Vol. 29

ISBN 978-969-8858-18-6

PROCEEDINGS

14th International Conference on Statistical Sciences:

Theme:
Statistics for Better Decision-Making and Development

March 14-16, 2016

at
Jinnah Sindh Medical University
Karachi, Pakistan

JOINTLY ORGANIZED BY

Jinnah Sindh Medical University
Rafiqui H.J. Shaheed Road, Karachi (Pakistan)
URL: http://www.jsmu.edu.pk/

Islamic Countries Society of Statistical Sciences
44-A, Civic Centre, Sabzazar, Multan Road, Lahore (Pakistan)
URL: http://www.isoss.net

SPONSORS

Published by: ISOSS, Lahore, Pakistan.

“All papers published in the

PROCEEDINGS

were accepted after formal peer review

by the experts in the relevant field.

Dr. Munir Ahmad
Editor

CONTENTS

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 1-14

COMPARATIVE ANALYSIS OF TEST MANAGEMENT TOOLS: SPIRA, HP QUALITY CENTER AND MERCURY QUALITY CENTER

Syed Fakhar Abbas, Mumtaz Ahmad, Zahid Miraj and **Mamoona Humayon**
University Institute of Information Technology, Rawalpindi, Pakistan

ABSTRACT

Software Testing plays an important role in improving software quality. Software engineers always have a goal: "There is always one more bug". To ensure quality, it is very hard to quantify amount of testing needed. Test management is one of the most crucial parts of the software life cycle but is often neglected or proper tools are not used which can make this process more efficient and manageable. Test management software's are available that can make the life of software engineers easier and are quite economical. In this paper, we will give the comparative analysis of different testing tools such as Spira, HP Quality Center and Mercury Quality Center on the basis of different test management activities.

KEYWORDS

Testing, Test Management Tools, Mercury QC, HP QC, Spira etc.

INTRODUCTION

Software Testing Process is carried out with intentions to find out errors. It is a major part of software development lifecycle and has a crucial role in the success of software systems. It is a technique aimed to assess the capabilities of a program and used to test other quality factors like integrity, security, reliability, usability, capability, efficiency, compatibility, portability etc [1]. The goal of Software Testing is to identify all possible defects in a software product.

Testing can be performed either manually or automated. In automated testing, the steps of manual testing are automated using different tools [2]. Automated Test Management keeps track of test plans, test cases, user requirements and defects. It also executes test cases and measures testing progress. **Table 1** shows the difference between manual and automated test management.

Table 1
Difference between Manual and Automated Test Management

Test Management (Manual)	Test Management (Automated)
Processes are ad-hoc and not repeatable across projects	Permanently recorded and reusable
Invisibility between test cases, requirements and defects	Test case, Requirement and defect are linked together
It is very difficult and time consuming to measure progress and productivity during testing.	Progress and productivity can be viewed in a graphical chart at a glance.
Sharing information across the project and getting real-time metrics related to the quality of the system being tested is difficult.	Associativity between the processes and a real time result.
Central repository that stores test results from all sources cannot be maintained	It's easy to maintain central repository.

It is crucial to maintain quality in development of software products and software testing is considered as an obvious way to achieve the goal of software quality. [6]. In process of software testing the most important task is test management. In test management activity tests are planned before their execution [4].

Test management is basically a process of managing test resources and objects such as test requirements, test cases, and test results [1]. In test management process every aspect of the testing cycle is properly managed [2]. Test management responsiveness of 3Ps: Processes (Processes involved in test generation), Products (Products of testing), and People (Team involved in testing) [1].

Ultimate goal of test management is to manage cost and quality [1].Management includes organizational configuration management, time, resources and quality management [2]. In addition the controlling and tracking process responsible of development of test cases their execution and management [8].

2. TEST MANAGEMENT PARAMETERS

From literature survey, we have identified eight (08) parameters for Test Management which includes Test Policy Management, User Management, Project Based test Management, Off Shore Support, Test Planning Support, Release Management, Bug Tracking and Analysis Report [1][3][4][5].

i) Test policy/Requirement management
This parameter is used to analyze whether test management tool has the capability to manage predefined requirements of the system. It relates to use case document management and provisioning.

ii) User Management
User management becomes really an important part when it comes to manage things and specifically while managing testing part. Tool for test management must have the capability to manage access rights of various users that is a specific user must only see what is relevant to him/her.

iii) **Project based test Management**
A test management tool must provision its users to manage testing of different projects in a specific time frame. In real life testing team has to manage testing of various projects at the same time.

iv) **Off Shore support**
Test management tool must provision user to email or report bugs online to the development team. This has become a basic requirement now a day.

v) **Test planning support**
Planning involves devising a time frame for testing of a project, module, and phase. It's important to allocate time for testing cycle and then assigning time to development for removal of bugs.

vi) **Release Management**
Tool for test management has to support management of different releases of projects and also management of bugs i.e. what bugs relate to which release.

vii) **Bug tracking**
Test management tool has to provide complete tracking of bugs reported by testing team i.e. the current status of bug whether fixed or not and currently with which user it is.

viii) **Analysis Reports**
Test Management tool has to present analysis reports for project. It must show different comparison analysis through different graphical charts and textual reports.

3. TOOLS FOR COMPARISON

Tools which we have included in our comparative analysis, that are currently used for test management are Spira, Mercury QC and HP QC. Some major features of these tools with screen shots are shown below and afterwards we have compared them on the bases of parameters that are discussed earlier.

3.1 Spira

A complete quality assurance solution that is responsible to manage requirement, test, bugs and issues with complete traceability is provided by Spira. Some common features of Spira are as under:

- ✓ To Manage Requirements
- ✓ To Manage Test Cases
- ✓ To Mange Defect Tracking
- ✓ To Mange Project
- ✓ To Release Management
- ✓ To Maintain Test Repository

Requirements Management

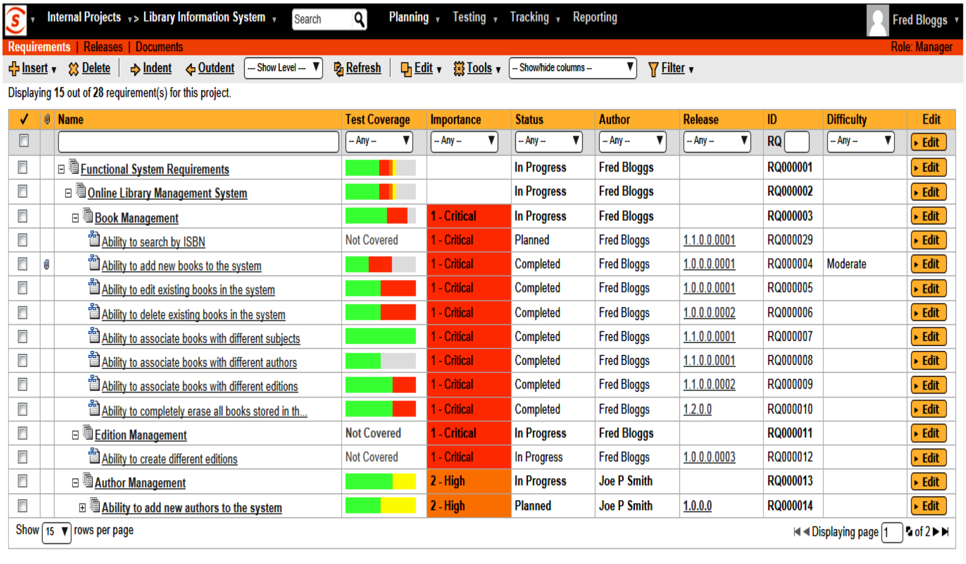


Fig. 1: Management of Requirements in Spira

Test Case Management

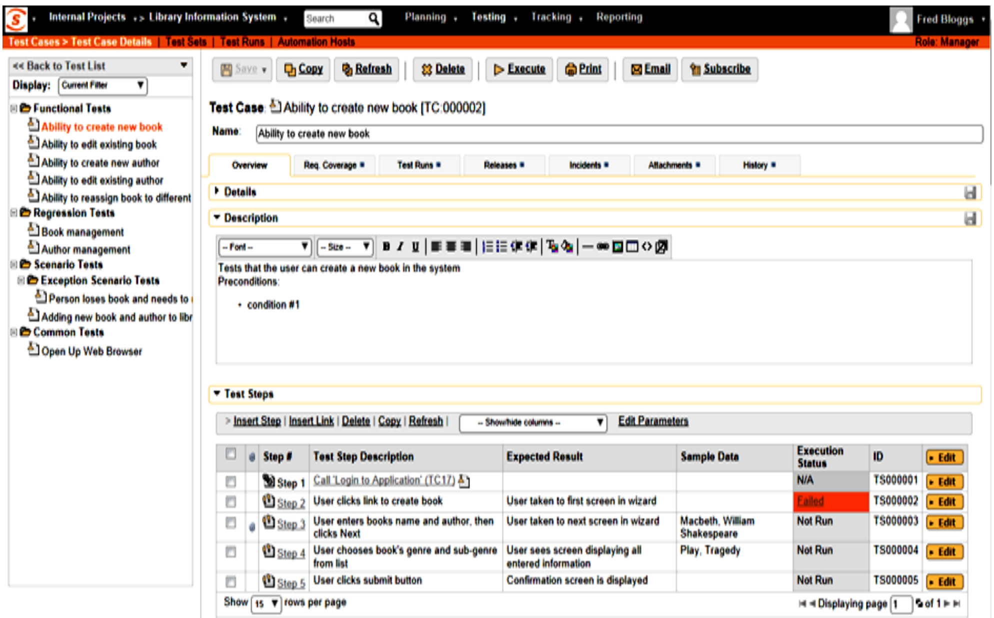


Fig. 2: Test Case Management in Spira

Defect Tracking

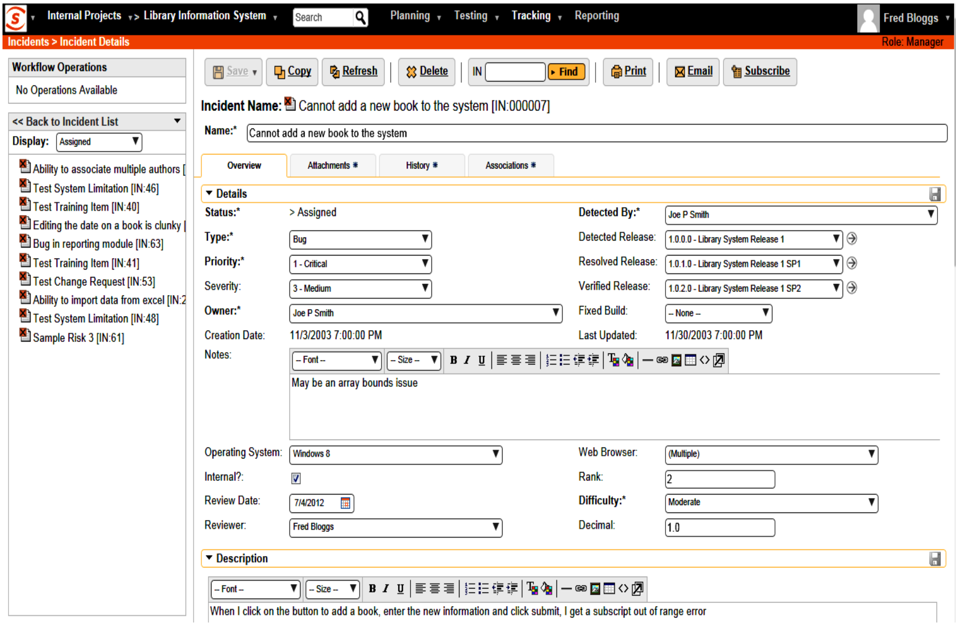


Fig. 3: Defect Tracking in Spira

Release Management

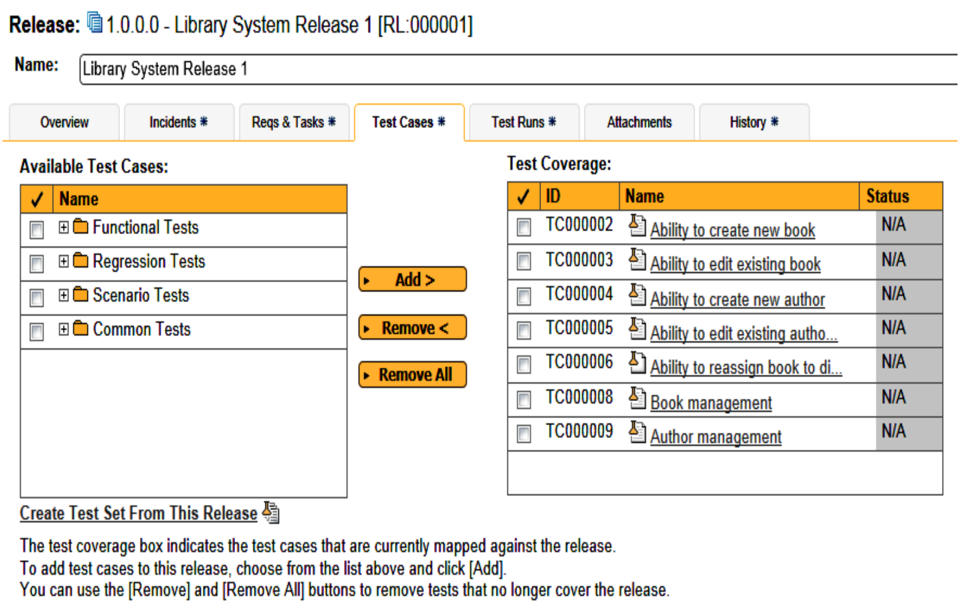


Fig. 4: Release Management in Spira

Reporting

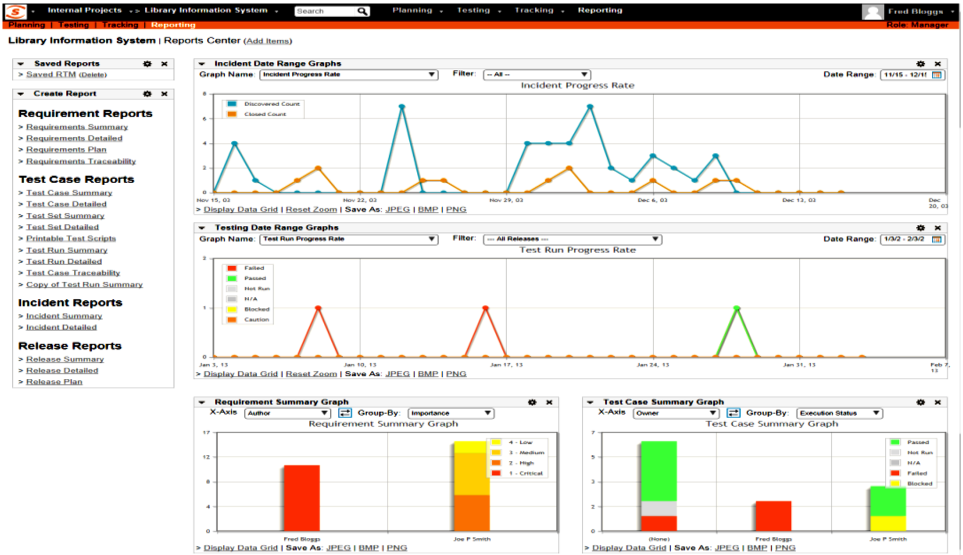


Fig. 5: Reporting in Spira

Project Management

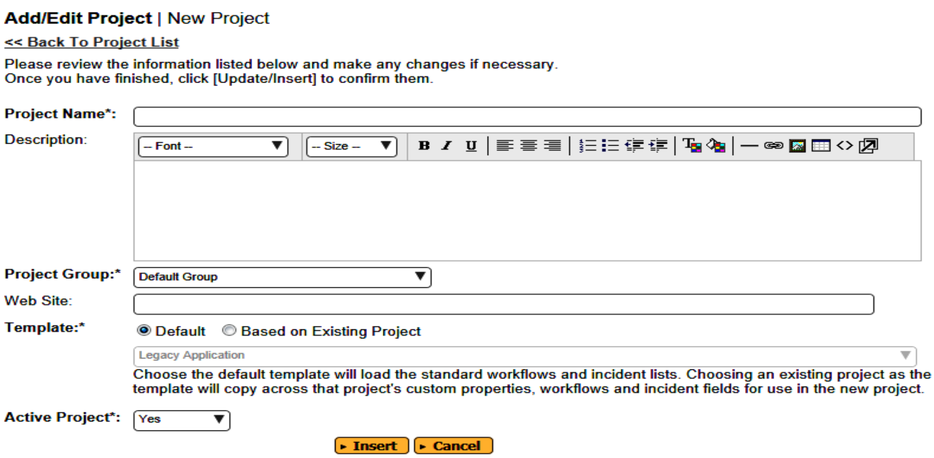


Fig. 6: Project Management in Spira

3.2 Mercury Quality Center (MQC)

Mercury Quality Center (MQC) is a web based tool for test management. It provides user privileges, test management along with analysis. MQC offers number of features and benefits. Some of them are:

- ✓ It supports complete testing process — requirements management; planning, building, scheduling, and executing tests; defect management.

- ✓ It allows teams to share and access testing assets globally, anywhere at any time via a browser interface.
- ✓ It supports and manages manual as well as automated tests
- ✓ It allows teams to investigate application readiness in testing process at any point in the testing process with integrated reports and graphs.

Requirement Management

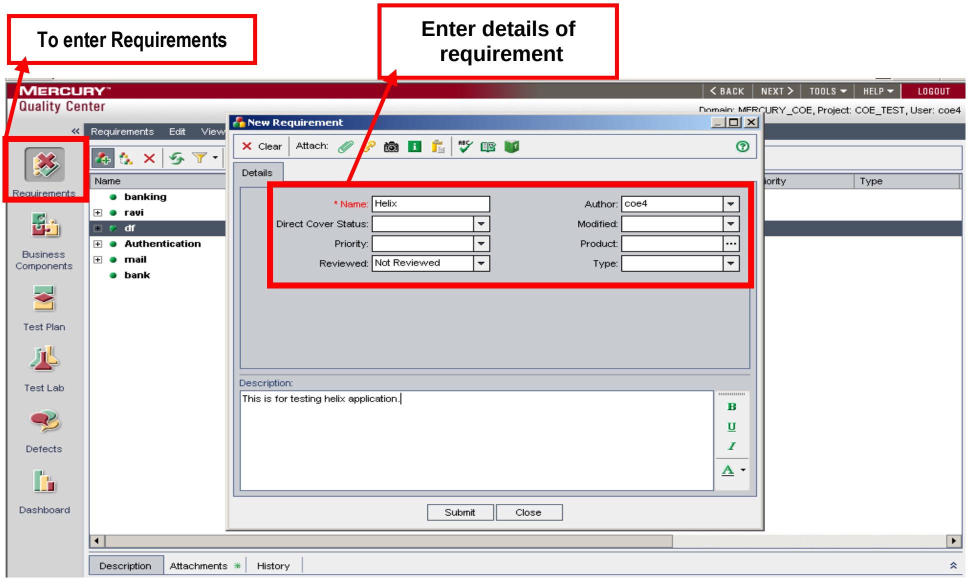


Fig. 7: Requirement Management in Mercury Quality Center

User Management (Login)

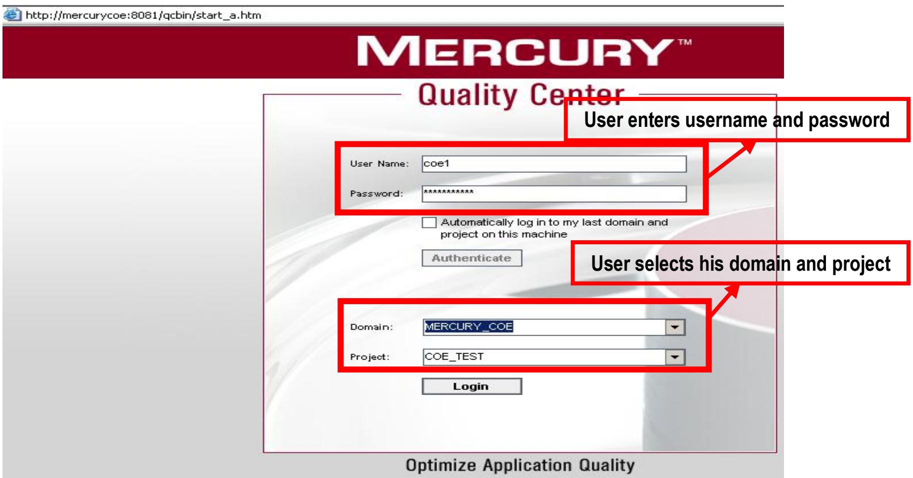


Fig. 8: Mercury Quality Center (Login)

Test Plan

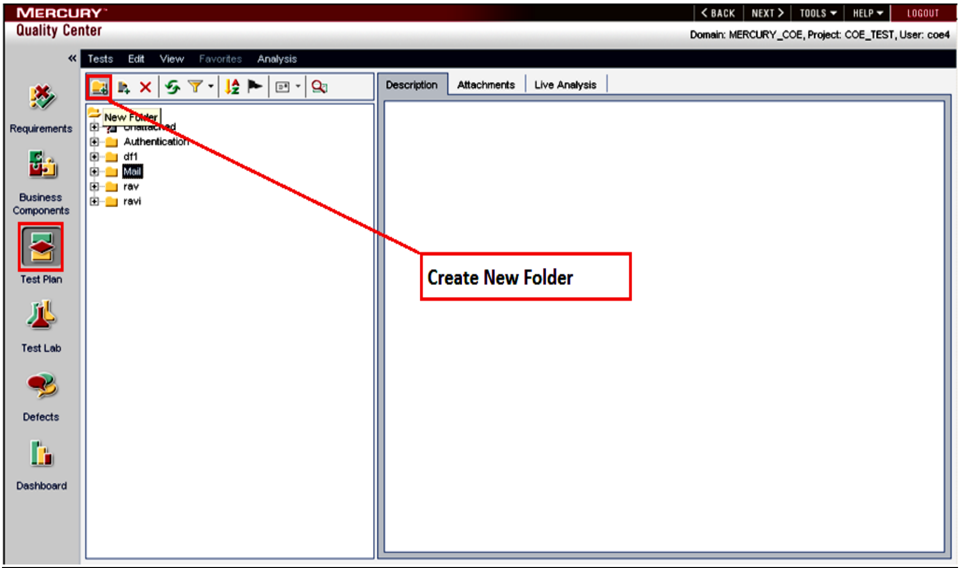


Fig. 9: Test Plan in Mercury Quality Center

Defect Module

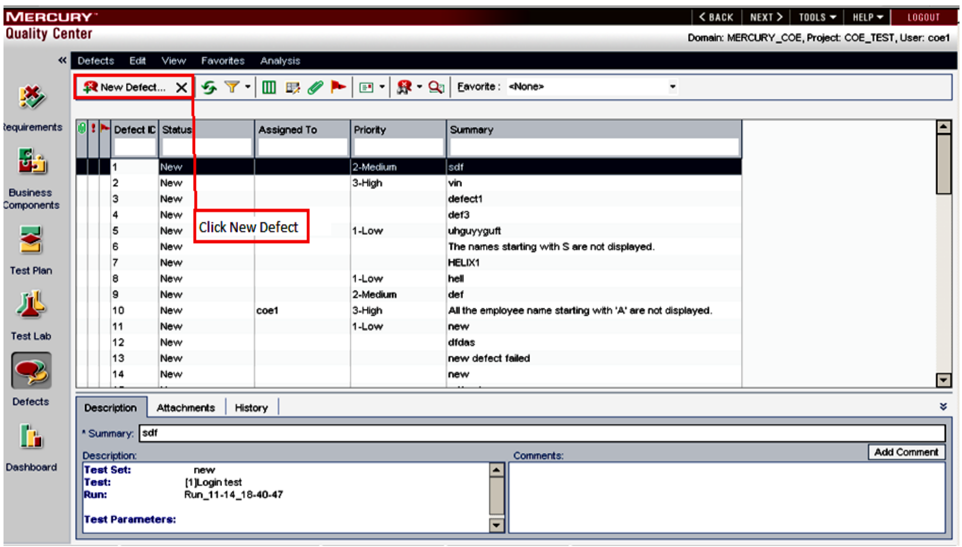


Fig. 10: Defect Management in Mercury Quality Center

Reporting

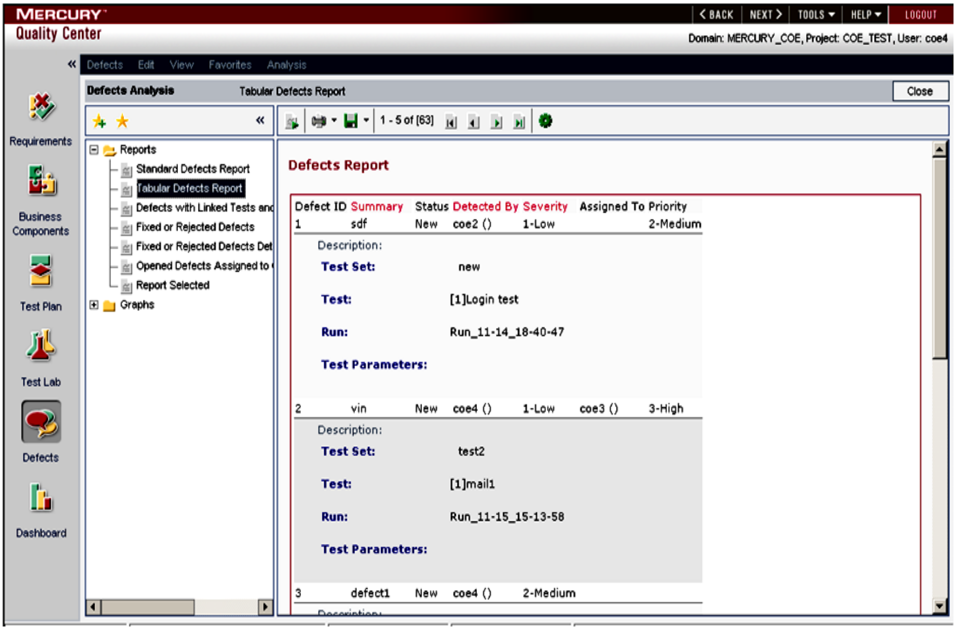


Fig. 11: Reporting in in Mercury Quality Center

3.3 HP Quality Center (HPQC)

HP Quality Center is developed in J2EE as front end and MS Sql Server as a backend data base. It runs on both UNIX and Windows environment. It supports activities like:

- ✓ It Manages Requirements
- ✓ It Test Design
- ✓ It Executes Tests
- ✓ It Manages Defects
- ✓ It Maintain Traceability of test cases
- ✓ Integration with other tools like WinRunner, QTP and Load Runner.

Releases and Cycles

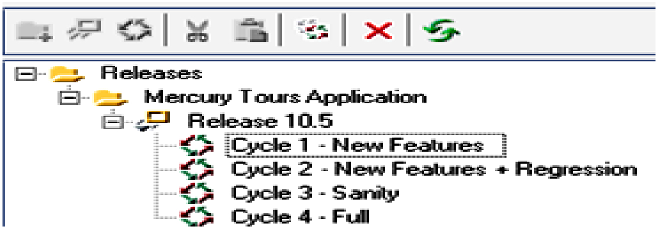
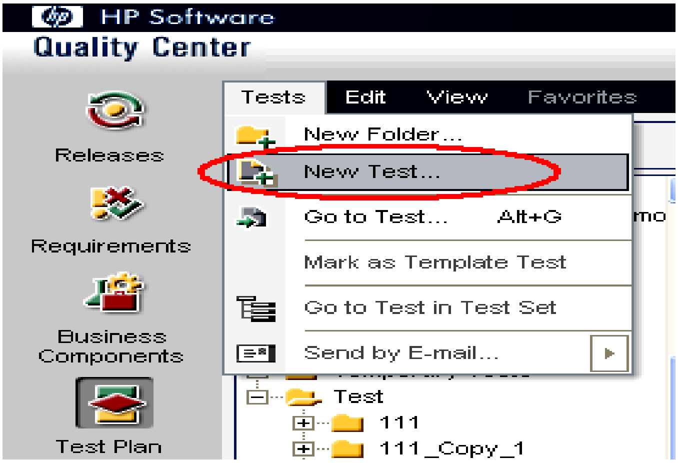


Fig. 12: Release Cycle in HP Quality Center

Adding a Test to the Subject Folder



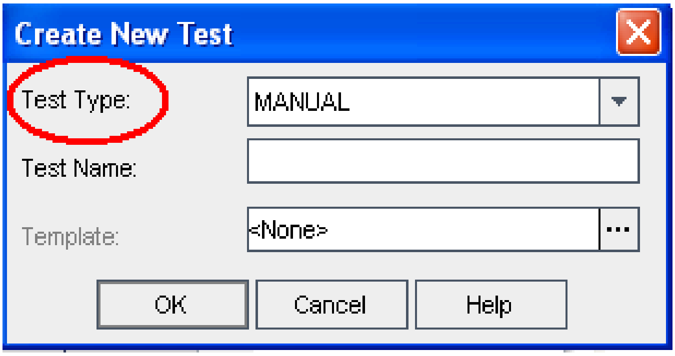


Fig. 13: Task Management in HP Quality Center

Assigning Defects

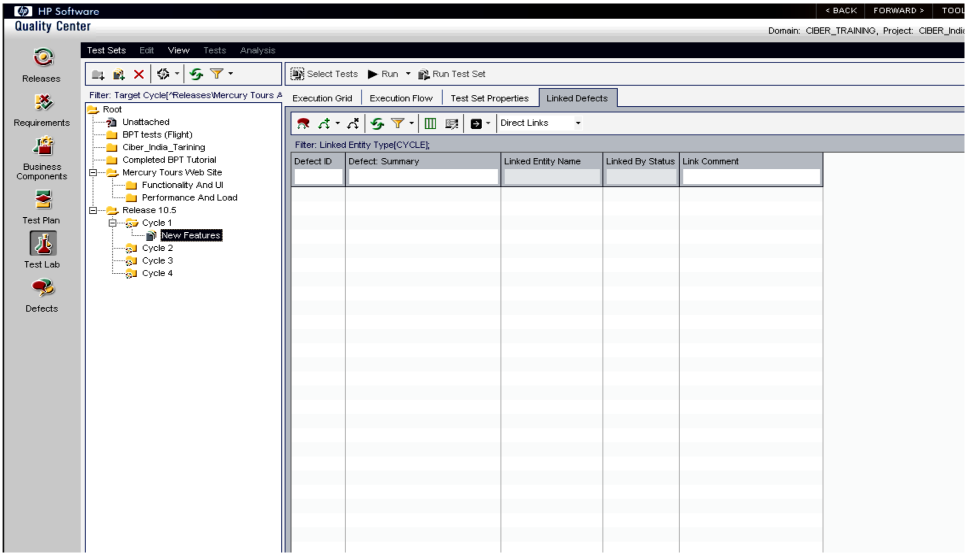


Fig. 14: Defect Management in HP Quality Center

Test Plan

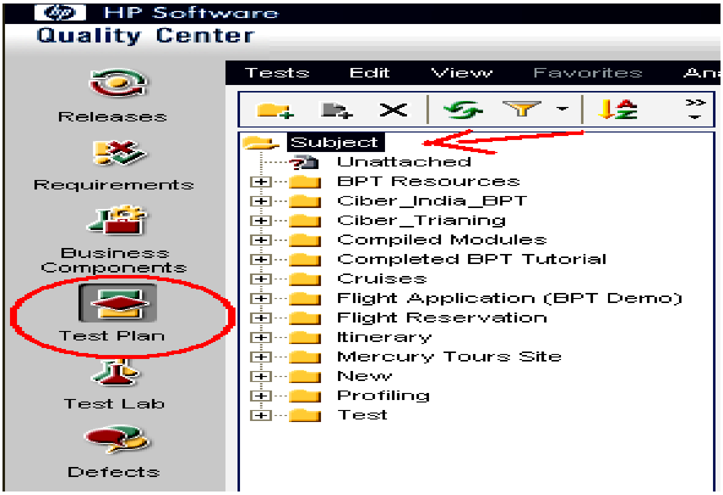


Fig. 15: Test Plan Management in HP Quality Center

4. TOOL COMPARISON

Table 2 below has shown the comparison of Spira, HP Quality Center and Mercury Quality Center on the basis of defined parameters used for test management:

Table 2
Overall summary of Spira, Mercury QC and HP Quality Center

S#	Parameters	Tools		
		Spira	**HP Quality Center**	**Mercury Quality Center**
1.	**Test policy/ Requirement management**	It has capability for creating, editing and managing project scope as well as requirements in a hierarchical organization like a typical scope matrix.	After defining the releases and cycles, the QA manager assigns the requirements from requirement module to releases and cycles.	It specifies requirements, enters detail of Requirements, and attaches Requirement File, naming the folder of Requirements and Checking the Status of requirements.
2.	**Test planning support**	Individual actions of a user must be taken in order to complete the test, which are represented in each step of test case.	In HP QC also the individual actions of a user must be taken to complete the test that is represented in each step of test case.	Requirements can be specified either by entering new data or by attachment of requirement document to MQC.
3.	**Bug Tracking**	All tasks are grouped in different categories like bugs, enhancements, issues, items of training, limitations, change of requests, and risks, and all these type have their own specific workflow and business rules.	If a bug is detected in an application while running a test set, the QA engineer can submit a defect. Quality Center automatically creates a link between the test run, the new defect as well as associated release and that particular cycle.	It raises the defects, searching defect based on conditions It changes the status of the defects and view views the status of the defects
4.	**Off Shore support**	**A powerful email integration capabilities is provided by Spira notify all the users regarding change in the system. Users may post comments directly from their email system.**	HP Quality Center notifies its users with email notification events that are being customized by project administrators.	Does not provide this feature of integrated email system.
5.	**Release Management**	All incidents which are detected, resolved and/or verified in a specific release are listed by drill down capability of Spira. This capability enables a quick	We might define the releases and cycles in the releases tree.	No support for this feature.

S#	Parameters	Tools		
		Spira	HP Quality Center	Mercury Quality Center
		determination regarding stability and readiness of a specific release.		
6.	**Reporting**	**An extensive library is provided by Spira that includes printable reports and graphical charts. These reports may be customized through XML based report templates. There are other supported report formats are also available like MS-Excel, MS-Word, HTML, MS-Project and Adobe Acrobat**	No support for this feature.	There is a sampling and defect reporting mechanism.
7.	**Project based test Management**	A web-based interface is provided by Spira that helps users in easily creation, edition, and deletion of a particular project. Moreover other option of creating a new project using the customized settings of an existing template project is also available.	No support for this feature.	Quality Center enables to manage multiple projects
8.	**User Management**	**An extensive administration also provided by Spira that can be exercised using just a web-browser. This capability play a vital role to reduce the burden on IT staff.**	No support for this feature.	Quality Center enables to manage user access to project. List of authorized users can be created in Quality Center. Quality Center administrator creates a password for the users to login. Each user in Quality Center is assigned to a user group.

