is accepted) with profit and loss sharing. There are many arguments that say that Islamic banking is not Islamic it is just merely change of name to attract people who want to live their life in Islamic way and afraid of riba, but as I studied different Islamic modes of financing I came to a conclusion that these methods follows Islamic principles and are riba free only the problem is that they follow the KIBOR and LIBOR as a benchmarks which needed to be replace by some Islamic benchmark, but only because of the use of these benchmarks we can't say these methods are not Islamic methods. These methods are introduced after many investigation by a shariah board and complies Islamic modes by adopting these methods by Islamic banks riba can be eliminated from the banking sector and consequently from the economy only there is a need of public awareness and from government sector to establish Islamic institutions that offer Islamic products all over the Muslim countries.

RECOMMENDATIONS

- There is a need to educate people about Islamic financing methods especially the bankers.
- Banks in Pakistan use KIBOR & LIBOR as benchmark for the determination of their rate of return, that needs to be replace by Islamic index.
- There should be continued dialogue b/w Islamic banks, Islamic economists, shariah scholars and those working in the conventional bank with an open mind.
- Shariah scholars must find out solution to conventional hedging, sale of deeds and securitization.

REFERENCES

1. Aziz Tayyebi (2008). *Discussion paper: Islamic Finance: An Ethical Alternatives to conventional Finance*? ACCA (The Association of Chartered Certified Accountants).
2. Alam, S.M. (2009). Islamic Injunctions against usury (Riba). *Pakistan and Gulf Economist*, 24.
3. Husain, Ishrat (2004). *Evolution of Islamic Banking*. State Bank of Pakistan.
4. Fahim M. Khan (2003). Guaranteeing Investment Deposits in Islamic banking System. *J.KAU: Islamic Econ.*, 16(1), 45-52.
5. Tariq S. Ahmed (2008). Understanding Riba-Free Islamic Banking. *Journal of Pakistan and Gulf Economist*, 16.
6. Iqbal, Z. (1997). Islamic financial systems. Finance & Development.
7. Mohammed El Oorchi, (2005). Islamic Finance gears up. *Finance & Development* – A quarterly magazine of the IMF, 42(4).
8. Siddiqui, S. Ahmad (2005). Understanding and Eliminating Riba: Can Islamic Financial Instruments be meaningfully implemented? *Journal of Management and Social Sciences*, 1(2), 187-203.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 179-186

BETA GENERALIZED PARETO DISTRIBUTION

Ayesha Roohi
Department of Statistics, Lahore College for Women University
Lahore, Pakistan
Email: ayesha_roohi_2004@yahoo.co.uk

ABSTRACT

The Pareto distribution is widely used to model data relating to income. In this paper a five parameter beta generalized Pareto distribution is being introduced. Different properties of this distribution are discussed. Expressions for the moment generating function, moments, mean deviation, hazard rate function, and entropy are derived. Distribution of the order statistic, its moment generating function and moment have also been obtained. Estimation has been done using the method of maximum likelihood.

KEYWORDS

Moments; hazard function; maximum likelihood estimators.

1. INTRODUCTION

The Pareto distribution is used to model income distributions. The Exponentiated Pareto distribution was derived by Gupta et al. (1998) as a lifetime model, whereas the beta Pareto distribution introduced by Akinsete et al. (2008) was used to model flood data. In this paper the beta generalized Pareto distribution is being proposed. This distribution is more flexible as the Pareto, Exponentiated Pareto and the beta Pareto distributions are its special cases. Eugene et al. (2002) proposed a method of deriving a new class of distributions. They suggested that if G(x) is the cumulative distribution function (c.d.f.) of a random variable X then

$$F\left(x\right)=\frac{1}{B(a,b)}\int_{0}^{G(x)} w^{a-1}\left(1-w\right)^{b-1}dw; \qquad a>0, b>0$$

Is the c.d.f. of a generalized class of distributions. By taking G(x) as the c.d.f. of the Exponentiated Pareto distribution, we obtain the distribution function of the beta generalized Pareto distribution as:

$$F\left(x\right)=\frac{1}{B(a,b)}\int_{0}^{\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]^{\theta}} w^{a-1}\left(1-w\right)^{b-1}dw; \quad x\geq k,\ a>0, b>0 \tag{1}$$

Equation (1) can be expressed in terms of hypergeometric function as:

$$F(x)=\frac{\left(1-\left(\frac{k}{x}\right)^{\alpha}\right)^{\theta}}{aB(a,b)}\,{}_2F_1\left[a,1-b;a+1;\left(1-\left(\frac{k}{x}\right)^{\alpha}\right)^{\theta}\right]$$

The density function of the new distribution corresponding to (1) is:

$$f(x)=\frac{\alpha\theta k^{\alpha}}{B(a,b)}\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]^{\theta a-1}\left[1-\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}\right]^{b-1}\frac{1}{x^{\alpha+1}},\ x\geq k \tag{2}$$

If X is a random variable with density function given in (2), we say that X~BGP(a,b,α,θ,k). If b>0 is a real non-integer, then equations (1) and (2) can be expressed as:

$$F\left(x\right)=\frac{1}{B(a,b)}\sum_{j=0}^{\infty}\frac{\Gamma b}{j!\Gamma(b-j)}\left(-1\right)^{j}\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]^{\theta(a+j)}\frac{1}{a+j} \tag{3}$$

which is the infinite weighted sum of the distribution functions of the Exponentiated Pareto distributions. Similarly

$$f(x)=\frac{\alpha\theta k^{\alpha}}{B(a,b)}\frac{1}{x^{\alpha+1}}\sum_{j=0}^{\infty}\frac{\Gamma b\left(-1\right)^{j}}{j!\Gamma\left(b-j\right)}\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]^{\theta(a+j)-1},\ x\geq k \tag{4}$$

This is the weighted sum of the Exponentiated Pareto density functions.

Special Cases:

Case1: When $a=\frac{1}{\theta}$ and b=1in (2), we obtain the Pareto distribution with parameters α & k.

Case2: When b=1, the beta generalized Pareto distribution reduces to the Exponentiated Pareto distribution with parameters α ,k, θ and a.

Case3: If θ=1, equation (2) reduces to the beta Pareto distribution with parameters α, k, a and b.

Case4: On substituting $y=1-\left(\frac{k}{x}\right)^{\alpha}$, we get the Weibullized beta distribution with parameters θ, a and b with density:

$$f\left(y\right)=\frac{\theta}{B(a,b)}y^{\theta a-1}\left(1-y^{\theta}\right)^{b-1},\ 0<y<1$$

Case 5: If $y=-\ln\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]$, we obtain the log-beta distribution given by:

$$f(y)=\frac{\theta}{B(a,b)}e^{-\theta a y}\left(1-e^{-\theta y}\right)^{b-1},\ 0<y<\infty.$$

2. PROPERTIES

2.1 Limit behavior of the beta generalized Pareto distribution

As $x\to\infty$, $f(x)\to 0$ and as $x\to k$, we have:

$$\lim_{x\to k} f(x)=\begin{cases}\infty, & 0<\theta<1/a\\ \dfrac{\alpha}{akB(a,b)}, & \theta=1/a\\ 0, & \theta>1/a\end{cases}$$

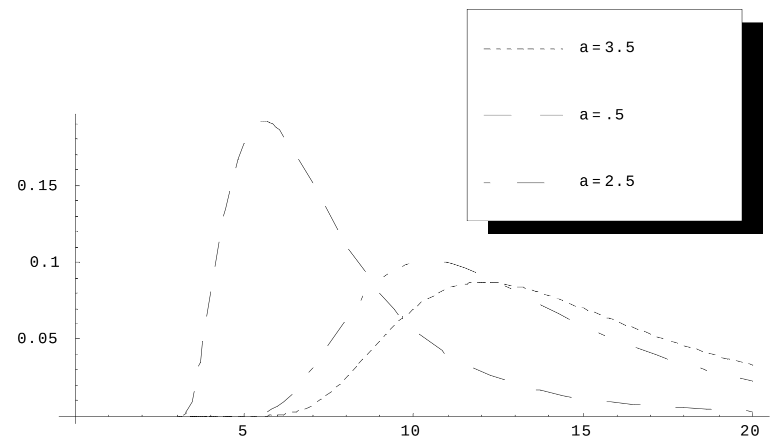


Fig. 1: Density function with θ=10, b=2, α=2, k=3 and different values of a.

2.2 Hazard Rate Function

The hazard rate function of a random variable X is:

$$h(x)=\frac{\alpha k^{\alpha}\theta\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]^{\theta a-1}\left[1-\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}\right]^{b-1}}{x^{\alpha+1}\left[B(a,b)-\sum_{j=0}^{\infty}\frac{\Gamma b(-1)^{j}}{j!\Gamma(b-j)}\frac{\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta(a+j)}}{a+j}\right]} \tag{5}$$

As $x \to \infty$, $h(x) \to 0$ and as $x \to k$, we have:

$$\lim_{x\to k} h(x) = \begin{cases} \infty, & 0<\theta<1/a \\ \dfrac{\alpha}{akB(a,b)}, & \theta = 1/a \\ 0, & \theta > 1/a \end{cases}$$

2.3 Moments

The moment generating function of the beta generalized Pareto distribution is:

$$M_x(t) = \frac{\theta}{B(a,b)} \sum_{j=0}^{\infty} \frac{\Gamma b(-1)^j}{j!\Gamma(b-j)} \sum_{i=0}^{\infty} \frac{t^i}{i!} k^i B\left(\theta(a+j), 1-\frac{i}{\alpha}\right) \quad for\, \alpha > i \tag{6}$$

$$\therefore \mu'_r = \frac{\theta}{B(a,b)} \sum_{j=0}^{\infty} \frac{\Gamma b(-1)^j}{j!\Gamma(b-j)} k^r B\left(\theta(a+j), 1-\frac{r}{\alpha}\right) \quad for\, \alpha > r \tag{7}$$

$$\mu'_1 = \frac{\theta}{B(a,b)} \sum_{j=0}^{\infty} \frac{\Gamma b(-1)^j}{j!\Gamma(b-j)} k\; B\left(\theta(a+j), 1-\frac{1}{\alpha}\right) \quad for\, \alpha > 1$$

$$\mu'_2 = \frac{\theta}{B(a,b)} \sum_{j=0}^{\infty} \frac{\Gamma b(-1)^j}{j!\Gamma(b-j)} k^2 B\left(\theta(a+j), 1-\frac{2}{\alpha}\right) \quad for\, \alpha > 2$$

$$\sigma^2 = \frac{\theta}{B(a,b)} \sum_{j=0}^{\infty} \frac{\Gamma b(-1)^j}{j!\Gamma(b-j)} k^2 B\left(\theta(a+j), 1-\frac{2}{\alpha}\right) - \left[\frac{\theta k}{B(a,b)} \sum_{j=0}^{\infty} \frac{\Gamma b(-1)^j}{j!\Gamma(b-j)} B\left(\theta(a+j), 1-\frac{1}{\alpha}\right)\right]^2 \quad \text{for} \quad \alpha > 2$$

2.4 Mean Deviation

Let the random variable X~BGP (a, b, α, θ, k) with mean μ and median M, then the mean deviation about mean and median are given as:

$$M.D.(\mu) = 2\left\{\mu F(\mu) - \int_k^{\mu} x f(x)\,dx\right\} \text{ (Akinsete et al. 2008)} \tag{8}$$

$$M.D.(M) = E(X) + 2MF(M) - M - 2\int_k^{M} x f(x)\,dx \text{ (Akinsete et al. 2008)} \tag{9}$$

Using equation (4), the integrals given in (8) and (9) can be solved as follows:

$$\int_k^m x f(x)\,dx = \frac{\alpha\theta}{B(a,b)} \sum_{j=0}^{\infty} \frac{\Gamma b}{j!\Gamma(b-j)} \sum_{i=0}^{\infty} \frac{(-1)^{i+j}\,\Gamma\theta(a+j)}{i!\Gamma[\theta(a+j)-i]} \left\{\frac{k^{\alpha(i+1)} m^{1-\alpha(i+1)} - k}{1-\alpha(i+1)}\right\} \tag{10}$$

Substituting (10) in (8) and (9) we get expressions for the mean deviation.

2.5 Shannon Entropy

Entropy is a measure of variation of uncertainty. Shannon entropy is derived as:

$$E\left(-\log f(x)\right) = -\log\left(\alpha\theta k^{\alpha}\right) + \log B(a,b) - \left(a - \frac{1}{\theta}\right)\left[\Psi(a) - \Psi(a+b)\right]$$

$$+(\alpha+1)\log k + \frac{b-1}{B(a,b)}\sum_{i=1}^{\infty}\left\{\frac{B(a+i,b)}{i}\right\} + \frac{1+\frac{1}{\alpha}}{B(a,b)}\sum_{i=1}^{\infty}\left\{\frac{B\left(a+\frac{i}{\theta},b\right)}{i}\right\} \quad (11)$$

3. ORDER STATISTICS

Let $X_{i:n}$ be the ith order statistic in a random sample of size n from the BGP distribution, then the density function of the ith order statistic is :

$$f\left(X_{i:n}\right) = \frac{1}{B(i,n-i+1)}\frac{\alpha\theta k^{\alpha}}{B(a,b)}\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]^{\theta a-1}\left[1-\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}\right]^{b-1}\frac{1}{x^{\alpha+1}}$$

$$\frac{1}{B^{i-1}(a,b)}\left[\sum_{j=0}^{\infty}\frac{\Gamma b}{j!\Gamma(b-j)}(-1)^{j}\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]^{\theta(a+j)}\frac{1}{a+j}\right]^{i-1}$$

$$\left[1-\frac{1}{B(a,b)}\sum_{j=0}^{\infty}\frac{\Gamma b}{j!\Gamma(b-j)}(-1)^{j}\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]^{\theta(a+j)}\frac{1}{a+j}\right]^{n-i}$$

$$f\left(X_{i:n}\right) = \frac{\alpha\theta k^{\alpha}}{B(i,n-i+1)}\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]^{\theta a-1}\left[1-\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}\right]^{b-1}\frac{1}{x^{\alpha+1}}$$

$$\frac{1}{B^{l+i}(a,b)}\sum_{l=0}^{n-i}\binom{n-i}{l}\left[\sum_{j=0}^{\infty}\frac{\Gamma b}{j!\Gamma(b-j)}(-1)^{j+l}\left[1-\left(\frac{k}{x}\right)^{\alpha}\right]^{\theta(a+j)}\frac{1}{a+j}\right]^{l+i-1}$$

Using the expansion $\left(\sum_{i=0}^{\infty} a_i\right)^{l} = \sum_{(m_1,\ldots,m_l)=0}^{\infty} a_{m_1}\ldots\ldots a_{m_l}$ for l a positive integer,(Souza et al. 2008) two expressions for the density of the ith order statistic are derived. If b is real non-integer, we have:

$$f\left(X_{i:n}\right) = \sum_{l=0}^{n-i}\sum_{m_1=0}^{\infty}\ldots\ldots\sum_{m_{l+i-1}=0}^{\infty}\delta_{l,i}^{1} f_{l,i}(x)$$

where $\delta_{l,i}^{1}=\dfrac{\binom{n-i}{l}(-1)^{l+\sum_{j=1}^{l+i-1}m_j}\,\Gamma b^{l+i-1}B\left(a(l+i)+\sum_{j=1}^{l+i-1}m_j,b\right)}{B^{l+i}(a,b)B(i,n+i-1)\Pi_{j=1}^{l+i-1}\left[m_j!(a+m_j)\Gamma(b-m_j)\right]}$ and $f_{l,i}(x)$ is the beta generalized Pareto density with parameters, $a(l+i)+\sum_{j=1}^{l+i-1}m_j,b,\theta,\alpha,k$. In case b is an integer, we have:

$$f\left(X_{i:n}\right)=\sum_{l=0}^{n-i}\sum_{m_1=0}^{b-1}.....\sum_{m_{l+i-1}=0}^{b-1}\delta_{l,i}^{2}f_{l,i}(x)$$

$$\delta_{l,i}^{2}=\frac{\binom{n-i}{l}(-1)^{l+\sum_{j=1}^{l+i-1}m_j}\,\Pi_{j=1}^{l+i-1}\left[\binom{b-1}{m_j}\left(\frac{1}{a+m_j}\right)\right]B\left(a(l+i)+\sum_{j=1}^{l+i-1}m_j,b\right)}{B^{l+i}(a,b)B(i,n+i-1)}.$$

The moment generating function and the rth moment of the order statistic in case b is real non-integer is:

$$M_{X_{i:n}}(t)=\sum_{l=0}^{n-i}\sum_{m_1=0}^{\infty}.....\sum_{m_{l+i-1}=0}^{\infty}\delta_{l,i}^{1}M_{X_{l,i}}(t),\ E\left(X_{i:n}^{r}\right)=\sum_{l=0}^{n-i}\sum_{m_1=0}^{\infty}.....\sum_{m_{l+i-1}=0}^{\infty}\delta_{l,i}^{1}E\left(X_{l,i}^{r}\right)$$

Similarly moments can be obtained for b an integer.

$E\left(X_{l,i}^{r}\right)$ can be obtained from equation (7) for the moments of the BGP distribution with parameters $a(l+i)+\sum_{j=1}^{l+i-1}m_j,b,\theta,\alpha,k$.

4. MAXIMUM LIKELIHOOD ESTIMATION

The log-likelihood function of the BGP distribution is given by:

$$\ln L=n\ln\alpha+n\ln\theta+n\alpha\ln k-n\ln\Gamma a-n\ln\Gamma b+n\ln\Gamma(a+b)$$
$$+(\theta a-1)\sum\ln\left(1-\left(\frac{k}{x}\right)^{\alpha}\right)-(\alpha+1)\sum\ln x+(b-1)\sum\ln\left\{1-\left(1-\left(\frac{k}{x}\right)^{\alpha}\right)^{\theta}\right\}$$

$$\frac{\partial\ln L}{\partial a}=-n\Psi(a)+n\Psi(a+b)+\theta\sum\ln\left(1-\left(\frac{k}{x}\right)^{\alpha}\right)=0 \qquad (12)$$

$$\frac{\partial\ln L}{\partial b}=-n\Psi(b)+n\Psi(a+b)+\sum\ln\left\{1-\left(1-\left(\frac{k}{x}\right)^{\alpha}\right)^{\theta}\right\}=0 \qquad (13)$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} + \sum \left[a - \frac{b-1}{\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{-\theta} - 1} \right] \ln\left(1-\left(\frac{k}{x}\right)^{\alpha}\right) = 0 \tag{14}$$

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} + n\ln k - \sum \ln x + \sum \left[\frac{\theta(b-1)}{\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{-\theta} - 1} - (\theta a - 1) \right] \frac{\left(\frac{k}{x}\right)^{\alpha} \ln\left(\frac{k}{x}\right)}{1-\left(\frac{k}{x}\right)^{\alpha}} = 0 \tag{15}$$

Since $x \geq k$, the MLE of k is $x_{(1)}$ the first order statistic.

Using the MLE of a, b, and α from the beta Pareto distribution, as the initial values, we can solve (14) for θ and then obtain the MLE of the parameters by using iterative techniques. The second partial derivatives are:

$$\frac{\partial^2 \ln L}{\partial a^2} = -n\Psi'(a) + n\Psi'(a+b), \quad \frac{\partial^2 \ln L}{\partial b^2} = -n\Psi'(b) + n\Psi'(a+b)$$

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{n}{\theta^2} - (b-1)\sum \left[\frac{\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta} \left[\ln\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}\right]^2}{\left[1-\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}\right]^2} \right]$$

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = -\frac{n}{\alpha^2} - (\theta a - 1)\sum \frac{\left(\frac{k}{x}\right)^{\alpha} \left\{\ln \frac{k}{x}\right\}^2}{\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^2}$$

$$+\theta(b-1)\sum \frac{\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta-2} \left(\frac{k}{x}\right)^{\alpha} \left\{\ln \frac{k}{x}\right\}^2}{\left[1-\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}\right]^2} \left[1-\theta\left(\frac{k}{x}\right)^{\alpha} - \left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}\right]$$

$$\frac{\partial^2 \ln L}{\partial a\ \partial b} = n\Psi'(a+b)\,,\ \frac{\partial^2 \ln L}{\partial a\ \partial \theta} = \sum \ln\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}$$

$$\frac{\partial^2 \ln L}{\partial a \partial \alpha} = -\theta \sum \frac{\left(\frac{k}{x}\right)^{\alpha}\left\{\ln \frac{k}{x}\right\}}{1-\left(\frac{k}{x}\right)^{\alpha}}$$

$$\frac{\partial^2 \ln L}{\partial b \partial \theta} = -\sum \left[\frac{\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta} \ln\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}}{1-\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}}\right],\ \frac{\partial^2 \ln L}{\partial b \partial \alpha} = \sum \left[\frac{\theta\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta-1}\left(\frac{k}{x}\right)^{\alpha}\ln\left(\frac{k}{x}\right)}{1-\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}}\right]$$

$$\frac{\partial^2 \ln L}{\partial \theta \partial \alpha} = -a\sum \left[\frac{\left(\frac{k}{x}\right)^{\alpha}\ln\left(\frac{k}{x}\right)}{1-\left(\frac{k}{x}\right)^{\alpha}}\right] + (b-1)\sum \frac{1-\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta} + \ln\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}}{\left[1-\left\{1-\left(\frac{k}{x}\right)^{\alpha}\right\}^{\theta}\right]^2}$$

The expectations of the second derivatives are not in closed form, hence numerical methods can be used to obtain Fisher’s information matrix.

REFERENCES

1. Akinsete, A., Famoye, F. and Lee, C. (2008). The beta Pareto distribution. *Statistics* 42(6), 547-563.
2. Eugene, N.L., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Commun. in Statist.: Theo. and Meth.*, 31, 497-512.
3. Gupta, R.C., Gupta, R.D. and Gupta, P.L. (1998). Modeling failure time data by Lehman alternatives. *Commun. in Statist.: Theo. and Meth.*, 27(4), 887-904.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 187-192

CITIES GROWTH AND HEALTH CHALLENGES OF POOR PEOPLES LIVING IN SLUMS OF URBAN AREAS - CASE FROM KARACHI

Muhammad Mazhar Manzoor[1], **Rabia Essa**[1], **Sangita Lal Chand Vaswani**[1], **Muhammad Usman Aleem**[3] and **M. Shariq Usman**[1]

[1] Department of Business Administration, Federal Urdu University of Science and Technology, Karachi, Pakistan Email: mmazher@gmail.com

[2] Department of Computer Sciences, Usman Institute of Technology, Karachi, Pakistan.

[3] Department of Management Sciences, KASBIT, Karachi, Pakistan.

ABSTRACT

Objective:

Rapid population growth reduced the urban living standards significantly and distressed real urban income levels that rapidly increase the poverty rate in urban areas. Cities are dealing with serious problems of housing, shortages of services, weak local governments, lack of medical and healthcare facilities and serious environmental issues. This growing problem creates inimitable challenge for policymakers and public health practitioners. This paper addressing the health challenges of poor urban peoples living in slums and lack of a comprehensive strategy on urbanization and urban poor has complicate the situation, resulting in poor environmental and living conditions in slum settlements, weak access to basic services and exposure to evictions and exploitation. It is clear that without effective and representative urban management it intricate to develop the living standard of slums and informal settlement residents that must require good local governance, involvement of all stake holder such as provincial and federal Government, NGO's and political commitment with all parties. Ensure provision of basic infrastructure and services such as water, education and also control of pollution are necessary to overcome the health challenges of urban poor.

Data Source:

Primary data source for this basal study is library based while secondary data obtained from publish sources and may be jolt down by utilizing existing survey, open and closed end questionnaires etc.

Methodology and Policy Relevance:

For this study we restrict population size to few goath and Kachi Abadis or informal Settlements of Karachi while random and stratified sampling is use. This Study is helpful to identify the health challenges of poor peoples living in slums of urban areas and meanwhile we opt such policies on national level which can overcome the health challenges of these areas.

KEYWORDS

Population Growth; Migration; Livelihood; Urbanization; Informal Settlements; Health Challenges.

INTRODUCTION

The history of urbanization in Pakistan can be alienated into four essential phases in which both the out migration and in migration have occurred. These phases are as. The first phase of urbanization started with the partition of British Indian Empire. When more than six million people migrated towards new born country. Majority of them settled in urban areas of Pakistan. The second phase started in the mid 1950s the government starts the process of industrialization. The opportunities of good life and employment encouraged million of rural peoples migrate to cities. The third phase stared with the breakup of East Pakistan in 1971. All the supporters of Pakistan in Bangladesh were forced to leave the country. The four but very complex and serious phases in 1980s, at this times changing in neighboring countries, such as Iran and Afghanistan greatly added to already continuing process of urbanization in Pakistan. Iran-Iraq war, Iranian rebellion and Soviet incursion on Afghanistan obliged many people to seek refuge in Pakistan. (Arif, 1997); Urbanization process has posed various critical and serious troubles. The high gap between the rural and urban development and income results into large percentage of rural to urban migration. Due to high population growth in urban areas there is a pressure for creation of new job opportunities, strengthening of present health facilities, construction of houses etc. The migration of peoples and families towards cities put incredible pressure on the resources of these localities. After sometimes it becomes a real challenge for the administration to preserve and maintain the essential facilities and the standard of life in the cities. (Dixon, 1997); the resources and services of these urban sectors are being divided over an increasing population. That increase the gap between the demand and supply of public services and rise to many problems such as the emergence of slums, shortage of resources such as water and electricity, inadequacy of public transport and increasing crime rate along with social costs such as ethnic conflicts also infrastructure, housing and shortage of other social services in urban areas may rise to many problems. Due to the distribution of resources over a larger population, people are facing delicate power and water shortages. Transportation sectors of the urban areas are also not able to manage with the increasing demand of public transport due to these problems the expectations of the migrants are not fulfill that frustrate them, thus giving rise to violent and aggressive behavior. Cross border illegal migration increase the additional risk of fraud and crimes due to unsatisfactory planning and management authorities' are unable to control cross border illegal migration. (Mazhar, 2010); In rural areas people face numerous problems not only regarding their income and jobs but they also lack access to basic needs of life, also unequal distribution of water, land & other resources and jagirdari system discourage rural peoples and Rural – urban income gap, industrialization and better paid jobs encourage them and they see urban areas as the solution of their problems. On other hand many of educated and skilled urban peoples are jobless in their homeland due to limited absorption capacity of both industrial and urban social services. Cross border and rural urban migration relative to economic status may create serious problems of unemployment and poverty.

LITERATURE REVIEW

In-migration play a key role in economic development of rural peoples and as well as In-migration can modify the conditions of supply of entrepreneurship in those areas where there is a strong and favorable demand side condition and In-migration can contribute to the process of economic development through entrepreneurship and increased rural urban interdependencies. This means that the role of new arrivals in the countryside is in the main cumulative, rather than transformational. (Kalantaridis Christos, 2010); Like many developing countries, the problems of unemployment and poverty has begin to appear in China as a result of rapid urbanization in the last two decades, along with other problems such as a progressive excess numbers of housing and social services, increased crime, pollution, and overcrowding. Given the serious problems, it is desirable for us to know the factors contributing to the urbanization and city ward migration so that corresponding policies may be derived. Urban Population Growth has increased through the contribution of rapid rural urban migration. Internal migrants were encouraged by the rural–urban income gap and depressed by their geographic distances from origin to destinations and the amount of migrants is positively related to rural–urban income gap and urban population in that province. (Zhang K., Song S., 2003); the political, social, and economic significances of the economic and social growth in the minority areas have a very close connection with the modernization of the whole country's politics, society, and economy. The movement of rural migrants to the urban areas or to the nonagricultural industries is one of the key issues in the economic and social development of both sending and receiving societies between which the migrant move. (Lu Qi, Yang Chunyue, 2008); Migration is an important process of transformation for rural populations in developing countries. Migration has been recognized as an independent individual risk factor for the acquisition of HIV. (Coast Ernestina, 2006); Pakistan has become a destination where large numbers of cross-border migrants and refugees are settled. These migrant groups, together with the growing number of rural people displaced by agricultural modernization and mechanization, have contributed to the extensive increase in the levels of urbanization in Pakistan, particularly in the more industrialized provinces of Punjab and Sindh. At the same time, like the people of so many low- and middle-income nations, Pakistani citizens have sought work abroad, and in the 1970s large-scale labor migration to the Middle East began in earnest. Remittances have since become an important component of the national economy and of the livelihoods of many households. These complex and substantial movements have resulted in profound changes in settlement patterns, and also in deep socioeconomic and cultural transformations. Smaller urban centers, such as the ones described in this paper, reflect the growing discrepancy between changing values and widening economic opportunities on the one hand, and the perseverance of a feudal system of political power often supported by a highly contentious managerial and political decentralization plan, on the other hand. (Hasan A., Raza M., 2009); Migration can cause of various socio-economic problems and factors of human security, housing, and unemployment, food that put extra burden on existing infrastructure and increase unhealthy competition of bread and butter. Due to industrialized vision, Karachi face rapid process of urbanization and lack of proper planning and management it become a critical and challenging task for the administration of Karachi. Many peoples from all over the Pakistan and cross border migrate to Karachi through proper and illegal channels for better employment and livelihood. Government enables to register them properly, that the reason they cannot create human securities for them. It increases the burden on infrastructure and resources of Karachi because services and resources of urban sector are divided into increasing population and create various socio economic problems. When expectation of having good life and better employment are not accomplish, they become frustrate and deprivation and they involve in illegal or wrong activities such as mobile snatching and organized crime such as trafficking in narcotics and arms also increase their exposure of joining political or religious extremist groups. (Manzoor, M., 2010); Majority of Migrants move from rural to urban areas for better employment, enhancing their source of income and themselves for social facilities is better health and better marketing facilities. Migration improves the house hold income, quality of life and promoted the socio economic status of the migrated families. also women empowerment increase after migration because in the absences of men, women take part in decision making process on the other hand migrants put pressure on the civic institutions because these institutions are not able to accomplish the demand of increasing population. (Farooq, M. and Cheema, A.M., 2005).

ANALYSIS & DISCUSSION

1. Security Relevance of Migrants and Emigrants:

Human security relevance, which is emerging paradigm for understanding global vulnerabilities, it holds the sense of possession and easy access to the needs of life; Food, Health, Environmental, Personal, Community, Political and Economic security.

Irregular movement of people generate result for human security - whether it is reasons such as poverty and lack of employment, threat of political violence and absence of rule and regulation and environmental insecurities. Moreover, the rise of cross-border issues such as the spread of infectious diseases (e.g. HIV/AIDS, Bird flu, Hepatitis ABC, Dangi virus etc.) and drug trafficking call for greater securitization. Yet, the methods adopted by states in securitizing these issues vary and thus have adverse impacts on the human security of migrants’ as well as emigrants.

2. Health Contagion:

As many people live in Katchi Abadi and slums the health problems are severe in those areas. As the case of polio continuously programs by Govt. of Pakistan are going on to eliminate this disease from the country but it is not eliminating due to Afghan mohajrins.

In general, some problems associated with or exacerbated by human over population;

- Inadequate fresh water for drinking water use as well as sewage treatment and effluent discharge. Some countries, like Saudi Arabia, use energy-expensive desalination to solve the problem of water shortages.
- Depletion of natural resources, especially fossil fuels.
- Increased levels of air pollution, water pollution, soil contamination and noise pollution. Once a country has industrialized and become wealthy, a combination of government regulation and technological innovation causes pollution to decline substantially, even as the population continues to grow.
- Deforestation and loss of eco-systems that sustain global atmospheric oxygen and carbon dioxide balance; about eight million hectares of forest are lost each year.
- Changes in atmospheric composition and consequent global warming.
- Irreversible loss of arable land and increases in desertification.

- High infant and child mortality. High rates of infant mortality are caused by poverty. Rich countries with high population densities have low rates of infant mortality.
- Starvation, malnutrition or poor diet with ill health and diet-deficiency. Famine is aggravated by poverty. Rich countries with high population densities do not have famine.
- Poverty coupled with inflation in some regions and a resulting low level of capital formation.
- Poverty and inflation are aggravated by bad government and bad economic policies. Many countries with high population densities have eliminated absolute poverty and keep their inflation rates very low.
- Low life expectancy in countries with fastest growing populations.
- Un-hygienic living conditions for many based upon water resource depletion, discharge of raw sewage and solid waste disposal.
- Elevated crime rate due to drug cartels and increased theft by people stealing resources to survive.
- Conflict over scarce resources and crowding, leading to increased levels of warfare.
- Over-utilization of infrastructure, such as mass transit, highways, and public health systems higher land prices.

3. Present Trends and Future Dangers

The general climate of insecurity pervading Karachi persists. Horizontal polarization of ethnic and religious groups is unabated, as is vertical polarization by economic stratum. The city's educational system is now crippled: some colleges have been forced to close, and others serve as armed strongholds for warring factions. Education is increasingly privatized and segregated along class lines.

More broadly, there are mounting concerns over the potential for conflict arising from major development projects that promote national economic growth at the expense of local communities. A good example is the Ghazi-Barotha hydropower project. Consisting of a barrage, a power channel, and a 1,425-megawatt generating complex, the project aims to use the drop of the Indus River between the tailrace of the Tarbela Dam and the confluence of the Indus and Haro Rivers to produce electricity. It promises to provide a much-needed renewable and emission-free source of energy to the country.

CONCLUSION & RECOMMENDATION

Conclusion:

Urban population growth has also increased through the contribution of rapid rural-urban migration, especially in the industrialized provinces of Sind and Punjab. In rural areas people face numerous problems not only regarding their earning and jobs but also lack access to basic needs such as education, health and proper sanitation, many of them see urban areas as the solution of their problems also rural-urban income gap, better paid jobs and industrialization encourage rural peoples migrate to urban areas. On the other hand lack of capital, shortage of power and iron, narrowness of home and foreign markets for the manufactured goods, frequent breakdown of electricity, lack of proper planning, use of backward technology are major obstacle in the expansion and growth of industrial sector in Pakistan, that's the reason many educated and skilled urban peoples are jobless in their homeland.

Recommendation:

It is recommended that for future policy making Governments must work to draw up proper Urban Development Plans together with the incorporation of communities and the private sectors in land delivery, and the devolution and reformation of procedures, increasing the assets of the urban poor, and provide access to basic health care, water and proper sanitation at subsidize rate and promote return migration on large scale.

Also take initiatives to control rural- urban migration by facilitating peoples with micro credit schemes, job opportunities, incentives oriented monetary policy, establishing schools, collages and other vocational training institutes at village level and also strictly imposes check on cross border illegal migrants with proper registration and also take action against corrupt agents that promote illegal migration

REFERENCES

1. Faooq, M., and Cheema, M.A. (2005). Likelihood Impact of Internal Migration on Family Institution in Punjab, Pakistan. *Journal of Agricultural and Social Sciences*, 1/2, 148-151.
2. Guest, P. (1998). The Dynamics of Internal Migration in Vietnam. *UNDP Discussion Paper 1*. UNDP, Hano.
3. Government of Pakistan (1998). *Population and Housing Census of Pakistan*. Census Bulletin-1. Population Census Organization Statistics Division, Islamabad
4. Goldscheider, C. (1971). *Population, Modernization, and Social Structure*. Little Brown & Co., Boston.
5. Hasan, A. and Raza, M. (2009). Migration and Small town in Pakistan. *Working paper series on Rural-Urban Interactions and Livelihood Strategies*, 15.
6. Harberle, R. (1983). The causes of rural-urban migration; A study survey of German Theories. *American Journal Sociology*, 932-950.
7. Jan, B., Iqbal, M. and Iftikharuddin (2008). Urbanization trend and urban population projections of Pakistan using weighted approach. *Sarhad J. Agric*., 42/6, 173-180.
8. Kalantaridis, C. (2010). In-migration, entrepreneurship and rural-urban interdependencies: The case of East Cleveland, North East England. *Journal of Rural Studies*, 1-10.
9. Lu, Qi. and Chunyue, Y. (2008). Characteristics of Rural Labor Emigration in Minority Areas of Honghe, Yunnan and Its Impact on New Socialist Countryside Construction in the 21st Century. *China Population Resources and Environment*, 18/4,085-088.
10. Managalam, I.J. (1978). Study of migration: A theoretical approach and some implication for research and planning. *Journal of Central Asia*, ½.
11. Mazhar, M.M., Ahmed, I. and Agha, S. (2010). Urban growth & environment-contemporary challenges and need for integration. *5th International Conference on Statistical Sciences*, Lahore, Pakistan.
12. *Microsoft Encarta Encyclopedia Deluxe*, 2002.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 193-198

CONTROLLING THE FALSE ALARM RATE OF *R* CHART FOR NORMAL AND NON-NORMAL PROCESSES

Muhammad Saeed[1], **Muhammad Ali Shah**[2] and **Saddam Akbar Abbasi**[3]

[1] Statistics & DWH Department, State Bank of Pakistan, Karachi, Pakistan. Email: muhammad.saeed2@sbp.org.pk

[2] Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan. Email: muhammadalishahbokhari@yahoo.com

[3] Department of Statistics, University of Auckland, Auckland, New Zealand. Email: sabb025@aucklanduni.ac.nz

ABSTRACT

Control chart introduced by Walter A. Shewhart in 1920, is the most important and widely used tool to monitor any industrial or analytical process. The basic purpose of implementing control chart procedures is to detect unfavorable variations in the process (location & scale) parameters or in other words identifying special causes. For monitoring process dispersion, *R* chart is the widely used statistical process control chart. Control chart constants (d_2 & d_3) and the quantile points of the relative range required for the design of the *R* chart are mostly available for parent normal distribution. Mahoney (1998) and Kao and Ho (2007) derived control chart constants for parent non-normal distributions for the purpose of controlling the false alarm rate. The results reported in Kao and Ho (2007) indicates that the false alarm rate of the *R* chart alters with the change in the sample size which seems to be due to the inappropriate use of 3-sigma limits. This study aims at providing the suitable probability limits of the *R* chart for non-normal processes which will help quality practitioners to achieve a desired false alarm rate.

1. INTRODUCTION

Statistical process control (SPC) deals with measuring and analyzing the variation in processes and also maintaining the processes to a fixed target by using statistical techniques. A primary tool used for SPC is the control chart, proposed by Walter A Shewhart in 1920 for the purpose of improving the economic effectiveness of the telephony transmission system during his work in Bell Labs. The basic objective of implementing control chart procedures is to distinguish between process variation resulting from common causes and special causes. There are three basic control chart types namely Shewhart, Cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) charts. Shewhart charts are quite good at detecting large shifts in process parameters, however for quick detection of small process shifts EWMA and CUSUM charts are more effective. A lot of research is available on control chart methodologies and applications in a wide variety of disciples such as in analytical laboratories (Masson (2007), Abbasi (2010)), nuclear engineering (Hwang et al. (2008), health care (Woodall (2006)), education (Wang and Liang (2008)) etc.

For monitoring of any process, samples are usually taken in the form of rational subgroups of size say *n*. The mean μ is generally estimated from the overall sample mean $\bar{\bar{X}}$ while process standard deviation σ can be estimated in many different ways; the most popular choices are based on sample range (*R*) or sample standard deviations (*S*) – see Montgomery (2001).

2. DESIGN OF *R* CHART

For monitoring process dispersion the *R* chart is the widely used control chart. *R* chart is based on estimating the process standard deviation from average of sample ranges ($\bar{R}$). *R* is defined as

$$R = X_{\max} - X_{\min} \tag{1}$$

where $X_{\max}$ and $X_{\min}$ represents the largest and smallest values of the observed sample $X_1, X_2, ..., X_n$ of size *n*. The design structure of *R* control chart is based on plotting the sample range *R* against time or sample number. The chart triggers an out of control signal when one or more *R* values lie outside the following control limits:

$$LCL = \bar{R} - 3\frac{d_3\bar{R}}{d_2}$$

$$UCL = \bar{R} + 3\frac{d_3\bar{R}}{d_2} \tag{2}$$

where d_2 and d_3 are the control chart constants depending upon the sample size *n* and are defined as the mean and standard deviation the distribution of sample range, i.e. $E(R) = d_2$ and $\sigma_R = d_3$.

In most SPC books, control chart constants for the *R* chart are provided under the assumption of normality of quality characteristics. When the assumption of normality is disturbed the use of these constants no longer remains valid as shown by Mahoney (1998). He considered different non-normal distributions and examine the effect of parent distribution on the values of d_2. He concludes that inappropriate use of d_2 values increases the false alarm rate of $\bar{X}$ chart. Following Mahoney (1998), recently Kao and Ho (2007) examined the influence of parent non-normal distributions on d_3 values, and have also shown through simulations that the use of d_2 and d_3 values that are computed under the assumption of normality, results in an increase in false alarm rate of *R* chart for non-normal processes. Results shown in Kao and Ho (2007) are a bit surprising as they could not be able to achieve a specified false alarm rate. In fact for some distributions such as uniform and triangular, there is an increase in false alarm rates due to their computed constants. This seems to be due to inappropriate use of 3-sigma control limit because 3-sigma limits should only be used in those situations where one can safely assume that the distribution of plotted statistic is symmetric; the distribution of range is highly asymmetric even for parent normal distribution as can be seen from Figure 1. Figure 1 represents the density plots of sample range under normal and a wide range of non-normal distributions for *n = 10*. The density plots clearly show the skewed nature of the range distribution for almost all the cases.

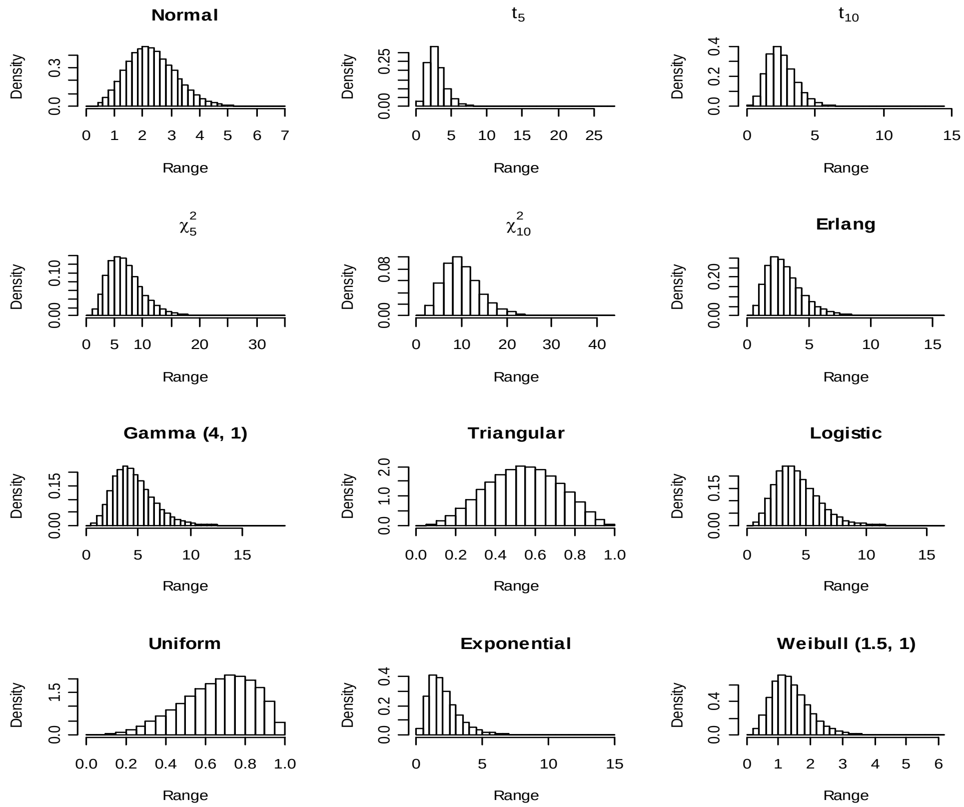


Fig. 1: Density plots of sample range for parent Normal and different Non-Normal distributions when *n = 10*

When the distribution of plotted statistic is asymmetric the use of 3-sigma limits results in an increase in the false alarm rate of *R* control charts. The same point has been raised by Ryan (2000), "Number of false alarms will be much greater with 3-sigma limits than that the user would expect". Hence a better approach is to use probability limits as compared to 3-sigma limits for the design structure of control charts. Probability limits for the *R* chart can be computed by using the quantile points of distribution of relative range (*W*) where *W* is defined as $W = R/\sigma$.

Let α be the specified probability of making Type-I error, denoting α-quantile of the distribution of *W* by W_α, the probability limits of the *R* chart are given as:

$$LCL = W_{\alpha/2}\bar{R}/d_2$$
$$UCL = W_{(1-\alpha/2)}\bar{R}/d_2 \qquad (3)$$

Similarly the use of quantile points that have been computed under the assumption of normality is not suitable for setting up probability limits for processes following non-normal distributions. Hence these quantile points must also be computed by giving proper consideration to the parent distribution. In most SPC books, quantile points of the distribution of relative range are only provided for parent normal distribution because for majority of the parent non-normal cases the distributional results of W are not well known. But one can estimate the characteristics of any distribution such as mean, standard deviation and cumulative probabilities through repeated generation of random samples. Hence we make use of comprehensive simulation routines for the computation of required results. In the next section we briefly describe the Monte Carlo simulation steps that have been performed for the computation of the quantile points of W for some non-normal cases portability limits and also compare the false alarm rates achieved by using the proposed probability limit approach and the 3-sigma limit approach used by Kao and Ho (2007).

3. SIMULATION AND COMPARISON

In this study we performed an extensive Monte Carlo simulation study to compute the probability limits of the R chart for non-normal parent distributions for the purpose of controlling the false alarm rate. Following Kao and Ho [2007] we used uniform distribution to represent the case of heavy tailed symmetric distributions and exponential distribution to represent the case of skewed distributions. The density function of these distributions is given below:

Uniform Distribution:

$$f(x) = 1 \qquad 0 \le x \le 1$$

Exponential Distribution:

$$f(x) = \lambda e^{-\lambda x} \qquad x \ge 0, \quad \lambda > 0$$

100,000 random samples of size n are simulated from a particular distribution and the distribution of W is obtained. For a specified Type-I error probability α, $(\alpha/2)th$ and $(1-\alpha/2)th$ quantile points have been computed from this empirical distribution of W and the results have been reported in Table 1 and 2 for uniform and exponential distributions respectively considering some representative values of n and α. We expect that the use of these quantile points that have been computed based on the parent distribution, helps in controlling the false alarm rate of the R chart for these non-normal cases. Next we evaluate the false alarm rate of the R chart by using our computed probability limits and make the comparison with the results reported in Kao and Ho (2007). Kao and Ho (2007) used the 3-sigma control limits as described in (2) hence for a valid comparison the probability limits have also been computed for $\alpha = 0.0027$ because when 3-sigma limits are used, a control chart user expects to observe a false alarm rate of 0.0027. The false alarm rate for both the approaches has been computed using 100,000 samples of size n and the results are reported in Table 3. For normal processes the quantile points of W reported in Montgomery (2001) have been used.

We can clearly see from the results of Table 3, that the false alarm rate of R chart reported in Kao and Ho [2007] changes with the change in the sample size n even for parent normal distribution. This seems to be due to inappropriate use of 3-sigma limits. The increase in the false alarm rate is even more significant for the non-normal processes. The false alarm rate achieved by the use of our proposed probability limits remain very close to the desired value of 0.0027. Although results are only provided for $\alpha = 0.0027$ and for a couple of non-normal cases, but similar results have also been observed for other values of n and different parent non-normal distributions. This study will help quality practitioners to choose the appropriate probability limits for non-normal processes.

CONCLUSIONS

This study investigates the design of R chart for non-normal processes. We have seen that it is difficult to achieve a desired false alarm rate by the use of 3-sigma limits particularly for non-normal processes. The probability limit approach proposed in this study helps in controlling the false alarm rate to the desired level for any value of n. Although results are provided for a couple of non-normal cases but the method can be generalized to any non-normal process.

Table 1:
Quantile points of the distribution of *W* for Uniform Distribution

n	$W_{0.0010}$	$W_{0.0014}$	$W_{0.0050}$	$W_{0.0100}$	$W_{0.0500}$	$W_{0.1000}$	$W_{0.9000}$	$W_{0.9500}$	$W_{0.9900}$	$W_{0.9987}$	$W_{0.9990}$
2	0.0018	0.0025	0.0087	0.0176	0.0871	0.1769	2.3675	2.6921	3.1174	3.3292	3.3457
3	0.0642	0.0735	0.1463	0.2071	0.4710	0.6847	2.7879	3.0009	3.2615	3.3909	3.4022
4	0.2128	0.2419	0.3891	0.4849	0.8559	1.1061	2.9683	3.1233	3.3154	3.4067	3.4163
5	0.4311	0.4647	0.6504	0.7815	1.1848	1.4355	3.0761	3.1985	3.3490	3.4233	3.4301
6	0.6494	0.6780	0.8751	1.0219	1.4454	1.6945	3.1456	3.2500	3.3717	3.4336	3.4377
7	0.8355	0.8724	1.0901	1.2377	1.6592	1.8991	3.1951	3.2809	3.3881	3.4374	3.4416
8	0.9718	1.0241	1.2744	1.4161	1.8331	2.0599	3.2258	3.3015	3.3953	3.4401	3.4427
9	1.1556	1.1941	1.4325	1.5841	1.9774	2.1884	3.2537	3.3220	3.4035	3.4429	3.4455
10	1.2728	1.3340	1.5702	1.7226	2.1005	2.3000	3.2758	3.3370	3.4102	3.4454	3.4481
20	2.1648	2.2024	2.3710	2.4646	2.7135	2.8359	3.3709	3.4012	3.4382	3.4547	3.4558
25	2.3844	2.4085	2.5551	2.6400	2.8509	2.9546	3.3911	3.4148	3.4431	3.4567	3.4578
50	2.8648	2.8882	2.9783	3.0246	3.1463	3.2018	3.4275	3.4396	3.4537	3.4604	3.4609

Table 2:
Quantile points of the distribution of *W* for Exponential Distribution

n	$W_{0.0010}$	$W_{0.0014}$	$W_{0.0050}$	$W_{0.0100}$	$W_{0.0500}$	$W_{0.1000}$	$W_{0.9000}$	$W_{0.9500}$	$W_{0.9900}$	$W_{0.9987}$	$W_{0.9990}$
2	0.0011	0.0013	0.0050	0.0102	0.0512	0.1057	2.2970	2.9817	4.5639	6.5074	6.7554
3	0.0346	0.0409	0.0756	0.1083	0.2537	0.3833	2.9748	3.6834	5.3245	7.3569	7.6134
4	0.1031	0.1196	0.1892	0.2412	0.4601	0.6257	3.3716	4.0800	5.7026	7.7530	7.9924
5	0.1870	0.2014	0.3051	0.3777	0.6361	0.8256	3.6502	4.3776	5.9424	7.7891	8.0864
6	0.2891	0.3073	0.4225	0.5054	0.7897	0.9916	3.8698	4.5867	6.2151	8.0804	8.3806
7	0.3731	0.4000	0.5271	0.6165	0.9303	1.1401	4.0562	4.7472	6.3819	8.3616	8.7901
8	0.4777	0.5035	0.6386	0.7293	1.0522	1.2691	4.2018	4.9243	6.5159	8.5889	8.8332
9	0.5341	0.5619	0.7193	0.8235	1.1601	1.3862	4.3413	5.0581	6.6723	8.5894	8.9544
10	0.6242	0.6513	0.8055	0.9126	1.2574	1.4873	4.4569	5.1675	6.8135	8.7372	8.9736
20	1.1838	1.2260	1.4222	1.5479	1.9341	2.1761	5.2118	5.9264	7.5854	9.5845	9.8768
25	1.4000	1.4374	1.6244	1.7412	2.1419	2.3926	5.4445	6.1577	7.7993	9.8400	10.0558
50	2.0391	2.0805	2.2763	2.4081	2.8244	3.0809	6.1397	6.8743	8.5380	10.5403	10.8497

Table 3:
False alarm rate comparison of the proposed method with Kao and Ho (2007)

n	**Normal**		**Uniform**		**Exponential**	
	Kao and Ho (2007)	Proposed	Kao and Ho (2007)	Proposed	Kao and Ho (2007)	Proposed
2	0.0091	0.0026	0.0000	0.0027	0.0184	0.0026
3	0.0059	0.0026	0.0000	0.0026	0.0155	0.0027
4	0.0049	0.0027	0.0000	0.0027	0.0144	0.0028
5	0.0046	0.0027	0.0014	0.0026	0.0138	0.0026
6	0.0045	0.0027	0.0035	0.0026	0.0135	0.0026
7	0.0044	0.0027	0.0052	0.0026	0.0132	0.0026
8	0.0043	0.0026	0.0066	0.0026	0.0132	0.0028
9	0.0043	0.0028	0.0074	0.0028	0.0130	0.0027
10	0.0043	0.0027	0.0080	0.0027	0.0129	0.0026
20	0.0045	0.0027	0.0114	0.0027	0.0123	0.0026
50	0.0050	0.0027	0.0132	0.0028	0.0121	0.0026

REFERENCES

1. Abbasi, S.A. (2010b). On the performance of EWMA chart in presence of two component measurement error. *Quality Engineering 22*(3), 199-213.
2. Hwang, S.L., Lin, J.T., Liang, G.F., Yau, Y.J., Yenn, T.C. and Hsu, C.C. (2008). Application control chart concepts of designing a pre-alarm system in the nuclear power plant control room. *Nuclear Engineering and Design*, 238(12), 3522-3527.
3. Kao, S.C. and Ho, C. (2007). Robustness of R-chart to non normality. *Commun. in Statist.: Simula. and Compu.*, 36(5), 1089-1098.
4. Mahoney, J.F. (1998). The influence of parent population distribution on d2 values. *IIE Transactions (Institute of Industrial Engineers)*, 30(6), 563-569.
5. Masson, P. (2007). Quality control techniques for routine analysis with liquid chromatography in laboratories. *Journal of Chromatography A*, 1158(1-2), 168-173.
6. Montgomery, D.C. (2001). *Introduction to statistical quality control* (4th Ed.). New York: Wiley.
7. Ryan, P.R. (2000). *Statistical methods for quality improvement* (2nd Ed.). New York: Wiley.
8. Wang, Z. and Liang, R. (2008). Discuss on applying SPC to quality management in university education. In *Proceedings of the 9th International Conference for Young Computer Scientists, ICYCS 2008*, 2372-2375.
9. Woodall, W.H. (2006). The use of control charts in health-care and public-health surveillance. *Journal of Quality Technology,* 38(2), 89-104.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 199-206

A COMPARISON OF THE CLUSTERING PATTERNS OF SOCIO-ECONOMIC FACTORS IN THE PUNJAB (PAKISTAN)

Shahla Ramzan[1], **M. Inayat Khan**[2], **Shumila Ramzan**[2] and **Sajid Rasul**[3]

[1] Government College University, Faisalabad, Pakistan.
Email: shahla_ramzan@yahoo.com
[2] University of Agriculture, Faisalabad, Pakistan.
[3] Punjab Bureau of Statistics, Lahore, Pakistan.

ABSTRACT

The main objective of this paper is to form and compare the clusters of the districts of Punjab on the basis of their socio-economic development. Therefore we use multivariate statistical technique – cluster analysis. For tracking the progress towards Millennium Development Goals for Education, Health, Water Supply & Sanitation, the government of Punjab has conducted two multiple indicator cluster surveys (MICS) so far (2003-04 & 2007-08). The comparison is made using the data of MICS (2003-04) and (2007-08). The results show various features. The government can use the results of this study for future policy making.

KEYWORDS

Socio-economic development; Cluster analysis; Multiple Indicator Cluster Survey; MDGs,

INTRODUCTION

The government of the Punjab is committed to reduce poverty by sustaining high growth in all aspects of provincial economy. An abiding challenge in maintaining such growth pattern is concurrent development of capacities in planning, implementation and monitoring which requires reliable and real time data on development needs, quality and efficacy of interventions and impacts. Being aware of this need, Planning and Development Department is consistently working on improved systems for generation of accurate information on development needs and impacts of the investments made through development budgets.

Social sectors remain a priority area for the government and development outlays for these sectors have grown multifold over the last five years. Government of the Punjab is committed to attain the Millennium Development Goals (MDGs) for education, health, water supply & sanitation and poverty. That would require not only provision of adequate resources but also a very robust system for ascertaining the area specific needs, efficient use of resources and regular monitoring of the results and impacts. Towards this end, government, with assistance of UNICEF, has embarked upon periodic conduct of Multiple Indicator Cluster Surveys (MICS). The first such survey was carried out in 2003-04 and proved to be the most important tool in determining government budgetary outlays for the next four years, particularly for the social sectors.

Multiple Indicator Cluster Survey (MICS) in Pakistan: A History

In Pakistan, a Multiple Indicator Cluster Survey was conducted in 1995 by a private sector at national level. The segregation was done at provincial level. By the assistance of UNICEF, first provincial Multiple Indicator Cluster Survey (MICS) was conducted in KPK in 2000-01 and the data was collected at district level. Then in the next phase, these surveys were conducted in other three provinces (Punjab, Sindh & Baluchistan). In FATA and AJK also, MICS was conducted. The second Multiple Indicator Cluster Survey has been conducted in Punjab in 2007-08. Now the third MICS (2010-11) is in progress in Punjab at district level.

Punjab MICS 2003-04 & 2007-08

The Punjab MICS 2003-04 was based on 40 indicators. This survey covers almost 30,000 households. The sample of the survey is large enough even to provide statistically representative results at district level. The usage of information generated by the survey pointed to the need of further improvements in the scope and coverage of selected indicators. Therefore, scope of MICS 2007-08 was further expanded to more than 70 indicators and the coverage was extended down to 'Tehsil' level. The information for more than 91,000 households was collected. This posed much greater challenges in terms of logistics and man-management but it is matter of great satisfaction that final product has proven to be worth that effort.

In this paper, the data from MICS 2003-04 & MICS 2007-08 is used for a comparison purpose. The main objective of this paper is to analyze and see whether the government has been successful in achieving its goals towards the progress of the country.

METHODOLOGY

Most of the observable phenomena in the empirical sciences are of a multivariate nature. Practically we have to deal with huge data sets with high dimensions. To make sense out of these data we have to handle the information based on lots of variables at the same time. Multivariate statistical techniques help us to look up these data in a more pleasant way. Multivariate statistical analysis is concerned with analyzing and understanding data in high dimensions.

Multivariate statistical analysis has been widely applied to socioeconomic problems and particularly to the classification of different types of administrative divisions (municipalities, counties or regions). Studies include Cziráky *et al.* (2005), Aragon *et al.* (2003), González and Morini (2000), Soares *et al.* (2003), Peschel (1998), Pettersson (2001) and Rovan and Sambt (2003). Those studies are restricted to a smaller area inside Europe, specifically Croatia, the Midi-Pyrénées Region, Tenerife Island, Portugal, the Baltic Sea countries, a Swedish county, Slovenia and the Spanish region of Galicia in their respective cases. There are other contributions outside of Europe e.g. Stimson *et al.* (2001) focused on the United States of America and Hill *et al.* (1998) on Australia.

Since our objective is to observe which districts are close to each other in sense of indicators. Therefore we need a technique of classifying the individual/objects into one of many mutually exclusive unknown groups on the basis of some characteristics of interest. Cluster analysis is concerned with forming groups of similar objects based on several measurements of different kinds made on the objects. The key idea is to identify classifications of the objects that would be useful for the aims of the analysis. Ward's method is used for making the groups of districts.

Data and descriptive statistics

The data used in this study is taken from the two MICS (2003-04 and 2007-08). Various indicators for education, health and water supply & sanitation are selected. The descriptive statistics of the socio-economic indicators are shown in Table 1. The mean, median and mode for almost all the indicators are not too different, therefore normality is not suspected in the data. Also the data is a random sample so we are not worried about the assumptions underlying the statistical procedures.

The average literacy rate for age 10 years and above is 56.26 in 2003-04 and increased to 57.74 in 2007-08. Median and mode of the literacy rate also increased in 2007-08. But this increase in the literacy rate is not as large as it should be. The maximum and minimum literacy rate is 78 and 34 respectively in 2003-04 whereas 81 and 33 respectively in 2007-08. The MDG set for literacy rate is 88 percent, which is not reached even in 2007-08. Similarly the descriptive statistics of all other variables are also not very good. None of the districts has achieved any of the millennium development goals.

RESULTS & DISCUSSION

Using Wards method for clustering the districts are classified into five groups for MICS 2003-04 & 2007-08. These clusters show interesting results. Cluster 1 consists of 7 districts in MICS 2007-08, previously in 2003-04 there were only 3 districts in this cluster. Some of the districts are shifted from one cluster to the other. For example Attok, Gujrat, Sialkot and Toba Tek Singh are in cluster 1, which was previously consisting of only three districts Chakwal, Jhelam and Rawalpindi. Especially Toba Tek Singh district is present in the first cluster of the most developed districts indicating much improvement. While according to MICS (2003-04) Toba is in the cluster 4. If we look at the indicators of Toba, clearly this district has improved the literacy rate, LHW, births by skilled attendants, use of contraceptive, water and sanitation facilities. The Infant mortality rate has been decreased from 73 to 64. The reported cases of Tuberculosis are also much decreased from 543 to 200 per 1000 population. Now the cluster-wise descriptive statistics for all the indicators used for clustering the districts are discussed.

The cluster-wise descriptive statistics for the literacy rate for both MICS data are presented in

Indicator	**MICS 2003-04**							**MICS 2007-08**						
	Min	Max	Mean	Median	Mode	Std. Dev.	Coeff. of variation	Min	Max	Mean	Median	Mode	Std. Dev.	Coeff. of variation
Literacy rate 10+ years	34	78	51.26	49.5	37	11.96	23.33	33	81	57.74	56	45	11.32	19.61
Infant Mortality Rate	55	97	79.62	83.5	86	12.66	15.90	40	110	76.86	78	78	17.93	23.33
Care provided by Lady Health Worker	8	89	39.91	38	23	18.93	47.42	18	83.1	56.66	60.5	31	18.32	32.33
Skilled attendant at delivery	7	65	31.65	31	22	12.98	41.01	12	68	41.06	39	35	14.52	35.36
Use of contraceptives	12	61	32.94	33	35	11.55	35.08	13	50	30.03	29	23	9.52	31.71
Reported tuberculosis	0.22	1.024	0.47	0.446	0.281	0.19	40.06	0.1	0.7	0.33	0.3	0.40	0.14	41.64
Physical access to drinking water	77	100	91.21	92	87	7.42	8.13	76	100	092.40	94	99	6.54	7.08

Use of sanitary means of excreta disposal	29	93	53.35	52.5	36	15.69	29.40	32	95	66.49	66	54	14.59	21.95

Table 2. The average literacy rate for cluster 1 in 2003-04 is 71.67 and 71.86 in 2007-08. Similarly for all other cluster there is no major change/increase in the average literacy. Overall cluster 1 and cluster 2 have areas with high education level as compared to others in both MICS data.