CONCLUSION

In this paper, we analyzed few automated test management tools with different aspects. The parameters were pre-defined. One can select a test management tool based on the type of application need to be tested, budget, and the efficiency required. The findings explain different tools according to defined criteria. Now it is up to user who will determine the usefulness of tool as per his application in hand rather than actual features of tool. There are some other test management tools which are quite efficient e.g. Jira, Load Runner etc. in future we are planning to perform a comprehensive survey to compare all test management tools that are currently available in the market that help users to select tool according to requirements and needs.

REFERENCES

1. *Spira Test management tool inflacter corporation* [online] are Available at https://www.inflactar.com retrieved on 1st November, 2015
2. Safana, A.I. and Ibrahim, s. (2010). Implementing Software Test Management Using SpiraTeam Tool. Fifth International Conference on Software Engineering Advances (ICSEA), 447-452. DOI: 10.1109/ICSEA.2010.76
3. MuWramala, R., Pedagani, R. and Keskar, H. (2004). Test Data Management System For Aviation Software. *Digital Avionics Systems Conference*, Vol. 2, 10.A.1-10.1-8. DOI: 10.1109/DASC.2004.1390794.
4. Ducasse, S., Gîrba, T. and Wuyts, R. (2006). Object-Oriented Legacy System Trace-based Logic Testing. In Proceedings of *10th European Conference on Software Maintenance and Reengineering* (CSMR'06), 35-44, IEEE Computer Society Press.
5. Parveen, T., Tilley, S. and Gonzalez, G. (2007). A Case Study in Test Management. Proceedings of the *45th Annual Southeast Regional Conference*, Winston-Salem, 82-87.
6. Yumig, P. and Mingna, X. (2009). Load Testing for Web Application. The *1st International conferences on information Sciences and Engineering* (ICISE 2009).
7. http://kosmaczewski.net/mercury
8. http://www.scribd.com/doc/6756194/Test-Director-Overview
9. http://smartbear.com/products/qa-tools/test-management/software-testing-management/
10. http://www.testmanagement.com/qacomplete.html
11. http://smartbear.com/products/
12. http://www.testlodge.com

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 15-26

PREVALENCE OF DEPRESSION, ANXIETY AND ASSOCIATED RISK FACTORS AMONG HYPERTHYROID PATIENTS IN KARACHI, PAKISTAN

Farah Zahra, Sidra Sharif, Samra Tariq, Raiya Mansoor, Sana Soomro, Asra Batool, Abdul Azeem, Fazail Zia, Sohiba Ghazal, Nazain Zaheer and **Syed Muhammad Usama**
Jinnah Sindh Medical University, Karachi, Pakistan
Email: farah.zahra.hazara@gmail.com; sidrasharif55@yahoo.com; samra_tariq@hotmail.com; raiya.mansoor@gmail.com; sanasoomro92@hotmail.com; asraghaffar@gmail.com; sazeem214@gmail.com; fazailzia@gmail.com; sohibaghazal@gmail.com; nazainz@yahoo.com; musama657@gmail.com

ABSTRACT

Background: It is estimated that 5-10% of Pakistani population is suffering from thyroid disorders. Depression and anxiety are two very common symptoms of hyperthyroidism and patients with these symptoms are often misdiagnosed as having any psychiatric disease. The objective of our study was to determine the prevalence of depression and anxiety among hyperthyroid patients in our population of Karachi, Pakistan.

Methods: Cross-sectional study was conducted in which 200 patients visiting the OPD of Jinnah Postgraduate Medical Centre, Karachi, Pakistan, were inducted. Subjects were classified as hyperthyroid and euthyroid. Convenient sampling technique was used. Data collection tool was a questionnaire in which sociodemographic questions and Hamilton Depression Rating Scale and Hamilton Anxiety Rating Scale were included. Logistic regression was used to determine the association of depression and anxiety with sociodemographic factors among hyperthyroid patients. OR, 95% CI and P values were calculated.

Results: Depression and anxiety among hyperthyroid patients were found to be 84% and 58% respectively. In multivariate analysis for hyperthyroid patients, age group of 18-40 years and females were more than two times (OR=2.716, CI=0.886-8.332), (OR=2.587, CI=0.827-8.095) depressed, compared to age group of 41-60 years and males, respectively, after adjusting for covariates. Similarly, females were found to have anxiety more than one times (OR=1.771, CI=0.720-4.356) compared to males while age group of 18-40 years was 30.5% (OR=0.695, CI=0.286-1.689) less likely to have anxiety compared to age group of 41-60 years.

Conclusion: The results of our study revealed that depression and anxiety are a common finding among hyperthyroid patients. Therefore hyperthyroidism should always be considered in the differential diagnosis of patients who present primarily with neuropsychiatric symptoms.

KEYWORDS:

Hyperthyroidism, Depression, Anxiety, Neuropsychiatric symptoms.

1. INTRODUCTION

Thyroid disorders are very common affecting 750 million people worldwide by recent World health organization (WHO) estimates, being possibly even more prevalent than diabetes [1]. All thyroid disease occur more frequently in women than in men. The most common forms of hyperthyroidism include Grave's disease, toxic multinodular goiter and toxic adenoma. Grave's disease, the most common of these has a male to female ratio of 15 to 10. It is estimated that 5-10% of Pakistani population is suffering from thyroid disorders [2].

Hyperthyroidism is a set of disorders involving excess synthesis and secretion of the thyroid hormones i.e. FT3, FT4 by the thyroid gland. TSH produced by the pituitary gland maintains the levels of these hormones via negative feedback mechanism. The excess of these unbound hormones in the peripheral circulation can cause thyrotoxicosis. These increased level of the hormones, regardless of the etiology would result in increased transcription in cellular protein hence increased metabolic rate. The most common sign and symptoms of hyperthyroidism are nervousness, anxiety, increased perspiration, heat intolerance, weight loss despite an increased appetite, tremors and palpitations.

It is well documented in the literature that the features of hyperthyroidism may be similar to those observed in patients with psychiatric disease. The most frequently features reported in common are depression and anxiety [3]. Numerous studies have been done in this regard. One study conducted in Rhode Island reported that 123 patients out of 170 presented with anxious mood [4]. In another study on patients with recently diagnosed untreated hyperthyroidism, depression and anxiety was found in approximately 1/3rd of the patients, bringing to the mind that concurrent presence of somatic thyroid symptoms artificially inflates level of depression and anxiety. They also suggested that a psychiatrist should be careful to exclude patient with hyperthyroidism prior to primary psychiatric diagnosis [5]. One study in Turkey concluded that hyperthyroidism and syndromal depression and anxiety have overlapping features that can cause misdiagnosis during acute phase. For differential diagnosis, one should follow up patient with hyperthyroidism with specific hormonal treatment and evaluate persistent symptoms thereafter [6]. Rodewig stated that psychologic symptoms in the hyperthyroidism are similar to neurotic anxiety symptomatology and anxious depression syndrome [7].

Studies regarding the prevalence of hyperthyroidism have been conducted in Pakistan. One study comprising of 500 participants, 77.2% were normal, 71% were hypothyroid and 8.6% were diagnosed as hyperthyroid [8]. Another study determined the frequencies of thyroid problems in different age groups, sex and in different seasons. They reported that hyperthyroidism in all age groups was 5.1%. Prevalence of hyperthyroidism was higher in females (3.85%) than males (1.2%) [9].

There have been numerous studies in our region regarding thyroid disorders but none have specifically targeted the neuropsychiatric symptoms in hyperthyroidism particularly depression and anxiety. This study of ours would highlight the above mentioned symptoms in hyperthyroid patients and compare these with euthyroid individuals. This would give us an idea about how frequently depression and anxiety are associated with hyperthyroidism and would indicate the importance of suspecting hyperthyroidism in patients reporting to their GPs with these symptoms who are often wrongly diagnosed as having a psychiatric disease, so that delays in diagnosis can be avoided. Moreover, this would help reduce the burden of the disease as well as the treatment cost.

2. OBJECTIVE

To determine the prevalence of depression and anxiety and associated risk factors among hyperthyroid patients in the population of Karachi, Pakistan.

3. METHODS

Operational Definitions: Hamilton depression rating scale is a 17 item rating scale designed to measure severity and symptomatology of depression [10,11]. The scores are interpreted as: 0-7=normal, 8-13=mild depression, 14-18=moderate depression, 19-22=severe depression, 23->23 =very severe depression [12].

Hamilton anxiety rating scale is a 14 item rating scale designed to measure severity, symptoms and pattern of anxiety. Each item is scored on a scale of 0 (not present) to 4 (severe), with a total score range of 0–56, where 14-17 indicates mild severity, 18–24 mild to moderate severity and 25–30 moderate to severe [13].

To make our comparisons more pronounced we labeled our HAM-D scoring as 0-7, no depression and >7, depression and our HAM-A scoring as 0-13, no anxiety and >13, anxiety.

Study setting: Study was carried out at Thyroid Clinic of Jinnah Postgraduate Medical Centre and Department of Atomic Energy in Karachi, Pakistan from June to December, 2015.

Study design: This was a cross-sectional study.

Inclusion and exclusion criteria: Our study participants included individuals aged between 18-60 years. Subjects were classified according to their thyroid status. Previously diagnosed hyperthyroid patients having high level of FT3 (>7.8pmol/l) and FT4 (>25pmol/l) and suppressed level of TSH(<0.4mU/l) were labeled as hyperthyroid whereas those having normal value of FT3 (3.5-7.8pmol/l), FT4 (9.0-25.0pmol/l) and TSH (0.4-4.5 mU/l) were labeled euthyroid [14]. The patients with known psychiatric disorders were excluded. We also excluded patients with chronic diseases such as diabetes, tuberculosis etc. and uncooperative patients as well as those with language barriers.

Data collection tool: Subjects were interviewed after getting written consent. Instrument used was a validated questionnaire which comprised of three sections. Section A was related to sociodemographic characteristics, section B was the Hamilton Depression rating scale and section C was the Hamilton Anxiety rating scale to assess respective symptoms in both groups. The Hamilton Anxiety Rating Scale and the Hamilton Depression Rating Scale were both developed by Max Hamilton in 1959 and 1960 respectively. These are standard scales used by clinicians to assess and rate the severity of a patient's anxiety and depression respectively. The questionnaire was prepared in English and for the sake of consistency, was translated to Urdu.

Variables: The two dependent variables used were anxiety and depression whereas the independent variables were: age, gender, occupation, marital status, residence and education.

Sample Size: We used the formula, nz2p(100-p)/d2(n-1)+z2p(100-p) for calculating the sample size which came out to be 497 but due to time restraints we had to limit our sample size to 200 i.e. 100 hyperthyroid and 100 euthyroid individuals. We also reviewed previous studies in which their sample size was on average 95 hyperthyroid patients and an equal proportion of euthyroid individuals.

Sampling technique: Non-probability convenient sampling was used.

Data Management and Statistical Analysis: Epidata for windows version 3 was used for data entry. The data was cleaned and coded after which Statistical Package for Social Sciences (SPSS) for windows, version 16 was used for data analysis. Frequencies of the sociodemographic characteristics was calculated. Comparison was done between sociodemographic characteristics and anxiety and depression, both for hyperthyroid and euthyroid individuals. Chi-square was run and their significance was assessed. Univariate and Multivariate analysis was also done in which OR and CI was calculated. To exhibit the proportions of the HAM-D and HAM-A items, pie charts were made in Microsoft Excel. EndNote for windows was used for citations.

Ethical considerations: First of all, study protocol was approved by the parent institution. Permission letters were signed from the respective authorities concerned. The study participants were initially informed about the nature and objective of our study and how their participation would benefit the community. Thereafter they were made to sign a written consent form after being ensured about the confidentiality of the information they provide.

4. RESULTS

Mean age of participants was 33.9±11.883(standard deviation) .73.5% subjects were of age group 18-40 years, 68% were females, 70.5% were married, 44% were housewives and 76% were literate. 70% of hyperthyroid individuals took less than a year to seek treatment after they initially experienced symptoms and 70% were correctly diagnosed within an year of seeking treatment (Table 1).

Table 1
Sociodemographic Characteristics of Study Participants (n=200)

Sociodemographic Characteristics	Frequency (n)	Percentage (%)
AGE (Mean=33.9, Std deviation=11.883)		
18-40 years	147	73.5
41-60 years	53	26.5
OCCUPATION		
Housewife	88	44
Student	52	26
Others *	60	30
GENDER		
Male	64	32
Female	136	68
RELIGION		
Islam	191	95.5
Others**	9	4.5
MARITAL STATUS		
Married	141	70.5
Single	59	29.5
RESIDENT		
Karachi	184	92
Others***	16	8
EDUCATION		
Illiterate ^	48	24
Literate	152	76
DURATION BETWEEN FIRST SYMPTOM OF HYPERTHYROIDISM AND SEEKING ITS TREATMENT (n=100)		
Less than 1 year	70	70
More than 1 year	30	30
DURATION BETWEEN TREATMENT & DIAGNOSIS OF HYPERTHYROIDISM (n=100)		
Less than 1 year	79	79
More than 1 year	21	21

*drivers, fruit vendors, businessmen, **Christians, Hindus,
***any city other than Karachi, ^cannot read/write

Table 2: Comparison* between Sociodemographic Characteristics and Depression among Hyperthyroid and Euthyroid Groups

	Hyperthyroid (n=100)			Euthyroid(n=100)		
	No Depression**	Depression**	P Value***	No Depression**	Depression**	P Value***
AGE						
18-40 yrs	8	61	0.08	57	21	0.21
41-60 yrs	8	23		13	9	
GENDER						
Male	7	20	0.107	26	11	0.964
Female	9	64		44	19	
EDUCATION						
Illiterate	7	27	0.372	10	4	0.9
Literate	9	57		60	26	
DURATION BETWEEN FIRST SYMPTOM OF HYPERTHYROIDISM AND SEEKING ITS TREATMENT (n=100)						
Less than 1 year	11	59	0.458			
More than 1 year	5	25				

*Chi-square test, **Hamilton Depression Rating Scale; 0-7= no depression; >7= depression, ***<0.05 is significant

In hyperthyroid group , 84% subjects were depressed out of which 61% were of age group 18-40 years, 64% were females, 57% were literate and 59% took less than an year to seek treatment after initially experiencing symptoms. In euthyroid group, 30% subjects had depression, in which 21% were of age 18-40 years, 19% were females and 26% were literate(Table 2).

Table 3: Comparison* between Sociodemographic Characteristics and Anxiety among Hyperthyroid and Euthyroid Groups

	Hyperthyroid (n=100)			Euthyroid(n=100)		
	No Anxiety**	Anxiety**	P Value***	No Anxiety **	Anxiety**	P Value***
AGE						
18-40 yrs	31	38	0.377	70	8	0.872
41-60 yrs	11	20		20	2	
GENDER						
Male	14	13	0.227	33	4	0.836
Female	28	45		57	6	
EDUCATION						
Illiterate	13	21	0.584	14	0	0.999
Literate	29	37		76	10	
DURATION BETWEEN FIRST SYMPTOM OF HYPERTHYROIDISM AND SEEKING ITS TREATMENT (n=100)						
Less than 1 year	29	41	0.663			
More than 1 year	13	17				

*Chi-square test, **Hamilton Anxiety Rating Scale; 0-13=no anxiety;>13= anxiety, *** <0.05 is significant

In hyperthyroid group, 58% people had anxiety. Majority (38%)was of age group 18-40 years , 45% were females, 37% were literate and 41% took less than an year to seek treatment after initially experiencing symptoms. While in euthyroid group, 90% people had no anxiety, and only 10% of them were identified to have anxiety out of which 8% were of age group 18-40 years, 6% were females and all of them were literate (Table 3).

Table 4
The Relationship between Sociodemographic Characteristics and Depression: Univariate Analysis

	HYPERTHYROID		EUTHYROID	
	CRUDE OR	**95% CI**	**CRUDE OR**	**95% CI**
AGE				
18-40	2.652	0.891-7.896	0.532	0.198-1.427
41-60	1		1	
GENDER				
Female	2.489	0.822-7.537	1.021	0.420-2.478
Male	1		1	
EDUCATION				
Literate	1.642	0.553-4.877	1.083	0.311-3.772
Illiterate	1		1	

Table 5
The Relationship between Sociodemographic Characteristics and Anxiety: Univariate Analysis

	HYPERTHYROID		EUTHYROID	
	CRUDE OR	**95% CI**	**CRUDE OR**	**95% CI**
AGE				
18-40	0.674	0.281-1.618	1.143	0.225-5.817
41-60	1		1	
GENDER				
Female	1.731	0.711-4.215	0.868	0.228-3.303
Male	1		1	
EDUCATION				
Literate	0.790	0.339-1.839	0.777	0.339-4.612
Illiterate	1		1	

In univariate analysis for hyperthyroid individuals, those in age group of 18-40 years were more than two times (OR=2.652, CI=0.891-7.896) more likely to have depression than age group of 41-60 years and females were also more than two times (OR=2.489, CI=0.822-7.537) more likely to have depression than males (Table 4). Similarly, females were found to have anxiety more than one times (OR=1.731, CI=0.711-4.215) than males while age group of 18-40 years was 32.6% (OR=0.674, CI=0.281-1.618) less likely to have anxiety compared to age group of 41-60 years(Table 5).But all these figures were statistically insignificant. There was no significant association of sociodemographic variables with depression and anxiety in euthyroid group(Table 4 &5).

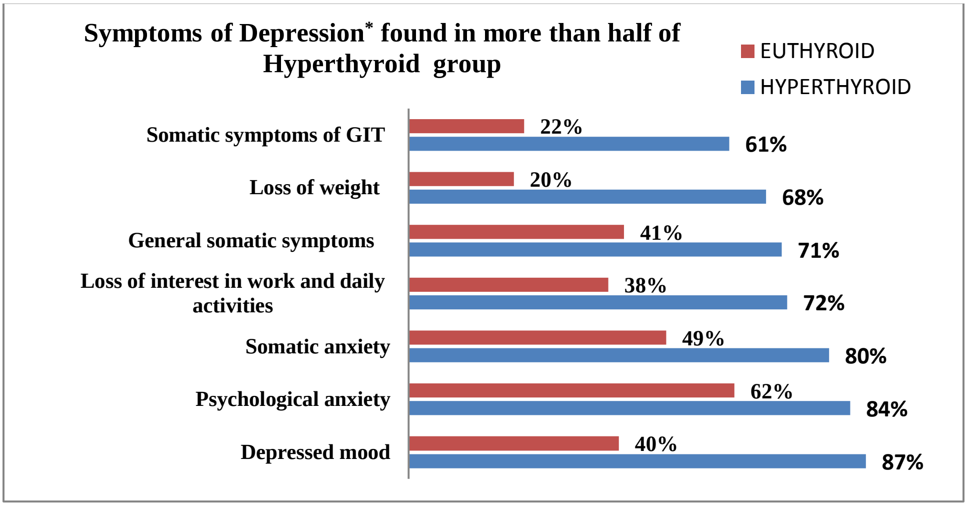
Table 6
The Relationship between Sociodemographic Characteristics and Depression: Multivariate Analysis

	HYPERTHYROID		EUTHYROID	
	ADJUSTED OR	95% CI	ADJUSTED OR	95% CI
AGE				
18-40	2.716	0.886-8.332	0.514	0.188-1.406
41-60	1		1	
GENDER				
Female	2.587	0.827-8.095	1.097	0.441-2.727
Male	1		1	
EDUCATION				
Literate	1.416	0.459-4.366	1.196	0.332-4.300
Illiterate	1		1	

Table 7
The Relationship between Sociodemographic Characteristics and Anxiety: Multivariate Analysis

	HYPERTHYROID		EUTHYROID	
	ADJUSTED OR	95% CI	ADJUSTED OR	95% CI
AGE				
18-40	0.695	0.286-1.689	1.003	0.192-5.247
41-60	1		1	
GENDER				
Female	1.771	0.720-4.356	0.780	0.201-3.020
Male	1		1	
EDUCATION				
Literate	0.778	0.327-1.848	0.788	0.314-3.688
Illiterate	1		1	

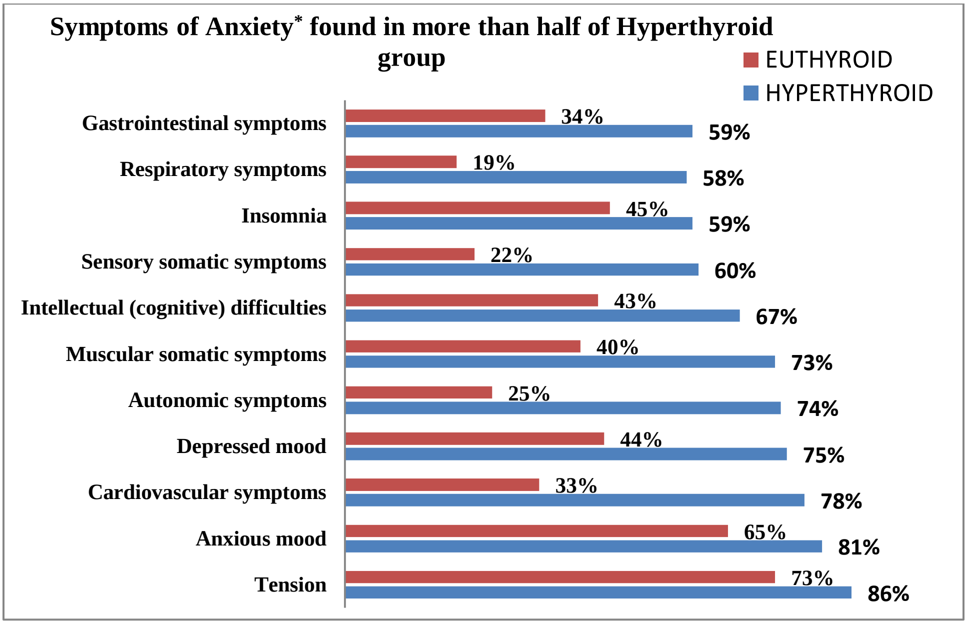
In multivariate analysis for hyperthyroid individuals, age group of 18-40 years was more than two times (OR=2.716, CI=0.886-8.332) more likely to have depression and females were also more than two times (OR=2.587, CI=0.827-8.095) more likely to be depressed after adjusting for covariates.(Table 6).Similarly, females were found to have anxiety more than one times (OR=1.771, CI=0.720-4.356) while age group of 18-40 years was 30.5% (OR=0.695, CI=0.286-1.689) less likely to have anxiety compared to age group of 41-60 years(Table 7).There was no significant association of sociodemographic variables with depression and anxiety in euthyroid group (Table 6 &7) on multivariate analysis as well.



*Hamilton depression rating scale

Figure 1: Comparison of Hamilton Depression Rating scale (HAM-D) items between Hyperthyroid and Euthyroid group

Regarding symptoms of depression, there was significant difference between hyperthyroid and euthyroid groups in terms of, loss of interest in work and activities, loss of weight and gastrointestinal symptoms. (Figure 1)



*Hamilton anxiety rating scale

Figure 2. Comparison of Hamilton Anxiety Rating Scale (HAM-A) items between Hyperthyroid and Euthyroid Group

Regarding symptoms of anxiety, there was significant difference between hyperthyroid and euthyroid groups in terms of, cardiovascular symptoms, autonomic symptoms, sensory somatic symptoms and respiratory symptoms.(Figure 2)

5. COMMENTS AND CONCLUSION

Discussion: Our study revealed that the symptoms of anxiety and depression were found in greater abundance among hyperthyroid patients when compared to euthyroid. Insomnia, autonomic symptoms, anxious mood and tension was also found in a significantly larger number of hyperthyroid participants.

Among hyperthyroid group in our study, patients between 18-40 years were more prone to have symptoms of depression (OR=2.65, CI=0.891-7.896). Mean age of study participants was 33.9, standard deviation: 11.883. Females were more likely to have depression when compared to their male counterparts (OR=2.587, CI=0.827-8.095). Similar statistics were found in another study according to which, 68% of the hyperthyroid patients among the group were females, mean age being 47 ± 14.8 [8]. Hyperthyroid patients who were literate were more likely to have depression and anxiety (OR=1.642, CI=0.553-4.877) than those who were illiterate.

In our study, 84% hyperthyroid had depression whereas 58% had anxiety in varying degrees of severity i.e. mild, moderate and severe. The figures were significantly low among euthyroid controls, 16% of which had no depression whereas, 90% of euthyroid individuals had no anxiety. In concordance with our study, anxiety disorder number of symptoms and depression disorder number of symptoms in another study was also higher among hyperthyroid patients when compared to euthyroid individuals [8]. This is because of the fact that high levels of serum T3 and T4 has a negative impact on a person’s psychological wellbeing. Serum T4 levels in the upper range of normal or slightly higher have been reported in depressed patients as compared to healthy or psychiatric controls [18].

Among symptoms of depression reported by most hyperthyroid patients in our study, included depressed mood (87%), loss of interest in daily work and activities (72%), loss of weight (68%) and somatic gastrointestinal symptoms (61%). It is also reported in certain studies that hyperthyroidism is an etiologic factor for depression and symptoms of anxiety [16,17]. In another study, it was reported that symptoms related to a high level of serum thyroid hormones is directly related to anxiety severity [15].

It was found in our study that 80% of the hyperthyroid patients suffered from symptoms of somatic anxiety, 84% of hyperthyroid patients had psychological anxiety and 87% complained of having depressed mood. A study conducted previously have reported that prevalence rate of major depression was 13-19%, generalized anxiety disorder was 46-61% and panic disorder was 30-61% among hyperthyroid patients [7,15]. A lot of these patients seeked medical help primarily because of the somatic and mood symptoms.

Many participants complained of having symptoms, like forgetfulness, difficulty concentrating in work and difficulty in carrying out everyday tasks. Frequent mood swings experienced by patients had a significant impact in their social life and marital life, with many complaining of having strained interpersonal relationships since their disease. These symptoms were so severe in certain patients that they had to quit their work.

Recommendation: It is necessary for medical practitioners to take into account the neuropsychiatric symptoms when treating a hyperthyroid patient. Appropriate medication for the management might be necessary depending upon the severity of the symptoms.

Strengths: To our knowledge, studies highlighting the importance of depression and anxiety in hyperthyroidism and comparing it with that of euthyroid individuals haven't been conducted in Pakistan, Karachi in particular. This study would serve as a source to emphasize the significance of anxiety and depression among the patients of hyperthyroidism in Pakistan.

Limitations: Since we are conducting this study as students and without any significant financial funding, our budget was limited due to which we could not conduct a large, multi-sector study. We had to limit the sample size due to our restricted time frame.

Conclusion: According to our study, neuropsychiatric symptoms such as, depression and anxiety were found in a significantly higher number of hyperthyroid patients as compared to euthyroid controls.

5. ACKNOWLEDGEMENT

We thank the staff and doctors of the Thyroid Clinics of Jinnah Postgraduate Medical Centre and Department of Atomic Energy who assisted us in conducting this research.

We are also thankful to our mentor, Dr. Mubashir and acknowledge him for supervising our project.

REFERENCES

1. Gharib, H. (2014). Emergent Management of Thyroid Disorders. *Endocrine and Metabolic Medical Emergencies, Endocrine Society*, 100-103.
2. Thyroiddiorders are common in Pakistan (2013). *The Nation*. Retrieved from www.thenation.com
3. Hendrick, V.C. and Garrick, T. (2000). Endocrine and metabolic disorders. *Comprehensive textbook of psychiatry*. 7th ed. Philadelphia, PA, USA, 1808-1818.
4. Stern, R.A., Robinson, B., Thorner, A.R., Arruda, J.E., Prohaska, M.L. and Prange, A.J. (1996). Neuropsychiatric Complaints in Graves' patients. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 8, 181-185.
5. Kathol, R.G. and Delahunt. J.W. (1986). The relationship of anxiety and depression to symptoms of hyperthyroidism using operational criteria. *Gen. Hosp. Psychiatry*, 8, 23-28.
6. Demet, M.M., Özmen, B., Deveci, A., Boyvada, S., Adıgüzel, H. and Aydemir, Ö. (2002). Depression and Anxiety in Hyperthyroidism. *Archives of Medical Research,* 33, 552-556.

7. Rodewig, K. (1993). Psychosomatic aspects of hyperthyroidism with special reference to Basedow' sdisease. *Psychother Med. Psychol.*, 43, 271-277.
8. Anwarullah, A.R., Farooq, N., Ansari, M.A. and Ashraf, S. (2012). Prevalence of hypo/hyperthyroidism cases in suspected population of Rawalpindi/Islamabad. *Pakistan. Journal of Public Health and Biological Sciences,* 1, 96.
9. Akhtar, S., Khan, A. Siddiqui, M.M. and Gul Nawab (2001). Frequencies of thyroid problems in different age, sex and seasons. *The Sciences,* 1, 153-156.
10. Aydemi, O. (2000). Instruments for psychiatric assessment: their properties and use. *Clinical Scales in Psychiatry*. Turkey.
11. Hamilton M. (1960). A rating scale for depression. *J. Neurol Neurosurg Psychiatry,* 20, 56-62.
12. Hamilton Depression Rating Scale (2010). Retrieved from http://www.serene.me.uk/.
13. Hamilton, M. (1959). The assessment of anxiety states by rating. *Br. J. Med. Psycho.*, 32, 50-55.
14. Thyroid Function Tests (2014). *British Thyroid Foundation*. Retrieved from http://www.btf-thyroid.org/information/quick-guides/97-thyroid-function-tests.
15. Trzepacz, P.T., Klein, I., Roberts, M., Greenhouse, J. and Levey, G.S. (1989). Graves' disease: an analysis of thyroid hormone levels and hyperthyroid signs and symptoms. *Am. J. Med.*, 87, 558-561.
16. Beyer, J., Burke, M., Meglin, D., Fuller A., Krishnan, K.R.R. and Nemeroff, C.B. (1993). Organic anxiety disorder iatrogenic hyperthyroidism. *Psychosomatics,* 34, 181-184.
17. Fardella C.E., Gloger, S., Figueroa, R., Gajardo, C., Salgado, C., Barroilhet, S. and Foradori, A. (2000). High prevalence of thyroid abnormalities in a Chilean psychiatric population. *J Endocrinol Invest.*, 23, 102-106.
18. Chopra, I.J. Solomon, D.H. and Huang, T.S. (2004). Serum thyrotropin in hospitalized psychiatric patients; evidence for hyperthyrotropinemia as measured by an ultra sensitive thyrotropinassay. *Metabolism*, 39, 41-49.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 27-32

A PRIMORDIAL APPROACH TO STATISTICAL RESEARCH

M. Inam Khokhar and **Zahida Habib**
[1] 5041 Silver Saddle Ct., Dublin, Ohio 43016, USA
Email: drinamus@gmail.com
[2] University of Education, Lahore, Pakistan
Email: zahidatariq_77@yahoo.com

ABSTRACT

A Primordial approach to statistical research may provide some clues and some challenges to statisticians and scientists. In the current paper an attempt is made to illustrate this view with specific examples in the context of the Quran and authentic Hadiths. The examples are provided in the domain of randomness, fetal development and instinctive evolution. The objective of writing this paper is to inspire the statistical researchers to pay due attention to this approach.

INTRODUCTION

There are only two sources of knowledge. The first source is the acquired knowledge by the use of five senses. The acquired knowledge is processed for logical and experimental investigation to formulate the scientific theories and laws. However all this knowledge is evolutionary and gradually advances towards the ultimate truth. The second source is the Divine knowledge that has been revealed by Allah (swt) to His chosen prophets and the last book of this Divine knowledge was revealed in the form the Quran. The knowledge revealed in the Quran is of absolute nature and consists of ultimate truths of moral and scientific realities. The Quran not only narrates many scientific truths of absolute nature, but it also inspires the intellectuals to explore and ponder over those verses. Some of the verses may provide some clues and challenges to scientific and statistical researchers. The objective of writing this paper is to inspire the researchers to pay due attention in this direction. This approach may be illustrated by the following examples.

1. The Primordial Randomness

There are many verses in the Quran that provide vital information about certain scientific facts. These verses appear to be randomly distributed in various chapters of the Quran. However despite their apparently random distribution their placement positions are of absolute nature because it is unanimously believed by the Islamic scholars that all the verses and chapters of the Quran are arranged and placed according to the will of Allah (swt). So these verses are placed at specific positions with planned purpose and wisdom of Allah (swt). Thus their placement distribution should not be considered as random by chance but as absolute randomness of primordial nature.

This Quranic concept of absolute randomness provides confirmation to the views of absolute randomness proposed by some statisticians for certain situations[1]. Such statisticians argue that if a randomness is persistent and of ‘irreducible’ nature then it should be treated as of absolute nature having a certain purpose. In this regard the random nature of quantum physics (which comprises the current and very successful understanding of the microscopic physical world) has been quoted by the eminent late physicist John A. Wheeler who named this randomness as ‘**Primordial or Irreducible’**. Thus *Primordial* should not be an unfamiliar term to statisticians.

The Primordial distribution of the Quranic verses indicates the following points:

1. Allah (swt) has created this universe and everything in it with certain planning and purpose of creating order and beauty. So with that objective it is perfectly justified to arrange few things randomly when it is required to serve the purpose. Sometimes things may appear to be random but the apparent randomness could be working with some purpose and certain rules set by the Creator. In statistical terminology they are of ‘Primordial’ nature.

2. A statistician should not accept any randomness to be meaningless just based upon apparent behavior of the things but should investigate their situation to explore if there is any purpose behind that randomness. This would be a kind of challenge to a statistical researcher.

Another Example of Primordial Randomness

The big bang explosion is often quoted as an explosion out of extremely dense and hot source. This explosion created randomly moving particles with extremely high speed in all directions. However, this random movement of the particles was primordial and of absolute nature because out of this randomness, galaxies, solar systems and planets including our earth were evolved.

2. SEQUENCE IN EMBRYO/FETAL FORMATION OF EARS, EYES AND THE HEART

The following Quranic verses indicate that in the embryo/fetal development Allah (swt) created first ears, then eyes and then heart as shown by their sequence in these verses. This sequence is persistent in all the quoted verses without any exception.