Table 1:
Descriptive Statistics of the Regional Indicators

Indicator	**MICS 2003-04**							**MICS 2007-08**						
	Min	Max	Mean	Median	Mode	Std. Dev.	Coeff. of variation	Min	Max	Mean	Median	Mode	Std. Dev.	Coeff. of variation
Literacy rate 10+ years	34	78	51.26	49.5	37	11.96	23.33	33	81	57.74	56	45	11.32	19.61
Infant Mortality Rate	55	97	79.62	83.5	86	12.66	15.90	40	110	76.86	78	78	17.93	23.33
Care provided by Lady Health Worker	8	89	39.91	38	23	18.93	47.42	18	83.1	56.66	60.5	31	18.32	32.33
Skilled attendant at delivery	7	65	31.65	31	22	12.98	41.01	12	68	41.06	39	35	14.52	35.36
Use of contraceptives	12	61	32.94	33	35	11.55	35.08	13	50	30.03	29	23	9.52	31.71
Reported tuberculosis	0.22	1.024	0.47	0.446	0.281	0.19	40.06	0.1	0.7	0.33	0.3	0.40	0.14	41.64
Physical access to drinking water	77	100	91.21	92	87	7.42	8.13	76	100	092.40	94	99	6.54	7.08
Use of sanitary means of excreta disposal	29	93	53.35	52.5	36	15.69	29.40	32	95	66.49	66	54	14.59	21.95

Table 2:
Cluster-wise Summary Statistics for Literacy Rate

Literacy rate (10+ years)	**MICS 2003-04**					**MICS 2007-08**				
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
Min	68	37	50	43	34	63	55	50	45	33
Max	78	74	60	58	47	81	74	69	53	58
Mean	**71.67**	**61.5**	**56.14**	**50.63**	**40.33**	**71.86**	**64.50**	**57.22**	**49.00**	**47.09**
Median	69	67.5	57	51.5	41	73	63	56	49	48
Std. Dev.	5.51	16.74	3.85	6.16	4.16	6.74	7.48	6.24	5.66	6.27
C.V	7.68	27.22	6.85	12.17	10.32	9.38	11.59	10.91	11.54	13.31

Table 3 shows the cluster wise summary of child mortality. In 2003-04, the minimum value of infant mortality rate in cluster 1 is 40 whereas maximum value is 70 which is 1:2. The minimum and maximum values in cluster 5 are 75 and 110. The values of means and medians are almost similar and monotonically increasing from cluster 1 to cluster 5. Although cluster 1 contains a district Rawalpindi which achieved the MDG (i.e. 40) but this cluster cannot be regarded as the developed regarding infant mortality rate as the value of CV is maximum in this cluster (i.e. 19.50). The average rate of infant mortality is least in cluster 1. It is because the average literacy rate for this cluster was high and more educated people will tend to take more care of their children as compared to less educated people. Cluster 5 contains the districts with very high values of infant mortality rate.

Care provided by lady health workers is significantly increased in cluster 3 and cluster 5 only (Table 4). The number of cases of Tuberculosis is also decreased in cluster 1, 2, 3 and 5 but this decrease is not so large to attain MDG. Summary statistics for reproductive health are given in

Table **5**. Cluster 1 & 2 have larger average value of skilled attendants at delivery in both surveys. One of the reasons is that these two clusters have more educated people. There is a small increase in the cluster-wise average values. Similarly the use of contraceptive has a little increase from 2003-04 to 2007-08. The cluster-wise summary for the two indicators of environment i.e., water supply and sanitation are presented in

Table 6. Water supply is better as compared to other indicators but sanitation has again low statistics. Some improvement regarding water supply and sanitation is made during these two surveys.

REFERENCES

1. Aragon, Y., Haughton, D., Haughton, J., Leconte, E., Malin, E., Ruiz-Gaen, A. and Thomas-Agnan, C. (2003). Explaining the Pattern of Regional Unemployment: The Case of Midi-Pyrenees Region. *Papers in Regional Science*, 82, 155-174.
2. Cziráky, D., Sambt, J., Rovan, J. and Puljiz, J. (2005). Regional development assessment: A structural equation approach. *Euro. J. Oper. Res.*, 174(1), 427-442.
3. González, J.I. and Morini, S. (2000). Posicionamiento socioeconómico y empresarial de los municipios de la Isla de Tenerife. (Socioeconomic and business affairs situation in Tenerife Island). *Working Paper 2000-07*. Universidad de La Laguna (in Spanish).
4. Peschel, K. (1998). Perspectives of Regional Development around the Baltic Sea. *The Annals of Regional Science*, 32, 299-320.
5. Pettersson, O. (2001). Microregional Fragmentation in a Swedish County. *Papers in Regional Science*, 80, 389-409.
6. Rovan, J. and Sambt, J. (2003). *Socio-economic Differences among Slovenian Municipalities: A Cluster Analysis Approach, in Development in Applied Statistics*, Ferligoj. A. and Mrvar A., eds., 265-278.
7. Soares, J.O., Marquês, M.L. and Monteiro C.F. (2003). A Multivariate Methodology to Uncover Regional Disparities: A Contribution to Improve European Union and Governmental Decisions. *Euro. J. Oper. Res.*, 145, 121-135.

Table 1:
Descriptive Statistics of the Regional Indicators

Indicator	MICS 2003-04							MICS 2007-08						
	Min	Max	Mean	Median	Mode	Std. Dev.	Coeff. of variation	Min	Max	Mean	Median	Mode	Std. Dev.	Coeff. of variation
Literacy rate 10+ years	34	78	51.26	49.5	37	11.96	23.33	33	81	57.74	56	45	11.32	19.61
Infant Mortality Rate	55	97	79.62	83.5	86	12.66	15.90	40	110	76.86	78	78	17.93	23.33
Care provided by Lady Health Worker	8	89	39.91	38	23	18.93	47.42	18	83.1	56.66	60.5	31	18.32	32.33
Skilled attendant at delivery	7	65	31.65	31	22	12.98	41.01	12	68	41.06	39	35	14.52	35.36
Use of contraceptives	12	61	32.94	33	35	11.55	35.08	13	50	30.03	29	23	9.52	31.71
Reported tuberculosis	0.22	1.024	0.47	0.446	0.281	0.19	40.06	0.1	0.7	0.33	0.3	0.40	0.14	41.64
Physical access to drinking water	77	100	91.21	92	87	7.42	8.13	76	100	092.40	94	99	6.54	7.08
Use of sanitary means of excreta disposal	29	93	53.35	52.5	36	15.69	29.40	32	95	66.49	66	54	14.59	21.95

Table 2:
Cluster-wise Summary Statistics for Literacy Rate

Literacy rate (10+ years)	MICS 2003-04					MICS 2007-08				
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
Min	68	37	50	43	34	63	55	50	45	33
Max	78	74	60	58	47	81	74	69	53	58
Mean	**71.67**	**61.5**	**56.14**	**50.63**	**40.33**	**71.86**	**64.50**	**57.22**	**49.00**	**47.09**
Median	69	67.5	57	51.5	41	73	63	56	49	48
Std. Dev.	5.51	16.74	3.85	6.16	4.16	6.74	7.48	6.24	5.66	6.27
C.V	7.68	27.22	6.85	12.17	10.32	9.38	11.59	10.91	11.54	13.31

Table 3:
Cluster-wise Summary statistics for Child mortality (Number/Thousand)

Infant Mortality Rate	MICS 2003-04					MICS 2007-08				
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Min	55	55	69	72	86	40	53	54	72	75
Max	66	61	86	89	97	70	81	92	86	110
Mean	**60.00**	**58.5**	**78.43**	**81.38**	**91.08**	**54.57**	**71.00**	**78.22**	**79.00**	**92.73**
Median	59	59	80	83.5	90.5	52	73	82	79	88
Std. Dev.	5.57	3.00	5.80	6.32	3.90	10.64	10.20	11.52	9.90	14.47
C.V	9.28	5.13	7.39	7.77	4.28	19.50	14.36	14.73	12.53	15.61

Table 4:
Cluster-wise Summary statistics for Adult health

Indictor	Statistic	MICS 2003-04					MICS 2007-08				
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Care provided by Lady Health Worker (LHW)	Min	36	18	11	41	8	44	18	57	65	31
	Max	73	48	55	89	60	82	42	83	77	73
	Mean	**55.00**	**32.5**	**30.29**	**59.38**	**31.25**	**67.00**	**28.83**	**68.67**	**71.00**	**52.91**
	Median	56	32	27	55	28.50	70	26.5	66	71	59
	Std. Dev.	18.52	14.29	16.79	15.12	12.61	12.75	9.72	8.25	8.49	14.31
	C.V	33.67	43.98	55.44	25.46	40.36	19.04	33.73	12.01	11.95	27.05
Reported tuberculosis	Min	0.22	0.277	0.223	0.45	0.233	0.1	0.3	0.1	0.6	0.3
	Max	0.54	0.331	0.518	1.024	0.721	0.3	0.5	0.4	0.7	0.5
	Mean	**0.39**	**0.30**	**0.36**	**0.65**	**0.48**	**0.21**	**0.40**	**0.24**	**0.65**	**0.39**
	Median	0.396	0.3045	0.342	0.5705	0.51	0.2	0.4	0.2	0.65	0.4
	Std. Dev.	0.16	0.03	0.09	0.22	0.15	0.07	0.06	0.11	0.07	0.08
	C.V	41.59	9.61	25.92	34.09	30.28	32.20	15.81	46.24	10.88	21.26

Table 5:
Cluster-wise Summary Statistics for Reproductive Health

Indictor	Statistic	MICS 2003-04					MICS 2007-08				
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
Skilled attendant at delivery	Min	44	43	28	22	7	44	44	31	20	12
	Max	65	60	42	38	26	68	66	50	26	42
	Mean	**51.67**	**48.75**	**36.29**	**30.13**	**19.25**	**58.00**	**53.83**	**38.33**	**23.00**	**28.82**
	Median	46	46	38	31	21	59	54.5	36	23	29
	Std. Dev.	11.59	7.68	5.12	6.29	5.17	9.43	8.42	6.24	4.24	7.69
	C.V	22.43	15.75	14.12	20.88	26.87	16.27	15.65	16.29	18.45	26.69
Use of contraceptives (any method)	Min	35	35	16	24	12	29	33	20	23	13
	Max	37	50	45	61	48	50	45	41	27	38
	Mean	**35.67**	**44.5**	**27.71**	**37.63**	**28.33**	**38.43**	**39.50**	**28.44**	**25.00**	**21.73**
	Median	35	46.5	27	35	28.5	37	40	26	25	21
	Std. Dev.	1.15	6.66	9.96	11.51	11.94	6.27	4.76	7.20	2.83	6.69
	C.V	3.24	14.96	35.94	30.60	42.15	16.31	12.06	25.30	11.31	30.81

Table 6:
Cluster-wise Summary statistics for water and sanitation

Indictor	Statistic	MICS 2003-04					MICS 2007-08				
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Physical access to drinking water (within dwelling)	Min	79	95	77	90	77	76	84	92	99	82
	Max	87	100	99	100	99	99	99	99	100	97
	Mean	**82.33**	**98.5**	**88.29**	**95.38**	**89.92**	**88.00**	**93.33**	**97.33**	**99.50**	**89.36**
	Median	81	99.5	91	96	88.5	87	96.5	98	99.5	91
	Std. Dev.	4.16	2.38	9.16	3.54	6.78	7.64	6.62	2.24	0.71	5.33
	C.V	5.06	2.42	10.38	3.71	7.53	8.68	7.10	2.30	0.71	5.97
Use of sanitary means of excreta disposal	Min	57	63	39	46	29	73	67	61	42	32
	Max	70	93	75	67	54	89	95	69	54	65
	Mean	**62.67**	**77.75**	**58.86**	**55.25**	**38.42**	**79.71**	**82.00**	**65.89**	**48.00**	**53.45**
	Median	61	77.5	61	54.5	37.5	80	83	66	48	55
	Std. Dev.	6.66	13.35	12.35	7.69	7.45	5.85	12.12	2.47	8.49	9.77
	C.V	10.62	17.17	20.98	13.91	19.40	7.34	14.78	3.75	17.68	18.28

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 207-212

PROJECT MENTOR

Imran Anwar Ujan[1] and **Imdad Ali Ismaili**[2]
Institute of Information & Communication Technology
University of Sindh, Jamshoro, Pakistan
Email: [1]iujan@yahoo.com
[2]iai_a@yahoo.com

ABSTRACT

The project Mentor is an idea that will provide an online platform for the management of software projects with the help of existing projects and ideas of various developers. This Online System will create an interactive environment among the developers. Each registered developer will be given a particular task with project title and description along with a deadline. This Online System will also provide a productive mechanism of communication among the developers that are in academies and industries so that the management of the projects can be done in an efficient and easy way.

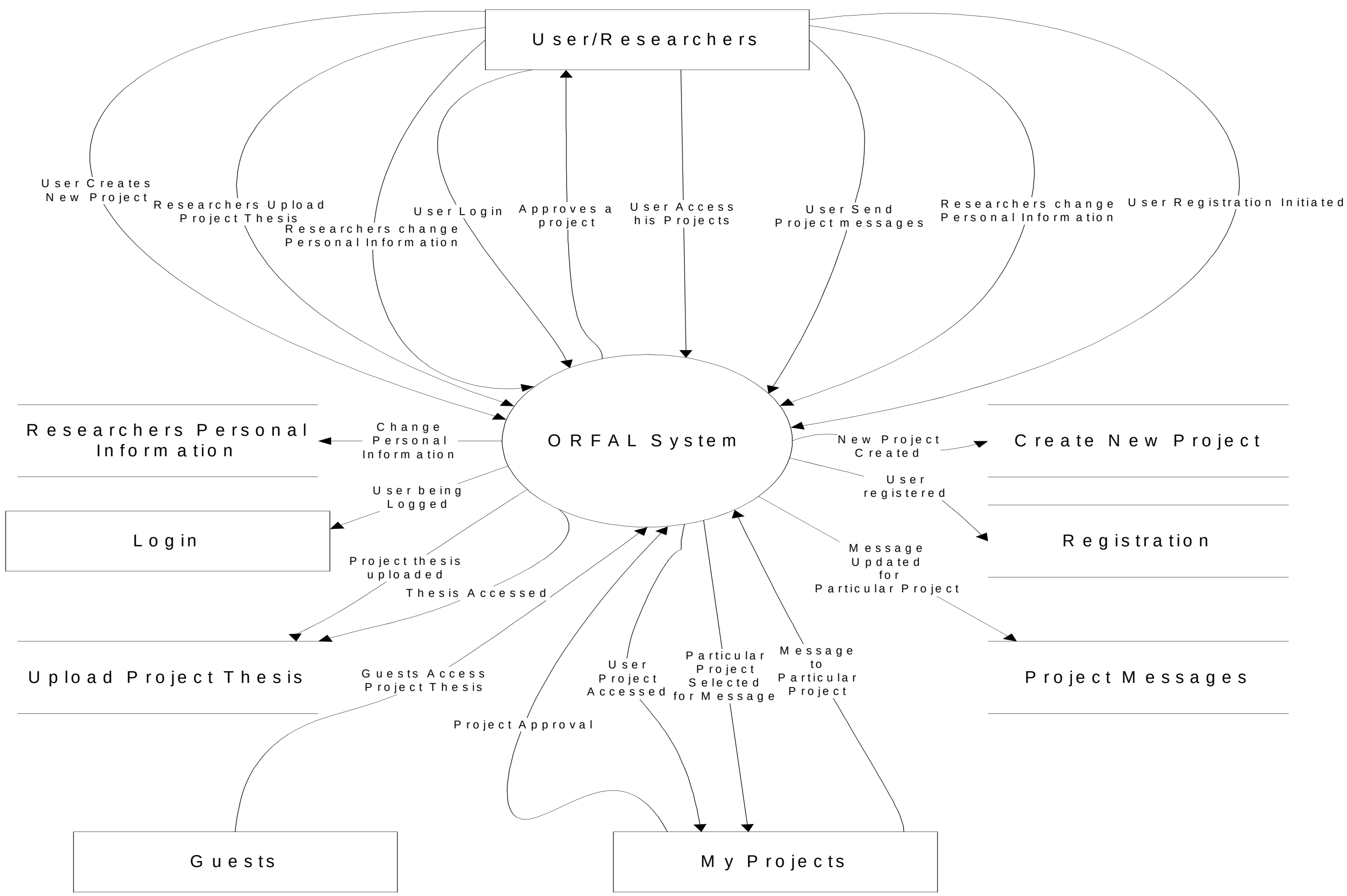
1. INTRODUCTION

1.1 Background

The project mentor is a web based application /Platform for the developers to work on different categories and sized project via internet to allow the maintenance, modification and up gradation of the online projects via an interactive environment.

With the advent of web based applications many web technologies and programming languages emerged from CGI to PHP and set of standards were being set in order to develop a web based application in which we have Model View Controller pattern to develop a standard web application in any of the programming language that supports construction of web based applications.

This web application connects the different users from different part of the world so when ORFAL is online it means that researchers from all institutions can collaborate their work through the ORFAL platform and easily they can search the library of projects for which until now no application is available this functionality is also available in Online Research Frame work and Library along with the Research Framework

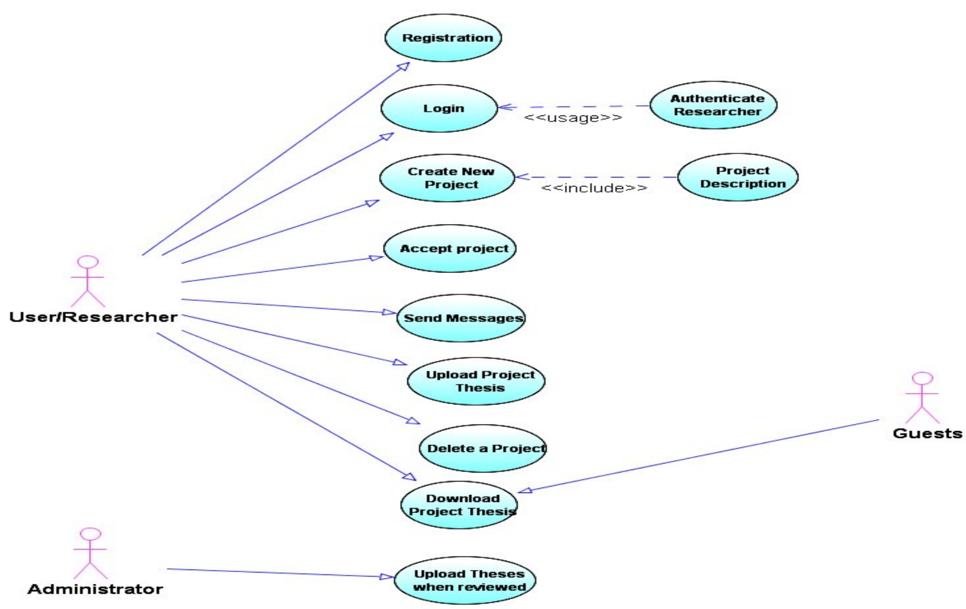

User/Researchers
User Creates New Project
Researchers Upload Project Thesis
Researchers change Personal Information
User Login
Approves a project
User Access his Projects
User Send Project messages
Researchers change Personal Information
User Registration Initiated
ORFAL System
Researchers Personal Information
Change Personal Information
Create New Project
New Project Created
User registered
Registration
User being Logged
Login
Project thesis uploaded
Thesis Accessed
Message Updated for Particular Project
Upload Project Thesis
Guests Access Project Thesis
User Project Accessed
Particular Project Selected for Message
Message to Particular Project
Project Messages
Project Approval
Guests
My Projects

1.2 Functional Requirements:

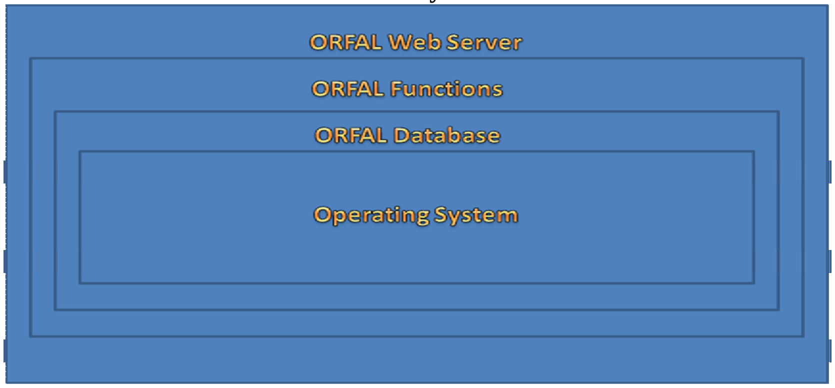
The Functional Requirements of the System are:

1) Initially the researcher will be able to Login into the system by entering his username and password
2) If the researcher is not registered he first have to go through the registration process then he can logs into the system
3) Projects will be visible to the researcher when he logs in
4) There would be a My Projects section for the researchers to see his own projects and there would be a My Accepted projects section that he accepts from other researcher in which he is participating.
5) There would be a functionality to create new projects along with the description which can be accepted by other researchers as well as researcher can delete his project.
6) A messaging system would be provided something like blog system in order to have communication between researchers for various ideas and to accomplish tasks.
7) Registered Researchers would have the functionality to upload the project thesis.
8) Functionality should be provided to download the thesis.
9) Researchers can be able to Logout from the system.
10) Users which are not registered can also download the project thesis from the default page of the application.
11) A mechanism should be provided in order to maintain the overall system through Administrator.
12) Thesis would be uploaded when reviewed by the Admin Staff.

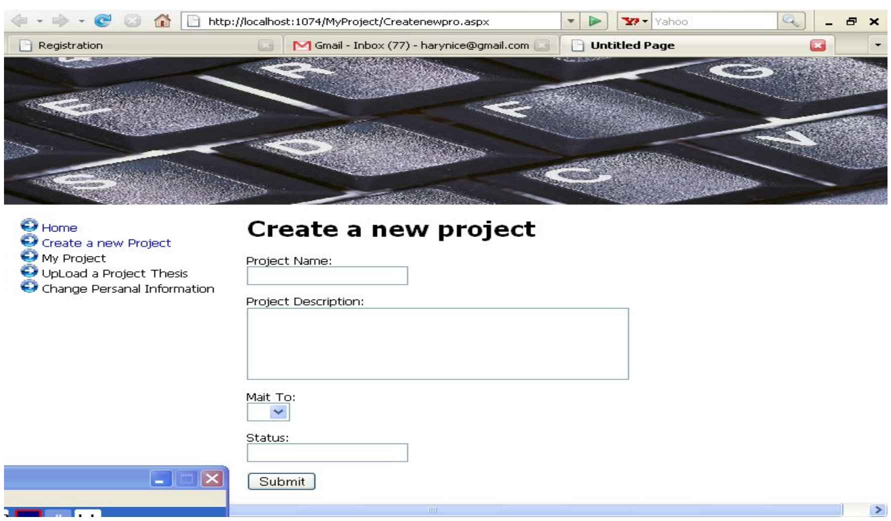
The functional requirements are implemented in use case model in Fig



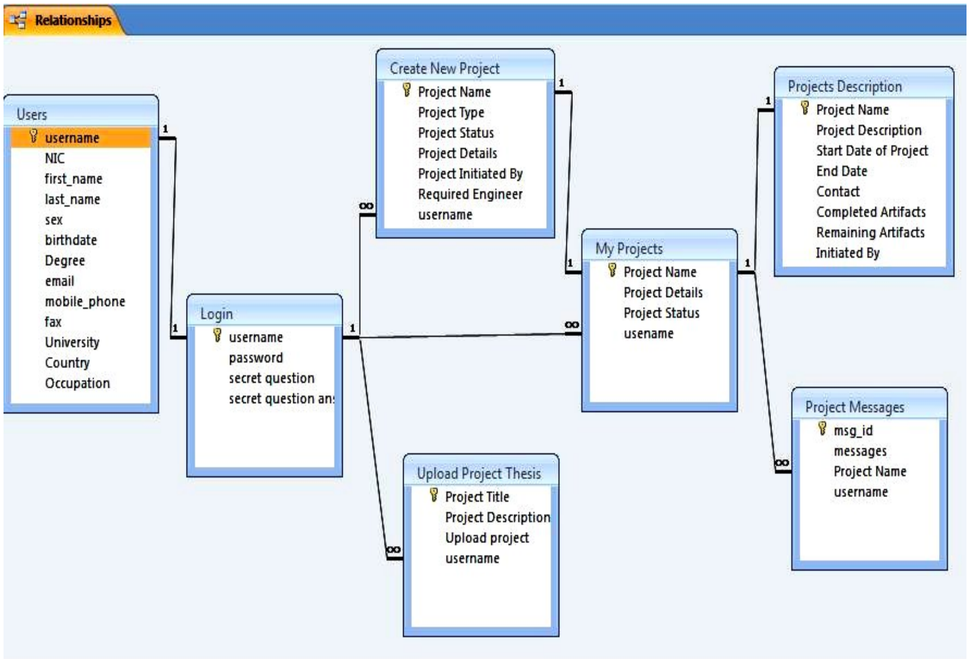
High –Level Software Architecture of the System:



User INTERFACE DESIGN:



Database DESIGN:



CONCLUSION

This Project is concerned with the providing the facility to the researchers to do their project work online for this a researcher is registered and then he starts the project and throw it and the other researchers accepts it and a mutual collaboration is started between different researchers to accomplish their project of mutual interest.

FUTURE WORK

There can be a lot of future work into the website a search facility can be provided to search the library of projects, more up gradation to be done in performance and efficiency as well as feedback should be enhanced along with the messaging system and a lot more future work can be decided.

REFERENCES

Books:

1. Avison, D.E. and Fitzgerald G. (1995). *Information Systems Development: Methodologies, Techniques and Tools*. 2nd Edition, McGraw-Hill.
2. Constantine and Lockwood (1999). *Software for Use: a practical guide to the models and methods of user-centered design*. ACM Press, Addison-Wesley.
3. Nielsen, J. (1993). *Usability Engineering*. AP Professional.
4. Pressman, R. (1997). *Software Engineering - A Practitioners Approach*. McGraw Hill.
5. Schneiderman, B. (1997). *Designing the User Interface*. Addison-Wesley.

6. Sommerville, I. (1997). *Software Engineering*. Addison-Wesley.
7. The British Computer Society (1995). *A Glossary of Computing Terms*. 9th Edition, Fred Brooks.
8. Brooks, F.P. (1995). The Mythical Man Month (Anniversary Edition with four new chapters) Ch. 16 (No Silver Bullet - Essence and Accident), Addison-Wesley.

Websites:

Wikipedia:

[1] http://en.wikipedia.org/wiki/Web_Engineering
[2] http://en.wikipedia.org/wiki/Web_Application
[3] http://en.wikipedia.org/wiki/File:Overview_of_a_three-tier_application.png
[4] http://en.wikipedia.org/wiki/Model-view-controller
[5] http://en.wikipedia.org/wiki/ASP.NET_MVC_Framework
[6] http://en.wikipedia.org/wiki/Three-tier_(computing)
[7] http://en.wikipedia.org/wiki/Multilayered_architecture
[8] http://en.wikipedia.org/wiki/ASP.NET

Web developers Journal:

[9] http://www.webdevelopersjournal.com/articles/development_team.html

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 213-222

VISUALIZING MULTIVARIATE DATA WITH ANDREWS' CURVES

Mudassir Uddin, Mehwish Hussain, Almas Izhar Fatmi
Department of Statistics, University of Karachi, Karachi, Pakistan
Email: mehvish.hussain@yahoo.com

ABSTRACT

The development of Multivariate analysis from the 1930s until recent years have concentrated on statistical methods primarily based on exploratory data analysis. Many methods are easily available in different statistical packages (Splus, SAS, SPSS, Minitab, etc) that produce the results in the form of a graph or some other type of visual display. These packages often allow the investigator to get a 'view' of the data that might have an impact impossible or at least difficult to achieve by means of examination of statistics derived from the data. With data sets having more than two dimensions some kind of transformation is always necessary. In this paper, it is reviewed in depth and we present different variants of Andrews' curves. In Andrews' curves we use a curve to represent each multidimensional data point. The aim of this paper is to present a new method which detects the out-of-control variables in an observed multivariate control chart signals. The proposed method is investigated which gives many interesting results. Furthermore, several modified Andrews' curves are accessible. These different variants of Andrews' curves, made by different statisticians later on, are compared using some real data sets and some conclusions are presented.

1. INTRODUCTION

Visualization techniques are always recommended as preliminary data exploration techniques. A strong desire of all data analysts is to have the ability to visualize data. Often this means taking a low dimensional projection of a data set and looking in turn at a variety of one, two or, at most, three dimensional projections of the data. An alternative is to transform the data in some way in order to make the data's properties visible via the transformation. Therefore the transformation has to maintain some inherent properties of the data if we are to be able to identify some inherent characteristics of the data after the transformation. In this paper we present a variation of Andrews' curves which performs that kind of transformation. We start with a review in depth of modification to Andrews' curves, then we focus on modification by Khattree and Naik (2002) and then on Wegman's curves, an extension of Andrews' curves that allows us to construct a two dimensional grand tour.

2. ANDREWS' CURVES

If statistical data are n-dimensional, then each set of n measurements can be presented as an n dimensional point. In order to plot high dimensional data in two dimensions, a graphical technique was developed by Andrews (1972). The method is a way to visualize and hence to find structure in high dimensional data. Each data point x = (x_1, x_2... x_n) defines a finite Fourier series as mentioned:

$$f_x(t) = \frac{x_1}{\sqrt{2}} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + ... \quad (1)$$

Andrews' curves can map each multi-response observation into a function, f(t), of a single variable, t. The function f(t) is defined as a linear combination of orthonormal functions in t, with the coefficients in the linear combination being the observed values of the responses and this function is then plotted for $-\pi < t < \pi$. This particular Fourier series contains the properties of mean, distance and variance preservation. As well, the curves are simultaneously showing all the projections onto that vector for the range $-\pi < t < \pi$ extracting one-dimensional projection and produces linear relationship for more than two dimensions. (For detail see Section 3).

These curves have been utilized in different fields such as biology, neurology, sociology and semiconductor manufacturing. Some of their uses include the quality control of products, the detection of period and outliers in time series or the visualization of learning in artificial neural networks. In this paper we applied the Andrews curves on the data sets obtained from agriculture, botany, demography and criminology.

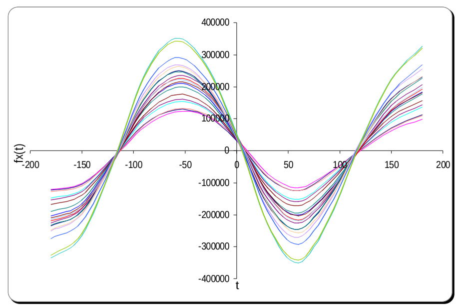
This technique has been useful for identifying external patterns such as outliers, clusters, etc., indicating that further investigation via some form of multidimensional scaling or cluster analysis might be fruitful. Nonetheless, a problem which arises when using this technique is that only a fairly limited number of observations may be plotted on the same diagram before it becomes too bemused to be accommodating. The Fourier plot suffers from a disadvantage that interchange of variables leads to a different picture (See Figure 1).

3. SOME PROPERTIES OF ANDREWS' CURVES

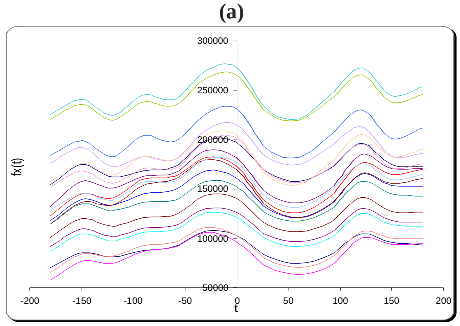
These curves have several useful properties (Cesar Garca-Osorio and Colin Fyfe, 2005), some of which are:

3.1 Mean preservation:

The function corresponding to the mean of a set of N-dimensional observations is the point-wise mean of the functions corresponding to the i^{th} observation vector. If $\mathrm{x}_i = (x_{1i}, \ldots, x_{mi})$ $(i = 1, \ldots, n)$ are n points in m-dimensional space, then if $\bar{x}$ stands for the mean vector, $f_{\bar{x}}(t) = \frac{1}{n}\sum_{i=1}^{n} f_{x_i}(t)$, So the curve representing the mean, look likes an *average* curve.



(a)



(b)

Figure 1:

(a) Andrews' plot for the crimes in Pakistan for 19 years with ten variables.

(b) A different picture for the same data while changing order of first two crimes.

3.2 Distance preservation:

The distance between two functions defined as $\left\| f_{x_j} - f_{x_l} \right\|_{L_2}^2 = \int_{-\pi}^{\pi} \{ f_{x_j}(t) - f_{x_l}(t) \}^2$ which is proportional to the Euclidean distance between the corresponding points, since $\left\| f_{x_j} - f_{x_l} \right\|_{L_2}^2 = \pi \sum_{k=1}^{m} \left(x_{kj} - x_{kl} \right)^2 = \left\| f_{x_j} - f_{x_l} \right\|_{L_2}^2 = \pi \left\| x - y \right\|^2$ Thus, the function preserves distance. The above function measure of distance between two functions is the L_2 norm so that points x and y that are close together lead to curves which are close together. Figure 3(a), demonstrates the points which are close together are apparent through curves forming clusters.

3.3 One-dimensional projections:

For a particular value of $t = t_o$, the function value $f_x(t_0)$ is proportional to the length of the projection of the vector $(x_1, x_2, x_3, \ldots, x_d)$ on the vector: $f_1(t_0) = \frac{1}{\sqrt{2}}, \left(\sin(t_0), \cos(t_0), \sin(2t_0), \cos(2t_0) \ldots . \right)$. This means that the curves are simultaneously showing all the projections onto that vector for the range $-\pi < t < \pi$, since $f_x(t_o) = X'a_o$. This projection onto a one-dimensional space may show up clustering or any data peculiarities that occur in this subspace and which may be otherwise obscured by other dimensions. The plot, therefore, provides a continuum of such one-dimensional projections all on the one graph.

3.4 Linear relationships:

If a point **y** lies on a line joining **x** and **z**, then for all values of t, $f_y(t)$ is between $f_x(t)$ and $f_z(t)$.

3.5 Variance preservation:

If the components of the data are uncorrelated with common variance σ^2, the Andrews' curves representations preserve that variance.

$$\operatorname{var}\{ f_x(t) \} = \sigma^2 (2^{-1} + \sin^2 t + \cos^2 t + \sin^2 2t + \cos^2 2t + \ldots)$$

$$= \begin{cases} 2^{-1} \sigma^2 m, \text{ if } m \text{ is odd} \\ 2^{-1} \sigma^2 \left\{ m - 1 + 2 \sin^2 \left(\frac{mt}{2} \right) \right\}, \text{ if } m \text{ is even} \end{cases}$$

where $2^{-1} \sigma^2 (m-1) \leq \operatorname{var}\{ f_x(t) \} \leq 2^{-1} \sigma^2 (m+1)$, $(-\pi < t < \pi)$

4. MODIFICATIONS TO ANDREWS' CURVES

To overcome the intricacies abide by the Andrews curves; several modification were proposed (see for example, Embrechts and Herzberg, 1991; Wegman and Shen, 1993 and Khattree and Naik, 2002).

4.1 Andrews' (1972):

The modified fastidious Fourier series by Andrews himself is articulated as

$$f_x(t) = x_1 \sin(n_1 t) + x_2 \cos(n_1 t) + x_3 \sin(n_2 t) + x_4 \cos(n_2 t) + \ldots \quad (2)$$

The restriction to integers in the progression is because of the distance preserving property; without integers, this property is lost. Andrews compared the curve with values $n_1 = 2,\ n_2 = 4,\ n_3 = 8, \ldots,$ with the original formulation and concluded that the former is more space filling but more difficult to interpret when it is used for visual inspection.

4.2 Khattree and Naik (2002):

Khattree and Naik (2002) attempt to modify the formula given by Andrews as follows:

$$f_x(t) = \frac{1}{\sqrt{2}}\{x_1 + x_2(\sin(t) + \cos(t)) + x_3(\sin(t) - \cos(t)) + x_4(\sin(2t)$$
$$+ \cos(2t)) + x_5(\sin(2t) - \cos(2t)) + \ldots.. \quad (3)$$

where, every x_i is exposed in terms of either a sine or cosine function.One of the advantages of this modification is that the trigonometric terms do not simultaneously vanish at any given t and avoids some of the shortcomings of Andrews' plots. The major disadvantage of this modification is that the odd numbered terms in the Andrews' plot given in equation simultaneously vanish at $t=0$; yet, for Andrews' plot around $t=0$, these similarities or dissimilarities are mostly due to even numbered variables. This is not so, for the modified Andrews' plots given by equation of original version of Andrews' curves.

4.3 Wegman and Shen (1993):

Wegman and Shen (1993) discussed the benefits of using a slightly different projection. They were concerned with the connection between Andrews' curves and the grand tour. They show that Andrews' curves are not a real one-dimensional grand tour. The problem is that Andrews' curves do not exhaust all possible orientations of a one-dimensional vector. Their generalization of Andrews' curves is more space filling, although it has lost the distance preservation property, and can be used to obtain a bi-dimensional *pseudo* grand tour. Clearly (w_1, w_2) form a set of 2 orthonormal basis vectors. If we define:

$$y_1 = w^T x \propto x_1 \sin(\lambda_1 t) + x_2 \cos(\lambda_1 t) + \ldots.. + x_d \cos(\lambda_{d/2} t) \quad (4)$$

$$y_2 = w^T x \propto x_1 \cos(\lambda_1 t) - x_2 \sin(\lambda_1 t) + \ldots.. - x_d \sin(\lambda_{d/2} t) \quad (5)$$

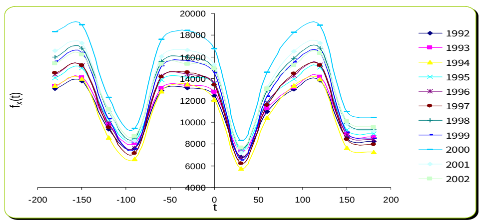
then we have a two dimensional display on which to project x so that we can look for structure by eye. From visual display of projections, we can identify clusters of points which are nearby and whose trajectories as we change t (i.e. as we move along the Andrews' curves) keep close together. When we use these curves in this way, we obtain a two dimensional "grand tour" of the data.

5. APPLICATIONS IN REAL LIFE PROBLEMS

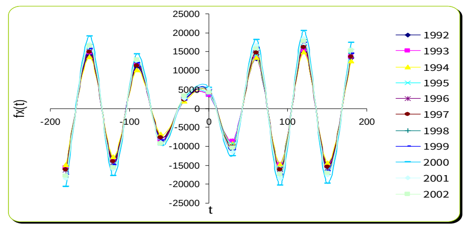
Real life data is always multivariate. The analysis of such data becomes quite complicated if the number of variables becomes more than two or three. Andrews, as described in the preceding sections, made endeavor to visualize the data in more interpretable mode. Few fascinating patterns will be displayed in the following case studies.

5.1 In Agricultural Research:

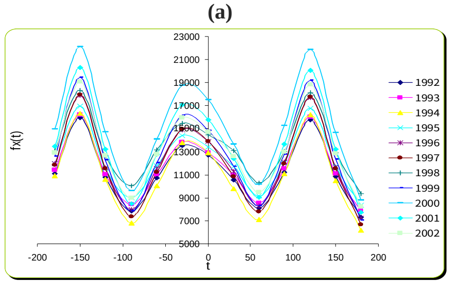
The data is taken from the Economic Review of Pakistan, 2002. It consists of production of major crops of Pakistan for the time period of 1992-2002. The curves have been plotted for eleven years data and the six agriculture crops of Pakistan (considered as variables), which are wheat, rice, bajra, jawar, maize and barley and are coded as x_1, x_2, x_3, x_4, x_5 and x_6 respectively. The different types of Andrews's curves are shown in Figure 2. Figure 2(a) depicts that there was minimum production in the year 1994. Moreover, there is gradual increase with nearly similar variations with the enhancement of time which has been exposed through similar curves whereas in the year 2000, the production of these crops is rather at a very large amount. Nonetheless, in the years 2001 and 2002, these crops were produced in less quantity as in the year 2000. The abrupt decision can be made by Figure 2(b) as it mislays the distance preserving property so detailed scrutinized cannot be feasible. The modifications made by Khattree and Naik (2002) is caused due to the smoothing in the trigonometric wave in Figure 2(c) and also it seems to be dissimilar to Figure 2(a) as losing distance preserving property for the revelation of particular data. The Wegman's algorithm seeks to develop reduction in the distance between each data sample, as shown in Figures 2(d, e) which demonstrate a clear insight of the data configuration.



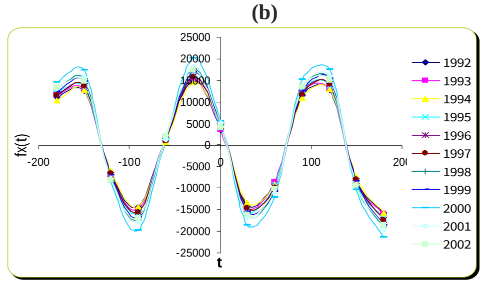
(a)



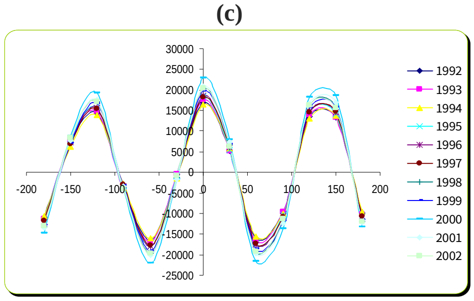
(b)



(c)



(d)



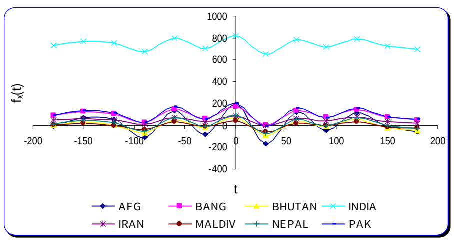
(e)

Figure 2:
(a) Andrews' plot of agricultural data
(b) The plot by implementation of equation 2
(c) The Khattree and Naik's modified plot
(d) & (e) The grand tour by Wegman's perspective.

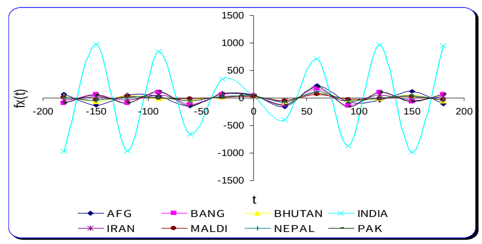
While comparing these variations, the visualization of Figure 2 extracts that the Andrews' original formula is seemed to be better to utilize than its other later described modified forms for the example of production of crops in Pakistan.

5.2 In Demographic Data:

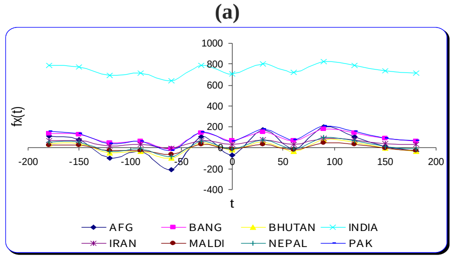
The demographic data consists of the social indicators for the countries of South Central Asia. The data has been taken from ESCAP population data sheet, 2003. The five social indicators chosen as variables are: x_1 : Mid-year population (per million), x_4 : Fertility Rate (per women), x_2 : Crude Birth Rate (per thousand), x_5 : Infant Mortality Rate (per thousand). x_3 : Crude Death Rate (per thousand).The main interest in this study is to visualize the affect of vital indicator for the population in Afghanistan, Bangladesh, Bhutan, India, Iran, Maldives, Nepal and Pakistan. Figure 3 gives a clear indication of the presence of an outlier (that is observations which emerge to be conflicting with the rest of the data's observations). In this prototype the outlier, colored in light blue, is a genuine observation demonstrative of an unusual observation. The outlier points towards the large population size of India and thus Andrews' curve has clearly detected an observation different from the rest of the data (a detailed account of the detection of outliers is given in Barnett and Lewis, 1978).



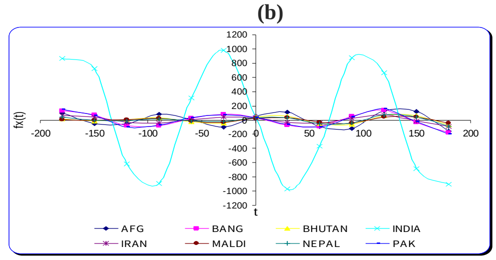
(a)



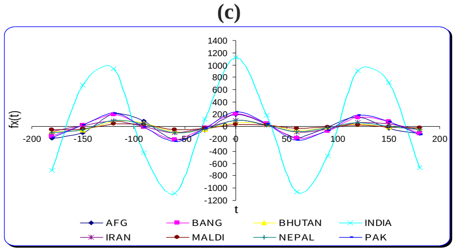
(b)



(c)



(d)



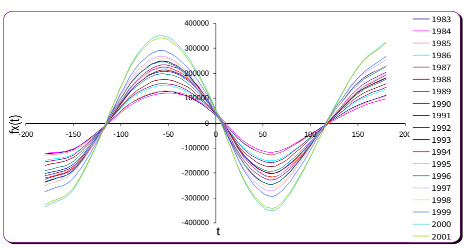
(e)

Figure 3:
(a) Andrews' plot of demographic data
(b) The plot by implementation of equation 2
(c) The Khattree and Naik's modified plot
(d) and (e) The grand tour by Wegman's perspective.

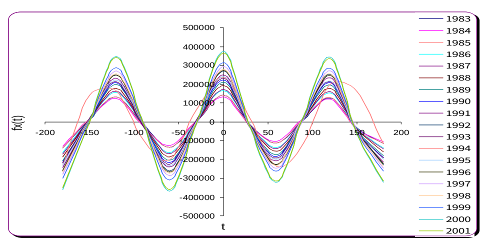
On the other hand the curves which are close together specified the similar points in the demographic data set. The curves of Bangladesh and Pakistan are almost same, demonstrating similarities between the observations of these two countries. Figure 3(b) is the visualization display after the application of formula expressed in **(2)**. The new revelation is less space filling. Figure 3(c) seems to be mirror image outlook of the Figure 3(a) indicating the same visual exposure by original version of Andrews' formula expressed and by the expression formulated by Khattree and Naik (2002). The Wegman's perspective does not seem as much differ in case of interpretation than the modified formula of Andrews' as displayed in Figure 3(b). Thus, the Andrews' original version of **(1)** seems to be better to explore this instance for visualization. Though, the modifications by Khattree and Naik (2002) can also be helpful for sake of interpretation.

5.3 In Criminology Study:

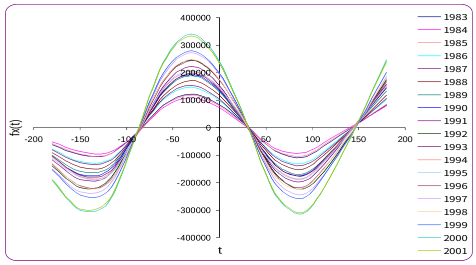
The third data set is pertained to yearly number of crimes in Pakistan. The data is taken from "*Pakistan Statistical Year Book, 2003*". It was accessible for 19 years from 1983 to 2001 on 10 variables that are murder, attempted murder, kidnapping/ abduction, child lifting, dacoity, robbery, burglary, cattle theft, other theft and others. The Andrews' curve and its modified forms for these 10 variables are displayed in Figure 4. Since the selected crimes data consists of 10 variables and 19 observations, the plot is difficult to study in detail as Andrews' plot becomes cluttered for greater number of observations. Well, an intense scrutinized of Figure 4(a) emerges the verdict that intensity of these crimes is higher in the year 2000 and 2001 which are plotted in aqua and olive green colors respectively. The space filling property by modified Andrews' formula plotted in Figure 4(b) exhibits an evident look of the unusual surveillances not only for the year 2000 and 2001 but also in the year 1994. The illustration in Figure 4(c) also establishes the same graphical view for crimes committed data as signified in Figure 4(a) just the difference is that the node shifts right forth. Here the trigonometric series approaches to zero at 90^{o}C at negative angle side and 150^{o}C at positive angle side whereas using original Andrews' formula the node looms to zero at 120^{o}C both sides. The Wegman's perspective does not stipulate clear clue for scrutinizing of crimes data due to the verity that crime's data consists of large sample observation with 10 variables. Thus, it is concluded that the modified formula of Andrews' curve **(2)** seems to visualize the behavior better than the other variants of the Andrews Curves. Nevertheless, the exploration using original formula of Andrews' and modified formula by Khattree and Naik (2002) cannot be disregarded.



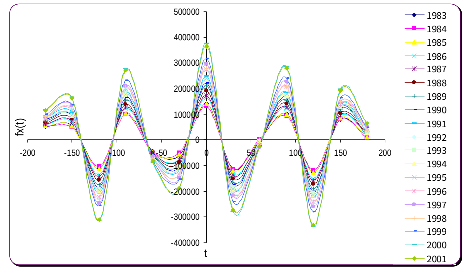
(a)



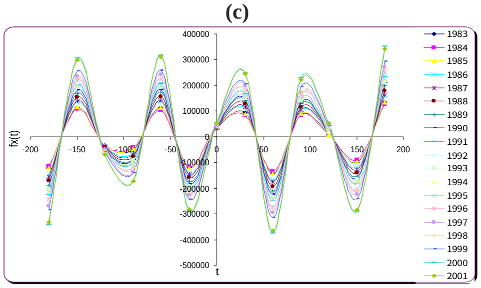
(b)



(c)



(d)

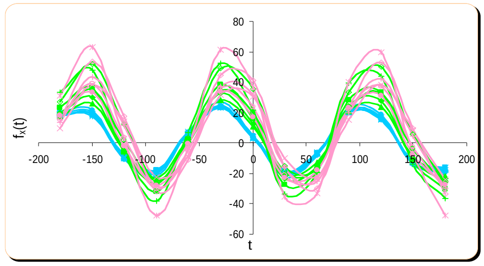


(e)

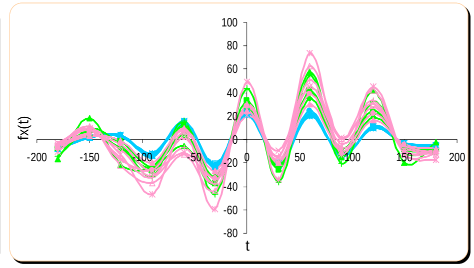
Figure 4:
(a) Andrews' plot of crimes data
(b) The plot by implementation of equation 2
(c) The Khattree and Naik's modified plot
(d) and (e) The grand tour by Wegman's perspective

5.4 In Botanical Research:

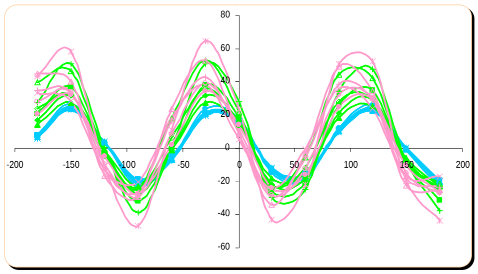
The data in this case study belongs to the laboratory experiment which investigate the effect on growth of paspalum grass at three different temperatures ($14\,^oC$, $18\,^oC$, and $22\,^oC$). Six three-dimensional observations were made for each temperature and growth was checked (Seber, 1983). For each pot of paspalum measurements were made on the fresh weight of roots (gm), the maximum root length (mm) and the fresh weight of tops (gm). Figure 5 illustrates the Andrews' plots based on trigonometric functions for the grass data. For three different temperatures the plot exhibits well-separated clusters indicating the difference in growth due to the changes in temperature. In this case Andrews' plot has shown groupings of curves which are almost alike and thus giving a clear interpretation of the data. Such a result may lead to some form of multidimensional scaling or cluster analysis for further investigation. In all the graphs of Figure 5, the plots clearly separate growth at $14\,^oC$ from the other two temperatures. Various other characteristics show up. For instance, the curves are very tightly banded near t=-30 and t=90, so a dimension reduction from three to two is likely. A canonical variate analysis would confirm this. It is hard to separate $18\,^oC$ from $22\,^oC$ although at some intervals for t, separation is clear. An interactively constructed plot would reveal this pattern much more clearly. Also some potential outlying curves in the cloud of $22\,^oC$ can be noted.



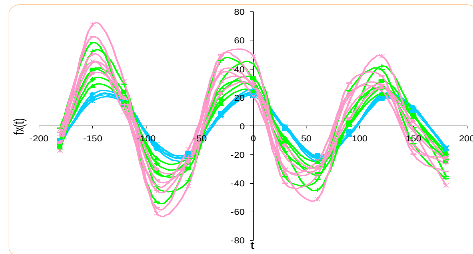
(a)



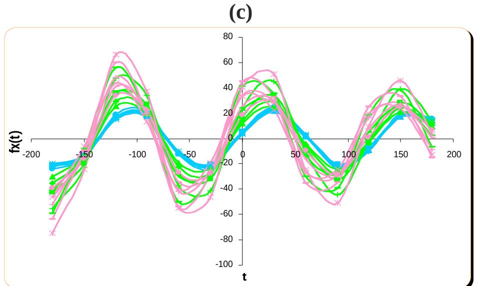
(b)



(c)



(d)



(e)

Figure 5:
(a) Andrews' plot of crimes data
(b) The plot by implementation of equation 2
(c) The Khattree and Naik's modified plot
(d) and (e) The grand tour by Wegman's perspective

The graphs of Figure 5(c) demonstrate that the modification suggested by Khattree and Naik is not as much differs from the original version of Andrews' curves in the case of visualization of Paspalum grass' growth data. The two-dimensional grand tour produces similar graphical display pattern in Figures 5(d) and 5(e). Thus, the two Wegman's grand tour assumed to be analogous on behalf of interpretation. The Andrews' curves in **(1)** and the modified Andrews' curves in **(3)** (Khattree and Naik, 2002) emerge to be better to explore the growth of Paspalum grass, whereas the Wegman's perspective can also generates significant outputs for investigating the data.

6. CONCLUSION

Andrews' plot is one of the familiar methods for exploratory data analysis. This paper elucidates this visual exploration technique with its theoretical framework pertinent to real life problems and compares it with different variants (see Wegman and Shen, 1993 and Khattree and Naik, 2002). Gnanadesikan (1977), and Everitt and Dunn (1990) affirm that Andrews' plot does appear to be the most simple and useful technique to plot high-dimensional data in two dimensions. Andrews' curve makes the interpretation of the results easier because it can map each multi-response observation into a function, *f(t)*, of a single variable, *t*. Thus, Andrews' plots form an ideal technique to start a general discussion on the topic of graphical representation of multi-dimensional observations. It is noted that the Andrews's Curve and its variations rely on the type of data set which is clearly demonstrated in Section 5.

This study suggests that in order to investigate the high dimensional data, the use of Andrews' curve can be beneficial. The ideal software and hardware environment has to allocate for interactive, high-resolution color plotting. Since, Andrews' curves are order dependent. The first few variables tend to dominate, so it is a recommended that to take the most crucial variables first. Also in these plots low frequencies are more readily perceived than high frequencies. The most essential variables can be identified, using either Principal Components Analysis (PCA) or Factor Analysis (FA). Such a plot of few numbers of retained variables would be probably more interpretable and investigation of data can subsist more readily. After initial examination, the variables should be re-arranged so as to take strongly associated variables together. Variables which emerge to discriminate groups should be set at the extreme frequencies in the Fourier series. The same data sets should be used for indulgent different variations because of the advantage of always using the same display is that, the analysts is more familiar with the map, and can interpret new information displayed on it effectively and rapidly.

REFERENCES

1. Andrews, D.F. (1972). Plots of High-Dimensional Data. *Biometrics*, 125-136.
2. Barnett and Lewis (1978). *Outliers in Statistical Data*. Chichester: John Wiley & Sons.
3. Embrechts, Paul & Herzberg, Agnes M. (1991). Variations of Andrews' plots, *International Statistical Review*, 59, 2, 175-194.
4. Economic Review of Pakistan 2002.
5. ESCAP Population Data Sheet 2003.
6. Everitt, B.S. and Dunn, G. (1990). *Applied Multivariate Data Analysis*, John Wiley and Sons, New York.
7. García-Osorio C. and Fyfe C. (2005). Visualization of High-Dimensional Data via Orthogonal Curves. *Journal of Universal Computer Science*, 11(11), 1806-181.
8. Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. John Wiley & Son, New York.
9. Khattree, R and Naik, D.N. (2002). Some New Suggestions and Applications. *J. Statist. Plann. and Infer.*, 100, 411-425.
10. Government of Pakistan (2003). *Pakistan Statistical Year Book*. Federal Bureau of Statistics, Islamabad.
11. Seber, G.A.F. (1983). *Multivariate Observations*. Wiley, New York.
12. Wegman, E.J. and Shen, J. (1993). Three Dimensional Andrews Plots and the Grand Tour. *Computing Science and Statistics*, 25, 284-288.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 223-234

A BAYESIAN LOOK AT THE PAIR COMPARISON MODEL WITH TIE AND ORDER EFFECT

Samina Satti[1] and **Muhammad Aslam**[2]
[1] Department of Statistics, University of Wah,
Wah Cantt, Pakistan. Email: samina.satti08@gmail.com
[2] Department of Statistics, Quaid-i-Azam University,
Islamabad, Pakistan. Email: aslamsdqu@yahoo.com

ABSTRACT

Barren (1978) introduces order effect in the Rao-Kupper model for the pair comparison experiments, which only consider tie effect. In present study Bayesian analysis of the Rao-Kupper model with tie and order effect is presented. We have calculated posterior estimates (mean, mode), preference probabilities and predictive probabilities for simulated data of four treatments, using two non- informative prior: Uniform prior and Jeffrey's prior. Results are also verified by Quadrature method and Gibbs sampling. Bayesian hypothesis testing, proposed by Aslam (2001), for comparison of treatments is also mentioned. The graphical display of the posterior (marginal) distributions and the appropriateness of the Rao Kupper model are also presented.

KEYWORDS

Non-informative prior; Bayes estimates; order effect; Posterior Distribution preference probabilities; Rao-Kupper model; Posterior estimates.

1. INTRODUCTION

The method of pair comparison may be considered as special rank order technique. This method is used in cases where objects are judged on subjective basis, it means that when it is impossible or impracticable to make relevant measurement in order to decide which of two objects is preferable. In our study we not only allow ties in our experiment but also consider order effect. The method of paired comparison has attracted the attention of people from a wide spectrum of interest. This method has found application in many areas such as; Statistics, Psychometrics, Marketing research, Preference measurement, Sports competition (Davidson and Farquhar, 1976). The present paper consists of the following sections. In Section 2, the introduction and mathematical model of the Rao-Kupper model with tie and order effect is discussed. Sections 3 and 4 deals with the notations, likelihood function. The Bayesian analysis is presented in Section 5 and 6. The appropriateness of the model is considered in section 7. Results (Posterior Estimates[mean, mode],Predictive Probabilities, Preference probabilities) related to Discussion on further area of research is given in Section 8 and 9.

2. THE RAO-KUPPER MODEL WITH TIE AND ORDER EFFECT

Rao-Kupper (1967) extends the Bradley-Terry model which allow for tied observation. They show that according to Bradley-Terry (1953) model for comparison of two treatments, results in a definite preference for one of the two treatments. But any model that does not allow for the possibility of ties is not making full use of information contain in the no preference class. A threshold parameter is introduced by them which permits ties in the model. The estimation and testing of the model parameter involve that of threshold parameter is also exhibited. They also justify the result for the use of model. Some comprehensive work has been done on Rao Kupper Model in different aspect, Aslam (2002, 2003). Barren (1978) presents a class of extensions to the Bradley-Terry model in pair comparison. Barren present special case of extension in Rao-Kupper model which is compare and evaluated by a numerical example. The extension is to accommodate ties in the individual comparisons or to estimate the effects of the order of presentation of the objects in these comparisons.

2.1 Mathematical Model

The mathematical model is given as below.[The construction of model can be seen in Barran, 1978]

$$\omega_{i.ij} = \frac{\gamma\theta_i}{\gamma\theta_i + \nu\theta_j}\ ,\ \omega_{j.ij} = \frac{\theta_j}{\gamma\nu\theta_i + \theta_j}\ ,\ \omega_{0.ij} = \frac{\theta_i\theta_j\gamma\left(\nu^2-1\right)}{\left(\gamma\theta_i+\nu\theta_j\right)(\gamma\nu\theta_i+\theta_j)}$$

Here $\omega_{i.ij}$ show the probability of selection of first treatment, $\omega_{j.ij}$ show the probability of selection of second treatment, $\omega_{0.ij}$ is the probability that when a judge declare tie, γ is order effect and ν denote tie.

Furthermore the sum of these probabilities is equal to 1. So the model follow a trinomial distribution. (Barren, 1978)

3. NOTATIONS OF THE MODEL

Here we define the notations to describe the data and the likelihood function for the model. If r_{ij} is the total number of comparisons then, $n_{ijk}(1)=1$ or 0 accordingly as treatment T_i is preferred to treatment T_j when treatment T_i is presented first in the k^{th} repetition of the comparison. $n_{ijk}(2)=1$ or 0 accordingly as treatment T_j is preferred to treatment T_i when treatment T_i is presented first in the k^{th} repetition of the comparison.

$n_{jik}(1)=1$ or 0 accordingly as treatment T_j is preferred to treatment T_i when treatment is T_j presented first in the k^{th} repetition of the comparison. $n_{jik}(2)=1$ or 0 accordingly as treatment T_i is preferred to treatment T_j when treatment T_j is presented first in the k^{th} repetition of the comparison.

$n_{ij}(i)=1$ or 0 accordingly as treatment T_i is preferred to treatment T_j when treatment T_i is presented first.

$n_{ij}(j)=1$ or 0 accordingly as treatment T_j is preferred to treatment T_i when treatment T_i is presented first.

$n_{ji}(j)=1$ or 0 accordingly as treatment T_j is preferred to treatment T_i when treatment T_j is presented first.

$n_{ji}(i)=1$ or 0 accordingly as treatment T_i is preferred to treatment T_i when treatment T_j is presented first.

$n_i = \sum_j \{n_{ij}(1)+n_{ji}(2)\}$ = is the total number of times T_i is preferred.

t_{ij} = 1 or 0 accordingly as treatment T_i is tied with T_j when T_i is presented first k^{th} repetition of the comparison.

t_{ji} = 1 or 0 accordingly as treatment T_i is tied with T_j when T_j is presented first k^{th} repetition of the comparison.