قُلۡ هُوَ ٱلَّذِىٓ أَنشَأَكُمۡ وَجَعَلَ لَكُمُ ٱلسَّمۡعَ وَٱلۡأَبۡصَـٰرَ وَٱلۡأَفۡـِٔدَةَ‌ۖ قَلِيلاً مَّا تَشۡكُرُونَ
(٢٣)(الملک)

Say it is He Who has created you, and endowed you with hearing (ears), seeing (eyes), and hearts. Little thanks you give. (Al-Mulk-67:23)

وَهُوَ ٱلَّذِىٓ أَنشَأَ لَكُمُ ٱلسَّمۡعَ وَٱلۡأَبۡصَـٰرَ وَٱلۡأَفۡـِٔدَةَ‌ۚ قَلِيلاً مَّا تَشۡكُرُونَ (٧٨)(المؤمنون)

It is He, Who has created for you (the sense of) hearing (ears), eyes (sight), and hearts (understanding). Little thanks you give. (Al-Moumenun-23:78)

ٱلَّذِىٓ أَحۡسَنَ كُلَّ شَىۡءٍ خَلَقَهُۥ‌ۖ وَبَدَأَ خَلۡقَ ٱلۡإِنسَـٰنِ مِن طِينٍ (٧) ثُمَّ جَعَلَ نَسۡلَهُۥ
مِن سُلَـٰلَةٍ مِّن مَّآءٍ مَّهِينٍ (٨) ثُمَّ سَوَّٮٰهُ وَنَفَخَ فِيهِ مِن رُّوحِهِۦ‌ۖ وَجَعَلَ لَكُمُ ٱلسَّمۡعَ
وَٱلۡأَبۡصَـٰرَ وَٱلۡأَفۡـِٔدَةَ‌ۚ قَلِيلاً مَّا تَشۡكُرُونَ (٩)(السجده)

Who made everything He has created good, and He began the creation of man from clay. (7) Then He made his offspring from semen of dispised water (male and female sexual discharge). (8) Then He fashioned him in due proportion, and breathed into him the soul (created by Allâh for that person), and He gave you hearing (ears), sight (eyes) and hearts. Little is the thanks you give!(9) (As-Sajda-32)

وَجَعَلۡنَا لَهُمۡ سَمۡعً۬ا وَأَبۡصَـٰرً۬ا وَأَفۡـِٔدَةً۬ فَمَآ أَغۡنَىٰ عَنۡہُمۡ سَمۡعُهُمۡ وَلَآ أَبۡصَـٰرُهُمۡ
وَلَآ أَفۡـِٔدَتُہُم مِّن شَىۡءٍ إِذۡ كَانُواْ يَجۡحَدُونَ بِـَٔايَـٰتِ ٱللَّهِ وَحَاقَ
بِہِم مَّا كَانُواْ بِهِۦ يَسۡتَہۡزِءُونَ (٢٦)

We made for them hearing and vision and hearts. But their hearing and vision and hearts availed them not from anything [of the punishment] when they were [continually] rejecting the signs of Allah ; and they were enveloped by what they used to ridicule. (46:26)

وَٱللَّهُ أَخۡرَجَكُم مِّنۢ بُطُونِ أُمَّهَـٰتِكُمۡ لَا تَعۡلَمُونَ شَيۡـًٔ۬ا وَجَعَلَ لَكُمُ ٱلسَّمۡعَ
وَٱلۡأَبۡصَـٰرَ وَٱلۡأَفۡـِٔدَةَ‌ۙ لَعَلَّكُمۡ تَشۡكُرُونَ (٧٨)(النحل)

And Allâh has brought you out from the wombs of your mothers while you know nothing. And He gave you hearing, sight, and hearts that you might give thanks (to Allâh). (An-Nahal-16:78)

The expression of these organs in the last verse (16:78) quoted above shows that their formation is being referred to in the womb of a pregnant mother (during embryo/fetal development).

These verses, first of all tell that these three organs are of utmost importance for the human personality. The heart is important because its failure means death. A person can survive without brain but not without his heart. The ears and eyes are also very crucial for fetal development in mother's womb.

A recent report on fetal development shows that the heart is formed at a much later stage of pregnancy. This may be in line with the absolute truth expressed in the above verses.

The ear embryology investigations show that the ears begin to appear during the 22nd day of embryo's development. And the heart begins to beat during the sixth weeks (after the last menstrual period) of pregnancy. Anyway there is challenge in these verses to do further investigation to elaborate the reality expressed in these verses of the Quran.

3. PRIMORDIAL INSTINCT AND EVOLUTION

Instinct is generally defined as an innate (inherent), typically fixed inborn pattern of behavior in animals and humans in response to certain stimuli. There is no learning involved in instinct; it happens naturally, without even thinking about it. Babies cry for pain or hunger, suck for food, try to grasp anything that comes close to their hands, and show pleasure with their facial expressions, all these by instinct. Among non-humans the common examples of instinctive actions include the spinning of web by a spider or building nests by birds. A chick at its birth breaks the shell of the egg mere by *instinct*. Since the chick does not learn from any external source, his act of breaking the shell is purely instinctive.

The point to be noted here is that the vital instinctive knowledge possessed by the species in the above quoted examples that inspires them to perform their crucial actions is of the category of absolute truth of primordial nature.

Different views and theories have been proposed by psychologists and biologists about instinct and instinctive behavior. The most popular seems to be the **Instinct Theory of Motivation that was first proposed by William McDougall (1908) and then elaborated by Sigmund Freud (1927). This theory states that** all activities, thoughts, and desires can be drawn from being caused by nature. Human beings do things because they are evolutionarily programmed to do so. Individuals have a programmed disposition at birth and genes are identified to motivate people.

However the Theistic concept would be that Allah (swt) created every entity with all of the essential commandments imbedded in it. These commandments inspire the entity how to behave, how and in which direction to evolve and with what purpose. Since these inspiring commandments are granted to the entities by Allah (swt) they qualify to be the Divine Primordial Instinct although we are not yet capable to identify their instinctive behavior. This Divine Primordial Instinct is programmed to evolve along with the overall evolution of life till it evolves to a level that we are able to identify it as an instinct with the available investigation tools. **The validity of this argument leads to the fact that the instinctive behavior is based upon absolute true Divine knowledge granted to the entities at their Primordial stage. Thus in their instinctive actions at an evolved stage there would be no chance at all for any mistake to occur because this knowledge being Divine is Absolute in nature and beyond the scope of statistical probability.**

Another important fact that can be deduced from the above discussion is that if we interpolate the instinctive behavior of newly born specie back to its Primordial stage, it becomes quite clear that in the course of Theistic Evolution of the Primordial instinct to the level of newly born baby-instinct there is no chance that the instinct of a certain species would change during its course of evolution to the instinct of another species. Thus it becomes quite evident that in the Grand Creation Process of Allah (swt) all species were created independently and they evolved independently without ever changing over to any other species. The evolution of independently created species always remained independent irrespective of time and space of their evolution.

Henri Bergson described in his book '*Creative Mind*'[2]: "The intellect is characterized by a natural inability to comprehend life. Instinct, on the contrary, is molded on the very form of life. While intelligence treats everything mechanically, instinct proceeds, so to speak, organically. If the consciousness that slumbers in it should awake, if it were wound up into knowledge instead of being wound off into action, if we could ask and it could reply, it would give up to us the most intimate secrets of life. For it only carries out further the work by which life organizes matter–so that we cannot say, as has often been shown, where organization ends and where instinct begins.

When a little chick is breaking its shell with pecks of its beak, it is acting by instinct, and yet it is just carrying on the movement which has borne it through embryonic life. Inversely, in the course of embryonic life itself (especially when the embryo lives freely in the form of a larva), many of the acts accomplished must be referred to instinct. The most essential of the primary instincts are really, therefore, vital processes."

4. FINDINGS OF DR. JOE L. SIMPSON

Recently Professor Dr. Joe Leigh Simpson[3] has pointed out how the Quranic revelations on the scientific facts can help the scientific researchers. He showed that the precision of time that the Quran has mentioned in certain cases brings another dimension in scientific research. This is valid at least in biological sciences. This provides a good clue to the scientific researchers and statisticians. Accepting that clue as a challenge the Statisticians can start their investigation in microbiology but should extend it in the other fields of scientific research too. In this regard a good knowledge of Quran and Hadith regarding the creation process would be essential.

Professor Simpson studied the following two Hadiths (sayings) of the Prophet Muhammad (saw):

> {*In every one of you, all components of your creation are collected together in your mother's womb by forty days*…}[4]
> {*If forty two nights have passed over the embryo, God sends an angel to it, who shapes it and creates its hearing, vision, skin, flesh, and bones*....}[5]

He studied these Hadiths extensively, noting that the first forty days constitute a clearly distinguishable stage of embryo-genesis. He was particularly impressed by the absolute precision and accuracy of those sayings of the Prophet Muhammad (saw). Then, during one conference, he gave the following opinion:

> "So that the two *Hadiths* that have been noted provide us with a specific time table for the main embryological development before forty days. Again, the point has been made, I think, repeatedly by other speakers this morning: these *hadiths* could not have been obtained on the basis of the scientific knowledge that was available [at] the time of their writing. It follows, I think, that not only there is no conflict between genetics and religion but, in fact, **religion can guide science by adding revelation to some of the traditional scientific approaches**, that there exist statements in the Quran shown centuries later to be valid, which support knowledge in the Quran having been derived from God."

The findings of Professor Simpson confirm the importance of Quranic study for both Muslim and non-Muslim researchers. His concluding statement (**religion can guide science by adding revelation to some of the traditional scientific approaches)** provides a clue to Islamic scholars as well as to the scientific researchers that probably the most significant purpose of placing the scientific verses in the Quran was to provide guidance to the scientific researchers of all ages to come. If the researchers adopt that approach (which may be called the **Primordial Approach)**, they will find an additional dimension that will help them to see their investigations with broader vision. With that process the realities of the nature will keep on unfolding according to the following verse of the Quran:

سَنُرِيهِمۡ ءَايَـٰتِنَا فِى ٱلۡأَفَاقِ وَفِىٓ أَنفُسِهِمۡ حَتَّىٰ يَتَبَيَّنَ لَهُمۡ أَنَّهُ ٱلۡحَقُّۗ
أَوَلَمۡ يَكۡفِ بِرَبِّكَ أَنَّهُۥ عَلَىٰ كُلِّ شَىۡءٍ شَهِيدٌ (فُصِّلَت۔٥٣)

We will show them our signs in the horizons and within themselves until it becomes clear to them that it is the truth. But is it not sufficient concerning your Lord that He is, over all things, a Witness? (Fussilat-41:53)

REFERENCES

1. *General Relativity and John Archibald Wheeler*, Editors: Ignazio Ciufolini, Richard A. Matzner, Springer Dordrecht, Heidelberg, London, New York, 2010. P.190.
2. Henri Bergson's *The Creative Mind: An Introduction to Metaphysics* (1946), Dover Books, May 2010.
3. *A Brief Illustrated Guide to Understanding Islam*, Second Edition, published by Darussalam, Houston, Texas, USA (1997) p-28
4. Narrated in *Saheeh Muslim*, #2643, and *Saheeh Al-Bukhari*, #3208. A *hadeeth* is a reliably transmitted report by the Prophet Muhammad's (saw) companions of what he said, did, or approved of.
5. Narrated in *Saheeh Muslim*, #2645.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 33-40

THE FRACTAL DIMENSION DESCRIPTION OF IONOSPHERE VIRTUAL HEIGHT OF F2 LAYER AT PAKISTAN ATMOSPHERIC REGION

Bulbul Jan[1] and **M. Ayub Khan Yusuf Zai**[2§]

[1] Institutes of Space and Planetary Astrophysics, University of Karachi Karachi, Pakistan. Email: bulbul.gilgit@gmail.com

[2] Department of Applied Physics and Solar-Terrestrial and Atmospheric Research Wing and Institute of Space and Planetary Astrophysics, University of Karachi, Karachi, Pakistan. Email: ayubzai@yahoo.com

[§] Corresponding author

ABSTRACT

The virtual height is an important variable of Ionospheric F2 layer. The estimation of virtual height is very important because it provides information about the frequencies at which variations occur, as well as the time where they occur.

This study discusses the estimation of virtual height using fractal dimension and Rescaled range analysis which estimated Hurst exponent to observe the complex nature of the data. It is known that the Hurst exponent associated with fractal dimension. Fractal dimension shows complex nature of the data while Hurst exponent represents smoothness or roughness of the data.

Our results concluded that the time series data of virtual height over Pakistan region shows anti-persistency. This means that a decreasing time series (negative correlation). It is more probable that an anti-persistent time series will exhibit higher noise and more instability.

This study will encourage the investigators working in the same field to verify the complex nature of Ionospheric signals.

KEY WORDS

Virtual Height, Hurst Exponent, Fractal Dimension, Rescaled Analysis, Ionosphere.

1. INTRODUCTION

The ionosphere is a very important to understand in any space weather study [1]. The ionosphere is a region of weakly ionized plasma which ranges from about 60 km to beyond 1000 km altitude within the Earth's atmosphere [2]. This region is important to sky-wave radio propagation and provides the basis for almost all HF communications [12]. The ionosphere is also essential in optimizing satellite communication systems since the satellite signals traverse the ionosphere, leading to attenuation, depolarization, refraction and dispersion as a result of scattering and frequency dependent group delay. When HF radio wave reaches the ionosphere, it can be refracted such that it is radiated back to the Earth at some horizontal distance beyond the horizon. This effect is due to refraction but it is often apparently considered to be a reflection [3].

Observations of the virtual height of the ionosphere and its variations carried out by Kirby et al. 1934. The virtual height of an ionosphere F2 layer could be explained as the wave is refracted bent down gradually rather sharply. Incident and refracted rays below the ionized layer follow the same path as if the refection has taken place from the greater height. The mark is called virtual height of this layer, and in other words when the incidence and returned rays are extrapolated to a vertex they meet at a height h' is called virtual height. The F layer has a stable existence; even from side to side the height varies on a daily basis. This layer is the most important for radio communications in the frequency range of 3 to 40 MHz [4].

Box-counting or box dimension is one of the most widely used dimensions. Its attractiveness is largely due to its relative simplicity of mathematical calculation and empirical estimation. The definition goes back at least to the 1930s and it has been variously termed Kolmogorov entropy, entropy dimension capacity dimension (a term best avoided in view of potential theoretic associations), metric dimension, logarithmic density and information dimension. We shall always refer to box or box-counting dimension to avoid confusion. This kind of analysis includes more insight into complexity and structure of the system [5].

Fractal dimensions are generally comparison of different numbers that are associated with fractals. The importance of Fractal dimensions is because they can be defined in connection with real world data which can be measured approximately by experiments.

Fractals are mathematical sets with geometrical shape that can model many natural phenomena. Almost all natural objects can be observed as fractals such as coastlines, trees, mountains, and clouds, etc. The most important feature of fractals is self-similarity and fractal dimension [6, 13].

The clouds, trees, feathers, coastlines, neurons networks in the body, dust in the air, the clothes, and the distribution of frequencies, the colors emitted by the sun, and the wrinkled surface of the sea during a storm are attached with the fractal dimensions system [5].

2. DATA AND METHODOLOGY

The time series data of virtual height of Ionospheeric region F2 layer of total 151 days from October-1986 to Februry-1987 and October-2015, to February-2016 of Pakistan air space region taken from Space and Upper Atmosphere Research Commission (SUPARCO) which updates daily space weather data.

In this communication we applied Fractal Dimension method to observe the complexity of the virtual height of Ionosphere F2 layer data. Higher the complexity of the data lower will be the smoothness or wise versa. Fractal dimension express the space filling ability of the data whereas, Hurst exponent represents the smoothness or roughness of the data [7].

The FD of self similar object with self similar pieces ‘r’ scaled down by a factor ‘s’ can be express mathematically expressed as

$$D = \frac{\ln r}{lns}$$

The fractal dimension (FD) and Hurst exponent associated as

$$D = 2\text{-}H$$

where D is fractal dimension, H represent Hurst exponent and 2 is Euclidean dimension. FD and H represent the dynamical behavior of the time series data [5].

The Hourst exponent, H can be estimated by rescaled range analysis

$$H = \frac{log(R/S)}{log(T)}$$

where, T is the duration of the sample data and R / S the corresponding value of the rescaled range. Then the Hurst exponent is estimated by plotting the values of $\log(R / S)$ versus $\log(N)$. The slope of the best fitting line gives the estimate of the Hurst exponent. This is done through an ordinary least squares regression procedure. We can generate 'least squares' regression line using linear regression method [6, 11].

Particularly, Hurst exponent compares the persistency, and anti persistency and Brownian nature of the time series data [5]. The Hurst exponent is limited to the range $0 \leq H \leq 1$. When the value of Hurst exponent is $H > 0.5$ we have a persistent behavior (positive correlation), if the value is $H < 0.5$ then the process is considered ant persistent (negative correlation) and if H=0.5 it means that the process is considered to be "white noise" (Brownian motion). In other words, when the value of H moves towards zero, it means that the roughness of the time series grows [8].

3. RESULTS AND DISCUSSION

Fractal dimension in one dimensional space are important for interpreting time series data of complex nature [9]. Fractal dimension is zero for a point, one for a line and two for a surface and so on [10]. The Analysis of data virtual height (h′) consists of total 151 days from October-1986 to February-1987 and October-2015, to February-2016.

The time series shown in Fig. 1 and Fig. 2 were taken for purpose of estimating the Hurst Exponent.

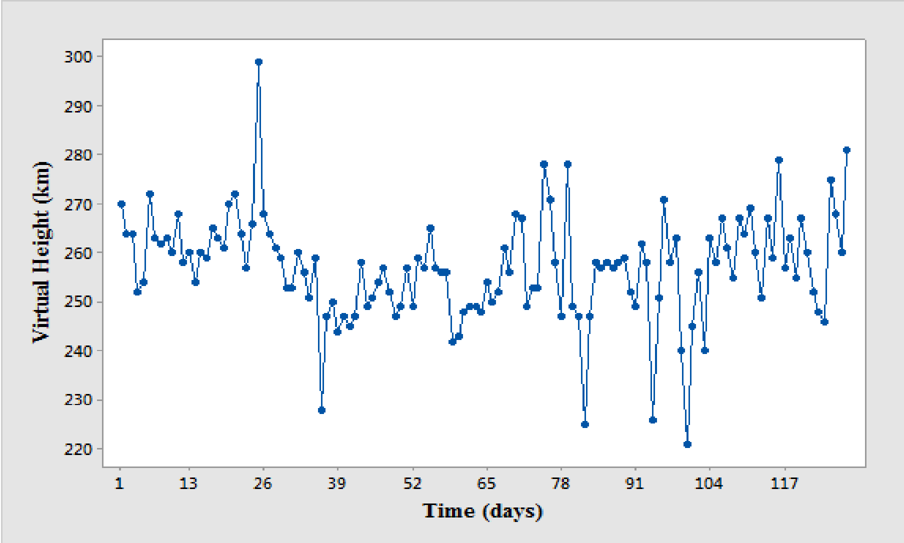


Fig. 1: Time series plot collected over 128 data values of Virtual Height of F2 layers for year 1986-87

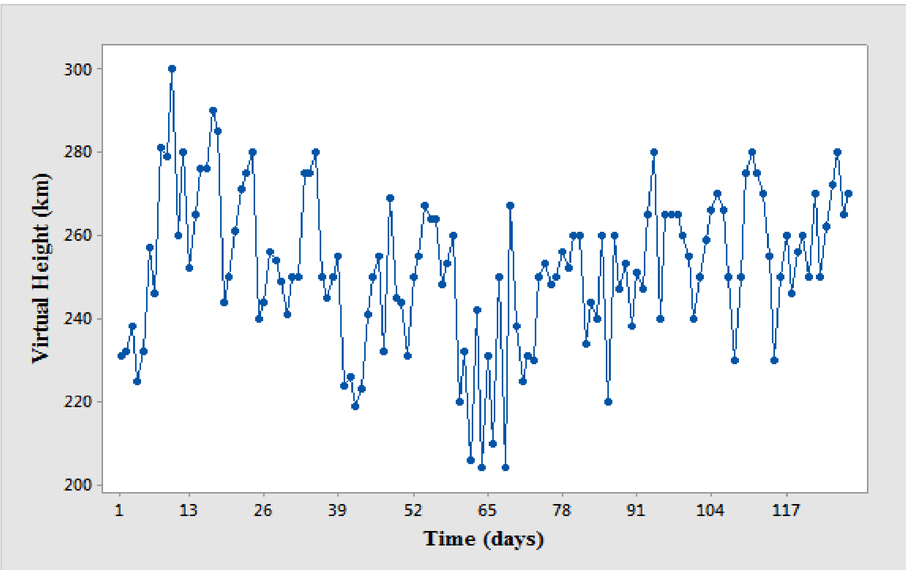


Fig. 2: Time series plot collected over 128 data values of Virtual Height of F2 layer for year 2015-16

The data were collected over 128 time intervals for virtual height over Pakistan region and the rescaled range analysis gave the in Table 1 and Table 2 shown below:

Table 1
Rescaled Range Analysis values Virtual Height of F2 Layer Over Pakistan Region to Estimate H for 1986-87

Segment size (N)	$log_2(N)$	$log_2(R/S)$
128	7	2.8511
65	6	2.8496
32	5	2.5528
16	4	1.8369

Table 2
Rescaled Range Analysis Values Virtual Height of F2 Layer Over Pakistan Region to Estimate H for 2015-16

Segment size (N)	$log_2(N)$	$log_2(R/S)$
128	7	2.3841
65	6	2.2345
32	5	1.9508
16	4	1.7299

Fig. 3 and Fig. 4 projected to estimate the slope of the log-log plot of R/S verses N gives the estimate of the Hurst exponent. The output gave the estimate of H for time series data for 1986-87 to be **0.3341** and for 2015-16 to be **0.2246**.

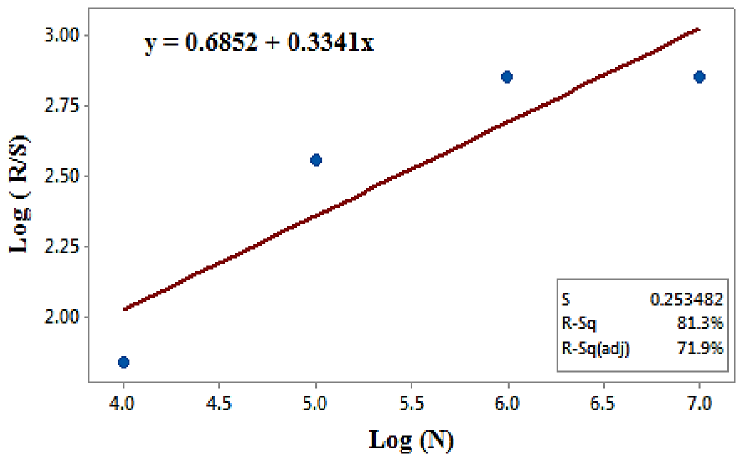


Fig. 3: Regression Line Fir Virtual Height of F2 Layer for Pakistan Region which Estimates the Hurst Exponent for the Year 1986-87

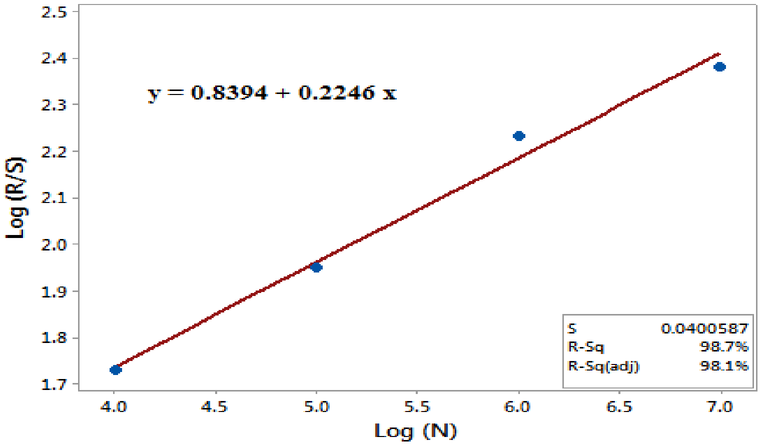


Fig. 4: Regression Line Fir Virtual Height of F2 Layer for Pakistan Region which Estimates the Hurst Exponent for the Year 2015-16

The Table 3 shown the results of FD and for time series data using rescaled analysis and we find H=0.3, D=1.7 (1986-87) and H=0.8, D=1.8 (2015-16). These results revealed that the nature of data is anti-persistent with negative correlation for both time intervals.

Table 3
FD and H of Virtual Height Data of F2 Layer for Pakistan Region over the Period 1986-87 and 2015-16

Years	D	H	Nature	Correlation
1986-87	1.7	0.3	Anti-persistent	Negative
2015-16	1.8	0.2	Anti-persistent	Negative

4. CONCLUSION AND COMMENTS

In this work, we have conducted fractal dimension to observed complex nature of Virtual Height of F2 layer time series data at Pakistan region and Rescaled analysis performed on each of the time series estimated the value of Hurst exponent.

- Fractal and Hurst exponent can detect changes in the roughness and negative correlation of the time series like on October (1986) to February (1987) and October (2015) to February (2016).
- As the value of Fractal dimension is $D > 1.5$ and $H < 0.5$ implies that we have anti-persistent behavior. This means that a decreasing time series (negative correlation). It is more probable that an anti-persistent time series will exhibit higher noise and more instability.
- Further work will involve the study of distribution properties (normality) of Hurst exponent for random walk model for time series.

5. ACKNOWLEDGMENT

We are thankful to the technical staff of Space and Upper Atmosphere Research Commission (SUPARCO) for updating daily Ionospheric data like critical frequency, maximum usable frequency (MUF) and we also thankful to organizing committee of ISOSS.

REFERENCES

1. López-Montes, R., Pérez-Enríquez, R. and Araujo-Pradere, E.A. (2012). The impact of large solar events on the total electron content of the ionosphere at mid latitudes. Advances in Space Research, 49(7), 1167-1179.
2. Jilani, A.A., Zai, M.A.K.Y., Afridi, F.A.K., Mian, K. and Alam, S.N. (2013). Seasonal Variability in Virtual Height of Ionospheric F_2 Layer at the Pakistan Atmospheric Region, *Proceedings of the Pakistan Academy of Sciences*, 50(2), 151-158.
3. Mudzingwa C., Nechibvute A. and Chawanda A. (2013). Maximum Useable Frequency Prediction Using Vertical Incidence Data. *International Journal of Engineering Research & Technology (IJERT)*, 2(8), 2050-2056.
4. Jilani, A.A. and Zai, M.A.K.Y. (2011). Quantifying the seasonal variation in virtual height of ionosphere F2 layer at Pakistan atmospheric region. *Journal of Information & Communication Technology*, 5(2), 52-60.
5. Hassan, D., Abbas, S., Ansari, M.R.K. and Jan, B. (2014). Solar Flares Data Analysis on application of Probability Distributions and Fractal Dimensions and a comparative analysis of North-South Hemispheric Solar Flares Data Behavior.
6. Kale, M. and Butar, F.B. (2011). Fractal analysis of time series and distribution properties of Hurst exponent. *Journal of Mathematical Sciences & Mathematics Education*, *5*(1), 8-19.
7. Zai, M.A.K.Y. (1989). Estimation of fractal dimension of snow and ice crystals, *Indian J. Phys*. 63A(6), 564-753.
8. López-Montes, R., Pérez-Enríquez, R., Araujo-Pradere, E.A. and Cruz-Abeyro, J.A.L. (2015). Fractal and wavelet analysis evaluation of the mid latitude ionospheric disturbances associated with major geomagnetic storms. *Advances in Space Research*, 55(2), 586-596.
9. Kaye, B.H. (1993). *Caos and Complexity: discovering the surprising patterns of science and technology*. Wiley, ISBN: 3527290397, 9783527290390.
10. Tong, H. (2011). Nonlinear time series analysis. In *International Encyclopedia of Statistical Science* (pp. 955-958). Springer Berlin Heidelberg.
11. Hamed, K.H. (2007). Improved finite-sample Hurst exponent estimates using rescaled range analysis. *Water Resources Research*, 43(4), 1-9.
12. Kirby, S.S., Berkner, L.V. and Stuart, D.M. (1934). Studies of the ionosphere and their application to radio transmission. *Proceedings of the Institute of Radio Engineers*, 22(4), 481-521.

13. Gneiting, T., Ševčíková, H. and Percival, D.B. (2012). Estimators of fractal dimension: Assessing the roughness of time series and spatial data. *Statistical Science*, 27(2), 247-277.
14. Mandelbrot, B.B. (1985). Self-affine fractals and fractal dimension. *Physica Scripta*, 32(4), 257-260.
15. Mošna, Z., Šauli, P. and Georgieva, K. (2009). Ionospheric response to the particular solar event as seen in the Ionospheric vertical sounding. WDS'09 Proceedings of Contributed Papers, Part II, 68-73, ISBN 978-80-7378-102-6 © *MATFYZPRESS*.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 41-50

WAVELET CHARACTERIZATION OF MAXIMUM USEABLE FREQUENCY (MUF) FOR IONOSPHERIC REGION OF PAKISTAN

M. Ayub Khan Yusuf Zai[1§] and **Bulbul Jan**[2]

[1] Department of Applied Physics and Solar-Terrestrial and Atmospheric Research Wing and Institute of Space and Planetary Astrophysics, University of Karachi, Karachi, Pakistan. Email: ayubzai@yahoo.com

[2] Institutes of Space and Planetary Astrophysics, University of Karachi Karachi, Pakistan. Email: bulbul.gilgit@gmail.com

[§] Corresponding author

ABSTRACT

In this communication, we have described that the propagation of electromagnetic waves in the ionosphere that is considered as complex by the fact that this medium is inhomogeneous, anisotropic, turbulent, non-stationary and multipath. One of the terms used in the Sky wave propagation is Maximum useable frequency (MUF) that refers to the highest possible frequency for high frequency (HF) communications that can be used to transmit over a particular path under given ionospheric conditions that fluctuate continuously due to temporal variations of the ionosphere.

The aim of this research is to study maximum usable frequency (MUF) using Wavelet I-D Haar 1-5 levels for approximation and decomposition over Pakistan region. Wavelet transform can map the power of a particular frequency at different times, giving an expansion of the signals in both time and frequency. For this analysis, a total 151 days of maximum usable frequency data (1st Oct-2015 to 29th Feb-2016) of Pakistan air space were analyzed collected from Space and Upper Atmosphere Research Commission (SUPARCO).

The results obtained in this investigation point out significant information about maximum usable frequency, such as scale variation of the signals and clearly detect the time localization of the period changes. The constructed models of Maximum usable frequency (MUF) data, the detailed and approximated parts at the lowest resolution several peaks are appeared that can be characterized. This study will encourage the investigators working in the same field to verify the complex nature of ionospheric signals.

KEYWORDS

Maximum useable frequencies (MUF), Ionosphere, Wavelet analysis, sky wave propagation.

1. INTRODUCTION

Applications of satellite navigation and satellite communication systems need to assess and monitor Ionospheric effects which may degrade their performance essential

(Kouris et al., 2004). Ionosphere corresponds to the weakly ionized plasma upper part of atmosphere that ranges from about 60km to beyond 1000km altitude within the earth atmosphere (Mudzingwa, 2013, Barclay, 2003) and electrically neutral to high degree of approximation, as positive and negative charged particles are always formed and destroyed together (Barclay, 2003). The term plasma is defined as “quasi-neutral gas of charged and neutral particles that exhibits in collective behavior” (Chen and Trivelpiece, 1974). Characterizing Ionospheric regulatory is of major importance to present both practical (radio-communications and navigation systems) and theoretical (Space–Earth coupling, climatological global change, and anthropogenic activities impact) issues. Therefore, establishing a statistical description of Ionosphere variations and relating them to possible driving sources such as global solar and geomagnetic behavior represent essential stakes. Often, Ionosphere variations are described in terms of long-term trends versus short-term fluctuations (Roux, 2012).

The upper part of atmosphere (ionosphere) is essential to sky-wave radio propagation and provides the basis for almost all HF communications beyond line of sight (Mudzingwa, 2013). Maximum useable frequency (MUF) is a significant Ionospheric parameter for satellite navigation and satellite communication systems and a detailed study of its behavior and especially of its variability from day-to-day, hour-to-hour (Kouris and Fotiadis, 2002) and within the- hour is essential. The term Maximum useable frequency (MUF) defined to the highest possible frequency for high frequency (HF) sky-wave transmission that can be used to broadcast over the given path under given Ionospheric conditions which fluctuate continuously due to temporal variations of the ionosphere (Malik ,2016, Oyekola, 2010). The sky wave Radio propagation users have great interest to know the daily behavior of the maximum usable frequency (MUF) for its applications. The MUF is obviously dependent of the Ionospheric F-layer critical frequency (foF2) and its corresponding propagation factor M (3000) of F2 layer (Souza et al, 2013).

The maximum usable frequency (MUF) is important for radio users in order to achieve better frequency execution (Kouris and Fotiadis, 2002). The MUF is the product of F2-layer critical frequency of F2 and propagation factor M (3000) F2 that is defined by the equation,

$$\text{MUF (3000) F2} = f_o F_2 * \text{M (3000) F2}$$

where, MUF (3000) is Maximum useable frequency received at 3000 km when reflected by ionosphere and $f_o F_2$ described the critical frequency or highest frequency reflected at vertical incidence from a given layer and M(3000)F2 is the propagation factor of F2 layer of which its influence is significantly less than of F2 (Fotiadis 2004, Ijayyil 2011, Malik 2016).

The aim of the wavelet approach in space weather application is the time-frequency decomposition (Domingues et al., 2004).Wavelet analysis is a particular time-or space-scale illustration of signals that has been found in atmospheric application i.e. maximum useable frequency (MUF) fluctuations due to ionosphere and the radio wave interactions. It is newly developed based signal processing approaches allows on several time scales of the local properties of complex signals that can present non-stationary areas (Zai and Mian, 2012). The wavelet approach used to find out the mode of variation and also to explore how it fluctuate with time by decomposition time series in to frequency domain and powerful multi resolution with respect to one dimensional Haar level. Wavelet analysis and synthesis are consider highly important in the field of signal processing which decomposes actual signal in to different signals to be analyzed in to principle and residual part (Mian and Zai, 2013).

2. DATA AND METHOD OF ANALYSIS

The MUF data of five months (1st Oct-2015 to 29th Feb-2016) of Pakistan air space region taken from Space and Upper Atmosphere Research Commission (SUPARCO) which updates daily space weather data.

There are several techniques which were used to study the Ionospheric irregularities based on solar activities and ionospheric properties. For this purpose, wavelet technique was applied to investigate the nonlinear behavior of Ionosphere. It is known that the coefficient of wavelet series realized the properties like strong transient, discontinuity and unforeseen of the function or distribution precisely thus wavelet analysis are considered as a mathematical microscope that detects the aspects of functions at different resolution (Domingues et al., 2004). Mathematically, wavelet analysis defined as;

(i) $S = A_j + \sum D_j$
(ii) $S = A_j + D_j$

where, S, A_j and D_j are defined by principle j level and residual part j level. Further, the relation between principle and residual part j level expressed as;

(iii) $A_{j-1} = A_j + D_j$

In 1910, A. Haar developed wavelet, considered the most suitable function to the introduction and of understanding of wavelet analysis. This study discussed Haar of level 3, 4 and 5 with dyadic scale a=2 and considered levels 3, 4 and 5 the resolution is given by 1/a or 2^{-j} (Mian and Zai, 2013). Further, the simple orthogonal mother wavelet called Haar wavelet presented as

(iv) $\psi(t) = \begin{cases} 1, & 0 \le t < \frac{1}{2} \\ -1, & \frac{1}{2} \le t < 1 \\ 0, & otherwise \end{cases}$

This detects signal abrupt variations i.e. one localized feature in the physical space (Domingues et al., 2004).