$t_i = \sum_j \left(t_{ij}+t_{ji}\right)=$ the total number of ties for the treatment T_i and T_j are tied $T=\sum\sum t_{ij}=$ the total number of times treatment T_i. $r_{ij}\left(i\right)=$ The number of times the treatment T_i and T_j are compared. T_i is selected, when T_i is presented first, plus T_i and T_j are tied. $r_{ij}(i)=n_{ij}(i)+t_{ij}$

$r_{ij}\left(j\right)=$ the number of times the treatment T_i and T_j are compared. T_j is selected when T_i is presented first, plus T_i and T_j are tied. $r_{ij}(j)=n_{ij}(j)+t_{ij}$

$r_{ji}\left(j\right)=$ the number of times the treatment T_i and T_j are compared. T_j is selected when T_j is presented first, plus T_i and T_j are tied. $r_{ji}(j)=n_{ji}(j)+t_{ji}$

$r_{ji}\left(i\right)=$ the number of times the treatment T_i and T_j are compared. T_i is selected when T_j is presented first, plus T_i and T_j are tied. $r_{ji}(i)=n_{ji}(i)+t_{ji}$

3.1 Likelihood Function for the Model

The likelihood function after observing the data of the trial is given by

$$l\left(x,\lambda,\nu,\theta_{1,}\theta_2....\theta_{_n}\right)=\prod_{i>j}^{m}\prod_{j=1}^{m}P_{ijk} \tag{1}$$

$$=\prod_{i=1}^{m}\prod_{j>i}^{m}\frac{r_{ij}}{n_{i.ij}\,!n_{j.ij}\,!t_{ij}}\left(\nu^2-1\right)^T\frac{\gamma^N\theta_i^{\,r_{ij}(i)}\,\theta^{\,r_{ji}(j)}}{\left(\theta_i\gamma+\theta_j\nu\right)^{r_{ij}(i)}\left(\theta_i\gamma\nu+\theta_j\right)^{r_{ji}(j)}} \tag{2}$$

where $r_{ij}(i)=n_{ij}(i)+t_{ij}$, $r_{ij}(j)=n_{ij}(j)+t_{ij}$, $r_{ji}(j)=n_{ji}(j)+t_{ji}$, $r_{ji}(i)=n_{ji}(i)+t_{ji}$, $N=\sum_i\sum_j n_{ij}(i)+T$,

4. ANALYSIS OF THE MODEL M=4

Now consider the case, when we have a number of treatment m=4, there are four parameters $\theta_1, \theta_2, \theta_3$ and θ_4 when three treatments are compared pair wise, ν is tie and γ is order effect. Then likelihood function can be written as

$$p(x;\theta_1,\theta_2,\theta_3,\theta_4,\gamma,\nu) = K^{-1}\left[\nu^2-1\right]^T\left[\gamma\right]^N\left[\theta_1\right]^{r_{121}+r_{212}+r_{131}+r_{312}+r_{141}+r_{412}}$$
$$\left[\theta_2\right]^{r_{211}+r_{122}+r_{231}+r_{322}+r_{241}+r_{422}}\left[\theta_3\right]^{r_{311}+r_{132}+r_{321}+r_{232}+r_{342}+r_{432}}$$
$$\left[1-\theta_1-\theta_2-\theta_3\right]^{r_{411}+r_{142}+r_{421}+r_{242}+r_{342}+r_{431}}\left[\frac{1}{A_1}\right]^{r_{121}}\left[\frac{1}{A_2}\right]^{r_{122}}\left[\frac{1}{A_3}\right]^{r_{211}}\left[\frac{1}{A_4}\right]^{r_{212}}\left[\frac{1}{B_1}\right]^{r_{131}}\left[\frac{1}{B_2}\right]^{r_{132}}$$
$$\left[\frac{1}{B_3}\right]^{r_{311}}\left[\frac{1}{B_4}\right]^{r_{312}}\left[\frac{1}{C_1}\right]^{r_{231}}\left[\frac{1}{C_2}\right]^{r_{232}}\left[\frac{1}{C_3}\right]^{r_{321}}\left[\frac{1}{C_4}\right]^{r_{322}}\left[\frac{1}{D_1}\right]^{r_{141}}\left[\frac{1}{D_2}\right]^{r_{142}}\left[\frac{1}{D_3}\right]^{r_{411}}\left[\frac{1}{D_4}\right]^{r_{412}}\left[\frac{1}{E_1}\right]^{r_{241}}$$
$$\left[\frac{1}{E_2}\right]^{r_{242}}\left[\frac{1}{E_3}\right]^{r_{421}}\left[\frac{1}{E_4}\right]^{r_{422}}\left[\frac{1}{F_1}\right]^{r_{431}}\left[\frac{1}{F_2}\right]^{r_{432}}\left[\frac{1}{F_3}\right]^{r_{432}}\left[\frac{1}{F_4}\right]^{r_{432}} \tag{4}$$

where K is normalizing constant., also

$A_1 = (\gamma\theta_1+\nu\theta_2)$, $A_2 = (\gamma\nu\theta_1+\theta_2)$, $A_3 = (\gamma\theta_2+\nu\theta_1)$, $A_4 = (\gamma\nu\theta_2+\theta_1)$,

$B_1 = (\gamma\theta_1+\nu\theta_3)$, $B_2 = (\gamma\nu\theta_1+\theta_3)$, $B_3 = (\gamma\theta_3+\nu\theta_1)$, $B_4 = (\gamma\nu\theta_3+\theta_1)$,

$C_1 = (\gamma\theta_2+\nu\theta_3)$ $C_2 = (\gamma\nu\theta_2+\theta_3)$, $C_3 = (\gamma\theta_3+\nu\theta_2)$, $C_4 = (\gamma\nu\theta_3+\theta_2)$,

$D_1 = (\gamma\theta_1+\nu(1-\theta_1-\theta_2-\theta_3))$ $D_2 = (\gamma\nu\theta_1+(1-\theta_1-\theta_2-\theta_3))$, $D_3 = (\gamma(1-\theta_1-\theta_2-\theta_3)+\nu\theta_1)$,

$D_4 = (\gamma\nu(1-\theta_1-\theta_2-\theta_3)+\theta_1)$, $E_1 = (\gamma\theta_2+\nu(1-\theta_1-\theta_2-\theta_3))$, $E_2 = (\gamma\nu\theta_2+(1-\theta_1-\theta_2-\theta_3))$,

$E_3 = (\gamma(1-\theta_1-\theta_2-\theta_3)+\nu\theta_2)$, $E_4 = (\gamma\nu(1-\theta_1-\theta_2-\theta_3)+\theta_2)$, $F_1 = (\gamma\theta_3+\nu(1-\theta_1-\theta_2-\theta_3))$,

$F_2 = (\gamma\nu\theta_3+(1-\theta_1-\theta_2-\theta_3))$, $F_3 = (\gamma(1-\theta_1-\theta_2-\theta_3)+\nu\theta_3)$, $F_4 = (\gamma\nu(1-\theta_1-\theta_2-\theta_3)+\theta_3)$

5: BAYESIAN ANALYSIS OF THE MODEL USING UNIFORM PRIOR FOR m=4

As there is no prior information available, thus according to Laplace's rule ,we use standard uniform distribution, based on the principle of in sufficient reason, as non-informative prior.

The prior distribution for the model when m=4, is given as

$$p(\theta_1,\theta_2,\theta_3,\gamma,\nu) \propto 1,\quad 0 \le \theta_i \le 1;\ \text{i=1,2,……m},\ \gamma \ge 0,\ \nu \ge 1$$

with $\theta_4 = 1-\theta_1-\theta_2-\theta_3$. (4)

For the purpose of statistical analysis and drawing graph we use the following simulated data

Table 1:Data for tie and order effect

Pairs	**(1,2)**	**(2,1)**	**(1,3)**	**(3,1)**	**(1,4)**	**(4,1)**	**(2,3)**	**(3,2)**	**(2,4)**	**(4,2)**	**(3,4)**	**(4,3)**
r_{ij}	30	30	30	30	30	30	30	30	30	30	30	30
$n_{i.ij}$	17	22	21	18	22	16	23	17	22	17	20	19
$n_{j.ji}$	10	6	7	9	7	12	6	10	6	11	7	8
$n_{0.ij}$	3	2	2	3	1	2	1	3	2	2	3	3

5.1 Posterior Distribution

By using likelihood function and Uniform distribution, the joint posterior distribution is given as

$$p\left(\theta_1,\theta_2,\theta_3,\gamma,\nu\right)\propto 1 \quad 0\le\theta_i\le 1 \tag{5}$$
$$\nu\ge 1,\gamma\ge 0,\theta_4=1-\theta_1-\theta_2-\theta_3;$$

$$p(\theta_{1,}\theta_{2,}\theta_3,\theta_4,\gamma,\nu|x) = K^{-1}\left[\nu^2-1\right]^T\left[\gamma\right]^N$$
$$\left[\theta_1\right]^{r_{121}+r_{212}+r_{131}+r_{312}+r_{141}+r_{412}}\left[\theta_2\right]^{r_{211}+r_{122}+r_{231}+r_{322}+r_{241}+r_{422}}$$
$$\left[\theta_3\right]^{r_{311}+r_{132}+r_{321}+r_{232}+r_{342}+r_{432}}\left[1\text{-}\theta_1\text{-}\theta_2\text{-}\theta_3\right]^{r_{411}+r_{142}+r_{421}+r_{242}+r_{342}+r_{431}}\left[\frac{1}{A_1}\right]^{r_{121}}\left[\frac{1}{A_2}\right]^{r_{122}}$$
$$\left[\frac{1}{A_3}\right]^{r_{211}}\left[\frac{1}{A_4}\right]^{r_{212}}\left[\frac{1}{B_1}\right]^{r_{131}}\left[\frac{1}{B_2}\right]^{r_{132}}\left[\frac{1}{B_3}\right]^{r_{311}}\left[\frac{1}{B_4}\right]^{r_{312}}\left[\frac{1}{C_1}\right]^{r_{231}}\left[\frac{1}{C_2}\right]^{r_{232}}\left[\frac{1}{C_3}\right]^{r_{321}}\left[\frac{1}{C_4}\right]^{r_{322}}$$
$$\left[\frac{1}{D_1}\right]^{r_{141}}\left[\frac{1}{D_2}\right]^{r_{142}}\left[\frac{1}{D_3}\right]^{r_{411}}\left[\frac{1}{D_4}\right]^{r_{412}}\left[\frac{1}{E_1}\right]^{r_{241}}\left[\frac{1}{E_2}\right]^{r_{242}}\left[\frac{1}{E_3}\right]^{r_{421}}\left[\frac{1}{E_4}\right]^{r_{422}}\left[\frac{1}{F_1}\right]^{r_{431}}\left[\frac{1}{F_2}\right]^{r_{432}}\left[\frac{1}{F_3}\right]^{r_{432}}\left[\frac{1}{F_4}\right]^{r_{432}} \tag{6}$$

Here K^{-1} is defined as the normalizing constant and is given as

The marginal distribution of θ_1 is given as

$$p\left(\theta_1|x\right)=\int_{\theta_2}^{1-\theta_1}\int_{\theta_3}^{1-\theta_1-\theta_2}\int_{\theta_4}^{1-\theta_1-\theta_2-\theta_3}\int_{\gamma=0}^{\infty}\int_{\nu=1}^{\infty}p\left(\theta_1,\theta_2,\theta_3,\theta_4,\gamma,\nu|x\right)d\nu d\gamma d\theta_4 d\theta_3 d\theta_2,\quad 0\le\theta_1\le 1 \tag{7}$$

Here $p\left(\theta_1,\theta_{2,}\theta_{3,}\theta_{4,}\lambda,\nu|x\right)$ is the joint posterior distribution. We also obtain other marginal posteriors distributions, namely $p\left(\theta_2|x\right)$, $p\left(\theta_3|x\right)$, $p\left(\theta_4|x\right)$, $p\left(\gamma|x\right)$ and $p\left(\nu|x\right)$.

5.2 The Posterior Estimates

We have calculated the posterior mean by using Quadrature method as well as Gibbs Sampling. The joint posterior mode of the parameters is calculated make a program in SAS package using command 'PROC SYSNLIN. The results are given in Table [3].

5.3 The Preference Probabilities

For the purpose of the ranking of the treatment parameters, we have calculated preference probabilities of the data set of table 1, which is based on the estimates of $\theta_1, \theta_2, \theta_3, \gamma, \theta_4$ and ν. These preference probabilities show the expected chances of preferring any one treatment to the other in any one comparison. The preference probabilities are calculated by using Gibbs sampling and are mentioned in table (2005)

5.4 Bayesian Testing of Hypotheses for m=4

Consider the two hypotheses for comparing the two treatment parameters are as follows $H_{12}: \theta_1 > \theta_2$ and $H_{21}: \theta_2 \geq \theta_1$. The posterior probability p_{12} for H_{12} is $p_{12} = p(\theta_1 > \theta_2)$ and $q_{12} = 1 - p_{12}$ is the posterior probability for H_{21}. The posterior probability $\{p_{12}\}$ for H_{12} is obtained as

$$p_{12} = p(\theta_1 > \theta_2) = p(\theta_1 - \theta_2 > 0) = p(\phi > 0)$$

where $\phi = \theta_1 - \theta_2$, $\xi = \theta_1$ and $1 - \xi - (\xi - \phi) = \theta_3$.

The decision rule used here, for accepting or rejecting the above hypotheses is

Let s=min ($p_{12,} q_{21}$), If p_{ij} is small then H_{ji} is accepted, if q_{ij} is small, H_{ij} is accepted and if s>0.1, the decision is inconclusive. (Aslam, 2001)

In the Bayesian hypothesis testing, the posterior probabilities of hypotheses are calculated using the posterior distribution. The Quadrature method is used in SAS package, and the posterior probability H_{12},

$$p_{12} = p\left[(\theta_1 - \theta_2) > 0 | x\right]$$

comes out to be 0.0001 and q_{12} is 0.9999 for the data set in Table 1. According to the rule discussed above, H_{12} is rejected. Similarly, the values of other posterior probabilities are p_{13}=0.87202 with q_{13}=0.12798, indicating that decision here is inconclusive, p_{23}=0.99246 with q_{23}=0.00754 show that H_{23} is accepted, p_{24}=0.99859 with q_{24}=0.00141 show that H_{24} is accepted, p_{34}=0.87206 with q_{34}=0.12794 show that decision is inconclusive and p_{14}=0.1753 with q_{14}=0.8247 indicate that decision is also inconclusive.

5.5 Predictive Probabilities

We use the Gibbs sampling for calculating the predictive probability of preferring the treatment T_i to the treatment in a single future comparison, then we get the future probability of declaring a tie when comparing treatment T_i over in a single comparison.

$$P_{12}(1) = \int_{\theta_1=0}^{1} \int_{\theta_2=0}^{1-\theta_1} \int_{\theta_3=0}^{1-\theta_1-\theta_2} \int_{\nu=1}^{\infty} \int_{\gamma=0}^{\infty} \psi_{12}(1)\, p(\theta_1, \theta_2, \theta_3, \theta_4 \gamma, \nu | x)\, d\gamma\, d\nu\, d\theta_3\, d\theta_2\, d\theta_1 \tag{8}$$

The calculated predictive probabilities are given in table (3).

6. BAYESIAN ANALYSIS OF THE MODEL USING JEFFREY'S PRIOR FOR m=4

We choose Jeffreys prior for the parameters. The Jeffreys is considered as the non-informative prior distribution. When we did not have much information about the parameters then the prior, which is most frequently used is Jeffery prior (1946). Berger (1985) argues that Bayesian analysis using non-informative priors is the single most powerful method of analysis. According to Jeffreys' rule the non-informative prior is defined as the density of the parameters proportional to the square root of the determinant of the Fisher information matrix. Symbolically, let $\phi = (\phi_1, \phi_2 .. \phi_m)^t$ is a vector of parameter, the Jeffrey's (1946,1961) prior distribution $\{p_j(\phi)\}$ is given as

$$p_j(\phi) \propto \sqrt{\det\{I(\phi)\}}, \tag{9}$$

in above expression 'det' denotes determinant and the $I(\phi)$ is the $(m \times m)$ Fisher information matrix which is the logarithm of likelihood function $L(\phi)$ of parameters ϕ and partially differentiating twice with respect to the parameters as given below,

$$I_{ij}(\phi) = -E\left\{\frac{\{\partial^2 \ln L(\phi)\}}{\partial\phi_i \partial\phi_j}\right\}, \tag{10}$$

Here E denotes expectation on data and i and j stand for rows and columns of determinants respectively.

We design a program in SAS package, and calculate the numerical form of the Jeffery's prior. The joint posterior distribution for the m=4 is given as

6.1 Posterior Distribution

$$p(\theta_1,\theta_2,\theta_3,\theta_4,\gamma,\nu \mid x) = K^{-1}(\theta_1,\theta_2,\theta_3,\theta_4,\gamma,\nu)\left[\nu^2-1\right][\gamma]^N$$
$$[\theta_1]^{r_{121}+r_{212}+r_{131}+r_{312}+r_{141}+r_{412}}[\theta_2]^{r_{211}+r_{122}+r_{231}+r_{322}+r_{241}+r_{422}}[\theta_3]^{r_{311}+r_{132}+r_{321}+r_{232}+r_{342}+r_{432}}$$
$$[1-\theta_1-\theta_2-\theta_3]^{r_{411}+r_{142}+r_{421}+r_{242}+r_{342}+r_{431}}\left[\frac{1}{A_1}\right]^{r_{121}}\left[\frac{1}{A_2}\right]^{r_{122}}\left[\frac{1}{A_3}\right]^{r_{211}}\left[\frac{1}{A_4}\right]^{r_{212}}$$
$$\left[\frac{1}{B_1}\right]^{r_{131}}\left[\frac{1}{B_2}\right]^{r_{132}}\left[\frac{1}{B_3}\right]^{r_{311}}\left[\frac{1}{B_4}\right]^{r_{312}}\left[\frac{1}{C_1}\right]^{r_{231}}\left[\frac{1}{C_2}\right]^{r_{232}}\left[\frac{1}{C_3}\right]^{r_{321}}\left[\frac{1}{C_4}\right]^{r_{322}}$$
$$\left[\frac{1}{D_1}\right]^{r_{141}}\left[\frac{1}{D_2}\right]^{r_{142}}\left[\frac{1}{D_3}\right]^{r_{411}}\left[\frac{1}{D_4}\right]^{r_{412}}\left[\frac{1}{E_1}\right]^{r_{241}}\left[\frac{1}{E_2}\right]^{r_{242}}\left[\frac{1}{E_3}\right]^{r_{421}}\left[\frac{1}{E_4}\right]^{r_{422}}$$
$$\left[\frac{1}{F_1}\right]^{r_{431}}\left[\frac{1}{F_2}\right]^{r_{432}}\left[\frac{1}{F_3}\right]^{r_{432}}\left[\frac{1}{F_4}\right]^{r_{432}} \tag{11}$$

where $p_j(\theta_1,\theta_2,\theta_3,\theta_4,\gamma,\nu)$ is the Jeffreys prior distribution, K is the normalizing constant. And $p(\theta_1,\theta_2,\theta_3,\theta_4,\gamma,\nu)$ is the joint prior probability distribution. Similarly the marginal posterior distribution for $\theta_1,\theta_2,\theta_3,\theta_4,\gamma$ and ν can be obtained.

6.2 The Posterior Estimates

By utilizing Jeffrey prior, we have calculated the posterior mean by using Quadrature method and Gibbs Sampling. The joint posterior mode of the parameters $(\theta_1,\theta_2,\theta_3,\theta_4,\gamma,\nu)$ is also calculated. The results are given in Table [3].

6.3 The Preference Probabilities

We also have calculated preference probabilities of the data set of table 1 by using Jeffry prior, which is based on the estimates of $\theta_1,\theta_2,\theta_3,\gamma,\theta_4$ and ν.The technique used for obtaining preference probabilities is Gibbs sampling, they are given in table[3].

6.4 Bayesian Testing Of Hypotheses FOR m=4

In the Bayesian hypothesis testing, the posterior probabilities are calculated using the posterior distribution. The Quadrature method is used in SAS package and the posterior probabilities is given as follows. p_{12} comes out to be 0.0001 and q_{12} is 0.9999, so H_{12} is rejected. Similarly, the values of other posterior probabilities are p_{13}=0.8735, with q_{13}=0.1265, indicating that decision here is inconclusive, p_{23}=0.9931 with q_{23}=0.0069 shows that H_{23} is accepted, p_{24}=0.99754 with q_{24}=0.00246 shows that H_{24} is accepted, p_{34}=0.87234 with q_{34}=0.12766 show that decision is inconclusive and p_{14}=0.17573 with q_{14}=0.82427 indicate that decision is also inconclusive.

7. APPROPRIATNESS OF THE MODEL FOR m=4

For testing the appropriateness of the model, we compared the observed number of preferences with expected number of preferences and if the discrepancies are small then, model is considered consistent. We employ the χ^2 statistics for testing the hypothesis given in Aslam [2005]. For testing the appropriateness of the model, we compared the observed number of preferences with expected number of preferences and if the discrepancies are small then, model is considered consistent. We employ the χ^2 statistics for testing the hypothsis. Let consider : $\hat{n}_{i.ij}$ = the expected number of times treatment T_i is preferred to to treatment T_j when T_i is presented first, and $\hat{n}_{0.ij}$ =the expected number of times treatment T_i and treatment T_j are tied. As the order effect is present, so we also have $\hat{n}_{j.ji}$ =the expected number of times treatment T_j is preferred to treatment T_i when T_j is presented first, and $\hat{n}_{0.ji}$ =the expected number of times treatment T_j and treatment T_i are tied when T_j is presented first. The χ^2 statistics has the following form:

$$\chi^2 = \sum_{i<j}^{m} \left\{ \begin{array}{l} \left\{ \dfrac{\left(n_{i.ij} - \hat{n}_{i.ij}\right)^2}{\hat{n}_{i.ij}} + \dfrac{\left(n_{j.ij} - \hat{n}_{j.ij}\right)^2}{\hat{n}_{j.ij}} + \dfrac{\left(n_{0.ij} - \hat{n}_{0.ij}\right)^2}{\hat{n}_{0.ij}} \right\} \\ + \left\{ \dfrac{\left(n_{j.ji} - \hat{n}_{j.ji}\right)^2}{\hat{n}_{j.ji}} + \dfrac{\left(n_{i.ji} - \hat{n}_{i.ji}\right)^2}{\hat{n}_{i.ji}} + \dfrac{\left(n_{0.ji} - \hat{n}_{0.ji}\right)^2}{\hat{n}_{0.ji}} \right\} \end{array} \right\} \tag{11}$$

with {2m(m-1)-(m+1)} degrees of freedom.

Calculations for expected number preference and ties are done by using the formula defined in (16). The results obtained are presented in the Table 2.

Table 2 Observed and Expected Number of Preferences

Pairs(i,j)	(1,2)	(2,1)	(1,3)	(3,1)	(2,3)	(3,2)	(1,4)	(4,1)	(2,4)	(4,2)	(3,4)	(4,3)
$n_{i.ij}$	17	22	21	18	22	16	23	17	22	17	20	19
$\hat{n}_{i.ij}$	18	22	21	19	23	17	22	22	24	16	21	19
$n_{j.ij}$	10	6	7	9	7	12	6	10	6	11	7	8
$\hat{n}_{j.ij}$	10	6	7	8	5	11	6	6	5	12	7	9
$n_{0.ij}$	3	2	2	3	1	2	1	3	2	2	3	3
$\hat{n}_{0.ij}$	2	2	2	2	2	2	2	2	2	2	2	2

Using the above calculations and the formula defined in (15), we obtain χ^2=9.24 with p-value 0.9692. So we can conclude that there is no evidence that the model does not fit.

8. COMPARISON OF RESULTS

The results obtained from both priors give approximately same results. A detail table is given below.

Table 3: Summary of the Results for the Model for m=4

	Using Quadrature Method		**Using Gibbs Sampling**		
Results / **Means**	**Using Uniform Prior**	**Using Jeffrey Prior**	**Results** / **Means**	**Using Uniform Prior**	**Using Jeffrey Prior**
γ	2.4322	2.5462	γ	2.4577	2.4915
θ_1	0.2872	0.2686	θ_1	0.2576	0.2714
θ_2	0.3119	0.3371	θ_2	0.3427	0.3491
θ_3	0.2008	0.2001	θ_3	0.2044	0.2007
θ_4	0.2008	0.1942	θ_4	0.1952	0.1788
ν	1.3073	1.2806	ν	1.3003	1.3191

Table 3 (continued)

Using Gibbs Sampling					
Predictive Probability	**Using Jeffrey Prior**	**Using Uniform Prior**	**Preference Probability**	**Using Jeffrey Prior**	**Using Uniform Prior**
$P_{1.12}$	0.5929	0.5818	$\psi_{1.12}$	0.5949	0.5869
$P_{2.12}$	0.2831	0.2977	$\psi_{2.12}$	0.2813	0.2939
$P_{0.12}$	0.1239	0.1204	$\psi_{0.12}$	0.1238	0.1192
$P_{2.21}$	0.7095	0.7102	$\psi_{2.21}$	0.7084	0.7155
$P_{1.21}$	0.1906	0.1943	$\psi_{1.21}$	0.1913	0.1904
$P_{0.21}$	0.0999	0.0955	$\psi_{0.21}$	0.1003	0.0941
$P_{1.13}$	0.7171	0.6997	$\psi_{1.13}$	0.7186	0.7043
$P_{3.13}$	0.1491	0.1607	$\psi_{3.13}$	0.1837	0.1989
$P_{0.13}$	0.0978	0.0982	$\psi_{0.13}$	0.0977	0.0968
$P_{3.31}$	0.5833	0.5944	$\psi_{3.31}$	0.5828	0.6
$P_{1.31}$	0.2914	0.2871	$\psi_{1.31}$	0.2915	0.2828
$P_{0.31}$	0.1252	0.1185	$\psi_{0.31}$	0.1257	0.1172
$P_{1.14}$	0.741	0.7145	$\psi_{1.14}$	0.7414	0.7138
$P_{4.14}$	0.1675	0.1908	$\psi_{4.14}$	0.167	0.1917
$P_{0.14}$	0.0915	0.0947	$\psi_{0.14}$	0.0916	0.0945
$P_{4.41}$	0.5543	0.5774	$\psi_{4.41}$	0.5544	0.5889
$P_{1.41}$	0.3161	0.3013	$\psi_{1.41}$	0.3159	0.2923
$P_{0.41}$	0.1296	0.1213	$\psi_{0.41}$	0.1296	0.1189
$P_{2.23}$	0.7665	0.7553	$\psi_{2.23}$	0.7666	0.7601
$P_{3.23}$	0.1491	0.1607	$\psi_{3.23}$	0.1489	0.1573
$P_{0.23}$	0.0843	0.084	$\psi_{0.23}$	0.0845	0.0826
$P_{3.32}$	0.5197	0.5245	$\psi_{3.32}$	0.5206	0.5299
$P_{2.32}$	0.3473	0.3484	$\psi_{2.32}$	0.3461	0.3441
$P_{0.32}$	0.133	0.127	$\psi_{0.32}$	0.1333	0.126
$P_{2.24}$	0.7876	0.7678	$\psi_{2.24}$	0.7867	0.7684
$P_{4.24}$	0.1343	0.1517	$\psi_{4.24}$	0.1348	0.1513
$P_{0.24}$	0.0781	0.0804	$\psi_{0.24}$	0.0785	0.0803
$P_{4.42}$	0.4899	0.5068	$\psi_{4.42}$	0.4917	0.5184
$P_{2.42}$	0.3745	0.3647	$\psi_{2.42}$	0.3727	0.3546
$P_{0.42}$	0.1356	0.1285	$\psi_{0.42}$	0.1356	0.127
$P_{3.34}$	0.6797	0.6648	$\psi_{3.34}$	0.6795	0.6643
$P_{4.34}$	0.2135	0.2294	$\psi_{4.34}$	0.2133	0.2301
$P_{0.34}$	0.1068	0.1058	$\psi_{0.34}$	0.1072	0.1056
$P_{4.43}$	0.6258	0.6327	$\psi_{4.43}$	0.6272	0.6435
$P_{3.43}$	0.2561	0.2549	$\psi_{3.43}$	0.2546	0.2468
$P_{0.43}$	0.1181	0.1123	$\psi_{0.43}$	0.1182	0.1097

9. SUMMARY

The results obtained under uniform prior, with those obtained under Jeffrey's prior are approximately same. The posterior estimates obtained by using two non-informative priors, show that $\theta_2 \rightarrow \theta_1 \rightarrow \theta_3 \rightarrow \theta_4$ The preference probabilities obtained by using Uniform prior and Jeffrey's prior show the same results. The predictive probabilities for future comparison also show approximately same results by using two approaches. As the both priors give approximately same results, so it can be concluded that either of the two priors can be used as a non-informative prior. Our most of the calculated, values agree up to two significant digits. By results it is also shown that the treatment, which is presented first, is always preferred.

We used Quadrature method for Bayesian analysis. We also used Gibbs sampling for the numerical solution. Here the results are based on 5500 iterations. A program is designed in SAS package by discarding the first 500 observation as Burn-in and taking a random sample of 500 random varieties after drawing every 10^{th} value. So it is suggested that the Gibbs sampling can be used for numerical solution like high dimensional integration.[for more detail Aslam (2007)].

Here we observe that the results from both approaches (Quadrature method and Gibbs sampling) are nearly very similar.

For further research, this work can be extended in many directions for the Bayesian analysis of paired comparison models. One may increase the number of treatments (i.e. more than 4). The results for paired comparison data using the Rao-Kupper Model (tie and order effect) can be compared with those obtained using other models. Efforts can be made the formulation of informative prior for the Bayesian analysis for the Rao-Kupper model.

REFERENCES

1. Aslam, M. (2002). Bayesian Analysis for Paired Comparison Model Allowing Ties, and Not Allowing Ties. *Pak. J. Statist.,* 18(1), 53-60.
2. Aslam, M. (2002). Reference Prior For the Parameters of Rao-Kupper Model. *Proc Pakistan Acad. Sci*., 39(2), 215-223.
3. Aslam, M. (2005). Bayesian Comparison of the Paired Comparison Models Allowing Ties. *J. Statist. Theo. and Applica.,* 4(2),161-171.
4. Aslam, M. (2007). Multidimensional Integration Using the Gibbs Sampling. *South Asian Bulletin of Mathematics*, 31(2), 423-432.
5. Barren, Van. A. (1978). On a Class of Extension to the Bradley-terry Model in Paired Comparisons. *Statistica Neerlandica*, 32, 57-67.
6. Berger, J.O. (1985). *Statistical Decision Theory and Bayesian analysis*. (2 Ed.). Springer-Verlag, New York.
7. Bradley, R.A. and Terry, M.E. (1952). The Rank Analysis of Incomplete Block Design, I. The Method of Paired Comparisons. *Biometrika*. 39, 324-345.
8. Bradley, R.A. (1953). Some Statistical Methods in Taste Testing and Quality Evaluation. *Biometrics*. 9, 22-38.
9. Davidson, R.R. and Beaver, R.J. (1977). On Extending the Bradley-Terry Model to Incorporate Within Pair Order Effects. *Biometrics*, 33, 693-702.

10. Jeffreys, H.S. (1946). An Invariant form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society*: A. 186, 453-461.
11. Rao, P.V. and Kupper. L.L. (1967). Ties in paired comparison experiment. A Generalization of Bradley-Terry Model. *J. Amer. Statist. Assoc.,* 62, 194-204.
12. Thurstone, L.L. (1927). Law of Comparative Judgment. *Psychological Review*, 34, 273-286.
13. Zermelo, E. (1929). Die Berechnung Der Turnier-Ergebnisse Als Ein Maximum Problem Der Wahrscheinlichkeitsrechnang. *Math. Zeit.,* 29, 436-496.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 235-254

PERFORMANCE EVALUATION OF DIFFERENT MODELS FOR VOLATILITY FORECASTING OF ASSET PRICES

Sehrish Saleem[1] and **Zahoor Ahmad**[2]
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: [1]sehrishsaying@gmail.com
[2]zahoor.ahmed@uog.edu.pk

ABSTRACT

Financial market forecast is an essential area of market research. Many researchers worked on volatility forecasting and there is a need to improve accuracy of forecast by trying different models as better forecast translates into better risk management and better pricing of options. In this paper, our aim is to compare performance of forecasting financial volatility by using Neural Network Models and classical GARCH family of models, of a major oil company, PSO (Pakistan State Oil). More precisely, the results suggest that Neural Network Model provides improved performances in forecasting the financial volatilities than other considered models.

KEYWORDS

Volatility; neural network models; GARCH type models, PSO.

INTRODUCTION

Precise predictions of stock market indexes play a vital role, for the financiers to hedge against apparent market risks, also for the market speculators and arbitrageurs, an opportunity to make profits through trading indexes. Undoubtedly, researchers and practitioners acquire intense implications and significance of, being able to precisely forecast stock market index.

Volatility in financial markets is an important concern to the theory and practice of risk management, asset pricing and asset allocation. It is also foremost for the reason that it shed further light upon the procedure of data generating. There has been extensive volatility (and uncertainty) in the long-ago in mature and promising financial markets worldwide. However earlier, it was considered that financial models assume volatilities to be constant, it is widely accepted that volatility varies along with time, amongst both practitioners and scholars. This acceptance initiated a wide-ranging research programs into the distributional along with vibrant properties of stock market volatility. Return volatility is, no doubt, fundamental to financial economics. Indeed, as illustrious by Campbell et al. (1997). “What differentiates financial economics is the essential role that uncertainty plays in both financial theories along with its experiential implementation. Certainly in the lack of uncertainty, the problems of financial economics condense to exercises in basic microeconomics”.

It is basically defined as a standard deviation or conditional variance of stock returns which is not directly observable. The most profitable decision of investors is based on variance of returns which can fluctuate over time, so it became imperative to model and also to forecast the conditional variance Campbell et al. (1997). In the production of goods and services oil is one of the essential inputs. Without the substitute inputs a rise in oil prices also increases the costs of production that in turns decrease cash flows along with stock prices. Moreover increase in oil prices will surely increase discount rate influenced by inflation pressures that further leads to take decision of increasing the inflation rate by the central bank. Consequently, change in the stock price along with change in discount rate relative to book value will directly affect the corporate investment. Though the change in the direction of stock price depends upon whether there is a consumer of oil or a producer of oil also of oil related products. As in the world market most of the companies are oil consumer, it became rational to model oil prices volatility and also forecasting the future prices (Sharma N, 1998).

The traditional econometric models for forecasting volatility comprise of GARCH family of models. GARCH type models were firstly introduced by Engle (1982) and Bollerslev (1986). Engle (1982) proposed the ARCH (p) model as financial time series model and used it to model the time series having volatility clustering along with fat tails. Bollerslev (1986) estimated variance which was based on last m observations and a long-run average variance *VL*, the less weight was given to the observation, as much older was that observation. Since by using ARCH (p) in forecasting volatility, the time lag p gets much larger than the model GARCH (p, q) was suggested by the author. Nelson (1991) proposed EGARCH model which consider leverage effect, a weakness of GARCH model that does not consider negative shocks effects. Variance of same magnitude is more heavily increased by negative return as compared to positive return. Also change in stock prices are inversely related to change in stock market volatility. Accordingly this asymmetric shock is known as leverage effect.

As the nature of market is vibrant therefore a variety of market variables must be considered during financial analysis. Traditional financial time series models require very strict assumptions regarding the distribution of time series, all these traditional models do not reflect entire market variables. Moreover they are often linear, consequently may not capture nonlinear behavior. Various computational techniques are available to capture modern behavior volatility analysis, as for the imperfection of traditional models; greater emphasis has been given for modeling and forecasting market variable by these techniques. Such as Genetic Algorithms (GAs), Principle Component Analysis, Artificial Neural Networks (ANNs) and many other designed for analyzing time series and multifarious systems.

LITERATURE REVIEW

Yochanan and Dorota (2000) conducted a study to observe the dynamic interrelations between major world stock markets by using artificial neural networks. Authors used the data of daily stock market indices of the major world stock markets of Canada, France, Germany, Japan, United Kingdom (UK), the United States (US), and the world excluding US (World). The results indicated that Neural Systems could be used as a substitute or supplemental method for forecasting financial variables and thus justified the potential use of this model by practitioners. Engle (2001) considered the ability of GARCH-family models to capture stylized facts such as mean-reversion and persistence, asymmetry such that the sign of an innovation also have an effect on volatility and the likelihood of exogenous or predetermined variables influencing volatility. Author found indication consistencies with the theoretical result that the empirical results attain are dependent on the sampling frequency a negative aspect of the GARCH specification. Jun Yu (2002) estimate the performance of nine different models for forecasting stock price volatility using daily New Zealand data. Author used different measures for evaluating accuracy of different models. The stochastic volatility model provided the finest performance among all other models. Among ARCH family models GARCH (3, 2) model performed most excellent, but a slight sensitive to the choice of evaluation measures. Exponentially weighted moving average models and regression models did not achieve fine outcome according to any assessment measure, in distinction to the consequences found in different markets. Jane et al, (2002) provided empirical evidence on the comparative macroeconomic forecasting performance of Linear along with Nonlinear Models. ARIMA and VAR models are used as Linear Forecasting Models while Neural Network Models are used as Nonlinear Forecasting Models. The results suggested that Nonlinear Models provided better within-sample and out-of sample forecasts and Linear Models are simply a subset of them. Shaikh (2004) present a primer in general for using Neural Networks for forecasting market variables, and in particular, forecasting volatility of the S&P 500 Index futures prices. Author of this study compared volatility forecasts from Neural Networks with oblique volatility from S&P 500 Index futures options using the Barone-Adesi and Whaley (BAW) model for pricing American options on futures. Results of the study showed that forecasts from Neural Network Models did better than Implied Volatility Forecasts. Implied Volatility Forecasts are found to be considerably different from Realized Volatility in two of three cases. Volatility forecasts from Neural Networks were not found to be significantly different as of Realized Volatility. Abdullah and Guven (2008) conducted a study to forecast volatility of daily stock market of ten different countries by employing seven GARCH Family Models. The conclusion of study put emphasis on that the category of asymmetric volatility models did better than the historical model in forecasting of stock market volatility. Fahimifard, et al. (2009) compared ANFIS and ANN as the Nonlinear Models and ARIMA and GARCH models as Linear Models for forecasting daily Iran Rial and Rial/US$. Authors of this study found that Nonlinear Models did better than Linear Models by using three forecast evaluation criteria (R square, MAD and RMSE), Also that ANFIS performed effective role to improve Iran’s exchange rate forecasting accuracy. Dhamija and Bhalla (2010) compared the predictive accuracy of Neural Networks and Conditional Heteroscedastic Models like ARCH, GARCH, GARCH-M, TGARCH, EGARCH and IGARCH, for forecasting the Exchange Rate Series. The results of study showed that both Neural Network and Conditionally Heteroscedastic Models can be effectively used for prediction. RBF networks did significantly better than MLP networks in Neural Networks case. IGARCH and TGARCH fare better than other Conditional Heteroscedastic Models. Neural Networks' performed better than that of Conditional Heteroscedastic Models in forecasting Exchange Rate and the implied volatility of NIFTY options.

MATERIAL AND METHOD

In this study the volatility behavior of individual stock return series of an oil company, Pakistan State Oil Stock index taken from Karachi Stock Exchange 100 index, is examined along with we compared the relative effectiveness of Econometric and Neural Network Methods in forecasting these oil prices. Data is collected from website *www.birecorder.com*. The data is sampled from first January 2003 to 26 of August 2008. The data is constrained to 26th of August 2008 for avoiding any biasness in kurtosis of returns of the data as after 26th of August 2008 no trading at KSE 100 index due to America's credit crunch policy, after that when market reopen prices jump down from price level so inclusion of data after 26th August 2008 does not give clear decision about the choice of model, therefore data is restricted to 26th of August 2008. The data spans continuous sequences of 1387 daily observation.

For econometric modeling data is divided into two parts one for the in-sample estimation and the other is to conduct out-of-sample tests. Data consists of 1387 observation among these observations, 1256 observations are used for in-sample estimation and remaining 122 observations are used for out-of-sample validation. The reason of this unequal split is that the GARCH specification required large sample sizes to produce good estimates.

For neural network models It is frequently agreed that, though in some applications many hidden layers are useful, one hidden layer is usually both sufficient moreover efficient. Therefore, we use only single hidden layer into our neural network model. Number of neurons in the hidden layer is selected by the intelligent problem solving method, different number of neurons is used in models, and only one neuron is used in the output layer in the analysis.

Direct statistical analysis of financial price is difficult because consecutive prices are highly correlated, and the variances of prices increase with time. This makes it usually more convenient to analyze changes in prices. The logarithm of relative prices changes are multiplied by 100, is used to calculate continually compound daily stock returns. The use of logarithm prices changes prevents non stationary of the level of stock prices from affecting stock return volatility. Let is defined as the share price then, continuously compound return at time t is defined as

$$= \quad [\ /\ ]$$

where and are the closing prices of PSO share price at days t and t-1 respectively.

GARCH-Family Models

Statistical properties of stock returns were primarily examined by Mandelbrot (1963) and Fama (1965). For the period of 1980s, Engle (2003) worked upon improving time-series analysis. In that time period statistical techniques mostly treated volatile variables, like stock prices, as constants. However these variables can change significantly along with time. Engle (2003) examine the variation in stock returns and developed ARCH (autoregressive conditional heteroscedasticity) models. These models predict future volatility by using previously observed patterns of variance. ARCH models (Engle, 1982;

Tsay, 2005 Dhamija A.K, 2010), characterized by mean along with volatility equations specified as

$$= \quad + \quad ; \quad =$$

$$= \quad + \sum$$

$$= \quad + \sum \qquad + \quad , \qquad = \quad + 1, \dots ,$$

This is ARCH model of order 'p' where , stands for the error term, sample size is denoted by *T*. Modifications of ARCH models are extensively being used in banking and finance. In empirical estimation of ARCH models, normally, a relative long lag is requisite as of the need to avoid negative variance parameters, furthermore persistency of volatility this calls for a huge number of parameters being subject to inequality limitations. To meet this particular problem and to ease the computational burden a practical extension of ARCH models is proposed by Bollerslev, the generalized ARCH (GARCH) model. Where *et* follows a GARCH model of order (p, q) (Bollerslev, 1986; Tsay, 2005, Dhamija.A.K, 2010). If

$$= \quad ,$$

$$= \quad + \sum \qquad + \sum$$

There are also assumptions on GARCH model, first is that our $\geq 0$and the second one is$\sum^{(\ ,)} \quad + \quad < 1$. The limitation on $+$ implies that, unconditional variance of *et* be finite, while its conditional variance changes over time. and are ARCH as well as GARCH parameters, correspondingly. The typical GARCH model is symmetric in its response to past innovations, it means that positive along with negative shock of have equivalent effect on the volatility irrespective of the sign of the shock (as it takes square of past innovations in the variance equation) which is mainly wrong for this reason asymmetric GARCH models were proposed that enables conditional variance to react asymmetrically to rises along with falls in innovations. Since good news and bad news have dissimilar effects on the volatility we considered alternative GARCH models in an attempt to capture the asymmetric nature of volatility responses. An EGARCH model can be specified as

$$= \quad ,$$

$$(\quad) = \quad + \sum \qquad \frac{}{} + \sum \qquad\qquad + \sum \qquad \frac{| \quad |}{} - \quad \frac{-}{-}$$

As *et* ~ *N* (0,), so standardized variable $\frac{}{}$ follows the standard normal distribution and accordingly $E \frac{| \ |}{} \;=\; \frac{-}{-}$. Parameters capture the effect of negative shocks known as leverage effect. As for $\frac{}{} > 0$ (a good news) the innovation impact *et–i* is$(\quad + \quad)\frac{}{}$ also for $\frac{}{} < 0$ (a bad news) the impact is$(- \quad + \quad)\frac{}{}$. $(\quad)$ Responds symmetrically to $\frac{}{}$ if becomes 0. TGARCH model with order of (p, q) was proposed by (Glosten, et al. 1993). The TGARCH model can also handle leverage affect (the effect of negative shocks) via assuming the following form

$$= \quad + \sum \qquad + \sum \qquad + \sum$$

where

$$= \begin{cases} 1, & < 0 \\ 0, & \geq 0 \end{cases}$$

Along with , , as non-negative parameters are satisfying assumptions parallel to those of GARCH models. It can be seen that a negative *et–i* with has a larger impact on ($+$) as compared to a positive contribution . Estimation of ARCH class models frequently involves maximizing likelihood function. Eviews version 6.software is used for estimating these econometric models. Estimation of models for stock index is done through Marquardt algorithm. There are also some assumptions before actually estimating these econometric models. In the pre-estimation analysis Ljung-Box test is used to classify the mean equation by examining the autocorrelation as well as partial autocorrelation of the return series. Significant Ljung-Box test for squared return series during the pre-estimation analysis is the indication of volatility clustering present in the financial time series. The ARCH test is used to test whether a time series is homoscedastic or else hetroskedastic, it is Lagrange Multiplier (LM) test for autoregressive conditional hetroskedasticity (ARCH) in the residual (Engle, 1982). Jarque-Bera test statistic is used for testing whether the series is normally distributed. In order to model a time series it must be stationary. The Augmented Dickey-Fuller test, or the ADF-test, for the stationary was derived in 1979 by Dickey and Fuller to test the null hypothesis that a time series is non-stationary against the alternative that is stationary.

Neural Networks

To understand the architecture of neuron, let us take a look on a biological neuron. Neuron is the essential part of a nervous system. All of the neurons of nervous system are having the same structure, but independent of their size. A biological neuron structure is shown in Figure 1. It consists of four main physical parts, including cell body, axon, dendrites and synaptic terminals. The dendrites are the signal receivers, which accept signal from outside or other neurons. Cell body generates impulses along with it is a place for message processing. Axon is basically the channel to transmit the messages that are generated by a neuron to the other neurons; also it can transmit messages to the outside, such as muscle fibers.

Once observing a biological neuron, we might not have any difficulty in understanding an artificial neuron in the neural networks that is mathematical neuron. It is no doubt a crucial block for building a neural network; furthermore it is fundamental to the functioning of a neural network. As the crucial building block of a neural network is the neuron**.** A neuron can be represented by mapping *y:* $\rightarrow$ *and* transforming n dimensional inputs in to the real numbers. Basically neuron consisting of propagation function *f*: $\rightarrow$ Along with activation function *g: R*$\rightarrow$ [0, 1] that is *g*(*x*) can take the output of f(x) as an argument.

Therefore, a neuron is represented in the simple form as

$y(x) = g(f(x; w))$.

Say if function $f(x)$ is a polynomial function then its degree might be called the order of that neuron also its coefficients are parameters of neuron. All these neurons are accumulating in layered structure for the construction of an artificial neural network (ANN). Theoretical framework of the neural networks has been adopted from Giacomini (2003). There are three main components that make up an artificial neuron as displayed in Figure-2. Almost all the neurons in artificial neural networks consist of these components without immunity, Weighting Factors, Summation function and Transfer Function. (Jianhua Chen, 2005).

An Artificial Neural Network Model:

An artificial neural network is produce by mapping $øNN$: → it can also be written in the form as

$g(y1, \ldots, ym) = øNN(X1, \ldots xn)$

Here input vector is $x = (x1,\ldots, xn)\ T$ along with output vector is $y= (y1,\ldots,ym)T$. When weighted sum ∑ will be greater than zero a particular neuron will be fired. for neurons to ire , is that threshold level. Moreover this threshold level can be built in to propagation function by weighting it $w0 = –1$. Thus, propagation function ( ) = ∑ − is actually the weighted summation of inputs. Activation function g of the neuron might assume various forms. It can either be linear function or might be a non-linear function. Usually a sigmoid function is used. All these interconnected neurons (Haykin, 1999) could be disposed through certain architecture. The network $øNN$ here the threshold values (Bishop, 1995) are included in the input vector $x = (x0, \ldots, xn)\ T$, $x0 = – 1$ along with the output vector is $y = (y1, \ldots, ym)\ T$ is represented in Figure 3 .

Multi Layer Perceptron Networks (MLP):

Neural networks where the hidden neurons have sigmoid activation function and the output neurons have sigmoid or identity function are called *Multi Layer Perceptrons* (MLP) Networks $øMLP$: → . This architecture consists of an input layer, an output layer and k-hidden layers, each containing jk neurons. Each p-component of $y = (y1, \ldots, ym)$ is released by the m-neuron at the output layer as a function of the input $x = (x1, \ldots, xn)$ and of the parameters w. Writing in compact from, with weights on the input vectors and $d – 1$ as total number of hidden layers.

= ∑ ∑ …∑ (∑) ...

Figure 3 shows the graph of a neural network $øMLP$, where $d = 3$, $n = 2$, $j1 = 4$, $j2 = 5$ and $m = 1$ or (2 – 4 – 5 – 1) MLP.

Radial Basis Function Networks (RBF):

Radial Basis Function (RBF) neurons are neurons in which the propagation function has the form $f(x)= \| x - w \|$, where $x = (x1, \ldots, xn)T$, $wr = (w1, \ldots, wn)\ T$ are the inputs and weights. The activation function $h(x)$ has the form of a radial symmetric function, commonly the Gaussians function. Networks with one hidden layer containing *r* RBF neurons and output neurons with propagation function *f*(*x*) = and identity activation function *g*(*x*) = *x* are called RBF networks *øRBF*: → with *r* RBF neurons on the hidden layer, displayed in Figure-4. Each *p*-component of the output *y* = (*y*1, …,*ym*)is given by

$$(\ ) = \sum \quad h \ (\| \quad - \quad \|)$$

The propagation function calculates how close (using the Euclidian distance) the input vector *x* is to the vector *wr*. The Gaussian activation function produces higher values for input vectors that are close to the vector *wr* and smaller values for inputs that are far away from it. Thus the weights form clusters in the input space.

Neural Networks in Volatility Estimation

Estimation of Conditional Volatilities: Neural Networks can be effectively used for estimation of the conditional volatilities (Giacomini, 2003; Eun and Resnick, 2004; Bhalla, 2008 Dhamija.A.K 2010) of the financial time series. Suppose that a time series with the stochastic volatility follows an AR (*p*)-ARCH (*p*) process of the following form

$$= \quad , \quad , \quad ,\ldots \quad , \quad ,\ldots, \quad , \quad , \quad ,\ldots \quad , \quad ,\ldots,$$

Here , is i.i.d. along with *E* ( ) = 0 and ( )= 1.

By defining

$$= \quad ,\ldots \quad , \quad ,\ldots, \quad ,$$

The AR (*p*)-ARCH (*p*) process could be written as

$$= \quad (\ ) + \quad (\ )$$

Now we can write

$$(\ ) = \ (\ ) - \quad (\ )$$

where

$$[\quad | \quad = \ ] = \quad (\ )$$

$$[\quad | \quad = \ ] = \quad (\ )$$

$$[\quad | \quad = \ ] = \quad (\ )$$

By using the neural network for approximating (*z*) along with Θ *NN* for approximating (*z*), we obtained

$$(\ ) = \quad (\ ; \ )$$

$$(\ ) = \Theta\, NN \ (\ ; \ )$$

where

$$= \quad \frac{}{}\sum \quad (\quad - \quad (\ ; \ ))2$$

$$= \;—\sum \;(\quad - \Theta\, NN \;(\; ; \;))2$$

The estimator of ( ) is obtained as

$$(\;) = (\;) - (\;)$$

This approach is used by Hardle et al. (2002) in which the residuals were substituted through the sample residuals. By approximating the residuals via the sample residuals

$$= \quad (\;)$$

Also the squared sample residuals with the neural network *NN* along with its parameters

$$—\sum \quad - \quad (\;;\;)\;2$$

So the estimation of the conditional volatility could be written as

$$(\;) = \quad (\;;\;)$$

Evaluation of Models for Prediction Accuracy:

The fundamental criterion for evaluating the performance of forecasting tools is the precision of the forecast. Conversely, the evaluation of forecast accuracy is not inconsequential, as accuracy can be measured in various ways, and the evaluation results might be inconsistent if different criteria are used. For better understanding the properties of the forecast methods, we used absolute error measures.

The Mean Absolute Error (MAE) is

$$= —$$

Mean Absolute Percentage Error (MAPE), measures the magnitude of absolute errors in relative terms.

$$\frac{\sum \quad —}{=}.$$

RESULTS AND DISCUSSIONS

Descriptive Statistics:

Table-1 shows the Daily return series of PSO stock index exhibit natural properties of financial data series. The unconditional mean is 0.025801 over the examined period. Fair to large daily fluctuation varies within the range of 9.531018 and -7.488594 (i.e. these are the maximum and minimum values respectively). The unconditional standard deviation of the PSO daily return series is 2.176264 which are corresponding to an annualized volatility (for 365 days in a year) of 41.5683. Skewness of PSO daily return series is, 0.016715, positively skewed, which shows excessive gains in return in favor of this time period. Kurtosis value 4.422164 is greater than 3, demonstrating that daily returns are thick tail. Historigram displayed in figure-5 is used for checking normality assumption. Jarque-Bera statistics for the data series is much greater than any critical value at usual confidence level, therefore rejecting the null hypothesis of normally distributed returns.

Pre-Estimation Analysis:

Pre estimation analysis is used to find out if the data is suitable according to certain assumptions and can be model using GARCH models. The main tests prior to actually estimating the conditional volatility are augmented Dicky fuller test to check stationary for the return series. Engle’s ARCH test to check hetroscedasticity and Ljung-box-pierce Q test whether volatility clustering is present. Test for normality in the raw returns is (presented in the descriptive statistic of return series) Jarque-Bera normality test.

Stationary of Return Series:

The initial step is to study the series properties in terms of stationarity. Augmented Dicky-Fuller test, (Dicky and Fuller, 1981) is used in order to test stationarity of return series. Table-2 demonstrates the consequences of this test. The optimal lag length is determined by both Schwarz Bayesian Information Criterion (SBIC) and Akaike Information Criterion (AIC). The null-Hypothesis for ADF test is that series is non stationary. Resulting p-value is much greater than any critical value at usual confidence level, consequently rejecting the null hypothesis of that daily stock return series is non-stationary.

Autocorrelation and Partial Autocorrelation:

Correlation of the return series is examined by the ACF (autocorrelation function) and PACF (partial autocorrelation function) over 20lags as identification tools. Table-3 reports the Ljung–Box–Pierce Q-test that quantifies qualitative checks for correlation present in the PSO return series up to 20 lags at the 0.01 and 0.05, level of significance. The null hypothesis for the test is that there exist no autocorrelation. Table-3 reports that there is no need to use correlation structure (ARX/ARMAX) in the conditional mean of the return series as the p-values are insignificant in all lags.

Volatility Clustering and Nonlinear Dependence in PSO Return Series:

Volatility clustering is defined as the autocorrelation in the variance of the return process. Volatility clustering can be detected by a simple method calculating first order autocorrelation function in squared return. Table-4 demonstrates ACF coefficient of the squared return series. ACF coefficient is significant for all lags.

Engle’s ARCH test of Heteroscedasticity:

Engle’s ARCH test is used to check for heteroscedasticity in returns. The null hypothesis of ARCH test is that there is no ARCH effect up to order q. Table-2 shows that for the return series ARCH effect is significant at. 01 and .05 level of significance.

Estimation of GARCH Models:

Table-5 reports the estimated coefficients along with p-values for the conditional variance equation of the return series under the normality assumption, for innovation process. Volatility of return series is estimated by using one symmetric GARCH model and two asymmetric GARCH models named EGARCH and GJR-GARCH models assuming normal distribution.

The GARCH (1, 1) parameterization is most rational specification because of the majority of significant coefficient for capturing symmetric dynamics. However, considering the significant parameter at 0.01 level of significance, GARCH (1, 2) and GARCH (2, 1) models are also used in forecast evaluation under normal distribution. EGARCH(1,1,2) model estimated under normal distribution shows that first and second lag of variance are significant at 0.01 and 0.05 level of significance. EGARCH (2, 1,1) model is not considered because its second lag is negative. GJR (1,1,1) is significant overall. While GJR (2, 1, 1) model shows second symmetric lag negative, so this is not considered for forecasting.

Evaluation of In-Sample Forecast of Volatility:

In-sample forecast of volatility is calculated through static forecasting. Forecast is evaluated through Mean Absolute Error and AMAPE. Table-6 shows that Mean absolute error is strongly in favor of GJR (1, 1, 1) model. Second best model is considered EGARCH because of least AMAPE.

Evaluation of Out-Of-Sample Forecast of Volatility:

Table-6 shows the out-of-sample forecast evaluation of volatility of the return series. Forecast is evaluated through MAE and AMAPE.

Both criteria are in favor of GJR models that these models performed better than all other considered models. Thus for stock return series GJR (1, 1, 2) is best forecasting model.

Neural Network Models:

In this section we estimate models by MLP and RBF networks through different architectures by varying the number of neurons in the hidden layer. Networks are designed through intelligent problem solver. As in time series input and output variables are same for neural networks modeling, so the input as well as output variable is definitely the return series of PSO stock index. Numbers of time steps to be used as input to the network are given 20 as PACF was significant on 20 lags. However software trained network by using 10 time steps as inputs. Look ahead parameter is to indicate how many time steps ahead of the last input case the output should be predicted. In the vast majority of cases, the Look ahead is set to 1, however the Steps varies according to the requirements of the problem domain, and the variable(s) used are all input/output.

Network Illustration Graphs:

Figure-6 displays the illustrated graphs of networks. Unit activation levels are (by default) displayed in color, red for positive activation levels, green for negative. Triangles pointing to the right indicate input neurons. These neurons perform no processing, and simply introduce the input values to the network. Squares indicate Dot Product synaptic function units (e.g. as found in Multilayer Perceptrons). Circles indicate Radial synaptic function units.

First three graphs are for multilayer perceptron networks. 10 input cases are used in all networks for introducing inputs to the network, represented in triangle shapes. In first graph two negative unit activation levels are used. In second graph one positive and one negative unit activation levels are used. And in the last network graph for multilayer perceptron used a single unit activation level.

Last three graphs are for radial basis functions networks. 10 input cases are used in all networks for introducing inputs to the network, represented in triangle shapes. In first graph three Radial synaptic function units are used in circles shapes. In second graph seven Radial synaptic function units are used. While in last graph 15 Radial synaptic function units are used. As there is insufficient space to display all the neurons in a layer, the middle ones are omitted and only 10 are displayed.

Models Summary:

Table-7 demonstrates the modes summery of models. Three networks are retained for Radial Basis function with minimum selection error (provided by K-Means, K-Nearest Neighbor and Pseudo Invert (Linear least square optimization)) and similarly three networks for Multi Layer Perception provided by Back Propagation and Conjugate Gradient Decent algorithms).

In the table the Train Performance, Select Performance and Test Performance explains the performance of the network on the subsets used during training. The performance measure depends on the type of network output variable. As the stock return series is a continuous variables the performance measure is the Standard Deviation Ratio.

Train Error/Select Error/Test Error is the error of the network on the subsets used during training. This is less interpretable than the performance measure, but is the figure actually optimized by the training algorithm (at least, for the training subset). This is the Root Mean Square of the network errors on the individual cases. Hidden 1 column describes the number of hidden units in a hidden layer. As a general rule only one hidden layer supposed to be used in MLP and RBF. All networks with lowest selection error and training errors in the run were stored.

MLP (s10 1:10-2-1:1) with 10 steps factor, 1 input variable, 10 input neurons, two hidden neurons and one hidden layer perform best and have less selection error in addition to train errors and test errors. Network was found on 42 epochs. Second best network of MLP (s10 1:10-1-1:1) with 10 steps factor, 1 input variable, 10 input neurons, 1 hidden neurons and 1 hidden layer was found on 22 epochs on the basis of minimum selection, training and testing error. Third best network with the same architecture of 1st network, MLP (s10 1:10-2-1:1) with 10 steps factor, 1 input variable, 10 input neurons, two hidden neurons and one hidden layer also performs best however the difference was that, on 24 epochs the best network with lowest selection error was stored.

Although RBF networks are also good in forecasting but may be for this data set they do not achieve good results in estimation of models seeing that MLP provided three times improved performance.

Here we are considering good network models among the three RBF networks on the basis of minimum errors in selection testing and training the networks. RBF (s10 1:10-3-1:1) with 10 steps factor, 1 input variable, 10 input neurons, 3 hidden neurons and 1 hidden layer having minimum selection error perform outstanding among RBF networks. RBF (s10 1:10-7-1:1) with 10 steps factor, 1 input variable, 10 input neurons, 7 hidden neurons and 1 hidden layer is considered second best network model according to minimum selection error. And last of all, RBF (s10 1:10-15-1:1) with 10 steps factor, 1 input variable, 10 input neurons, 15 hidden neurons and 1 hidden layer is considered third fine model according to selection error, or in other words having largest selection error among RBF networks. But according to less training error RBF (s10 1:10-15-1:1) with 10 steps factor, 1 input variable, 10 input neurons, 15 hidden neurons and 1 hidden layer performed tremendous. Furthermore according to less testing error RBF (s10 1:10-7-1:1) with 10 steps factor, 1 input variable, 10 input neurons, 7 hidden neurons and 1 hidden layer performed exceptional.

One Period Ahead Forecast:

Figure-7 shows the First graph of the one period ahead forecast of stock return series. All the six models are plotted on single graph for comparison of prediction.

Figure-8 shows the Second graph of the one period ahead forecast of stock return series which is displaying the prediction of two models MLP and RBF which have least error among all models, for comparison purpose. MLP perform better among all six models.

Forecast Evaluation of Neural Network Models:

Forecast through neural network models is evaluated on the basis of minimum mean absolute error. Table-10 presents these results.

Network found on 42 epochs i.e. MLP(s10 1:10-2-1:1) with 10 steps factor, 1 input variable, 10 input neurons, two hidden neurons and one hidden layer perform best and have less absolute mean error among all the six models of neural network. Second best model in forecasting oil price, according to mean absolute error is RBF (s10 1:10-15-1:1) with 10 steps factor, 1 input variable, 10 input neurons, 15 hidden neurons and 1 hidden layer. This also indicates that RBF models can effectively applied for forecast purpose. On third rank network model MLP (s10 1:10-2-1:1) with 10 steps factor, 1 input variable, 10 input neurons, two hidden neurons and one hidden layer found on 24 epochs forecast the time series very well. Fourth best model on the basis of MAE is MLP (s10 1:10-1-1:1) with 10 steps factor, 1 input variable, 10 input neurons, 1 hidden neurons and 1 hidden layer was found on 22 epochs. RBF (s10 1:10-3-1:1) with 10 steps factor, 1 input variable, 10 input neurons, 3 hidden neurons and 1 hidden layer along with RBF (s10 1:10-7-1:1) with 10 steps factor, 1 input variable, 10 input neurons, 7 hidden neurons and 1 hidden layer are on fifth and sixth position respectively according to minimum MAE, in forecasting oil prices.

Comparison of Forecast of Econometric Models and Neural Network Models:

Table-8 shows that Forecast is evaluated through mean absolute error. Results showed that neural network models performed better than all other considered econometric models with minimum error for forecasting volatility.

CONCLUSION

The phenomenon's of stationarity, volatility clustering, autocorrelation and heteroscedasticity are observed in PSO daily stock returns. Neural Networks do a fairly good job in forecasting volatility in return series. MLP networks do considerably better than RBF networks in this case. Conditional heteroscedastic models also can be efficiently used to predict mean and volatility of PSO daily stock returns. Within the conditional heteroscedastic models, the performance of GJR and GARCH(1,1) was better than other heteroscedastic models. The sum of the ARCH and GARCH coefficient is close to unity that indicates explosive volatility process. This implies that innovations in the conditional variance will be highly persistent. Neural Networks performance is better (around 10-15% improvement) than conditional heteroscedasticity models in forecasting oil prices. Neural network can be effectively employed to estimate the conditional volatility (besides existing methods of conditional heteroscedasticity models like GARCH, EGARCH, TGARCH).