3. RESULTS AND DISCUSSION

The applications of wavelet technique have been widely used in the atmospheric sciences in last decades. Now, the wavelet transforms had turned out to be a very valuable approach in atmospheric signal analysis, creating an encouraging new perspective to the research activities. By comparison, wavelet transform is better than Fourier transforms due to high-resolution features, good localization both in time and scale domains and its capacity of analyzing signals at multi time scales,. Furthermore, wavelet analysis also has the ability of diagnosing the jump features of signals. Therefore, wavelet analysis has been applied to the studies of atmospheric multi-time scales analysis (Singha and Bhatlab, 2012, Domingues et al., 2004).

Fig. 1 shows the variation of different resolution at level 1-3 of Haar wavelet form in the detailed and approximated part and the cyclic variation also presented at different level. In the detailed and approximation part at the lowest resolution several peaks are appeared and the upper panel of this figure the signal representation is given. This conformed that the time localization of the period changes can be clearly detected.

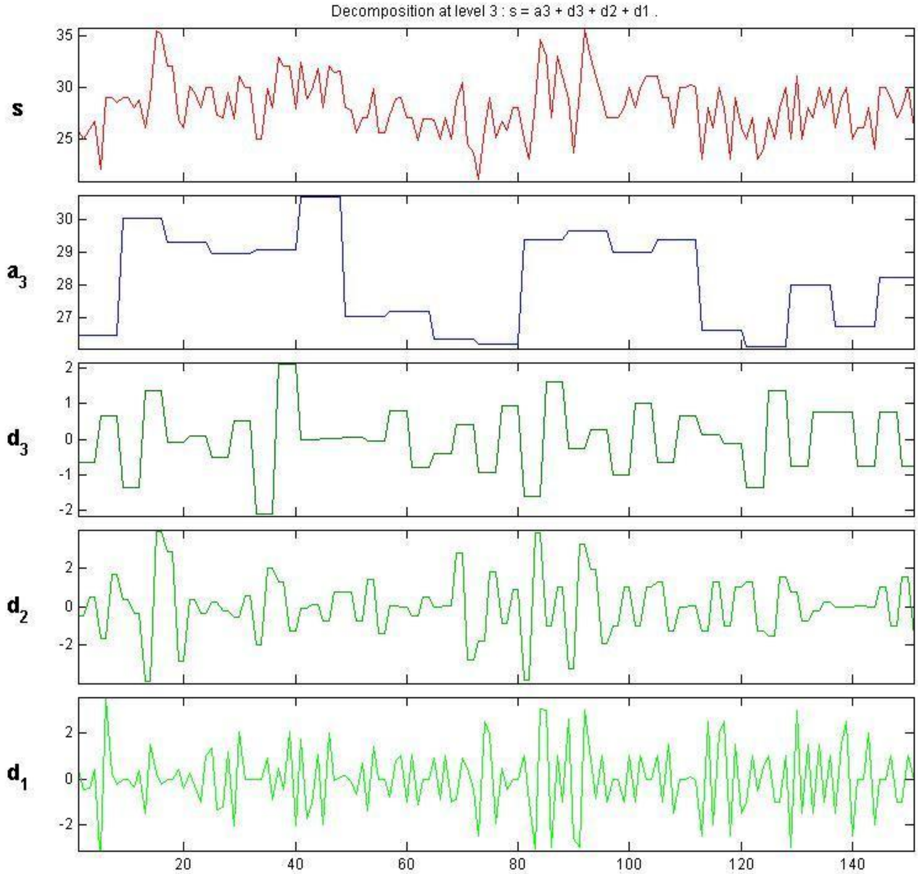


Fig. 1 Decomposition at level 3, s = a_3+ d_3 + d_2 + d_1 of Ionosphere of F2 layer MUF at Pakistan air space

For more convenient discussion about the behavior of wavelet, we compared this illustration to level 4 and level 5 of Haar wavelet transform as revealed in Figure 2.

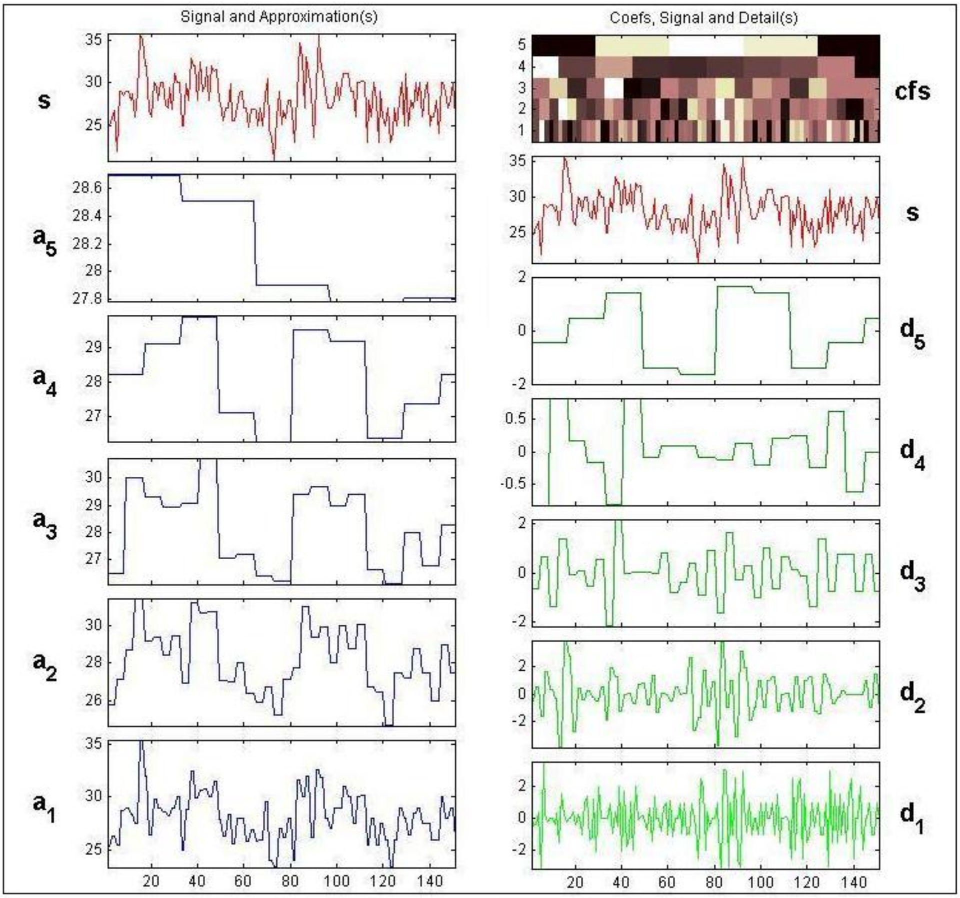


Fig. 2 Plot of decomposition at level, a_1, a_2, a_3, a_4, a_5 and d_1, d_2, d_3, d_4, d_5 of F2 layer MUF at Pakistan air space

The wavelet irregularity can be used to find out the exact values of the distinctive frequencies. This can be only possible in Haar wavelet transform which has the best localization in frequency domain and also to observe the high values of the modulus of the wavelet coefficients indicate a transition region among unusual nature of movement (Domingues et al., 2004). Figure: 3 (a-c) indicate original signal , the upper panel approximation level 1 and lower panel details level 1 of MUF of F2.

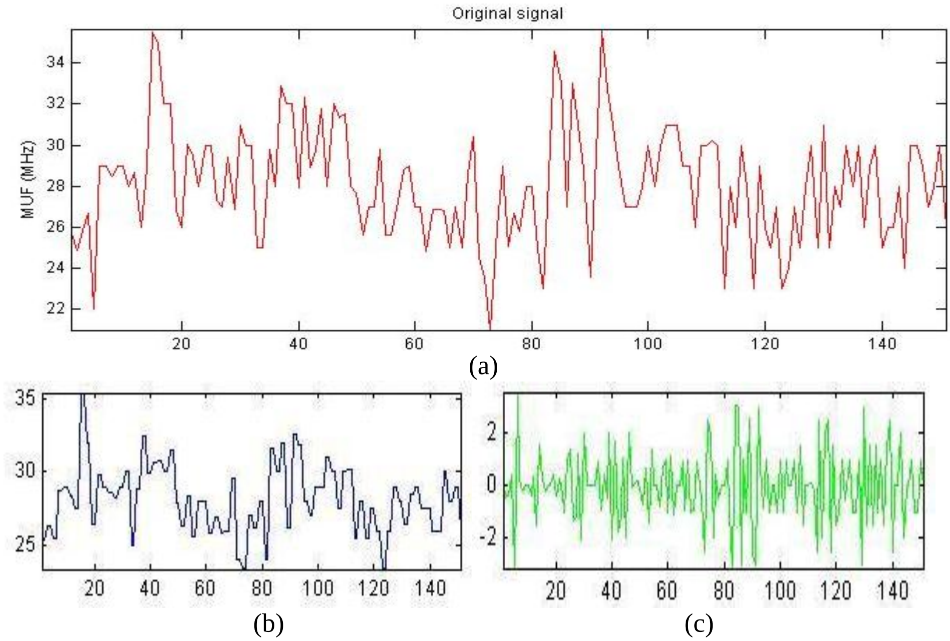


Fig. 3 (a-c): Plot of Original Signals, Upper Panel Approximation Level 1 and Lower Panel Detail level 1 of MUF Ionosphere of F2 Layer at Pakistan Air Space

The Figure 4(a-b) illustrated the upper panel approximation level 2 and lower panel detail level 2 of MUF of F2. Similarly, Figure 5(a-b), 6(a-b) and 7(a-b) depict the plots of upper panel approximation level of 3, 4, 5 and lower panel details level of 3, 4, 5 of MUFF2 at Pakistan region respectively.

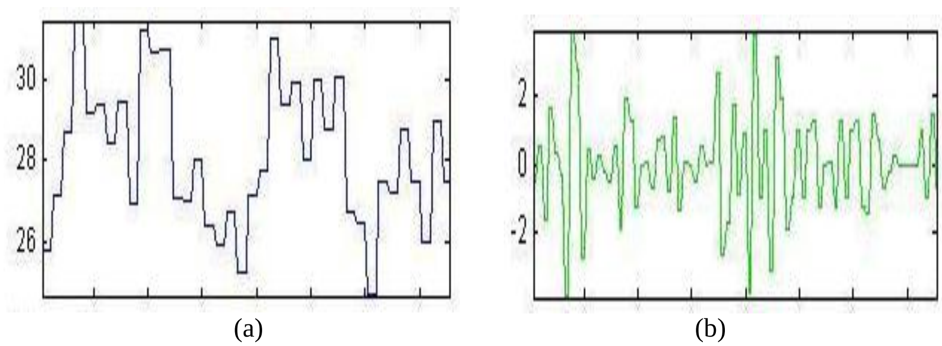
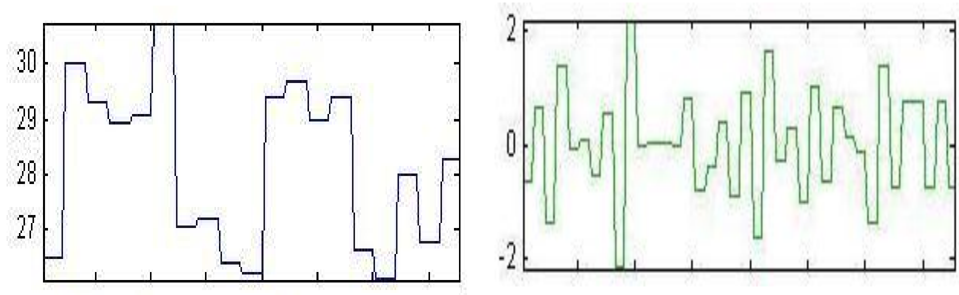
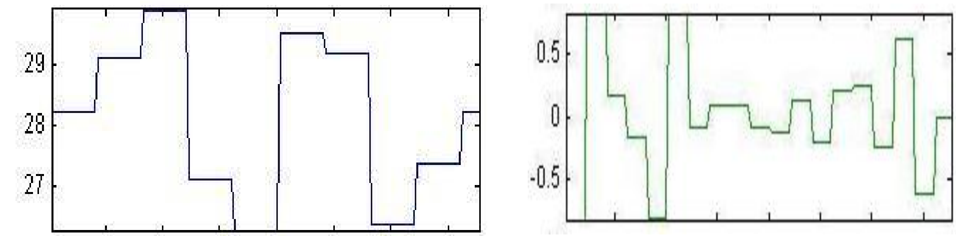


Fig. 4 (a-b): Plot of Upper Panel Approximation Level 2 and Lower Panel Detail Level 2 of MUF Ionosphere of F2 Layer at Pakistan Air Space



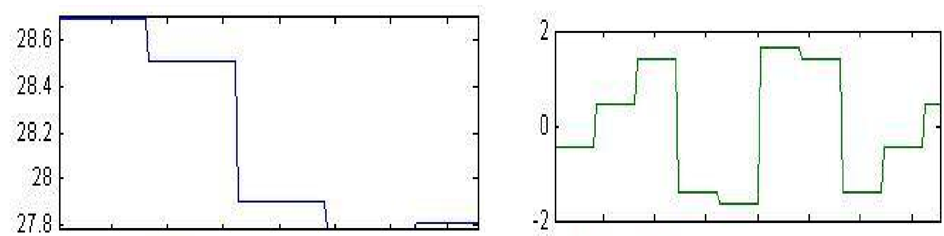
(a) (b)

Fig. 5 (a-b) Plot of Upper Panel Approximation Level 3 and Lower Panel Detail Level 3 of MUF Ionosphere of F2 Layer at Pakistan Air Space



(a) (b)

Fig. 6 (a-b): Plot of Upper Panel Approximation Level 4 and Lower Panel Detail Level 4 of MUF Ionosphere of F2 Layer at Pakistan Air Space



(a) (b)

Fig. 7 (a-b): Plot of Upper Panel Approximation Level 5 and Lower Panel Detail Level 5 of MUF Ionosphere of F2 Layer at Pakistan Air Space

4. COMMENTS AND CONCLUSION

This study elaborates the facts about the variations of MUF for radio wave transmission of F2 layer using wavelet approach at Pakistan air space region. The meaningful data of maximum useable frequency was collected from Space and Upper Atmosphere Research Commission (SUPARCO) Pakistan which updates on daily basis. Wavelet, I-D Haar 1-5 levels for approximation and decomposition details are achieved.

To observe the behavior of MUF, we have compared illustration of level 3, 4 and level 5 of Haar wavelet transform. To declare a comprehensive sympathetic and to ensure the consistency of the obtained results in the investigation of space weather phenomena using this technique, a good knowledge of the wavelet implementation is required.

The outcomes obtained in this paper point out significant information about maximum usable frequency or scale variation of the signals and clearly detect the time localization of the period changes. The constructed models of Maximum usable frequency (MUF) data, the detailed and approximated parts at the lowest resolution several peaks are appeared that can be characterized.

This study will support new horizon to the research activities working in the field of Ionospheric signals to verify their complex nature using wavelet transforms.

5. ACKNOWLEDGMENT

We are thankful to the technical staff of Space and Upper Atmosphere Research Commission (SUPARCO) for updating daily Ionospheric data like critical frequency, maximum usable frequency (MUF) and other variables on their respective website from where the space weather data can be freely downloaded on daily basis. Also the authors of this communication appreciate the perspectives of the organizers of this 14th International Conference on Statistical Sciences that was jointly organized by Jinnah Sindh Medical University Karachi and Islamic Countries Society of Statistical Sciences from 14-16 March, 2016.

REFERENCES

1. Oyekola, O.S. (2010). Variation in the ionospheric propagating factor M (3000) F2 at Ouagadougou, Burkina Faso. *Advances in Space Research*, 46(1), 74-80.
2. Malik, R.A., Abdullah, M., Abdullah, S. and Homam, M.J. (2016). Comparison of maximum usable frequency (MUF) variability over Peninsular Malaysia with IRI model during the rise of solar cycle 24. *Journal of Atmospheric and Solar-Terrestrial Physics*, 138, 87-92.
3. Hadi, K.A. and Aziz, A.Z. (2012). Studying the Impact of the Solar Activity on the Maximum Usable Frequency Parameter over Iraq Teritory. *IOSR Journal of Computer Engineering (IOSRJCE)*, 5(3), 35-39.
4. Kouris, S.S. and Fotiadis, D.N. (2002). Ionospheric variability: a comparative statistical study. *Advances in Space Research*, 29(6), 977-985.
5. Fotiadis, D.N., Baziakos, G.M. and Kouris, S.S. (2004). On the global behaviour of the day-to-day MUF variation. *Advances in Space Research*, 33(6), 893-901.
6. Wen, Z., and Xun-Jie, Z. (1999). Predictions of HF communication MUF in the region of the South China Sea. *Antennas and Propagation Magazine, IEEE*, 41(4), 35-38.
7. Malik, R.A., Khamis, K., Zakariya, A.S., Tompang, A., Mohamed, H.T., Kamal, M.R., Rajab, A.R., Lay, L.S., Kasmoni, H., Ramli, I., Zahari, N., Jaafar, J., Hambali, H., Manaf, Z.A. and Mustafa, M. (2010). Determination of Maximum Usable Frequency (MUF) for High Frequency (HF) communications. *Def. Sci. Technol. Tech. Bull*, 3(1), 1-13.

8. Chen, C., Wu, Z.S., Sun, S.J., Ding, Z.H. and Zhao, Z.W. (2008). MUF variability at Haikou, China. In *8th International Symposium on Antennas, Propagation and EM Theory, 2008. ISAPE 2008* (pp. 400-403). IEEE.
9. Ijayyil, M, (2011). Investigation of the variation of maximum usable frequency (MUF) to ensure a good communications link between two locations, Thesis Master of Science, College of Science, University of Baghdad.
10. Roux, S.G., Knížová, P.K., Mošna, Z. and Abry, P. (2012). Ionosphere fluctuations and global indices: A scale dependent wavelet-based cross-correlation analysis. *Journal of Atmospheric and Solar-Terrestrial Physics*, 90, 186-197.
11. Gilliland, T.R., Kirby, S.S., Smith, N. and Reymer, S.E. (1938). Maximum usable frequencies for radio sky-wave transmission, 1933 to 1937. *Proceedings of the Institute of Radio Engineers*, 26(11), 1347-1359.
12. Mudzingwa C., Nechibvute A. and Chawanda A. (2013). Maximum Useable Frequency Prediction Using Vertical Incidence Data. *International Journal of Engineering Research & Technology* (IJERT), 2(8), 2050-2056.
13. Barclay, L.W. (2003). *Propagation of Radio Waves*, Institution of Electrical Engineers, London.
14. Chen, F.F. and Trivelpiece, A.W. (1976). Introduction to plasma physics. *Physics Today*, 29, 54.
15. Zai, M.A.K.Y. and Mian, K., (2012). Wavelet Characterization of Turbulence for Ionospheric Region of the Pakistan Air Space, *Proceedings of the Pakistan Academy of Sciences*, 49(4), 295-300.
16. Domingues, M.O., Mendes Jr, O. and Da Costa, A.M. (2004). On wavelet techniques in atmospheric sciences. *First Latin American Advanced School on Space Environment*.
17. Mian, K. and Zai, M.A.K.Y. (2013). Comparison of Wavelet Characterization of Computed Perturbed and Un-perturbed Parcel Velocities of F. *Proceedings of the Pakistan Academy of Sciences*, 50(1), 71-76
18. Singha, R.C. and Bhatlab, R. (2012). Wavelets in meteorology. In *AIP Conf. Proc.*, 1463, 176-193.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 51-56

MEASURING RETURN IN PORTFOLIO

Ammara Nawaz Cheema[1] and **Ahmad Nawaz Cheema**[2]

[1] Department of Mathematics, Air University, Islamabad Pakistan. Email: ammara.au@gmail.com

[2] Financial Analyst, Snl Financial, Islamabad, Pakistan Email: ahmadnawaz01@gmail.com

[§] Corresponding author

ABSTRACT

In this study we increased awareness regarding the benefits of investment in a well diversified portfolio and the investment options ever. We took two types of investor, one who invest his capital in one company, the other invest his capital in more than one company. For drawing the inference we used regression model, best subset selection, and proportion test. After applying these tests we found that the return on investment in a well diversified portfolio for a short period of time as well as for a long period of time is greater than the return on investment in an undiversified portfolio.

KEY WORDS

Investment, portfolio, regression, best subset selection, proportion test, investor.

1. INTRODUCTION

It is well known that stock market investing is risky. Both practitioners and theoreticians recommend holding a well-diversified portfolio to reduce risk. While mutual funds offer a quick and relatively inexpensive way to diversify. It is also important that the diversification in portfolio must be done properly which means the company that are taken in the portfolio should have no relation between each other. If there will be no relation between the companies in such portfolio only then it will be possible; that if the return from one company goes down; other might goes up.

Another important thing in portfolio investment is the share of investment that should be higher in such portfolio which gives maximum returns and the share of investment should be lower in the portfolio which gives low returns. It is obvious that a group of companies giving high returns should get more investment share. Among the more incommodious problems of a business organization, are decisions; relating to the optimal allocation of resources among competing investment opportunities.

The classical theory of portfolio choice rests on strong assumptions as no transaction costs, investor's awareness of menu of asset available and knowledge of their risk and return, no uninsurable risks, such as human capital. If all investors face the same distribution of returns and have the same information set, in equilibrium they select the same menu of risky assets (Arjan, et al. 2004).

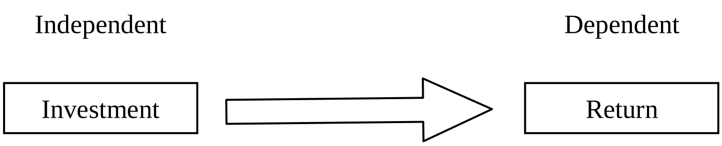
A changing trend in the financial sector is observed during the past decade, which is now focusing its attention towards lending to the common salaried individuals. The objective of this policy is to encourage the common people to invest in small business ventures and other portfolio which will help increasing the economic activity in the country. This policy has opened many new options for investment and portfolio diversification. Among the better known such approaches are those attributable to (Markowitz 1952; Drucker 1963; Basu 1977; Wright 1978; Allan 1979, Wind 1975, and Sheth and Frazier 1983). At a more strategic level, researchers concerned with Research and Development have proposed a variety of tools for use in the selection of individual projects. Examples of these approaches include scoring procedures (Dean and Nishry 1965), and various process models of innovation (Utterback and Abernathy 1975). In addition, different portfolio models appear to produce different managerial implications (Wind, Mahajan, and Swire 1983). A number of such tools and their particular strengths and weaknesses are also discussed by Anderson (1981).

Earlier studies exploring the problem of optimal portfolio size mostly focus on how many stocks are sufficient to make an efficient portfolio. For example, Statman (1987) reveals that a portfolio including 30-40 stocks can effectively achieve efficient diversification. Related researches can be found in Chung (2000) and Shawky and Smith (2005).

Goetzmann & Kumar (2008) find substantially concentrated portfolios in the individual brokerage accounts. The current empirical literature on diversification has two main limitations. First, the samples used often are not representative of the entire population. Second, diversification is considered only for a part of the household portfolio. In recent years, there has been a growing interest in incorporating financial planning tools in product portfolio approaches and project selection techniques (Rezayat and Yavas 2006).This research is aim to observing the level of profit available to the people in portfolio investment and the portfolio options available to the people. There are many opportunities to make sound investment that have a positive financial social impact. Learning from the financial intermediaries is just one of the ways that people become aware of investment opportunities.

2. METHODOLOGY

As far as our research is concerned in the theoretical framework we have taken the return as a dependent variable and independent variable is investment.



In this research we have done quantitative research study of two different types of investors who are investing their capital, out of these two one person invest his capital in one company, the other invest his capital in more than one company. Basically from this data we have measured the effect of investing in just one company with respect to investing the capital in more than one company that means in a diversified portfolio firm.

Population of our research is the listed companies of Pakistan. These companies are as follows; Chemical Industry, Fertilizer Industry, Vehicle Industry, Cement Industry, Insurance Industry, Oil & Gas Industry, Glass Industry, Banks. In this research we calculate the actual return from each investment done in companies. Basically we have taken two different types of investors. From the portfolio of these two investors we have checked the effect of investments in a group of companies as well as investment in only one company.

Share prices of different companies are taken for drawing inference about the return on investment. There are 36 observations related to diversified and undiversified portfolio (3-Year dataset on monthly basis), and data for longer time period is also collected (3-Year dataset on 3-month basis), these observations are taken with 15 different industries related to diversified and 15 companies of a single industry related to an undiversified portfolio.

3. RESULTS AND DISCUSSION

In our research return (Y) as response variable and investment (X) is predictor. Regression analysis is performed to estimate fitted values of portfolio return (Y).

The results of regression equation of undiversified portfolio are as:

$Y = -0.0272\,X + U$ (3-Year dataset on monthly basis)

and

$Y = -0.0774\,X + U$ (3-Year dataset on 3-month basis)

These models indicate that investment in undiversified portfolio is inversely related with return. The regression model is appropriate as it satisfied all the basic assumptions like; normality of residuals, homoscedastic and autocorrelation.

The results of regression equation of undiversified portfolio are as:

$Y = 0.00160\,X + U$ (3-Year dataset on monthly basis)

and

$Y = 0.0051\,X + U$ (3-Year dataset on 3-month basis)

These models indicate that investment in diversified portfolio is directly related with return. Value of adjusted R^2 indicates that 92% of change in return is due to investment and remaining 8% is due to other factors either whose individual impact on return is not significant or highly correlated with the investment included in the equation.

To improve the model, an attempt has been made; we apply the Akaike's information criterion (AIC) and Schwarz' Bayesian criterion (SBC) on the same predictor which used in the regression procedure.

The table 1 reports these values of AIC and SBC.

Table 1
AIC and SBC

Model	**AIC**	**SBC**
Undiversified Portfolio	7.46	7.51
Diversified Portfolio	7.43	7.47

As we know that the model having small values of AIC or SBC is considered best model, so the Diversified Portfolio model having minimum value of AIC (7.43) is:

$$Y = 0.00160\ X$$

Here note that the SBC have minimum value (7.47). One can choose the same model for prediction of the best model and may conclude that investment in well diversified portfolio will increases the return.

Both the criterions discussed above select the same prediction model of return with predictors as investment in well diversified portfolio.

From the above prediction models, it is clear that investment in undiversified portfolio, the return on the average decrease. Hence we can conclude the investment in undiversified portfolio contributes negatively to return. On the other hand, return increases as the investment in well diversified portfolio increases. The effect of investment in well diversified portfolio is positive on return. So our study recommends that any investment in well diversified portfolio is suitable for better return.

Proportion test is performed on data set with the following hypothesis:

$$H_1 : R_{dp} > R_{udp}$$

or $H_1 : p > 1$

where

$$p = R_{dp} / R_{udp}$$

Hypothesis for investment in long time period as follows:

$$H_{2:} Rl_{dp} > Rl_{udp}$$

or $H_{2:}$ p > 1

where

$$p = Rl_{dp} / Rl_{udp}$$

Our research hypothesis can be tested in this case by t-statistics as:

$$t = \left(p - p_0\right) \Big/ \sqrt{\left(p_0 q_0 / n\right)}\ .$$

Table 2
Proportion Test

Test	For Short period	For Long period
t-statistic	0.11627	0.9081
P-Value	0.50577	0.623

From the results it is clear that we do not reject H_1 and H_2 at the 5% level of significance and conclude that the investment in well diversified portfolio increases the return (for both time periods).

From the above hypothesis testing, it is clear that investment in undiversified portfolio for a long period (3-Months) and short period (1-Month), the return on the average decreases. On the other hand, return increases as the investment in well diversified portfolio increases for both (1 and 3 Months) time periods. So our study recommends that any investment for both time periods in well diversified portfolio can increases ones return.

4. COMMENTS AND CONCLUSION

On the basis of above results we conclude that investing in a well-diversified portfolio results in increased overall profitability and the risk involved is significantly reduced to a bare minimum level. Diversifying is basically a risk management technique that attempts to provide some insurance against the unexpected. The rule of thumb for diversification is to combine investments that have low or negative correlation. In order to have maximum return a person must invest his total capital in the form of a well-diversified portfolio investment that is investing in two or more firm belonging to different industries.

If a person is investing in just one company he/she has to face more risk because if that particular industry goes down then it will affect his/her total investment and as the result the lost he/she may be facing can be maximum.

REFERENCES

1. Allen, M.G. (1979). Diagnosing GE's Planning for What's Watt. *Corporate Planning: Techniques and Applications*, AMACOM, New York, 211-220.
2. Anderson, P.F. (1981). Marketing Investment Analysis, in J. N. Sheth (Ed.), *Research in Marketing*, Vol. 4, JAI Press, Greenwich, CT, 1-38.
3. Berkelaar, A.B., Kouwenberg, R. and Post, T. (2004). Optimal Portfolio Choice under Loss Aversion. The Review of Economics and Statistics, 86(4), 973-987.
4. Basu, S. (1977). Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. *The Journal of Finance*, 32(3), 663-682.
5. Chung, H.K. (2000). *How Many Securities Make a Well-Diversified Portfolio: KLSE Stocks*. PhD Thesis, Universiti Sains Malaysia.
6. Dean, B.V. and Nishry, M.J. (1965). Scoring and Profitability Models for Evaluating and Selecting Engineering Projects. *Oper. Res.*, 13, 550-570.

7. Drucker, P.F. (1963). Managing for Business Effectiveness. *Harvard Business Rev*., 41, 59-60.
8. Goetzmann, W.N. and Kumar, A. (2008). Equity Portfolio Diversification. Review of Finance. *Oxford University Press for European Finance Association*. 12(3), 433-463
9. Markowitz, H.M. (1952). Portfolio selection. *J. Finance* (7), 77-91. Wiley, New York.
10. Rezayat, F. and Yavas, B.F. (2006). International Portfolio Diversification: A Study of Linkages among the U.S., European and Japanese Equity Markets. *Journal of Multinational Financial Management*, 16(4), 440-458.
11. Shawky, H.A. and Smith, D.M. (2005). Optimal Number of Stock Holdings in Mutual Funds Portfolio Based on Market Performance. *Financial Review*, 40, 481-495.
12. Sheth, J.A., Frazier, G.L. (1983). A Margin-Return Model for Strategic Planning. *J. Marketing*, 47, 100-109.
13. Statman, M. (1987). How Many Stocks Make a Diversified Portfolio? *Journal of Financial and Quantitative Analysis*. 22, 353-363.
14. Utterback, J.M. and Abernathy, W.J. (1975). A Dynamic Model of Process and Product Innovation, *Omega*, 3, 639-655.
15. Wind, Y. (1975). Product Portfolio: A New Approach to the Product Mix Decision. Combined Proceedings, American Marketing Association, Chicago.
16. Wind, Y., Mahajan, V. and Swire, D.J. (1983). An Empirical Comparison of Standardized Portfolio Models. *J. Marketing*, 47, 89-99.
17. Wind, Y. and Robertson, T.S. (1983). Marketing Strategy: New Directions for Theory and Research. *J. Marketing*, 47, 12-25.
18. Wright, R.V.L. (1978). *A System for Managing Diversity*. *Marketing Management and Administrative Action*, (McGraw-Hill, New York, 46-60.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 57-66

IMPROVED RATIO-TYPE ESTIMATORS OF POPULATION MEAN IN RANKED SET SAMPLING USING TWO CONCOMITANT VARIABLES

Lakhkar Khan[1&2] and **Javid Shabbir**[2]

[1] Department of Statistics, Government College Toru, Khyber Pathunkhwa Pakistan. Email: lakhkarkhan.stat@gmail.com

[2] Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan Email: javidshabbir@gmail.com

ABSTRACT

In this paper, we propose an efficient class of ratio-in-exponential-type estimators with two concomitant variables using Ranked Set Sampling (RSS) scheme which improves the available estimators. The biases and mean square errors (MSEs) of the proposed estimators are obtained to first degree approximation. Comparison among the proposed and competitor estimators are made both theoretically and through simulation study. It turned out that when the variable of interest and the concomitant variables jointly followed a tri-variate gamma distribution, the proposed class of estimators dominates all other competitor estimators.

KEYWORDS

Ranked set sampling, *MSE*, Bias, concomitant variables.

Mathematics subject classification 62 D05

1. INTRODUCTION

Rank set Sampling (RSS) is a sampling technique which is used to reduce cost and increase efficiency in that situation where the measurement of survey variable is costly and time consuming, but it can be ranked easily at no cost or at very little cost. The technique of RSS was first introduced by Mclntyre (1952) to increase efficiency of the estimator of the population mean. The general method of a RSS can be described as follows, first, *m* subsamples, each of size m, are drawn at random from a population. Next, for each sub sample, the elements in the subsample are ranked relating to the concomitant variables, and then one and only one element of the subsample rank is measured. The procedure produces a sample of n measurement of independent order statistics.

Takahasi and Wakimoto (1968) proved the mathematical theory that the sample mean under RSS is an unbiased estimator of the finite population mean and more precise than the sample mean estimator under simple random sampling (SRS). In some situations, ranking may not be done perfectly. To tackle this problem, stokes (1977) considered the case where the ranking of elements is done on basis of the auxiliary variable X instead of judgment. Singh et al. (2014) proposed an estimator for population mean and ranking of the elements is observed on the basis of auxiliary variable. The use of the auxiliary information plays an important role in increasing efficiency of the estimators. Samawi and Muttlak (1996) have suggest an estimator for population ratio in RSS and showed that it has less variance as compared to usual ratio estimators in SRS. Khan and Shabbir (2015) suggested a class of Hartley-Ross type unbiased estimator in *RSS*. Khan and Shabbir (2016) have also suggested Hartley-Ross type unbiased estimators in *RSS* and stratified ranked set sampling (*SRSS*). Khan et al. (2016) proposed unbiased ratio estimator of finite population mean in *SRSS*.

Munoz and Rueda (2009) used relative bias (*RB*) and the relative root mean square error (*RRMSE*), for the comparison of different estimators. For the more detail see chambers and Dunstan (1986), Rao et al. (1990), Silva and skinner (1995) and Harms and Duchesne (2006).

In this article, we investigate the properties of the usual mean estimator in RSS and propose and efficient class of the ratio-in-exponential type estimators using two concomitant variables under RSS scheme.

2. RSS PROCEDURE WITH TO CONCOMITANT VARIABLES

In ranked set sampling m independent random samples each of size m are chosen and the items in each sample are selected with equal probability and without replacement from a finite population of size N. the items of each random sample are ranked with respect to the characteristic of the study variable or concomitant variable. Let Y is the study variable and X and Z are the two concomitant variables. Then randomly select m^2 trivariate sample elements from the population and allocate them into m sets, each of size m. each sample is ranked with respect to one of the concomitant variable X or Z. Here, ranking is done on the basis of the concomitant variable X. An actual measurement from the first sample is then taken on the item with the smallest rank of X, together with variable Y and Z associated with smallest rank of X. From the second sample of size m, the variables Y and Z values associated the highest rank of X and measured from m^{th} sample. This completes a cycle of sampling. The procedure is repeated r times to obtain a sample size $n = mr$ items. Thus in a RSS scheme, a total of $m^2 r$ items have been drawn from the population and only mr of them are selected for analysis. To estimate population mean $\left(\bar{Y}\right)$ in RSS using regression type estimator with two concomitant variables, the procedure can be summarized as follows:

Step 1: Randomly select m^2 trivariate sample items from the population.

Step 2: Allocate these m^2 items into m sets, each of size m .

Step 3: Each set is ranked with respect to the concomitant variable X.

Step 4: Select the i^{th} ranked item from the i^{th} $\left(i = 1, 2, ..., m\right)$ set for actual magnitude.

Step 5: Repeat steps 1 through 4 for r cycles until the desired sample size $n = mr$, is obtained.

The usual RSS estimators $\bar{y}_{(RSS)}$ and its variance, are given by

$$\bar{y}_{(RSS)} = \frac{1}{mr}\sum_{j=1}^{r}\sum_{i=1}^{m} y_{[i:m]j}, \tag{1}$$

$$V(\bar{y}_{(RSS)}) = \bar{Y}^2\left(\gamma C_y^2 - W_{[y]}^2\right), \tag{2}$$

where $W_{[y]}^2 = \dfrac{1}{m^2 r\bar{Y}^2}\sum_{i=1}^{m} \mathrm{T}_{y[i:m]}^2$, $\mathrm{T}_{y[i:m]}^2 = \left(\mu_{y[i:m]} - \bar{Y}\right)$, $\gamma = \left(\dfrac{1}{mr}\right)$ and C_y is the coefficient of variation of y. The value of $\mu_{y[i:m]}$ depends on order statistics from some specific distribution (see Arnold et al, 1992).

3. PROPOSED CLASS OF ESTIMATORS IN RSS

We propose a class of ratio-in-exponential type estimators having two auxiliary variables X and Z in RSS as

$$\bar{y}_{G(RSS)} = \bar{y}_{[rss]}\left(\frac{\bar{X}}{\bar{x}_{(rss)}}\right)^{\alpha_1}\left(\frac{\bar{Z}}{\bar{z}_{[rss]}}\right)^{\alpha_2}\left[k\exp\left(\frac{\bar{X}-\bar{x}_{(rss)}}{\bar{X}+\bar{x}_{(rss)}}\right) + (1-k)\exp\left(\frac{\bar{Z}-\bar{z}_{[rss]}}{\bar{Z}+\bar{z}_{[rss]}}\right)\right] \tag{3}$$

where

$$\bar{y}_{[rss]} = \frac{1}{mr}\sum_{j=1}^{r}\sum_{i=1}^{m} y_{[i:m]j},\ \ \bar{x}_{(rss)} = \frac{1}{mr}\sum_{j=1}^{r}\sum_{i=1}^{m} x_{(i:m)j},\ \ \bar{z}_{[rss]} = \frac{1}{mr}\sum_{j=1}^{r}\sum_{i=1}^{m} z_{[i:m]j},$$

α_1 and α_2 are unknown constant whose values are to be determined so that MSE of $\bar{y}_{G(RSS)}$ is minimized and k is a scalar quantity which can take 0 or 1 values. Also $\bar{X}$, $\bar{Y}$ and $\bar{Z}$ are the population means of X, Y and Z respectively.

To find the bias and MSE of the estimators, we define the following error terms:

Let $\bar{y}_{[rss]} = \bar{Y}\left(1+e_0\right)$, $\bar{x}_{(rss)} = \bar{X}\left(1+e_1\right)$, $\bar{z}_{[rss]} = \bar{Z}\left(1+e_2\right)$

Such that $E\left(e_p\right) = 0$, (p=0, 1, 2,), and

$$E\left(e^2\right) = \gamma C_y^2 - W_{[y]}^2 = V_{200},\ E\left(e_1^2\right) = \gamma C_x^2 - W_{(x)}^2 = V_{020},$$

$$E\left(e_2^2\right) = \gamma C_z^2 - W_{[z]}^2 = V_{002},$$

$$E\left(e_0 e_1\right) = \gamma C_{yx} - W_{(yx)} = V_{100},\ E\left(e_0 e_2\right) = \gamma C_{yz} - W_{[yz]} = V_{101},$$

$$E\left(e_1 e_2\right) = \gamma C_{xz} - W_{(xz)} = V_{011},$$

$$W_{[y]}^2 = \frac{1}{m^2 r\bar{Y}^2}\sum_{i=1}^{m}\mathrm{T}_{y[i:m]}^2,\ W_{(x)}^2 = \frac{1}{m^2 r\bar{X}^2}\sum_{i=1}^{m}\mathrm{T}_{x(i:m)}^2,\ W_{[z]}^2 = \frac{1}{m^2 r\bar{Z}^2}\sum_{i=1}^{m}\mathrm{T}_{z[i:m]}^2$$

$$W_{(yx)}=\frac{1}{m^2 r\overline{Y}\,\overline{X}}\sum_{i=1}^{m}\mathrm{T}^2_{yx(i:m)}\ ,\ \ W_{[yz]}=\frac{1}{m^2 r\overline{Y}\,\overline{Z}}\sum_{i=1}^{m}\mathrm{T}^2_{yz[i:m]}\ ,$$

$$W_{(xz)}=\frac{1}{m^2 r\overline{X}\,\overline{Z}}\sum_{i=1}^{m}\mathrm{T}^2_{xz(i:m)}$$

$$\mathrm{T}^2_{y[i:m]}=\left(\mu_{y[i:m]}-\overline{Y}\right),\quad \mathrm{T}^2_{x(i:m)}=\left(\mu_{x(i:m)}-\overline{Y}\right),\quad \mathrm{T}^2_{z[i:m]}=\left(\mu_{z[i:m]}-\overline{Y}\right)$$

$$\mathrm{T}^2_{yx(i:m)}=\left(\mu_{y[i:m]}-\overline{Y}\right)\left(\mu_{x(i:m)}-\overline{X}\right),\quad \mathrm{T}^2_{yz[i:m]}=\left(\mu_{y[i:m]}-\overline{Y}\right)\left(\mu_{z[i:m]}-\overline{Z}\right),$$

$$\mathrm{T}^2_{xz(i:m)}=\left(\mu_{x(i:m)}-\overline{X}\right)\left(\mu_{z[i:m]}-\overline{Z}\right).$$

Here $C_{xy}=\rho_{xy}C_xC_y$, $C_{yz}=\rho_{yz}C_yC_z$, $C_{xz}=\rho_{xz}C_xC_z$, where C_x , C_y and C_z are the coefficient of variation of X, Y and Z respectively. The value of $\mu_{x[i:m]}$ and $\mu_{z[i:m]}$ depend on order statistics from some specific distributions.

In terms of e's upto first order of approximation, we have

$$\bar{y}_{G(RSS)}=\overline{Y}\left[1+e_0-\alpha_1e_1-\alpha_1e_0e_1+\frac{1}{2}\alpha_1\left(\alpha_1+1\right)e_1^2\right]\left[1-\alpha_2e_2+\frac{1}{2}\alpha_2\left(\alpha_2+1\right)e_2^2\right]$$

$$\left[k\left(1-\frac{1}{2}e_1+\frac{3}{8}e_1^2\right)+\left(1-k\right)\left(1-\frac{1}{2}e_2+\frac{3}{8}e_2^2\right)\right],$$

or

$$\left(\bar{y}_{G(RSS)}-\overline{Y}\right)=\overline{Y}\begin{bmatrix} e_0-\frac{1}{2}\left(k+2\alpha_1\right)e_1-\frac{1}{2}\left(1-k+2\alpha_2\right)e_2 \\ +\frac{1}{8}\left\{k\left(4\alpha_1+3\right)+4\alpha_1\left(\alpha_1+1\right)\right\}e_1^2 \\ +\frac{1}{8}\left\{(1-k)\left(4\alpha_2+1\right)+4\alpha_2\left(\alpha_2+1\right)\right\}e_2^2 \\ -\frac{1}{2}\left(k+2\alpha_1\right)e_0e_1-\frac{1}{2}\left(1-k+2\alpha_2\right)e_0e_2 \\ +\frac{1}{2}\left\{\alpha_1(1-k)+\alpha_2k+2\alpha_1\alpha_2\right\}e_1e_2 \end{bmatrix}. \qquad (4)$$

The bias of $\bar{y}_{G(RSS)}$, is given by

$$Bias\left(\bar{y}_{G(RSS)}\right) \cong \bar{Y}\begin{bmatrix} \frac{1}{8}\left\{k\left(4\alpha_1+3\right)+4\alpha_1\left(\alpha_1+1\right)\right\}V_{020} \\ +\frac{1}{8}\left\{(1-k)\left(4\alpha_2+1\right)+4\alpha_2\left(\alpha_2+1\right)\right\}V_{002} \\ -\frac{1}{2}\left(k+2\alpha_1\right)V_{110}-\frac{1}{2}\left(1-k+2\alpha_2\right)V_{101} \\ +\frac{1}{2}\left\{\alpha_1(1-k)+\alpha_2 k+2\alpha_1\alpha_2\right\}V_{011} \end{bmatrix} \quad (5)$$

Taking square of Eq. (4) and then expectations, the MSE of $\bar{y}_{G(RSS)}$, is given by

$$MSE\left(\bar{y}_{G(RSS)}\right) \cong \bar{Y}^2\begin{bmatrix} V_{200}+\frac{1}{4}\left(k+2\alpha_1\right)^2 V_{020}+\frac{1}{4}\left(1-k+2\alpha_2\right)^2 V_{002} \\ -\left(k+2\alpha_1\right)V_{110}-\left(1-k+2\alpha_2\right)V_{101} \\ +\frac{1}{2}\left(k+2\alpha_1\right)\left(1-k+2\alpha_2\right)V_{011} \end{bmatrix}. \quad (6)$$

The optimum values of α_1 and α_2 are

$$\alpha_{1(opt)} = \frac{\left[2C_y\left(\rho_{yx}-\rho_{xz}\rho_{yz}\right)-kC_x\left(1-\rho_{xz}^2\right)\right]}{2C_x\left(1-\rho_{xz}^2\right)}$$

and

$$\alpha_{2(opt)} = \frac{\left[2C_y\left(\rho_{yx}-\rho_{xz}\rho_{yz}\right)-(1-k)C_z\left(1-\rho_{xz}^2\right)\right]}{2C_x\left(1-\rho_{xz}^2\right)}.$$

It is remarked that for different values of α_1 and α_2 in Eq (3), we can get various exponential ratio-type estimators from the proposed family of estimators $\bar{y}_{G(RSS)}$. Some are given below as:

(i) For $\alpha_1=1, \alpha_2=0$ and $k=0$, we get

$$\bar{y}_{1(RSS)} = \bar{y}_{[rss]}\left(\frac{\bar{X}}{\bar{x}_{(rss)}}\right)\exp\left(\frac{\bar{Z}-\bar{z}_{[rss]}}{\bar{Z}+\bar{z}_{[rss]}}\right) \quad (7)$$

The bias and MSE of $\bar{y}_{1(RSS)}$ are given respectively

$$Bias\left(\bar{y}_{1(RSS)}\right) \cong \bar{Y}\left[V_{020}+\frac{1}{8}V_{002}-V_{110}-\frac{1}{2}V_{101}+\frac{1}{2}V_{011}\right] \quad (8)$$

and

$$MSE\left(\bar{y}_{1(RSS)}\right) \cong \bar{Y}^2\left[V_{020}+V_{020}+\frac{1}{4}V_{002}-2V_{110}-V_{101}+V_{011}\right]. \quad (9)$$

(ii) For $\alpha_1 = 0$, $\alpha_2 = 1$ and $k = 1$, we get

$$\bar{y}_{2(RSS)} = \bar{y}_{[rss]}\left(\frac{\bar{Z}}{\bar{z}_{[rss]}}\right)\exp\left(\frac{\bar{X} - \bar{x}_{(rss)}}{\bar{X} + \bar{x}_{(rss)}}\right) \tag{10}$$

The bias and MSE of $\bar{y}_{2(RSS)}$, are given respectively

$$Bias\left(\bar{y}_{2(RSS)}\right) \cong \bar{Y}\left[\frac{3}{8}V_{020} + V_{002} - \frac{1}{2}V_{110} - \frac{1}{2}V_{101} + V_{011}\right] \tag{11}$$

and

$$MSE\left(\bar{y}_{2(RSS)}\right) \cong \bar{Y}^2\left[V_{200} + \frac{1}{4}V_{020} + V_{002} - V_{110} - 2V_{101} + V_{011}\right]. \tag{12}$$

(iii) For, $\alpha_1 = 1, \alpha_2 = 0$ and $k = 1$, we get

$$\bar{y}_{3(RSS)} = \bar{y}_{[rss]}\left(\frac{\bar{X}}{\bar{x}_{(rss)}}\right)\exp\left(\frac{\bar{X} - \bar{x}_{(rss)}}{\bar{X} + \bar{x}_{(rss)}}\right) \tag{13}$$

The bias and MSE of $\bar{y}_{3(RSS)}$ are given respectively

$$Bias\left(\bar{y}_{3(RSS)}\right) \cong \bar{Y}\left[\frac{15}{8}V_{020} - \frac{3}{2}V_{110}\right] \tag{14}$$

and

$$MSE\left(\bar{y}_{3(RSS)}\right) \cong \bar{Y}^2\left[V_{200} + \frac{9}{4}V_{020} - 3V_{110}\right]. \tag{15}$$

(iv) For $\alpha_1 = 0$, $\alpha_2 = 1$ and $k = 0$, we get

$$\bar{y}_{4(RSS)} = \bar{y}_{[rss]}\left(\frac{\bar{Z}}{\bar{z}_{[rss]}}\right)\exp\left(\frac{\bar{Z} - \bar{z}_{[rss]}}{\bar{Z} + \bar{z}_{[rss]}}\right) \tag{16}$$

The bias and MSE of $\bar{y}_{4(RSS)}$ are given respectively

$$Bias\left(\bar{y}_{4(RSS)}\right) \cong \bar{Y}\left[\frac{13}{8}V_{002} - \frac{3}{2}V_{101}\right] \tag{17}$$

and

$$MSE\left(\bar{y}_{4(RSS)}\right) \cong \bar{Y}^2\left[V_{200} + \frac{9}{4}V_{002} - 3V_{101}\right]. \tag{18}$$

(v) For $k = 1$, Eq. (3) becomes

$$\bar{y}_{5(RSS)} = \bar{y}_{[rss]}\left(\frac{\bar{X}}{\bar{x}_{(rss)}}\right)^{\alpha_1}\left(\frac{\bar{Z}}{\bar{z}_{[rss]}}\right)^{\alpha_2}\exp\left(\frac{\bar{X} - \bar{x}_{(rss)}}{\bar{X} + \bar{x}_{(rss)}}\right) \tag{19}$$

The bias and MSE of $\bar{y}_{5(RSS)}$, are given respectively

$$Bias\left(\bar{y}_{5(RSS)}\right) \approx \bar{Y}\begin{bmatrix} \frac{1}{8}\left(4\alpha_1^2+8\alpha_1+3\right)V_{020}+\frac{1}{2}\alpha_2\left(\alpha_2+1\right)V_{002} \\ -\frac{1}{2}\left(1+2\alpha_1\right)V_{101}-\alpha_2 V_{101}+\frac{1}{2}\alpha_2\left(1+2\alpha_1\right)V_{011} \end{bmatrix}$$

and

$$MSE\left(\bar{y}_{5(RSS)}\right) \cong \bar{Y}^2\begin{bmatrix} V_{200}+\frac{1}{4}\left(1+2\alpha_1\right)^2 V_{020}+\alpha_2^2 V_{002}-\left(1+2\alpha_1\right)V_{110} \\ -2\alpha_2 V_{101}+\alpha_2\left(1+2\alpha_1\right)V_{011} \end{bmatrix}.$$

The optimum values of α_1 and α_2 are

$$\alpha^*_{1(opt)} = \frac{2C_y\left(\rho_{yx}-\rho_{xz}\rho_{yz}\right)-C_x\left(1-\rho_{xz}^2\right)}{2C_x\left(1-\rho_{xz}^2\right)}$$

and

$$\alpha^*_{2(opt)} = \frac{C_y\left(\rho_{yx}-\rho_{xz}\rho_{yz}\right)}{2C_z\left(1-\rho_{xz}^2\right)}.$$

The minimum bias and MSE of $\bar{y}_{5(RSS)}$, are given respectively

$$Bias\left(\bar{y}_{5(RSS)}\right)_{\min} \cong \frac{1}{8}\bar{Y}\left(\gamma C_x^2-W_{(x)}^2\right) \tag{20}$$

and

$$MSE\left(\bar{Y}_{5(RSS)}\right)_{\min} \cong \bar{Y}^2 V_{200}\left[1-\frac{C_y^2}{C_x^2(1-\rho_{xz}^2)}\left(\rho_{yx}-\rho_{xz}\right)^2\right]. \tag{21}$$

(vi) For $k=0$, Eq.(3) becomes

$$\bar{y}_{6(RSS)} = \bar{y}_{[rss]}\left(\frac{\bar{X}}{\bar{x}_{(rss)}}\right)^{\alpha_1}\left(\frac{\bar{Z}}{\bar{z}_{[rss]}}\right)^{\alpha_2}\exp\left(\frac{\bar{Z}-\bar{z}_{[rss]}}{\bar{Z}+\bar{z}_{[rss]}}\right) \tag{22}$$

The bias and MSE of $\bar{y}_{6(RSS)}$, are given respectively

$$Bias(\bar{y}_{6(RSS)}) \cong \bar{Y}\begin{bmatrix} \frac{1}{2}\alpha_1\left(\alpha_1+1\right)V_{020}+\frac{1}{8}\left(4\alpha_2^2+8\alpha_2+1\right)V_{002} \\ -\alpha_1 V_{110}-\frac{1}{2}\left(1+2\alpha_2\right)V_{101}+\frac{1}{2}\alpha_1\left(1+2\alpha_2\right)V_{011} \end{bmatrix}$$

and

$$MSE\left(\bar{y}_{6(RSS)}\right) \cong \bar{Y}^2\begin{bmatrix} V_{200}+\alpha_1^2 V_{020}+\frac{1}{4}\left(1+2\alpha_2\right)^2 V_{002}-2\alpha_1 V_{110} \\ -\left(1+2\alpha_2\right)V_{101}+2\alpha_1\alpha_2 V_{011} \end{bmatrix}.$$

The optimum values of α_1 and α_2

$$\alpha^{**}_{1(opt)} = \frac{C_y\left(\rho_{yx} - \rho_{xz}\rho_{yz}\right)}{2C_x\left(1-\rho^2_{xz}\right)}$$

and

$$\alpha^{**}_{2(opt)} = \frac{2C_y\left(\rho_{yx} - \rho_{xz}\rho_{yz}\right) - C_z\left(1-\rho^2_{xz}\right)}{2C_z\left(1-\rho^2_{xz}\right)}.$$

The minimum bias and MSE of $\bar{y}_{6(RSS)}$, are given respectively

$$Bias\left(\bar{y}_{6(RSS)}\right)_{\min} \cong \frac{1}{8}\bar{Y}\left(\gamma C^2_z - W^2_{[z]}\right) \tag{23}$$

and

$$MSE\left(\bar{Y}_{6(RSS)}\right)_{\min} \cong \bar{Y}^2 V_{200}\left[1 - \frac{C^2_y}{C^2_z(1-\rho^2_{xz})}\left(\rho_{yz} - \rho_{xz}\right)^2\right]. \tag{24}$$

4. EFFICIENCY COMPARISON

We obtain the conditions under which the proposed estimators are more efficient than the usual RSS mean estimator.

(i) Comparison: By Eq.(2) and Eq.(9)

$$MSE\left(\bar{y}_{1(RSS)}\right) < MSE\left(\bar{y}_{(RSS)}\right), \text{ if } \frac{V_{020} + \frac{1}{4}V_{002}}{2V_{110} + V_{101} - V_{011}} < 1$$

(ii) Comparison: By Eq.(2) and Eq.(12)

$$MSE\left(\bar{y}_{2(RSS)}\right) < MSE\left(\bar{y}_{(RSS)}\right), \text{ if } \frac{\frac{1}{4}V_{020} + V_{002}}{V_{110} + 2V_{101} - V_{011}} < 1$$

(iii) Comparison: By Eq.(2) and Eq.(15)

$$MSE\left(\bar{y}_{3(RSS)}\right) < MSE\left(\bar{y}_{(RSS)}\right) \text{ if } \frac{3C_x}{4C_y\rho_{yx}} < 1$$

(iv) Comparison: By Eq.(2) and Eq.(18)

$$MSE\left(\bar{y}_{4(RSS)}\right) < MSE\left(\bar{y}_{(RSS)}\right), \text{ if } \frac{3C_z}{4C_y\rho_{yz}} < 1$$

(v) Comparison: By Eq.(2) and Eq.(21)

$$MSE\left(\bar{y}_{5(RSS)}\right) < MSE\left(\bar{y}_{(RSS)}\right) \text{ , if } \left(\rho_{yx} - \rho_{xz}\right)^2 > 0$$

(vi) Comparison: By Eq.(2) and Eq.(24)

$$MSE\left(\bar{y}_{6(RSS)}\right) < MSE\left(\bar{y}_{(RSS)}\right), \text{ if } \left(\rho_{yz} - \rho_{xz}\right)^2 > 0$$

5. SIMULATION STUDY

To obtain *MSE*, Relative Bias (*RB*) and Relative Root Mean Square Error (*RRMSE*) of the proposed class of exponential ratio-type estimators, a simulation study is conducted. Ranking is performed on basis of the concomitant variable X. Trivariate random observation $\left(X, Y, Z\right)$ are generated from a trivariate gamma distribution with known population correlation coefficients $\rho_{yx} = 0.90$, $\rho_{yz} = 0.80$ and $\rho_{xz} = 0.70$. Using 20,000 simulations, estimates of *MSE*, *RRMSE* and *RB* for different estimators are computed using ranked set sampling scheme as described in Section 2. Estimators are then compared in the term of *MSE*, *RB*, *RRMSE* and percentage relative efficiency (*PRE*). We have computed the *PRE* of different ratio-in-exponential type estimators of population mean $\left(\bar{Y}\right)$ with respect to usual unbiased estimator $\bar{y}_{(RSS)}$ for different values of m and r. The results are shown in Tables 1, 2, 3 and 4. The findings indicate that with increase in sample size, *MSEs*, *RB*, *RRMSEs* decrease which are expected results. We used the following expressions to obtain the *MSEs*, *RB*, *RRMSEs* and *PREs*:

$$RB\left(\bar{y}_{G(RSS)}\right) = \frac{1}{\bar{\bar{Y}}}\left[\frac{1}{20000}\sum_{i=1}^{20000}\left(\bar{y}_{G(RSS)i} - \bar{Y}\right)^2\right], \quad G = 1, 2, \ldots, 6$$

$$RRMSE\left(\bar{y}_{G(RSS)}\right) = \frac{1}{\bar{\bar{Y}}}\left[\frac{1}{20000}\sum_{i=1}^{20000}\left(\bar{y}_{G(RSS)i} - \bar{Y}\right)^2\right]^{\frac{1}{2}},$$

$$MSE\left(\bar{y}_{G(RSS)}\right) = \frac{1}{20000}\sum_{i=1}^{20000}\left(\bar{y}_{G(RSS)i} - \bar{Y}\right)^2 \text{ and}$$

$$PRE = \frac{MSE\left(\bar{y}_{(RSS)}\right)}{MSE\left(\bar{y}_{G(RSS)}\right)} \times 100, \quad G = 1, 2, \ldots, 6.$$

6. CONCLUSION

In Table 1 and 3, we see that the proposed ratio-in-exponential type estimators $\bar{y}_{G(RSS)}$, have less *MSE* and *RRMSE* values as compared to $\bar{y}_{G(RSS)}$. Also, *MSE* and *RRMSE* decrease with increase in the sample size. The simulation result of Table 4 indicate that the proposed estimators have reasonable biases, since the values of percentage *RB* are all less than 2% in absolute terms. Also, the value of percentage *RB* decreases with increase in sample size $n = mr$. So, we conclude that the proposed ratio-in-exponential type estimators are preferable than the usual *RSS* mean estimators under *RSS* scheme.

6. ACKNOWLEDGEMENT

The authors wish to thank the cheif editor and the referees for their valuable suggestions regarding improvenment of the paper.

REFERENCES

1. Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1992). *A First Course in Order Statistics*. Vol. 54, Siam.
2. Chambers, R.L. and Dunstan, R. (1986). Estimating the distribution function from survey data. *Biometrika*, 73, 597-604.
3. Harms, T. and Duchesne, P. (2006). On calibration estimation for quantile. *Survey Methodology*, 32, 37-52.
4. Khan. L. and Shabbir, J. (2015). A class of Hartley-Ross type unbiased estimators for population mean using ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, DOI:10.15672/HJMS. 20156210579.
5. Khan, L. and Shabbir, J. (2016). Hartley-Ross type unbiased estimators for population mean using ranked set sampling and stratified ranked set sampling. *North Carolina Journal of Mathematics and Statistics*, 2, 10-22.
6. Khan.L. Shabbir, J. and Gupta, S. (2016). Unbiased ratio estimators of the mean in stratified ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, DOI: 10.15672/HJMS. 20156210579.
7. Mclntyre, G. (1952). A method for unbiased selective sampling using ranked sets. *Crop and Pasture Science*, 3(4), 385-390.
8. Munoz, J.F. and Rueda, M. (2009). New imputation method for missing data using quantiles. *Journal of Computational and Applied Mathematics*, 232, 305-317.
9. Rao, J.N.K., Kovar, J.G. and Mantel, H.J. (1990). On estimating distribution function and quantiles from survey data using auxiliary information. *Biometrika*, 77, 365-375.
10. Samawi, H.M. and Muttlak, H.A. (1996). Estimation of ratio using ranked set sampling. *Biometrical Journal*, 38(6), 753-764.
11. Silva, P.L.D. and Skinner, C.J. (1995). Estimating distribution function with auxiliary information using post stratification. *Journal of Official Statistics*, 11, 277-294.
12. Singh, H.P., Tailor, R. and Singh, S. (2014). General procedure for estimating the populations mean using ranked set sampling. *Journal of Statistical Computation and Simulation*, 84(5), 931-945.
13. Stokes, S.L. (1977). Ranked set sampling with concomitant variables. *Communication in Statistics: Theory and Methods*, 6(12), 207-1211.
14. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, 20(1), 1-31.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 67-76

IMPULSE RESPONSE FUNCTION ANALYSIS: AN APPLICATION TO MACROECONOMIC DATA OF PAKISTAN

Fariha Shafiq, **Muhammad Arif** and **Muhammad Yaseen**
Department of Mathematics & Statistics, University of Agriculture, Faisalabad, Pakistan
Email: fari_198@live.com; arif_uaf@hotmail.com; myaseen208@gmail.com

ABSTRACT

Impulse response function (IRF) is increasingly common in financial markets; it is significantly related to policy changes of Government or the firm. An IRF of any dynamic model is its reaction to some external fluctuations emerged with brief input signals over time. It is a way to examine and evaluate the impact of external shocks. To make a comprehensive view of Pakistan's macroeconomy, six typical macroeconomic indicators focused for analysis. In this study the impact of one variable to all other variables with respect to external shocks has valued description with the help of VAR model. By the help of IRF it can be concluded that exchange rate, Import rate, export rate and money supply in Pakistan have strong relation between themand show a considerable reaction in future for the impulses occur in each other so the model of these four variables have sound predictive validity and used for better forecast.

KEY WORDS

Impulse response function, VAR model.

1. INTRODUCTION

The concept of simultaneous equations introduced by Sims (1980) in a very elegant and flexible manner called vector autoregressive (VAR) model in empirical statistics, in which all variables are treated like endogenous variables no exogenous variables included in the model. Although the VAR model have very attractive features but sometimes it found to be difficult to deal with the model when it suffering from external shocks in one or more variables. Because all variables are related to each other and have the interaction between all of them so it was difficult to assume them as independent. It is often interesting to know the reaction of all macroeconomic variables when an external shock occurs to any variable. All the economist, statistician and policy makers want to know that how persistent will be the effect of any external shock on the Nation's economy. In financial markets the IRF have the importance like a back bone. Gujrati and Porter (2008)stated that the IRF find out the reaction of the dependent variable in the VAR dynamic system for those shocks that are present in the error terms. He explained that the IRF is used to interpret the VAR model when there are several lags in each equation of the model and also have the alternate signs. The main area of VAR analysis centered on the calculations of IRF which evaluate the impact of shocks on economic system over the time. It is widely accepted that error terms are correlated to each other, there for residual matrix of error terms in VAR model is not the diagonal matrix. Then to overcome this problem Swanson and Granger (1997) found that the residual matrix converted in to orthognalize matrix by using Cholesky decomposition in Vector Moving average representation.

It is claimed by the Jorda (2004) that the standard method of getting Impulse response estimates from VAR model by VMA representation is not robust to the misspecifications, he suggested a natural alternate to overcome all the problems called Local linear projection (LP) method which considered to be good for its robustness to misspecifications, estimates can easily obtain by ordinary least square methods, good for nonlinear method as well as for linear method. But Meier (2005) criticized that the LP method not much better than ordinary VAR method because it gives the greater bias and greater variances so the Standard VAR method considered to be good for all type of data sets at short to long horizon over the time to get the Impulse response estimates.

For more precise forecasting Pesavento and Rossi (2006) gave the idea of Impulse response Confidence intervals around the point estimates. It is a more defined way to forecast about the macroeconomic variables in future that how will all other variables react when any variable from them suffering from external shocks like: flood, earthquake, drought, and extreme temperatures. The graphical representation of Impulse Response Analysis is easy and familiar for user to understand the changes that will occur in future horizon as compare to numerical results.

There are many tools available to check the economic growth in Pakistan but some are more important and basic by which we can take an idea of development condition of our country that is it high or low. Here the purpose of research is to estimate the effect of response of economic indicators when some unusual changes suddenly occur in the country. Normal ups and down are the part of the life but the problem occurs when a great wave of external shocks or natural disasters comes and destroy everything of economy just like floods droughts and earthquakes. During the visit of Pakistan 2010 Ban Ki-moon the U.N Secretary General said that he never saw a flood disaster just like seeing in Pakistan, and it was actually a worst disaster in the history of Pakistan. 20 million people effected from this flood that is the worse than the earth quack disaster in North West areas of Pakistan in 2005. So when this type of shocks occurs in the country it destroy all the current situation of economy and this is very critical situation for the analyst, economist and policy makers that how to estimate it and then what should be the future predictions.

2. METHODOLOGY

Monthly data from 1971 to 2015 collected from State Bank of Pakistan and Pakistan bureau of statistics. Six macroeconomic variables are ; Exchange Rate (ER), Consumer Price Index (CPI), Money Supply(MS), Wholesale Price Index (WPI), Import rate and export rate.

Vector auto regressions (VAR) were introduced by Sims (1980) who demonstrated that VAR model is the best representation of multi equation model where all the variables are related to each other and treated like the endogenous; Vector auto regressive model is the extension of univariate autoregressive model. It is one of the successful models to deal with the multivariate time series with serially correlated nature data. VAR model is suitable for large data sets. VAR model does not require any strong identification or assumptions. A VAR model has a set of variables and their lag values it can easily be estimated by the OLS (Ordinary Least Square) method.

Vector autoregressive model (VAR) of order p is written as

$$\boldsymbol{y_t} = \boldsymbol{c} + \boldsymbol{\Phi_1 y_{t-1}} + \cdots + \boldsymbol{\Phi_p y_{t-p}} + \boldsymbol{\varepsilon_t} \quad \cdots \tag{1}$$

where

$\boldsymbol{y_t} = (y_{1t} \; \dots \; y_{Nt})'$

$\boldsymbol{\epsilon_t} = (\epsilon_{1t} \; \dots \; \epsilon_{Nt})'$ are (N×1) and $\boldsymbol{\Phi}_i$ are(N×N) $\forall_i$

Vector generalization of a scalar autoregression is

$$E\,(\boldsymbol{\epsilon_t}) = \boldsymbol{0}) \quad \boldsymbol{E}(\boldsymbol{\epsilon_t}\;\boldsymbol{\epsilon_t'}) = \begin{cases} \boldsymbol{\Omega}, & t = t \\ \boldsymbol{0}, & t \neq t \end{cases}$$

Ω is a positive definite matrix, explained by Wang (2008). For estimating the IRF from VAR model a Vector moving average representation is necessary because shocks that are occur in different variables are assumed to be independent only when shock occur in one variable at a time, so a decomposition take place in VMA representation which is a well-known choleskey decomposition.

VMA representation is

$$\boldsymbol{y_t} = \textstyle\sum_{\boldsymbol{i=0}}^{\infty} \boldsymbol{\Psi_i \epsilon_{t-i}} \quad \cdots \tag{2}$$

Then the Impulse Response Function

$$\boldsymbol{\Psi_{n\{i,j\}}} = \frac{\boldsymbol{\partial y_{it+n}}}{\boldsymbol{\partial \varepsilon_{jt}}} \cdots \tag{3}$$

Describe by Lu and Xin (2010).

The response of $\boldsymbol{y_{i,t+n}}$ to a one-time impulse in $\boldsymbol{\varepsilon_{j,t}}$ with all other variables dated t or earlier held constant.In that case, setting all other errors to zero may provide a misleading picture of the actual dynamic relationships between the variables. So to overcome this problem Choleskey decomposition has been used.

$$\boldsymbol{y_t} = \boldsymbol{\varepsilon_t \Psi_1} + \boldsymbol{\varepsilon_{t-1} \Psi_2 \varepsilon_{t-2}} + \cdots \tag{4}$$

$\boldsymbol{\Omega} = \mathbf{PP'}$Cholesky Decomposition

$$\boldsymbol{y_t} = \boldsymbol{\Theta_0 w_t} + \boldsymbol{\Theta_1 w_{t-1}} + \boldsymbol{\Theta_2 w_{t-2}} + \cdots \tag{5}$$

With $\boldsymbol{\Theta_i} = \boldsymbol{\Psi_i P}$, $\boldsymbol{w_t} = \boldsymbol{P^{-1} \varepsilon_t}$

The change in one component of w_t has no effect on the other components. By choleskey decomposition all error terms become orthognalize. The model can again converted from VMA representation to VAR representation by an iterative method.