APPENDIX'S

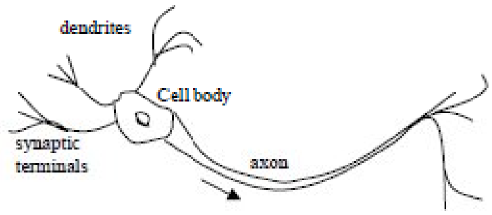


Fig. 1: (Structure of A Biological Neuron)

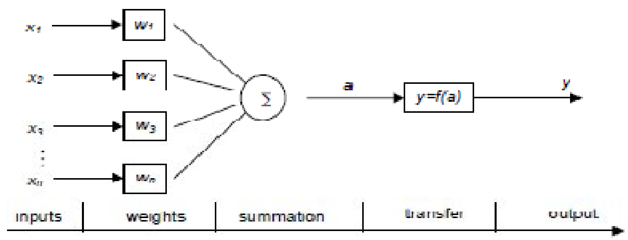


Fig. 2: (Structure of an Artificial Neuron)

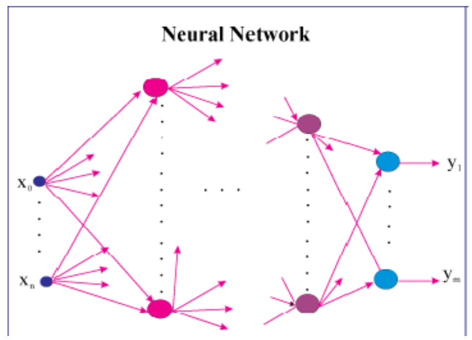


Fig. 3: (Artificial Neural Network)

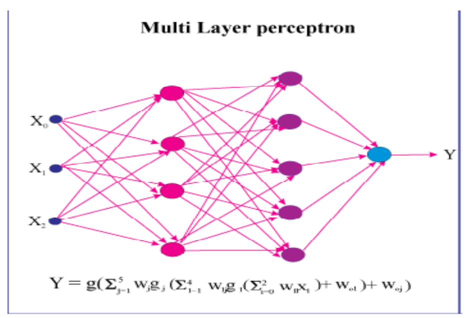


Fig. 4: (multi layer perceptron)

Table-1 (Descriptive Statistics)

Series PSO-LN-RETURNS	
Sample	1-1387
Observations 1387	
Mean	0.025801
Median	0.066494
Maximum	9.531018
Minimum	-7.488594
Std. Dev.	2.176264
Skewness	0.016715
Kurtosis	4.422164
Jarque-Bera	116.9511
Probability	0.000000

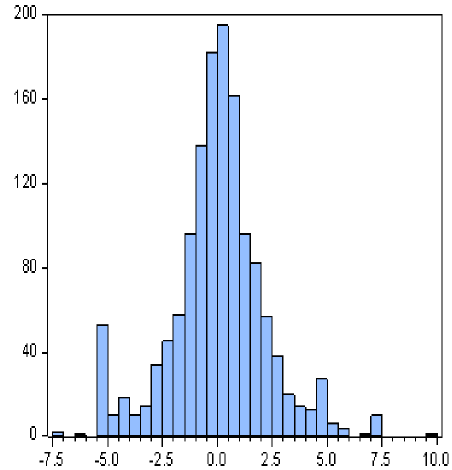


Fig. 5: (Histogram)

Table-2

Inclusion IN ADF test	ADF t-statistic		P-Values	
With constant term	**AIC**	**SBC**	**AIC**	**SBC**
	-27.86988	-36.76085	0.0000	0.0000
With constant and trend term	-27.89519	-36.78220	0.0000	0.0000
Heteroscedasticity Test: ARCH				
F-statistic	123.4862		Prob. F(2,1382)	0.0000
Obs*R-squared	209.9832		Prob. Chi-Square(2)	0.0000

Table-3

	AC	PAC	Q-Stat	Prob
1	0.011	0.011	0.1560	0.693
2	-0.067	-0.067	6.3956	0.041
3	0.011	0.013	6.5632	0.087
4	0.024	0.020	7.3831	0.117
5	-0.007	-0.005	7.4423	0.190
6	-0.018	-0.016	7.9151	0.244
7	0.026	0.025	8.8243	0.266
8	-0.039	-0.043	10.965	0.204
9	0.010	0.015	11.110	0.268
10	0.051	0.046	14.772	0.141
11	-0.011	-0.011	14.951	0.185
12	-0.013	-0.005	15.184	0.232
13	0.016	0.014	15.562	0.274
14	-0.019	-0.025	16.093	0.308
15	-0.018	-0.012	16.557	0.346
16	-0.012	-0.015	16.767	0.401
17	0.059	0.056	21.626	0.200
18	0.007	0.009	21.686	0.246
19	0.043	0.051	24.304	0.185
20	0.006	0.000	24.352	0.227

Table-4

	AC	PAC	Q-Stat	Prob
1	0.360	0.360	180.44	0.000
2	0.261	0.151	275.37	0.000
3	0.257	0.145	367.67	0.000
4	0.239	0.103	446.96	0.000
5	0.262	0.129	542.87	0.000
6	0.214	0.046	606.90	0.000
7	0.211	0.062	669.04	0.000
8	0.206	0.051	728.31	0.000
9	0.131	-0.038	752.43	0.000
10	0.218	0.110	818.81	0.000
11	0.180	0.018	864.05	0.000
12	0.186	0.051	912.76	0.000
13	0.142	-0.018	941.12	0.000
14	0.166	0.051	979.69	0.000
15	0.176	0.032	1023.0	0.000
16	0.134	-0.010	1048.2	0.000
17	0.124	-0.004	1069.8	0.000
18	0.152	0.035	1102.5	0.000
19	0.192	0.088	1154.2	0.000
20	0.170	0.019	1194.9	0.000

Table-5

model	μ						
GARCH(1,1)	0.078847 (0.1067)	0.175903 (0.0000)	0.136255 (0.0000)		0.823323 (0.0000)		
GARCH (1,2)	0.063336 (0.1941)	0.233010 (0.0000)	0.191850 (0.0000)		0.239324 (0.0165)	0.514675 (0.0000)	
GARCH(2,1)	0.067287 (0.1628)	0.082477 (0.0000)	0.254986 (0.0000)	-0.174225 (0.0001)	0.900190 (0.0000)		
EGARCH(1,1,1)	0.033356 (0.4853)	-0.119240 (0.0000)	0.268919 (0.0000)		-0.016391 (0.3270)		0.935795 (0.0000)
EGARCH(1,1,2)	0.016917 (0.7321)	-0.174595 (0.0000)	0.366343 (0.0000)		-0.003772 (0.8688)	0.399677 (0.0000)	0.521360 (0.0000)
EGARCH(2,1,1)	0.018122 (0.7088)	-0.100937 (0.0000)	0.413175 (0.0000)	-0.206237 (0.0002)	-0.007887 (0.5724)		0.957335 (0.0000)
GJR(1,1,1)	-0.00044 (0.9920)	0.19932 (0.0022)	0.11980 (0.0001)		0.81139 (0.0000)		0.04524 (0.3005)
GJR(1,1,2)	-0.01270 (0.7722)	0.25493 (0.0034)	0.17226 (0.0002)		0.26238 (0.0881)	0.48221 (0.0003)	0.04747 (0.4131)
GJR(2,1,1)	-0.00598 (0.8884)	0.08878 (0.2248)	0.24430 (0.0000)	-0.16565 (0.0001)	0.89576 (0.0000)		0.00990 (0.7574)

Table-6

In-Sample forecast evaluation			Out-of-sample forecast evaluation		
Model	**MAE**	**AMAPE**	**Models**	**MAE**	**AMAPE**
GARCH(1,1)	4.4487	59.05	**GARCH(1,1)**	6.6271	57.30
GARCH (1,2)	4.4290	59.06	**GARCH (1,2)**	6.6045	57.21
EGARCH(1,1,1)	4.4470	59.04	**EGARCH(1,1,1)**	6.6673	57.47
EGARCH(1,1,2)	4.4387	53.13	**EGARCH(1,1,2)**	6.6475	57.37
GJR(1,1,1)	4.2443	58.98	**GJR(1,1,1)**	6.6032	57.18
GJR(1,1,2)	4.4063	58.96	**GJR(1,1,2)**	6.5895	57.12

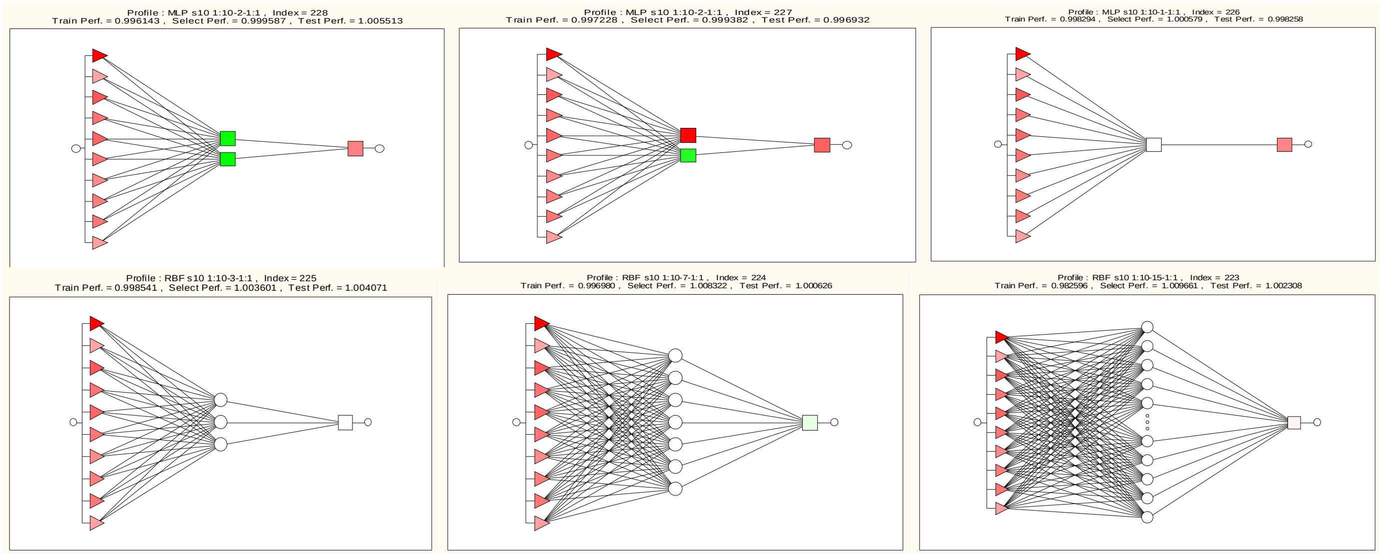


Fig. 6: (Network Illustration Graph)

Table-7

MODELS	Training members	Train Perf.	Train Error	Select Perf.	Select Error	Test Perf.	Test Error	Hidden(1)	Abs E. Mean
RBF s10 1:10-15-1:1	KM,KN,PI	0.982596	0.439277	1.009661	0.423026	1.002308	0.415598	15	1.551569
RBF s10 1:10-7-1:1	KM,KN,PI	0.996980	0.445707	1.008322	0.422471	1.000626	0.415324	7	1.563675
RBF s10 1:10-3-1:1	KM,KN,PI	0.998541	0.446405	1.003601	0.420472	1.004071	0.416916	3	1.560688
MLP s10 1:10-1-1:1	BP100,CG 22b	0.998294	0.131109	1.000579	0.122973	0.998258	0.122399	1	1.555957
MLP s10 1:10-2-1:1	BP100,CG 42b	0.997228	0.130874	0.999382	0.122895	0.996932	0.121510	2	1.552521
MLP s10 1:10-2-1:1	BP100,CG 24b	0.996143	0.130805	0.999587	0.122853	1.005513	0.123127	2	1.548577

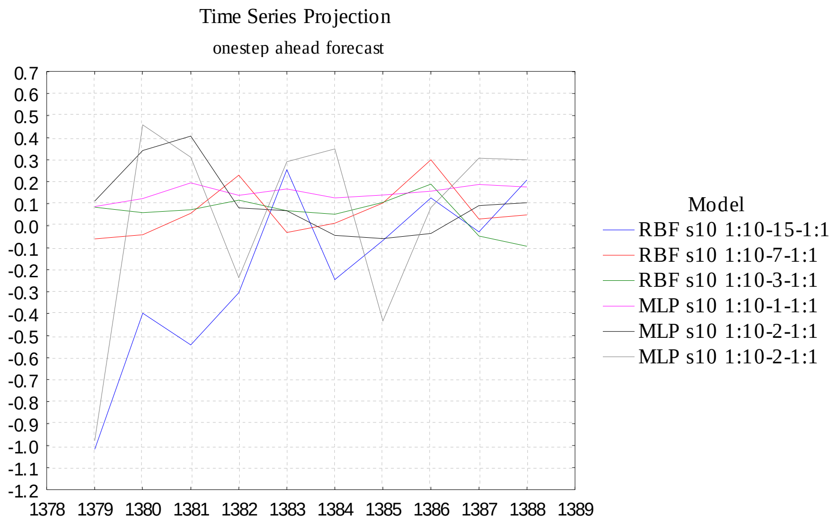


Fig. 7: (Time Series Projection)

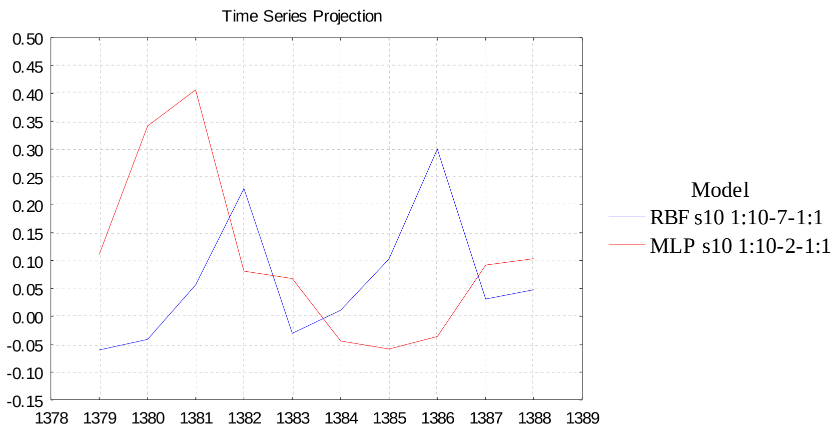


Fig. 8: (Time Series Projection)

Table-8

MODELS	MAE	MODELS	MAE
GARCH(1,1)	4.4487	RBF s10 1:10-15-1:1	1.551569
GARCH (1,2)	4.4290	RBF s10 1:10-7-1:1	1.563675
EGARCH(1,1,1)	4.4470	RBF s10 1:10-3-1:1	1.560688
EGARCH(1,1,2)	4.4387	MLP s10 1:10-1-1:1	1.555957
GJR(1,1,1)	4.2443	MLP s10 1:10-2-1:1	1.552521
GJR(1,1,2)	4.4063	MLP s10 1:10-2-1:1	1.548577

REFERENCES

1. Altman, I.E., Marco, G. and Varetto, F. (1994). Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (the Italian Experience). *Journal of Banking and Finance*, 18, 505-529.
2. Binner, J.M., Bissoondeeal, R.K., Elger, T., Gazely, A.M. and Mullineux, A.W. (2002). *A Comparison of Linear Forecasting Models and Neural Networks: An Application to Euro inflation and Euro Divisia*. 1-29.
3. Balkin, S.D. and Ord, J.K. (2000) Automatic neural network modeling for univariate time series, *International Journal of Forecasting*, 16, 509-15.
4. Cheng, B. and Titterington, D.M. (1994). Neural Networks: A Review from a Statistical Perspective. *Statistical Science*, 9(1), 2-54.
5. Dhamija, A.K. and Bhalla, V.K. (2010). Financial Time Series Forecasting: Comparison of Neural Networks and ARCH Models. *International Research Journal of Finance and Economics*, 49, 185-202.
6. Duzgun, R. (2008). A comparison of artificial neural networks and ARIMA models success in GDP forecast. *I.I.B.F Dergisi*, 25(2), 165-176.
7. Erba, M.E., Pratt, J.E., Aplin, R.D. and Stephenson, M.W. (1996). Comparison of Costs and Efficiencies between Cooperative, Proprietary, and Captive Fluid Milk Processors: A Neural Network Approach. *Work Paper of the Department of Agricultural, Resource and Managerial Economics, Cornell University.*
8. Fausett, L. (1994). Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. *Englewood Cliffs, NJ: Prentice Hall*.
9. Haykin, S. (1999). Neural Networks: A Comprehensive Foundation. *Prentice Hall*.
10. Hagan, M., Demuth, H.B. and Beale, M. (1996). *Neural Network Design*. Boston: PWS Publishing Company.
11. Fahimifard, S.M., Homayounifar, M., Sabouhi, M. and Moghaddamnia, A.R. (2009). Comparison of ANFIS, ANN, GARCH and ARIMA techniques to Exchange rate forecasting. *Journal of Applied Sciences*, 9(20), 3641-3651.
12. Hamid, S.A. (2004). *Primer on Using Neural Networks for Forecasting Market Variables*. Working Paper No. 2004-03.
13. Joerding, W.H., Li, Y. and Young, D.L. (1994). Feed forward Neural Network Estimation of a Crop Yield Response Function. *J. Agri. and Applied Econ.*, 26(1), 252-63.
14. Jordan, M.L. (1995). Why the logistic function? A tutorial discussion on probabilities and neural networks. *Technical Report of Computational Cognitive Science*, Massachusetts Institute of Technology.
15. Kastens, L.T., Featherstonem, A.M. and Biere, A.W. (1995). A Neural Networks Primer for Agricultural Economists. *Agricultural Finance Review*, 55, 54-73.
16. Li, Y. (1995). Macroeconomics modeling on UK GDP growth by neural computing. Technical Report CSC.
17. Refenes, N.A., Zapranis, A. and Francis, G. (1994). Stock Performance Modeling Using Neural Networks: A Comparative Study with Regression Models. *Neural Networks*, 7(2), 375-388.
18. Shachmurove, Y. (2004). Applying Artificial Neural Networks to Business, Economics and Finance. *Work paper of Department of Economics, The University of Pennsylvania.*

19. Shachmurove, Y. and Witkowska, D. (2000). Utilizing Artificial Neural Network Model to Predict Stock Markets. CARESS Working Paper #11.
20. Swanson, N.R. and White, H. (1995). A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. *Review of Economics and Statistics*, 79, 540-550.
21. Verleger, Philip K. (1993). *Adjusting to Volatile Energy Prices*, Institute for International Economics, Washington DC.
22. Zurada, M. Jacek (1992). Introduction to Artificial Neural Systems. West Publishing Company, New York.
23. Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 157-175.
24. Oil Price History and Analysis (1998). *WTRG Economics Website*.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 255-268

A STUDY ON ACADEMIC PERFORMANCE OF UNIVERSITY STUDENTS

Erum Shahzadi[1] and **Zahoor Ahmad**[2]
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: [1]erum_stat@yahoo.com
[2]zahoor.ahmed@uog.edu.pk

ABSTRACT

Earlier research has focused to explore the factors that are related to the academic performance of university students [Hijazi and Naqvi (2006); Vandamme, et al. (2005); Cheesman, et al. (2006)]. In this paper we want to develop a model for academic performance of students of university of Gujrat, Gujrat, Pakistan. To develop this model we consider the independent factors like home environment, study habits, hardworking, learning skills, and academic interaction. For the development of model using Structure equation modeling, first of all we use Confirmatory factor analysis to confirm the considered factors. For this study the population was all students of social sciences and now studying in 4th semester and who enrolled in 2007 and we select a sample of 300 students using stratified sampling with proportional allocation. The fitted model shows that academic performance depends on learning skills and learning skills depends on home environment. Also academic performance depends on academic interaction and academic interaction depends on study habits and home environment. It means academic performance can be estimated for any student by its home environment and learning skills and also by its academic interaction, study habits, and home environment. By examining the three possible paths of estimating academic performance, the strongest path is the home environment which affects the learning skills and ultimately learning skills lead to affect the academic performance. According to our model students can achieve high academic performance by focus on home environment and learning skills.

KEYWORDS

Academic Performance, Home Environment, Study Habits, Learning Skills, Academic Interaction, Confirmatory Factor Analysis, Structural Equation Modeling.

1. INTRODUCTION

Student's academic performance and graduation rates have been the area of interest for higher education institutions. Investigation of factors related to the academic performance of university students become a topic of growing interest in higher educational circle. Many recent studies were carried out to explore factors that affecting university student's academic performance. Hanson (2000) reported that Student performance is affected by different factors such as learning abilities, gender and race. Simmons, et al. (2005) concluded that family income level, attending full time, receiving grant aid and completing advanced level classes in high school having statistically significant effects on college persistence among first generation college students. Garton, et al. (2000) carried out a study with freshmen college students to evaluate the efficiency of student learning style and other university admission variable in predicting student academic performance and retention. Act composite score, high school class rank, high school core GPA, and learning style were used as predictors. Results showed that core GPA and Act score were best predictors for predicting academic performance of first year of college. Mckenzie and Schweitzer (2001) conducted a prospective study to explore the psychosocial, cognitive, and demographic predictors of academic performance of first year Australian university students. Results demonstrate that previous academic performance was identified most significant predictors of university performance. Integration into university, self efficacy, and employment responsibilities were also predictors of university performance. Hijazi and Naqvi (2006) conducted a study to find out the factors which affecting college students' performance. In this study researcher mainly focus to explore the factors that associated with performance of students in intermediate examination. This study conclude that attitude towards attendance in classes, time allocation for studies, parents level of income, mother's age and mother's education were main factors that affect performance of students of private colleges.

There are numerous factors which affect the academic performance and retention of students in higher education institutions. We discuss those important factors which we used in this study. The justification of the factors with existing literature is given below.

1.1.1 Home Environment:

Reviewed literature indicated that there is an awareness of the importance of the home environment or family on pupil's/students academic performance. The home has a great influence on the students' psychological, emotional, social and economic state. In the view of Ajila and Olutola (2007), the state of the home affects the individual since the parents are the first socializing agents in an individual's life. This is because the family background and context of a child affect his reaction to life situations and his level of performance.

Ichado (1998) stated that parent's constant disagreement affects children emotionally and this could lead to poor academic performance. Taylor, et al. (1995) showed that parenting style (nature and control) and parental involvement significantly predicted academic outcomes. In Saudi Arabia, Kritam, et al. (2004), reported that the family financial support, encouragement and following up have positive impact on students' performance as measured by their GPA.

1.1.2 Study Habits:

Study habits of students may be relevant to the prediction of grades because it is possible that student's grades may be related to their study habits. That is, students with poor study habits may obtain lower grades than those students with better study habits. The importance of the relationship between grades, instructor ratings and study habits has not been determined [Middleton (1979)].

Study skills and learning approaches include, for example, time management, using information resources, taking class notes, communicating with teachers, preparing for and taking examination, and several other learning strategies. The research shows a significant correlation between such learning behavior and approaches and academic achievement in higher education [Soares, et al. (2009)]. Students who create their own study aids are spending time making them, whereas those who use others' study aids or not. It may also be that the process of creating study aids helps the learner gain more meaningful knowledge through the process of synthesizing disparate pieces of information into new knowledge, as has been shown with note taking. We wondered if students who used study aids made by others rather than making their own might be missing out on the benefits of time-on-task and concept mapping [Sleight and Mavis (2006)].

Estes and Richards (1985) developed a survey of study habits for use with high school and college students. Their study skills index measured three factors for both homework and test situations. Distractibility items assess the degree to which students report being unable to maintain their attention or concentrate on their task. Inquisitiveness items measure how well students try to make sense of the material they are studying- do they look for essential concepts or deeper meaning? Compulsiveness items assess the degree to which students attend to details and try to remember facts.

1.1.3 Learning Skills:

Recent research has considered student behavior and learning to be important factors in student's academic success and retention. Hattie, et al. (1996) conclude that if we aim to increase student's academic success in higher education institutions, we must focus on interventions directed towards learning strategies, a fact which suggests the need to develop programs of this kind [Soares, et al. (2009)]. The influence of learning strategies on academic achievement, on the other hand, has been much less widely investigated, in spite of its theoretical importance and prevalence in international reports [Martin, et al. (2008)]. In 1998, Jere Brophy demonstrated that increased time spent on learning activities yields increased learning, provided that the teacher was competent and that the learning activities were effectively designed and implemented. Another theory that guided us was concept mapping. Concept mapping is a method in which the learner links new knowledge to a framework of relevant concepts that the learner already knows. Ausubel (1963) maintained that this linking of new with existing knowledge was a key factor in successful learning and that it was the difference between meaningful learning and rote learning [Sleight and Mavis (2006)].

1.1.4 Academic Interaction:

Research on college students suggests that activities like advising could increase students' involvement in their college experiences. Colleges and universities could use strategic planning to design advising programs based on relationships of shared responsibility and focused on students' success. Research on positive outcomes of college and on the diverse needs of students making up today's student population suggests that a new look at advising is needed. Findings link academic advising directly and indirectly to contact between faculty and students and persistence in college. For example, involvement influences learning and defines effective institutions as those having the capacity to involve students [Astin (1984)]. Research also indicates that frequent and meaningful contact with faculty members, especially contact focusing on intellectual or career-related issues, seems to increase students' involvement and motivation [Astin (1984); Pascarella (1980, 1985); Terenzini, Pascarella, and Lorang (1982); Tinto (1987)]. These results can be important to advisers, for they have the capacity to increase meaningful contact with students and to encourage them to persist in college. When a broad base of the college community plans for, implements, and evaluates advising services, advising can become a systematic enterprise of the institution that enhances the educational outcomes of college. Another very important factor in establishing high retention rates at a college is the degree to which students establish close and supportive personal and professional relationships with faculty and other significant people on campus [Tinto (1987)].

2. MATERIAL AND METHODS

2.1 Population:

Population of study consisted of all students of social sciences and now studying in 4^{th} semester and who enrolled in 2007. Size of target population in this study is 708 students.

2.2 Sample Selection:

The students of social sciences (Statistics, Sociology, CSIT, Business Administration, English) are not homogeneous with respect to academic performance across disciplines and programs (BS and MA/Msc). We have used stratified random sampling with proportional allocation method to select a sample.

We Calculated the Sample size using Yamane (1967) as:

$$n = \frac{N}{1 + Ne^2}$$

where n and N are sample and population size respectively and 'e' is margin of error. Let the e = 0.04 and N = 708 then our required sample size is 300.

2.3 Research Instrument:

Questionnaire is used for data collection. First part of the questionnaire is designed to obtain information on the demographic characteristics of university students, like gender, age, region, family system, profession of father of respondent. Next part designed to obtain information on some quantitative variables related to student performance. Then there are 39 items that consist of a combination of two categories nominal items, and 37 items using a 5-point Likert-Scale. Items are designed to assess six dimensions associated with student academic performance. (Previous Achievements, Home Environment, Study Habits, Learning Skills, Hardworking, Academic Interaction).

2.4 Data Analysis Techniques:

2.4.1 Confirmatory Factor Analysis:

CFA is used to provide a confirmatory test of our measurement theory. A measurement theory specifies how measured variables logically and systematically represent constructs involved in a theoretical model. In confirmatory factor analysis (CFA), theory is a systematic set of casual relationships that provide the comprehensive explanation of a phenomenon. In confirmatory factor analysis (CFA), model is a specified set of dependant relationships that can be used to test the theory. In confirmatory factor analysis (CFA), is used to test structural equations. The path diagram shows the graphical representation of cause and effect relationships of the theory. In confirmatory factor analysis (CFA), endogenous variables are the resulting variables that are a causal relationship and exogenous variables are the predictor variables.

In confirmatory factor analysis (CFA), identification is used to test whether or not there are a sufficient number of equations to solve the unknown coefficient. In confirmatory factor analysis (CFA) identifications are of three types: (1) under identified, (2) exact identified, and (3) over-identified. In confirmatory factor analysis (CFA), goodness of fit is the degree to which the observed input matrix is predicted by the estimated model.

2.4.2 Structural Equation Modeling:

Structural equation modeling (SEM) is a relatively new analytical tool, but its roots extend back to the first half of the twentieth century. During the late 1960s and early 1970s, the work of Joreskog and Sorbom led to simultaneous maximum likelihood estimation of the relationship between constructs and measured indicator variables as well as among latent constructs.

Structural equation modeling (SEM) is a family of statistical models that seek to explain the relationship among multiple variables. In doing so, it examines the structure of interrelationships expressed in a series of equations, similar to a series of multiple regression equations. These equations depict all of the relationships among constructs (the dependent and independent variables) involved in the analysis. Constructs are unobservable or latent factors represented by a multiple variables (much like variables representing a factor in factor analysis). SEM can be thought of as a unique combination of both types of techniques (interdependence, dependence) because SEM's foundation lies in two familiar multivariate techniques: factor analysis and multiple regression analysis. SEM is the only multivariate technique that allows the simultaneous estimation of multiple equations.

With large sample sizes, the χ^2 test statistic is known to always reject in any formal test of significance (Byrne, 1998). Hilton et al. (2004), more focused on the Root Mean Square Error of Approximation (RMSEA), the Goodness-of-Fit Index (GFI), the Non-Normed Fit Index (NNFI; termed Tucker-Lewis Index or TLI in Marsh & Yeung, 1996), the Comparative Fit Index (CFI) and the Relative Fit Index (RFI, termed Relative Noncentrality Index or RNI in Marsh & Yeung, 1996), and the normed version of the χ^2 test statistic: $\chi^2/d.f$. For the last index, no clear-cut guidelines exist; values in the range of 2.0 to 5.0 are acceptable, with lower values indicating better fit. For RMSEA, values $\leq$ 0.05 indicate good fit, values $\leq$ 0.08 indicate reasonable fit. The indices GFI, NNFI, CFI, and RFI, all normally lie in the range 0.0 – 1.0, with higher values indicating better fit. As a benchmark for good fit, the value 0.90 is often used (Kline, 2005).

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics:

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summary statistics about different variables. Descriptive statistics of all the variables are given in table1 to table 3. Table1 contains the summary statistics (minimum, maximum, mean and SD) of all quantitative variables. It shows that the average age of students is 20.93 with 1.538 SD. Average father income is Rs. 31738/- with 23411.566 SD, indicating that there is a lot of variation in father income of students it is so because the university located in Gujrat and more the parents of students are working abroad and also gujrat is an industrial area. Mean CGPA after two semesters is 2.88 with SD 0.49. Average study time of the students is 145 minutes with 80.068 S.D.

Table-2 contains the percentages of nominal scale variables. Results show that percentage of female respondents (68.7%) is high as compared to male respondents (31.3%). On the basis of these results we can say mostly respondent win educational prizes but scholarship is not offered to them. 19.7% students leave the university before receiving a degree because they accept a good job. 9.7% leave because of financial problems, 17% leave because they get married, 11% leave due to lack of interest, 8.7% leave because lack of academic ability. 34% leave due to some other reasons.

Table-3 depicts the percentages of ordinal scale variables. Table-3 shows that average rank of respondents on the statement that previous degree marks greatly influence your current academic abilities is 3.72, which is close to 4, it means on the average respondents are agree with that statement. Average rank of the respondents on the statement that previous degree marks really reflect what you can do is 3.62, it means on the average respondents are agree on the above statement. Average rank of respondent on the statement that your home environment supports you to enhance your academic abilities is 4.34, it means on the average respondents are agree with that statement. Average rank of the respondents on the statement that your family provides you facilities which are required for attaining your educational goals is 4.52, it means on the average respondents are Strongly agree with that statement. Average rank of respondents on the statement that your family encourages you on your academic achievement is 4.50, it means on the average respondents are strongly agree with that statement. In the light of above results about the home environment factor that home environment supports the students to enhance their academic performance. Average rank of respondents on the statement that for getting a good grade to organize your time and to set aside time each day for studying is important is 3.98, it means on the average respondents are agree with that statement. Average rank of respondents on the statement that you schedule definite study times and outline specific goals for your study time is 3.53, it means respondents are agree with that statement. Average rank of respondents on the statement that you avoid activities which tend to interfere with your planned schedule of study is 3.24, it means on the average respondents are neutral with that statement. Average rank of respondents on the statement that I am confident with the level of concentration, I am able to maintain in study is 3.73, it means on the average respondents are agree with that statement. Average rank of the respondents on the statement that you take notes in class, refine and study them soon after class, and review them frequently is 3.34, it means on the average respondents are neutral on the above statement. Table-3 also contains the percentages and average rank of other variables; which can be interpreted in similar manner.

3.2 Confirmatory Factor Analysis:

Confirmatory factor analysis is a special type of factor analysis and is the first part of a complete set of a structural equation model. We confirm all the factors which we considered such as Study Habits, Learning Skill, Academic Interaction, Academic Performance and Home Environment by using confirmatory factor analysis.

Table-4 shows the model estimates of confirmatory factor analysis of all the factors. Goodness of fit measures for all the factors is given in table-5. Table-4 contains the model estimates of confirmatory factor analysis of Study Habits. p-values of all the items are significant so we reject the null hypothesis that all items are not confirm for that factor. So we conclude that all items of that factor are confirmed for that factor. Individual parameter estimates exhibited that proper study time has a parameter estimate value of 1.105, which is high as compared to other variables in the factor; it means that variable is most important for the factor. In other words giving proper time to studies is an essential variable for the development of student's study habits. Concentration level has a parameter estimate value of .421, which is low as compared to other variables in the factor; it means that variable is less important for the factor. Goodness of Fit measures is used to assess the model fitness. Almost all goodness of fit measures meets the recommendation level for this factor. So, goodness of fit measures supports our estimated model.

Table-4 also contains the model estimates of other factors; which can be interpreted in similar manner. After Confirmatory Factor Analysis our next step is to fit the structure equation model on those factors (including items) that are confirmed by Confirmatory Factor Analysis.

3.3 Structure Equation Modeling:

In this paper, we use Structure Equation Modeling to develop the academic performance model of students of social sciences at University of Gujrat. We use Home environment, Study Habits, Learning Skills, Hardworking, Academic Interaction, and Academic Performance as a constructs.

A structure model involves specifying Structural relationships between latent constructs. Table-6 contains the Parameter estimates of Structure equation model. p-values of all the parameters are significant so we reject the null hypothesis that the coefficients are zero. So we conclude that all relations are significant and positive.

Goodness of Fit measures of is used to assess the fitness of structure equation model. p-value of Chi-Square test is significant. So our model is fit. Recommended value of ($\chi^2/d.f$) is less than 3. In this case, the value of ($\chi^2/d.f$) is 2.42 that is less than 3. So it also supports our estimated model. In this case, GFI value is .94 the value of AGFI is .90, which supports our estimated model. In this model, the value of RMSEA is .06 that is less than .08. So RMSEA is supported to fitted model. All the important Goodness of fit measures indicates that our estimated model is best fitted.

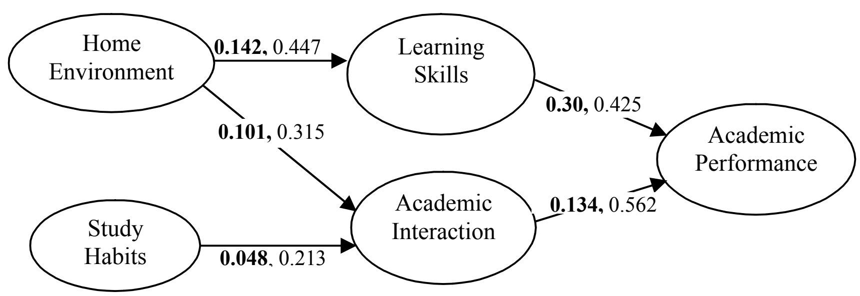


Fig. 1: Fitted Structural Equation Model

The fitted possible equations for estimating academic performance form figure 1 can be written as:

Academic Performance = .447 (Home environment) + .425 (Learning skills)
Academic Performance = .315 (Home environment) + .562 (Academic interaction)
Academic Performance = .213 (Study habits) + .562 (Academic interaction)

Form figure 1, the fitted model shows that academic performance depends on learning skills and learning skills depends on home environment. Also academic performance depends on academic interaction and academic interaction depends on study habits and home environment. It means academic performance can be estimated for any student by its home environment and learning skills and also by its academic interaction, study habits, and home environment. By examining the three possible paths of estimating academic performance, the strongest path is the home environment which affects the learning skills and ultimately learning skills lead to affect the academic performance.

3.4 Conclusions

In this paper, we determine the different factors that are related to Academic performance and develop an academic performance model of social science students at University of Gujrat by using Structure equation modeling. They should do efforts to take initiatives in academic activities like (Presentation, Quiz and Assignments) and up-to-date their self with academic matters (Course objectives, Course outlines, Week plan), then they can enhance their academic Performance.

In the light of these results we can say that home environment is more important construct for student academic performance at university level. The contribution of previous achievement and Home environment to the academic performance of students is parallel with decades of research on the importance of these factors for university student's academic performance. Rollins and Thomas (1979) found that high parental control were associated with high achievement. Religiosity as an aspect of the family environment is another independent variable possibly influencing academic achievement [Bahr, et al. (1993)]. Cassidy and Lynn (1991) investigated how family environment impacts motivation and achievement. Furthermore results indicated that motivation served as a mediating variable between home background, personal characteristics, and educational attainment.

Since most of the literature emphasizes the importance of faculty-student out-of-classroom interaction [Astin, (1993); Milem and Burger, (1997); Pascarella, (1980); Pascarella and Terenzini, (2005)]. In our research work, results demonstrate that academic interaction with teachers and peers have significant affect on the academic performance of them. We can also conclude from descriptive statistics that students must improve their Study Habits especially give proper time to study per day and revise the lecture notes daily. The present results are supported by the earlier findings by Aluja and Blanch (2004).Moreover; the findings of our research work brings to a close that the better study habits, the higher the academic interaction which ultimately lead to higher academic performance.

For high Academic Performance, we suggest that students basically improve their Home Environment and Study Habits like give proper study time, proper revision of lecture notes, avoid activities interfere in planned schedule of study, then their Learning Skills like especially presentation skills, reading the material related to course content besides the lecture notes and Academic Interaction especially academic discussions with other class fellows and teachers.

REFERENCES

1. Ajila, C. and Olutola, A. (2007). Impact of parents' socio-economic status on university students' academic performance. *Ife Journal of Educational Studies,* 7(1), 31-39.
2. Aluja, A. and Blanch, A. (2004). Socialized Personality, Scholastic Aptitudes, Study Habits, and Academic Achievement: Exploring the Link. European. *Journal of Psychological Assessment*, 20.
3. Astin, A.W. (1993). *What matters in college: Four critical years revisited*. San Francisco, CA: Jossey Bass.
4. Astin, A.W. (1984). Student Involvement: A Developmental Theory for Higher Education. *Journal of College Student Personnel*, 25, 298-307.
5. Ausubel, D. (1963). *The Psychology of meaningful verbal learning*. New York: Grune and Stratton.
6. Bahr, S.J., Hawks, R.D. and Wang, G. (1993). Family and religious Influences on adolescent substance abuse. *Youth and Society,* 24.
7. Brophy, Jere (1998). *Motivating students to learn.* New York: McGraw-Hill.
8. Byrne, B.M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.
9. Cassidy, T. and Lynn, R. (1991). Achievement motivation, educational attainment, cycles of disadvantage and social competence: Some longitudinal data. *British Journal of Educational Psychology*, 61, 1-12.
10. Cheesman, J., Simpson, N. and Wint, A.G. (2006). *Determinants of Student Performance at University: Reflections from the Caribbean.* Unpublished Manuscript.
11. Estes, T.H. and Richards, H.C. (1985). Habits of Study and Test Performance. *Journal of Reading Behavior*, 17, 1-13.
12. Garton, B.L., Dyer, J.E. and King, B.O. (2000). The Use of Learning Styles and Admission Criteria in Predicting Academic Performance and Retention of College Freshmen. *Journal of Agriculture Education*, 41(2), 46-53.
13. Hanson, J.B. (2000). *Student performance and student growth as measure of success: A evaluator's perspective.* Paper presented at annual meeting of the American educational research association New Orleans. Louisiana. 25 April.
14. Hattie, J., Biggs, J. and Purdie, N. (1996). Effects of Learning Skills Interventions on student learning: A Meta-analysis. *Review of Educational Research*, 66, 99-136.
15. Hijazi, S.T. and Naqvi, S.M.M.R. (2006). Factor Affecting Students' Performance: A case of Private Colleges. *Bangladesh e-journal of Sociology*, 3(1).
16. Hilton, S.C., Schau, C. and Olsen, J.A. (2004). Survey of Attitudes Toward Statistics: Factor structure invariance by gender and by administration time. *Structural Equation Modeling*, 11(1), 92-109.

17. Ichado, S.M. (1998). Impact of broken home on Academic Performance of Secondary School Students in English Language. *Journal of Research in Counseling Psychology,* 4(1), 84- 87.
18. Kline, R.B. (2005). *Principles and practice of structural equation modeling* (2nd Ed.). New York: Guilford Press.
19. Kritam, Abou Rakbah and Al-Awassawi (2004). *Factor Impact on Students' Motivation*. Published in the Second International Conference on Administrative Science Proceedings. King Fahd University of Petroleum and Minerals-Saudi Arabia.
20. Marsh, H.W., Hau, K.T., Balla, J.R. and Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33, 181-220.
21. Martin, E., Arias-Martinez, R., Marchesi, A. and Perez, E.M. (2008). Variables that Predict Academic Achievement in the Spanish Compulsory Secondary Educational System: A Longitudinal, Multi-Level Analysis. *The Spanish Journal of Psychology*, 11(2), 400-413.
22. Mckenzie, K. and Schweitzer, R. (2001). Who succeeds at University? Factors Predicting Academic Performance in First Australian University Students. *Higher Education Research and Development*, 20(1), 21-33.
23. Middleton, M.A. (1979). *An Exploratory Study of the Relationship of Study Habits, Expected Grade and Instructor Effectiveness Ratings to Academic Performance*. Simon Fraser University.
24. Milem, J.F. and Berger, J.B. (1997). A modified model of college student persistence: Exploring the relationship between Astin.s theory of involvement and Tinto.s theory of student departure. *Journal of College Student Development*, 38 (4), 387-399.
25. Pascarella, E.T. (1980). Student-faculty informal contact and college outcomes. *Review of Educational Research*, 50 (4), 545-595.
26. Pascarella, E.T. and Terenzini, P.T. (2005). *How college affects students: A third decade of research.* San Francisco, CA: Jossey-Bass.
27. Pascarella, E.T. (1980). Student-Faculty Informal Contact and College Outcomes. *Review of Educational Research,* 50, 545-95.
28. Pascarella, E.T. (1985). *College Environmental Influences on Learning and Cognitive Development*. In Higher Education: Handbook of Theory and Research, Vol 4, edited by J.C. Smart. New York: Agathon Press.
29. Rollins, B.C. and Thomas, D.L. (1979). Parental Support, Power and Control Techniques in the Socialization of Children. *Contemporary Theories about the Family*, 1, 317- 364.
30. Simmons, A.B., Musoba, G.D. and Choong, C.G. (2005). *Persistence Among First-Generation College Students in Indiana: The Impact of Precollege Preparation, College Experiences, and Financial Aid.* Indian Project on Academic Success (IPAS) Research Report # 05-01.
31. Sleight, A.D. and Mavis, E.B. (2006). Study Skills and Academic Performance among Second-Year Medical Students in Problem-Based Learning. *Med Educ*.
32. Soares, A.P., Guisande, M.A., Almeida, L.S. and Paramo, M.F. (2009). Academic Achievement in first-year Portuguese College Students: The role of Academic Preparation and Learning Strategies. *International Journal of Psychology*, 44(3), 204-212.

33. Taylor, L.C., Hinton, I.D. and Wilson, M.M. (1995). Parental influences of academic performance in African-American Students. *Journal of Child and Family Studies*, 4, 293-302.
34. Terenzini, P.T., Pascarella, E.T., Pascarella, and Lorang, W.G. (1982). An Assessment of the Academic and Social Influences on Freshman Year Educational Outcomes. *Review of Higher Education,* 5, 86-109.
35. Vandamme, J.P., Meskens, N. and Superby, J.F. (2005). *Fresher's Achievement: Prediction Methods and Influencing Factors*. Higher Education, Multijuridictionality and Golbalisation Conference, Ateliers des Fucam, Mons, Belgium. Dec 16-17.
36. Yamane, T. (1967). *Statistics, An Introductory Analysis*, 2nd Ed., New York: Harper and Row.

APPENDIX

Table 1: **Descriptive Statistics of Quantitative Variables**

Variables	Minimum	Maximum	Mean	Std.Deviation
Age in years	18	28	20.93	1.538
Father income	3000	200000	31738	23411.566
CGPA at the end of second semester	1.58	3.89	2.8842	0.49566
Study time in minutes	0	360	144.88	80.068

Table 2: Percentages of Nominal Scale Variables

Gender		**Discipline**	
Male	Female	Business Administration	31.3%
31.3%	68.7%	English	11.3%
Program		Statistics	12.7%
BS	MS	CSIT	29.7%
57.3%	42.7%	Sociology	8.3%
Region		Psychology	6.7%
Rural	Urban	**Reasons to leave university**	
44.7%	55.3%	To accept a good job	19.7%
Educational Prize		It would cost more than my family could afford	9.7%
Yes	No		
71.7%	27.3%	Marriage	17%
Scholarship Offers		Lack of interest	11%
Yes	No	Lack of academic ability	8.7%
33%	67%	Other	34%
Father Profession			
Business man	33.3%	Doctor	2.3%
Government employ	10.3%	Late	3.0%
Engineer	5.0%	Others	35.7%
Lawyer	3.7%		

Table 3: Percentages, Mean and Standard Deviation of all Ordinal Scale Variables

Factor/Variables	SA%	A%	N%	D%	S.D%	Mean	S.D
Previous Achievements							
Influence of previous degree marks	24.7	43.3	17	9.3	5.7	3.72	1.10
Reflection of abilities	19	46.7	17.3	11.3	5.7	3.62	1.09
Home Environment							
Support by home environment	50.7	38	6.7	3.7	1.0	4.34	.83
Facilities provided by family	61.7	31.3	5.0	1.0	1.0	4.52	.72
Encouragement by family	62.7	27.7	7.0	2.0	0.7	4.50	.76
Study Habits							
Time management for getting a good grade	31.7	45.3	14.3	6.7	2.0	3.98	.95
Schedule Proper time for study	12	40.3	26.7	16.7	4.3	3.53	2.5
Avoid interference in planned schedule of study	13.7	30.3	29	20	6.7	3.24	1.12
Fully concentrated during study	18.3	52	16.3	11.3	2.0	3.73	.95
Proper revision of notes	13	39.3	21	21.7	5.0	3.34	1.10
Learning Skills							
Critical attitude towards new concepts	15	51.3	24	8.7	1.0	3.71	.86
Presentation skills	17.3	46.7	26	8.7	1.3	3.70	.90
Influence of presentation skills on academic performance	25.3	49.7	17.3	6.7	1.0	3.96	.84
Reading of material on course content	15.7	37	24.3	18.7	4.3	3.48	1.52
Express concept through Writing	27	48.7	17.7	4.3	2.0	3.95	.89
Confidence as UOG student	23.3	43.3	20.7	10	2.3	3.76	.99
Hardworking							
Focus on work during study	20.3	51	17.7	9.3	1.7	3.79	.92
Class participation	22	36.3	24	15.7	2.0	3.61	1.05
Competing environment	20.3	51	20.7	6.7	1.3	3.82	.87
Initiatives in academic activities	9.3	46	32	10.7	2.0	3.50	.87
Academically up-to-date yourself	15.7	43.7	20.7	15.7	4.3	3.51	1.06
Academic Interaction							
Influence of Interpersonal relationship on academic growth	24.3	48	19	8.3	0.3	3.88	.88
Role of Peer group support on academic growth	26	49.3	15	8.7	1.0	3.91	.91
Effect of academic interaction with Students	19	52.7	20.7	6.7	1.0	3.82	.85
Opportunities to meet faculty members	19.3	46.3	19.7	11.7	3.0	3.67	1.01
Interaction with teachers outside the classroom	32.3	41.3	15	7.3	4.0	3.91	1.05
Academic Performance							
Satisfaction with academic performance	13	44.3	16.7	21.3	4.7	3.40	1.10
Performance according to academic abilities	10.7	50	17.7	17.3	4.3	3.45	1.03
Satisfaction with academic experience	11.3	49	24.7	12.3	2.7	3.54	.94
Growth of academic performance	18	5	16.7	12	1.3	3.73	.93
Importance of grades	42	43.7	8.3	4.3	1.7	4.20	.88
Confidence after joining that university	33.7	47.3	11	5.3	2.7	4.04	.94
Ability to convey knowledge	18.7	53.3	19	7.0	2.0	3.80	.89
Faith on own perception	13.7	54.3	22	9.3	0.7	3.71	.84
Critical discussion	18.7	47	23.7	9.3	1.3	3.72	.91

SA: Strongly Agree, **A:** Agree, **N:** Neutral, **D:** Disagree and **SD:** Strongly Disagree

Table 4: Model Estimates of Confirmatory Factor Analysis of Different Factors

Variables	Parameter Estimate	Standard Error	T Statistic	Prob. Level
Study Habits				
Time management for getting a good grade	0.454	0.066	6.839	0.000
Schedule Proper time for study	1.105	0.179	6.184	0.000
Avoid interference in planned schedule of study	0.711	0.080	8.927	0.000
Fully concentrated during study	0.421	0.067	6.316	0.000
Proper revision of notes	0.607	0.077	7.867	0.000
Learning Skills				
Critical attitude towards new concepts	0.309	0.061	5.049	0.000
Presentation skills	0.531	0.066	8.072	0.000
Influence of presentation skills on academic performance	0.560	0.063	8.868	0.000
Reading of material on course content	0.441	0.109	4.039	0.000
Express concept through Writing	0.286	0.063	4.564	0.000
Confidence as UOG student	0.358	0.071	5.051	0.000
Hardworking				
Focus on work during study	0.409	0.064	6.424	0.000
Class participation	0.580	0.072	8.036	0.000
Competing environment	0.579	0.061	9.520	0.000
Initiatives in academic activities	0.454	0.060	7.561	0.000
Academically up-to-date yourself	0.430	0.074	5.847	0.000
Academic Interaction				
Influence of Interpersonal relationship on academic growth	0.395	0.057	6.928	0.000
Effect of academic interaction with Students	0.526	0.053	9.863	0.000
Opportunities to meet faculty members	0.755	0.064	11.795	0.000
Interaction with teachers outside the classroom	0.675	0.066	10.159	0.000
Academic Performance				
Performed Academically	0.392	0.072	5.443	0.000
Confidence After Joining that University	0.518	0.065	7.918	0.000
Ability to Convey through Knowledge	0.538	0.062	8.703	0.000
Faith on own Concepts	0.426	0.058	7.342	0.000
Critical Discussion on Concept	0.497	0.063	7.864	0.000
Home Environment				
Support by Home	0.554	0.053	10.550	0.000
Facilities by Family	0.530	0.047	11.391	0.000
Encouragement by Family	0.524	0.049	10.792	0.000

Table 5: Measure of Goodness of Fit of CFA Model of Different Factors

Factors	χ^2	d.f	p-value	χ^2 /d.f	GFI	AGFI	RMSEA
Study Habits	9.50602	5	0.09	1.90	.98	.96	0.05
Learning Skills	17.1534	9	0.04	1.90	.98	.95	0.05
Hardworking	12.9228	5	0.02	2.58	.98	.94	0.07
Academic Interaction	2.93344	2	0.23	1.46	.99	.97	0.03
Academic Performance	35.204	5	0.00	7.04	.95	.86	.14
Recommended				≤ 3	$\geq .90$	$\geq .90$	≤ 0.08

Table 6: Model Estimates of Structure Equation Model

Structural Relationships	Parameter Estimate	Standard Error	T Statistic	Prob. Level
Home Environment-->Learning Skills	0.447	0.142	3.141	0.002
Home Environment-->Academic Interaction	0.315	0.101	3.126	0.002
Study Habits-->Academic Interaction	0.213	0.048	4.396	0.000
Learning Skills-->Academic Performance	0.425	0.130	3.267	0.001
Academic Interaction-->Academic Performance	0.562	0.134	4.197	0.000

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 269-276

BAYESIAN ANALYSIS OF THE RAYLEIGH MODEL ASSUMING SINGLE AND MIXTURE PRIORS

Muhammad Aslam[1] and **Muneeb Javed**[2]
[1] Department of Statistics, Quaid-i-Azam University,
Islamabad, Pakistan Email: aslamsdqu@yahoo.com
[2] Department of Statistics, University of Gujrat,
Gujrat, Pakistan Email: muneeb.javed@uog.edu.pk

ABSTRACT

The comparison of informative priors for different sample sizes and parameter values are discussed. The Bayesian analysis for Rayleigh model assuming the Square root inverted gamma and Mixture of two component Square root inverted gamma priors is presented. The comparison is made on the basis of Bayes Estimators, Posterior Variances, Credible Intervals and Highest Posterior Density (HPD) Regions for evaluating the best prior distribution among the informative priors in Bayesian statistics. To see the trend of hyper-parameters the expression for the Bayesian Predictive intervals for informative priors are also calculated by assuming different combinations of the hyper-parameters and then take the expert suggestions on those combinations.

KEYWORDS

Square Root Inverted Gamma Prior; Mixture of Two Components Square Root Inverted Gamma Prior; Credible Intervals; Highest Posterior Density (HPD) Regions; Bayesian Predictive Intervals.

1. INTRODUCTION

In this paper, comparison of uninformative and informative priors is presented on the basis of Bayes Estimators, Posterior Variances, Credible Intervals and Highest Posterior Density Regions. Sinha (1998) has discussed the credible Intervals and HPDs for the parameter of Rayleigh distribution, exponential distribution and normal distribution etc. Sinha and Howlader (1983) & Lalitha and Mishra (1996) have quoted useful references on the Rayleigh model. In informative priors, we also assume the mixture priors that combine with likelihood to give mixture posterior distribution. Bose (1994) considered neighborhood classes of mixture priors. Results are presented for two kinds of “mixture classes”, which yield different types of neighborhoods. Bolstad (2004) has made robust Bayesian inference by using the two components of mixture priors.

The trend of hyper-parameters is determined by calculating the 95% Bayesian predictive intervals. Sinha (1998) has obtained the 95% Predictive Intervals for various sets of values of the hyper-parameters using the sample size of $n=100$ from the Mendenhall and Harder (1958) mixture model. Dey and Das (2005) discussed Bayesian prediction interval for a Rayleigh distribution. Saleem and Aslam (2008) have evaluated the Bayesian Predictive Intervals of the Rayleigh mixture assuming the Inverted Chi, the Inverted Rayleigh and the Square Root Inverted Gamma priors.

2. BAYESIAN ESTIMATION OF RAYLEIGH MODEL USING MIXTURE OF TWO COMPONENTS SQUARE ROOT INVERTED GAMMA PRIOR

Let $x_1, x_2, ..., x_n$ be a random sample of size n, drawn from Rayleigh distribution with unknown parameter θ. We assume that the θ follows mixture of two components Square Root Inverted Gamma distribution

$$p(\theta) = p\frac{2b_1^{a_1}}{\Gamma(a_1)}\theta^{-(2a_1+1)}e^{-\frac{b_1}{\theta^2}} + (1-p)\frac{2b_2^{a_2}}{\Gamma(a_2)}\theta^{-(2a_2+1)}e^{-\frac{b_2}{\theta^2}} \quad \theta > 0$$

The posterior distribution

$$p(\theta \mid x) = \frac{1}{C}\left\{ p\frac{b_1^{a_1}}{\Gamma(a_1)}\theta^{-\left\{2(a_1+n)+1\right\}}e^{-\frac{(b_1+\sum x^2)}{\theta^2}} + (1-p)\frac{b_2^{a_2}}{\Gamma(a_2)}\theta^{-\left\{2(a_2+n)+1\right\}}e^{-\frac{(b_2+\sum x^2)}{\theta^2}} \right\} \theta > 0 \tag{2.1}$$

where $C = p\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n)}{2(b_1+\sum X^2)^{a_1+n}} + (1-p)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n)}{2(b_2+\sum X^2)^{a_2+n}}$

where a_1, a_2 and b_1, b_2 are the hyper-parameters to be estimated.

2.1 THE BAYES ESTIMATOR AND VARIANCE USING THE MIXTURE OF TWO COMPONENTS SQUARE ROOT INVERTED GAMMA PRIOR

The expression for the Bayes estimator of θ by using the squared error loss function is:

$$E_{\theta|x}(\theta) = \frac{1}{2C}\left\{ p\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n-1/2)}{(b_1+\sum X^2)^{a_1+n-1/2}} + (1-p)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n-1/2)}{(b_2+\sum X^2)^{a_2+n-1/2}} \right\} \tag{2.2}$$

while expression for its variance is

$$V_{\theta|x}(\theta) = \frac{1}{2C}\left\{ p\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n-1)}{(b_1+\sum X^2)^{a_1+n-1}} + (1-p)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n-1)}{(b_2+\sum X^2)^{a_2+n-1}} \right\} - \left[\frac{1}{2C}\left\{ p\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n-1/2)}{(b_1+\sum X^2)^{a_1+n-1/2}} + (1-p)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n-1/2)}{(b_2+\sum X^2)^{a_2+n-1/2}} \right\}\right]^2 \tag{2.3}$$

where $C = p\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n)}{2(b_1+\sum X^2)^{a_1+n}} + (1-p)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n)}{2(b_2+\sum X^2)^{a_2+n}}$

2.2 PREDICTIVE INTERVALS USING THE MIXTURE OF TWO COMPONENTS SQUARE ROOT INVERTED GAMMA PRIOR

The predictive distribution of the future observation y is

$$p(y|\mathrm{x})=\frac{y}{C}\left\{P\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n+1)}{(b_1+\sum x_i^2+y^2)^{(a_1+n+1)}}+(1-P)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n+1)}{(b_2+\sum x_i^2+y^2)^{(a_2+n+1)}}\right\}$$

$y>0$ (2.4) where $C=p\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n)}{2(b_1+\sum X^2)^{a_1+n}}+(1-p)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n)}{2(b_2+\sum X^2)^{a_2+n}}$

$(1-\alpha)100\%$ Bayesian Predictive Interval (L,U) is obtained by solving the following two equations:

$$\int_0^L p(y|\mathrm{x})dy=\frac{\alpha}{2},\quad \int_U^\infty p(y|\mathrm{x})dy=\frac{\alpha}{2}$$

On simplification these equations become

$$\frac{\alpha}{2}=-\frac{1}{2C}\left[P\frac{b_1^{a_1}}{\Gamma(a_1)}\Gamma(a_1+n)\left\{\frac{1}{(b_1+\sum x_i^2+L^2)^{a_1+n}}-\frac{1}{(b_1+\sum x_i^2)^{a_1+n}}\right\}\right.$$
$$\left.+(1-P)\frac{b_2^{a_2}}{\Gamma(a_2)}\Gamma(a_2+n)\left\{\frac{1}{(b_2+\sum x_i^2+L^2)^{a_2+n}}-\frac{1}{(b_2+\sum x_i^2)^{a_2+n}}\right\}\right] \quad (2.5)$$

$$\frac{\alpha}{2}=\frac{1}{2C}\left[P\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n)}{(b_1+\sum x_i^2+U^2)^{a_1+n}}+(1-P)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n)}{(b_2+\sum x_i^2+U^2)^{a_2+n}}\right] \quad (2.6)$$

2.3 CREDIBLE INTERVAL USING THE MIXTURE OF TWO COMPONENTS SQUARE ROOT INVERTED GAMMA PRIOR

From equation (2.1) it follows that

$$\frac{2\left(b_1+\sum_{i=1}^{n}X_i^2\right)}{\theta^2}\sim\chi^2\left(2(a_1+n)\right)+\frac{2\left(b_2+\sum_{i=1}^{n}X_i^2\right)}{\theta^2}\sim\chi^2\left(2(a_2+n)\right)=\overset{*}{\chi}{}^2 .$$

Hence $1-\alpha=P\left[\overset{*}{\chi}{}_1^2<\frac{1}{\theta^2}<\overset{*}{\chi}{}_2^2\right]$

Thus, $\left[C_L^{(\theta)},C_U^{(\theta)}\right]=P\left[\sqrt{\frac{1}{\overset{*}{\chi}{}^2_{\alpha/2}}}<\theta<\sqrt{\frac{1}{\overset{*}{\chi}{}^2_{1-\alpha/2}}}\right]$ (2.7)

Is the $(1-\alpha)100\%-$ credible interval for θ.
where

$$\chi^{*2}_{\alpha/2} = \frac{1}{4C}\left[P\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n)}{(b_1+\sum_{i=1}^{n}X_i^2)^{a_1+n+1}}\chi^2_{2(a_1+n),\alpha/2}\right.$$
$$\left.+(1-P)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n)}{(b_2+\sum_{i=1}^{n}X_i^2)^{a_2+n+1}}\chi^2_{(a_2+n),\alpha/2}\right]$$

$$\chi^{2\,*}_{1-\alpha/2} = \frac{1}{4C}\left[P\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n)}{(b_1+\sum_{i=1}^{n}X_i^2)^{a_1+n+1}}\chi^2_{2(a_1+n),1-\alpha/2}\right.$$
$$\left.+(1-P)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n)}{(b_2+\sum_{i=1}^{n}X_i^2)^{a_2+n+1}}\chi^2_{2(a_2+n),1-\alpha/2}\right]$$

$$C = p\frac{b_1^{a_1}}{\Gamma(a_1)}\frac{\Gamma(a_1+n)}{2(b_1+\sum X^2)^{a_1+n}}+(1-p)\frac{b_2^{a_2}}{\Gamma(a_2)}\frac{\Gamma(a_2+n)}{2(b_2+\sum X^2)^{a_2+n}}$$

3. NUMERICAL STUDY

In order to compare estimates we conducted a numerical study, taking a random sample of different sizes and different parameter values and using the computer software. As one data was unable to clarify performance of the method, we used random sample of size $n = 50$, $n = 100$ and $n = 150$ from the Rayleigh distribution with parameter values $\theta = 6.5$ and $\theta = 5.5$

3.1 Bayesian Predictive Intervals using the Square Root Inverted Gamma Prior

Bayesian Predictive Interval using the Square Root Inverted Gamma Prior are evaluated using equation $\frac{\alpha}{2} = 1-\left(\frac{b+\sum X^2}{b+\sum X^2+L^2}\right)^{a+n}$ and $\frac{\alpha}{2} = \left(\frac{b+\sum X^2}{b+\sum X^2+U^2}\right)^{a+n}$ for different combinations of the hyper-parameters, a and b, and those are arranged in the following table.

Table 3.1: Bayesian Predictive Intervals for different Values of the Hyperparameters

	$a=5$	$a=10$	$a=15$	$a=20$	$a=25$
$b=1$	$L=1.90465$ $U=23.0734$ $\delta=21.16875$	$L=1.88625$ $U=22.84895$ $\delta=20.9627$	$L=1.86837$ $U=22.6309$ $\delta=20.76253$	$L=1.850990$ $U=22.4189$ $\delta=20.56791$	$L=1.83408$ $U=22.2128$ $\delta=20.37872$
$b=5$	$L=1.90486$ $U=23.07597$ $\delta=21.17111$	$L=1.88646$ $U=22.8514$ $\delta=20.96494$	$L=1.86858$ $U=22.6332$ $\delta=20.76462$	$L=1.851193$ $U=22.4213$ $\delta=20.570107$	$L=1.83429$ $U=22.2151$ $\delta=20.38081$

The lower and upper limits of the predictive intervals are observed as a function of various combinations of a and b. It is interesting to note that higher values of a while lower values of b lead towards higher precision.

3.2 Bayesian Predictive Intervals using the Two Components Square Root Inverted Gamma Mixture Prior

Bayesian Predictive Interval using the mixture of two components Square Root Inverted Gamma priors are evaluated using equation (2.5) and (2.6) for different combinations of the hyper-parameters are arranged in the following table.

Table 3.2: Bayesian Predictive Intervals for different values of the Hyperparameters

	$a_1=5$ $a_2=5$	$a_1=10$ $a_2=10$	$a_1=15$ $a_2=15$	$a_1=20$ $a_2=20$	$a_1=25$ $a_2=25$
$b_1=10$ $b_2=10$	$L=1.90512$ $U=23.0791$ $\delta=21.17398$	$L=1.88671$ $U=22.8546$ $\delta=20.96789$	$L=1.86883$ $U=22.6363$ $\delta=20.76747$	$L=1.85144$ $U=22.4243$ $\delta=20.57286$	$L=1.83453$ $U=22.2181$ $\delta=20.38357$
$b_1=20$ $b_2=20$	$L=1.90564$ $U=23.0854$ $\delta=21.17976$	$L=1.88723$ $U=22.8609$ $\delta=20.97367$	$L=1.86934$ $U=22.6426$ $\delta=20.77326$	$L=1.85196$ $U=22.4306$ $\delta=20.57864$	$L=1.83503$ $U=22.2242$ $\delta=20.38917$

It is observed from the table that keeping the lower values of b_1 & b_2 and higher values of a_1 & a_2 to make predictive intervals of minimum spread, finally we select values of the hyperparameters as $a_1=5, a_2=25, b_1=15, b_2=10$.

3.3 Comparison of Uninformative and Informative Priors

In this section we compare Bayes estimates, posterior variances, credible intervals and highest posterior density (HPD) region for uninformative and informative priors.

Let us assume hyper-parameters as $a=25$ and $b=1$ (say) in case of Square root inverted gamma prior and the hyper-parameters as $a_1=5, a_2=25, b_1=15, b_2=10$ (say) in case of mixture of two component square root inverted gamma prior.

3.3.1 Comparison of Bayes Estimates

Bayes estimates are evaluated for uninformative (Jeffreys and Uniform) and informative (Square root inverted gamma and Mixture of two component square root inverted gamma) priors, for different sample sizes and different parameter values. We summarize these estimates in Table 3.3.

Table 3.3: Comparison of Bayes Estimates

Parameter point	n	p	θ	Bayes Estimates			
				Jeffreys Prior	Uniform Prior	Informative Prior	Mixture Prior
1	50	0.1	6.5	6.91	6.95	5.67	6.6
2	100	0.05	6.5	6.451	6.47	5.79	6.31
3	150	0.1	5.5	5.27	5.28	4.90	5.19
4	250	0.1	6.5	6.52	6.522	6.22	6.45

These estimates tend towards true value of parameter as we increase the sample size. In uninformative priors Jeffreys prior estimates are slightly more accurate than the uniform prior estimates. As compared to the single prior, the mixture prior gives us better result.

3.3.2 Comparison of Variances

In this section variances of uninformative and informative prior, comparing with different sample sizes and different parameter values. We summarize these variances in Table 3.4.

Table 3.4: Comparison of Posterior Variances with Uninformative and Informative Priors

Parameter point	n	P	θ	Posterior Variances			
				Jeffreys Prior	Uniform Prior	Informative Prior	Mixture Prior
1	50	0.1	6.5	0.244118	0.249171	0.110128	0.202335
2	100	0.05	6.5	0.105225	0.106295	0.0682012	0.095702
3	150	0.1	5.5	0.046719	0.047034	0.034668	0.043893
4	250	0.1	6.5	0.042647	0.042819	0.0354838	0.041043

Clearly, as we increase the sample size, variances are tending to decrease. Variances of informative priors are less than that of the uninformative priors, although the variance of mixture distribution is greater than the single informative distribution, and this can be further decrease by adopting some more suitable combination of hyper-parameters.

3.3.3 Comparison of Credible Intervals and Highest Posterior Density (HPD) Regions

Credible intervals and highest posterior density (HPD) regions are given for uninformative and informative priors. Comparison of the 95% intervals and regions are shown in the following tables.

Table 3.5: Comparison of 95% Credible Intervals with Uninformative and Informative Priors

Parameter point	n	p	θ	Credible intervals			
				Jeffreys Prior	Uniform Prior	Informative Prior	Mixture Prior
1	50	0.1	6.5	(6.02, 7.94)	(6.05, 8.00)	(5.06, 6.36)	(5.79, 7.56)
2	100	0.05	6.5	(5.85, 7.13)	(5.87, 7.14)	(5.30, 6.33)	(5.74, 6.95)
3	150	0.1	5.5	(4.88, 5.73)	(4.88, 5.72)	(4.55, 5.28)	(4.81, 5.63)
4	250	0.1	6.5	(6.12, 6.93)	(6.13, 6.94)	(5.87, 6.60)	(6.07, 6.87)

95% credible intervals for θ assuming uninformative and informative priors for different sample sizes and parameter values are given in table 3.5. It is evident that as we increase the sample size, the interval length becomes narrower. The informative prior gives narrower intervals than those of uninformative priors; therefore interval estimates based on informative priors are more efficient than ones based on the uninformative priors.

The highest posterior density (HPD) regions are summarized in Table 3.6.

Table 3.6: Comparison of 95% HPD with Uninformative and Informative Priors

Parameter point	n	p	θ	Highest Posterior Density (HPD) Regions		
				Jeffreys Prior	Uniform Prior	Informative Prior
1	50	0.1	6.5	(5.97, 7.893)	(5.999, 7.94)	(5.04, 6.334)
2	100	0.05	6.5	(5.829, 7.097)	(5.84, 7.116)	(5.2895, 6.3112)
3	150	0.1	5.5	(4.86, 5.702)	(4.861, 5.720)	(4.5385, 5.2738)
4	250	0.1	6.5	(6.12, 6.92)	(6.12, 6.93)	(5.863, 6.5943)

95% HPD for θ assuming the uninformative and informative priors for different sample sizes and parameter values are given in above table. As we increase the sample size the regions become narrower. In case of informative prior regions have narrower length than that of the uninformative priors.

4. CONCLUSION

The comparisons are made between uninformative and informative priors on the basis of Bayes estimators, posterior variances, credible intervals and highest posterior density (HPD) regions for different sample size and parameter values.

The Bayes estimates assuming both the uninformative priors gave very close results to the corresponding true parameter for Rayleigh model. The Jeffreys prior gave slightly more accurate result as compare to the Uniform prior. These estimates give more accurate information when we increase the sample size. The posterior variances of the estimator for Rayleigh model assuming Jeffreys prior are slightly lesser than the posterior variances of the Uniform prior distribution. So Jeffreys prior is more accurate and efficient as its Bayes estimates are more close to the true parameters and its posterior variances are lesser than the Uniform prior. These points make us clear that Jeffreys prior is slightly better than the uniform prior.

The Interval estimates assuming both the uninformative priors contain the corresponding true parameter but the Jeffreys prior gave narrower interval as compared to the Uniform prior. Also the credible intervals and highest posterior density (HPD) regions become narrower when we increase the sample size.

In interval estimation, the informative prior gave us better results as compare to the uninformative priors. This is depicted from the evaluated credible intervals and highest posterior density (HPD) regions which are narrower than their uninformative counterparts.

REFERENCES

1. Bolstad, W.M. (2004). *Introduction to Bayesian Statistics*. John Wiley & Sons.
2. Bose, S. (1994). Bayesian Robustness with Mixture Classis of Priors. *The Annals of Statistics*, 22(2), 652-667.
3. Dey, Sanku and Das, M.K. (2005). A Note on Prediction Interval for a Rayleigh Distribution: Bayesian Approach. Presented at *International Workshop/Conference on Bayesian Statistics and its Applications (*January 6-8, 2005) Banaras Hindu University, Varanasi, India.
4. Lalitha, S. and Mishra, A. (1996). Modified maximum likelihood estimation for Rayleigh distribution. *Comm. Statist. Theory & Math*. 25, 389-401.
5. Mendenhall, W. and Hader, R.A. (1958). Estimation of Parameters of Mixed Jeffreys, H. (1967). *Theory of Probability*, London: Oxford University Press.
6. Mendenhall W. and Hader R.J. (1958). Exponentially Distributed Failure Time Distributions from Censored Life Test Data. *Biometrika*, 45(3/4).
7. Saleem, M. and Aslam, M. (2008). On Prior Selection for the Mixture of Rayleigh Distribution using Predictive Intervals. *Pak. J. Statist.*, 24(1), 21-35.
8. Sinha, S.K. (1998). *Bayesian Estimation*. New Age International (P) Limited, Publishers, New Delhi.
9. Sinha, S.K. and Howlader, H.A. (1983). Credible and HPD intervals of the parameter and reliability of Rayleigh distribution. *IEEE Trans. Reliability*, 32, 217-220.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 277-284

FACTORS INFLUENCING ACADEMIC FAILURE OF UNIVERSITIES STUDENTS

Yousaf Ali Khan[1], **Zahoor Ahmad**[2] and **Sadia Kousar**[3]
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: [1]sehr.iman@yahoo.com
[2]zahoor.ahmed@uog.edu.pk
[3]friend.stat@gmail.com

ABSTRACT

There is a close link between education and development. Education plays a vital role in human capital formation .Academic failure from universities is a problem that has became a serious concern for higher education institutions. This study presents the result of a recent investigation at the University of Gujrat that attempted to identify the post enrollment factors that students perceived as having important influences on students failures in their university studies. Data is collected by using questionnaire. The Independent-Samples t-test procedure is used to compares means for two groups of cases. Structural equation modeling (SEM) analysis is also performing to test the relationship among latent factor represented by the actual data from100 drop out students of university. Finding indicates that male and female groups are significantly different in their perception of causes of academic failure, but perception of students from different programs (Bs, M.A /M.Sc) is not significantly different. In our estimated model class participation and lack of family support is directly affected on students drop out from universities, but wrong choice of subject and change of system which is basically link with class participation and indirectly effect on students drop out.

KEYWORDS

Identify, Lack of Commitment, Appropriate, Gujrat.

1. INTRODUCTION

Education plays the role of leadership in the society. In Quran (Holy book of Muslims) the first revelation from God to Prophet Mohammad was an order of ***Iqra……*** means read. It indicates the importance of education in Islam. In our age education plays a essential role in transforming the society. The role of university in transforming the society is particularly critical because it educates virtually all the people who are destined for leadership in our society from all walks of life. Its main duty is to prepare leadership through sound liberal learning not only during the period of formal education but also throughout their lives. So university is responsible to turnout well-educated population for the leadership of the nation. All over the world universities are guiding and co-operating with the industrial an agricultural development organizations and they are developing their economics rapidly and meaningfully

Students entering universities in Pakistan come from a wide range of social and cultural backgrounds that give them very different life experiences, different educational opportunities and a great variety of expectations, needs and academic potential .This situation also occurs in other countries that have shifted the focus of higher education from exclusiveness to mass opportunity [McKenzie & Schweitzer (2001)]. When students are admitted to a higher education institution there is a understood assumption that they will be capable of successfully completing the course in which they are permitted to enroll or not. Therefore, it is necessary to have entry requirements that permit valid student selection decisions to be made. However, there can be no guarantee that these students will eventually satisfy the requirements for graduation. There are many factors that can contribute to the high drop out rates such as low commitment with study, drug use, the lack of motivation or interest or the lack of parental involvement and supervision. However, these individual reasons are not the only factors that can lead teens to dropout. University was traditionally considered a ladder out of poverty for young working class people. According to Peters (1992) dropping out from classes is an old thing. He reported that the completion of even the first arts course leading to the Bachelor's degree is not the rule but the exception. This is so because one quarter to one third of registered students left the university as Bachelors, hardly one twentieth to one sixteenth as Masters. There are various reasons that made them leave, that is mostly personal characteristics, environmental characteristics and academic factors. He further observes that students drop out because of lack of motivation, lack of interest, boredom, apathy, rejection of the teaching programmed, the feeling of not receiving anything from the course, lack of purpose and uncertainty in the occupational orientation.

The fact that so many factors can be important is probably the main reason that single measures based on previous academic success, particularly at school, are not strong predictors of success at university. Killen (1994) suggested that no matter how carefully students are constructed school matriculation examinations or special university entrance examinations are not likely to be strong predictors of success at university because they do not measure no intellective factors that are related to many of the important influences on success that students encounter after they enroll at university. Given the complexity of the problem, it seems unlikely that there is much value in trying to find simple pre-enrolment predictors of success at university. Rather, it might be more useful to focus on post-enrolment factors. Studies that have taken this approach have identified a limited number of factors that appear to have a strong influence on academic success.

The purpose of the study is to identify the post enrolment factors that lecturers and students see as having important influences on student failure at university. I think if we are able to uncover the views of students have about these different adversities that exist which may allow us to better understand what types of programs will be most effective. Identifying these factors has the potential to be useful in several important ways. First, it can provide a basis for helping students to reflect on their perceptions and expectations of university study so that they can gain more control over their learning and Approach University studies in a way that will maximize their chances of success. Second, it can provide a basis for helping lecturers reflect on their expectations of and about students so that they will be better informed about ways in which they can facilitate student learning, enhance the influence of positive factors and minimize the influence of negative factors on student success. Third, the results can be used by university administrators to help them provide a learning environment that will maximize their chance of success.

Purposed model of the study

With using all the factor of our study we generate a purposed model of this study by the help of literature which is given below. In this model lack of commitment to study, Time management and class participation have direct affect on dropout and all other factors are affecting indirectly.

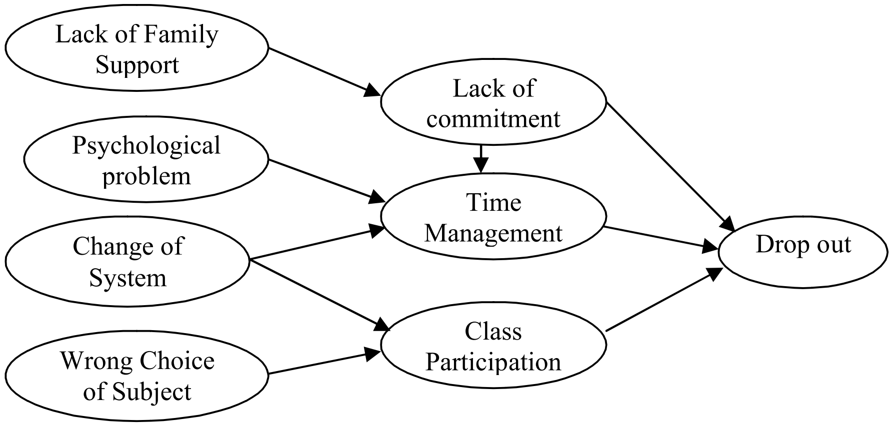


Fig. 1.1: Purposed model of the study

2. LITERATURE REVIEW

Tinto (1975) was the first investigator to propose a longitudinal model of student drop out, that predicted, rather than simply explained this behavior. His theory was based on Durkheim's (1961) theory on suicide, which suggested that suicide was most likely when an individual was insufficiently integrated into society, both in terms of insufficient collective affiliation, and also insufficient congruence with the moral values of society .He viewed drop out as analogous to suicide, and suggested it occurred when an individual had insufficient social integration with others at university, or when they did not fit in with the norms and value patterns of the university, and were unable to achieve academically to the required level (i.e. insufficient academic integration). Social integration occurs through informal peer group associations, semiformal extracurricular activities, and also through interaction with university staff. Noth and Neill (1981) many research studies have specified that family factors are significantly related to the decision of students to drop out. Studies found that the dropout's family was less solid, less influenced by a father, less likely to interact in leisure activities, and less able to communicate than the persister's family. Research studies also indicated that loss of a family member due to death or divorce and other family problems influence a student's decision to drop out In addition, the level of education and the occupation of dropouts' parents were significant factors. The finding shows that the focus of the study is on the failure due to family maters. The study is enormously related to our research, because both are discussed in the issues of failures. Makki (2007) In this study researcher found that majority of drop outs have no intention of continuing their studies either because they have given up for social or economic reasons, they have gone to study in another country, they have changed disciplines or have already built a satisfactory career without the need for educational qualifications. The largest percentage of non-active students appear to come from socially and economically weaker classes who attend less glamorous courses whose degrees have rather dubious value in the wider labour market system. This study showed the positive relationship between social and economic reasons in decision making to drop out. It also stressed that students should shows the willingness to maintain the sense of responsibility towards there academic goal. Young, et al. (2008) the purpose of this study is to develop a scale to measure drop out for cyber-university students in e-learning. In this study five dimensions of learners drop out were identified; 1) students background, 2) students characteristic, 3) curriculum and contents, 4) educational environment, and 5) educational outcome. And each dimension consists of several items. For items analysis, EFA (exploratory factor analysis) will be performed in Study 1. In order to validate the scale CFA (confirmatory factor analysis) will be conducted. This research will present the reliability and validity of the new scale of drop-out in Cyber University. In this study the writer discussed the five reasons of dropping out on which our study based, the most important are students back ground, curriculum and contents. These fetchers are also elaborated in our study.

3. OBJECTIVE OF THE STUDY

1. To explore the demographic characteristics of the respondents.
2. To develop a drop out model of structural equation modeling.
3. To provide estimated drop out model or to find the estimate of suggested model
4. To conform the factors those can effect on drop out of students which we have suggested.

4. DATA AND METHODOLOGY

Questionnaire.

The questionnaire consisted of eight factors: time management, wrong choice of subject, family problems, the lack of commitment or interest, low understanding level, class participation, psychological problem, because of semester system.

Design and Analysis.

The data were analysis with two major statistical methods: structural equation modeling and independent sample t-test. First with in each factor structural equation modeling was use to examine the influence of factors on student dropout. Second independent sample t- test were performed to examine mean difference for two groups of cases. (Bs/M.A/M.Sc, male/female).

Confirmatory Factor Analysis.

A confirmatory factor analysis (CFA) is perform to test the adequacy of the measurement model for the latten factor of this study through STATISTICA 7.0 software, to determine the goodness of fit. The factor which is conformed by confirmatory factor analysis is than the conformed factor is use in structured equation modeling. Since chi-square test is based against large sample sizes (Byrne, 1994) the goodness of fit index is more reliable for testing model fit. The factors Wrong choice of subject, change of system, lack of family support, class participation, time management, psychological problem and overall dropout are conformed. Because there goodness of fit indicate that these factors are fulfill the given criteria. But the factor lack of commitment is none conformed.

4.1. Structural Equation Modeling Results

Structural equation modeling (SEM) analysis is perform to test the relationship among latent factor represented by the actual data from 100 drop out students of university with the STATISTICS 7.0 statistical package on the questions from five likert-scale.Different structure run on this data but these path are not significant there goodness of fit are not fulfill the given criteria. Finally we run a Figure 1.2 which is appropriate according to given criteria. In Figure 1.2, circles represent basic factors. Lines connecting variables imply a direct effect. The wrong choice of subject factor measures four variables, change of system measures four variables, class participation measure four variables, lack of family support measures four variables, and overall dropout measures eight variables but in model fitting we use only those variables which have high value of parameter estimates.

Table 1:
Goodness of fit indexes of structural equation modeling

Fit statistics	Values	Criterion
Chi-Square	153.049	Smaller is better
GFI	0.964	≥0.95
RMSEA Index	0.027	≤0.08
Non-Normed Fit Index	0.997	≥0.95
Comparative Fit Index	0.999	≥0.95
RMS Standardized Residual	0.079	≤0.08

The model is specified by the following equations,

The estimated equations are

Class participation =0.34 wrong choice of subject + 0.713change of system
Drop out =B class participation + $\acute{\Gamma}_3$ lack of family support + ς_2
Drop out =0.404 class participation + 0.404 lack of family support
Drop out =$\acute{\Gamma}_1$ wrong choice of subject + $\acute{\Gamma}_2$ change of system
+ $\acute{\Gamma}_3$ lack of family support + B class participation + $\varsigma3$
Drop out =0.404 (0.34 wrong choice of subject + 0.713 change of system)
+0.404 lack of family support

Finally estimated equation of drop out model is

Drop out = 0.1373 wrong choice of subject + 0.2881 change of system
+ 0.404 lack of family support

Several indices describe the overall fit between the observed data and a model, including Chi - square, GFI, AGFI, CFI, NNFI and RMSEA .Factor class participation and lack of family support is directly affected on students drop out from universities, but wrong choice of subject and change of system which is basically link with class participation and indirectly effect on students drop out. So in this way we have a path to find the actual affected factors that affects student’s dropout from universities directly or indirectly.

Table 2:
Estimated model of structural equation modeling

	Parameter	Stander Error	T Statistics	Prob. level
(WCS)-15->(CP)	0.340	0.124	2.738	0.006
(COS)-16->(CP)	0.713	0.137	5.211	0.000
(CP)-17->(DO)	0.404	0.157	2.573	0.010
(FS)-18->(DO)	0.726	0.119	6.105	0.000

The table-2 show that in this model we have four relationships which show that all the p-values of the relationship are significance. For checking the goodness of fit of the model we use the results of table-1.The goodness of fit indices indicate that the model was adequate: the GFI = 0.964 which is greater than 0.95 and also NNFI=0.997, CFI=0.999, which is also greater than 0.95,all indices are fulfill the given criteria.(Hu and Bentler 1999) which indicated the model fit the data adequately. The RMSEA = 0.027, which is smaller than .08, indicated a fair goodness of fit. In resulting we find a proper model fitting by using structure equation mode

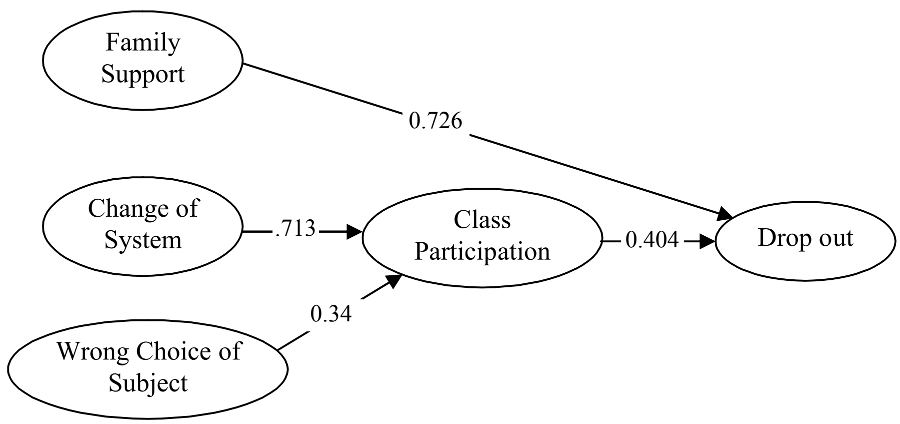


Fig. 1.2: Hypothesized Structural Equation Modeling of factor affecting student's dropout from universities with parameter estimates.

Figure 1.2 summaries the relationship among the factors and contain the path coefficients. All most all path coefficients between the laten factors are significant (at the .05 level).The results shows that the wrong choice of subjects has positive influence on the class participation (β=.34, t =2.738, $p < .05$) it implying that as students select wrong subject in other word not interesting subject they don't perform better in the class and at last he/she will drop from universities. Result also shows that the change of system has positive influence on the class participation (β=.713, t =5.211, $p < .05$) it implying that change of system play a role in student drop out. Family support has positive influence on the students drop out (β=.404, t =2.573, $p < .05$) it implying that as students family support and home environment has great impact on students success or failure. Class participation has great influence on the students drop out (β=.404, t =2.573, $p < .05$) it shows that as students don't perform well in class they cannot get academic success. Finally it shows that all the factor effecting on the student's academic performance directly or indirectly.

T-test. The Independent-Samples t-test procedure is use to compares means for two groups of cases. Since the data is approximately normal so in this study we use independent sample t–test to check the mean difference of the total score of all the factors among the students of male and female groups .Results table-3 show that males and female groups are significantly differ in their perceptions of cause of academic failure in the total score of the two factors. Namely, Lack of commitment to study and time management. It also shows that males and female groups not significantly differ in their perceptions of cause of academic failure in the total score of the six factors. Namely, Wrong choice of subject, Class participation, Semester system, Lack of family support, Psychological problem.

Table 3:
The Independent-Samples t-test for compares means of gender

Factor with total scores	Male		Female		T value	P value
	Mean	Std. Deviation	Mean	Std. Deviation		
Lack of commitment study	4.3684	.64453	1.5349	.66722	-1.259	.011
Wrong choice of subject	2.4386	.80217	2.5116	.70279	-.475	.636
Class participation	8.8421	4.30859	9.2558	4.69089	-.458	.648
semester system	2.6842	.46896	2.6977	.46470	-.143	.887
Time management	5.5088	.57080	2.6279	.57831	-1.027	.031
Lack of Family support	2.6491	.58221	2.7674	.42746	-1.123	.264
Psychological	2.0702	.45746	1.9767	.34423	1.121	.265

Table 4:
The Independent-Samples t-test for compares means of BS and MA/MSC classes

Factor with total scores	MA/MSc		BS		T value	P value
	Mean	Std. Deviation	Mean	Std. Deviation		
Lack of commitment study	1.4394	.61090	1.4412	.74635	-.013	.990
Wrong choice of subject	2.5455	.70562	2.3235	.84282	1.393	.487
Class participation	8.7273	4.47401	9.5882	4.43893	-.914	.363
semester system	2.6667	.47502	2.7353	.44781	-.698	.487
Time management	2.5758	.58337	2.5294	.56329	.381	.704
Lack of Family support	2.6818	.58221	2.7353	.42746	-.483	.630
Psychological	2.0152	.41109	2.0588	.42220	-.499	.619

Independent sample t –test is also performed to see the mean difference of the total score of all the factors among the students from Bs and M.A/M.Sc classes. Result In table-4 shows that when the students were classified in term of their classes no significance differences are found among them in term of total score of the entire instrument (factors).

5. CONCLUSION

Student effort is prominent in student's expectation of success and failure. Active study, setting appropriate goals, a good study environment, and effective time management is considered important for academic success. Academic failure is attributed primarily to lack of study, poor time management, and inadequate goal setting. we can say that a better understanding of the factors or variables that would influence on academic failure and those factors that would motivate students to engage persistently in their studies might also hold, the key towards improved student performance at institutions of higher learning.

REFERENCES

1. Akaike, H. (1974). A new look at the statistical Model Identification. *Research and Development*, 25(1), 935-940.
2. Astleitner, H. (2000). Dropout and distance education. *Journal of Adolescence*, 20(1), 163-176.
3. Barrett, P. (2007). Structural equation modeling: Adjusting model fit. *Personality and Individual Difference,* 42(50), 815-824.
4. Bartz, D. and Miller, L. (1991). *Teaching methods to enhance student learning: What research says to the teacher. Washington DC National Education Association.* ERIC Document Reproduction Service No. ED34068.
5. Bentler, P. (1990). Comparative Fit Indexes in Structural Model. *Psychological Bulletin,* 88(3), 588-606.
6. Bentler, P.M. and Bonnet, D.C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin,* 88(3), 588-606.
7. Chyung, Y., Winiecki, D. and Fenner, J.A. (1999). Evaluation of effective interventions to solve the dropout problem in adult distance education. *Childhood and Society*, 10(2), 51-55.
8. Crowley, S.L. and Fan, X. (1997). Structural equation modeling: Basic Concepts and Application in Personality Assessment Research. *Personality Assessment,* 68(3), 508-513.
9. Christic, A. and Smith, K. (2004). Individual study style and their association with learning outcome. *Higher Education,* 20, 167-89.
10. Durkheim, E. (1951). *Suicide: a study in sociology*. Glencoe, IL, Free Press.
11. Demo, D., Small, S. and Savin, W. (1987). Family relations and the self-esteem of adolescents and their parents. *Marriage and the Family*, 49(1), 705-715.
12. Farmer, S.H. and Vispoel, P.W. (1990). Attribution of Female and Male adolescents for real-life failure experiences. *Experimental Education*, 58(2), 41-52.
13. Fan, X., Thompson, B. and Wang, L. (1999). Effect of sampling size, estimation model and model specification on structural modeling fit indexes. *Structural Equation Modeling,* 6(1), 56-83.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 285-294

A STUDY ON ACADEMIC DISHONESTY OF UNIVERSITY STUDENTS

Muafia Munir[1], Zahoor Ahmad[2] and **Erum Shahzadi[3]**
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: [1]muafia08_munir@yahoo.com
[2]zahoor.ahmed@uog.edu.pk
[3]erum_stat@yahoo.com

ABSTRACT

Academic dishonesty is a prevalent problem that crosses all disciplines at the university level. Academic dishonesty, a serious concern, cuts to the heart of the purpose of higher education. The finished products of the university, its students, may not possess the fundamental information and skills implied by the transcript. Academic dishonesty is an affront to academically honest students as well as most college professors whose purpose is to teach. In this paper the main objective is to classify the cases in to three groups of free-riders (never, sometimes, always) on the basis of different factors of academic dishonesty by using discriminant analysis. Secondly to investigate the relative importance of different factors with respect to discriminating power. A sample of almost 200 students was selected from the population of 6749 students of University of Gujrat. The results show that all considered independent variables included in this study are playing significant role for discriminating the cases into free-riders groups. Classification table show that 74.5% cases are correctly classified using the developed discriminant function. The most important factors are Sabotage, Electronic Cheating and Outside Help.

KEYWORDS

Academic Dishonesty; Plagiarism; Fabrication; Deception; Cheating; Sabotage; Electronic Cheating; Unethical Behavior; Free-riders; Discriminant Analysis.

1. INTRODUCTION

Dishonesty is a word which, in common usage, and it may be defined as the act or to act without honesty. It is used to describe a lack of probity (virtue), cheating, lying or being deliberately deceptive (misleading) or a lack in integrity (truth). Dishonesty is the fundamental component of a majority of offences (violation of law) relating to the acquisition (achievement).

Academic dishonesty or academic misconduct is any type of cheating that occurs in relation to a formal academic exercise. Academic dishonesty has been documented in most every type of educational setting, from elementary school to graduate school, and has been met with changeable degrees of respect throughout history. Today, educated society tends to take a very negative view of academic dishonesty. The University of Newcastle defines academic dishonesty as comprising two major categories, academic fraud and plagiarism. Academic fraud is making a false representation to gain an unjust advantage. The presentation of the thoughts or works of another as one's own is plagiarism. Some definitions include the objective of the person engaging in the dishonest behavior [Tibbetts (1998), (1999)]. Academic dishonesty “is defined as an intentionally unethical behavior.” Studies defined academic dishonesty based upon a particular violation behavior, such as cheating in a test or plagiarism [McCabe and Bowers (1994)].

Academic/Student dishonesty, a prevalent and permanent phenomenon, has been an issues of considerable interest among researchers in ethics, psychology, sociology, and higher education. Many in the educational system are concerned with the problem of academic dishonesty and the rate at which it is increasing [Ameen, et al. (1996)].

Researchers have performed studies to find the motivations behind cheating and have pointed out factors that influence this dishonest behavior. Works by all of these scholars have led to a similar conclusion: that cheating is widespread on college campuses. Academic dishonesty, a serious concern on most college campuses, cuts to the heart of the purpose of higher education. The finished products of the university, its students, may not possess the fundamental information and skills implied by the transcript. Academic dishonesty is an injury to academically honest students as well as most college professors whose purpose is to teach [Newstead, et al. (1996); Graham, et al. (1994)].

Student academic dishonesty includes, but is not limited to, lying, cheating on exams, copying or using other people’s work without permission, shifting or forging documents, buying papers, plagiarism, purposely not following the rules, shifting research results, providing false excuses for missed tests and assignments, making up sources, and so on [Arent (1991); Pratt and McLaughlin (1989)].

Alienation (unfriendliness and separation) may cause students to engage in academic dishonesty [Eve and Bromley (1981)]. Similarly, a small but statistically significant association between anomie (i.e., lack of being tied to society and rules) and cheating was observed among business majors at an Australian university [Caruana, et al. (2000)]. Cheating may also occur because of low levels of commitment to the ideals of higher education and learning orientations (i.e., wanting to learn versus earning a grade/degree) [Haines, et al. (1986); Whitley (1998)].

Finally, past behavior tends to be the best predicator of future behavior, and this appears to be true for academic dishonesty. It was observed that those who cheated in high school are more likely to cheat in college or university level [Whitley (1998)]. Additionally, cheating was sometimes argued to be justified because of the course is too hard or the instructor is unfair [Diekhoff, et al. (1996)]. While there has been significant research on the subject of academic dishonesty, why students cheat and what types of cheating they typically engage in and has not been fully answered. “Little research appears to have been done to try and identify variables that have an effect on academic dishonesty” [Caruana, et al. (2000)].

In order to effectively combat cheating, it is necessary to understand how it is done, who does it, its forms, and why it is done. Factors of Academic Dishonesty:

1. **Plagiarism:** The adoption or reproduction of original creations of another author (person, collective, organization, community or other type of author, including anonymous authors) without due acknowledgment.
2. **Fabrication:** The falsification of data, information, or citations in any formal academic exercise.

3. **Deception:** Providing false information to an instructor concerning a formal academic exercise—*e.g.,* giving a false excuse for missing a deadline or falsely claiming to have submitted work.
4. **Cheating:** Any attempt to give or obtain assistance in a formal academic exercise (like an examination) without due acknowledgment.
5. **Sabotage:** Acting to prevent others from completing their work. This includes cutting pages out of library books or willfully disrupting the experiments of others.
6. **Outside Help:** use the other resources to complete assignments or notes.
7. **Electronic Cheating:** using the cells phones for getting help during the test or quiz.
8. **Unethical Behavior:** When want a get a high grade?
9. **Free-Rider:** free-riders were, ineffect, cheating their group members when they failed to contribute their fair share of work toward the project.

Today, educated society tends to take a very negative view of academic dishonesty. It has been reported that academic dishonesty is a prevalent problem that crosses all disciplines at the university level. But, how prevalent is it in University of Gujrat. As academic dishonesty is a multidimensional construct, so the results of this study will be beneficial to explore that which factor is more serious than other i.e.; cheating is more serious than plagiarism.

In this paper the main objective is to classify the cases in to three groups of free-riders (never, sometimes, always) on the basis of different factors of academic dishonesty by using discriminant analysis. This study will be helpful for the policy makers to develop the character development programs. As dishonest behavior is not good in professional life so by this study students “character building” can be done towards an honest behavior by the prevalence of the particular factor of dishonesty.

So I want to conduct this study in UOG as this type of study has not conducted before at this level.

2. LITERATURE REVIEW

Lim and See (2001) examined in his study students attitudes toward cheating and whether they would report instances of cheating they witnessed. Data were collected from 518 students, in three educational institutions in Singapore. Findings suggested that students perceived cheating behaviors involving exam-related situations to be serious, whereas plagiarism was rated as less serious. Majority of the students admitted having engaged in such behavior. As finding about the prevalence of academic dishonesty they found that students were morally hesitating about academic cheating and tolerant among their peers and the majority of students ignore the problem rather than to tell other peer groups or teachers.

Whitley (2001) focused on gender differences alone, finding that women students demonstrated significantly more negative attitudes towards cheating than male students even though reported frequencies of cheating behavior for both male and female were nearly the same.

Underwood and Szabo (2003) reported the issue of Academic offences and e-learning: individual propensities in cheating. It has been argued that the use of

Information and Communications Technologies has made academic dishonesty easier. Data were collected from 291 typical UK undergraduates' students. Multiple linear regression analyses showed that to achieve the academic goals students accept that they involve in plagiarism, acceptance of acts such as plagiarism as a legal way to achieve academic goals. There was a positive correlation between use of information from the Internet for assignments and gender ($r = 0.16$, $p < 0.01$) with female students (mean = 3.17) reporting more use than males (mean = 2.79). New undergraduates were more likely make mistakes than students in later years of their degree. There were number of interrelated factors impacting individuals' willingness to commit academic offences. While fear of failure is important there were other factors that might encourage students to cheat such as an inability to manage their level of work set, dullness of the work and time management problems.

Etter, S. et al. (2006) discussed Ethical Orientations and Personality Factors Associated with Attitudes about Cheating with Information Technology on the origins of academic dishonesty (*e.g., cutting and pasting excerpts from Web sites without acknowledgment at two institutions).* Take 237 undergraduate students at a private church-affiliated college and their counterparts at a regional campus of a research university. Questionnaire was used as a study instrument. The ordinal rankings of academically dishonest behaviors at both institutions were similar (rho = 0.90).

Nadelson (2007) carried out a study on academic misconduct by university students, faculty perceptions' and responses and academic misbehavior faculty member's suspect occur in their classrooms, methods they use to prevent such misconduct, and the factors that influence instructor's decisions to act on suspected inappropriate behavior. Data were collected using survey from faculty members at a public university; a seven point likert scale questionnaire is used. Information was collected about 300 full-time graduate and undergraduate faculty members. For relationship between variables correlation analysis is performed and t-test analysis is performed for mean comparison. Faculty members indicate that they use a variety of measures to improve students' behavior. However some teachers did not act on suspected academic dishonesty due in part to their anxiety about the process.

Eastman, et al. (2008) measured the reason of unethical behavior. Data were collected from 421 students by using convenience sampling from state universities in the southern region. Five point likert scale questionnaire was used as survey tool. There was significant evidence links between those students who feel they had stronger reasons for committing unethical academic behaviors. Reporting that they are more involve in academic dishonesty in the area of cheating, seeking outside help, plagiarism and electronic cheating than those students with weaker reasons for unethical behaviors through a series of t-test.

3. MATERIAL AND METHODS

3.1 Population:

Our study population consisted of BS and MS students (social sciences and basic sciences). Students were selected from Academic block, Science block, Medical block and Engineering block of University of Gujrat (Hafiz Hayat Campus) and select those students who were studying in summer section 2010, from the total population of 6749 students.

3.2 Sample Selection:

For the selection of an appropriate sample simple random sampling is used. For this study we take six blocks of University of Gujrat for selection of sample and than from these blocks, four blocks were selected at random (**Academic block, Science block, Medical block and Engineering block**). A sample of almost 200 students was selected from the population of 6749 students of University of Gujrat **ground and first floor of each block.**

3.3 Research Instrument:

In this study questionnaire is used as a data collection tool and is consist of two parts. First part of the questionnaire consisted of information on the demographic characteristics of University students like age, gender, program, discipline, region, current GPA, CGPA and family system of the respondent. Second part consisted of 59 questions using 5 point Likert Scale (never, rarely, sometimes, usually, always) related to academic dishonesty information and nine factors are used for collecting this information. These factors are Plagiarism, Fabrication, Deception, Cheating, Sabotage, Outside Help, Electronic Cheating, Unethical Behavior and Free-Riders.

3.4 Data Analysis Technique:

3.4.1 Discriminant Analysis:

When nonmetric dependent variables may be used for prediction and explanation then discirminant analysis is used. When researcher are interested in the prediction and explanation of the relationships that affect the category in which an object is located they use discriminant analysis. The purpose of discriminant analysis function is an understanding of the data set, as a careful examination of the prediction model that results from the procedure can give insight into the relationship between group membership and the variables used to predict group membership. The basic purpose of discriminant analysis is to estimate the relationship between a single nonmetric (categorical) dependent variable and a set of metric independent variables, in this single form:

= + + +………+
(nonmetric) (metric)

Discriminant analysis is used to predict model the value of a dependent categorical variable based on its relationship to one or more predictors. Discriminant analysis can determine which variables are the best predictors of group membership. It determines which groups differ with respect to the mean of a variable, and then uses that variable to predict new cases of group membership.

Given a set of independent variables, discriminant analysis attempts to find linear combinations of those variables that best separate the groups of cases. These combinations are called discriminant functions and have the form displayed in the equation.

= + +…………..+

where

is the value of the kth discriminant function for the ith case

p is the number of predictors

bjk is the value of the jth coefficient of the kth function

xij is the value of the ith case of the jth predictor

The number of functions equals min (#groups-1). The procedure automatically chooses a first function that will separate the groups as much as possible. It then chooses a second function that is both uncorrelated with the first function and provides as much further separation as possible. The procedure continues adding functions in this way until reaching the maximum number of functions as determined by the number of predictors and categories in the dependent variable.

3.4.2 Assumptions:

The discriminant model has the following assumptions: All the independent variables (predictors) should fellow normal distribution. Cases should be independent and the predictors are not highly correlated with each other. Within-group variance-covariance matrices should be equal across groups. The correlation between two predictors is constant across groups. When two classifications are involved, the technique is referred to as two-group discriminant analysis. When three or more classifications are identified, the technique is referred to as multiple discriminant analysis (MDA).

4. RESULTS AND DISCUSSION

From Table-1 we see means of each cases of each predictor. This table shows the mean values of all the predictors according to their case in which they lie. Plagiarism, cheating, sabotage and unethical behavior have high mean value in the always category as compare to other categories. This means that those respondents who have unethical behavior were always doing plagiarism, cheating and sabotage.

Assessing the Contribution of Individual Predictors

From tests of equality of group means, the discriminant function coefficients and the structure matrix we can assess the contribution of each variable to the model.

Table-1 show the test of equality of group means measure each independent variable's potential before the model is created. Each test displays the results of a one-way ANOVA for the independent variable using the grouping variable as the factor. If the significance value is greater than 0.10, the variable perhaps does not contribute to the model. It also gives significant results at the 0.05 level of significance. According to the results in this table, every variable in discriminant model is significant. This shows that all the variables play significant role, all the p-values are less than the level of significant 0.10 and 0.05.

Wilks' lambda is a further measure of a variable's potential. Smaller values point out that the variable is better at discriminating between groups. In this Sabotage has the smaller value which indicates that sabotage is the most important factor for this study, than electronic cheating and outside help. The most important factors are Sabotage, Electronic Cheating and Outside Help according to Wilks' lambda values.

Table 1: Group Descriptive Statistics and Tests of Equality in the Three Group Discriminant Analysis

Independent Variables	Dependent variable Group Mean			Test of Equality of Group Means		
	Group 1: never (n=139)	Group 2: sometimes (n= 49)	Group 3: always (n= 16)	Wilks' Lambda	F Value	p-value
Plagiarism	6.53	9.24	9.50	.823	21.644	.000
Fabrication	15.16	19.57	18.62	.864	15.803	.000
Deception	5.51	7.39	7.12	.914	9.448	.000
Cheating	16.99	21.92	22.44	.877	14.089	.000
Sabotage	6.26	10.18	10.25	.715	40.079	.000
Outside Help	16.06	22.20	20.94	.765	30.847	.000
Electronic Cheating	5.60	9.20	7.81	.758	32.072	.000
Unethical Behaviour	12.60	16.76	17.25	.858	16.618	.000

Table 2: Summary Statistics of Three-Group Discriminant Analysis

Overall Model Fit : Canonical Discriminant Function

Function	Eigen Value	Percent of variance		Canonical Correlation	Wilks' Lambda	Chi-Square	df	Sig
		Function %	Cumulative %					
1	.696[a]	95.6	95.6	.641	.571	110.564	16	.000
2	.032[a]	4.4	100.0	.175	.969	6.177	7	.519

a. First 2 canonical discriminant functions were used in the analysis.

Discriminant Function and Classification Function Coefficients

Discriminant Function

Independent Variables	Unstandardized Discriminant Function		Standardized Discriminant Function		Classification Functions		
	Function 1	Function 2	Function 1	Function 2	never	sometimes	always
Plagiarism	.113	.145	.317	.407	.310	.491	.580
Fabrication	-.059	-.086	-.293	-.427	.134	.041	-.013
Deception	-.075	-.024	-.209	-.066	-.095	-.227	-.234
Cheating	-.002	.066	-.010	.422	.138	.124	.173
Sabotage	.179	.185	.525	.541	.087	.382	.490
Outside Help	.093	-.056	.460	-.278	.365	.541	.487
Electronic Cheating	.121	-.322	.335	-.892	.115	.386	.135
Unethical Behaviour	.050	.076	.249	.376	.261	.340	.388

Fisher linear discriminant functions

Table-2 shows the classification functions that are used to assign cases to groups. There is a separate function for each group. The discriminant model assigns the case to the group whose classification function obtained the highest score. The coefficients for plagiarism, cheating, sabotage and unethical behavior are less for never and for sometimes classification, these coefficients has the highest scores for always group, which means that those students who were doing plagiarism, cheating, sabotage and unethical behavior were more likely to be academically dishonesty. The coefficients for fabrication, outside help, and electronic cheating are more likely to in sometimes group and deception is in never category. The standardized coefficients allow you to compare variables measured on different scales. Coefficients with large absolute values communicate to variables with greater discriminating ability.

Table-3 shows correlation of each variable with its each function. The structure matrix shows the correlation of each predictor variable with the discriminant function. The ordering in the structure matrix is the same as that suggested by the tests of equality of group means and is different from that in the standardized coefficients table. This disagreement is likely due to the collinearity between electronic cheating and deception the correlation matrix. The astarike shows *Largest absolute correlation between each variable and any discriminant function. Since the structure matrix is unaffected by collinearity, it's safe to say that this collinearity has inflated the importance of electronic cheating and deception in the standardized coefficients table. Thus, sabotage best discriminates between categories of free-riders (never, sometimes, and always).

Table 3: Structure Matrix

Structure Matrix[a]		
Independent Variables	Function 1	Function 2
Sabotage	.756*	.182
Electronic Cheating	.666*	-.557
Outside Help	.662*	-.219
Plagiarism	.554*	.235
Unethical Behaviour	.485*	.234
Fabrication	.474*	-.167
Cheating	.447*	.202
Deception	.367*	-.057
* Largest absolute correlation between variables		

Assessing Model Fit:

To see that how well the discriminant model as a whole fits the data, here eigenvalues and wilks lambda were used. From table-2 the eigenvalues provides information about the relative worth of each discriminant function. Eigenvalues tells the importance of the function in terms of variation. If the function value is zero it means that there is multicolinearity exists in the data and importance of that variable is reduced. The canonical correlation is equivalent to Pearson's correlation between the discriminant scores and the groups. Wilks' lambda is a measure of how well each function separates cases into groups. Smaller values of Wilks' lambda indicate greater discriminatory ability of the function. So the function 1 through 2 is most important and that is a significant function, its Wilks lambda value is 0.571. The associated chi-square statistic tests the hypothesis that the means of the functions listed are equal across groups. The small significance value indicates that the discriminant function does better than chance at separating the groups. The discriminant function 1 through 2 has a small significant value i.e; 0.000 which indicates that these discriminants function does better than chance at separating the groups.

Checking Homogeneity of Covariance Matrices

For checking homogeneity of covariance matrices here log determinants and Box's M Test of Equality of Covariance Matrices is used, if the Box’s M test is significant and smaller difference is the groups log determinants than covariance matrices is same and need to run discriminant analysis by separate groups covariance’s. And compare the classification table results of both within and separate groups’ covariance’s.

Table-5 shows log determinants of each case. Log determinants are a measure of the variability of the groups. Large differences in log determinants indicate groups that have different covariance matrices. The log determinants of the groups indicate that groups have the same covariance matrices; there are smaller differences in the groups log determinants. Box's M tests the null hypothesis of equal population covariance matrices. This shows that the variances across groups are same. Since Box's M is significant, so we should run separate matrices to see different classification results.

Table 5: Checking Homogeneity of Covariance Matrice

Checking Homogeneity of Covariance Matrice (Within)					**Checking Homogeneity of Covariance Matrice(Separate)**			
Free-riders	Rank	Log Determinant	Box's Test of Equality of Covar. Matrices		Rank	Log Determinant	Box's Test of Equality of Covar. Matrices	
Never	8	18.583	Boxes M	123.81	2	-.222	Boxes M	10.821
Sometimes	8	20.272	F Approx.	1.482	2	.239	F Approx.	1.744
Always	8	17.086	Sig.	.005	2	.552	Sig.	.106

Table 6: Classification Results for Multiple Discriminant Analysis

Predicted Group Membership (Within Group)						**Predicted Group Membership (Separate Group)**			
Sample	Free-Riders	Never	Sometimes	Always	Total	Never	Sometimes	Always	Total
Original Count	Never	109	17	13	139	113	16	10	139
	Sometimes	5	30	14	49	6	32	11	49
	Always	4	6	6	16	4	5	7	16
%	Never	109	12.2	9.4	100.0	81.3	11.5	7.2	100.0
	Sometimes	5	61.2	28.6	100.0	12.2	65.3	22.4	100.0
	Always	4	37.5	37.5	100.0	25.0	31.2	43.8	100.0

The classification table shows the practical results of using the discriminant model. The classification table shows the values and percentage of correctly classified cases. 71.1% of original grouped cases correctly classified within group. Of the cases used to create the model, 6 of the 16 people who are always free-riders

By applying separate group covariance matrix the results shows that there is the large difference in the log determinants values and the Boxs M is insignificant. The classification results are more batter then the within group covariance and this shows that 74.5% cases are correctly classified using the developed discriminant function.

5. CONCLUSION

Our findings suggest that the most significant predictor of academic dishonesty (free-riders) in students of University of Gujrat (Hafiz Hayat Campus) was sabotage. The most important factors are sabotage, electronic cheating and outside help. Results of this paper reveal that sabotage, electronic cheating and outside help are most important variable for prediction of academic dishonesty of University of Gujrat students. In classifying the students at the categories of free-riders, we achieve 74.5% cases are correctly classified using the estimated discriminant function.

REFERENCE

1. Arent, R. (1991). To tell the truth. *Learning,* 19 (6), 72-73.
2. Ameen, E.C., Guffey, D.M. and McMillan, J.J. (1996). Gender differences in determining the ethical sensitivity of future accounting professionals. *Journal of Business Ethics*, 15, 591-597.
3. Caruana, A., Ramaseshan, B. and Ewing, M.T. (2000). The effect of anomie on academic dishonesty among university students. *The International Journal of Educational Management*, 14, 23-30.
4. Diekhoff, G.M., LaBeff, E.E., Clark, R.E., Williams, L.E., Francis, B. and Haines, V.J. (1996). College cheating: Ten years later. *Research in Higher Education*, 37, 487-502.
5. Eastman, J.K., Iyer, R. and Reisenwitz, T.H. (2008). The Impact of Unethical Reasoning on Different Types of Academic Dishonesty: An Exploratory Study. *Journal of College Teaching & Learning,* 5(12), 7-15.
6. Etter, S., Cramer, J.J. and Finn, S. (2006). Origins of Academic Dishonesty: Ethical Orientations and Personality Factors Associated with Attitudes about Cheating with Information Technology. *J. Res. on Tech. in Edu.*, 39(2), 133-155.
7. Eve, R. and Bromley, D.G. (1981). Scholastic dishonesty among college undergraduates: Parallel test of two sociological explanations. *Youth and Society*, 13, 3-22.
8. Graham, M. A. (1994). Cheating at small colleges: An examination of student and faculty attitudes and behaviors. *Journal of College Student Development*, 35, 255-260.
9. Haines, V.J., Kiefhoff, G.M., LaBeff, E.E. and Clark, R. (1986). College cheating: Immaturity, lack of commitment, and neutralizing attitude. *Research in Higher Education*, 25, 342-354.
10. Jendrek, M.P. (1992). Students' reactions to academic dishonesty. *Journal of College Student Development*, 33, 260-273.
11. Lim, V.K.G. and See, S.K.B. (2001). Attitudes Toward, and Intentions to Report, Academic Cheating Among Students in Singapore. *Ethics & Behavior*, 11(3), 261-274.
12. Livosky, M. and Maline, M.B. (1993). Views of cheating among college students and faculty. *Psychology in the Schools,* 31, 72-82.
13. Nadelsom, S. (2007). Academic Misconduct by University Students: Faculty Perception and Responses. *Plagiary: Cross-Disciplinary Studies in Plagiarism, Fabrication and Falsification*, 2(2), 1-10.
14. Newstead, S.E., Franklyn-Stokes, A. and Armstead, P. (1996). Individual differences in student cheating. *Journal of Educational Psychology*, 88, 229-241.
15. Pratt, C. and McLaughlin, G. (1989). An analysis of predictors of students' ethical inclinations. *Research in Higher Education*, 30, 195-219.
16. Roig, M. (1997). Can undergraduate students determine whether text has been plagiarized? *The Psychological Record,* 47, 113-22.
17. Underwood, J. and Szabo, A. (2003). Academic offences and e-learning: individual propensities in cheating. *British Journal of Educational Technology,* 34(4), 467-477.
18. Whitley, B.E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education*, 39, 235-274.
19. Whitley, B.E. (2001). Gender differences in affective responses to having cheated: The mediating role of attitudes. *Ethics & Behavior*, 11, 249-259.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 295-304

EFFECT OF LEARNING APPROACHES ON STUDENT'S ACADEMIC ACHIEVEMENT

Hafsa Inayat[1] and **Zahoor Ahmad**[2]
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: [1]eeta.online@gmail.com
[2]zahoor.ahmed@uog.edu.pk

ABSTRACT

One of the most influential concepts that have been emerged from research into teaching and learning in higher education is "Approaches to Learning" framework that helps to understand how students go about learning in higher education. Three approaches to learning of students: deep, strategic and surface has been identified until now. This study aims at predicting students' CGPA levels, Low (2.00-2.99), Medium (3.00-3.50) and High (3.51-4.00) on the basis of different factors that define three approaches to learning. A sample of almost 200 students has been of UOG. To meet the objectives of this study, neural networks model has been applied.

INTRODUCTION

Approaches to learning are considered as the individual differences in students' intentions when they are faced with a learning task and these reflect the strategies an individual uses to acquire a particular goal [Diseth (2003); Burton and Nelson (2006)]. Student's approaches to learning are one of the most widely used frameworks to understand how students learn in higher education [Ramburuth and Mladenovic (2004); Tight (2003)]. Research into the student's approaches to learning began in the 1970's at the University of Gothenburg in Sweden [Byrne et al. (2009)]. Two approaches to learning were identified initially known as deep and surface approaches to learning, that are clearly related with different levels of understanding achieved [Marton (1975); Marton and Saljo (1976)]. Ramsden (1979) identified another approach to learning: the strategic approach to learning.

Students take different approaches to learning, with the outcomes of learning closely associated with the chosen approaches [Ramsden (2003)]. Prediction of students' performance accurately is worthwhile as it helps in identifying those students who are likely to give poor performance in their studies, and this information can be used in a variety of contexts for example strategic planning can be made for maintaining or improving students' performance during their period of studies in the institutions [Paris, et al. (2010); Ibrahim and Rusli (2007)]. This study focuses on predicting student's academic performance, measured by cumulative grade point average (CGPA).

OBJECTIVES OF THE STUDY

The main objective of this study is the prediction of student's academic performance measured by CGPA on the basis of their learning approaches.

LITERATURE REVIEW

Swanberg and Martinsen (2010) carried out a study in Norway on to investigate the relationship between the five-factor model of personality, approaches to learning and academic achievement. 687 business students participated in this study. Structural equation modeling technique was used to test the proposed model with an objective of testing the relationship between the five-factor model of personality, approaches to learning and studying, and the effects of these variables on academic achievement [RMSEA = .042, Chi-square = 710.441, p = .000, d.f. = 324, NNFI = .938, CFI = .948, GFI = .932].

Byrne, et al. (2009) conducted a comparative study on students studying first course in accounting belonging to United States and Irish University. 204 U.S and 309 Irish students were included in the sample for the study. By using Approaches and Study Skills inventory for Students (ASSIST), their approaches to learning were measured and compared using Mann-Whitney U-test. The results indicated significant difference between learning approaches of the students of both universities. The U.S. students attained significantly higher scores on the deep and strategic scales as compared to the Irish students and scores on surface scale for both groups did not differ significantly.

Papinczak (2009) conducted a study with an aim of exploring whether medical students that can be categorized as having deep and strategic approach to learning find problem-based learning (PBL) pleasant and helpful of their learning, and achieve well in the first-year course. The data was collected from 213 first year medical students on approaches to learning, preferences for different learning environments, self-efficacy, and perceptions of learning within PBL tutorials. A two step cluster analysis identified 5 clusters, out of which 3 clusters represented three approaches to learning (deep, strategic and surface) and two other clusters were unusual combinations. The results showed that the deep and strategic learners appeared to be less exposed to the stresses of PBL in a medical course.

Byrne, et al. (2002) performed a study in Ireland in which ASSIST research instrument was used and the data was collected from 74 Irish and 61 overseas students of BA in European Business (BAEB) program at Dublin City University (DCU). Comparison of approaches to learning of both groups (Irish and Overseas students) was made using independent sample t-test. Results indicated that there was significant difference between two groups on strategic and surface scale that overseas students were inclined towards a strategic approach and Irish students showed a tendency towards surface approach. There was no significant difference on Deep approach between two groups.

Burton and Nelson (2006) performed a research on "*The relationships between personality, approaches to learning, and academic success in first-year psychology distance education students*". The sample consisted of 119 first year psychology distance students from the University of Southern Queensland. Only 97 students could be available and they completed an online questionnaire. ASSIST questionnaire was used to measure the learning approaches and the short form of the International Personality Item Pool (IPIP) [Goldberg (1999), (2001)] was used to measure the Big-Five factors of personality: Extroversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect. Correlation and regression analysis were used to analyze the data obtained. The results showed negative relationship between surface learning approach and students predicted achievement, which was, measured by Grade Point Average. Neither deep nor strategic approach predicted the academic success. Also the results provided support that different learning approaches can be predicted by personality traits.

MATERIALS AND METHODS

Survey methodology:

The population consisted of 6749 students of BS (Hons) (4-year program) and Masters (2-years program) studying in University of Gujrat (Hafiz Hayat Campus) in summer 2010. Six academic blocks of University of Gujrat were taken as clusters in which total number of students was 6749 and from these six blocks three blocks: Social science, basic Science and Engineering block were selected at random. The sample contained 200 students.

Data collection instrument:

For this study, "Approaches and Study Skills Inventory for Students" or simply the ASSIST research instrument is used to collect data from students. This instrument contains 52-items, all on 5-point likert-scale that measure student's approaches to learning in three dimensions: deep, strategic and surface based on 13 sub-scales. Along with the section of demographic information of students, CGPA and academic skills (Analytical, communication, presentation, confidence and critical thinking ability) were added in the questionnaire. A description of approaches to learning and their subscales is outlined in table 1.

Table 1: ASSIST – Approaches to learning subscales and their characteristics

Approaches to learning & subscales	Meaning
• Deep Approach	
Seeking meaning	Intention to understand
Relating ideas	Relating to other topics and courses
Use of evidence	Relating evidence to conclusions
Related Motive	
Interest in ideas	Interest in learning for learning's sake
• Strategic Approach	
Organized studying	Able to work regularly and effectively
Time Management	Organize time and distribute effort to greatest effect
Alertness to assessment demands	Being alert to cues regarding the assessment
Related Motive	
Achieving	Competitive and confident
Monitoring effectiveness	Checking progress to ensure achievement of aims
• Surface Approach	
Lack of purpose	Lack of direction
Unrelated memorizing	Not understanding material and relying on memory
Syllabus-boundness	Relying on lecturers to define learning tasks
Related Motive	
Fear of failure	Pessimism and anxiety about academic outcomes

Source: Byrne, et al. (2002, 2009)

DATA ANALYSIS TECHNIQUE

1. **Confirmatory Factor Analysis:**
   A series of relationships that suggest how measured variables represent a construct not measured directly is known as measurement theory. Confirmatory factor analysis (CFA) is used to test a proposed measurement theory. This

technique enables us to test how well the measured variables represent the construct.

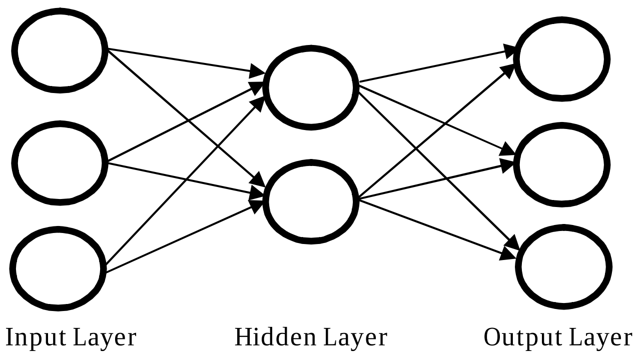
2. **Artificial Neural Network:**
   An artificial neural network is a computer-intensive, algorithmic procedure for transforming inputs into desired outputs using highly connected networks of relatively simple processing units (neurons or nodes). Neural networks are modeled after the neural activity in the human brain.

In statistical applications, the computing units are arranged in a series of layers with connections between nodes in different layers, but not between nodes in the same layer. The layer receiving the initial inputs is called the input layer. The final layer is called the output layer. Any layers between the input and output layers are called hidden layers.

The Multilayer Perceptron Neural Network Model

Multilayer perception stands for a neural network with one or more hidden layer (s) between the input and output layers. Figure-1 illustrates a Perceptron network with three layers:

Figure-1



This network has an input layer (on the left) with three neurons, one hidden layer (in the middle) with two neurons and an output layer (on the right) with three neurons. There is one neuron in the input layer for each predictor variable.

Activation Functions:

The activation function "links" the weighted sums of units in a layer to the values of units in the succeeding layer.

Artificial neural networks (ANNS) are objective and efficient classification methods. Thus, they have been successfully used in a large numbers of classification fields [Feng and Hong (2008)].

RESULTS AND DISCUSSION

Descriptive statistics:

The average age of the respondents of this study is 21.09 years with a standard deviation of 1.49. Average CGPA is obtained to be 3.06. 53% of the respondents are females and 47% are males. 57% students are from BS-Hons program, and 41% are of Masters program.

Validity and reliability of the instrument:

Cronbach alpha value for 52-items of ASSIST was obtained to be .896 indicating high internal reliability. Richardson (1994) has stated that when a questionnaire is to be employed in a situation different from that in which it was actually developed; factor analysis should always be performed to guarantee its intended constituent (an essential part) structure so that it can be reconstructed in a new context. Confirmatory factor analysis was performed for validation of ASSIST for the present study. All of the subscales were confirmed except “Fear of failure” and “Alertness to assessment demands”, and therefore these subscales are not used in the further analysis. Table 2 represents the criterion for a factor to be confirmed and values for each subscale along with the status that whether a subscale is confirmed or not.

Table 2:

Criteria of confirmation	$\chi^2/_{d.f} \leq 3$	RMSEA ≤ .08	GFI ≥ .90	AGFI ≥ .90	Status
Scales					
• **Deep Approach**					
Seeking meaning	1.3767	.0439	.993	.965	Confirmed
Relating ideas	.52408	.000	.997	.987	Confirmed
Use of evidence	1.7876	.0663	.991	.953	Confirmed
Related Motive					
Interest in ideas	.95013	.000	.995	.975	Confirmed
• **Strategic Approach**					
Organized studying	2.127295	.0746	.989	.947	Confirmed
Time Management	1.157515	.029	.994	.971	Confirmed
Alertness to assessment demands	11.5076	.219	.950	.749	Not Confirmed
Related Motive					
Achieving	.137732	.000	.999	.997	Confirmed
Monitoring effectiveness	.891835	.000	.995	.977	Confirmed
• **Surface Approach**					
Lack of purpose	.354	.000	.998	.991	Confirmed
Unrelated memorizing	.202207	.000	.999	.995	Confirmed
Syllabus-boundness	1.229425	.000	.994	.970	Confirmed
Related Motive					
Fear of failure	4.7358	.142	.975	.877	Not Confirmed

Prediction of Categories of CGPA:

Three categories of CGPA are to be predicted that are: Low (2.00-2.99), Medium (3.00-3.50) and High (3.51-4.00) using Neural Network Multilayer perceptron analysis technique.

The information of the input and output variables used to develop the model for predicting categories of CGPA along with the number of units in each (input, hidden and output) layer and the activation functions used is displayed in Table 3. There are 11 covariates in input layer that are used for the prediction of Low medium and high categories of CGPA. There is only 1 hidden layer having 8 units in it. Output layer has one dependent variable, i.e. categories of CGPA and number of units in output layer are 3 that are Low, Medium and High categories of CGPA.

Figure-2 illustrates the three layered neural network model graphically, in which the left most layer is the input layer having 11 units (that are the independent variables, subscales of the learning approaches). Middle layer is the hidden layer having 8 units and the right most layer is the output layer, having 3 units that are the low, medium and high categories of the dependent i.e. CGPA. The colored lines connecting the units of input, hidden and output layers have different meanings. Grey lines show that the weights assigned to a particular relation are positive. Blue lines are the indication of negative weights.

Table 3: Network Information

Input Layer	**Hidden Layer**	**Output Layer**
1. Lack of purpose 2. Unrelated Memorizing 3. Syllabus Boundness 4. Organized Studying 5. Time Management 6. Achieving 7. Monitoring effectiveness 8. Seeking Meaning 9. Relating Ideas 10. Use of Evidence 11. Interest in Ideas	1 hidden layer Number of Unit in Hidden layers = 8	Low, Medium, High categories of CGPA Number of Units in Output Layer = 3
Activation Function:	Sigmoid	Sigmoid

Table 4 is the classification table of the neural networks that elaborates how the neural network model has classified the cases correctly and incorrectly into the Low, Medium and High categories of CGPA. The model has correctly classified 93.8%, 94.6% and 69.0% cases for Low, Medium and High categories of CGPA respectively, and overall correctly classification is 89.8% which is considerably a good correct classification percentage.

Figure 3 is the graphical representation of the classification table that is predicted-by-observed chart of the categories of CGPA. It plots the observed vs. predicted pseudo-probability of categories. The left most box plot is for the correct classification for Low category of CGPA. 5th box plot from the left is for the correct classification of Medium category and similarly the last box plot is for the High category of CGPA. The portion of the boxplot above the 0.5 mark on the y axis represents correct predictions shown in the classification table. The portion below the 0.5 mark represents incorrect predictions. And this model has predicted medium category more correctly followed by low and high categories.

Table 4: Classification

Sample	Observed	Predicted			
		Low	Medium	High	Percent Correct
Training	Low	60	1	3	93.8%
	Medium	3	70	1	94.6%
	High	3	6	20	69.0%
	Overall Percent	39.5%	46.1%	14.4%	89.8%

Dependent Variable: CGPA_cat

Figure 1 is the graphical representation of the classification table that is predicted-by-observed chart of the categories of CGPA. It plots the observed vs. predicted pseudo-probability of categories. The left most box plot is for the correct classification for Low category of CGPA. 5th box plot from the left is for the correct classification of Medium category and similarly the last box plot is for the High category of CGPA. The portion of the boxplot above the 0.5 mark on the y axis represents correct predictions shown in the classification table. The portion below the 0.5 mark represents incorrect predictions. And this model has predicted medium category more correctly followed by low and high categories.

Figure 1

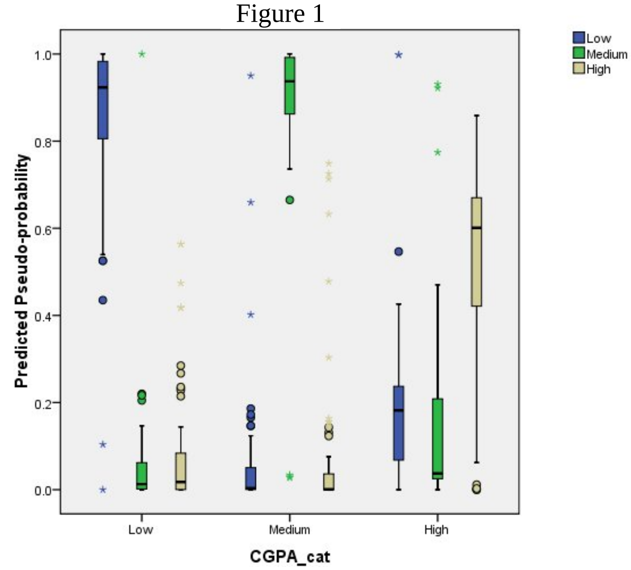


Figure 2 is the Receiver Operating Characteristics (ROC) curve that provides a visual display of the **Sensitivity** (is the probability that a "positive" case is correctly classified) and **Specificity** (is the probability that a "negative" case is correctly classified.) for all possible cut offs in a single plot. This curve shows the trade-off between these two. Three curves are in the ROC curve each for a category of the dependent variable. Area under each curve is demonstrated in table-5. Highest area under the curve is for the medium category i.e. .971 followed by Low category having .964 and then High category, the area under which is .872. The highest area is under the curve of that category for which the model has the highest percentage of correct classification.