3. RESULTS

Consumer price index, Exchange rate, Money supply, whole sale price index, import rate and export rate are the important components of any country and assumed like wheels of the economic bus. The study about IRF is actually designed to check the effect of natural disasters on country ‘economy that how all the variables react when any external shock occur in one or more variables. This chapter presents the estimation of all the desired tests necessary for drawing any results about the behavior of the variables towards the shocks in their error terms, first of all the past and present behavior of economical time series observed, then all variables set in Vector auto regressive model in which lagged values involved, and the last step contained impulse response analysis to check the deep insight of causality and impact of external shocks in error terms.

Behavior of Time Series Variables

Test each of the time-series to determine their order of integration. All variables presented in log form for compression the data and for the removal of autocorrelation.

Augmented Dickey Fuller (ADF) Unit Root test Results for all variables at 0 differences:

Null hypothesis: variable has a unit root**:**

Table 1
Augmented Dickey Fuller (ADF) Unit Root test Results at 0 Differences

Variable	**P value**	Non-stationary
LCPI	0.4314	Non-stationary
LWPI	0.4371	Non-stationary
LEXP	0.5705	Non-stationary
LIMP	0.6129	Non-stationary
LER	1.0000	Non-stationary
LMS	1.0000	Non-stationary

where

LCPI = log of Consumer Price Index
LWPI = log of Whole Sale price Index
LEXP = log of Export Rate
Limp = log of Import Rate
LER = log of Exchange Rate
LMS = log of Money Supply

Here the P values are greater than the significance values so that null hypothesis cannot be rejected. It can be concluded that these variables have unit roots (non-stationary behavior).

Augmented Dickey Fuller (ADF) Unit Root test Results at 1st difference

Null hypothesis: D (variable) has a unit root

Table 2
Augmented Dickey Fuller (ADF) Unit Root test Results at 1st Difference.

Variable	P value	Conclusion
LCPI	0.0000	Stationary
LWPI	0.0000	Stationary
LEXP	0.0000	Stationary
LIMP	0.0000	Stationary
LER	0.0000	Stationary
LMS	0.0000	Stationary

Here the P values are less than the significance values so that null hypothesis can be rejected. It can be concluded that these variables have no unit roots (stationary behavior).

VAR Model

From above discussion it is observed that m=1 shows that whole group of time series variables have *maximum* integration order 1 .At this time, set a six equation VAR model in the levels (for (1 2) interval)of the data, without considering that what was the integration order of the all six time series. Most significantly, difference the data is not the requirement for this step, no matter what is originate at Step.

Lag Length Criteria

Now decide the suitable lag length, might be p, by using the standard methods. Choice of p base on the common information criteria, such as AIC, SIC. According to Lvanov and Kilian (2005) Akaike information criteria is best for the lag selection of monthly data for all type of reduced form or unreduced form VAR models.

Table 3
Lag Length Criteria

AIC	SC	HQ
86.83638	86.88620	86.85591
67.28660	67.63531	67.42332
66.94194	67.58956*	67.19585
66.77316	67.71967	67.14425
66.78523	68.03064	67.27351
66.70681	68.25112	67.31228
66.23875	68.08196	66.96141
66.02292*	68.16503	66.86277*
66.02962	68.47063	66.98665
66.05086	68.79078	67.12509
66.12762	69.16643	67.31904
66.19114	69.52885	67.49974

The several information criteria recommend that we must have a lag length of 7 for each time series variable. Particularly if individual AIC is taking as lag criteria then for the 7th lag VAR model have the smallest value and theoretically it has been proved that if the model has a lesser value for all lag length criteria than it is a good fit.

To check that the expected VAR model is dynamically established and well identified

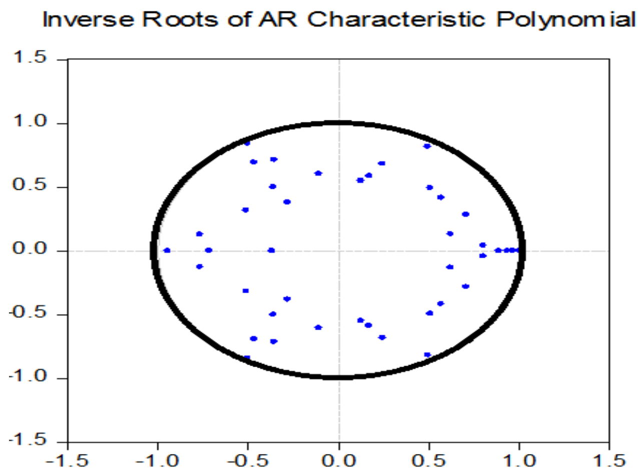


Fig. 1: Inverse Roots of AR Characteristic Polynomial

The expected or estimated VAR is stationary/ stable if all roots are inside the circle unit and have mod less than 1. Roots sometimes lie at broader line of the circles and it is still stationary or stable on the other hand if the VAR model is not stable, IRFs are not valid.

Impulse Response Analysis

To check the interdependence between the all the six variables impulse response analysis is under taken which describe the complete story of relationship, dependency and causality. It is possible to identify which variables we need to offer an impulse, and which variables we need to get the response from. By using multiple graphs with confidence intervals by using Monte Carlo approach for 20 periods in future with choleskey decomposition. Here is the first graph show the response of CPI, the second graph show the response of exchange rate and similarly the third, fourth, fifth and sixth show the response of export, import, money supply and WPI for innovations in all variables.

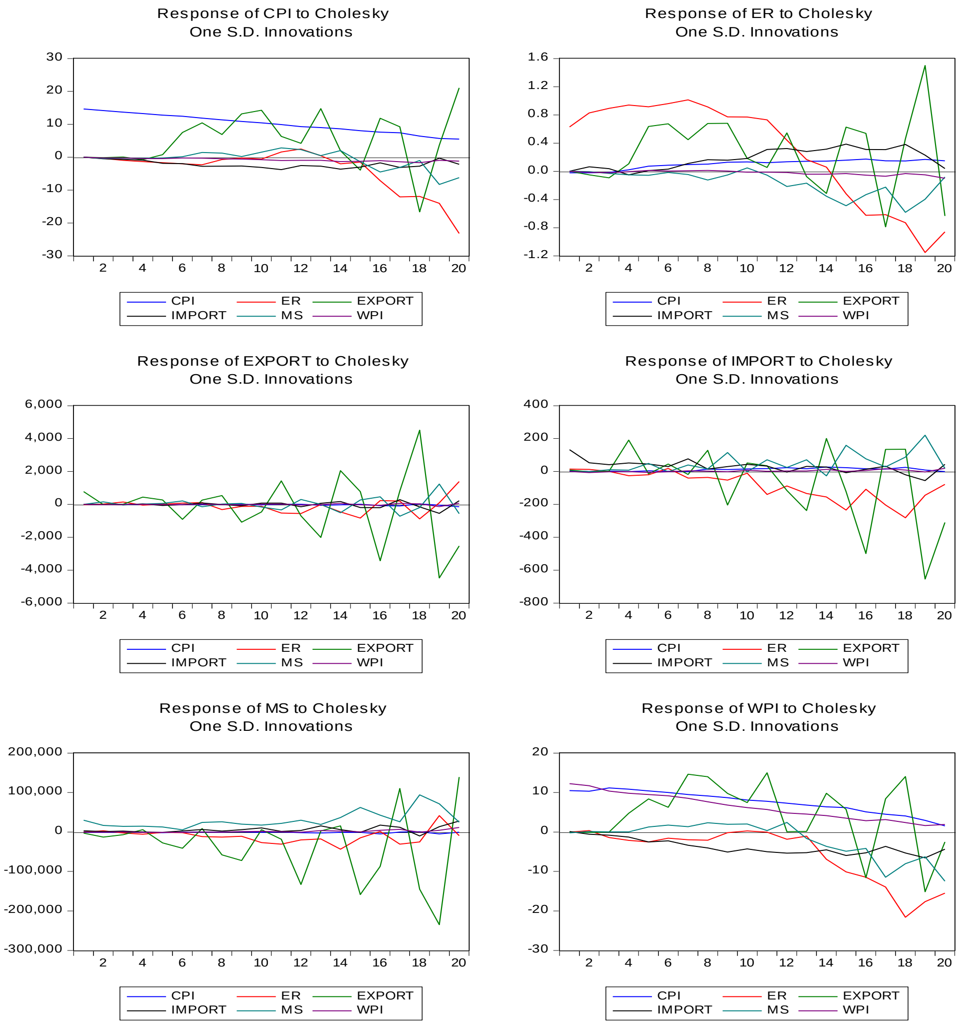


Fig 2: Response of Variables to Choleskey One SD. Innovations

4. COMMENTS AND CONCLUSION

From above results it can be concluded that when there is one standard deviation innovation or external shock occur in the error term of CPI then response of EXPORT shows a slight increase at 1st month but at 2nd month there will no change occur in exchange rate and after that the effect of CPI impulse on exchange rate will remain almost insignificant for all 20 months. When one standard deviation impulse occur in exchange rate then at the start it effect negatively and very small but as time passes this negative effect become large. Response of Export to impulse in export has a great effect from the first month but for fifteen to eighteenth month the effect will goes insignificant after that it will start to be negative. Response of export to impulses in Import, Money supply and WPI behave alike in start but after that export response very different for all three variables, Money supply have positive effect imports have mix effect but WPI have insignificant effect on Export for the whole time period.

Response of imports shows small increase at the very 1^{st} period but after 7 months there will be a rise occur in import rates of Pakistan and remain the positive till the end. If the shocks occur in exchange rate then it badly and negatively affects the imports of Pakistan. Response of imports for imports and exports shocks almost same, positive at the start and negative at the end. Money supply starts from zero but have a positive attitude but WPI has no significant effect for all time period.

Except of exchange rate all other variables have zero effect on exchange rate at the start when any external shock occur in CPI, export, import, money supply and WPI. Response of exchange rate shows no effect of CPI impulse at first four months but from the fifth month the graph shows a positive but very small change in exchange rate of Pakistan.

When any external shock occur in CPI it will significantly affect only the CPI and remain the positive for the whole time period all other variables have no considerable effects on CPI when any natural disaster occur in them.

WPI shows the same behavior for the external shock occurs in CPI and WPI that will remain positive for all the 20 months all other variables start from zero and after that show a falling behavior.

Money supply shows an insignificant response for the seven months but after that there seems a gradually fall in MS due to change occur in CPI. Same attitude Money supply will show for the WPI for the one standard deviation impulse in exchange rate, export and import rate will greatly effect from start to end and shoe a negative trend at the end. Response of money supply for the shock in money supply has a positive and increasing effect. From all the results and discussions it can be concluded that consumer price index and whole sale price index have a strong relationship between them and significantly affected when impulse occur in one of them but have no considerable reaction for the all other variables. On the other hand exchange rate, Import rate, export rate and money supply in Pakistan have strong relation between them and show a considerable reaction in future for the impulses occur in each other so theses four variables have sound predictive validity and used for better forecast.

5. ACKNOWLEDGEMENT

Bounteous praise for ALMIGHTY ALLAH, the magnificent, the merciful, and all the praises for the HOLY PROPHET MUHAMMAD (P.B.U.H.) for enlightening our conscience of faith in ALLAH, converging all His kindness and mercy upon him. I feel much honor to express my deepest sense of gratitude and indebtedness to my honorable supervisor, Mr. Muhammad Arif and Dr. Muhammad Yaseen from the core of my heart for their dynamic supervision, marvelous guidance, keen interest and encouraging behavior.

REFERENCES

1. Gujarati, D.N. and Porter, D.C. (2008). Vector auto regression (VAR). Some problems with VAR modeling. Basic Econometrics. Douglas Reiner. 5, 806-811.
2. Jorda, O. (2004). Model-free impulse responses. Department of economics. *UC Davis Working Paper*. 1-44.
3. Lu, C. and Xin, Z. (2010). Impulse-response function analysis. An application to macroeconomic data of China. D level Essay in Statistics.
4. Lvanov, V. and Kilian, L. (2005). A practitioner's guide to lag order selection for VAR impulse response analysis. *Studies in Nonlinear Dynamics & Econometrics*, 9, 1-36.
5. Meier, A. (2005). How big is the bias in estimated impulse responses? A horse race between VAR and local projection methods preliminary. *Journal of Econometrics*, 83, 57-88.
6. Pesavento, E. and Rossi, B. (2006). Small sample confidence intervals for multivariate impulse response functions at long horizons. *Journal of Applied Econometrics*, 21, 1135-1155.
7. Sims, C.A. (1980). Macroeconomics and Reality. *Econometrica*, 48, 1-48.
8. Swanson, N.R. and Granger, C.W. (1997). Impulse response functions based on a causal approach to residual orthogonalization in vector auto regressions. *Journal of the American Statistical Association*, 92, 357-367.
9. Wang, P. (2008). Shock persistence and impulse response analysis. *Financial econometrics. Routledge*, 2, 106-116.
10. http://www.pbs.gov.pk/content/price-statistics
11. http://www.sbp.org.pk/ecodata/index2.asp

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 77-84

BLS KNOWLEDGE: A COMPARISON BETWEEN CLINICAL AND BASIC MEDICAL PROFESSIONALS

Anam Rehman[§], Ujala Shujat, Nuha Mahmood, Hafsa Naseem, Maryam Siddiqui, Sohaima Samad, Amim ul Haq, Aatra Mahfooz, Sadaf Athar, Bareera Akhtar and **Ammara Shaikh[1]**
Dow University of Health Sciences, Karachi, Pakistan
[§]Corresponding Author Email: anam.rehman92@gmail.com

ABSTRACT

Introduction: Cardiopulmonary resuscitation (CPR) is performed on sudden cardiac arrest patients. There is significant morbidity and mortality due to SCA and choking. This high incidence raises a question regarding knowledge of BLS in health care professionals. Therefore, we will be assessing knowledge of BLS among clinical and basic medical health care professionals.

Methodology: A cross-sectional study was conducted on the Medical professionals of Jinnah Sindh Medical University (JSMU) and Jinnah Postgraduate Medical Centre (JPMC) to assess their knowledge of BLS, within a period of 10 months. Data was collected from 94 participants using a pretested questionnaire through convenient sampling technique. Adequate knowledge was assessed and compared between medical professionals of JSMU and JPMC. P-value was taken significant if it were <0.05 after applying chi-square test.

Result: The mean age of participants was 33.0 ± 8.66 years. 48% candidates were related to Academics Faculty whereas 52% were Clinical Faculty. Overall 60 (63.8%) responders had "Adequate Knowledge". Out of those, 23 (38.3%) were from JSMU and 37 (61.6%) were from JPMC. There was a significant difference of knowledge of BLS among the institutes (p-value= 0.003). There was also significant difference of adequate knowledge of participants of JSMU who have had previous BLS training and those who haven't (p value =0.05).

Conclusion: BLS knowledge of medical professionals should be frequently revised and updated because they should be competent enough to handle any emergency situation, be it in a hospital or outside of it. Furthermore, this updated knowledge will have a positive effect on their students.

1. INTRODUCTION

According to American Heart Association (AHA), Cardiopulmonary resuscitation (CPR) is performed on patients with sudden cardiac arrest and Heimlich manoeuvre on patients with foreign body obstruction. In United States, annually >66,000 deaths are due to SCA. The survival of the victim decreases by 7 to 10% per minute if immediate and effective CPR is not given and it increases by 2-3 times when given promptly and adequately (Harris, 2008). Choking due foreign body obstruction accounts for more than 500 emergency calls in San Diego County Emergency Department (Soroudi, 2007).

Studies have been conducted to assess knowledge of Basic Life Support (BLS) in health care professionals. One such study in Nepal declared that lack of adequate knowledge of BLS/CPR is found in medical and paramedical professionals (Roshana, 2012). Similar results were seen in a local study, conducted in Rawalpindi and Islamabad, in which only 40.09% of the participants, which included medical professionals and students, knew what CPR stood for and only 10% knew the first step of CPR (Quratulain, 2012). This shows that ample studies have been conducted to assess knowledge of clinical medical professionals and paramedical staff. To the best of our knowledge, however, no previous study has been conducted on non-clinical Academic medical professionals nationally. Furthermore, no comparison of knowledge of BLS among clinical and non-clinical medical professionals has been established. This gap needs to be filled as both medical fields are expected to have equal knowledge of BLS. Therefore, as a preliminary step, this study is aimed to assess and compare the level of knowledge of BLS among clinical and non-clinical healthcare professionals of a basic and applied institute of Karachi, Pakistan.

2. METHODOLOGY

Our objectives for this research are the following:

1. To compare knowledge of BLS between JSMU and JPMC
2. To compare knowledge of BLS between JSMU and JPMC with regards to BLS training.
3. To compare knowledge of BLS between JSMU and JPMC with regards to past BLS practice on victim.

Operational Definition for "Knowledge level/Adequate Knowledge": If a participant scores more than or equal to 50% of the total score, i.e., 12 or more out of 23, his/her knowledge will be considered adequate.

Study Design & Setting: A cross-sectional study was conducted on the Medical professionals of Jinnah Sindh Medical University (JSMU) and Jinnah Postgraduate Medical Centre (JPMC) to assess their knowledge of BLS

Duration of Study: The study was covered within a time duration of 10 months.

Sample Size and sampling technique: Using convenient sampling technique, we calculated our data using WHO software (with the following equation) for sample size determination, with confidence level of 95% and margin of error to be 5%. Our recommended sample size came out 100 (Roohi, 2014). However, data of 94 participants was collected.

$$n = \left(\frac{Z\alpha}{2}\right)^2 \frac{pq}{E^2}$$

Sample Selection:

- **Inclusion Criteria:** Amongst JSMU candidates, the main focus was upon the Teaching Faculty, whereas, within JPMC: Consultants, Registrars, Residents and Postgraduate trainees were included. The data was assembled from Neuro Medicine, Psychiatry, General Medicine, General Surgery, Orthopaedics, Dermatology, E.N.T, Ophthalmology, Chest and Pulmonology, Gynaecology and Obstetrics wards.
- **Exclusion Criteria:** Medical students and other staff members of each department were excluded in JSMU. In JPMC setting house offices, nursing staff and those practitioners working at the cardiology and Accident & Emergency Department were excluded.

Data Collection:

- Study Variables:

  Dependant Variable*:* Knowledge of BLS

  Independent Variable:
  1. **Socio-Demographic:** Age, Institution and Faculty member.
  2. **Exposure to BLS:** Previous BLS training, number of BLS training received, BLS done on a victim, BLS taught to others.

- **Data collection Tool and Method:** A written consent was obtained, and name of the contenders were not included in the questionnaire. A 32-questions based self-structured questionnaire, on AHA guidelines for BLS 2010, was handed over to the participants. It was to be interviewed but only those participants who gave consent and were unable to fill the questionnaire themselves were interviewed, while the rest chose to fill it themselves.

- **Data Analysis Plan:** We will enter data in Microsoft Excel 2013 and then transport it to Statistical Package for Social Sciences (SPSS) version 16.0 for analysis. We will calculate mean and standard deviation for continuous variables (age) and frequency and percentages for categorical variables (Institution, faculty member, previous BLS training, time since last BLS session, BLS training given, BLS attempted on any victim and overall correct answers for CPR and choking). We will also compare knowledge of BLS between JSMU and JPMC; between participants in each group with past BLS training & between participants in each group with past BLS practice on a victim using Chi-square test. P-value <0.05 will be considered statistically significant. If in any contingency table, an expected count falls less than 5, then Fisher-exact test will be applied.

- **Ethical Considerations:** Ethical permission was granted by the Jinnah Sindh Medical University.

3. RESULTS

Table 1 shows the socio-demographic statistics of the participants. The mean age of participants was 33.0 ± 8.66 years. 78 (83%) out of 94 participants were less and or equal to 40 years. Our data included 47 (50%) participants from JSMU and 47 (50%) from JPMC.

Table 2 describes the previous BLS exposure of the participant. When asked for previous training of BLS, 83 (88.3%) of participants responded that they received BLS training. 11 (11.7%) of medical professionals responded that they have never attended any BLS session. Out of those who have previous BLS training, 73 (76.6%) candidates said they have received their training more than 1 year ago. Also, when asked about BLS performed on any victim, 52 (55.3%) replied with affirmation.

Table 1
Socio- Demographic Status of Participant

Variable	**Frequency (%)**
Age	**Mean 33.0 ± 8.66**
≤ 40 yrs	78 (83)
> 40 yrs	16 (17)
Institute	
JSMU	47 (50)
JPMC	**47 (50)**
Faculty Member	
Academics Faculty	45 (48)
Clinical Faculty	49 (52)

Table 2
Previous BLS exposure of Participant

Variable	**Frequency (%)**
BLS training received	
Yes	83 (88.3)
No	11 (11.7)
Completion of Last BLS training	
Less than or one year ago	10 (10.6)
More than one year ago	73 (77.7)
Never did it	1 (11.7)
BLS training given	
Yes	25 (26.6)
No	69 (73.4)
BLS on Victim	
Yes	52 (55.3)
No	42 (44.7)

Participants were investigated with reference to their knowledge to manage emergency situations such as sudden cardiac arrest and choking. With regards to Figure 1, our results confirm that there is a significant difference of knowledge of BLS among the institutes (p-value = 0.003) with doctors of JPMC Centre (Clinical Institute) with 72.3% adequate knowledge than that of JSMU (Academic institute) 44.7 %.

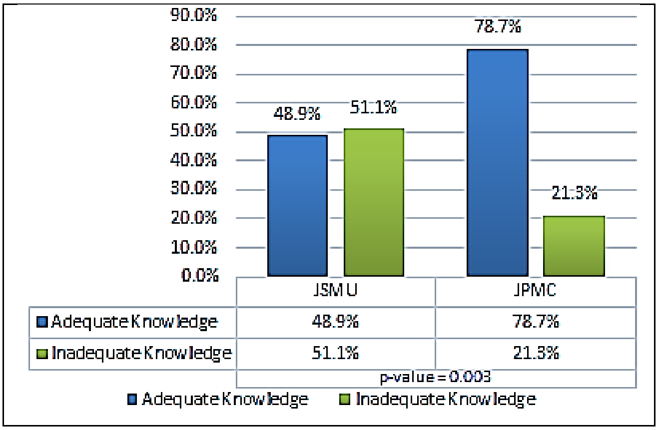


Fig. 1: Comparison of Knowledge of BLS: JSMU vs. JPMC (in percentages)

Fig. 2 shows that there was a significant difference of adequate knowledge, with p-value 0.05, between participants from JSMU who have had previous BLS training and those who haven't. Contrary to which, insignificant difference (p-value 0.11) was found between adequate knowledge of JPMC professionals who've attended BLS training and those who have not. According to Figure 3, BLS done on victim has not proved to cause any significant difference in the knowledge of BLS within JSMU and JPMC, with p-values 0.25 and 0.67 respectively:

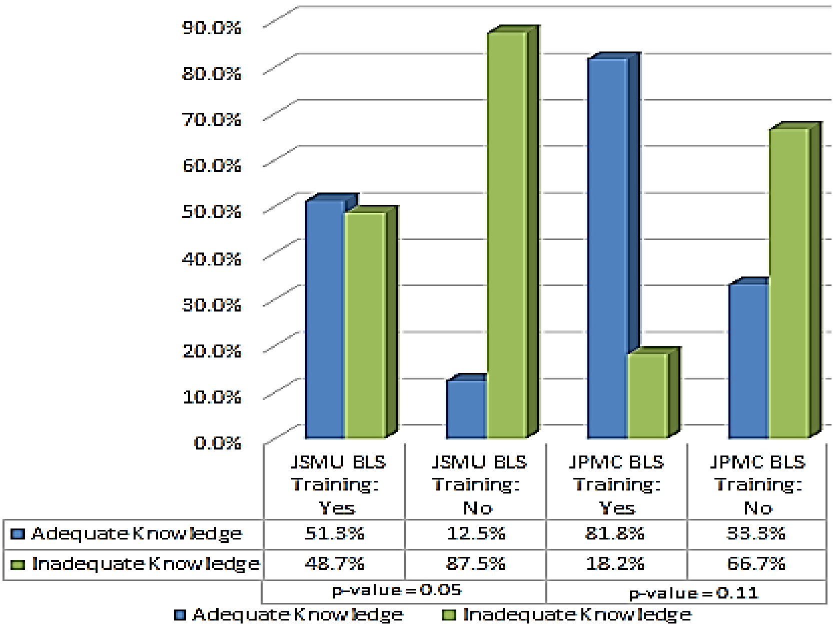
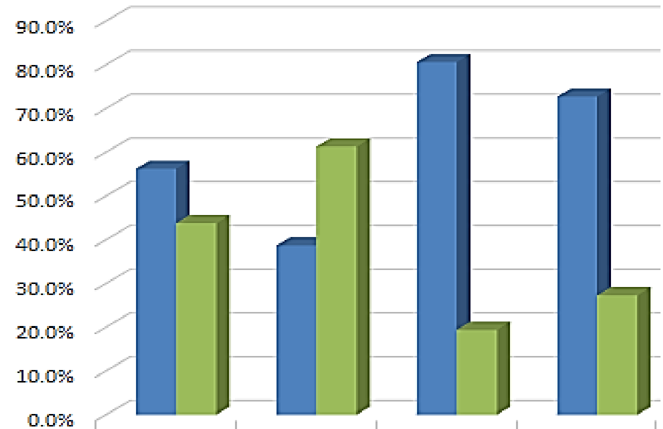


Fig. 2: Comparison of Knowledge with regard to previous BLS training: JSMU vs. JPMC



	JSMU: BLS on Victim: Yes	JSMU: BLS on Victim: No	JPMC: BLS on Victim: Yes	JPMC: BLS on Victim: No
Adequate Knowledge	56.2%	38.7%	80.6%	72.7%
Inadequate Knowledge	43.8%	61.3%	19.4%	27.3%
	p-value = 0.25		p-value = 0.67	

Fig. 3: Comparison of Knowledge with regard to BLS attempt on a victim: JSMU vs. JPMC

4. COMMENTS AND CONCLUSION

Our results confirm that there is a difference of level of knowledge about BLS between the two institutes, with doctors of Jinnah Postgraduate Medical Centre (Clinical Institute) having better knowledge than that of Jinnah Sindh Medical University (Academic institute) with a p-value of 0.003. The result was similar to previous study done at Nepal (Roshana, 2012). This is an important finding of our research as both institutes comprise of qualified doctors with exceptional working experience, and hence are expected to have a good grasp on the knowledge of BLS. In fact, non- clinical practitioners are one of the most cited crowd for primary source regarding information on CPR at the time of emergency (Jones, 2000). Secondly, Pakistan Medical & Dental Council (PMDC) also require doctors to renew their BLS knowledge every 5 years. Hence, doctors, whether practicing in clinics or teaching in medical university are required to be well-equipped with lifesaving skills.

In our study, the overall adequate knowledge came out to be 63.8%, which was close to the study done at Civil Hospital, Karachi (Rubeen, 2013). Since our criteria for “adequate knowledge” was 50% correct answers, our result became inconsistent with study done in Nepal. Although 52% of their participants answered 50% of questions correctly, they did not take this as their criteria for adequate knowledge (Roshana, 2012).

Within JSMU, participants with previous BLS training had significant difference of BLS knowledge compared to professionals with no previous BLS training. This result is coherent with previous studies (Chaudhary 2012; Roohi 2012; Saraç 2010).

As concluded in a research done in Turkey, candidates who acquire CPR skills through traditional and case-based learning prove to retain their knowledge far better than those who self-learn through videos (Saraç 2010). Thus, proper BLS training workshop should be provided to medical professionals teaching at medical universities. Difference in result was insignificant between both factions of BLS trained and untrained personnel in JPMC.

BLS done on a real victim was expected to show significant difference in level of knowledge of BLS, but no such result was seen in both institutes. Each institute was divided into “BLS done on Victim: Yes” and “BLS done on Victim: No”. Answer ‘yes’ to BLS done on victim has not proved to cause any significant difference in the knowledge of BLS within JSMU and JPMC, with p-values 0.25 and 0.67 respectively. This invalidates any claim that clinical professionals have better BLS knowledge due to their practice on real patients.

The sample selection of our study was done on convenient technique, thus, our result cannot be generalized to all other medical institutes. Moreover, it did not include data from private institutions and other government institutions to perhaps compare the knowledge of BLS of health professionals among them. Furthermore, since our study was on the theoretical knowledge of BLS and it did not include the practical knowledge of BLS steps should be taken to analyze that part as well. Lastly, our study only tested for knowledge of choking and CPR, not other components of BLS. Therefore, further studies should be conducted on federal and provincial level to recognize the lack of BLS knowledge among medical personnel and paramedics all over Pakistan.

BLS is a skill that every person of the medical field should be equipped with. Learning it frequently would help the medical personnel to act relevant and fast when he/she faces a situation requiring it. Medical teaching staff should be made aware of the relevance of BLS training and importance of effective and adequate bystander CPR for a victim’s survival. Moreover, compulsory BLS course should be added to their training workshops because they are an important part of our society so they should be competent enough to handle any emergency situation, be it in a hospital or outside of it.

REFERENCES

1. American Heart Association Survey Reveals Americans Lack Confidence in Lifesaving Skills for Common Cardiac Emergency. 2015 [cited 5 December 2015]; http://www.prnewswire.com/news-releases/american-heart-association-survey-reveals-americans-lack-confidence-in-lifesaving-skills-for-common-cardiac-emergency-57311027.html
2. Chaudhary, A., Parikh, H. and Dave, V. (2011). Current Scenerio: Knowledge of Basic Life Support in Medical College. *Natl. J. Med. Res*., 1(2), 80-82.
3. Jones, G., Brewer, K. and Garrison, H. (2000). Public Expectations of Survival Following Cardiopulmonary Resuscitation. *Acad. Emergency Med*., 27(1), 48-53.
4. Zammer, Q., Nadeem, A. and Rizvi, A.H. (2012). Awareness of cardiopulmonary resuscitation in medical-students and doctors in Rawalpindi-Islamabad, Pakistan. *J. Pak. Med. Assoc*., 62(12), 1361-4.

5. Ilyas, R., Shah-e-Zaman, K., Pradhan, N., Feroz, H., Jamal, Z., Amjad, N., Saghir, T., Khwaja, A., Ghauri, S. and Ahmed, Q. (2014). Retention of Knowledge And Skills of Basic Life Support among Health Care Provided Trained in Tertiary Care Hospital. *Pakistan Heart J*., 47(1), 39-45.
6. Roshana, S., Batajoo, K.H., Piryani, R.M. and Sharma, M.W. (2012). Basic life support: knowledge and attitude of medical/paramedical professionals. *World J. Emerg. Med*., 3(2), 141-5.
7. Rubeen, R., Zareen, N., Khan, F., Rattan, S., Moiz, M. and Nousheen, N. (2013). Awareness and Practice of Basic Life Support among Doctor in Civil Hospital Karachi. *Medical Channel*, 19(1), 15-8.
8. Saraç, L. and Ok, A. (2010). The effects of different instructional methods on students' acquisition and retention of cardiopulmonary resuscitation skills. *Resuscitation*, 81(5), 555-561.
9. Soroudi, A., Shipp, H.E., Stepanski, B.M., Ray, L.U., Murrin, P.A., Chan, T.C., Davis D.P. and Vilke, G.M. (2007). Adult Foreign Body Airway Obstruction in the Prehospital Setting. *Prehospital Emergency Care*, 11(1), 25-29.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 85-92

IMPACT OF FACULTY SATISFACTION ON STUDENTS' PERFORMANCE-A CASE STUDY OF FAST-NATIONAL UNIVERSITY

Farrukh Idrees
FAST National University of Computer and Emerging Sciences, Islamabad, Pakistan
Email: farrukh.idrees@nu.edu.pk

ABSTRACT

The research intends to explore the impact of faculty satisfaction on students' performance in higher education. This paper takes a statistical approach by applying descriptive statistics, correlation and ANOVA to assess the impact of satisfaction of 47 faculty members involved in teaching more than 130 courses at undergraduate and post graduate levels in year 2015. The unit of analysis of the study is FAST National University, Islamabad Campus. The paper provides insight for future evaluation of higher education learning environments by linking faculty satisfaction on student's performance. Strong correlation between faculty satisfaction and student performance is found in the faculty of Electrical Engineering. Further research may be done to explore the impact of other factors impacting learning behaviour of students. In addition, more rigorous quantitative techniques e.g. factor analysis, mediating tests, design of experiments etc. may be used to triangulate the findings. Further, sample size may be enhanced to generalize the findings of the study.

KEY WORDS

Faculty Satisfaction, Outcome Based Education, Higher Education, Critical Factors, Engineering School, Student Performance, Class Average Score.

1. INTRODUCTION

Most of the TQM champions claim that TQM helps an institution in improving both tangible and intangible aspects of an organisation like faculty satisfaction, staff satisfaction, better enrolments, lowering academic programs' costs, conformance to requirements and increasing productivity. That in turn, contributes to performance excellence by bringing a focused approach on both soft and hard aspects of the institution (Ramsden 1991, Barnett 1992, Seymour 1992, Lewis and Smith 1994, Trigwell et al. 1999, Parri 2002)

The TQM programs implementation results are more reported to be unsuccessful rather winning fate fates for organizations. Several authors and practitioners have come up with various factors for successful deployment of TQM programs, for example, leadership commitment, faculty satisfaction, faculty development, focus on continual improvement etc. Paradox in quality management success factors in enhancing learning environments is widely discussed in literature of quality in higher education. However, ultimate consensus to the declared factors have remain an enigma. This study intends to explore the impact of faculty satisfaction on students' performance. Faculty, being the internal customers to support the quality management program, are pivotal to the institutions' success for the continuity of quality operations. As Gomes (1996) suggests, the world class leaders work for their people and hence should continuously explore methods and alternatives to provide a conducive learning environment for their faculty and students. Satisfied faculty are more inclined to be motivated and deliver quality for the university.

The present study contends satisfied faculty may play a pivotal role for the successful implementation of QM programs particularly in the context of higher educational institutions and specifically in Pakistani educational sector.

2. BACKGROUND OF FAST SCHOOL OF ENGINEERING

FAST National University of Computer and Emerging Sciences Islamabad which was constituted in year 2000. The university offers BS, MS and PhD in various disciplines. The university has a well-structured quality management cell. Through that cell, various periodic quality management survey questionnaires are launched with the mandate to analyze conformance of practices with the standards e.g. Student course evaluation questionnaire, Faculty course review report, survey of graduating students, research students' progress review, faculty satisfaction survey, alumni survey, employers survey, faculty resume, and teachers' evaluation. Reports against these survey questionnaires are prepared and discussed with relevant bodies for uplifting the quality of education through alignment of practices with objectives of programs. Apart from these surveys, the data generated through various information gathering activities e.g. online admission form filling, semester performance of faculty, CGPA performance of students etc. are also used to generate statistical reports for aiding decision making process.