Table-5 Area Under the Curve

		Area
CGPA_cat	Low	.964
	Medium	.971
	High	.872

Figure 2

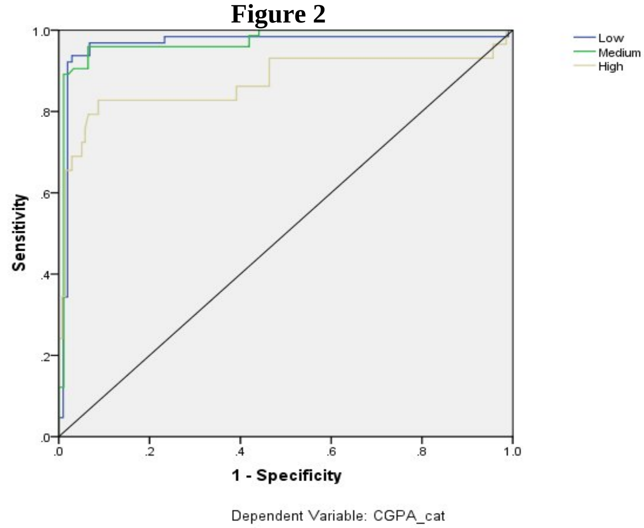
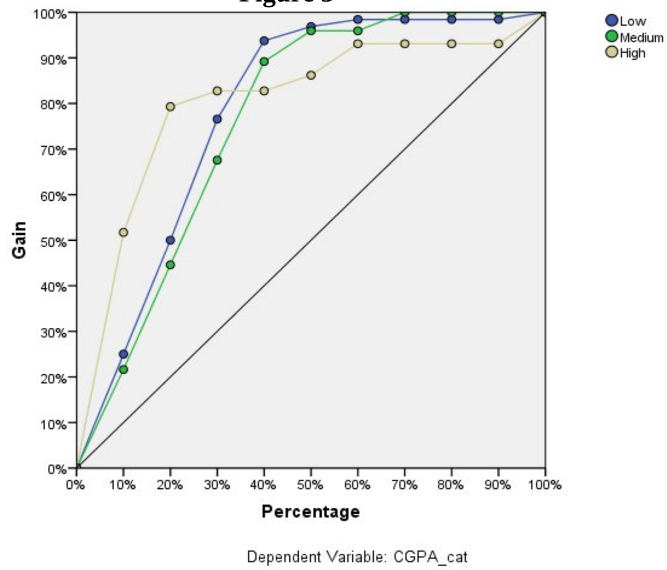


Figure 3



re-3

The cumulative gains chart shown in figure-3 is the percentage of the overall number of cases in a given category "gained" by targeting a percentage of the total number of cases. For example, the first point on the curve for the Low category is at (10%, 25%), meaning that if we score a dataset with the network and sort all of the cases by predicted pseudo-probability of Low, we would expect the top 10% to contain approximately 25% of all of the cases that actually take the category of Low CGPA. Similarly for High category of CGPA the first point is at (10%, 52%), means that if the data set is scored with the network and all the cases are sorted by predicted pseudo-probability of High CGPA, the top 10% are expected to contain approximately 52% of the cases that actually have the category of High CGPA.

Table 6: Independent Variable Importance

	Importance	Normalized Importance
Lack_of_purpose	.099	92.8%
Unrelated Memorizing	.099	93.2%
Syllabus Boundness	.082	77.3%
Organized Studying	.093	86.9%
Time Management	.089	83.5%
Achieving	.107	100.0%
Monitoring effectiveness	.077	71.9%
Seeking Meaning	.082	77.2%
Relating Ideas	.084	79.0%
Use of Evidence	.095	88.8%
Interest in Ideas	.093	87.4%

The importance of the independent variables for the prediction of the dependent is shown in table-6 and the figure followed by, figure-4, graphically represents the normalized importance of each predictor in ascending order. The importance of an independent variable is a measure of how much the network's model-predicted value changes for different values of the independent variable. Normalized importance is simply the importance values divided by the largest importance values and expressed as percentages.

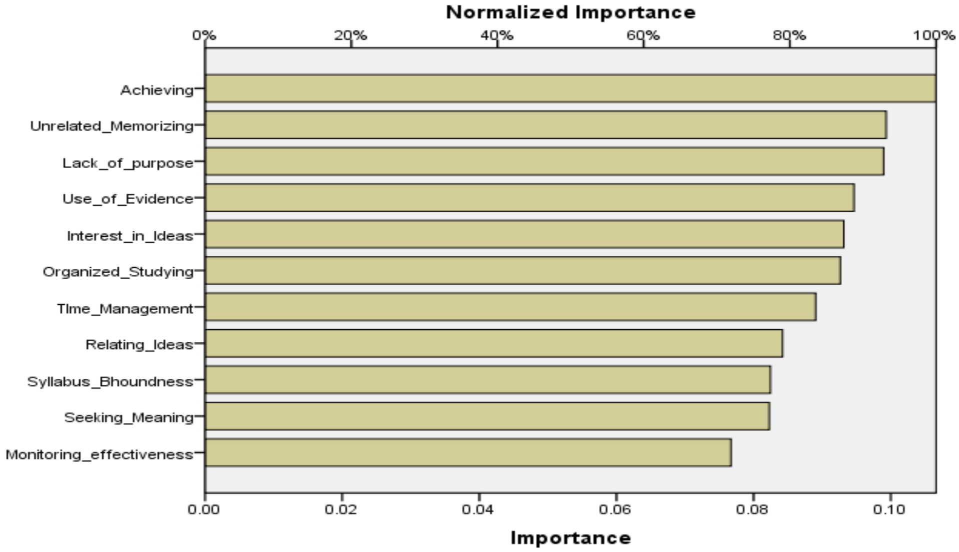
CONCLUSION

The instrument used in the present study is validated using confirmatory factor analysis. Only two subscales: Fear of Failure and Alertness to assessment Demand did not produce reasonable goodness of fit indices, rest of the subscales were confirmed and were best fitted. The neural network model that has been developed to predict the Low, Medium and High categories of CGPA, performs well that is has a correct classification percentage of 89.8%, which is quite good.

LIMITATIONS AND RECOMMENDATIONS

The data collected comprised of students from all disciplines at University of Gujrat Hafiz Hayat campus. Research on approaches to learning of individual disciplines can also be accomplished in order to have good generalizations of the results for students in each discipline. Also, inclusion of other factors that are likely to effect students' academic performance along with the factors of approaches to learning is recommended so that accurate prediction of students' academic performance can be made.

Figure 4



ACKNOWLEDGEMENT

A bundle of thanks and gratefulness is for Miss Erum Shehzadi and my friend Marwa Anwar whose motivation, moral support, and affection made me able to complete this research paper in critical circumstances.

REFERENCES

1. Burton, L.J. and Nelson, L.J. (2006). The relationships between personality, approaches to learning, and academic success in first-year psychology distance education students. *Paper presented in Higher Education Research and Development Society of Australasia (HERDSA) Annual Conference*, Perth, Australia.
2. Byrne, M., Flood, B. and Willis, P. (2002). Approaches to Learning of European Business Students. *Journal of Further and Higher Education,* 26(1), 19-28.
3. Byrne, M., Flood, B. and Willis, P. (2009). An Inter-Institutional Exploration of the Learning Approaches of Students Studying Accounting. *International Journal of Teaching and Learning in Higher Education*, 20(2), 155-167.
4. Diseth, A. (2003). Personality and approaches to learning as predictors of academic achievement. *European Journal of Personality,* 17, 143-155.

5. Feng, L. and Hong, W. (2008). Classification error of multilayer perceptron neural networks. *Neural Computing & Applications*, 18(4), 377-380.
6. Goldberg, L.R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt & F. Ostendorf. (Eds.). *Personality Psychology in Europe*, 7, 7-28.
7. Goldberg, L.R. (2001). *International Personality Item Pool*. Retrieved from http://ipip.ori.org/
8. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. (2006). *Multivariate Data Analysis*, Sixth Edition. Pearson Education, Inc.
9. Ibrahim, Z. and Rusli, D. (2007). *Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression.* 21st Annual SAS Malaysia Forum, 5th September 2007.
10. Johnson, R.A. and Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis.* Pearson Education, Inc.
11. Marton, F. (1975). On non-verbatim learning: I. Level of processing and level of outcome. *Scandinavian Journal of Psychology* 16 (1), 273-79.
12. Marton, F. and Saljo, R. (1976). On qualitative differences in learning: I Outcome and process. *British Journal of Educational Psychology* 46 (1), 4-11.
13. Papinczak, T. (2009). Are deep strategic learners better suited to PBL? A preliminary study. *Adv in Health Sci. Edu.*, 14, 337-353.
14. Paris, I.H.M., Affendey, L.S. and Mustapha, N. (2010). Improving Academic Performance Prediction using Voting Technique in Data Mining. *World Academy of Science, Engineering and Technology*, 62, 820-823.
15. Ramburuth, P. and Mladenovic, R. (2004). Exploring the relationship between students' orientations to learning, the structure of students' learning outcomes and subsequent academic performance. *Accounting Education: An International Journal*, 13(4), 507-527.
16. Ramsden, P. (1979). Student learning and perceptions of the academic environment. *Higher Education*, 8, 411-427.
17. Ramsden, P. (2003). *Learning to teach in higher education*. London: Routledge Falmer.
18. Richardson, J.T.E. (1994). Using questionnaires to evaluate student learning, in: G. GIBBS (Ed.) *Improving Student Learning through Assessment and Evaluation* (Oxford, The Oxford Centre for Staff and Learning Development).
19. Swanberg, A.B. and Martinsen, Ø.L. (2010). Personality, approaches to learning and achievement. *Educational Psychology*, 30 (1), 75-88.
20. Tight, M. (2003). *Researching higher education*. Maidenhead, Berkshire, United Kingdom: McGraw Hill.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 305-312

ACADEMIC PERFORMANCE OF ATHLETE AND NON-ATHLETE COLLEGE STUDENTS: A STATISTICAL STUDY

Safoora Samuel[1] and **Saleha Naghmi Habibullah**[2]
Department of Statistics, Kinnaird College for Women, Lahore, Pakistan
Email: [1]thefinch_717@hotmail.com; [2]salehahabibullah@hotmail.com

ABSTRACT

To predict the academic performance of college athletes and non-athletes has been a topic of interest in the literature over the past few decades. According to Parmer (1994) aspirations of playing professional sports have been linked to the failure of some athletes to balance academic and athletic tasks. According to Gaston (2003), all of the athletic demands leave little time to interact socially or participate in other campus activities. A student athlete who has the desire to pursue a career in athletics or a challenging degree program can easily become overwhelmed by athletic demands. As a result, less emphasis, time, and energy is placed on academic related tasks. The purpose of the current study is to examine the relationship between gender, athletic participation and academic performance (Divisions/GPA's) in FA/FSc and BA/BSc student athletes and non-athletes. Male and female non-athlete students and student-athletes from all basic sports have been surveyed regarding the academic performance of athlete and non-athlete college students. The collected data has been analyzed using univariate, bivariate and multivariate techniques included visual depictions such as pie charts and indicators such as the odds ratio. The statistical package SPSS 16 has been used for purposes of data-analysis. Implications for future research are also discussed.

1. INTRODUCTION

Athletics is a collection of sporting events that involve competitive running, jumping, throwing, and walking. It is also known as track and field or track and field athletics. The word is derived from the Greek word "athlos" meaning "contest". The simplicity of the competitions, and the lack of a need for expensive equipment, makes athletics one of the most commonly competed sports in the world. (Wikipedia, the free encyclopedia, Athletics – Definition (http://www.wordiq.com/definition/Athletics)).

Dating back to the Ancient Greeks, athletics was the only competition to be held in the first Olympic Games which took place in Athens in 776 BC. At that time, the single athletic event was known as the 'stade,' a foot race which covered the length of the Athenian Olympic stadium. The Olympic Games continued to take place in Athens every four years. Schools began to organise competitive meets, the earliest of which took place in Exeter College, Oxford in 1850. The first modern Olympic Games took place in 1896 and athletics were the key element of the games. In 1928 further progressions were made when women were allowed to partake in athletics competitions for the first time. (Wikipedia, the free encyclopedia, History (http://www.talk-athletics.co.uk/guides/history_of_athletics.html)).

For many college students, the experience of becoming a student athlete is a dream comes true. However, a college sport is a huge commitment that will affect many aspects of your life. College athletes deal with many opportunities, experiences, and even sometimes downfalls. (Wikipedia, the free encyclopedia, Athletics – sport (http://en.wikipedia.org/wiki/Athletics_(sport))).

There are many different advantages and disadvantages of being a college athlete. Advantages include: (Directessays, Nzuahp, 5th May 2010, Rockler-Gladen, N.)

- **Part of team:** A general advantage of being a college athlete is that you get to be part of a team, which means new teammates, and new friends. Teammates are probably the first friends at college. It's always nice to meet new people when you first come to college.
- **Fit and healthy life:** Another general advantage of being a college athlete is that by playing a sport, one will stay fit and healthy. Athletes stay in shape by running, working hard on the field, and also by lifting weights. By staying fit at a young age in life, will be helpful later on in life.
- **Profession, Publicity and Wealth:** First and foremost, some students practice sports as a profession. For instance, athletics is one of the highest paid occupations in the world today. This has influenced many students who excel in various sports such as: football, Hockey, basket ball, Cricket just to name a few, for publicity and for wealth.

Disadvantages comprise:

- **Effect on Studies:** Students who do a lot of sports have small amount of time to concentrate on their course work. Some students put in much of their time in doing too much sport and spend less time in studying, and as a result, they end up performing very poor in their various courses.
- **Imitation:** In addition, there are some students who do sports not because they are very good at doing sports, but because they see others doing so, therefore they want to imitate others which may not be very worthwhile.

Today there are numerous events which combine to make up the sport of athletics. The most common types of athletics competitions are track and field, road running, cross country running, and race walking. Throughout this paper, we are going to discuss many different issues (academics/athletics) that college athletes deal with as compared to college non-athletes.

2. LITERATURE REVIEW

Melendez (2006-2007) examined the relationship between race/ethnicity, gender, athletic participation, and college adjustment in 207 freshmen and sophomore college student athletes and non-athletes. All participants completed questionnaires assessing background demographics, athletic status, and college adjustment. Findings revealed gender and athletic status were significantly related to college adjustment. In addition, gender and race/ethnicity significantly interacted, and were related to scores on college adjustment.

Aries, et al. (2004) investigated that student-athletes were studied over 4 years at a highly selective liberal arts college and an Ivy League university. Students spending 10 or more hours per week in athletic activities had lower entering academic credentials and academic self-assessments than non-athletes, but the academic performance of athletes was not below what would be expected based on their entering profiles. Athletes surpassed non-athletes on sociability/extraversion and self-reported well-being in each annual wave of the study. Athletes were not isolated from the rest of the student body; they spent over 50% of their time with non-group members and belonged to non-athletic extracurricular groups every year. Athletes perceived group membership to pose greater difficulties to academic performance and involvements outside the group than did members of other types of extracurricular groups. Athletes drank more heavily on weekends that non-athletes, but did not differ in growth or well-being. Comparisons by athletic status were similar for men and women.

Joy L. (2004) examined the utility of academic and athletic motivation as a key variable in predicting academic performance among 211 college athletes at a Division I institution in the Midwest so therefore academic motivation as a predictor of academic performance for college athletes has been debated in the literature. After controlling for background characteristics, results revealed that ACT score, ethnicity, and academic motivation were significant in the regression model.

Whitley and Pressley (1995) determined if there is a difference in the educational performance of high school athletes and non-athletes in the North Carolina high schools. Through the voluntary compliance of 133 (44%) of the 301 member schools of the North Carolina High School Athletic Association who were eligible to participate in this study, quantitative data were gathered on 126,700 students in grades 9 through 12 from across the state for the 1994-95 school year. The grade point averages (GPAs), attendance rates, discipline referral rates, dropout rates, and graduation rates for the athletes and non-athletes from the schools who participated in the study were compared. Paired t-tests were used to determine if there were any differences in grade point averages and attendance between: athletes and non-athletes. Paired t-tests were used rather than independent t-tests, because of the dependent measures within each school were correlated. The results of this study led to the conclusion that those students who participated in interscholastic athletics in the North Carolina high schools surveyed in this study outperformed their non-athlete counterparts in the high school setting. This finding was consistent with the related literature on other studies conducted relative to this topic.

Adler and Adler (1985) examined the relationship between athletic participation and academic performance among athletes involved in big-time college sports. Drawing on four years of participant observation of a major college basketball program, they traced athletes' involvement in academics throughout their college careers. Studies showed that, contrary to popular belief, most athletes enter college with optimistic and idealistic goals and attitudes about their impending academic careers. However, their athletic, social, and classroom experiences lead them to become progressively detached from academics. They concluded that the structure of universities with big-time athletic programs and the athletes' patterned experiences within these universities undermine their attainment of the professed goals of the educational system.

3. RATIONALE OF THE STUDY

This research study intends to examine the (Divisions/GPA’s) of FA/FSc and BA/BSc athletes in comparison to FA/FSc and BA/BSc non-athletes with population of roughly 150 students and the factors that influence college athlete success because it is difficult to balance both academics and athletics. So the study focused on the academic performance of college athletes as compared with the academic performance of non-athletes in the college.

4. OBJECTIVES OF THE STUDY

To investigate the relationship between gender, athletic participation and academic performance by comparing male and female, to access which causes are associated to gender and athletes/non-athletes, to observe which gender or athlete/non-athlete is likely to perform academically good/bad than the other one and as well as to identify and determine the prominent factors that are related to academic performance among both athletes and non-athletes.

5. RESEARCH METHODOLOGY

The study design used in this research is an observational cross-sectional study. The study has been conducted by taking responses from the two institutes of Lahore (Kinnaird College for Women and FC College University). A semi-structured questionnaire consisting of 13 questions was designed and was administered on 150 students selected randomly.

The questionnaire included questions on a number of variables such as age, gender, study year, the extent to which the respondent participated in sports, whether the respondent was an athlete or non-athlete, the respondent's academic performance as depicted by Divisions/GPA's, the number of hours spent by the respondent on his/her studies after college hours, whether or not the respondent missed classes, academic honors received by the respondent (if any), the respondent's preference vis a vis study career/sports career, the number of hours spent on sports, primary sport the respondent played or was interested in and athletic honors received by the respondent (if any).

The respondents included FA/FSc and BA/BSc students in the following variety of sports: Hockey, Cricket, Open tennis, Table tennis, football, basketball, volleyball, swimming and badminton. Statistical analysis carried out using programs available in SPSS version 16.0 including univariate, bivariate, and multivariate analyses.

6. RESULTS

Analysis of the collected data revealed that the respondents were in the age-group 16-24, and that 70% of them were male and 30% female. The respondents included students of F.A/F.Sc. as well as B.A/B.Sc.,

To check the association between variables, Pearson's Chi-square test / Fisher's exact test was applied. Table 1 presents the p-values pertaining to the Fisher's Exact test applied to test the association of key variables with gender.

Variables found to be associated with gender are:

- to what extent do you participate in sports (p=0.050)
- How many hours you spend on sports (p=0.002),
- Primary sport you play or interested (p=0.000), and
- Athletic honors (p=0.030).

Similarly, Fisher's Exact test applied to test the association of key variables with the question "Are you an athlete/non-athlete?". Variables found to be associated with the variable athlete/non-athlete are:

- to what extent do you participate in sports (p=.000),
- what do you prefer, study career, sports career or both ? (p=0.000),
- how many hours you spend on sports (p=0.000),
- primary sport you play or interested (p=0.000) and
- athletic honors (p=0.000).

Table 1: Association of key variables with Gender

Variables	Fisher's exact test P-value
To what extent do you participate in sports?	.050
Are you athlete/non-athlete?	.722
Divisions/GPA's	.414
How many hours you spend on your studies after college hours?	.796
Do you miss any classes?	.399
Academic honors?	.620
What do you prefer?	.430
How many hours you spend on sports?	.002
Primary sport you play or interested?	.000
Athletic honors?	.030

The **Odds Ratio** was calculated to observe the effect of athletics on the academic performance of males as compared to female students and athletes as compared to non-athletes. To obtain this measure, the scale of key variables was changed to binary response. Table 2 presents the odds ratio with reference to gender.

Table 2: Odds Ratio (Gender)

Variables	Odds Ratio
To what extent do you participate in sports?	2.818
Are you athlete/non-athlete?	1.180
Divisions/GPA's	0.655
How many hours you spend on your studies after college hours?	0.859
Do you miss any classes?	1.543
Academic honors?	1.395
What do you prefer?	0.784
How many hours you spend on sports?	1.041
Primary sport you play or interested?	1.462
Athletic honors?	0.847

From the observed data, it can be said that:

- For the variable "To what extent do you participate in sports?" the effect of athletics on the academic performance of female students is 2.818 times more than those of male students,
- For the variable (Divisions/GPA's) athletics affects less, almost .655 times on the academic performance of female students as compared with those of male students, and
- For the variable "How many hours you spend on sports?" the effect of athletics on the academic performance is equally likely 1.041 times on the academic performance of female students as compared with those of male students.

Similarly, the odds ratio was computed with reference to athletes/non-athletes. On the basis of this computation, it can be said that:

- for the first variable that the effect of athletics on the academic performance of athlete students is .357 times less than those of non-athlete students,
- for the variable (Divisions/GPA's) athletics affects less almost .700 times on the academic performance of athlete students than those of non-athlete students,

- for the variable (how many hours you spend on your studies after college hours) odds show that the effect of athletics on the academic performance of athletes is 1.684 times more than those of non-athlete students and
- the odds of last variable (athletic honors) also shows that the athletics affects 15.960 times more on the academic performance of athletes than those of non-athlete students.

MULTIVARIATE ANALYSIS

Factor analysis was applied to find the prominent factors which may affect the academic performance of male and female student athletes as compared to non-athlete students. Requirement of minimum sample size of 150 recommended by Tabachnick and Fidell (1996) was met and there were no missing values. Another requirement of the factor analysis is that some of the correlations between the variables should be 0.30 or more. In this analysis most of the correlations were 0.30 or greater, so this requirement was also met. Bartlett's Test of Sphericity is used to test the null hypothesis that the correlation matrix is an identity matrix. This hypothesis ought to be rejected in order to attempt the factor analysis. KMO is the measure of sampling adequacy it tests the amount of variance that could be explained by the factors within the data. In this case KMO's MSA is 0.699, which is larger than the limit 0.50 said to be mediocre by Kaiser (1970). Bartlett's Test of Sphericity, the probability associated with Bartlett's Test of Sphericity (χ^2 (df = 78) = 530.491, $p < 0.000$) is less than the level of significance (0.05) so sufficient number of correlations exist between the variables.

The communality measures the percent of variance in a given variable explained by all the factors jointly and may be interpreted as the reliability of the indicator. If the communality exceeds 1.0, there is a spurious solution, which may reflect too small a sample or the researcher has too many or too few factors. As seen in Table 3, 0.715 i.e. 72% variation was explained by the variable "To what extent do you participate in sports?" Similarly 0.788 i.e. 79% of the variance was explained by "Are you athlete/non-athlete?", 0.643 i.e. 64% by "Primary sport you play or interested?" and 0.571 i.e. 57% variation was explained by the last variable "Athletic honors". The initial eigenvalues, extraction sums of squared loadings and rotation sums of squared loadings showed that four components have an eigenvalue greater than one. These 4 components explained 61.717% of the data variation. In rotation sums of squared loadings the percentage of variance, eigenvalues and cumulative percentage of variance explained by all the components were changed after varimax rotation.

Table 3: Communalities

	Initial	Extraction
Age	1.000	.835
Gender	1.000	.559
Study year	1.000	.868
To what extent do you participate in sports?	1.000	.715
Are you athlete/non-athlete?	1.000	.788
Divisions/GPA's	1.000	.293
How many hours you spend on your studies after college hours?	1.000	.676
Do you miss any classes?	1.000	.452
Academic honors?	1.000	.499
What do you prefer?	1.000	.565
How many hours you spend on sports?	1.000	.561
Primary sport you play or interested?	1.000	.643
Athletic honors?	1.000	.571

The scree plot graphs the eigenvalues against the number of components. As one moves to the right, toward later components, the eigenvalues drop. When the drop ceases and the curve makes an elbow toward less steep decline, scree plot says to drop all further components after the one starting the elbow. This rule is sometimes criticized for being amenable to researcher-controlled “fudging” because the curve can be a smooth curve or it can have multiple elbows. Therefore we will retain all factors with eginvalues greater than 1 as there is more than one elbow and also there are less than 30 variables and average communality is greater than 0.6 (Kaiser’s criterion). From the scree plot in Figure 1, it is clear that first four components have eigenvalues greater than 1 so they are selected.

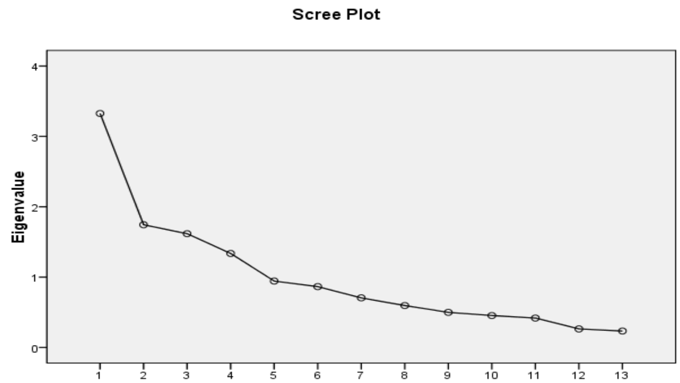


Fig. 1: Scree Plot

Thus from the rotated component matrix four factors were obtained; Factor 1 comprised of 5 variables i.e. Are you athlete/non-athlete, to what extent do you participate in sports, athletic honors, what do you prefer and how many hours you spend on sports. Factor 2 comprised of 2 variables i.e. study year and Age. Factor 3 comprised of 4 variables i.e. how many hours you spend on your studies after college hours, academic honors, do you miss any classes and Divisions/GPA's. Factor 4 comprised of 2 variables i.e. primary sport you play or interested and gender.

DISCUSSION AND CONCLUSION

The study presented in this paper highlights the factors contributing to academic performance of athletes and non-athletes in the two educational institutes of Lahore. The purpose of the study was to assess the academic performance of athlete and non-athlete college students (comparing male, female students). Based on a thorough review of the literature, the study was comprised of different constructs such as the age group of athletes and non-athletes, their study year, their preference in terms of career, does athletics effect academic performance of athletes as compared to non-athletes, their primary interest of sport, study hours and hours spent on sports, how frequently student athletes miss their classes as compared to non-athletes and academic/athletics honors.

Analysis of the collected data revealed that the primary sports that students play or are interested in are basketball, cricket, football and swimming. Boys’ primary interest lies in basketball and cricket whereas females’ primary interest is in basketball and swimming. Whereas the ratio of winning athletic honors is obviously higher among athletes, the ratio of achieving academic honors is the same in both athlete/non-athlete students. The odds ratio with reference to gender shows that the effect of athletics is more on the academic performance of female students as compared to male students and the odds ratio with reference to athletes/non-athletes shows that athletics affects almost equally likely on the academic performance of athlete and non-athlete students.

To conclude, athletics does not affect much the academic performance of athletes, their academic achievements (Divisions), activities (study hours, academic honors) are almost equally likely to non-athlete students and the success of such balanced academic performance and integration depends on more than one factor.

In this study the findings are opposite to the generally accepted statement that the academic performance of athlete students is not as good as the non-athletes but the results of this study support many of the findings of previous studies.

RECOMMENDATIONS

The practical implication of this research is that; being athlete or non-athlete, participation in the sports should be encouraged and that the athletic program/activities should be continued in the educational institutes among both male and female students. Further research can be conducted on the relatively large sample size and similar studies can be carried out in other big cities of Pakistan such as Karachi, Islamabad and Peshawar.

REFERENCES

1. Melendez, M.C. (2006-2007). The influence of athletic participation on the college adjustment of freshmen and sophomore student athletes. *Journal of College Student Retention: Research, Theory and Practice*, 8(1) 39-55.
2. Aries E., McCarthy D., Salovey P. and Banaji M.R. (2004). A Comparison of Athletes and Non-Athletes at Highly Selective Colleges: Academic Performance and Personal Development. *Research in Higher Education*, 45(6) 577-602.
3. Joy L., G.G. (2004). Examining Academic and Athletic Motivation among Student Athletes at a Division I University. *Journal of College Student Development*, 45(1) 75-83.
4. Adler P. and Adler P.A. (1985). The Academic Performance of College Athletes. *Sociology of Education*, 58(4) 241-250.
5. Whitley R.L. and Pressley J.S. (1995). *A Comparison of the Educational Performances of Athletes and Nonathletes in 133 North Carolina High Schools*.
6. *Wikipedia*, the free encyclopedia, Athletics (sport), http://en.wikipedia.org/wiki/Athletics_(sport), retrieved on 24th Jan 2011
7. *Wikipedia*, the free encyclopedia, Athletics – Definition, http://www.wordiq.com/definition/Athletics, retrieved on 24th Jan 2011
8. *Wikipedia*, the free encyclopedia, History, http://www.talk-athletics.co.uk/guides/history_of_athletics.html, retrieved on 22nd Jan 2011.
9. *Directessays*, Disadvantages and Advantages of Being a College Athlete, http://www.directessays.com/viewpaper/90164.html, retrieved on 22nd Jan 2011.
10. *Nzuahp*, Advantages and Disadvantages of doing sports, http://www.english-test.net/forum/ftopic54443.html, Issue date 5th May 2010, retrieved on 25th Jan 2011.
11. Rockler-Gladen, N., *Should You Become a Student Athlete? Advantages and Disadvantages of College Sports Participation*. http://www.suite101.com/content/should-you-become-a-student-athlete-a105576, Issue date 5th Apr 2009, retrieved on 25th Jan 2011.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 313-320

DETERMINANTS OF MONTHLY INFLATION IN PAKISTAN

Mian Muhammad Irfan[1] and **Ahmed Nadeem**[2]
Statistics & DWH Department, State Bank of Pakistan, Karachi, Pakistan
Email: [1]muhammad.irfan@sbpo.org.pk; [2]ahmed.nadeem@sbpo.org.pk

ABSTRACT

This study is an approach to find out the determinants of monthly inflation in Pakistan. Several studies are available for finding out the determinants of inflation. The monthly data on inflation, money supply, exchange rates, imports, wheat support prices, government revenues and large scale manufacturing index from August, 1991 to August, 2010 have been used for analysis. A linear model using multiple regressions has been developed to capture most of the factors of inflation in Pakistan which might be useful in formulation and implementation of monetary policy in Pakistan. In the short run, the most important determinants of inflation are expectations of inflation, money supply, wheat support price, increase in government revenue and seasonal factors. There doesn't exist a meaningful long-run relationship among variables.

1. INTRODUCTION

A persistent increase in the average price level in the economy is called inflation. An average inflation rate of 12 percent in Pakistan during the period of 1970-80 has typically been below double digit levels with the exception of two brief periods following the 1973 and 1979 due to oil price shocks. Unlike the earlier experience, inflation in the 1990s was not only higher and more persistent; its' causes were also less readily understood. The present yearly rate of inflation 11.7 percent for FY10 in Pakistan could be explained in terms of factors such as low rate of output growth, monetary expansion, higher dollar price of imports, exchange rate depreciation, increase in excise and sales taxes, and changes in administrative prices such as fuel prices, utility charges and procurement price of wheat. While cost-push factors such as increase in the price of fuel, can have temporary effect on the general level of prices, these effects cannot be sustained without an accommodating monetary policy. The inflationary impact of the depreciation of the exchange rate can similarly be regarded as an indirect effect of an escalation of money supply. Thus money supply would appear to be a key determinant of inflation in an economy. It's therefore, surprising that some of the recent studies on inflation attribute a minor role to monetary growth as an explanation of the recent inflation in Pakistan.

Control of inflation should be a matter of priority due to a number of reasons. It is important from the point of view of poverty alleviation and social justice. Inflation is a regressive form of taxation and among the most vulnerable to the inflation tax are the poor and fixed income groups. Inflation also causes relative price distortion as some prices adjust more slowly than others. Another form of distortion takes place during inflationary periods when absolute price changes are mistaken for relative price changes. These distortions cause efficiency losses and lower the productive base of the economy.

Inflation can discourage savings if the rate of return on savings does not reflect the increase in the level of prices. The uncertainty about future prices can cause unexpected gains and losses in trade and industry and thus discourage long term contracts and investments. Resources are likely to be channeled into less productive activities like speculation in stock exchange and real estate. Some of these speculative activities, such as speculation on the exchange rate, can have serious macroeconomic consequences.

Four different price indices are published in Pakistan: the consumer price index (CPI) captures the movement in prices of the urban workers; the whole sale price index (WPI) provides an early signal of the trend in prices, the sensitive price index (SPI) reflects the movement in prices of the consumption basket of low income employees and the GDP deflator. In most countries including Pakistan, the main focus for assessing inflationary trends is placed on the CPI because it most closely represents the cost of living.

2. REVIEW OF LITERATURE

Cottarelli *et al* (1998) have explained inflation performance in a sample of industrial and transition economies by looking at policymakers' incentives to inflate the economy, and the perceived cost of disinflation. They have found a significant effect of fiscal deficits on inflation, particularly (but not exclusively) in countries where the government securities market have not well developed. Other factors that have affected inflation significantly were relative price changes, central bank independence, the exchange rate regime, and the degree of price liberalization; there has been only limited evidence that other structural factors, such as those influencing the natural rate of unemployment, have a significant effect on inflation.

Durevall (1998) has developed an error-correction model with the aim of analyzing the behavior of prices during a period of chronic inflation in Brazil during 1968 to 1985. The degree of inflationary inertia has been estimated, and tested for the importance of disequilibrium the domestic-money, foreign-goods, labor, and domestic goods markets on inflation have been carried out.

Ericsson (1998) has examined several central issues in the empirical modeling of money demand. These issues include economic theory, data measurement, parameter constancy, the opportunity cost of holding money, cointegration, model specification, exogeneity and inferences for policy. Review of these issues at a general level is paralleled by discussion of specific empirical applications, including some new results on the demand for narrow money in the United Kingdom.

Liu and Adedeji (2000) have established a framework for analyzing the major determinants of inflation in Islamic Republic of Iran. They have used an empirical model by taking into consideration disequilibrium in the markets for money, foreign exchange and goods. Results have strongly support the need for a sustained prudent monetary policy in order to reduce inflation and stabilized the foreign exchange market. The estimation has shown that an access money supply generates an increase in the rate of inflation that, in turn intensifies asset substitution (from money to foreign exchange), thereby weakening real demand for money and exerting pressures on the foreign exchange market. The study has also found that a permanent rise in income tends to increase the real demand for money and reduces inflation in the long run.

Hahn (2003) has investigated the pass-through of external shocks, i.e. oil price shocks, exchange rate shocks, and non-oil import price shocks to euro area inflation at different stages of distributions (import prices, producer prices and consumer prices). The analysis was based on VAR model that includes the distribution chain of pricing. According to their results the pass-through was largest and forecast for non-oil import price shocks, followed by exchange rate chocks and oil price shocks. The size and the speed of the pass through of theses shocks declined along the distribution chain. External shocks explained a large fraction of the variance in all price indices. They seemed to have contributed largely to inflation in the euro area since the start of the European Monetary Union. The results on the size and the speed of the pass-through in the euro area appeared to be robust over time and different identification schemes.

Diouf (2007) has investigated that how consumer price inflation has determined in Mali for 1979–2006 along three macroeconomic explanations: (1) monetarist theories, emphasizing the impact of excess money supply, (2) the structuralist hypothesis, stressing the impact of supply-side constraints, and (3) external theories, describing the effects of foreign transmission mechanisms on a small open economy. The analysis has made use of cointegration techniques and general-to-specific modeling. Average national rainfall, and to a lesser extent deviations from monetary and external sector equilibrium have found to be the main long-run determinants of inflation. The paper has offered policy recommendations for controlling inflation in Mali.

Khan *et al* (2007) have used an econometric framework for the identification of the main determinants of recent inflation trends in Pakistan. They have used the data from the 1972-73 to 2005-06 period, applied ordinary least square method and verified results through Breusch-Godfrey Serial Correlation LM and Augmented Dickey-Fuller tests. They have found that the most important determinants of inflation in 2005-06 were adaptive expectations, private sector credit and rising import prices. However, the fiscal policy's contribution to inflation was minimal.

Kandil and Morsy (2009) have studied determinants of inflation in GCC, using an empirical model that included domestic and external factors. Inflation in major trading partners appeared to be the most relevant foreign factor. In addition, oil revenues have reinforced inflationary pressures through growth of credit and aggregate spending. In the short-run, binding capacity constraints also explained higher inflation given increased government spending. Nonetheless, by targeting supply-side bottlenecks, the increase in government spending has eased capacity constraints and would ultimately help to moderate price inflation.

3. MATERIALS AND METHODS

3.1 Data and Data Sources

The following monthly variables from August, 1991 to August, 2010 have been used for analysis

CPI = Consumer Price Index (2000-01=100)

WPI = Wholesale Price Index (2000-01=100)

SPI = Combined Sensitive Price Index (2000-01=100)

M2 = Money Supply (Million Rupees)

ER= Pak. Rupees per US$
M = Imports (Million US$)
W = Wheat Support Price (Pak. Rupees)
T = Federal Board of Revenue Tax Collection (Million Rupees)
LSM = Large scale manufacturing index (2000-01=100)
S = Seasonal factor generated through X-12 ARIMA

Price Indices, Wheat Support Prices and Imports: Federal Bureau of Statistics “Monthly Statistical Bulletin”, website: http://www.statpak.gov.pk

Money Supply, Exchange Rate and Federal Board of Revenue Tax Collection: State Bank of Pakistan “Monthly Statistical Bulletin”, website: http://www.sbp.org.pk

3.2 Multiple Regression Time Series Models

We have used the following model for modeling of inflation in Pakistan

$$y_t = a_1 + b_1 x_{1t} + \ldots + b_n x_{nt} + \varepsilon_t$$

where y_t is dependent variable, a_1 is an intercept, *b’s* are regression coefficients, *x’s* are explanatory variables and ε_t is error term and $\varepsilon_t \sim N(0, \sigma^2)$.

Consequently $y_t \sim N(\mu, \sigma_y{}^2)$

and $b_t = \left(x_t{}' x_t\right)^{-1} x_t{}' y_t$

Method of least squares has been used to estimate the parameters.

4. RESULTS AND DISCUSSION

The correlation matrix is as follows:

Correlation Between Variables at Levels

Variable	CPI	WPI	SPI	M2	Exchange Rates	Imports	Wheat support price	FBR Tax Collection	LSM
CPI	1.00	1.00	1.00	0.98	0.94	0.86	0.97	0.90	0.87
WPI	1.00	1.00	1.00	0.98	0.93	0.88	0.97	0.90	0.87
SPI	1.00	1.00	1.00	0.98	0.93	0.87	0.97	0.90	0.86
M2	0.98	0.98	0.98	1.00	0.89	0.92	0.95	0.91	0.92
Exchange Rates	0.94	0.93	0.93	0.89	1.00	0.71	0.88	0.81	0.79
Imports	0.86	0.88	0.87	0.92	0.71	1.00	0.84	0.86	0.90
Wheat support price	0.97	0.97	0.97	0.95	0.88	0.84	1.00	0.87	0.80
FBR Tax Collection	0.90	0.90	0.90	0.91	0.81	0.86	0.87	1.00	0.83
LSM	0.87	0.87	0.86	0.92	0.79	0.90	0.80	0.83	1.00

All the variables are stationary at 1st difference with log.

Test of Stationary of Variables in Log-Levels

Variable	Phillips Perron Test		Augmented Dickey-Fuller Test		Critical Value
	Log Level	First Difference	Log Level	First Difference	
CPI	0.65	-12.28	0.76	-5.40	-2.87
WPI	0.56	-10.37	0.58	-10.37	-2.87
SPI	0.90	-10.99	1.01	-10.72	-2.87
M2	-1.16	-17.03	0.47	-3.30	-2.87
Exchange Rates	-1.02	-9.65	-1.08	-9.63	-2.87
Imports	-1.25	-30.82	-0.23	-4.30	-2.87
Wheat support price	0.07	-15.59	-0.06	-15.50	-2.87
FBR Tax Collection	-2.35	-118.33	0.71	-5.44	-2.87
LSM	-1.94	-19.31	0.15	-7.03	-2.87

Tests includes intercept

Critical value at the 5 percent level based on MacKinnon (1996).

The multiple regression model is used to find the determinants of inflation in Pakistan. Several models have been tried for determinants of inflation. Significance of exchange rates, imports, large scale manufacturing have not been found for determinants of inflation.

The estimated regression model using ordinary least square is as follows:

Multiple Regression Model for CPI at Log Difference

Variable	Coefficient	Standard Error	T-Statistic	P-Value
Constant	-0.3141	0.1191	-2.638	0.0090
CPI_{t-11}	0.1912	0.0623	3.071	0.0024
WPI_{t-1}	0.1425	0.0356	4.004	0.0001
SPI_{t-3}	0.1273	0.0436	2.921	0.0039
$M2_{t-10}$	0.0567	0.0292	1.941	0.0536
W_{t-10}	0.0353	0.0089	3.964	0.0001
T_{t+2}	0.0047	0.0016	3.005	0.0030
S_t	0.3167	0.1191	2.659	0.0084

Important Statistics

Statistic	Value	Statistic	Value
R-squared	0.2971	Mean dependent variable	0.0071
Adjusted R-squared	0.2733	S.D. dependent variable	0.0073
S.E. of regression	0.0062	Akaike info criterion	-7.2936
Sum squared residuals	0.0079	Schwarz criterion	-7.1682
Log likelihood	792.0627	Hannan-Quinn criterion	-7.2429
F-statistic	12.4978	Durbin-Watson statistic	2.0637
Prob. (F-statistic)	0.0000		

Consumer Price Index (CPI) is most widely used for measure of inflation in Pakistan and is also important for formulation and implementation of monetary policy. However, the State Bank of Pakistan keep into consideration of core inflation which in non-food, non-food-non-oil and trimmed inflation for formulation of monetary policy. The model showed that the higher inflation in preceding year develops expectations for hike in future inflation and it's highly significant at 1% level of significance as also observed by Khan *et al* (2007).

The increase in wholesale prices is immediately transmitted into consumer prices which ultimately increase general level in the country and it has highly significant affect on CPI at 1% level of significance. The price hike in essential consumer items which is measured by sensitive price indices is translated in general level on quarterly basis and has significant impact on CPI at 1% level of significance. The State Bank of Pakistan may also consider other price indices rather than depending only on CPI.

The money supply which is measured through M2 is key factor for increase in prices. The impacts of loose and tight monetary policy are not immediately transmitted into increase in price level. The model showed that M2 has significant effect on prices at 5% level of significant at 10 months lag. Khan and Axel (2006), using monthly data from January 1998 to June 2005, also conclude that the lagged growth of private sector credit and lagged growth of money supply (M2) are two significant causes of inflation in Pakistan in recent years.

The increase in wheat support prices one way increases the income level of the rural population especially farmers which creates the demand of goods and services ultimately increasing the price level. On the hand, the wheat is a basic necessity used as food for the human population, the consumer have to pay more to satisfy their basic needs. The impact of increase in wheat support prices is not immediately transmitted into increase in price level. It has been found that it takes usually ten months to transmit into price increase. It might be due to that increase in income of farmers is not immediately created demand in the market or most of the population in rural areas has already enough stocks of wheat for six months. It has also significant impact on CPI at 1% level of significance. The wheat support price has also been identified as an important determinant of inflation in Pakistan by Khan and Qasim (1996) and Hasan et al (1995).

It has been observed that the Federal Board of Revenue tax collection has also importance for determinants for inflation. It may be interpreted in different way that the shortfall in revenue collection is translated into fiscal deficit and creating pressure of Govt. borrowings from banking system untimely curding out and increase in prices. It has also a significant impact on CPI at 1% level of significance after seven months lag.

Seasonal factors for CPI are obtained by using X-12 ARIMA and used as a regressor. The CPI is highly affected with season at significant at 1% level of significance. The strong seasonality has been found from July to January of each fiscal year.

In the selected model, the R square is 0.29 and DW statistic is 1.99 which is close to 2.0 showing that there is now serial correlation in the selected model of CPI.

The stationary of residuals is shown below:

Stationary Tests of Residuals

Augmented Dickey-Fuller Test of Residuals*	None	Intercept	Intercept and Trend
T-Statistic	-14.34	-14.31	-14.27
Probability	0.0000	0.0000	0.0000

*Optimal lag length is selected by SIC

The residuals are stationary at 1% level of significance.

We have tried to find out the long run relationship has been found in the variables as the all the variables of I(1). For testing co-integration, we have used trace test and maximum Eigen test. The tests show that there are five co-integration equations.

Co-integration Test

Hypothesized No. of CE(s)	Eigen value	Trace Test		Maximum Eigen Test	
		Statistic	Critical Value	Statistic	Critical Value
None *	0.63	449.96	95.75	220.61	40.08
At most 1 *	0.30	229.36	69.82	77.98	33.88
At most 2 *	0.22	151.38	47.86	54.81	27.58
At most 3 *	0.17	96.57	29.80	41.87	21.13
At most 4 *	0.15	54.70	15.49	37.40	14.26
At most 5 *	0.07	17.30	3.84	17.30	3.84

Critical value at the 5 percent level based on MacKinnon (1996).
Co-integration Test includes seasonal factor as exogenous variable and other variables at 1^{st} log difference.

No single long run relationship exists to interpret the results.

5. CONCLUSIONS

The model found that the most important determinants of monthly inflation in Pakistan are expectations of inflation, exchange rate, imports, wheat support price, money supply and expected shortfall in government revenue. The State Bank of Pakistan may intervene in the foreign exchange market to stabilize the exchange rates but it depends on the foreign exchange reserves level. The steady and natural increase in exchange rates might be useful to increase the exports but abrupt increase in exchange rates might be harmful for payments of external debt, imports and foreign direct investment. The policy makers may devise a strict policy to reduce the imports of luxuries which will save the precious foreign exchange of the country and to stabilize the foreign exchange market. The wheat support price must be increase under strong grounds of increase in prices of agriculture inputs such as fertilizer, agriculture machinery, pesticides, seeds, labor, oil prices, shortage of water etc. rather than as political decision to have the sympathies of the rural population. The increase in money supply is also important factor for inflation. M2 may be kept at optimum level which will ultimately reduce the general price level in the country. The hoarding, corruption, speculations etc. are also important causes of inflation which might be controlled through administrative measures, new legislation through parliament and strong independent judicial system. The shortage of energy, production of energy using furnace oil and pass through of international oil prices have also increased overall cost of production ultimately increasing prices. The policy makers may develop long-term policies for cheap production of energy such construction of new dams and production of energy through coal, solid waste, wind and nuclear technology. A reasonable rate of inflation, around 3 to 6 percent for Pakistan (Khan, 2005 and Hussain, 2005) is often viewed to have positive effects on the economy, since it encourages investment and production and allows growth in wages.

REFERENCES

1. Cottarelli, Carlo, Mark Griffiths, and Reza Moghadama (1998). The Monetary and Nonmonetary Determinants of Inflation: A Panel Data Study. *IMF Working Paper* 98/23, (Washington: International Monetary Fund).
2. Durevall, D. (1998). The Dynamics of Chronic Inflation in Brazil, 1968-1985. *Journal of Business & Economic Statistics*, 16(4), 423-432.
3. Diouf, M. (2007). Modeling Inflation for Mali. *IMF Working Paper* 07/295, (Washington: International Monetary Fund).
4. Hahn, E. (2003). Pass-through of external shocks to euro area inflation. *European Central Bank, working paper* 243. http://www.ecb.int
5. Hasan et al. (1995). What Explains the Current High Rate of Inflation in Pakistan? *The Pakistan Development Review*, 34: 4 Part III (Winter 1995) 927-943.
6. Hussain, Manzoor, (2005). Inflation and Growth: Estimation of Threshold Point for Pakistan, *Pakistan Business Review*, October.
7. Kandil, M. and H. Morsy (2009). Determinants of Inflation in GCC. *IMF Working Paper* 09/82, (Washington: International Monetary Fund).
8. Khan, Ashfaque H., and Mohammad Ali Qasim (1996). Inflation in Pakistan Revisited, *The Pakistan Development Review*, 35(4), Part II, 747–59.
9. Khan, Mohsin (2005). *Inflation and Growth in MCD Countries*, Mimeo. International Monetary Fund.
10. Khan et al. (2007). *Determinants of recent inflation in Pakistan*. Social Policy and Development Centre.
11. Liu, O., and Olumuyiwa Adedeji (2000) Determinants of Inflation in the Islamic Republic of Iran-A Macroeconomic Analysis. *IMF Working Paper* 00/127, (Washington: International Monetary Fund).

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 321-326

CRITICAL REVIEW OF ISLAMIC BANKING INDUSTRY IN PAKISTAN

Muhammad Mazhar Manzoor[1], Asad Ali Siddiqui[2], Sanober Sheikh[3], Aisha Bashir Shah[3] and **Mehmood Qasim[4]**

[1] Department of Business Administration, Federal Urdu University of Science and Technology, Karachi, Pakistan. Email: mmazher@gmail.com

[2] Federal Urdu University, Islamabad, Pakistan

[3] Department of Business Admin., University of Sindh, Jamshoro, Pakistan

[4] University of Karachi, Karachi, Pakistan

ABSTRACT

As Islamic banking and finance is still growing and emerging field of finance as compare to that of its counterpart i.e. conventional finance, but its growth and market share is drastically positive and rapid one in both type of countries such as Muslims and non Muslim countries. This paper is a attempt to critically analyze the growth of Islamic banking industry in Pakistan especially. Furthermore, it examines six year data i.e. from 2003 to 2008, which highlight the true growth of Islamic banking industry in Pakistan. Secondary data shows ample progress of IB in Pakistan especially the population of urban areas are highly motivated due to several religious , Social and economical factors.

INTRODUCTION

As at end of the year 2003 only one bank operated as a full-fledged Islamic bank (Meezan Bank) and three conventional banks were operating Islamic banking branches. Today there are 6 full fledge licensed Islamic banks (IBs) and 12 conventional banks have licenses to operate dedicated Islamic banking branches (IBBs). All of the five big banks in Pakistan are providing Islamic banking services. The total assets of the Islamic banking industry are over Rs. 225 billion as of 30th June, 2008 which accounts for a market share of 4.5% of total banking industry assets. The market share of deposits stands at 4.2%. Total branch network of the industry comprises of more than 330 branches with presence in over 50 cities and towns covering all the four provinces of the country and AJK.

It is also important to compare progress of Islamic Banking in Pakistan with the progress in other Muslim countries in order to define an expected path for our industry. Malaysia’s first Islamic Bank commenced operations in 1983 and the market share of the Malaysian Islamic banking system today stands at approximately 13%. The Malaysia’s official target for the Islamic Banking is 20% market share by 2015. Bahrain’s Islamic banking system is said to have gained a share of around 8% in over 30 years. Similarly in Indonesia Islamic banking was introduced somewhere in mid 90’s. According to Bank Indonesia, Islamic banking assets reached in March, 2007 to 1.67 percent of total assets as compared with 0.7 percent in 2003 and currently the deposits stand at 1.69% and financing at 2.60% of the total banking sector.

For the Islamic Banking Institutions preferred modes of financing belong to the low risk category i.e. those which usually generate a fixed return. Among these, murabahah (cost plus mutually agreed profit margin) and ijarah (leasing) are the most attractive and popular modes of financing, However, lately the share of diminishing musharakah has also grown at a rapid pace leading towards diversification of the Islamic banks' financing portfolio. Furthermore Sukuks are also gaining popularity in exponential fashion.

STRATEGIES OF ISLAMIC BANKING INDUSTRY OF PAKISTAN

Pakistan has several comparative and competitive advantages in the Islamic Banking space. We have a large population of 160 million people with 97% population being Muslim. This provides a huge domestic market base, second only to Indonesia. Given the fact that our Banking Companies Ordinance, 1962 (BCO) has been amended to suit Islamic Banking principles and concepts; there is a fairly well developed regulatory and legal framework to underpin our current developments. We also have an established institution in the form of National Institute of Banking and Finance (NIBAF) through which a good quality Islamic Banking Certification Course is being offered. Work is also underway towards building a specialised institution for imparting education and training in Islamic Banking. A comprehensive operational shari'ah compliance mechanism is in place which allows us a base to further strengthen our shari'ah compliance framework. Above all Pakistan's Islamic banking offering has been accepted by all the stakeholders as evidenced by an issueless launch of the recent initiative and the consequent growth in the market. This means that we now have the winning combination and a base that can be used for future development of the Islamic banking industry. Pakistan's Islamic banking offering has also received international acceptance as evidenced by the oversubscription of Sukuk floated in the international market and the number of international players/investors which applied for Islamic banking licenses in the country including multinational banks. This enables well for establishing Pakistan as a good destination for the belief sensitive Foreign Direct Investment, especially from the Middle East and Far East.

Five Pillars of Islamic banking strategy

To achieve the stated objectives and to capitalize on the strengths, SBP's strategy for promotion of Islamic Banking in Pakistan focusing on five pillars, which are:

1. Extension of outreach - both breadth and depth.
2. Strong shari'ah compliance mechanism
3. Robust Regulatory framework able to accommodate the unique aspects of Islamic banking transactions.
4. Capacity building through human resource development
5. Internal and external relations.

INDUSTRY PROGRESS AND MARKET SHARE OF ISLAMIC BANKING INDUSTRY OF PAKISTAN

The Islamic banking industry continued its progress during the quarter March-June 2008. This has resulted in increasing their share of assets in the overall banking system by 0.2 percentage points (pp) to 4.3% during the quarter. The growth in IBIs is also reflected in increased share of Islamic banking deposits, and financing and investment that stood at 4.0% and 4.1% respectively at the end of quarter. Given the small base the progress seems quite impressive.

The Islamic banks have also continued their efforts to extend the outreach of Islamic banking. Specifically, the number of branches reached 326 from 289 branches since December 2007. Though the pace of branch extension seems impressive, IBIs have largely failed to penetrate in the rural areas. The agreement on extension of rural branch network and increasing focus on SME and Microfinance is available in the form of Islamic Banking Strategic Plan that was chalked out by taking extensive feedback and buy-in from the IBIs. Now IBIs should work on extending rural branch network and catering to the needs of financially excluded segments of the economy, there is no real efforts made by IBIs in this direction.

Table 1:
Industry Progress & Market Share of Islamic Banking Industry of Pakistan

Description	Rs in Billion					
	Dec. 03	Dec. 04	Dec. 05	Dec. 06	Dec. 07	Jun. 08
Total Assets	13	44	71	119	206	235
% of Banking Industry	0.5%	1.5%	2.0%	2.8%	4.0%	
Deposits	8	30	50	84	147	169
% of Banking Industry	0.4%	1.3%	1.8%	2.6%	3.8%	
Financing and Investment	10	30	48	73	138	166
% of Banking Industry	0.5%	1.3%	1.7%	2.3%	3.5%	
Full Fledge Islamic Banks	1	2	2	4	6	6
Branches of Islamic Banks	10	23	37	93	186	
Conventional Banks with Islamic Banks Branches	3	9	9	12	12	12
Branches of Conventional Banks	7	25	33	57	103	
Total Islamic banking Institutions	4	11	11	16	18	
Total Number of Branches	17	48	70	150	289	326

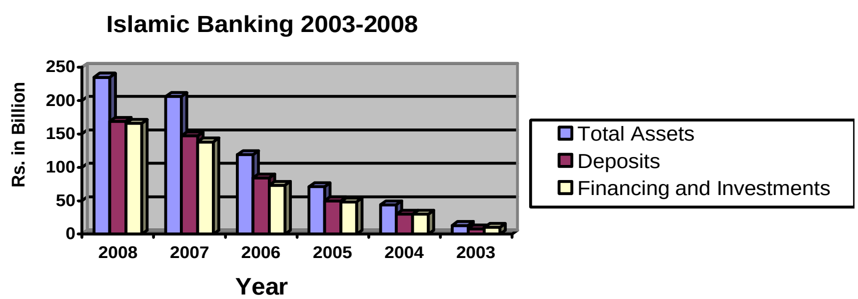


Figure 1. Graph showing Islamic Banking Status from 2003 to 2008.

DATA ANALYSIS AND DISCUSSION

Since its secondary data base research, out of several resources ample data was taken from State bank of Pakistan’s hand book showing how many Islamic financial institutions are working in Pakistan and which financial instruments these banks are using for financing which are helping in elimination of Riba. As well as these data also show the growth of Islamic banks till Sep, 2009. Islam was the basis of creation of an independent state within the undivided Indo-Pak Sub-Continent. Since its creation, the people of Pakistan have held the demand for elimination of Riba from the financial system of Pakistan on the basis of Islamic precepts. All Constitutions of Pakistan have incorporated, within the principles of policy, the elimination of Riba as an important objective of the State policy.

Since Pakistan started with an approach to convert the whole system into Islamic one, a number of amendments in relevant laws were introduced providing legal cover for Islamic financial products and services. Similarly, some new laws were introduced to allow new financial institutions or facilitate the existing ones.

Description	**Dec. 03**	**Dec. 04**	**Dec. 05**	**Dec. 06**	**Dec. 07**	**Dec. 08**
Total Assets	13	44	72	118	276	276
%age of Banking Industry	0.5%	1.4%	2.1%	2.9%	4.2%	4.9%
Deposit	8	30	50	83	202	202
%age of Banking Industry	0.4%	1.2%	1.9%	2.8%	4.1%	4.8%
Financing& Investment	10	30	48	72	186	187
%age of Banking Industry	0.5%	1.3%	1.8%	2.4%	3.6%	4.4%
Conventional bank with Islamic bank branches	3	7	9	12	12	12
No. of branches (including sub branches)	17	48	70	150	289	514

The legal and regulatory infrastructure developed during that era has proved to be invaluable asset as we keep on charting the present landscape of the industry today on the same. Islamic Banking Industry of Pakistan continued its progress during the year 2008. They have increased their share of assets in the overall banking system to 4.9% up to December 2008. The growth is also reflected in increased share of deposits and financing & investment that stood at 4.8% and 4.4% respectively at the end of Dec 2008.

Table 3: Industry Progress and market share

Rupees in billion & industry share in percent								
	Sep. 09*	**Jun. 09***	**Dec. 08**	**Dec. 07**	**Dec. 06**	**Dec. 05**	**Dec. 04**	**Dec. 03**
Total Assets	323	313	276	206	119	72	44	13
Share in industry	5.3	*5.1*	*4.9*	*4.0*	*2.8*	*2.0*	*1.5*	*0.5*
Deposits	245	238	202	147	84	50	30	8
Share in industry	5.5	*5.2*	*4.8*	*3.8*	*2.6*	*1.8*	*1.3*	*0.4*
Net Financing & Investment	198	194	186	138	73	48	30	10
Share in industry	4.2	*4.2*	*4.3*	*3.5*	*2.3*	*1.7*	*1.3*	*0.5*
Total Islamic Banking Institutions	19	18	18	18	16	11	11	4
Total No. Of Branches**	551	528	515	289	150	70	48	17

**Source: Annual Accounts except for June and Sept 09, data for which is based on Un-audited Quarterly Accounts*

Currently, there are 6 full-fledge banks and 12 conventional Banks are offering Islamic Banking products through their Islamic Banking Branches. Hence, branch network of IBIs comprises of around 514 branches as on Dec 31, 2008.

Table 4: Islamic Banking Branch Network as on Sep'30, 2009

Type	Bank Name	No. of Branches
Full Fledge Islamic Banks	Albaraka Islamic Bank B.S.C. (E.C.)	30
	BankIslami Pakistan Limited	70
	Dawood Islamic Bank Limited	25
	Dubai Islamic Bank Pakistan Limited	24
	Emirates Global Islamic Bank Limited	42
	Meezan Bank Limited	140
	Sub-total	**331**
Islamic Bank Branches of Conventional Banks	Askari Bank Limited	22
	Bank AL Habib Limited	6
	Bank Alfalah Limited	48
	Habib Bank Limited	1
	Habib Metropolitan Bank Limited	4
	MCB Bank Limited	11
	National Bank of Pakistan	8
	Soneri Bank Limited	6
	Standard Chartered Bank(Pakistan)	11
	The Bank of Khyber	16
	The Royal Bank of Scotland	3
	United Bank Limited	5
	Faysal Bank Limited	1
	Sub-Total	**142**
Sub Branches	Askari Bank Limited	2
	BankIslami Pakistan Limited	32
	Dawood Islamic Bank Limited	6
	Dubai Islamic Bank Pakistan Limited	2
	Meezan Bank Limited	35
	Emirates Global Islamic Bank Ltd	1
	Sub Branches Total	**78**
	Grand Total	**551**

CONCLUSION

The Islamic financing system replaces the concept of riba with profit and loss sharing. There are many arguments that say that Islamic banking is not Islamic it is just merely change of name to attract people who want to live their life in Islamic way and afraid of riba, but as I studied different Islamic modes of financing I came to a conclusion that these methods follows Islamic principles and are riba free only the problem is that they follow the KIBOR and LIBOR as a benchmarks which needed to be replace by some Islamic benchmark, but only because of the use of these benchmarks we can't say these methods are not Islamic methods. These methods are introduced after many investigation

by a shariah board and complies Islamic modes by adopting these methods by Islamic banks riba can be eliminated from the banking sector and consequently from the economy only there is a need of public awareness and from government sector to establish Islamic institutions that offer Islamic products all over the Muslim countries.
Furthermore , we can conclude that till now the data which is available for analysis shows that Islamic banking is turning into a powerful financial remedy which continuously increasing growth rate and market share.

REFERENCES

1. Alam S.M. (2008). Riba or Interest: Islam Prohibits in absolute terms. *Pakistan and Gulf Economist*, 43-44.
2. Alam, S.M. (2009). Islamic Injunctions against usury (Riba). *Pakistan and Gulf Economist*, 24.
3. Aziz Tayyebi (2008). *Discussion paper: Islamic Finance: An Ethical Alternatives to conventional Finance*? ACCA (The Association of Chartered Certified Accountants).
4. DTZ Research (2008). *Islamic Finance and real state*.
5. Shamshad Akhter (2007). *Pakistan Islamic Banking: Past, Present and Future Outlook*. Speech Governor State Bank of Pakistan.
6. El-Gamal, Mahmoud Amin (2006). Over view of Islamic Finance. Department of Treasury, Office of International Affairs, *Occasional Paper* No. 4, June 2006.
7. El-Gamal, Mahmoud Amin (2000). *A Basic Guide to Contemporary Islamic Banking and Finance*.
8. El Oorch Muhammad (2005). Islamic Finance gear up. *Finance and development a quarterly magazine of the IMF* December, 42(4).
9. El Gulam, M.Amin "*Permissible financing method*".
10. *Frequently ask Question on Islamic Banking*, State Bank of Pakistan, Islamic Banking Department Karachi
11. Hasan, Zubair (2002). Mudaraba as a Mode of Financing in Islamic Banking: Theory, Practice and Problem. *Munich Personal RePEc Archive*.
12. *Handbook of Islamic Product and Services*, Islamic Banking Department, State Bank of Pakistan.
13. *Hand Book on Islamic SME Financing*, Islamic Banking Department, State Bank of Pakistan.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 327-342

CLASSROOM ACADEMIC PERFORMANCE BASED ON MOTIVATIONAL AND SELF REGULATING LEARNING FACTORS

Memoona Khalid[1] and **Zahoor Ahmad**[2]
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: [1]mamoona.stat008@gmail.com; [2]zahoor.ahmed@uog.edu.pk

ABSTRACT

In every educational institution, success is measured by academic performance. Intelligence and education is not the only key to success, much effort is made to identify, evaluate, and encourage the progress of students. The motivation and self-regulated learning of students is an essential issue in higher education, particularly due to significance of academic performance in their professional life. In this research we have studied the effect of motivational and self regulating factors on class room academic performance. We consider two main independent factors motivation and self Regulation & Learning Abilities and then we consider sub-factors of motivation that are self efficacy, intrinsic value and test anxiety, and sub-factors for self regulation are cognitive strategy, self regulation. The main purpose is to examine the relationship between motivational and self-regulated learning components Also we want to study that which factor is playing a role for clusters of below average and above average performers. For this study a sample of 200 students of University of Gujrat was selected using cluster sampling technique and Motivated Strategies for Learning Questionnaire [**Pintrich, et al. (1991)]** is used for data collection purpose. Reliability of the data is checked before the data analysis. Firstly, confirmatory factor analysis is used to confirm the factors; Correlation Analysis, and Two Step Cluster Analysis are used for data analysis purpose. The results show that there is a significant relationship between the three motivational components and self-regulated learning components. The results of cluster analysis showed that there are four clusters of students with respect to academic performance on the basis of considered independent factors.