3. CONCEPTUAL FRAMEWORK

Following conceptual framework can be derived to depict the stance of the research:

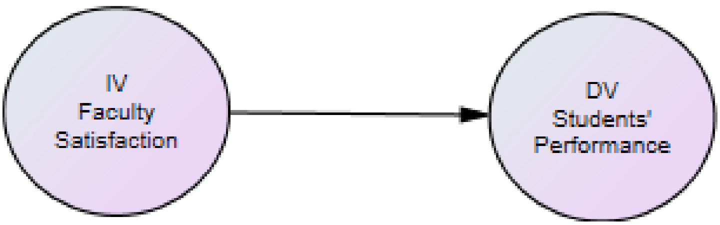


Fig. 1: Conceptual Framework

4. HYPOTHESIS

H1: Satisfied faculty contributes positively to higher performance of students in educational institutions.

5. METHODOLOGY

Table 1
Methodology

What?	How?
Measurement approach	Exam assessment results, entry test questionnaire
Dimensions measured	Faculty Satisfaction, Learning Ability
Population surveyed	90 faculty members
Measurement instrument	On line likert scaled questionnaire
Mode	Anonymous
Sample size	47 out of population of about 90 faculty members
Data analysis software	SPSS
Data analysis technique	Frequency tables/bar charts, Correlation, Regression
Communication of findings to questioned people	Online(charts plus text)
Research Assumption	High class average scores depicts demonstration of higher learning abilities.

6. DEMOGRAPHICS

61 out of 90 faculty members filled the questionnaire. However, 47 questionnaires were sorted for taking in the study. The respondents belong to Electrical engineering, computer science, and management sciences department of FAST-NU. Annexure 1 shows frequency bar charts of each dimension of faculty satisfaction survey.

7. CONFIRMATORY FACTOR ANALYSIS OF FACULTY SATISFACTION CONSTRUCT

Following figure shows the loadings for faculty satisfaction (FS) range from 0.38 to 0.73:

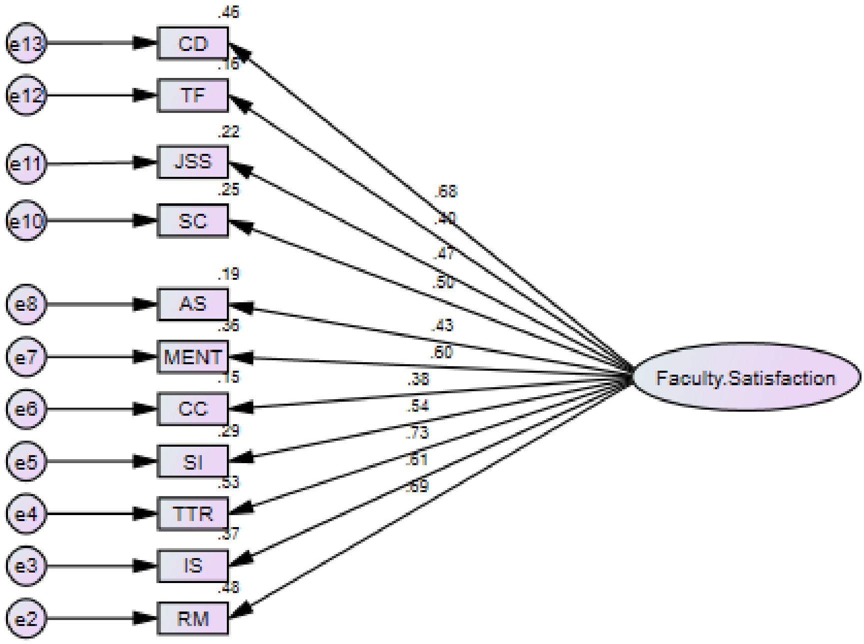


Fig. 2: Loadings for Faculty Satisfaction

Following table summarizes FS model fit indices. FS depicts good model fit with χ^2 statistics of 73.86. The value of CMIN/ DFis 1.679 indicating adequate fit. The P value is 0.03 showing adequate validity of model. The CFI is 0.81, and IFII is 0.72. These scores are ≥ 0.8 and are very close to 1.0 indicating better fit. The TLI is 0.72 indicating adequate fit. The RMSEA value is 0.1 depicting adequate fit.

Table 2
Model Fit Indices for Faculty Satisfaction

Factors	CMIN	P	CMIN/DF	IFI	TLI	CFI	RMSEA
Values	73.86	.003	1.679	0.84	0.72	0.81	0.1

8. CONVERGENT VALIDITY OF FACULTY SATISFACTION SURVEY

Correlation coefficients among faculty satisfaction dimensions indicate which will be used in next chapter of the study. Mostly variables have modest correlation coefficients that is between 0.40-0.69.

Table 3
Correlation among Faculty Satisfaction Variables

	RM	IS	TTR	SI	CC	MENT	AS	SC	JSS	TF	CD	DEK
RM	1											
IS	.596**	1										
TTR	.703**	.525**	1									
SI	.458**	.467**	.513**	1								
CC	.276	.125	.383**	.175	1							
MENT	.409**	.400**	.419**	.331*	.368*	1						
AS	.389**	.433**	.298*	.370*	.143	.57**	1					
SC	.385**	.376**	.301*	.277	.103	.52**	.443**	1				
JSS	.346*	.360*	.308*	.116	.096	.359*	-.084	.446**	1			
TF	.285	.256	.207	.035	.272	.312*	.157	.458**	.355*	1		
CD	.428**	.422**	.493**	.432**	.47**	.48**	.441**	.393**	.387**	.407**	1	
DEK	.482**	.364*	.461**	.297*	.360*	.64**	.339*	.457**	.422**	.263	.477**	1

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

9. RELIABILITY

The Cronbach's Alpha value of Faculty Satisfaction survey was 0.863 indicating good reliability of the scale.

10. DATA ANALYSIS

Following table indicates insignificant ANOVA results of consolidated datasheet covering all departments.

Table 4
ANOVA Table

ANOVA Avg.FS (Binned)					
	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Between Groups	2.078	5	.416	.888	.498

Following table depicts Pearson Correlation between faculty satisfaction and class average score in three departments. Management sciences and computer science departments are found to have insignificant relationships among faculty satisfaction and class average scores. However, Electrical Engineering department is found to have a significant relationship (0.504) between study variables. The learning environment in Electrical Engineering department is already aligned on outcome based educational philosophy. That may be the factor behind higher relationship between faculty satisfaction and class average score because through OBE faculty is continuously being trained to make structured efforts in aligning their lectures, assessment mix, and teaching methodologies to impact students' learning abilities.

Table 5
Department wise Correlation between Faculty Satisfaction and Class Average Score

Factors	Management Sciences	Computer Science	Electrical Engineering
Pearson Correlation	-0.050	-0.048	0.504*
Sig. (2-tailed)	0.876	0.883	0.033
Sample	12	13	18

*. Correlation is significant at the 0.05 level (2-tailed).

10. CONCLUSION AND RECOMMENDATIONS

i) As per the findings of the study, there is a dire need to dig out the causes for low satisfaction on parameters like mix of research, teaching and community service, Salary and compensation package, mentoring available to faculty, Type of teaching/ research, utilization of experience and knowledge of faculty by department etc.

ii) A significant relationship among faculty satisfaction and class average score is found in Electrical Engineering department of the University. This may be because of following rigorous outcome based education management philosophy, as per the requirements of the concerned accreditation council’s requirements.

iii) Further analysis may be conducted by exploring the individual dimensions of the class average scores i.e. assignments, quizzes, sessional 1, sessional 2 results etc. through their correlation with faculty satisfaction construct of the study.

iv) The procedures may be developed to standardize marking criteria to further improve accuracy of the collected data i.e. class average score.

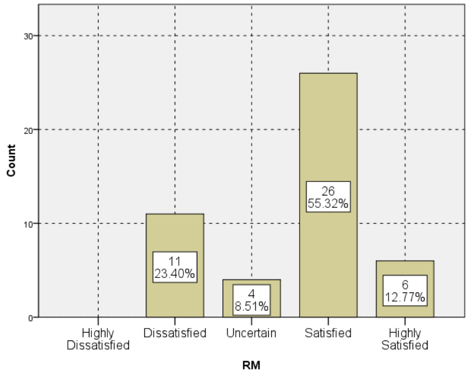
v) A comparative report for multi campuses of FAST National University may be prepared to base the thesis of the impact of faculty satisfaction on student performance.

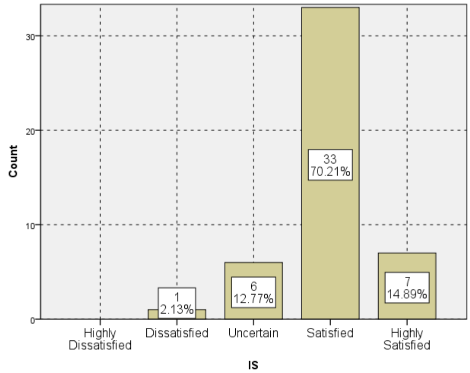
REFERENCES

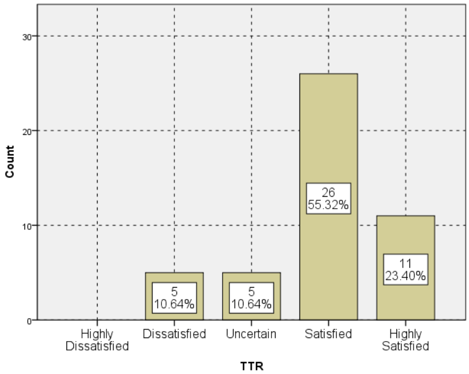
1. Barnett, R. (1992). *Improving Higher Education*: Total Quality Care, ERIC.
2. Clewes, D. (2003). A student-centred conceptual model of service quality in higher education. *Quality in Higher Education*, 9(1), 69-85.
3. Coates, H. (2005). The value of student engagement for higher education quality assurance. *Quality in Higher Education*, 11(1), 25-36.
4. Lewis, R.G. and Smith, D.H. (1994). Total Quality in Higher Education. Total Quality Series, ERIC.
5. Parri, J. (2002). Quality in higher education. Vadyba Management, Citeseer.
6. Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education*,16(2), 129-150.
7. Ruben, B.D. (1995). *Quality in higher education*, Transaction Publishers.
8. Seymour, D.T. (1992). On Q: Causing Quality in Higher Education, ERIC.
9. Trigwell, K., Prosser, M. and Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. Higher education37(1): 57-70.

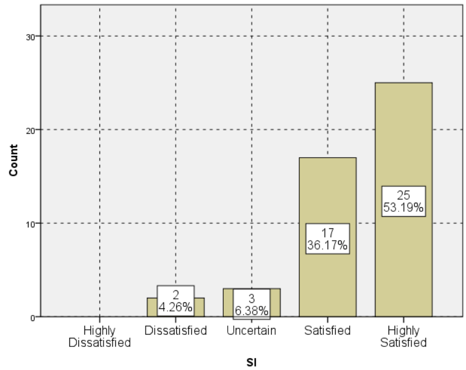
ANNEX 1

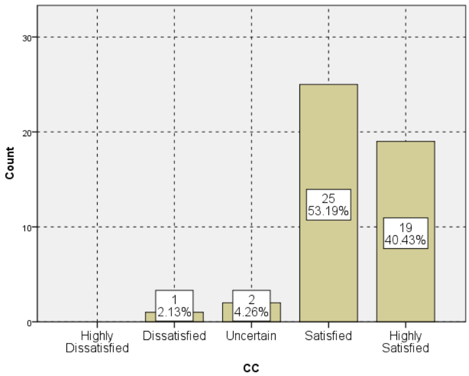
Faculty Satisfaction Survey Results

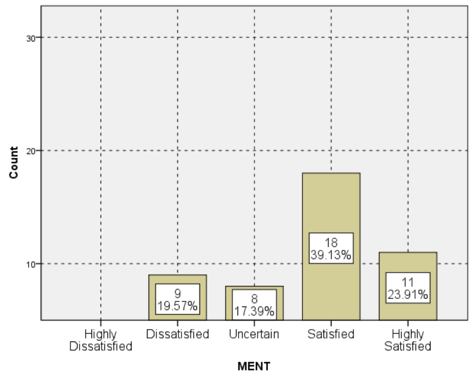


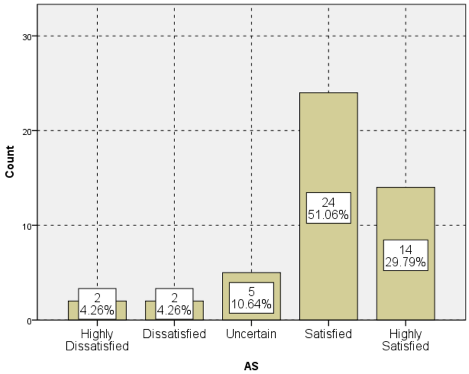


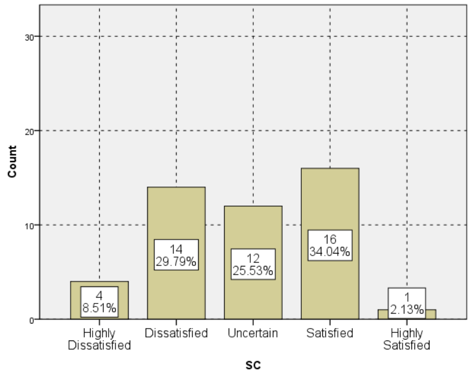


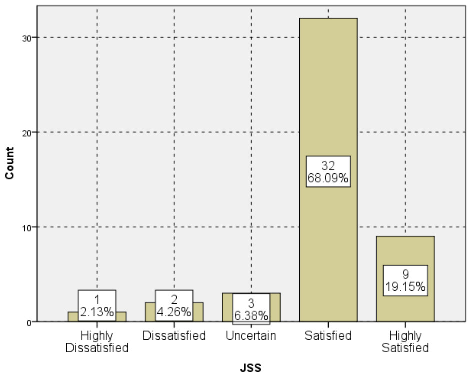


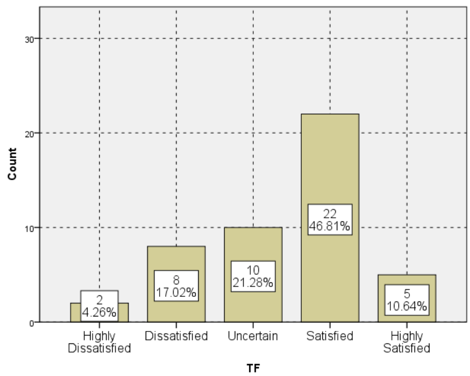


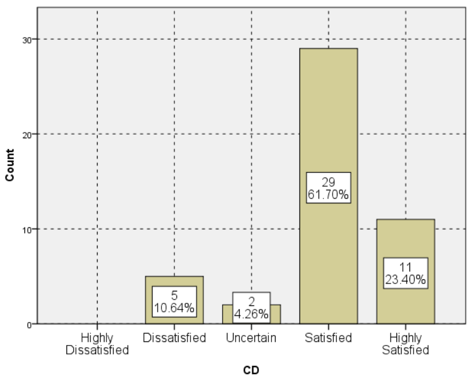


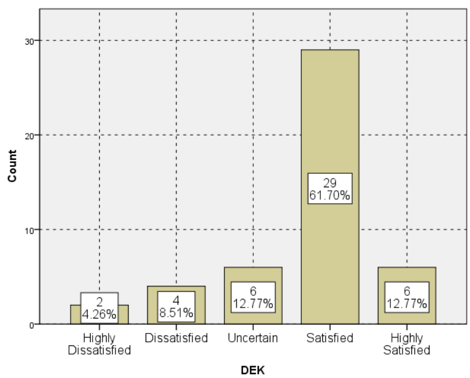

Count
0
10
20
30
2
4.26%
2
4.26%
5
10.64%
24
51.06%
14
29.79%
Highly Dissatisfied
Dissatisfied
Uncertain
Satisfied
Highly Satisfied
AS


Count
0
10
20
30
4
8.51%
14
29.79%
12
25.53%
16
34.04%
1
2.13%
Highly Dissatisfied
Dissatisfied
Uncertain
Satisfied
Highly Satisfied
SC


Count
0
10
20
30
1
2.13%
2
4.26%
3
6.38%
32
68.09%
9
19.15%
Highly Dissatisfied
Dissatisfied
Uncertain
Satisfied
Highly Satisfied
JSS


Count
0
10
20
30
2
4.26%
8
17.02%
10
21.28%
22
46.81%
5
10.64%
Highly Dissatisfied
Dissatisfied
Uncertain
Satisfied
Highly Satisfied
TF


Count
0
10
20
30
5
10.64%
2
4.26%
29
61.70%
11
23.40%
Highly Dissatisfied
Dissatisfied
Uncertain
Satisfied
Highly Satisfied
CD


Count
0
10
20
30
2
4.26%
4
8.51%
6
12.77%
29
61.70%
6
12.77%
Highly Dissatisfied
Dissatisfied
Uncertain
Satisfied
Highly Satisfied
DEK

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 93-102

MANAGING STUDENTS' INTAKE QUALITY- A CASE STUDY OF AN ENGINEERING SCHOOL OF PAKISTAN

Farrukh Idrees
FAST National University of Computer and Emerging Sciences, Islamabad, Pakistan
Email: farrukh.idrees@nu.edu.pk

ABSTRACT

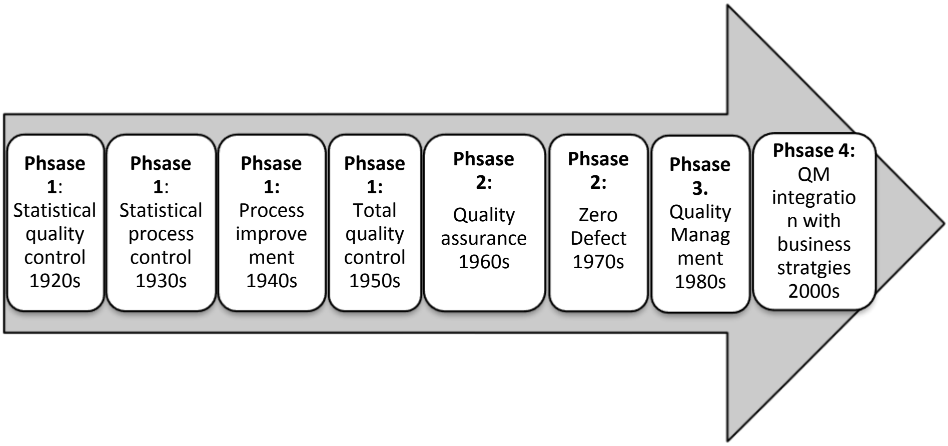
The research intends to explore the impact of students' intake quality on their entry test performance in higher education which is a paradox among educators and researchers. This paper takes a statistical approach by applying descriptive statistics, correlation and regression to assess the impact of intake quality measures in an engineering school through assessing a dataset of 130 students. The unit of analysis of the study is FAST School of Engineering. The paper provides insight for future evaluation of higher education learning environments by linking student intake quality measures with their entry test performance. Further research may be done to explore the impact of geographical factors on students' entry test performance. In addition, more rigorous quantitative techniques e.g. factor analysis, ANOVA, design of experiments etc. may be used to triangulate the findings. Moreover, sample size may be enhanced to generalize the findings of the study.

KEY WORDS

Student Intake Quality Measures, Outcome Based Education, Higher Education, Critical Factors, Engineering School.

1. INTRODUCTION

Quality management reemerged in 20th century in 1920s with the flourishing of statistical quality control concept in manufacturing sector. With development through various phases as described in the following figure, eventually it was integrated in an organizations' strategic mix in the early years of 21st century. Since 1990s after tremendous success of quality management in manufacturing sector it started diffusing in services sector e.g. hospitality, healthcare, education etc. (Ramsden 1991, Barnett 1992, Seymour 1992, Parri 2002, Harvey and Askling 2003, Mansor, Chakraborty et al. 2012, Popa and Bochis 2015). With gradual success in services sector, it is accepted as a robust strategy in higher education sector. There are many similarities of the application of quality and productivity management in higher education and industries e.g. functions in higher education are not so different from counterparts in business or elsewhere in the public sector (Al-Lawatia 2006); higher education is a system for mass production (Al-Lawatia 2006); structure in higher education is also similar to a structure which formerly was common in yesterday's industry: specialized units with limited contacts between them. In higher education institutions, the mandate of quality management is to produce graduates who will be equipped with skills and expertise as required by the market/industry (Dexelar et al. 2000; Jaideep et al. 1997).



Source: Terry (1996)

Figure 1: Development of QM in 21st Century

Currently mostly followed vision of quality approach in higher education is the transformation of existing "input focused" traditional education model in to output based "students focused" education model. The new model favors making students demonstrate that they know and are able to do whatever the required outcomes are. It emphasizes setting clear standards for observable, measurable outcomes through which students' performance can be empirically measured (King and Evans 1991, Spady and Marshall 1991, Spady 1994, Harden 2002, Lozano, Ceulemans et al. 2015, Wick, Galante et al. 2015). Further, output focused education is followed by Universities Ranking agencies (such as QS World Universities Ranking, Financial Times Ranking etc.) and Accreditation bodies (e.g. Accreditation Board for Engineering and Technology.

Johnson and Weinstein (2004) argue that educational institutions flourish because of producing quality graduates who are capable to demonstrate pertinent necessary abilities. Without focusing on students' learning outcomes, educational institutions will not grow and serve their purpose. However, focusing on students' learning outcomes may need to be started right from the beginning of the admission process that is admission tests. The admission tests may be focused enough to reflect the probability of high performing students. One of the most pertinent performance parameters is cumulative grade point average (CGPA) of students, which is calculated at the end of the semester.

2. CONCEPTUAL FRAMEWORK

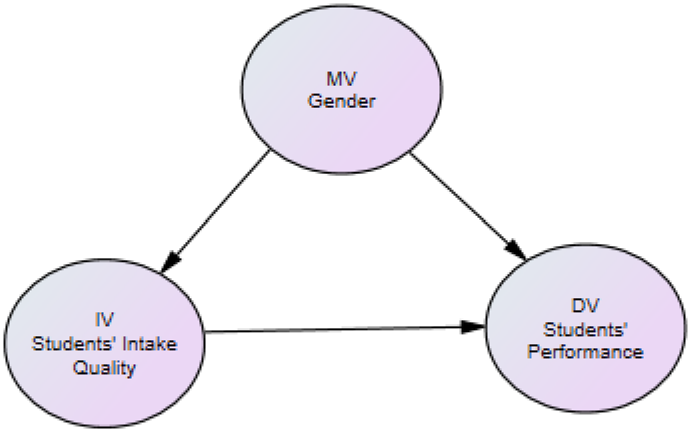


Figure 2: Conceptual Framework

3. DIMENSIONS OF THE STUDY

Table 1
Dimensions of the Study

Dimensions of the Study	
Variables	**Dimensions**
Independent Variable (IV) Students' Intake Quality	• Matric Score • FSc Score • FAST-NU Admission Test Score
Mediating Variable (MV) Gender	• Male/Female
Dependent Variable (DV) Students' Performance	• Commutative Grade Point Average (CGPA)

4. HYPOTHESIS

H1: Student intake Quality has a direct positive impact on students' performance in higher education institutions.

H2: Gender moderates the impact of student intake quality on students' performance in higher education institutions.

5. METHODOLOGY

Table 2
Methodology

	How?
Measurement approach	Exam assessment results, entry test questionnaire
Dimensions measured	Learning ability
Population surveyed	132 students of Batch 2015
Measurement instrument	On line Likert scaled questionnaire
Mode	Anonymous
Sample size	45 out of population of about 65
Data analysis software	SPSS
Data analysis technique	Frequency tables/bar charts, Correlation, Regression
Communication of findings to questioned people	Online(charts plus text)
Research Assumption	Students with high CGPA demonstrate higher learning abilities

6. DATA ANALYSIS

Descriptive statistics of the study variables are given in Annex 1 of the study. Following table indicates that there is significant correlation coefficient between Matric and CGPA, FSc and NET, NET and CGPA.

Table 3: Correlation Table among IV and DV

		Matric.	FSc.	NET	CGPA
Matric.	Correlation	1.00			
	Sig				
FSc.	Correlation	0.10	1.00		
	Sig	0.24			
NET	Correlation	0.01	.319**	1.00	
	Sig	0.92	0.00		
CGPA	Correlation	-.314**	0.14	.275**	1.00
	Sig	0.00	0.12	0.00	
**. Correlation is significant at the 0.01 level (2-tailed).					

Following regression tables indicate that FSc is a significant contributor with B values of 0.32 towards NET.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	44.179	6.258		7.060	.000
Matric	-.002	.005	-.026	-.312	.756
FSc	.024	.006	.326	3.774	.000
Gender	-.247	1.111	-.019	-.223	.824
a. Dependent Variable: NET					

Following regression tables indicate that Matric score and NET score are significant contributors with B values of -0.32 and 0.24 towards CGPA.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1.068	1.485		.719	.473
Matric	-.004	.001	-.326	-4.053	.000
FSc	.001	.001	.092	1.085	.280
NET	.053	.018	.249	2.949	.004
a. Dependent Variable: CGPA					

6.1 Andrew Hyes Test of Moderation

Following tables indicates that gender is playing a moderating role in FSc and NET, NET and CGPA, Matric and CGPA (partially).

Y: NET
X: FSc
Conditional effect of X on Y at values of the moderator(s):

Gender	Effect	se	t	p	LLCI	ULCI
.0000	.0214	.0068	3.1471	.0021	.0079	.0348
1.0000	.0386	.0168	2.2931	.0235	.0053	.0719

Y: NET
X: Matric
Conditional effect of X on Y at values of the moderator(s):

Gender	Effect	se	t	p	LLCI	ULCI
.0000	.0009	.0057	.1665	.8680	-.0103	.0122
1.0000	.0002	.0113	.0147	.9883	-0.223	.0226

Y: CGPA
X: NET
Conditional effect of X on Y at values of the moderator(s):

Gender	Effect	se	t	p	LLCI	ULCI
.0000	.0481	.0185	2.5973	.0105	.0114	.0847
1.0000	.1240	.0539	2.3023	.0229	.0174	.2307

Y: CGPA
X: FSc
Conditional effect of X on Y at values of the moderator(s):

Gender	Effect	se	t	p	LLCI	ULCI
.0000	.0011	.0015	.7130	.4772	-.0019	.0040
1.0000	.0046	.0037	1.2589	.2103	-.0026	.0118

Y: CGPA
X: Matric
Conditional effect of X on Y at values of the moderator(s):

Gender	Effect	se	t	p	LLCI	ULCI
.0000	-.0039	.0011	-3.4472	.0008	-.0061	-.0016
1.0000	-.0025	.0022	-1.1388	.2569	-.0070	.0019

7. DISCUSSION AND CONCLUSION

Quality in higher education has a vital role to play in transforming institutes in to a true place of learning and innovation by conforming to the best practices and standard in the dimensions of enhancing intake quality. There is significant correlation coefficient between Matric and CGPA, FSc and NET, NET and CGPA. Further FSc score is found to be a significant contributor with B values of 0.32 towards NET. Moreover, Matric and NET scores are found to be significant contributors with B values of -0.32 and 0.24 respectively towards CGPA. However, the negative significant correlation of Matric score with CGPA may lead to further investigation. In addition, gender is playing a moderating role in FSc and NET, NET and CGPA, Matric and CGPA (partially).

The findings may be used to take measures in facilitating prospective students to score higher in NET through special preparatory classes. The significant role of gender may be factor in setting quotas as per the gender in available seats for admission.

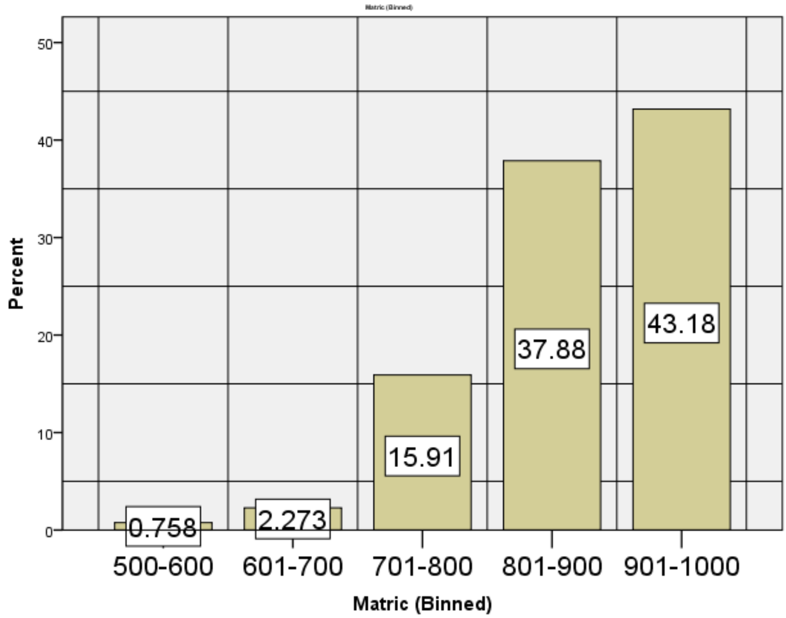
As a part of research limitations, more rigorous quantitative techniques e.g. factor analysis, design of experiments, six sigma etc. may be used to triangulate the findings. Further, sample size may be enhanced to generalize the findings of the study.

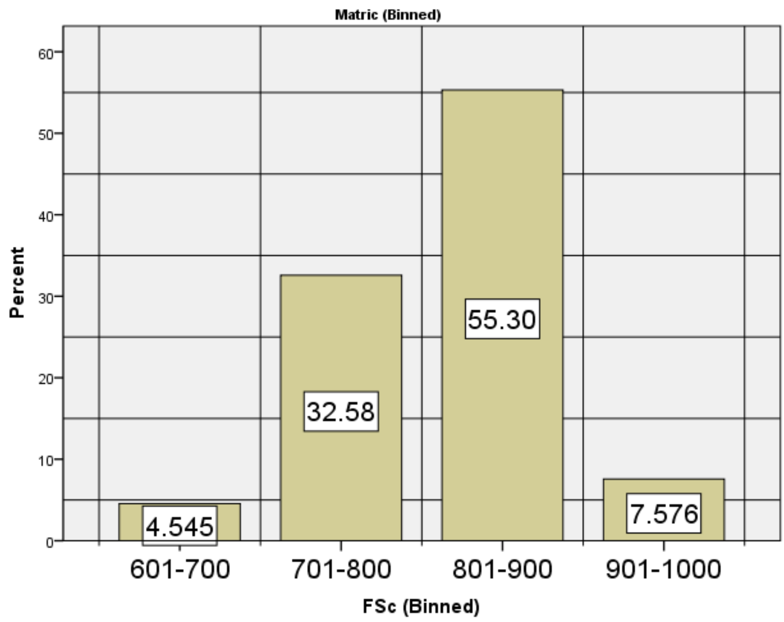
REFERENCES

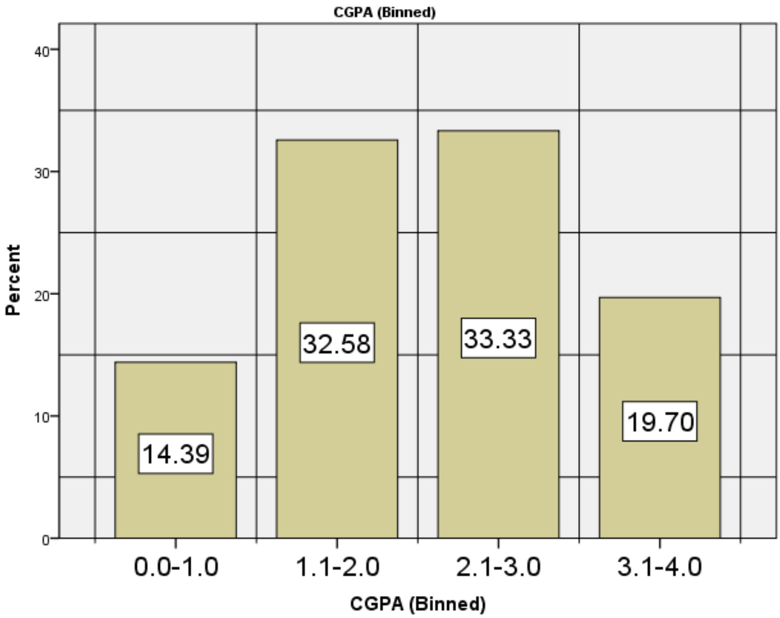
1. Barnett, R. (1992). *Improving Higher Education: Total Quality Care*, ERIC.
2. Harden, R.M. (2002). Developments in outcome-based education. *Medical teacher*, 24(2), 117-120.
3. Harvey, L. and Askling, B. (2003). Quality in higher education. The dialogue between higher education research and practice, *Springer*, 69-83.
4. King, J.A. and Evans, K.M. (1991). Can we achieve outcome-based education. *Educational Leadership*, 49(2), 73-75.
5. Lozano, R., Ceulemans, K. and Seatter, C.S. (2015). Teaching organisational change management for sustainability: designing and delivering a course at the University of Leeds to better prepare future sustainability change agents. *Journal of Cleaner Production*, 106, 205-215.
6. Mansora, N.N.A., Chakrabortyb, A.R., Tay Ke Yinc, T.K. and Mahitapoglud, Z. (2012). Organizational Factors Influencing Performance Management System in Higher Educational Institution of South East Asia. *Procedia - Social and Behavioral Sciences*, 40, 584-590.
7. Parri, J. (2002). *Quality in higher education*. Vadyba Management, Citeseer.
8. Popa, C. and Bochis, L. (2015). Students' Satisfaction Towards Academic Courses in Blended Weekend Classes Program. *Procedia - Social and Behavioral Sciences*, 191, 2198-2202.
9. Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education*, 16(2), 129-150.
10. Seymour, D.T. (1992). *On Q: Causing Quality in Higher Education*, ERIC.
11. Spady, W.G. (1994). *Outcome-Based Education: Critical Issues and Answers*, ERIC.
12. Spady, W.G. and Marshall, K.J. (1991). Beyond Traditional Outcome-Based Education. *Educational Leadership*, 49(2), 67-72.
13. Wick, E.C., Galante, D.J., Hobson, D.B., Benson, A.R., Lee, K.H.K., Berenholtz, S.M., Efron, J.E., Pronovost, P.J. and Wu, C.L. (2015). Organizational Culture Changes Result in Improvement in Patient-Centered Outcomes: Implementation of an Integrated Recovery Pathway for Surgical Patients. *Journal of the American College of Surgeons*, 221(3), 669-677.