1. INTRODUCTION

In every educational institution, success is measured by academic performance. The performance of students in universities should not concern only to the educators and administrators, but also to corporations in the labour market because the core assets of the universities are the students. Intelligence and education is not the only key to success, much effort is made to identify, evaluate, and encourage the progress of students.

In classroom setting academic performance is determined by the demonstration of declarative and procedural knowledge after having engaged in several other complex and distracted tasks: that is, the knowledge was recently attained through many different and difficult tasks that occur within and outside of the classroom [Hezlett, et al. (2004)].

Academic performance refers to how students deal with their studies and how they achieve different tasks given to them by their teachers. Academic performance is the skill to learn and remember information and being able to communicate your knowledge verbally or down on paper. Academic performance was hypothesized to be determined by a number of factors, such as student ability, motivation, the quality of secondary education obtained as well as the gender of the student [Cheesman, et al. (2006)].

Motivation is a theoretical construct used to explain initiation, direction, intensity, persistence, and quality of behavior especially goal directed behavior [Maehr and Meyer (1997)]. According to Maehr (1984) motivation is one of the most important components of learning in any educational environment. It is considered to be one of the best formative factors of students' success. There are many theoretical models proposed about motivational beliefs.

Yukselturk and Bulut (2007) proposed a model that involves three basic components: value component that involves task value and goal orientation, expectancy component that involves self efficacy and control beliefs; and the effective factors of test anxiety in which he explore the motivational effect.Based on Yukselturk and Bulut's (2007) theoretical framework, motivational beliefs focused on in this study are intrinsic goal orientation, extrinsic goal orientation, task value, and control of learning beliefs, self-efficacy for learning and performance and test anxiety.

Pintrich and De Groot (1990) explain three factors that characterize student motivation: an expectancy component (self-efficacy), a value component (intrinsic value), and an affective component (test anxiety). The expectancy component involves students' beliefs about their ability to perform a task. The value component involves students' goals and beliefs about the importance and interest of the task. The affective component involves students' emotional responses to the task.

Self-efficacy refers to the abilities of the students for success in a given task [Bandura, (1997)]. It is a student's "I can" or "I cannot" belief. Many studies show that self-efficacy may be a good predictor of performance [Marie, (2006); Larkin, et al. (1987)]. Self-efficacy was positively related to student cognitive engagement and performance [Pintrich and De Groot (1990)].

The value component of student motivation involves students' goals for the task and their beliefs about the importance and interest of the task [Pintrich and De Groot (1990)]. Although this component has been conceptualized in a number of ways (e.g., learning vs. performance goals, intrinsic vs. extrinsic orientation, task value, and intrinsic interest), this motivational component mostly focus on the reasons why students involve in an academic task [Pintrich and De Groot (1990)]. In this paper, the value component is studying in terms of students' intrinsic task value.

Test anxiety is a worry, or a cognitive component which refers to students' negative thoughts that disturb performance, and emotional components which refer to affective and psychological arousal aspects of anxiety. In the study of Pintrich and De Groot (1990) higher levels of test anxiety are related to lower levels of performance on exams.

In academic context self-regulated learning is an important aspect of learning and achievement. Self-regulated learning is a fairly new construct in study on student performance and attainment in classroom settings. [Corno and Mandinach (1983); Corno and Rohrkemper (1985)].In previous studies, there is a variety of definitions of self-regulation. Self-regulation refers to feelings 'thoughts, and actions that are planned and adapted to the achievement of personal goals' [Zimmerman, (2000)]. Self-regulated learning refers to the processes by which individual learners set goals for their learning and then try to monitor, regulate and control their cognition, motivation and behavior [Pintrich, (2000)].

The relationship between self-regulated learning and academic performance has been empirically established involving high school students in traditional settings [Zimmerman and Martinez-Pons, (1986)]. All theories of self-regulated learning treat motivational processes used by students interdependent to their learning processes. Self-regulated learning has been related to high academic performance [Zimmerman and Martinez-Pons, (1992)]. In this paper, we take the self-regulated learning components as predictors of students' academic performance in the classroom. These are cognitive strategy use and self regulation.

Cognitive strategy is a term used in cognitive psychology "to describe the way individuals think, perceive and remember information, or their preferred approach to using such information to solve problems". The use of cognitive strategies can increase the efficiency with which the learner approaches a learning task. These academic tasks can consist of, but are not limited to, remembering and applying information from course content, constructing sentences and paragraphs, editing written work, paraphrasing, and classifying information to be learned.

Self-regulation refers "ones' ability to prioritize which information is to be learned and how it will be learned". Zimmerman and Martinez-Pons, (1988) concluded that students who have a tendency to regulate their learning are usually more successful than those who do not.

In the process of self-regulated learning, learners need to set their learning goals, make their learning plans, choose their learning strategies, monitor their learning processes, evaluate their learning outcomes and suppress interference. Lei, et al. (2002) believe that high achievers' self-regulation ability is higher than that of low achievers. The importance of self-regulated learning lies in providing students with successful experience in order to enhance their intrinsic motivation and promote their self-regulation ability [Zeidner, et al. (2000)].

Students' motivational beliefs and self-regulated learning are directly connected to their academic performance [Pintrich and de Groot, (1990); Zimmerman and Martinez-Pons, (1990)]. So in this study, we examine the relationship between motivational and self-regulated learning components and academic performance of the university students. The research objectives are:

1. To confirm the factors of motivational beliefs and self-regulated learning that influence on classroom academic performance.
2. To find out the relationship between the three motivational components and self-regulated learning components.
3. To find out the relationship between three motivational components and academic performance of students.
4. To explore that which factor is playing a significant role for clusters of below average and above average performers.

2. LITERATURE REVIEW

Ames, et al. (1988) conducted a research to measure achievement goals in the classroom through students' learning strategies (salience of mastery and performance goals) and motivational process. The results shows that students who perceived an emphasize on mastery goals using more effective strategies preferred challenging tasks, had a more positive attitude toward the class and had a stronger believe that success follows from ones effort where as the students who perceived performance goals as salient tended to focus on their ability, evaluating their negatively and attributing failure to lack of ability. The strength and pattern of the findings reveal that the classroom goal orientation may help the maintenance of adaptive motivation patterns when mastery goals are salient and are adopted by students. Pintrich and Groot (1990) conduct a correlational study to check the relationship between motivational believe, self-regulated learning and classroom academic performance at University of Michigan. A sample of 173 seventh graders select from eight science and seven English classes. The results show that self efficacy and intrinsic value, both were positively related to the cognitive engagement and performance. By using regression analysis, results show that self-efficacy, self-regulation and test anxiety is the best predictor of academic performance. Intrinsic value was strongly related to self-regulation and cognitive strategy use, regardless of prior achievement level but did not have a direct influence on performance. Zimmerman, et al. (1992) conduct a study to examine the role of students' self-efficacy beliefs and academic goals in self-motivated academic success by using path analysis procedures. Student's beliefs in their self efficacy for self-regulated learning affected their perceived self efficacy for academic achievement, which in turn influenced the academic goals they set for themselves and their final academic achievement. The results show that students' self efficacy and academic goals were correlated to their self motivated academic success. Schultz (1993) conducted a correlational study to examine the relationship between socioeconomic advantage, achievement motivation and academic performance in an urban elementary school sample size of 130 minorities (African-American –Hispanic). A self report measure of students self efficacy, intrinsic value and self regulatory leaning styles were used to determine the level of achievement motivation (High or Low). Multivariate revealed that socioeconomic advantage and achievement motivation are significant mediators of academic performance in minority children independent of intellectual ability.

Mousoulide and Philippou (2005) conduct a study to examine the relationship between motivational beliefs, self regulation strategies use and mathematics achievement in Cypriot pre-service teachers. A sample of 194 pre-service teachers by using some modification in MSLQ and a mathematics achievement test to developed a model representing connection and casual relations between cognitive and affecting factors. The findings reveal that the theoretical model fits very well, means that self-efficacy playing a significant role for mathematics achievement and self regulation strategies use playing a negative role for achievement.

3. SURVEY METHODOLOGY

Our study population is consisted of the students of the students of BS and MSc programe of Social and basic sciences of university of Gujrat at Hafiz Hayat campus. The total population of the university is 6749. We take those students who were registered in summer session 2010 and the total population of summer session is 1990.

A sample of almost 200 students was selected from the population of 6749 students of University of Gujrat. We calculated sample size as

$$n = \frac{N}{1 + Ne^2} \qquad \textbf{(Yamane 1967)}$$

For the selection of an appropriate sample, cluster sampling is used. We take six blocks of university as Clusters and then from these blocks we select three blocks at random which are Academic block, Science block and Engineering block, and we collect our data from those students who were sitting out of the classes. . We use Motivated Strategies for Learning Questionnaire developed by Pintrich and his colleagues at University of Michigan [Pintrich, et al. (1991)] after some modification. Confirmatory factor analysis, Correlation Analysis and Two-Step Cluster Analysis techniques are used for data analysis purpose.

4. RESULTS AND DISCUSSION

In this Chapter, we interpreted all the results according to the different multivariate data analysis techniques to meet our objectives. First of all, we use reliability analysis for the reliability of the scale use confirmatory factor analysis to confirm the different factors; results are shown in Appendix-B. Second, we use correlation analysis to find out the relationship between the three components of motivation and components of self-regulated learning, also find the relationship between three motivational components and academic performance of students'. Third, we use two step cluster analysis to explore that which factor is playing a significant role for cluster of below and above average performers.

4.1 Reliability Analysis:

Reliable data provides the reliable and significant results. The value of Cronbach's alpha is reported in the Reliability Statistics table. For the reliability of data coefficient of reliability statistics should be greater than 0.70.

Table 1 shows the reliability statistics, the value of Cronbach's alpha based on standardized item is 0.957 which indicates that the data is highly reliable for our study and the results drawn from this data would be reliable.

Table 1: Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.956	.957	55

4.2 Confirmatory Factor Analysis:

We use confirmatory factor analysis to confirm the motivational believe and self regulated learning factors. Table 2 shows the parameter estimates of self efficacy, which indicates that all the variables are playing a significant role for the factor of self-efficacy. Because all the p-values are less than 0.05, so reject the null hypothesis that all variables are not confirm for that factor and conclude that all variables are confirmed for the factor self-efficacy.

Table 2: Model Estimates (Motivational Believe: Self-Efficacy)

	Parameter Estimate	Standard Error	T Statistic	Probability Level
(Self Efficacy)-1->[A_1]	1.266	0.121	10.457	0.000
(Self Efficacy)-2->[A_2]	1.190	0.100	11.903	0.000
(Self efficacy)-3->[A_3]	1.213	0.094	12.958	0.000
(Self Efficacy)-4->[A_4]	1.135	0.103	11.025	0.000
(Self efficacy)-5->[A_5]	1.004	0.097	10.309	0.000
(Self Efficacy)-6->[A_6]	1.291	0.102	12.600	0.000
(Self Efficacy)-7->[A_7]	1.135	0.091	12.464	0.000
(Self Efficacy)-8->[A_8]	1.025	0.098	10.472	0.000
(Self Efficacy)-9->[A_9]	1.093	0.087	12.605	0.000

Table 3 shows the parameter estimates of intrinsic value, which indicates that all the variables are playing a significant role for the factor of intrinsic value. Because all the p-values are less than 0.05, so reject the null hypothesis that all variables are not confirm for that factor and conclude that all variables are confirmed for factor of intrinsic value.

Table 3: Model Estimates (Motivational Believe: Intrinsic Value)

	Parameter Estimate	Standard Error	T Statistic	Probability Level
(Intrinsic Value)-1->[B_10]	1.348	0.113	11.917	0.000
(Intrinsic Value)-2->[B_11]	1.369	0.100	13.624	0.000
(Intrinsic Value)-3->[B_12]	1.251	0.094	13.374	0.000
(Intrinsic Value)-4->[B_13]	1.030	0.090	11.458	0.000
(Intrinsic Value)-5->[B_14]	0.884	0.103	8.605	0.000
(Intrinsic Value)-6->[B_15]	1.176	0.100	11.744	0.000
(Intrinsic Value)-7->[B_16]	1.249	0.103	12.162	0.000
(Intrinsic Value)-8->[B_17]	1.006	0.097	10.417	0.000
(Intrinsic Value)-9->[B_18]	1.202	0.101	11.874	0.000

Table 4 shows the parameter estimates of test anxiety, which indicates that all the variables are playing a significant role for the factor of test anxiety. Because all the p-values are less than 0.05, so reject the null hypothesis that all variables are not confirm for that factor and conclude that all variables are confirmed for the factor of test anxiety.

Table 4: Model Estimates (Motivational Believe: Test Anxiety)

	Parameter Estimate	Standard Error	T Statistic	Probability Level
(Test Anxiety)-1->[C_19]	1.364	0.125	10.871	0.000
(Test Anxiety)-2->[C_20]	1.666	0.112	14.903	0.000
(Test Anxiety)-3->[C_21]	1.283	0.119	10.788	0.000
(Test Anxiety)-4->[C_22]	1.000	0.122	8.198	0.000

Table 5 shows the parameter estimates of cognitive strategy use, which indicates that all the variables are playing a significant role for the factor of cognitive strategy use. Because all the p-values are less than 0.05, so reject the null hypothesis that all variables are not confirm for that factor and conclude that all variables are confirmed for the factor of cognitive strategy use.

Table 5: Model Estimates (Self-Regulated Learning: Cognitive Strategy Use)

	Parameter Estimate	Standard Error	T Statistic	Probability Level
(Cognitive Strategy Use)-1->[D_23]	1.463	0.113	12.933	0.000
(Cognitive Strategy Use)-2->[D_24]	1.282	0.100	12.760	0.000
(Cognitive Strategy Use)-3->[D_25]	0.756	0.125	6.041	0.000
(Cognitive Strategy Use)-4->[D_26]	1.172	0.098	11.944	0.000
(Cognitive Strategy Use)-5->[D_27]	1.007	0.113	8.909	0.000
(Cognitive Strategy Use)-6->[D_28]	1.212	0.101	12.018	0.000
(Cognitive Strategy Use)-7->[D_29]	1.087	0.110	9.849	0.000
(Cognitive Strategy Use)-8->[D_30]	1.428	0.113	12.601	0.000
(Cognitive Strategy Use)-9->[D_31]	1.273	0.093	13.660	0.000
(Cognitive Strategy Use)-10->[D_32]	1.051	0.105	10.030	0.000
(Cognitive Strategy Use)-11->[D_33]	1.091	0.100	10.953	0.000
(Cognitive Strategy Use)-12->[D_34]	1.000	0.089	11.188	0.000
(Cognitive Strategy Use)-13->[D_35]	0.985	0.103	9.530	0.000

Table 6 shows the parameter estimates of self-regulation, which indicates that all the variables are playing a significant role for the factor of self regulation. Because all the p-values are less than 0.05, so reject the null hypothesis that all variables are not confirm for that factor and conclude that all variables are confirmed for factor of self-regulation.

Table 6: Model Estimates (Self-Regulated Learning: Self Regulation)

	Parameter Estimate	Standard Error	T Statistic	Probability Level
(Self Regulation)-1->[E_36]	0.998	0.129	7.735	0.000
(Self Regulation)-2->[E_37]	1.042	0.124	8.384	0.000
(Self Regulation)-3->[E_38]	1.007	0.109	9.204	0.000
(Self Regulation)-4->[E_39]	1.131	0.107	10.565	0.000
(Self Regulation)-5->[E_40]	0.805	0.108	7.427	0.000
(Self Regulation)-6->[E_41]	0.927	0.133	6.985	0.000
(Self Regulation)-7->[E_42]	0.760	0.105	7.224	0.000
(Self Regulation)-8->[E_43]	0.890	0.118	7.547	0.000
(Self Regulation)-9->[E_44]	0.766	0.121	6.331	0.000

Table 7 shows the parameter estimates of academic performance, which indicates that all the variables are playing a significant role for the factor of academic performance. Because all the p-values are less than 0.05, so reject the null hypothesis that all variables are not confirm for that factor and conclude that all variables are confirmed for factor of academic performance.

Table 7: Model Estimates (Academic Performance)

	Parameter Estimate	Standard Error	T Statistic	Probability Level
(Academic performance)-1->[F_45]	1.373	0.111	12.416	0.000
(Academic performance)-2->[F_46]	1.421	0.089	16.025	0.000
(Academic performance)-3->[F_47]	1.327	0.086	15.498	0.000
(Academic performance)-4->[F_48]	1.385	0.093	14.910	0.000
(Academic performance)-5->[F_49]	1.220	0.086	14.129	0.000

Table 8 shows the measures of goodness of fit criteria's of confirmatory factor analysis for all the factors. P-values for all the factors are significant. The recommended criteria for χ^2/df is less than or equal to three, we see that only one factor barely meet the criteria that is self efficacy and remaining factors are not meet the recommended criteria. The recommended criteria for goodness of fit indices (GFI) is greater than or equal to 0.09, we see that self efficacy, test anxiety and academic performance are meet the recommended criteria so we say that these factors are important for the analysis and the remaining factors are acceptable. The recommended criterion for AGFI is also greater than or equal to 0.09, we see that all the factors are not greater than or equal to 0.09 but are acceptable. Recommended range for root mean square error approximation (RMSEA) is less than or equal to 0.03, we see that only one factor is meet the criteria that is test anxiety and the remaining factors are not meet the recommended criterion.

Table 8: Measures of Goodness of Fit Criteria's of Confirmatory Factor Analysis

Factors	Recommended Criteria's						
	Chi-Square	df	p-value	χ^2/df	GFI	AGFI	RMSEA
				$\leq$	$\geq$.	$\geq$.	$\leq$.
Self Efficacy	90.834	27	0.000	3.36	0.901	0.836	0.09
Intrinsic Value	135.42	27	0.000	5.01	0.856	0.760	0.129
Test Anxiety	14.230	2	0.001	7.115	0.969	0.846	0.0859
Cognitive Strategy Use	237.46	65	0.000	3.65	0.839	0.775	0.103
Self Regulation	126.83	27	0.000	4.69	0.871	0.784	0.117
Academic performance	31.249	5	0.000	6.24	0.937	0.810	0.11

4.3 Correlation Analysis:

Correlation Analysis is a statistical procedure by which we can determine the degree of association or relationship between two or more variables. We use correlation analysis to find out the relationship between the three motivation components and self-regulated learning components.

Table 9 shows a zero-order correlation between the three motivational and self-regulated learning components. Table 9 shows that self-efficacy (r =.689, p-value=.000) and intrinsic value (r =.823, p-value =.000) are positively and significantly correlated with cognitive strategy use. Test anxiety (r =.336, p-value =.000) is also correlated with cognitive strategy use. Table 9 also shows that self-efficacy (r =.553, p-value=.000) and intrinsic value (r = .677 p-value =.000) are positively and significantly correlated with self-regulation. Test anxiety is also correlated with self-regulation.

Table 10 shows the Pearson correlation between three motivational components and academic performance of students. Self-efficacy (r =.616, p-value=.000) is significant and positively correlated with students' academic performance. Intrinsic value (r = .652, p-value=.000) is also significant and positively correlated with students' performance. Test anxiety (r = .139, p-value=.05) is less correlated with students' performance.

Table 9: Zero-Order Correlation Matrix

Variables/ Factors		Self Efficacy	Intrinsic Value	Test Anxiety	Cognitive Strategy Use	Self Regulation
Self Efficacy	Pearson Correlation	1	.668**	.191**	.689**	.553**
	Sig. (2-tailed)		.000	.007	.000	.000
Intrinsic Value	Pearson Correlation	.668**	1	.300**	.823**	.677**
	Sig. (2-tailed)	.000		.000	.000	.000
Test Anxiety	Pearson Correlation	.191**	.300**	1	.336**	.452**
	Sig. (2-tailed)	.007	.000		.000	.000
Cognitive Strategy Use	Pearson Correlation	.689**	.823**	.336**	1	.740**
	Sig. (2-tailed)	.000	.000	.000		.000
Self Regulation	Pearson Correlation	.553**	.677**	.452**	.740**	1
	Sig. (2-tailed)	.000	.000	.000	.000	

**. Correlation is significant at the 0.01 level (2-tailed)

Table 10: Pearson Correlation

Variables/ Factors		Academic Performance
Self Efficacy	Pearson Correlation	.616
	Significance (2-tailed)	.000
Intrinsic Value	Pearson Correlation	.652
	Significance (2-tailed)	.000
Test Anxiety	Pearson Correlation	.139
	Significance (2-tailed)	.050

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

4.4 Two-Step Cluster Analysis:

We use two-step cluster analysis *to explore that which factor is playing a significant role for cluster of below and above average performers.* Auto-clustering table summarizes the process by which the number of clusters is chosen.

In Table 11 the clustering criterion (in this case the BIC) is computed for each potential number of clusters. Smaller values of the BIC indicate better models and in this situation the "best" cluster solution has the smallest BIC. In table F-1, the smallest value of BIC is 590.971 on Cluster 4. However, there are clustering problems in which the BIC will continue to decrease as the number of clusters increases, but the improvement in the cluster solution, as measured by the BIC Change, is not worth the increased complexity of the cluster model, as measured by the number of clusters.

In Table 11, the value of BIC Change is -18.272. This value is on cluster 4 which shows that we can select 4 clusters. In such situations, the changes in BIC and changes in the distance measure are evaluated to determine the "best" cluster solution. A good solution will have a reasonably large Ratio of BIC Changes 1.000 and a large Ratio of Distance Measures 2.351. The value of large Ratio of BIC Changes lie in the cluster 2 and the large Ratio of Distance Measures lies in the cluster 4, so it is clear that we select four clusters for our data.

Table 11: Auto-Clustering

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change[a]	Ratio of BIC Changes[b]	Ratio of Distance Measures[c]
1	878.959			
2	681.726	-197.233	1.000	1.954
3	609.243	-72.482	.367	1.708
4	590.971	-18.272	.093	2.351
5	616.693	25.722	-.130	1.415
6	651.969	35.276	-.179	1.036
7	688.038	36.068	-.183	1.447
8	730.967	42.929	-.218	1.022
9	774.233	43.267	-.219	1.136
10	819.300	45.067	-.228	1.145
11	866.038	46.738	-.237	1.069
12	913.516	47.478	-.241	1.153
13	962.426	48.909	-.248	1.302
14	1013.506	51.080	-.259	1.178
15	1065.677	52.171	-.265	1.011

a. The changes are from the previous number of clusters in the table.
b. The ratios of changes are relative to the change for the two cluster solution.
c. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

The cluster distribution Table 12 shows the frequency of each cluster. There are 200 total cases; all the cases were assigned to cluster. Of the total 200 cases, 64 were assigned to the first cluster, 58 to the second, 58 to the third and 20 to the fourth.

Table 12: Cluster Distribution

Cluster	N	% of Combined	% of Total
1	64	32.0%	32.0%
2	58	29.0%	29.0%
3	58	29.0%	29.0%
4	20	10.0%	10.0%
Combined	200	100.0%	100.0%
Total	200		100.0%

In Table 13, centroids show that the clusters are well separated by the continuous variables. In cluster-1 we see that the mean of all variables are high as compares to others clusters, so we can say that high academic performers are in first cluster, above average performers are in second cluster and average performers are lies in third cluster. In the fourth clusters the mean of all the independent variables is less than the average value, so we can say that low average performers are lies in the fourth cluster.

Table 13: Centroids

Variables		Cluster				
		1	2	3	4	Combined
Self efficacy	Mean	50.5000	48.6207	36.7931	29.5500	43.8850
	Std. Deviation	6.30696	6.69597	8.11055	13.5704	10.85268
Intrinsic Value	Mean	53.7812	51.8448	39.1724	28.8500	46.4900
	Std. Deviation	5.75276	5.20392	7.68260	12.0056	11.01073
Test Anxiety	Mean	21.1250	10.8103	15.9138	11.2500	15.6350
	Std. Deviation	3.80684	3.62959	4.32571	4.96170	5.87404
Cognitive Strategy Use	Mean	76.4219	73.4483	55.0345	38.8500	65.6000
	Std. Deviation	7.25675	6.44320	10.0768	14.6657	15.43539
Self Regulation	Mean	50.7344	42.1034	36.7414	28.7000	41.9700
	Std. Deviation	4.88354	5.91664	6.05982	8.89766	9.27520

The cluster frequency Table 14 by academic performance further clarifies the properties of clusters. Cluster 1, 2 and 3 shows that those types of students whose academic performance is above average, with frequencies of 64, 58 and 58 respectively. Whereas cluster 4 shows those types of student whose academic performance is below average with frequency of 20.

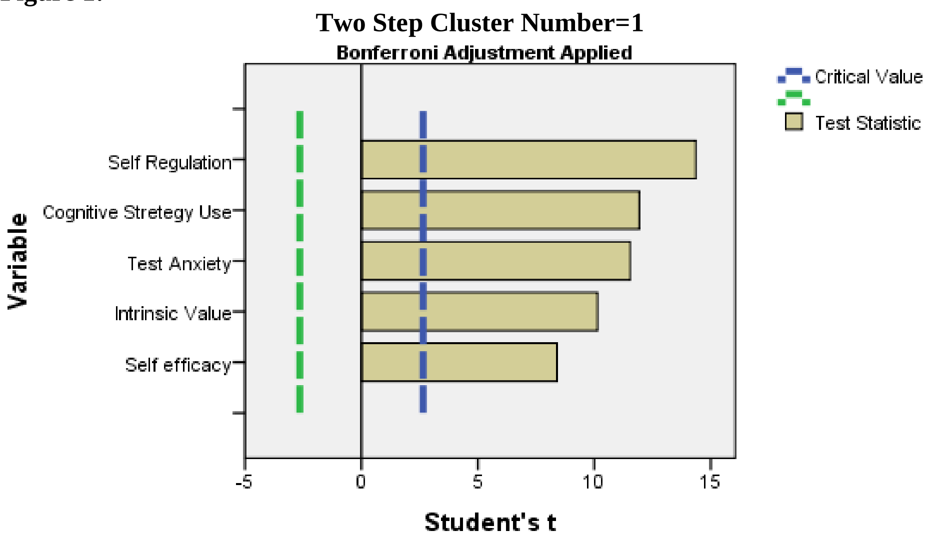
Table 14: Cluster Frequency

		Below Average		Above Average	
		Frequency	Percent	Frequency	Percent
Cluster	1	0	.0%	64	35.6%
	2	0	.0%	58	32.2%
	3	0	.0%	58	32.2%
	4	20	100.0%	0	.0%
	Combined	20	100.0%	180	100.0%

The "by variable" importance charts are produced with a separate chart for each cluster. The variables are lined up on the Y axis, in descending order of importance. The dashed vertical lines mark the critical values for determining the significance of each variable. For a variable to be considered significant, its t statistic must exceed the dashed line in either a positive or negative direction. A negative t statistic indicates that the variable generally takes smaller than average values within this cluster, while a positive t statistic indicates the variable takes larger than average values. If the importance measures for all of the variables exceed the critical value in the chart, we can conclude that all of the continuous variables contribute to the formation of the cluster.

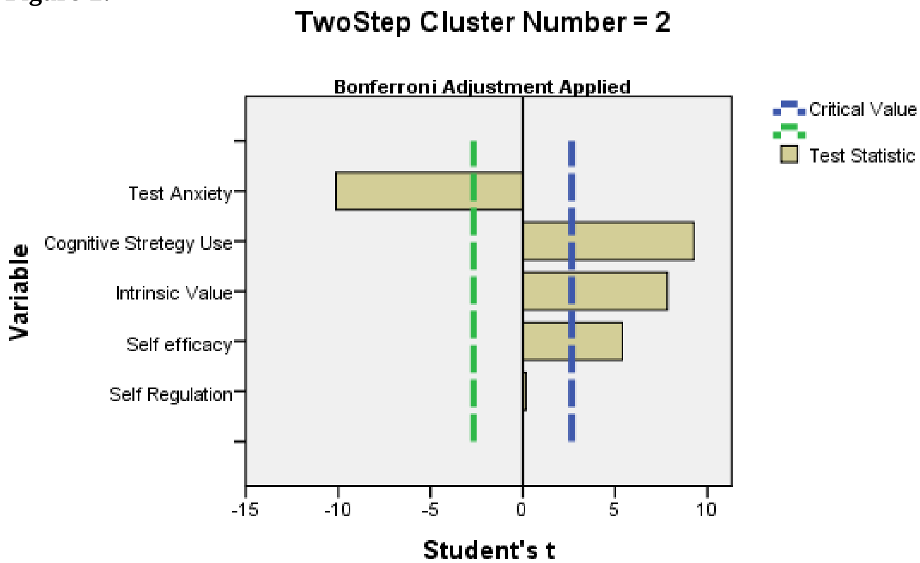
In figure-1, the chart for Cluster 1 shows that all the factors (self regulation, cognitive strategy, test anxiety, intrinsic value and self efficacy) take larger than average values. The highest bar in cluster 1 is Self regulation that supported in the construction of cluster 1, which means that this is the most significant factor of classroom academic performance.

Figure 1:



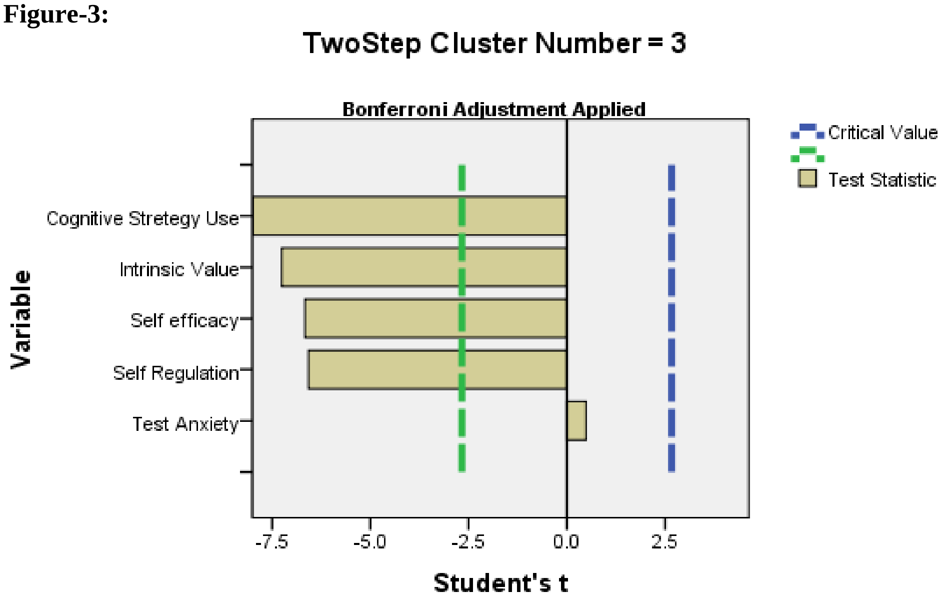
In figure-2, the chart for Cluster 2 shows that the factor of self regulation is not important to the formation of this cluster. Thus for cluster 2 the factors (cognitive strategy, intrinsic value and self efficacy) take larger than average values while the factor test anxiety takes smaller than the average values.

Figure-2:



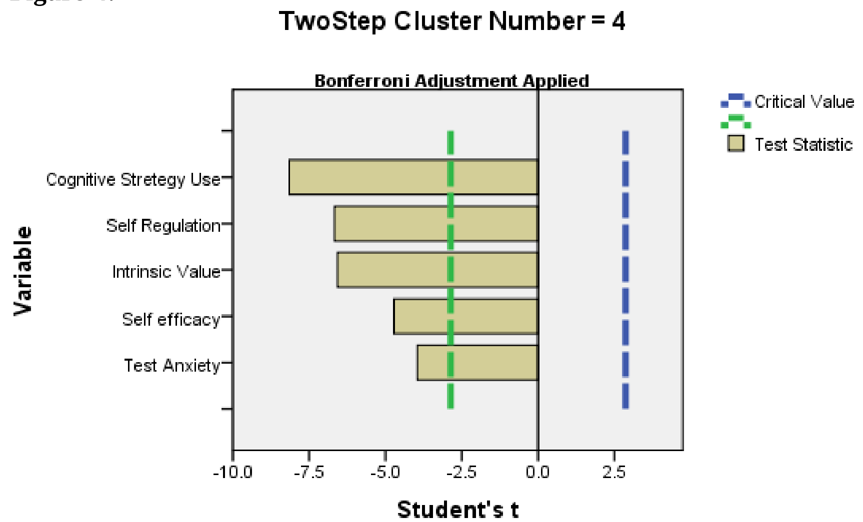
In figure-3, the chart for Cluster 3 shows that the factor of test anxiety is not important to the formation of this cluster, and all of the other variables take smaller than average values.

Figure-3:



In figure-4, the chart for Cluster 4 shows that all of the factors take smaller than average values and all the factors are important to the formation of this cluster.

Figure-4:



4.5 Conclusion:

It is generally accepted that, students who are able to successfully regulate their effort initiate learning task, set goals, decide an appropriate strategies to achieve their goals, then monitor and evaluate their progress will likely to do better who are not [Pintrich, (2003)].However, attaining knowledge of cognitive and metacognitive self-regulating learning is not sufficient to enhance student learning and academic performance, students must also be motivated to use their metacognitive strategies to build upon their understanding of instructional material [Pintrich, (2000)].

In an attainment context, self-efficacy involves students' confidence in their cognitive skills to learn and perform the academic course work (Pintrich, 1999). Although Pintrich (1999) declared that self-efficacy was strongly related to academic performance including examinations. Pintrich, et al. (1993) suggested that intrinsic task value is correlated to performance but those relations are not as strong as those for self- efficacy. On the basis of given results, our findings reveals that motivational and self-regulated learning factors are positively and significantly correlated with each other. Self efficacy and intrinsic value are positively and significantly related with academic performance.

As we know about the factor *which is playing a significant role for clusters of below and above average performers. On the basis of the self efficacy, cognitive strategies and self regulation we can say that these factors playing an important role for the achievement of high academic performance in the classroom context and test anxiety plays a significant role for the low performance in the classroom context.* Also our findings suggested that self-efficacy; intrinsic value, cognitive strategy use and self regulation are most significant factors are associated with academic performance of University of Gujrat Students in the classroom context.

REFERENCES

1. Ames, C. and Archer, J. (1988). Achievement goals in the classroom: Student learning strategies and motivation processes. *Journal of Educational Psychology*, 80, 260-267.
2. Bandura, A. (1977). Self-efficacy toward unifying theory behaviour change. *Psychological Review*, 84, 191-215.
3. Cheesman, J., Simpson, N. and Wint, A.G. (2006). *Determinants of Student Performance at University: Reflections from the Caribbean*. Unpublished Manuscript.
4. Corno, L. and Mandinach, E. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist*, 18, 88-100.
5. Corno, L. and Rohrkemper, M. (1985). The intrinsic motivation to learn in classrooms. In C. Ames and R. Ames (Eds.), *Research on motivation: Vol. 2. The classroom milieu* (pp. 53-90). New York: Academic Press.
6. Kuncel, N.R., Hezlett, S A. and Ones, D.S. (2004). Academic Performance, Career Potential, Creativity, and Job Performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86(1), 148-161.
7. Lei, L., Wang, L. and Tanjia C (2002). Comparative study of self-regulated learning between high achievers and low achievers. *Psychological Development and Education*, 2, 6-11.
8. Lent, R.W., Brown, S.D. and Larkin, K.C. (1987). Comparison of three theoretically derived variables in Predicting career and academic behaviour: Self-efficacy, interest congruence, and consequence thinking. *Journal of Counseling Psychology*, 34, 293-298.
9. Maehr, M.L. (1984). Meaning and Motivation: toward a theory of personal investment. In R.E. Ames and C. Ames (Eds.). *Research on Motivation in education* (Vol. 1, 115-144). New York: Academic Press.
10. Maehr, M.L. and H.A. Meyer (1997) Understanding motivation and schooling: where we've been, where we are, and where we need to go, *Educational Psychology Review*, 9, 371-403.
11. Marie, K. (2006). Examining an instructional strategy: relationship between feedback, performance and self-efficacy in a first year mathematics course, Conference Proceedings of AERA, available at www.sfu.ca/rethinkingteaching/publications/ krbavacAERA.pdf. Retrieved on January 03, 07.
12. Mousoulides, N. and Philippou, G, (2005). Students' motivational beliefs, self-regulation strategies and mathematics achievement. *Procedures of the 29th conference of the interaction group for the psychology of mathematics education*, (3), 321-328. Melbourne: PME.
13. Pintrich, P.R. (1999). The Role of Motivation in Promoting and Sustaining Self-Regulated Learning. *International Journal of Educational Research* 31, 459-470.
14. Pintrich, P.R. (2000). The role of motivation in self-regulated learning. In P.R. Pintrich & P. Ruohtie (Eds.), *Cognitive constructs and self-regulated learning.* Saarijarvi: Offset 51-66.
15. Pintrich, P.R. (2003). A motivational science perspective on the role of student motivation in learning and teaching context. *Journal of Educational Psychology*, 95, 667-686.

16. Pintrich, P.R., Marx, R.W. and Boyle, R.A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63, 167-199.
17. Printich, P.R. and de Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33-50.
18. Pintrich, P.R. (2000). The role of motivation in self-regulated learning. In P.R. Pintrich & P. Ruohtie (Eds.), *Cognitive constructs and self-regulated learning.* Saarijarvi: Offset 51-66.
19. Pintrich, R., Smith, D.A.F., Garcia, T. and McKeachie, W.J. (1991). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, 53, 801-813.
20. Pintrich, R., Smith, D.A.F., Garcia, T. and McKeachie, W.J. (1991). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, 53, 801-813.
21. Schultz, G.F. (1993). Socioeconomic advantage and achievement motivation: important; mediators of academic performance in minority children in urban schools. *Urban Review*, 25, 221-232.
22. Yukselturk, E. and Bulut, S. (2007). Predictors for students' success in an on-line course. *Educational Technology & Society*, 10(2), 71-83.
23. Zeidner, M. and Matthews, G. (2005). Evaluation anxiety: Current theory and research. In A. Elliott & C. Dweck (Eds.). *Handbook of competence and motivation* (141-166). New York: Guilford.
24. Zimmerman, B.J. (2002). Becoming a self-regulated learner: An overview. *Theory Into. Practice*, 41(2), 64-70.
25. Zimmerman, B. and Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, 23, 614-628.
26. Zimmerman, B.J. and Martinez-Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of educational psychology*, 80(3), 284-290.
27. Zimmerman, B.J. and Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, 82(1), 51-59.
28. Zimmerman, B., Bandura, A. and Martinez-Pons, M. (1992) Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29, 663-676.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 343-346

ON THE ROLE OF VALIDATED HR ACCOUNTABILITY FOR SUSTAINABLE DEVELOPMENT IN THE DEVELOPING COUNTRIES

Muhammad Tufail Jarral
PEPCO (WAPDA), Lahore, Pakistan
Email: mtjarral@hotmail.com

ABSTRACT

For sustainable development in a country good governance is pivotal. The governance of a country refers to sum total of exploitation of its available natural resources in the country plus external resources effectively through compatible decision making by the managers / state functionaries. At micro functional level, the managers achieve the set target(s) in the organizations through optimal management of such resources which includes identification of potential resources (Physical, HR, and Financial), infrastructure capacity building, strategic formulation and developing short / medium / long term operational mechanism for the same.

Good governance among other differentials like infrastructure, resources and organizational system environment is based on effective HR functional roles specially at the monitoring level. The challenges of corruption and bad governance in the developing countries which have direct negative impact on the sustainable development, can also be addressed through better functional capacity building specially in control management.

This paper seeks to quantify importance of HR capacity building specially the accountability management in the organizations and to suggest mechanism for ensuring validated HR accountability system that will lead to sustainable development and good governance in the developing countries.

KEY WORDS

Decision making, capacity building, data power, human resource, governance differentials, validation, accountability, infocracy.

1. INTRODUCTION

1.1 Overall development in a country depends on effective micro functioning of the organizations. Decision making has far reaching impact on the internal as well as external environment of the organization and resultantly on the overall economic development of a country at each managerial level right from planning to control.

1.2 Decisions making is imperative for promoting a business activity like banking, industry, production, agriculture, population, healthcare etc. and even in the human resource management itself.

1.3 The functional management activity may be operational or monitoring requires apart from other management differentials, judicious decision making.

2. ANALYTICAL DISCUSSION

2.1 Quality Services

2.1.1 Like quality of finished products which is obtained through application of statistical techniques, the quality services in the organizations are associated with the proper capacity building in any managerial activity may be policy making, operational, or administrative.

2.1.2 Capacity building in any managerial activity generally refers to taping of various systems differentials and resources applied in a business scenario for achieving the set objective(s) of that activity with validated deliverables.

2.1.3 Functional capacity building refers to intra capcitation of various departments of the organization and HR role playing with given responsibility. If HR do not perform their functions then there should be some inbuilt mechanism in the system that makes them responsive to their actions i.e. what is referred as accountability. It is the effective role playing of HR that yields quality results. The effective role playing is based on decision management capacity building.

2.1.4 Effective controlling activity in the organization results in quality services which requires decision management capacity building.

2.2 Decision Management Capacity Building

2.2.1 Monitoring or control activity which determines deviations between the actual actions / results and the desired targets for rectifying the same through some corrective / preventive actions involving accountability process of the potential HR which is based on valid decision making.

2.2.2 The quality of data and its validity are special features of decision management [6]. Statistics or info indicators like economic, financial, socio-economic provide rationale for establishing valid decision making capacity building in all business and socio-economic sectors. We see that official statistics and economic indicators in the developing countries are open to certain observations from quality end like customization of data input formats, data reporting systems [1].

3. ACCOUNTABILITY CAPACITY MANAGEMENT

3.1 Accountability refers to the mechanism of making individual role(s) responsive to their decisions / actions. Valid accountability based on compatible decision making yields sustainable development in the organization. Public managers must be concerned with issues of accountability and equity. Performance based organizations operate closer to public sector models. [4].

3.2 The managers involved at monitoring level or accountability process are to make compatible decision making. In developing countries, it is an established fact that monitoring activities are not conducted properly e.g.

- The process itself is not transparent and specified procedures are not sometimes followed in true spirit may be due to lack of knowledge management in the field.
- Decisions are not made timely by the competent authority and they take years and years.
- The decisions of the competent authorities are not compatible and are challengeable in sister / higher legal forums.
- The officers given penalty by the competent authorities are exonerated later on by the appellate forums on account of say procedural flaws, or with no data support.

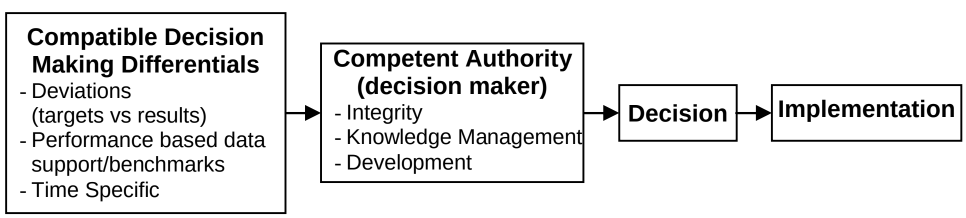
3.3 Beyond the principles of accountability and legitimacy, merit must include problem solving and management [3]. This leads to transparent accountability system.

3.4 Compatible decision making can be made if proper decision management capacity building [6] is developed in the organizations within following correlates leading to proper accountability process:

a. Accountability HR are independent of the other managerial activities.
b. HR involved in monitoring / accountability assignments are men of integrity.
c. Decisions are made within specific time schedule.
d. Performance based targets are taken as benchmarks for monitoring activity i.e. decisions are based on valid data and taken judiciously which will also lead to infocracy culture in the organizations.

Foregoing discussion boils down to the following suggested accountability model:

Accountability Model



The above accountability model will develop merit culture, transcend growth environment in the organization(s) leading to overall development and governance of the country.

4. RECOMMENDATIONS

4.1 Decision management capacity building be developed in the organizations.

4.2 Proposed accountability model be introduced in the organizations.

4.3 Managers responsible for monitoring / accountability system in the organizations be equipped with knowledge of statistics, business / public organizations laws and groomed through tailor made courses / seminars / workshops for better analysis and interpretation of data profiles for making transparent and compatible decisions.

REFERENCES

1. SESRTCIC (2004). WEB Usage in the National Statistical Organizations of the Member Countries of the Organization of the Islamic Conference – Report of Fourth meeting of the Heads of the NSOs of the Member Countries of the OIC (11-13 Feb), Lahore, Pakistan.
2. Ross Bernard H, (1988). Public and Private Sectors: The Underlying Differences, Management Review, 77 (May).
3. Ingraham, Patricia W. (1995). *The Foundation of Merit: Public Service In American Democracy*, Baltimor, John Hopkins University Press.
4. Roberts, Alasdair (1997). Performance Based Organizations: Assessing the Gore Plan. *Public Administration Review*, 57 (Nov. Dec).
5. Jarral, M.T. (2010). On the need of Establishing Data Validation Research Houses in the Developing Countries. Proc. *Two Day International Conference on World Statistics Day "Statistics for Society"* organized by ISOSS (Oct 21-22) at Superior University, Lahore, Pakistan, 31-34.
6. Jarral, M.T. (2010). Data as Resource Differential for Sustainable Development and governance in a Developing Country. *IUB Journal of Social Sciences and Humanities*, 8(1), The University of Bahawalpur, Pakistan.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 347-360

AN INVESTIGATION OF MORTALITY IN ELIZABETHAN ENGLAND USING THE BSD

Sam C. Saunders[1] and **Saleha Naghmi Habibullah**[2]
[1] Prof. Emeritus, Washington State University, Washington, USA.
[2] Department of Statistics, Kinnaird College for Women, Lahore, Pakistan. Email: salehahabibullah@hotmail.com

ABSTRACT

This is an application of the Fatigue-Life, or the Birnbaum-Saunders distribution [hereafter we use BSD] which has been successful in describing the service life of materials determined by the gradual accretion of incremental damage, to depict human mortality during the 16^{th} century in England. It was a time when no mortality data, as such, was collected hence only anecdotal evidence is presently available. However by utilizing the "plateau- effect", one of the characterizations of the BSD, we are able to explain and account for many of the anomalies of that time.

1. INTRODUCTION

A lecture on this subject was presented at the *Shakespeare Authorship Studies Conference* in 2008 held at Concordia University in Portland, Oregon, USA. It was initially an investigation principally to estimate the population size of Stratford-upon-Avon so as to determine the likelihood that a boy raised there could have attended the King's Free Grammar School. But here it is applied to another question.

It has been long known that virtually all populations of units, biological, mechanical or electrical, have a hazard (mortality) rate that is "bath-tub shaped". But only recently has it been discovered that the late-life mortality levels off and shows a plateau effect. It is theoretically known that the hazard function determined from the distribution of the waiting time until the accumulation of random incremental damage exceeds a critical level (the so-called Fatigue life or Birnbaum-Saunders distribution) does exhibit this non-intuitive behavior. Consequently these distributions have here been applied to model the distribution of human mortality in England during the 16^{th} century to see if the anecdotal evidence of remarkable exceptionally long lives can be accounted for during a time when the median life-length was only seven years and the mean life only seventeen. This talk was then modified and presented at the *8^{th} International Conference on Recent Advances in Statistics* held in Lahore, Pakistan in 2011.

It is not always true that large amounts of data are needed to answer every statistical question that arises. Sometimes when the correct distribution is being used even when data is sparse valid answers can be obtained. The most striking example perhaps occurred in the first year of the 19^{th} century when three observations were made in a short period of what was then believed to be a new planet. Subsequently bad weather then obscured observation and a few weeks later when the weather cleared the `planet' could not be located. But using these observations and the distribution now named for him Carl F.

Gauss computed by hand, an enormous task encompassing several months' labor, the estimated orbit predicting planets' location, so that what is now known as Ceres, the largest asteroid, was again located by astronomers in September. A similar accomplishment is what is intended here as regards human mortality in Elizabethan England, an almost legendary time of English history.

2. ESTIMATION OF POPULATION SIZE

We begin with a Theorem in Stochastic Processes which applies to population size:

Theorem 2.1:

The size of any human population, under stationary conditions (that is, maintained only by birth and death without migration, epidemics or wars), can be estimated simply as the product of the average life of its members, in years, times the birth rate of babies per year.

This is a fact that can be used in every town in which there is a record of the number of births each year (say, kept by the church) and the average length of life can be estimated from local tombstones. In order to estimate well the average length of life, one would need only the date of birth and the date of death of each person from a small random sample of people within the population.

Thus, hypothetically, if the records maintained at Trinity Church in Stratford-upon-Avon were faithfully kept and reveal an average of 40 births per year, when the average life-length is about 25 years then the population size of the township would be estimated at about 1,000 persons.

Note that in a stationary process, with the expected population size remaining constant, the death rate per year must equal the birth rate per year on average or the population size would change and the population not be stationary. So we can use what available statistics there are from either source, death records or birth records, to estimate the population size. (The estimate due to the birth rate should match closely on average the estimate using the death rate. If they don't, it means that perhaps in some period many men were killed because of a war or a large segment of the population was wiped out because of some natural disaster or an epidemic of the plague occurred.)

3. MATHEMATICAL PRECIS

The two-parameter Fatigue-life distribution, label it F, with density $f = F'$ and hazard function $H = -\ln[1-F]$, can be represented, for all $t > 0$; $\alpha, \beta > 0$, by

$$F(t) = \Phi[\tfrac{1}{\alpha}(\xi(t/\beta)]$$

where, for $-\infty < x < \infty$ we have utilized notation for the standard Gaussian distribution as

$$\Phi(x) = \int_{-\infty}^{x} \frac{e^{-t^2/2}\,dt}{\sqrt{2\pi}}$$

and we have written $\xi(t) = t^{1/2} - t^{-1/2}$ for $t > 0$. for short.

Let X denote the random time until infant mortality [early failure] due to birth defects or childhood disease [poor quality control or mis-assembly] or accidents independent of age. Let Y be the time until death caused by the gradual deterioration during adult life with a weakening immune system which facilitates disease and infection and ultimate death. [or fatigue, wear or abrasion cause machine failure].

The observed life will be $L = \min\{X,Y\}$ and so the survival distribution of life is then given by

$$\Pr[L > t] = \Pr[X > t, Y > t] = \exp\{-H_1(t) - H_2(t)\} \text{ for } t > 0.$$

For early life mortality, that is in H_1, we take $\alpha \approx 7, \beta \approx 3$ while for adult mortality in H_2 we take $\alpha \approx 0.3, \beta \approx 25$. These four parameters will produce a life-length distribution having a hazard rate which ultimately exhibits not only the traditional bath tub shape but one that also displays the counter-intuitive `plateau-effect' phenomenon of extreme old age. Such a good fit to data representing the whole-life hazard is possible here using only four parameters while competing models require many more, such as one composed from Weibull hazards, requires at least seven.

4. DATA ON LIFE-LENGTH IN 16TH CENTURY ENGLAND

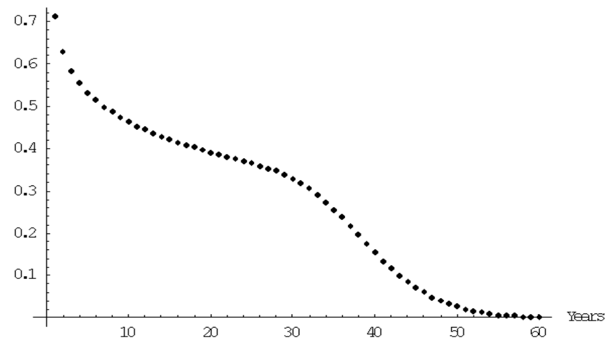
The crucial question is: What anecdotal data do we have on average life length? In order to answer this question, we refer to Hanson (2004), Riley (2001), White (2002), or Isaac Asimov and others regarding some aspects of life in 16th century England. Riley (2001) opines: "There are, in human history, virtually no valid mortality data until the 19th century." On the other hand, (a) Hanson (2004) talks about "Elizabeth's reign and the short life expectancy of her subjects – an average of 25 years for the poor and 35 for the rich", (b) White (2002) states: "The year 1600 was little different from 1450. The average life expectancy was 24 years for a woman and perhaps 27 for a man", and (c) Isaac Asimov & others mention "half the children born in 16th century England die before age 7".

Nevertheless ---- valid data is a requirement for valid conclusions; fuzzy data compels fuzzy answers. The question is: How can one obtain absolutely reliable survival data for a specific population during a particular time of interest? One answer is as follows:

Identify a cohort, a large number, say 10,000 persons, all born at known times within one year in the population, and then keep track of the time until the death of every person, until the last one has expired. This is the procedure that was used in Sweden: about the year 1900, a large group of say 10,000 people, all born within one year, and they kept track of them until they all expired. If this had been done in the 16th century, we then would have a Cohort Survival Graph which for Elizabethan England might look like the following:

Figure 2.1

Generic (Typical) Cohort Survival for Elizabethan England Survival Fraction



But the question then arises that, if there were no data, where did this graph come from? Is it all fictitious. Not quite, it comes from those few benchmarks quoted and the BSD. (We have proceeded with the assumption that the statements in the cited references are, at least approximately, true.)

5. IMPORTANT GUIDELINES

In order to predict a complex effect resulting from the presence of multiple independent causes, we must establish the laws stating the effect each factor would have were it isolated from the other operative factors.

The choice of the correct distribution of life length is very important, and life length is determined by the multiple operative mechanisms which end or shorten life, especially when predictions must extrapolate beyond the applicable data.

6. SOME WELL-KNOWN LAW OF MORTALITY

a. **The Gompertz Law of Mortality**
In 1825 Gompertz introduced his "New Law of Human Mortality" which said that the chance of death during each additional year of life can be expressed by his law which increases with age and does so exponentially i.e. if *h(t)* is Gompertz law [or force] of mortality then

$$h(t) = kt^{\alpha}$$

In other words, at any age $t > 0$, the logarithm of Gompertz force-of-mortality is linear.

b. **The Gompertz-Makeham Law**
Gompertz & Makeham, about 1860, published "On a (Mathematical) Law Expressive of Human Mortality". This was a refinement of the Gompertz Law called the Gompertz-Makeham law. Under it there exists an acceleration in late-life aging. This

law implies that the older a person becomes the faster will increase his/her likelihood of death. In particular, the differences in the force of mortality each year not only increase but the differences between the differences also increase (acceleration).

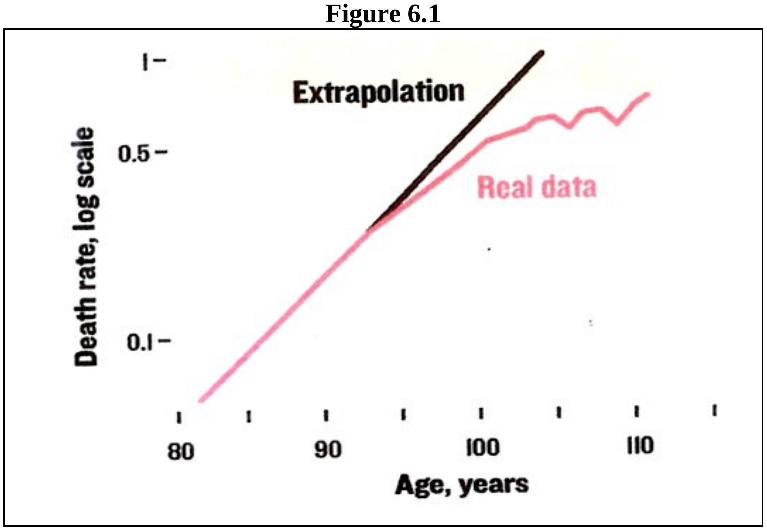
There now exist many other IFR or IHR (Increasing Failure, or Hazard Rate) statistical distributions which can be used to describe human mortality.

7. EVIDENCE FROM SWEDEN

Recent data from Sweden on female mortality for the years 1990-2000, which are given in the Kannisto-Thatcher Database on Old Age Mortality, show that in fact death rates slow at advanced ages. [This now seems contrary to our intuition, having accepted Gompertz –type laws for two two centuries] After age 95 the observed risk of death deviates from the value predicted by mortality laws which specify that the force of mortality accelerates with age. This evidence is depicted in Figure 6.1.The fluctuations in the data toward the RHS of the graph are due to the paucity of data at the advanced ages. (For example; in USA, Only two persons are presently alive at the age of 114.) The important point is: The projected line is way off!

Proper prediction of death rate at advanced ages is important for insurance and health providers for the elderly. And, as such, the choice of the correct distribution at extreme old age is critical.

Figure 6.1



8. COMPARISONS BETWEEN MEN AND MACHINES

The failure rate for components (or materials) which in service perform repeated duty-cycles, is similar to the force of mortality for humans enduring a daily-grind of work and stress during maturity during which their age gradually increases and strength decreases. (Note: The “force of mortality” in biological life corresponds to the hazard, or failure, rate in machines and electronic systems.)

These similarities include:

(1) Infant mortality in biological life is the same as manufacturing error in human products.
(2) Mid-life for humans and mid-service-life failure in machines both often result from accidents which are independent of age.
(3) Late-life failure in both men and machines is most often due to accumulated damage from stress, fatigue or wear with immune system weakening by age.
(4) Correspondence of Extreme-Old-Age-Mortality Behavior in men and machines

In view of the similarities between men and machines as well as man-made artifacts, such as steel, relays, thermal insulation ----- perhaps reliability theory may be able to provide a plausible explanation.

9. THE BIRNBAUM-SAUNDERS DISTRIBUTION (BSD)

The BSD applies when failure is determined by the gradual accumulation of incremental damage during the repetition of a duty-cycle (or diurnal cycle) causing wear, fatigue, abrasion, creep or galling until a critical level is exceeded, which in metals is also often accompanied by a relaxation of the yield stress (i.e. aging). The weakening may be induced by oxidation, UV-exposure, chemical attack of pollutants until eventually failure occurs.

The BSD has been shown to predict well the reliability and service lives of many materials, and machines. Its applicability to human mortality is what is hoped to be demonstrated here.

But --- without hard, and only anecdotal, data?

10. THE HAZARDS DURING FOUR EPICS OF LIFE

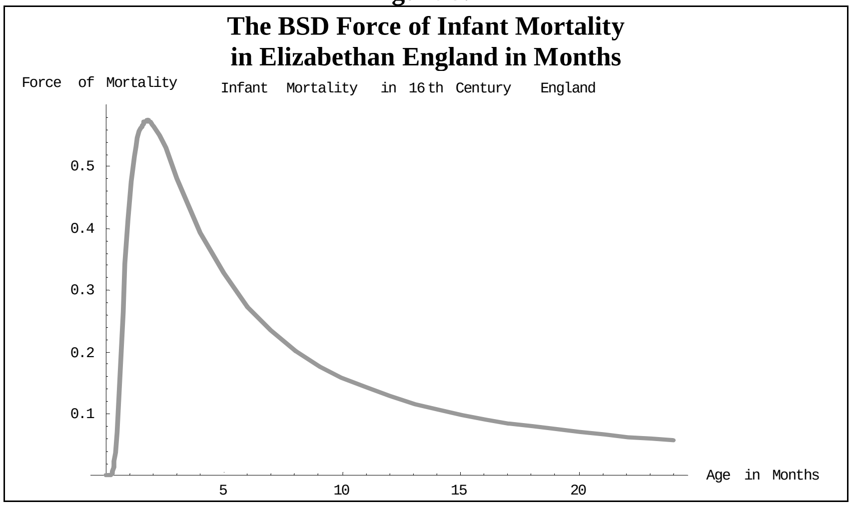
The hazard rate can be separated into the following four categories:

- Infant mortality and childhood disease
- mid-life accidents and Acts of God
- late-life aging and a weakening immune system, and&
- extreme old age

Figure 9.1

The BSD Force of Mortality for persons over age 21 in 16th Century England Force of Mortality

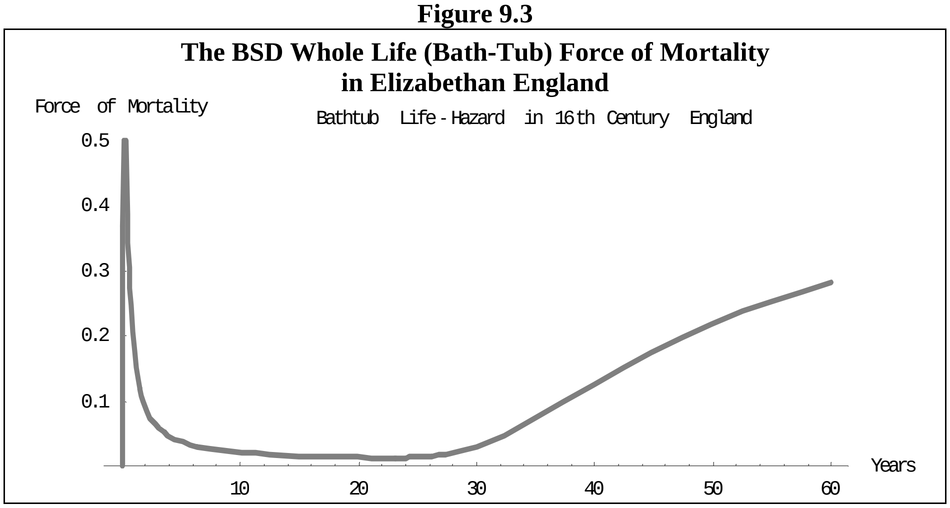
Figure 9.2



where does this last graph come from? It comes from using the BSD and the single presumption that half the children perish before their seventh birthday.

Combining the two graphs by adding the two hazard rates, we obtain the following graph:

Figure 9.3



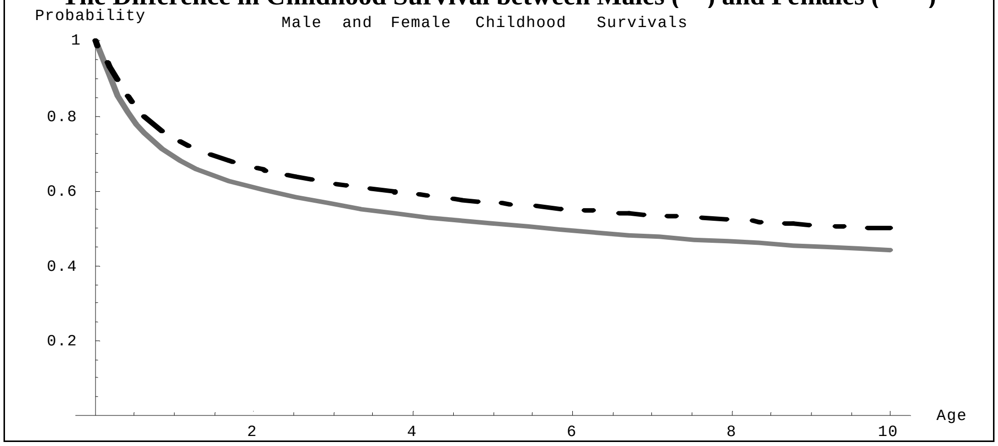
The bath-tub hazard rate for the service life of manufactured components has been observed now for over fifty years. (The term bath-tub originated in the context of machine life.) Here we see it in the context of human life.

(NB: When we contract the axis from months to years, the spike in infant mortality looks like it starts from zero; the sharp peak on the extreme left results from the contraction of the X axis.)

11. COMPARISON BETWEEN MALES AND FEMALES

Figure 10.1

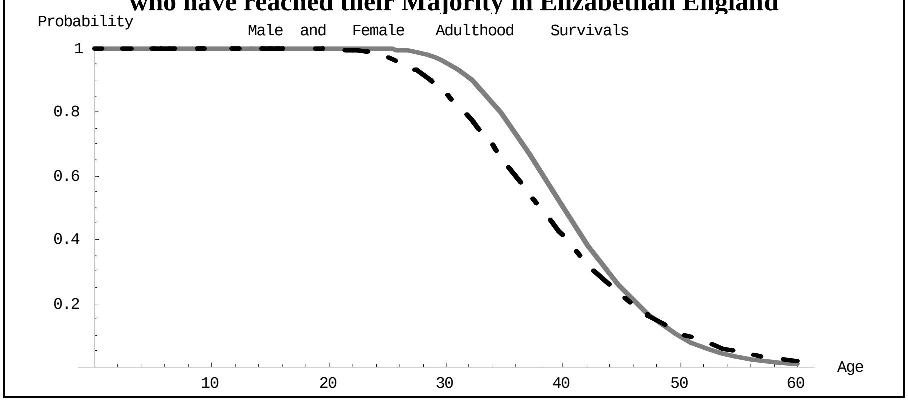
The Difference in Childhood Survival between Males (—) and Females (- - -)



A Possible Explanation: Men have greater variability in every inherited characteristic than do females. This is because men have XY chromosomes whereas women have XX. (Women's second chromosome is duplicated whereas men's is varied. This means there are more men with low resistance and hence more die from disease. (It's evolution's plan since 52% of the births are males in order to equalize during marriageable years those males that were lost in disease.)).

Figure 10.2

Survival Probabilities for Males (—) and Females (- - -)
who have reached their Majority in Elizabethan England



NB: "Majority" => age >21 (as opposed to being Minor).

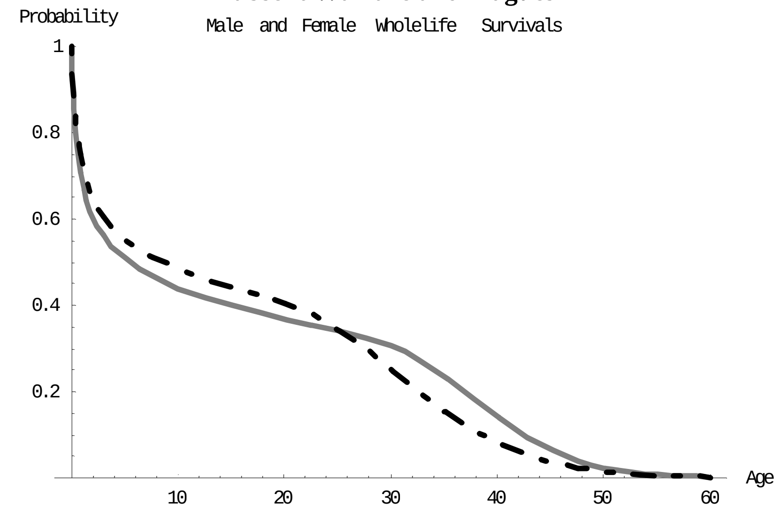
NB: Given one is alive at age 21 means that, for ages 0 to 21, the probability of survival is 1:

We see that the increase in mortality for mature females, due to childbirth, causes a diminution in survival relative to males. For ages beyond 50, we note that the survival probabilities for females are slightly higher than those for the males. Why? Because women are the healthier creature --- greater emotional stability, fewer systemic problems such as heart attacks, etc.

Combining the previous two graphs, we obtain the following graph:

Figure 10.3

Male (—) and Female (- - -) Survival in Elizabethan England, absent Warfare and Plagues

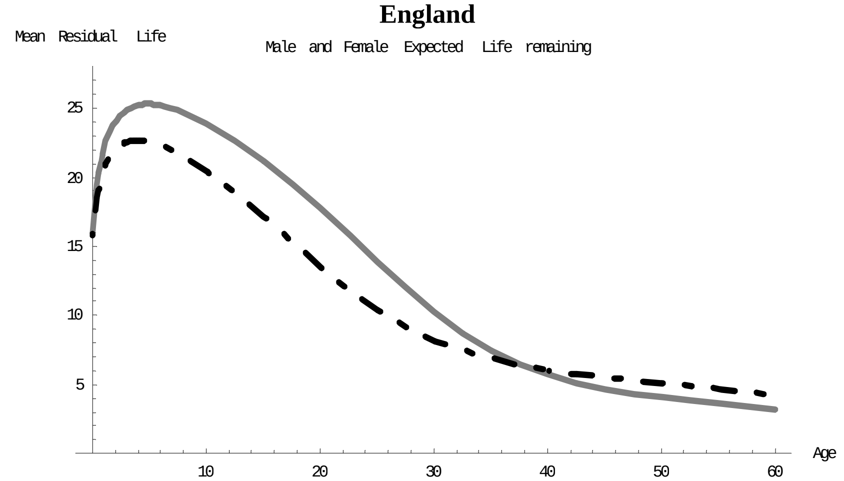


12. MEAN RESIDUAL LIFE

MRL at age x is defined to be the expected life remaining if one lives to age x, We obtain the following graph for the MRL for the two sexes:

Figure 11.1

The Mean Residual Life of Males (—) and Females (- - -) in Elizabethan England



We note that in Elizabethan England after a baby was born, there was a great probability that he/she will die right away. If he/she lives to age 7, the expected residual life rises dramatically.

In addition, we see that:

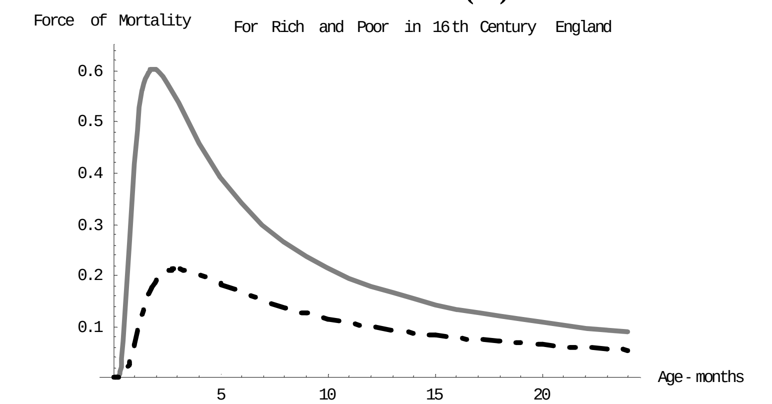
- At age 40, both men and women have a expected residual life of about 7 years.
- Prior to age 40, men have a longer expected residual life.
- Beyond age 40, women have a longer expected residual life.

13. COMPARISON OF INFANT MORTALITY FOR THE NOBILITY AND COMMONERS

The population of England during the mid 16th century was about 3 million people of which maybe only 2% were nobility and hence rich.

Figure 12.1

Comparison of Infant Mortality for the Nobility (- - -) and Commoners (—)



A Possible Explanation:

Every child born to the nobility had a nurse, was better fed, dressed, and housed and not exposed to so many deadly diseases carried by lice or rats as well as other agents of infection than was a child of the commoners. Mortality would equalize when a male commoner was bright enough and lucky enough to finally be educated to read and write.

14. THE FOURTH EPIC OF LIFE

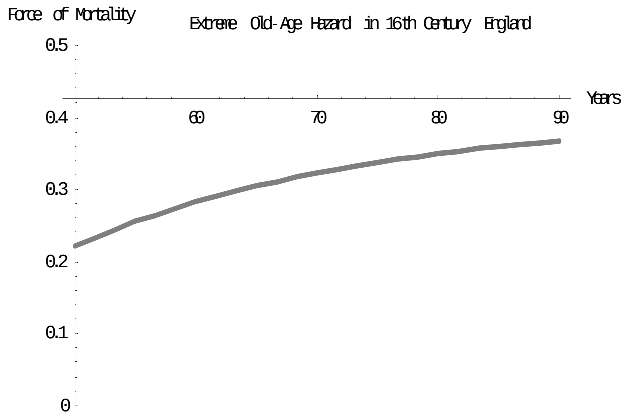
The fourth epic in life is another one that we and our machines share. The late-life deceleration of mortality means that death rate stops increasing and ultimately “plateaus.” According to IEEE Spectrum (Nov. 2004):

- If you live to be a 110, your chances of seeing your next birthday are not very good but paradoxically they are neither better nor worse than they were at 102.
- There have been numerous attempts to explain the biology behind this but since this is found in man-made stuff, such as steel, relays, thermal insulation perhaps reliability theory may provide a better answer.

WE now demonstrate how the BSD depicts the behavior of life in extreme old age:

Figure 13.1

The BSD Deceleration of Mortality in Extreme Old Age in the 16th Century



This is the hazard rate for extreme old age, and, clearly, it has started to flatten out NB that the chosen scale and aspect ratio of the graph make the curvature seem far less than it would be if the scale matched that of the preceding graphs.

15. SOME POINTS OF CONFIRMATION OF THE BSD AND A FEW FACTS OF ELIZABETHAN LIFE:

The following points follow from the BSD Model fitted to the 16th century population in Elizabethan England where there is little data:

- **The equality of sexes in mating potential:** There are 2% more males born than females. 102% of fraction of men aged 20 to 50=0.351; 98% of fraction of women aged 15 to 45=0.354. Hence, in the reproductive [marriageable] ages of people, the numbers of men and women are about equal.[Nature's plan]
- **Statistics for the general population:**
  - The median life (when exactly half have died) is 7 years
  - Life expectancy at birth is only 17 years
  - Modal life (most frequent age at death) is seen to be about two months, 1/6 yr.

As such, we see that the model fits the accepted facts about Elizabethan life.

16. THE MIRACULOUS SURVIVAL OF A FEW?

We find the following remarkable instances of longevity quoted in the literature:

- William Cecil (1520-1598) lived to age 79
- Beth of Hardwick (1521-1607) lived to age 85±1.
- William Parry, executed for treason, claimed his father died in 1566 at age 108

- George Beeston was knighted in 1588 by Lord Admiral Howard (his cousin), at the age of 89, for his service as a Naval Commander against the Spanish Armada
- Eleanor of Acquitaine (1122-1204) , wife to two Kings of France, mother of 11 children including two Kings of England, lived to age 81±1

Under the Gompertz-Makeham law (19th century) which has an accelerating aging, these facts were miraculous!

How well does the BSD model account for these long lives in Elizabethan England? We compute the probability of a person attaining the stated age using the BSD.

William Cecil (1520-1598) lived to age 79:

- Cecil: 8.06×10^{-6} or 8 in a million
- There were 3 million people in Elizabethan England implying that 24 persons could be expected to live to that long (age 79) each year.
- So this seems clearly possible and is not a miracle at all.

Beth of Hardwick (1521-1607) lived to age 85±1:

- Beth of Hardwick: 14.5×10^{-6} or 14 in a million
- Thus about 42 persons could be expected at that age each year in England
- Even more likely than Cecil's long life. Not only possible but to be expected.

Eleanor of Acquitaine (1122-1204), wife to two Kings of France, mother of 11 children including two Kings of England, lived to age 81±1:

- Eleanor of Acquitaine: 36.5×10^{-6} or 36 persons in a million
- This seems possible despite the fact life in the 12th century was even more brutish and shorter than life in the 16th.

George Beeston was knighted in 1588 by Lord Admiral Howard (his cousin), at the age of 89, for his service as a Naval Commander against the Spanish Armada:

- George: 0.23×10^{-6} or 2.3 in 10 million
- There were 3 million people in Elizabethan England implying that about 1 male could be expected to be alive at that age (89) each year.
- That must have been George himself! But it is a rare event

William Parry, executed for treason, claimed his father died in 1566 at age 108:

- Parry's dad: 1.73×10^{-6}
- Extremely improbable
- It is either a miracle or a lie, take your choice!

How well does the BSD model do to account for these long lives in Elizabethan England?
They were rare events and hence notable but they were not miracles.

17. OVERVIEW AND SUMMARY

The perspective presented in this paper has been based on one canard: “half the children born died before the age of seven” and a few facts, e.g., “William Cecil lived to age 79” as well as one assumption: “Life-length in Elizabethan England can be described and analyzed by using the BSD to break up life’s hazard rate into two terms one representing the force-of-mortality for childhood and one representing the force of mortality during maturity. (The probability of life being greater than age t equals the product of two independent probabilities, the probability of not dying due to childhood disease and infant mortality, and the probability of not dying due to cumulative damage as a result of work and stress of life.) before age t.

In conclusion we can say that the BSD seems to be the `Best Scientific Description.’

REFERENCES

1. Asimov, Isaac, (1978). *Asimov’s Guide to Shakespeare,* in Commentary on *Troilus and Cressida*. Avenel Books, NY.
2. Budiansky, S. (2005). *Her Majesty’s Spymaster,* Penguin Group, NY.
3. Gidlow, C. (2007). *Life in a Tudor Palace*, Sutton Publishing Ltd, Gloucestershire.
4. Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality and a new mode of determining the value of life contingencies. *Philos. Trans. Roy. Soc.*
5. Hanson, N. (2004). *The Confident Hope of a Miracle.* Doubleday, London.
6. IEEE (2004). *Spectrum.*
7. Riley, J.C. (2001). *Rising Life Expectancy, A Global History,* Cambridge Univ. Press.
8. Saunders, S.C. (2007). *Reliability, Life Testing and the Prediction of Service Lives,* Springer, NY.
9. Smith, L.B. (2006). *Treason in Tudor England*, Pimlico Edition.
10. White, M. (2002). *The Pope and the Heretic,* Harper Collins Publishers Inc., NY.

APPENDIX

A computational formula that Mathematica uses in order to compute the BSD hazard rate

```
NIntegrate[(1 - capf[u, α, β])*(1 - capf[u, ∈, δ]), {u, x, 150}]/
 ((1 + (-1 - Erf[(-(1/Sqrt[x/β]) + Sqrt[x/β])/(Sqrt[2]*α)])/2)*
  (1 + (-1 - Erf[(-(1/Sqrt[x/δ]) + Sqrt[x/δ])/(Sqrt[2]*∈)])/2))NIntegrate[(1. + 0.5*(-1. -
1.*Erf[(-1./Sqrt[u/β] + Sqrt[u/β])/(Sqrt[2.]*α)]))*
  (1. + 0.5*(-1. - 1.*Erf[(-1./Sqrt[u/δ] + Sqrt[u/δ])/(Sqrt[2.]*∈)])), {u, x, 150.}]/
 ((1. + 0.5*(-1. - 1.*Erf[(0.7071067811865475*(-1./Sqrt[x/β] + Sqrt[x/β]))/α]))*
  (1. + 0.5*(-1. - 1.*Erf[(0.7071067811865475*(-1./Sqrt[x/δ] +
Sqrt[x/δ]))/∈])))NIntegrate[(1 - capf[u, α, β])*(1 - capf[u, ∈, δ]), {u, x, 150}]/
 ((1 + (-1 - Erf[(-(1/Sqrt[x/β]) + Sqrt[x/β])/(Sqrt[2]*α)])/2)*
  (1 + (-1 - Erf[(-(1/Sqrt[x/δ]) + Sqrt[x/δ])/(Sqrt[2]*∈)])/2))NIntegrate[(1 - capf[u, α,
β])*(1 - capf[u, ∈, δ]), {u, x, 150}]/
 ((1 + (-1 - Erf[(-(1/Sqrt[x/β]) + Sqrt[x/β])/(Sqrt[2]*α)])/2)*
  (1 + (-1 - Erf[(-(1/Sqrt[x/δ]) + Sqrt[x/δ])/(Sqrt[2]*∈)])/2))NIntegrate[(1 - capf[u, α,
β])*(1 - capf[u, ∈, δ]), {u, x, 100}]/
 ((1 + (-1 - Erf[(-(1/Sqrt[x/β]) + Sqrt[x/β])/(Sqrt[2]*α)])/2)*
  (1 + (-1 - Erf[(-(1/Sqrt[x/δ]) + Sqrt[x/δ])/(Sqrt[2]*∈)])/2))NIntegrate[(1 - capf[u, α,
β])*(1 - capf[u, ∈, δ]), {u, x, 150}]/
 ((1 + (-1 - Erf[(-(1/Sqrt[x/β]) + Sqrt[x/β])/(Sqrt[2]*α)])/2)*
  (1 + (-1 - Erf[(-(1/Sqrt[x/δ]) + Sqrt[x/δ])/(Sqrt[2]*∈)])/2))NIntegrate[(1 - capf[u, α,
β])*(1 - capf[u, ∈, δ]), {u, x, 150}]/
 ((1 + (-1 - Erf[(-(1/Sqrt[x/β]) + Sqrt[x/β])/(Sqrt[2]*α)])/2)*
  (1 + (-1 - Erf[(-(1/Sqrt[x/δ]) + Sqrt[x/δ])/(Sqrt[2]*∈)])/2))NIntegrate[(1 - capf[u, α,
β])*(1 - capf[u, ∈, δ]), {u, x, 150}]/
 ((1 + (-1 - Erf[(-(1/Sqrt[x/β]) + Sqrt[x/β])/(Sqrt[2]*α)])/2)*
  (1 + (-1 - Erf[(-(1/Sqrt[x/δ]) + Sqrt[x/δ])/(Sqrt[2]*∈)])/2))NIntegrate[(1 - capf[u, α,
β])*(1 - capf[u, ∈, δ]), {u, x, 150}]/
 ((1 + (-1 - Erf[(-(1/Sqrt[x/β]) + Sqrt[x/β])/(Sqrt[2]*α)])/2)*
  (1 + (-1 - Erf[(-(1/Sqrt[x/δ]) + Sqrt[x/δ])/(Sqrt[2]*∈)])/2))NIntegrate[(1 - capf[u, α,
β])*(1 - capf[u, ∈, δ]), {u, x, 150}]/
 ((1 + (-1 - Erf[(-(1/Sqrt[x/β]) + Sqrt[x/β])/(Sqrt[2]*α)])/2)*
  (1 + (-1 - Erf[(-(1/Sqrt[x/δ]) + Sqrt[x/δ])/(Sqrt[2]*∈)])/2))
```

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 361-366

UNIVERSALLY OPTIMAL NEIGHBOR BALANCED CIRCULAR DESIGNS

Naqvi Hamad[1] and **Muhammad Hanif**[2]
[1] National College of Business Administration & Economics, Lahore, Pakistan.
[2] Department of Mathematics, Lahore University of Management Sciences, Lahore, Pakistan. Email: hanif@lums.edu.pk

ABSTRACT

In this paper, we have developed some series of nearest neighbor balanced circular designs which are universally optimal. Series of $v = m$, m is odd prime, is constructed using primitive roots. A series for nearest neighbor balanced designs for all number of treatments, i.e. for odd and even number of treatments, (n>2) is first time constructed in literature. In all developed series, for any fixed treatment θ, all other treatments appear once as left and once as right neighbor. In our designs all treatment differences are estimated with equal precision.

KEYWORDS

Binary blocks; Universally optimal design; Primitive roots; Nearest neighbor binary circular designs

1. INTRODUCTION

A neighbor design is one-dimensional if neighbor effects are controlled in only one direction. One-dimensional neighbor designs are used in circular plates in biometrics and in block design setup in the field of agriculture where each block is a single line of plots and blocks are well separated. In block design setup, border plots are needed on both ends of every block to make the design circular. An example of one-dimensional neighbor circular design can be seen in the agriculture on mountains where crops are cultivated on terraces in such a way that these form a natural circular blocks, each cultivated plot is nearest neighbor to the next and hence along with the effect of its own it may yield the effects of its neighboring plots. Another example of one-dimensional neighbor circular design can be seen in the biometrics where a test called the ousterlogy/ouchterlony gel diffusion test is used to investigate the relationship between antigens. Rees (1967) introduced one-dimensional neighbor balanced designs for the ouchterlony gel diffusion test. Then several researcher worked on it see for reference Misra et al. (1991), Chaure and Misra (1996), Druilhet (1999), Bailey and Druihet (2004), Mingyao et al. (2007), Nutan (2007), Hamad et al. (2010). Ahmed and Akhtar (2011).

2. UNIVERSALLY OPTIMAL NEIGHBOR DESIGNS

It has been shown by various researchers that circular binary neighbor balanced designs are universally optimal for the estimation of treatment effects and neighbor effects. Druilhet (1999) proved that a design which is circular, binary, equi-replicated, pair wise balanced and equi-neighbor balanced is universally optimal. Universal optimality given by Keifer (1975) is a strong family of optimality criteria which includes

A-optimality, D-optimality and E-optimality criteria as particular cases. This criterion requires maximum trace and complete symmetry from the information matrix of a design. One can see Druilhet (1999), Bailey and Druihet (2004) and Mingyao et al. (2007) for the universal optimality of one-dimensional neighbor balanced circular designs. Our constructed series generate neighbor balanced designs that meet the universal optimality criterion and conditions given in those research papers.

The constructed designs are universally optimal for one-sided and two-sided neighbor effects under models M_1 and M_2 given below;

$$Y_{ij} = \mu + \tau_{(i,j)} + \beta_j + \varphi_{(i-1,j)} + \varepsilon_{ij} \qquad (M_1)$$
$$Y_{ij} = \mu + \tau_{(i,j)} + \beta_j + \varphi_{(i-1,j)} + \varphi_{(i+1,j)} + \varepsilon_{ij} \qquad (M_2)$$

Y_{ij} is the response from the ith plot in the jth block, μ is overall mean, $\tau_{(i,j)}$ is direct effect of the treatment in the ith plot of jth block, β_j is the effect of the jth block, $\varphi_{(i-1,j)}$ is the left neighbor effect due to the treatment in the (i-1)th plot of jth block, $\{\varphi_{(i-1,j)}, \varphi_{(i+1,j)}\}$ is the un-differentiated neighbor effect due to the treatment in (i-1)th plot and (i+1)th plot of jth block, i.e., neighbor effect due to left treatment is same to the neighbor effect of right treatment and ε_{ij} is error assumed to be independent and normal.

3. UNIVERSALLY OPTIMAL NEAREST NEIGHBOR BALANCED DESIGNS FOR ODD PRIME USING PRIMITIVE ROOTS

We have used the smallest primitive root x of prime number m to construct neighbor designs. A binary series of neighbor balanced designs for odd prime number of treatments is generated through smallest primitive root, when each treatment has every other treatment as neighbor exactly once to its left and exactly once to its right. No series of binary neighbor balanced circular design for odd prime number of treatments is ever developed in literature for which each treatment has remaining treatments as neighbor exactly once to its left and exactly once to its right.

3.1 Theorem 1

Let $v = m$ treatments, where m be the prime number and x be the primitive root of m. The initial block consisting of following treatments is $\mathrm{I} = \left\{ 1, x^{m-2}, x^{m-3}, \ldots, x \right\}$. This initial block when developed under mod (v) generates a series of incomplete binary nearest neighbor balanced circular design with parameters $k = r = m$-1, $b = v$, λ = 2. In this design, for a fixed treatment θ, every other treatment occur as neighbor exactly once to the left and exactly once to the right.

Proof: Let the k = (m-1) distinct treatments appearing in circular initial block $\mathrm{I} = \left\{ 1, x^{m-2}, x^{m-3}, \ldots, x \right\}$ under modulo v. From the initial block the forward and backward differences are $\pm(x^{m-2} - 1) \pm (x^{m-3} - x^{m-2}), \pm(x^{m-4} - x^{m-3}), \ldots\ldots\ldots, \pm(1 - x)$. The remaining blocks are derived from the initial block by cycling the treatment. Among the totality of forward and backward differences all differences appears twice giving λ = 2. In each forward and backward difference of all blocks there exists positive difference for each negative difference and it shows that for any fixed treatment θ, all other treatments occur as neighbor once as left and once as right neighbor. The above initial block develops a series of binary neighbor balanced circular designs with parameters $k = r = m$-1, $b = v$, λ = 2.

Corollary 3.1.

Designs developed through theorem 1 are balanced incomplete block designs (BIBD). These designs fulfill following requirements of BIBD;

(1) $bk = rv = N$;

(2) $r(k\text{-}1) = \lambda\ (v - 1)$, ($\lambda$ in BIBD is quite different from λ of neighbor design);

(3) $b = v$.

When design is neighbor balanced and pair wise balanced then all treatment differences are estimated with equal precision

Example 3.1

Let $v = m = 17$, $x = 3$. The binary initial block of size 16 is;

I = (1, 6, 2, 12, 4, 7, 8, 14, 16, 11, 15, 5, 13, 10, 9, 3).

Forward and backward differences of initial block are;

$$= \pm(5), \pm(-4), \pm(10), \pm(-8), \pm(3), \pm(1), \pm(6), \pm(2) \pm (-5), \pm(4), \pm(-10),$$
$$\pm(8), \pm(-3), \pm(-1), \pm(-6), \pm(-2)\,.$$

Each difference is repeated twice, irrespective of mathematical signs, which shows that each treatment would occur as neighbor with every treatment equally. In backward and forward differences, there is positive difference for each negative difference which shows that for fixed treatment θ, the rest treatments occur as neighbor once to the left and once to the right giving $\lambda = 2$. Sum of all these differences is equal to zero. The remaining binary blocks can be obtained cyclically under modulo 17 through initial block as;

(2, 7, 3, 13, 5, 8, 9, 15, 0, 12, 16, 6, 14, 11, 10, 4), (3, 8, 4, 14, 6, 9, 10, 16, 1, 13, 0, 7, 15, 12, 11, 5), (4, 9, 5, 15, 7, 10, 11, 0, 2, 14, 1, 8, 16, 13, 12, 6), (5, 10, 6, 16, 8, 11, 12, 1, 3, 15, 2, 9, 0, 14, 13, 7), (6, 11, 7, 0, 9, 12, 13, 2, 4, 16, 3, 10, 1, 15, 14, 8), (7, 12, 8, 1, 10, 13, 14, 3, 5, 0, 4, 11, 2, 16, 15, 9), (8, 13, 9, 2, 11, 14, 15, 4, 6, 1, 5, 12, 3, 0, 16, 10), (9, 14, 10, 3, 12, 15, 16, 5, 7, 2, 6, 13, 4, 1, 0, 11), (10, 15, 11, 4, 13, 16, 0, 6, 8, 3, 7, 14, 5, 2, 1, 12), (11, 16, 12, 5, 14, 0, 1, 7, 9, 4, 8, 15, 6, 3, 2, 13), (12, 0, 13, 6, 15, 1, 2, 8, 10, 5, 9, 16, 7, 4, 3, 14), (13, 1, 14, 7, 16, 2, 3, 9, 11, 6, 10, 0, 8, 5, 4, 15), (14, 2, 15, 8, 0, 3, 4, 10, 12, 7, 11, 1, 9, 6, 5, 16), (15, 3, 16, 9, 1, 4, 5, 11, 13, 8, 12, 2, 10, 7, 6, 0), (16, 4, 0, 10, 2, 5, 6, 12, 14, 9, 13, 3, 11, 8, 7, 1), (0, 5, 1, 11, 3, 6, 7, 13, 15, 10, 14, 4, 12, 9, 8, 2).

These blocks yield a binary neighbor balanced circular design with parameters: $b = v = 17$, $r = k = 16$ and $\lambda = 2$. The above design is balanced incomplete block design with $\lambda = 15$.

Note: Catalogue of above theorem 1 is given for $v = 43$. For $v > 43$, neighbor designs can be generated by the theorem.

Table 1 (Theorem 3.1)
Binary Neighbor balanced Circular Designs for $v = m$ & $\lambda = 2$

m	x	Initial block
5	2	(1, 3, 4, 2)
7	3	(1, 5, 4, 6, 2, 3)
11	2	(1, 6, 3, 7, 9, 10, 5, 8, 4, 2)
13	2	(1, 7, 10, 5, 9, 11, 12, 6, 3, 8, 4, 2)
17	3	(1, 6, 2, 12, 4, 7, 8, 14, 16, 11, 15, 5, 13, 10, 9, 3)
19	2	(1, 10, 5, 12, 6, 3, 11, 15, 17, 18, 9, 14, 7, 13. 16, 8, 4, 2)
23	5	(1, 14, 12, 7, 6, 15, 3, 19, 13, 21, 18, 22, 9, 11, 16, 17, 8, 20, 4, 10, 2, 5)
29	2	(1, 15, 22, 11, 20, 10, 5, 17, 23, 26, 13, 21, 25, 27, 28, 14, 7, 18, 9, 19, 24, 12, 6, 3, 16, 8, 4, 2)
31	3	(1, 21, 7, 23, 18, 6, 2, 11, 14, 15, 5, 12, 4, 22, 28, 30, 10, 24, 8, 13, 25, 29, 20, 17, 16, 26, 19, 27, 9, 3)
37	2	(1, 19, 28, 14, 7, 22, 11, 24, 12, 6, 3, 20, 10, 5, 21, 29, 33, 35, 36, 18, 9, 23, 30, 15, 26, 13, 25, 31, 34, 17, 27, 32, 16, 8, 4, 2)
41	6	(1, 7, 8, 15, 23, 38, 20, 17, 37, 13, 9, 22, 31, 12, 2, 14, 16, 30, 5, 35, 40, 34, 33, 26, 18, 3, 21, 24, 4, 28, 32, 19, 10, 29, 39, 27, 25, 11, 36, 6)
43	3	(1, 29, 24, 8, 17, 20, 21, 7, 31, 39, 13, 33, 11, 18, 6, 2, 15, 5, 16, 34, 40, 42, 14, 19, 35, 26, 23, 22, 36, 12, 4, 30, 10, 32, 25, 37, 41, 28, 38, 27, 9, 3)

4. UNIVERSALLY OPTIMAL NEIGHBOR DESIGN FOR TOTAL EFFECTS

Let n distinct treatments be assigned numbers 0 to n-1 and are arranged in a following base block $B_1 = [0, 1,..., n-1\}$. The remaining blocks are obtained under modulo n by multiplying base block with 2, 3, …, n-1 respectively. Each block is kept binary. When n is odd prime then all blocks have equal block size ($k = n$) and when n is not odd prime then blocks have different block sizes. These blocks give exactly r replicates of each treatment. To obtain neighbor balanced-ness some other blocks are obtained by cycling the treatments of those blocks whose size is less than n ($k < n$).

These blocks generates a circular neighbor balanced design if:

1. The difference of any two neighbor treatments in a given direction; say clockwise, is from ±1 to ±(v-1);
2. There is positive difference for each negative difference;
3. The sum of differences is equal to zero;
4. Treatments in each block are circularly ordered;
5. All ordered pair of distinct treatments appear once in all blocks.

4.1 Theorem 4.1

Let n be any positive number, $n \geq 3$. Then there exists a neighbor balanced circular design which is universally optimal for total effects. In such designs for any fixed treatment, all other treatments appear as neighbor once as left and once as right neighbor. This design will be proper if n and all n-1 numbers have greatest common divisor (gcd) = 1, otherwise design will be non-proper, i.e., design with unequal block sizes.

Proof:

Let we have $v = n$ treatments numbered in a set $X = [0, 1 \ . \ . \ . \ n-1]$. The collection of blocks under modulo n is obtained as;

$B_i = [0, i, 2i, 3i, \ldots, (n-1)i]$ where $i = 1, 2, 3, \ldots, n$-1.

For each i there is a block and each block is binary (no repetition of treatment in a block is allowed). The design will have exactly n-1 blocks if each block size is equal to n and this will be only possible when each i and n have greatest common divisor (gcd) =1. When i and n have more than one common divisor then B_i blocks have unequal block sizes. More blocks are generated cyclically under modulo n from those blocks whose size is less than n. Thus there are more than n-1 blocks in a non-proper design.

Example 4.1

Let $v = n = 9$ and $i = 1, 2, 3, 4, 5, 6, 7, 8$. The collection of blocks under modulo n is obtained as;

$B_i = [0, i, 2i, 3i, \ldots, (n-1)i]$.
$B_1 = [0, 1, 2, 3, 4, 5, 6, 7, 8]$, $B_2 = [0, 2, 4, 6, 8, 1, 3, 5, 7]$, $B_3 = [0, 3, 6]$,
$B_4 = [0, 4, 8, 3, 7, 2, 6, 1, 5]$, $B_5 = [0, 5, 1, 6, 2, 7, 3, 8, 4]$, $B_6 = [0, 6, 3]$,
$B_7 = [0, 7, 5, 3, 1, 8, 6, 4, 2]$, $B_8 = [0, 8, 7, 6, 5, 4, 3, 2, 1]$,

Other blocks generated cyclically under modulo n from those blocks whose size is less than n are;

$B_9 = [1, 4, 7]$, $B_{10} = [2, 5, 8]$ generated cyclically from B_3,
$B_{11} = [1, 7, 4]$, $B_{12} = [2, 8, 5]$ generated cyclically from B_6.

These 12 blocks generate neighbor balanced circular design and in this design each ordered pair of distinct treatments occur once in all circular blocks which shows that the treatments in all blocks appear with every other treatment once as a left and once as a right neighbor. These blocks yield a binary neighbor balanced circular design with parameters: $b = 12$, $v = 9$, $r = 8$ and $\lambda = 2$.

Note: In following catalogue, all blocks up to $v = 10$ are given. For other v, neighbor balanced designs can be generated by the theorem.

Table 2 (Theorem 4.1)
Neighbor Balanced Circular Designs for $v = n$ and $\lambda = 2$

v	B	All blocks
3	2	(0, 1, 2), (0, 2, 1)
4	4	(0, 1, 2, 3), (0, 2), (0, 3, 2, 1), (1, 3)
5	4	(0, 1, 2, 3, 4), (0, 2, 4, 1, 3), (0, 3, 1, 4, 2), (0, 4, 3, 2, 1)
6	9	(0, 1, 2, 3, 4, 5), (0, 2, 4), (0, 3), (0, 4, 2), (0, 5, 4, 3, 2, 1), (1, 3, 5), (1, 4), (2, 5), (1, 5, 3)
7	6	(0, 1, 2, 3, 4, 5, 6), (0, 2, 4, 6, 1, 3, 5), (0, 3, 6, 2, 5, 1, 4), (0, 4, 1, 5, 2, 6, 3), (0, 5, 3, 1, 6, 4, 2), (0, 6, 5, 4, 3, 2, 1)
8	12	(0, 1, 2, 3, 4, 5, 6, 7), (0, 2, 4, 6), (0, 3, 6, 1, 4, 7, 2, 5), (0, 4), (0, 5, 2, 7, 4, 1, 6, 3), (0, 6, 4, 2), (0, 7, 6, 5, 4, 3, 2, 1), (1, 3, 5, 7), (1, 5), (2, 6), (3, 7), (1, 7, 5, 3)
9	12	(0, 1, 2, 3, 4, 5, 6, 7, 8), (0, 2, 4, 6, 8, 1, 3, 5, 7), (0, 3, 6), (0, 4, 8, 3, 7, 2, 6, 1, 5), (0, 5, 1, 6, 2, 7, 3, 8, 4), (0, 6, 3), (0, 7, 5, 3, 1, 8, 6, 4, 2), (0, 8, 7, 6, 5, 4, 3, 2, 1), (1, 4, 7), (2, 5, 8), (1, 7, 4), (2, 8, 5)
10	16	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9), (0, 2, 4, 6, 8), (0, 3, 6, 9, 2, 5, 8, 1, 4, 7), (0, 4, 8, 2, 6), (0, 5), (0, 6, 2, 8, 4), (0, 7, 4, 1, 8, 5, 2, 9, 6, 3), (0, 8, 6, 4, 2), (0, 9, 8, 7, 6, 5, 4, 3, 2, 1), (1, 3, 5, 7, 9), (1, 5, 9, 3, 7), (1, 6), (2, 7), (3, 8), (4, 9), (1, 9, 7, 5, 3)

5. DISCUSSION

Neighbor effects (either natural or due to layout of plots) can deprive the results from its representative-ness. One-dimensional neighbor designs are important tools to control it. Neighbor designs for 4*n*+3 (power of prime) treatments and 4*n*-1 (power of prime) treatments exist in literature but no attention has been given for the construction of all odd prime numbers using primitive roots. Universally optimal neighbor balanced deigns are developed for odd prime using primitive roots.

Bailey and Druihet (2004) showed that a circular design neighbor-balanced at distances 1 and 2, having all ordered pair of distinct treatments is universally optimal for total effects. Mingyao et al. (2007) generalized this result to distance 2γ. When each *i* and *n* have greatest common divisor (gcd) =1 then theorem 4.1 generates neighbor balanced circular designs which are same as given by Mingyao et al. (2007) and universally optimal for the estimation of the total effects. The optimality is under assumption that all treatments are uncorrelated and have common variance. When *i* and *n* have more than one common divisor then one condition of pair-wise balanced for universally optimal design is not achieved. We can say that theorem 3 of Mingyao et al. (2007) for the construction of circular neighbor balanced design for odd prime becomes a special case of our theorem 4.1.

REFERENCES

1. Ahmed, R. and Akhtar, M. (2011). Designs balanced for neighbor effects in circular blocks of size six. *J. Statist. Plann. Infer.,* 141, 687-691.
2. Bailey, R.A. and Druilhet, P. (2004). Optimality of neighbor-balanced designs for total effects. *The Ann. Statist.,* 32, 4, 1650-1661.
3. Chaure, K. and Misra, B.L., (1996). On construction of generalized neighbor designs. *Sankhya* 58(B, pt 2), 245-253.
4. Druilhet, P. (1999). Optimality of neighbor balanced designs. *J. Statist. Plann. Infer.,* 81, 141-152.
5. Hamad, N., Zafaryab, M. and Hanif, M. (2010). Non-binary neighbor balance circular designs for $v = 2n$ and $\lambda = 2$. *J. Statist. Plann. Infer.,* 140, 3013-3016.
6. Kiefer, J. (1975). Construction and optimality of generalized Youden designs. *A Survey of Statistical Designs and Linear Models*.
7. Mingyao, A., Gennian, G. and Lingyau, C. (2007). Circular neighbor-balanced designs universally optimal for total effects. *J. Science in China Series A: Mathematics*, 50, 821-828.
8. Misra, B.L., Bhagwandas and Nutan, (1991). Families of neighbor designs and their analysis, *Commun. Statist. Simula. Comput.*, 20 (2 and 3), 427-436.
9. Nutan, S.M. (2007). Families of proper generalized neighbor designs, *J. Statist. Plann. Infer.,* 137, 1681-1686.
10. Rees, D.H., (1967). Some designs of use in serology, *Biometrics*, 23, 779-791.

Proc. 8th International Conference on Recent Advances in Statistics
Lahore, Pakistan – February 8-9, 2011, 367-376

APPLICATION OF FUZZY LOGIC IN ACADEMIC SETUP

Sehrish Waris[1] and **Zahoor Ahmad**[2]
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: [1]sehr.iman@yahoo.com; [2]zahoor.ahmed@uog.edu.pk

ABSTRACT

In past fuzzy logic has been used mostly for classification and control. In electrical engineering to deal with such problems, fuzzy logic use for control of Rotary cranes, hybrid system controller and vector control method has been optimized by using fuzzy controller, Multi factorial Fuzzy Approach for the Assessment of Ground water Quality. Fuzzy logic has been used with great success to control machines and consumer product. Fuzzy logic is not the answer of all technical problems but for control problem so in this paper we have been use fuzzy logic in academic setup in which we deal with logical variables if age is a variable and we say 60 years age person is old. Is age of 59 years and 3 month person is young? Such type of problems we can solve by using fuzzy logic technique and fuzzy logic is best one with dealing of such variables, And we discuss Fuzzy logic technique, and why this is important for some logical variables and we take a issue of academic performance of students, many variables have effect on students performance (GPA) but we discuss only such variables where fuzzy logic is required, and these variables are previous marks, study timing and their final GPA. We have estimate student GPA at any point by using fuzzy logic with the help of fuzzy rules based on information of previous marks and study timing.

INTRODUCTION

Why concept of fuzzy logic is required. In everyday life most of problems involved imprecise concept and order to handle the imprecise concepts, the conventional methods of set theory and numbers are insufficient and need to be some other concept and fuzzy logic is one of the concepts for this purpose.

Fuzzy logic systems are widely used for control, system identification and pattern recognition, but in this paper we use fuzzy logic for social variables, such variables which a computer can't explain better. Computers are very well known for counting numbers, applying various arithmetic operations, like multiply, division, subtraction also use for reasoning and also for detecting reasons. Detecting reasoning in which case either given values are true or false but truth values are given. In our daily life we have a lot of concert. That we have humans can easily describe, understand and explain to others but traditional mathematics including the set theory, fails to handle in a rational way, the concept "young" is an example for any specific person his/her age is precise. However, relating a particular age to "young" involves fuzziness and is sometimes confusing and difficult. What age is young and what age in not?

The nature of such question is deterministic and has nothing to do with stochastic concepts such as probability or possibility. Why is it that a 34.9 year old person is completely "young" while a 35.1 year old person is not "young" at all? Fuzzy logic is a new way of express the probability. Fuzzy logic is specifically designed to deal with imprecision of facts.

Prof. Manohar Lal says, fuzzy logic handles this type of variables for example "a healthy person" if we say a person is depress, so there are no. of reasons, and degree of depress is again a variable so depression is not something which can be deterministic, we take a example of everyday Old and Young, now we can't say person is old on a particular age, we can't say a person is young at 25 year old and next day he will be old, never. And he says how language can vary? Dinosaurs lived at the earth for a long period (about millions of years). It has not rained for a long period (say about six month). I had to wait for the doctor for a long period (about six hours). So how much variation in period representation by fuzzy logic we make a computer precise for these types of words.

The concept of Fuzzy Logic (FL) was conceived by Lotfi Zadeh, a professor at the University of California at Berkley, and presented not as a control methodology, but as a way of processing data by allowing partial set membership rather than crisp set membership or non-membership. This approach to set theory was not applied to control systems until the 70's due to insufficient small-computer capability prior to that time. Professor Zadeh reasoned that people do not require precise, numerical information input, and yet they are capable of highly adaptive control. If feedback controllers could be programmed to accept noisy, imprecise input, they would be much more effective and perhaps easier to implement.

Fuzzy logic deals with uncertainty in engineering by attaching degree of certain to the answer to logical questions. Commercially fuzzy logic has been used with great success to control machines and consumer products, fuzzy logic systems are simple to design, can be understood and implemented by non-specialist in control theory. Fuzzy logic is not the solution of all technical problems but for speed of implementation is important then fuzzy logic is useful technique. [James Vernon]

M. Hellmann **says** Basically, Fuzzy Logic (FL) is a multivalued logic that allows intermediate values to be defined between conventional evaluations like true/false, yes/no, high/low, etc. Notions like rather tall or very fast can be formulated mathematically and processed by Computers in order to apply a more human-like way of thinking in the programming of computers. Fuzzy systems are an alternative to traditional notions of set membership and logic that has its origins in ancient Greek philosophy.

Lotfi Zadeh says fuzzy logic is not the waves of the future. It is now! There are already hundreds of millions of dollars of successful, fuzzy logic based commercial products, everything from self-focusing cameras to washing machines that adjust themselves according to how dirty the clothes are, automobile engine controls, anti-lock braking systems, color-film developing systems, subway control systems and computer programs trading successfully in the financial markets.

Fuzzy logic was conceived by Steven D. Kachler as a better method for sorting and handling data but has proven to be a excellent choice for many control system applications since it mimics human control logic. It can be built into anything from small, hand-held products to large computerized process control systems. It uses an imprecise but very descriptive language to deal with input data more like a human operator. It is very robust and forgiving of operator and data input and often works when first implemented with little or no tuning.

Fuzzy logic is also use for dealing with complicated systems in simple way, and it is also possible for classification.

Ricketts, J.H.1 (July 2009) defined fuzzy rule base system and in which he discuss Mumdani fuzzy knowledge rule system. He says, people are required to make trip-scheduling decisions in their daily lives. They deal with a rich set of uncertainties – changes of venue, priority, and time of activities – choices in mode, route, and time of trips – information horizons – delays and changes in current activities, and in trips. Populations both evolve and contain a rich set of applicable strategies. And rule based system is also used for both to explain and predict responses of population or individuals and they can be constructed by human experts or by machine learning techniques.

Fuzzy logic is useful for modeling inference under imprecision and fuzzy logic can be used in a rule-based system. One system is a Mamdani Fuzzy Rule Based System (FRBS) with a database and a rule-base, and these have been previously adapted to classification problems although mostly using the Michigan approach (Cordon et al. 2004)

Miroslav Hudec and Mirko Vujosevic works on selection and classification of statistical data using fuzzy logic. And he defined two often used processes are data selection (database queries) and data classification and situations when classical {true, false} logic is not adequate in these two processes and offers fuzzy logic because the fuzzy logic is an approach to computing based on "degrees of truth" rather than the usual "true or false" logic. Fuzzy approach is suitable for statistical databases. Linguistic expressions like: high rate of unemployment or medium migration level etc., are very often used and it is useful to catch them and use in database queries and classification.

Ravi. Jain and Ajith. Abraham works on (*A Comparative Study of Fuzzy Classification Methods on Breast Cancer Data*) and in which they examine the performance of four fuzzy rule generation methods on Wisconsin breast cancer data. The first method generates fuzzy if-then rules using the mean and the standard deviation of attribute values. The second approach generates fuzzy if-then rules using the histogram of attributes values. The third procedure generates fuzzy if-then rules with certainty of each attribute into homogeneous fuzzy sets. In the fourth approach, only overlapping areas are partitioned. The first two approaches generate a single fuzzy if-then rule for each class by specifying the membership function of each antecedent fuzzy set using the information about attribute values of training patterns. The other two approaches are based on fuzzy grids with homogeneous fuzzy partitions of each attribute. The performance of each approach is evaluated on breast cancer data sets.

Salah Bouhouche et al. (2007) published a study based on **"***Quality Monitoring Using Principal Component Analysis and Fuzzy Logic Application in Continuous Casting Process* "and in which they deal with non linear system monitoring, based on a combined use of Principal Components Analysis (PCA) and fuzzy logic to process and quality monitoring. In which fuzzy logic was used to estimate the fault or defect according to the dynamic changes in the process inputs –outputs characterized by T2 Hoteling and Squared Prediction Error (SPE).

The motivation for fuzzy logic was expressed by Zadeh (1984) in the following way: "The ability of the human mind to reason in fuzzy terms is actually of a great advantage. Even though a tremendous amount of information is presented to the human senses in a given situation – an amount that would choke a typical computer – somehow the human mind has the ability to discard most of this information and to concentrate only on the information that is task relevant. This ability of the human mind to deal only with the information that is task relevant is connected with its possibility to process fuzzy information. By concentrating only on the task-relevant information, the amount of information the brain has to deal with is reduced to a manageable level.

H.Chris Tseng (2007) defined many internet applications need to deal with large amount of data collected from non-technical users and is imprecise and incomplete in nature. Well structured rules are hardly available in general applications and the nature and the pattern of the users can never be fully accounted. Soft computing techniques like fuzzy logic ideal for coping with this type of problems. And H.Chris Tseng defined fuzzy logic, the ability to model imprecise and qualitative knowledge and handle uncertainty are distinguished characteristics of fuzzy sets. Fuzzy logic is capable of addressing approximate or vague notions that are inherent in many information retrieval (IR) tasks.

Mostly on high level or university level some variables effect on students GPA. In which study timing, attendance, previous degree marks etc are effected variables. And these variables are logical variable, if we want to observe these variables affect and we categories these variable for easy interpretation, suppose we say if study timing is 5 hr then GPA is low and between 5 hr and 8 hr GPA is average and if timing is above 10 hr then GPA would be high. But logically it is not right because if study timing is 4 hr and 59 minutes so GPA is low and if study timing is 5 hr and 1 minute so GPA is average. By this means only 2 minutes low GPA convert into average. Here we want different GPA on different timing, because 2 minutes effect convert GPA low to average .Computer have not sense of such output, but human can understand such type of variables. Similarly other variables have this type of problem. For this type of problem we use fuzzy logic technique because it allows us a natural description.

MATERIAL AND METHODS

Logical variables:

In our data we have many variables, such variables that a human can better understand. And we have many variables that have effect on Academic performance of students in university. Here we take two input variables study hour and previous degree marks and output is GPA. Now we make a fuzzy inference system. Fuzzy inference system is a process of mapping from given input to an output by using fuzzy logic. There are five primary GUI tools for building fuzzy inference systems in the Fuzzy Logic Toolbox: the Fuzzy Inference System or FIS Editor, the Membership Function Editor, the Rule Editor, the Rule Viewer, and the Surface Viewer.

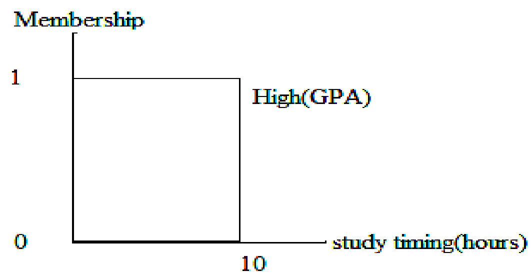
So firstly we make FIS Editor, The first step is to take the inputs and determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions. The input is always a crisp numerical value and the output is a fuzzy degree of membership (always in interval between 0 and 1).

We enter three variables in FIS editor in which two as a input and one is output, Study hour, percentage of previous degree marks, these are input variables and GPA is output variable and computer cannot explain well, for example if answer in the form of No or Yes and we say if study hour 6 so GPA will be average. Logically questions arise if study hour are 6hr and 5 min so what would be GPA? Are GPA should be change? And, if GPA is 5 hr and 56 min so GPA will be low? So for this type of problem in fuzzy logic we make if then rules. By rule we human can determine such variables better than computer. So in this technique in each point of input we have output. On 6 hr we have output of GPA and on the point of 5 hr and 56 min we have

FIS editor:
The FIS editor provide general information about a fuzzy inference system
Fuzzy inference system (FIS):

Fuzzy inference is the actual process of mapping from a given input to an output using fuzzy logic. The process consist of all the parts that we'll discuss next i.e., fuzzy sets, membership functions, fuzzy logic operators, and if-then rules

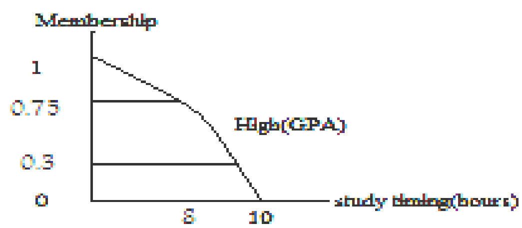
Firstly we explain fuzzy sets; mostly we have two types of sets. Crisp or classical sets and fuzzy Crisp sets:



(Figure 1)

In classical set theory membership of an object is always 0 or 1.This set is not reasonable because the change on membership function at 10 (hours). Here a fundamental problem exists, on the different hours membership value changes from 1 to 0 may be used.

Fuzzy sets:

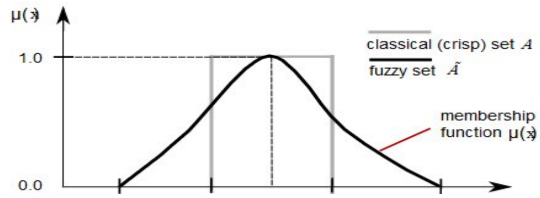


(Figure 2)

In fuzzy set theory membership values lies also between 0 and 1.

Membership function:

A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. and membership function is graphical representation of inputs and output. In which our fuzzy sets have different properties and membership functions represents the degree of truth. And peak of distribution shown in (figure: 3) shows the highest degree means close to 1 or 1. and both tails of distribution shows smallest degree close to 0 or 0.



(Figure 3)

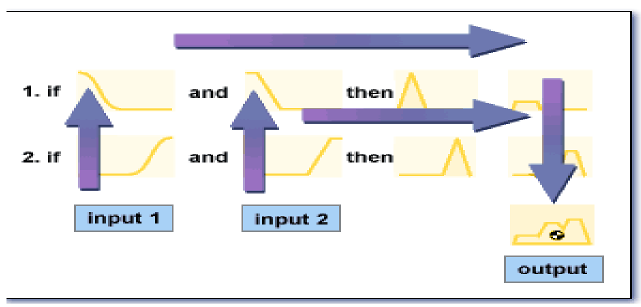
Fuzzy rules:

Fuzzy rules are the conditional statements that make fuzzy logic useful. A single fuzzy if-then rule assumes the form:

If x is A then y is B

where A and B are linguistic values defined by fuzzy sets on the ranges X and Y, respectively. The if-part of the rule *"x is A"* is called the antecedent or premise, while the then-part of the rule *"y is B"* is called the consequent or conclusion. In our study we take some rules by the help of finding of irum shehzadi 2009. In this finding our logical variables effect on students GPA but study timing is more effected then previous degree marks like this and by this information we make rules and gets final outputs.

RULES



(Figure 4)

1. If study timing is low and previous marks low then GPA is low
2. If study timing is average and previous marks low then GPA is low
3. If study timing is high and previous marks low then GPA is average
4. If study timing is low and previous marks average then GPA is average
5. If study timing is average and previous marks average then GPA is average

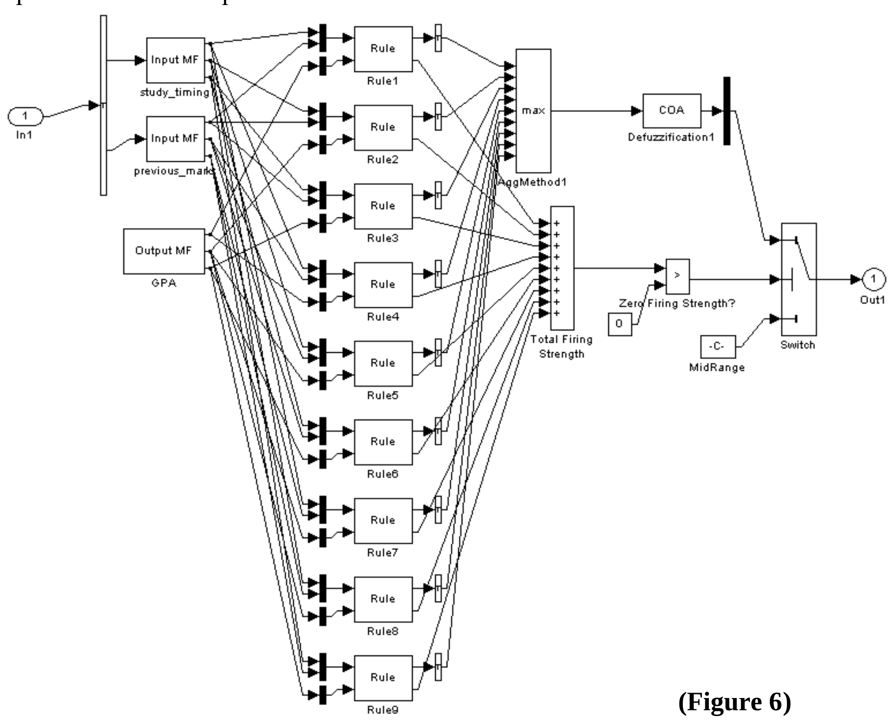
6. If study timing is high and previous marks average then GPA is high
7. If study timing is low and previous marks high then GPA is average
8. If study timing is average and previous marks high then GPA is high
9. If study timing is high and previous marks high then GPA is high

Simulation:

We can evaluate our FIS performance by using fuzzy logic controller block (figure: 5) in simulink model. The Fuzzy Logic Controller block generates a hierarchical block diagram that represents our fuzzy inference systems. Simulink is a control library which allows easy implementation of any control algorithm, including linear control, fuzzy logic, neural networks, etc.

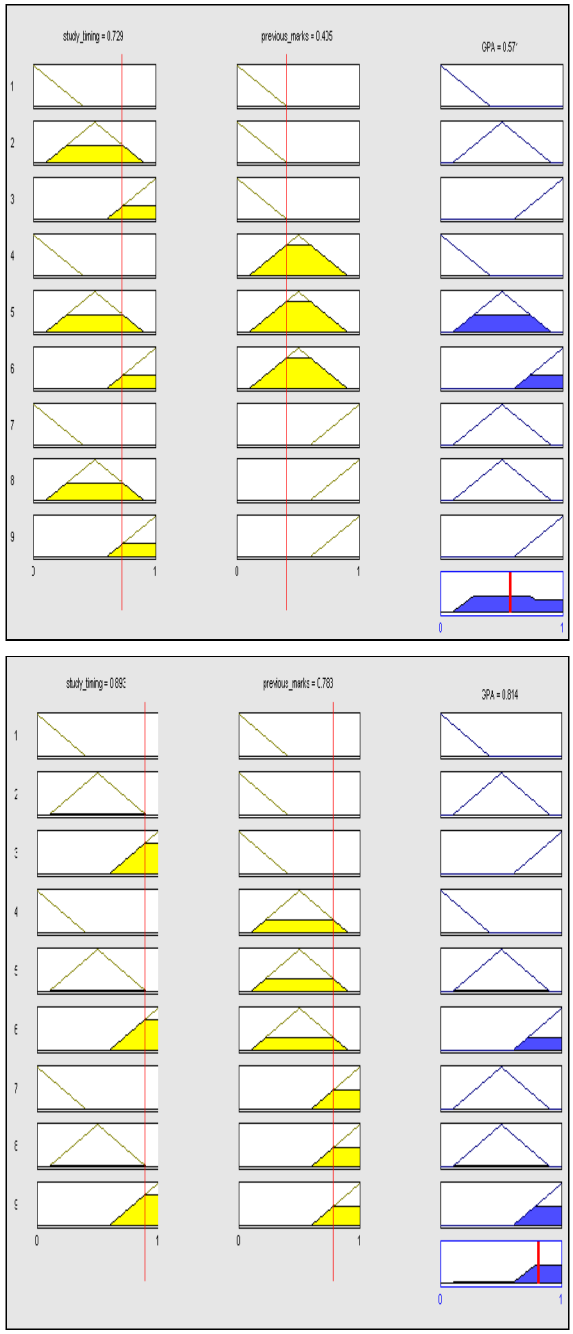
(Figure 5)

In (Figure 6) this simulink block use for evaluation of FIS and this block shows us how our rules attach with our inputs and output and after defuzzification gives us a single value output and by moving the both inputs we find a change output similarly each point of input we can find output value.

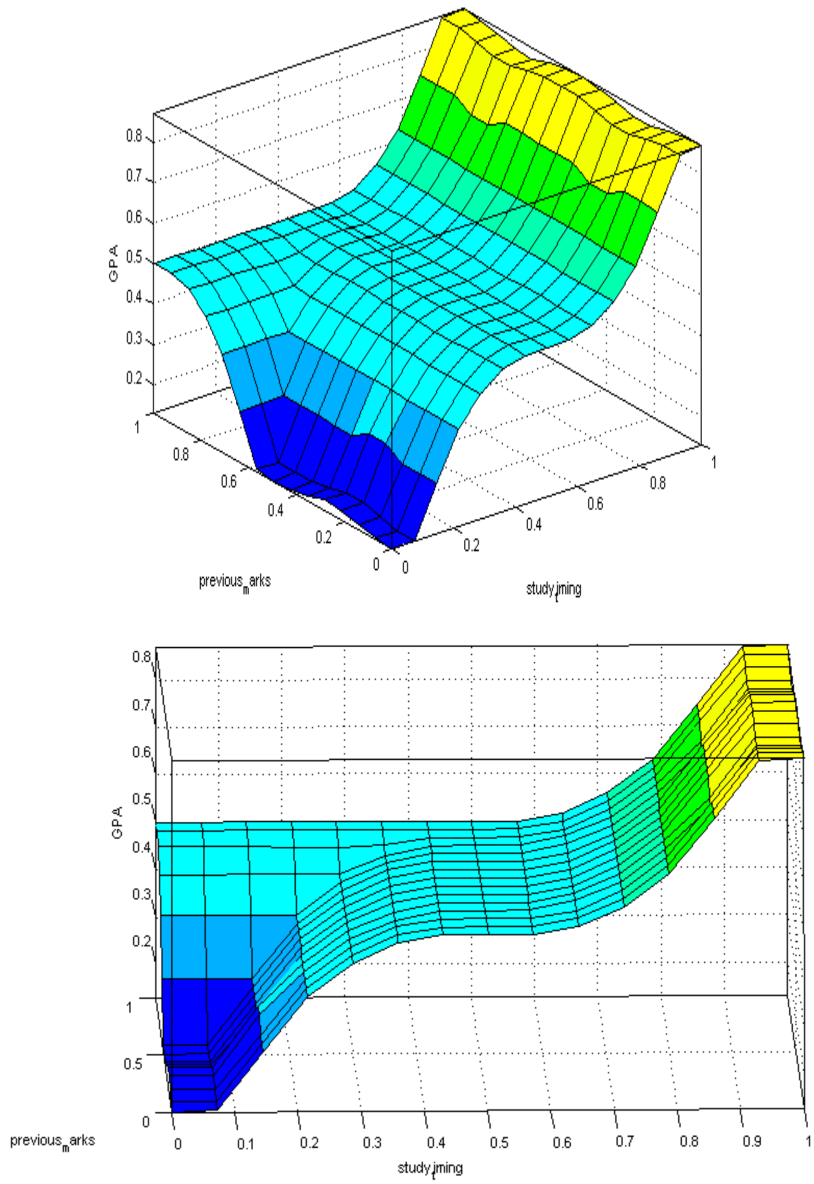


(Figure 6)

RESULTS AND DISCUSSION



(Figure 7)



(Figure 8)

In Figure 8, we generate 3-D surface from two inputs and one output. we can see that where lower marks and low timing then GPA is low and average and by increasing the study timing our GPA increase and move up to yellow colure where GPA high. And these all points combine by the information of rules and by changing each point we can find a different output. By this we can find weight of inputs variable.

REFERENCES

1. Sung-Wook Kim, Kwangsoo Kim, Joo-hyung Lee and Dong-il Cho (2001). Application of fuzzy logic to vehicle classification Algorithm in Loop. *Asian Journal of Control*, 3(1), 64-68.
2. Sourabh Dash, Raghunathan Rengaswamy and Venkat Venkatasubramanian (2002). Fuzzy logic based trend classification for fault diagnosis of chemical processes. *Computers and Chemical Engineering* 27, 347-362.
3. Paolo Dadone (2001). Design Optimization of Fuzzy Logic Systems. Ph.D. Dissertation, Virginia Polytechnic Institute and State University, USA.
4. Ravi Jain and Ajith Abraham (2003). A Comparative Study of Fuzzy Classification Methods on Breast Cancer Data. Presented at the *7th International Work Conference on Artificial and Natural Neural Networks*, IWANN'03, Spain.
5. H. Chris Tseng (2007). Internet Applications with Fuzzy Logic and Neural Networks, *Journal of Engineering and computing and Architecture*, 1(2).
6. Salah Bouhouche (2007). Quality Monitoring Using Principal Component Analysis and Fuzzy Logic Application in Continuous Casting Process. *Amer. J. App. Sci.*, 4(9), 637-644.
7. Miroslav Hudec and Mirko Vujošević "*Selection and Classification of Statistical Data Using Fuzzy Logic*" Faculty of Organizational Sciences, Jove Ilića 154, Beograd, Serbia.
8. Amjed A. Al-Mousa (2000). *Control of Rotary Cranes Using Fuzzy Logic and Time-Delayed Position Feedback Control*. Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University.
9. Ricketts, J.H. (2009). Tuning a modified Mamdani fuzzy rule base system with A genetic algorithm for travel decisions. *18th World IMACS* Australia
10. Natarajan Venkat Kumar, Samson Mathew and Ganapathiram Swaminathan (2010). Multifactorial Fuzzy Approach for the Assessment of Groundwater Quality. *J. Water Resource and Protection*, 2, 597-608.
11. Vinod Kumar and R.R. Joshi, (2005). Hybrid Controller based Intelligent Speed Control of Induction Motor. Journal *of Theoretical and Applied Information Technology*, 71-75.
12. Umit UNCU (2001). A Fuzzy Rule-Base Model for Classification of Sipormetic Fvc Graphs in Chronical Obstructive Pulmonary Diseases. *23rd Annual Conference Istanbul*, Turkey.

Author Index