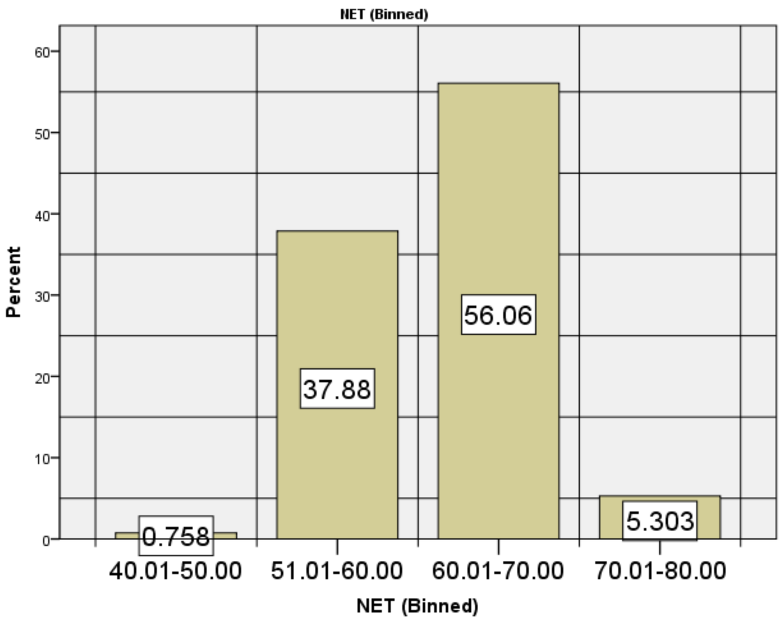
ANNEXURE 1

Descriptive statistics of independent and dependent variables






CGPA (Binned)
Percent
40
30
20
10
0
14.39
32.58
33.33
19.70
0.0-1.0
1.1-2.0
2.1-3.0
3.1-4.0
CGPA (Binned)


NET (Binned)
Percent
60
50
40
30
20
10
0
0.758
37.88
56.06
5.303
40.01-50.00
51.01-60.00
60.01-70.00
70.01-80.00
NET (Binned)

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 103-108

INTERRELATIONSHIP AMONG SPECIES AND ENVIRONMENTAL VARIABLES USING SPATIAL ANALYSIS BY DISTANCE INDICES

Imtiaz Husain and **M. Khalid Siddiqui**
Federal Urdu University of Arts, Science and Technology, Karachi
Email: imtiazbc@gmail.com

ABSTRACT

Spatial Analysis by Distance Indices methods are widely used to measure the interrelationship among the species and environmental variables and their neighboring grid points. Identification of the location and movement pathway enable us to detect association in several scales. Arbitrary coordinates systems were applied to rule out the initial and boundary positions of the species and their associated environmental variables and their distances in various block sizes. We observed each grid unit and measures the distances of one hundred and forty four nearest neighboring grid units. Mean distances among species and environmental variables were calculated in different blocks. The distance between centroids of the grid unit and the counts of species population were found to be 0.644 with maximum distance between sample units is 202.233 units with mean =39.173, variance 6.032 and Index of dispersion was determined as 22.021. The absolute over all flow in the study area is found to be 2.213 at probability ($p<0.003$) with index of aggregation 1.960.

1. INTRODUCTION

Various distance methods were developed and analyzed to study the spatial patterns of the interacting communities and populations in several scale. One of them is Spatial Analysis by Distance Indices (SADIE) which is primarily used to identify the clustering patterns and their indices. This technique first shifts the field data to regularize the information of each grid units to the distance matrix of the sample grid (Perry *et al.*, 1999). Its uses different mathematical and statistical method to determine autocorrelation, which is the ratio between arithmetic mean distance achieving the regular distribution of the selected samples and the distance used to shift to get the regular pattern for the studied data. This new technique provides the information between the two interacting communities and populations and their inter-specific linkage (Perry & Dixon, 2002). Researchers widely applied this method in their respective fields such as root infections (Oerker *et al.*, 2010), soil sciences (Oveisi, Yousefi & Gonzalez-Andujar, 2010), ecosystem managements (Barbeito *et al.*, 2009), and plant ecology (Rodriguez *et al.*, 2009) and plant-soil interaction (Maestre *et al.*, 2003).

Clustering of the species and environmental variables and to calculate their indices are considered as the integral elements of SADIE technique (Perry *et al.*, 1999). Diggle in 1983, explained this mechanism more clearly and stated that analysis of particular species in the scattered field observe the ecological processes and pattern detection techniques have been very informative to determine the types of pattern either are random, clumped (aggregated) or regular (Imtiaz *et. al*., 2015). Spatial analysis has been addressed by researchers from all disciplines and showed their interest to study the interrelationship among several species and their interacting environmental parameters as well as the trend they exhibit.

2. MATERIALS AND METHODS

Spatial Analysis by distances Indices used to determine association among different species and their interacting environmental variables within the field of various blocks. The localized primary and secondary positions of the species and their associated environmental parameters denoted by $\left(x_i, y_i\right)$ while the secondary position is represented as $\left(x_j, y_j\right)$ and their internal distance obtained by:

$$D_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}$$

where D_{ij} denoted by the distance between the neighboring species. However, the average overall distance can be determined by:

$$X_{ij} = \frac{\sum_j D_{ij} P_{ij}}{\sum_j P_{ij}}$$

and concluded as:

$$X_{ij} = \frac{D_i}{x_i - \mu} \qquad [1]$$

Dimensionless clustering indices of all available species, distance can be evaluated as:

$$D = \sum_i D_i$$

While the dimensionless clustering index obtained by:

$$P_i = \frac{X_{io} X}{iX . cX} \qquad [2]$$

where iX and cX are denoted by the average calculated values obtained by the following procedure:

1. First measure initial position of the species variables and their average distances be calculated for each iteration $X_{i,k}$ where $k = 1, 2, 3, ..., n$.
2. The average distance cX for the observed data of the species and associated environmental variables x_i throughout the iteration mentioned as:

$$cX = \sum_{k=1}^{n} \frac{X_{i,k}}{n}$$

3. While species of each block for i^{th} unit (x_i, y_i) is recorded for every iteration, whereas, the average distance of the particular position of the species is calculated in each randomization ${}_{k,i}X$ where $k = 1, 2, 3, ..., n$
4. By the average distance $({}_iX)$ of the species and environmental parameter is elaborated as:

$$ {}_iX = \sum_{k=1}^{n} \frac{|{}_{k,i}X|}{n} $$

where ${}_0X$ consider as the absolute average value of ${}_cX$ for the species as well as environmental data sets throughout the grid units and almost equal to the average value of ${}_iX$ for all blocks (grid units) respectively.

The idea behind the SADIE technique is that a value of an index closer to unity indicates a random pattern. Since an index is determined by the ratio of average value of the sample grid and an observed distance between the species (Table-I).

3. RESULTS AND DISCUSSION

A subarea of designed field plot of 14.4 by 14.4 meter square grid further divided into 1.2 by 1.2 meter square for the study of the interrelationship among species and their environmental variables. The ability to move the species variables from initial position to their neighboring *x* and *y* coordinates blocks (determined as distance), and exists to discriminating patterns. Correlation among the indices agrees that SADIE method provide sufficient information to describe the patch pattern exhibits insight the field data. Approximate distance of the species and their environmental variables is given in Table-I. A Perfect association found between the species and their associated environmental variables around the grid in all blocks (Figure-I and II).

Table I
Distance between the *x* and *y* Coordinate of Species and their Environmental Variables

x-coordinate	y-coordinate	Distances
1	1	4.400
2	1	1.260
1	2	1.260
2	2	2.900
1	5	2.150
2	5	5.600
1	6	5.600
2	6	3.400

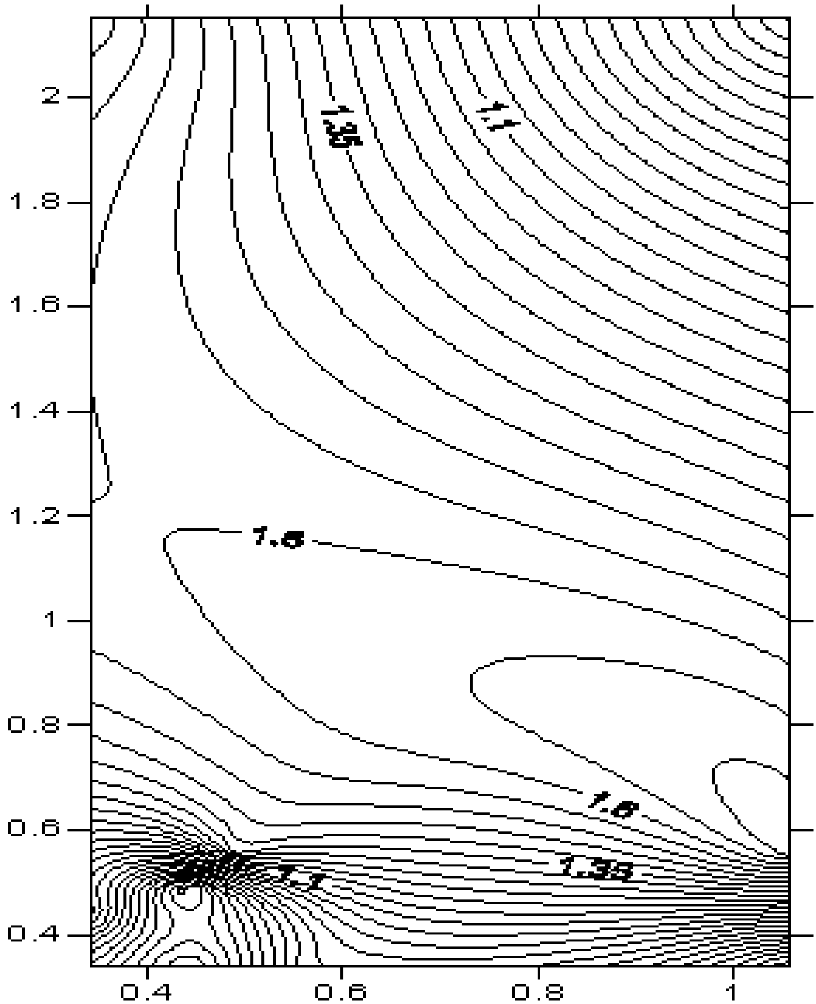


Fig. I: Species Shows Clumping in Lower to Middle Side in the Grid

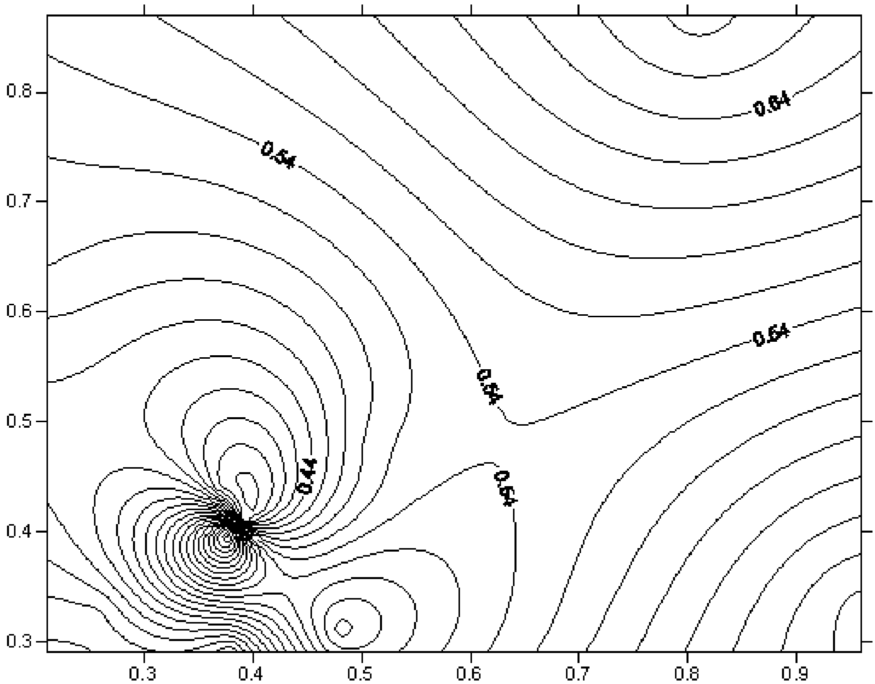


Fig. II: Contour Map of the Environmental Variables Found in Lower Left Side of the Grid at Average Value 0.44

4. COMMENTS AND CONCLUSION

In our study, distance between centroids of the units and counts of species variable found to be 0.644 with maximum distance between sample units is 202.233, mean =39.173, variance = 6.032 and Index of dispersion are determine as 22.021. Whereas the distance between the centroids of the units and counts of associated with environmental variable were found to be 2.947 with maximum distance sample units is 202.232, mean=32.048, variance=19.720 and their index of dispersion is found to be 18.914.

5. ACKNOWLEDGEMENT

This research was supported by Higher Education Commission of Pakistan on the Research Project (Start Up Research Grant Program SRGP#832).

First authoralso acknowledged Higher Education Commission of Pakistan and The Vice Chancellor, Federal Urdu University of Arts, Science and Technology, Karachi Prof. Dr. Suleman D. Muhammad for providing the opportunity to work under the Interim placement for fresh Ph.D. Program (IPFP).

REFERENCES

1. Barbeito, I., Fortin, M.J., Montes, F. and Canellas, I. (2009). Response of pine natural regeneration to small-scale spatial variation in a managed Mediterranean forest. *Applied Vegetation Science*, 12, 488-503.
2. Diggle, P.J. (1983). *Statistical analysis of spatial point patterns*. Academic Press, 27-32.
3. Husain, I., Khan, N.U., and Shaukat, S.S. (2015). Mathematical techniques to study the Spatial Pattern of *Meloidogyne javanica* in relation to soil fungi and soil characteristics in a tomato grown field. *Research Journal of Recent Sciences*, 4(12), 1-9.
4. Maestre, F.T., Cortina, J., Bautista, S., Bellot, J. and Vallejo, R. (2003). Small-scale environmental heterogeneity and spatiotemporal dynamics of seeding establishment in a semiarid degraded ecosystem. *Ecosystems*, 6, 630-643.
5. Oerker, E.C., Meier, A., Dehne, H.W., Sulyob, M., Krska, R. and Steiner, U. (2010). Spatial variability of Fusarium head blight pathogens and associated mycotoxins in wheat crops. *Plant Pathology*, 59, 671-682
6. Oveisi, M., Yousefi, A.R. and Gonzalez-Andujar, J.L. (2010). Spatial distribution and temporal stability of crenate broomrape (*Orobanchecrenata* Forsk) in faba bean (*Viciafaba* L.): a long-term study at two localities. *Crop Protection*, 29, 717-720.
7. Perry, J.N., Winder, L., Holland, J.M. and Alston, R.D. (1999). Red-blue plots for detecting clusters in count data. *Ecological Letters*, 2, 106-113.
8. Perry, J.N. and Dixon, P.M. (2002). A new method to measure spatial association for ecological count data. *Ecoscience*, 9, 133-141.
9. Rodriguez, A., Duran, J., Fernandez-Palacios, J.M. and Gallardo, A. (2009). Wildfire changes the spatial pattern of soil nutrient availability in *Pinuscanariensis* forests. *Annals of Forest Sciences*, 66, 210, 1-7.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 109-116

MEASURING KURTOSIS OF WEIBULL DISTRIBUTION BY VARIOUS APPROACHES

Rahila Hafiz[1] and **Ahmed Zogo Memon**[2]
The Institute of Management Sciences (Pak-AIMS), Lahore, Pakistan
Email: [1]rahila@pakaims.edu.pk
[2]azogo.memon@hotmail.com

ABSTRACT

In statistical literature there exists a controversy about the real purpose of Pearson's measure of Kurtosis. Measures of this characteristic of a distribution have been proposed by a number of authors including Horn (1983), Bonett and Seier (2003), and Zenga (2005). In this paper we investigate the kurtic behavior of Weibull family of distributions by using the approaches suggested by these authors.

1. INTRODUCTION

There has been a controversy in statistical literature over the real purpose of Pearson's measure of peakedness of a probability distribution. Kaplansky (1943) gave four examples of standardized symmetric distributions which when compared with standard normal distribution; show that there is no logical connection between the value of the density of the standardized distribution and the sign of Pearson's coefficient. A common argument is that tail, shoulder and centre being the important aspects exercise their impact. As a consequence a number of its measures have been proposed using the argument. Some of these measures are due to Chissom (1970), Darlington (1970), Horn (1983), Bonett and Seier (2003), Hosking (1990), Blest (2003), and Fiori and Zenga (2005).

Rahila (2011) applies to Weibull distribution various measures of kurtosis and analyses their performance. In this paper we consider Weibull distribution for application of the measures of kurtosis proposed by Horn (1983), Bonett and Seier (2003), and Fiori and Zenga (2005). Remarks are given on the similarities when compared to Pearson's coefficient.

Horn's measure (1983) is limited to symmetrical unimodel distributions. He argues that this measure is more precise than the traditional measure of kurtosis and is also applicable to Cauchy distribution; and supports the claim of most of the statisticians that Cauchy distribution is more peaked than normal distribution. He measures the peakedness by the coefficient

$$Mt_p(f) = 1 - \frac{p}{A_p(f)} \tag{1}$$

$$A_p(f) = f(Mx)F^{-1}(p+.5) \tag{2}$$

where is the area under the pdf f(x) contained in the rectangle $R_p(f)$. The measure Mtp (f) varies between 0 and 1. Its larger value indicates more peakedness of f(x). For characterizing peakedness he examines his measure under $p \leq 0.25$. This measure is 0.07 for normal distribution when p=0.25.

Bonett and Seier (2003) suggest measures K_1 and K_2 of kurtosis by the following expected values.

$$K_1 = E\{ab^{-|z|}\} \ , \ 0 < b \leq 1 \tag{3}$$

$$K_2 = E\{a(1-|Z|^b)\} , \ 2 < b \leq 20 \tag{4}$$

where a and b are constants, Z is the standardized random variable. They compare a number of distributions based on these expectations.

Fiori and Zenga (2005) find

$$\partial\alpha_4 / \partial f_r \tag{5}$$

to determine the relative importance of tail heaviness and peakedness, where α_4 and f denote Pearson's coefficient and the density function. They show that α_4 changes when an observation is added at some particular point X_r. Their work is essentially useful to have an idea of the impact of an additional observation on the **kurtic behavior** of a distribution.

In this paper we apply the above measures to the Weibull distribution and make general remarks relative to the traditional measure of kurtosis.

2. MEASURES OF KURTOSIS OF WEIBULL DISTRIBUTION

Weibull family of distributions has a broad division of classes A, B and C with respect to the skewness of their density curves. These classes consist of positively skewed distributions, about symmetric and negatively skewed depending on the shape parameter $\lambda \leq 3.4$, $3.4 < \lambda < 3.8$, and $\lambda \geq 3.8$ respectively. We apply the above mentioned three methods to calculate the coefficients of kurtosis when the shape parameter of Weibull distribution assumes various values. The corresponding coefficients of kurtosis by Pearson's method are also given for the purpose of comparison.

2.1 Measure of Kurtosis by Horn

The pdf of Weibull distribution is

$$f(x) = \lambda x^{\lambda-1} e^{-x^{\lambda}} \quad 0 < \text{x} \tag{6}$$

We determine Horn's measure for a range of the shape parameter λ's values including 3.5 where f(x) is nearly symmetrical - (Table in Appendix, Rahila (2011). The graph of Horn's coefficient against the selected value of p is drawn below. Four curves relating to four values are included in the graph. The fifth curve corresponds to N (0, 1) distribution. The coefficient increases with the increase in p for each selected value of λ. Also as the shape parameter increases Horn's coefficient decreases at each fixed p level.

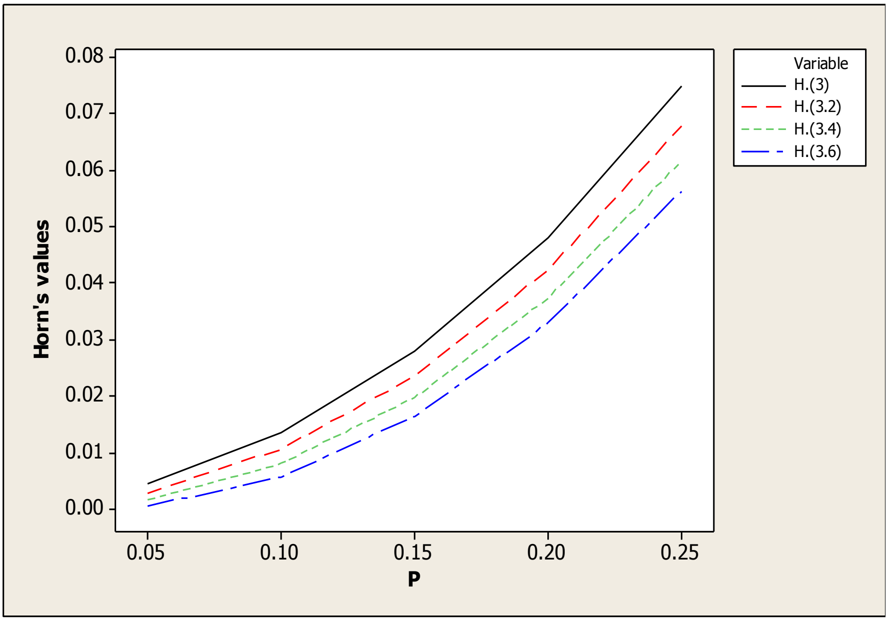


Fig. 1: Horn's Value (Mt_p) for Weibull Distribution

2.1.1 Remarks

Horn's coefficient increases with the increase in p for a selected value of λ. Also as the shape parameter increases, Horn's coefficient decreases at each fixed p level. The relative behavior of these curves is also comparable with the curve based on N(0, 1); that is, the Weibull distributions with larger Horn's coefficients are more peaked. On the contrary if we use Pearson's measure of kurtosis it can be seen that the Weibull family for $3.4<\lambda<3.8$ has approximately the same degree of kurtosis as a normal distribution but this family in not symmetrical.

2.2 Measure of Kurtosis by Bonett and Seier

These measures in Eqs. (3, 4) assign relatively more importance to the central part of the distribution.

We have

$$K_1 = \int_0^\infty ab^{-\left|\frac{X-\mu}{\sigma}\right|} f(x)dx \tag{7}$$

$$K_2 = \int_0^\infty a\left(1-\left|\frac{X-\mu}{\sigma}\right|^b\right) f(x)dx \tag{8}$$

where μ and σ are the mean and S.D of the Weibull distribution. The constant 'a' is selected as so that

K_1 = 3, K_2 = 3 for the normal distribution. We take up the values b= e, 5, 8, 20 and b=0.5, 0.8, 1 to find K_1 and K_2. Their values are given in Tables [Appendix- Rahila (2011)]. We present here the graphs on these measures.

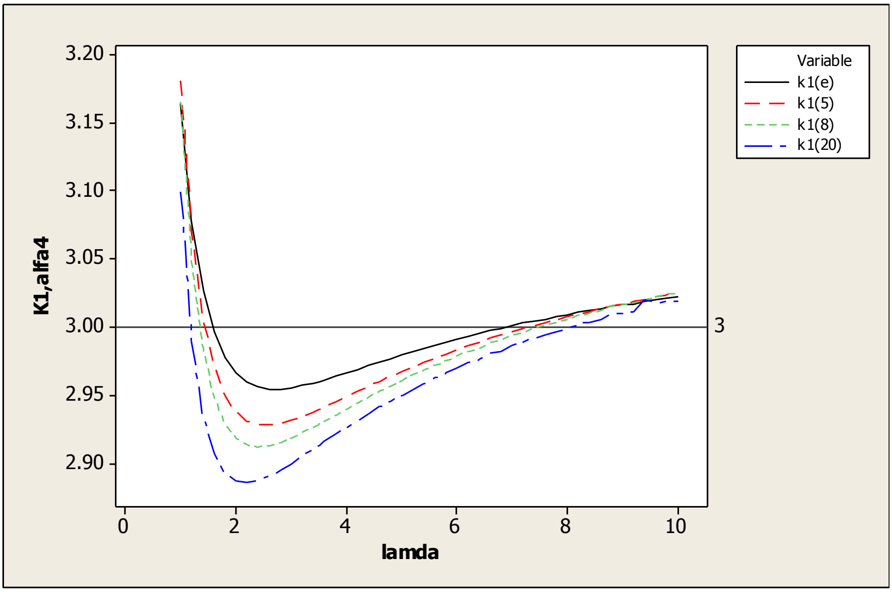


Fig. 2: Combined Graph of K_1 (e, 5, 8, 20) for Three Classes A, B, C

The four values of K_1 present a similar behavior and assume a value less than three between (λ=1.8 to 7). The other classes A and C show opposite patterns based on this measure.

We present a correlation table to display a general relation of kurtic behavior between the two approaches, but since Bonett & Seier propose their measure for symmetrical distributions we focus on its comparison with Pearson's measure with reference to Class B. Clearly a disparity exists implying that Pearson's measure has an added effect of tail characteristic.

Correlation Table
K_1 with Pearson's coefficient

K (a)	α_4 (A)	α_4 (B)	α_4 (C)
$K_1(e)$	**0.999**	**-0.628**	**0.995**
$K_1(5)$	**0.999**	**-0.652**	**0.994**
$K_1(8)$	**0.997**	**-0.664**	**0.993**
$K_1(20)$	**0.987**	**-0.676**	**0.992**

Measure of Kurtosis K_2

Here we consider the second measure K_2. It is calculated only for Class B of Weibull distributions, that is, which have a minor degree of skewness. The following graph includes the kurtic behavior of the distributions by K_2 and Pearson's coefficient α_4 .

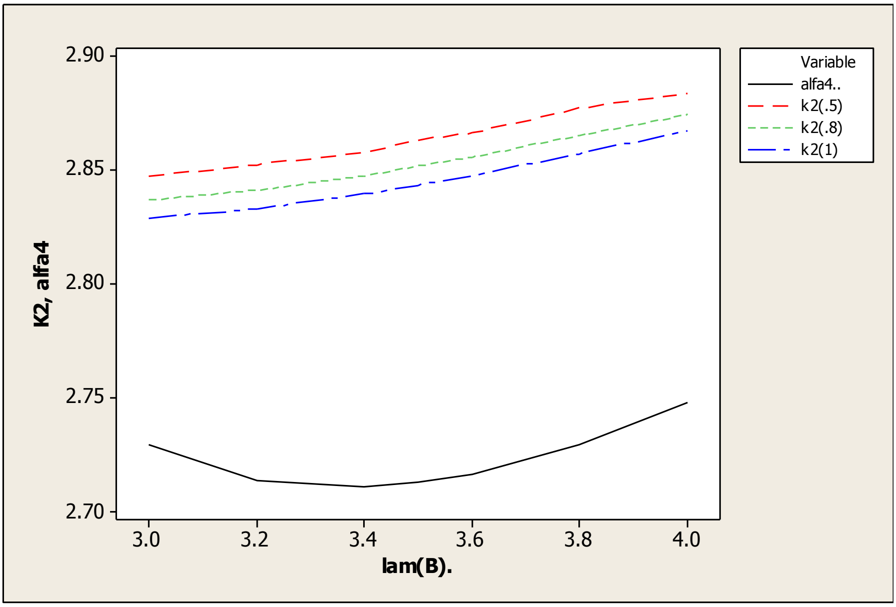


Fig. 3: Graph of K_2 (0.5, 0.8, and 1) and Pearson's kurtosis for Class B

The correlations of K_2 with Pearson's coefficient are given in the following table.

Correlation Table
K_2 with Pearson's coefficient in Class B

Kurtosis	α_4 (B)
$K_2(.5)$	0.612
$K_2(.8)$	0.637
$K_2(1)$	0.656

Remarks

The measures K_1 and K_2 are specially developed to give larger importance to kurtosis at the centre of symmetrical distributions. Class B of Weibull family of distributions was considered here for the comparison of these measures with Pearson's measure of kurtosis. We conclude that these **Bonett & Seier** measures K_1 and K_2 reveal significant disparities when compared with Pearson's measure.

2.3 Measure of Kurtosis by Fiori and Zenga's:

To determine the relative importance of tail heaviness and peakedness they find the partial derivative of Pearson's coefficient α_4 with respect to the frequency f_r (r =1, 2, 3…) of X_r

$$\frac{\partial\alpha_4}{\partial f_r} = \frac{1}{p}\left[(z_r^2-\alpha_4)^2-\alpha_4(\alpha_4-1)-4\alpha_3 z_r\right] \tag{9}$$

where p denotes the total frequency, α_3 and α_4 are skewness and Pearson coefficients of skewness and kurtosis.

For optima change $\frac{\partial\alpha_4}{\partial f_r}=0$. The roots of Eq. (9) for a Weibull distribution derived through Zenga's approach are given in Table 6.3 of Rahila (2011), which in fact forms a basis for the formation of regions. At these points $X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}$ the derivatives are zero.

Let Z be the standardized Weibull random variable. If z_r is the standardized value and α_3 is the skewness coefficient. The values of X are

$$X_r^{(2,3)} = mean \pm z_r SD \tag{10}$$

$$X_r^{(1,4)} = mean \pm z_r SD \tag{11}$$

These points create five regions of the Weibull distribution starting from extreme left. Area of the first region (i) covers the extreme left below X_1.

$$F\left(X_1\right)=1-e^{-X_1^{\alpha 4}}$$

Area of the region "ii" between X_1 and X_2

$$F(X_2)-F(X_1)=e^{-X_1^{\alpha 4}}-e^{-X_2^{\alpha 4}}$$

Area of the region "iii" between X_2 and X_3

$$F(X_3)-F(X_2)=e^{-X_2^{\alpha 4}}-e^{-X_3^{\alpha 4}}$$

Area for region "iv"

$$F(X_4)-F(X_3)=e^{-X_3^{\alpha 4}}-e^{-X_4^{\alpha 4}}$$

Area for region "V"

$$1-F(X_4)=e^{-X_4^{\alpha 4}}$$

The tails, are below X_1 and above X_4. The regions "ii, iv" represent the shoulders and, "iii" shows the center of the distribution. The graphical representation of these regions of

Weibull distribution shows the changes that occur in the shape of the distribution. For $\lambda = 1$ the distribution has three region (iii, iv, v). For $1.2 \le \lambda \le 2.4$ the distribution has four regions (ii, iii, iv, v) and when $\lambda > 2.4$ the distribution has five regions (i, ii, iii, iv, v). Regions "i, v" represent tails.

2.3.1 Remarks

Zenga's work is essentially useful for understanding the effect of an additional observation on the **kurtic behavior** of Weibull distribution. This observation may arise within a left tail, a right tail, one of two shoulders, or the centre. Probability of an additional observation in a region depends on the value of the shape parameter. Critical points in the Kurtic behavior change their positions depending on the parameter. Below we provide a broad picture of the kurtic behavior of a Weibull distribution in the perspective of Zenga's work based on tables and graphs in Rahila (2011).

i) **Class A:** In this class the probability of a new observation is higher on the right tail (right shoulder) when compared with that on the left tail (left shoulder) of the distribution. An additional observation in right tail and shoulder produces negative changes in Pearson's coefficient. If it appears in the centre of distribution, the change in kurtosis is slightly positive.

ii) **Class B**: The right tail (shoulder) and left tail (shoulder) have about the same probability of an additional observation. Its effect on kurtosis remains just moderate, negative though. The observation in the centre has slightly a positive change in the coefficient of kurtosis.

iii) **Class C**: In this class the probability of a new observation is higher on the left tail (left shoulder) when compared with that on the right tail (right shoulder) of the distribution. In this class the effect of an additional observation in the right tail is negative on kurtosis, and in the center it is moderate.

REFERENCES

1. Blest, D.C. (2003), A New Measure of Kurtosis Adjusted for Skewness. *Australian & New Zealand Journal of Statistics*, 45(2), 175-179.
2. Balanda, K.P. and MacGillivray, H.L. (1988). A Critical Review. *The American Statistical Association*, 42(2), 111-119.
3. Balanda, K.P. and MacGillivray, H.L. (1990). Kurtosis and Spread. *Canadian Journal of Statistics*, 18, 17-30.
4. Dyson, F.J. (1943). A Note on Kurtosis. *Journal of the Royal Statistical Society*, 106(4), 360-361.
5. Finucan, H.M. (1964). A Note on Kurtosis. *Journal of the Royal Statistical Society*, 26(1), 111-112.
6. Fiori, A.M. and Zenga, M. (2005). The Meaning of Kurtosis, The Influence Function and an Early Intuition by L. Faleschini. *Statistica*, Anno 65(2), 131-140.
7. Hosking, J.R.M. (1992). Moments or L moments? An example comparing two measures of distributional shape. *The American Statistician Association*, 46(3), 186-189.

8. Horn, P.S. (1983). A Measure for Peakedness. *The American Statistician*, 37(1), 55- 56.
9. Kaplansky, I. (1945). A common error concerning Kurtosis. *The American Statistician Association*, 46(230), 259.
10. Pearson K. (1905). Das Fehlergesetz Und Seine Verallgemeinerungen Durch Fechner Und Pearson. *Biometrika*, 4(1/2), 169-212.
11. Rahila, H. (2011). Measuring kurtosis of Weibull Distribution by Various Approaches. M.Phil. Thesis, University of the Punjab, Pakistan.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 117-122

ARE WE READY TO ACCEPT THE CHALLENGE OF EBOLA VIRUS: PERCEPTION OF HEALTHCARE PROVIDERS

Syed Arif Ali, Masood Hussain Rao, Summaiya Siddiqui
and **Razia Siddiqui**
Dow University of Health Sciences, Lahore, Pakistan

ABSTRACT

Introduction: Ebola virus disease (EVD), formerly known as Ebola hemorrhagic fever, is a severe, often fatal illness in humans. It is a rare and deadly disease caused by infection with Ebola virus. The virus is transmitted to people from wild animals and spreads in the human population through human-to-human transmission. The average EVD case fatality rate is around 50%. Case fatality rates have varied from 25% to 90% in past outbreaks. Because of its high mortality rate, EBOV is also listed as a select agent, World Health Organization Risk Group 4 Pathogen (requiring Biosafety Level 4-equivalent containment).

Objective: To determine the clinical manifestations of viral hemorrhagic fever, and to assess the awareness and knowledge to accept the challenge of Ebola virus in health care providers.

Methodology: This cross sectional non-probability convenience sampling study was conducted on 370 health care providers which includes doctors, nurses and paramedical staff of all Medical, surgical and pediatric units, ICU and emergency department of Civil Hospital Karachi. Ethical approval from the Institutional Review Board of DUHS and a writing consent from each participant were taken. The questionnaire was filled through interview. Data was entered in SPSS 17.0. Mean ± S.D. was calculated for all quantitative data whereas frequency and percentages were calculated for all the qualitative variables. Chi-square test was applied for statistically significance level of the data at P=<0.05.

Results: Out of 370 participants, 249 (67.3%) were ever heard about Ebola virus. Out of these, majority know the symptoms as fever, bleeding severe headache and muscle pain. Regarding appearance of these symptoms, only 53% of 249 participants have the correct knowledge. The way of transmit of the Ebola virus was known by 69.5% participants. Only 27.7% of health care providers know that it is spread in the community through animals and infected persons. Physicians were less ready for handling the Ebola virus patient as well as to get training materials for it as compared to nursing staff (P<0.001). Similarly year of service is directly correlated with the owning the responsibility of handling the patients and to get training for handling the Ebola virus as compared to nursing and other health care providers. (P<0.001)

Conclusion: The Ebola virus is lives threaten but the knowledge and proper prevention procedures of health care providers was not up to the standard. Their behavior shows that our health care providers are not yet ready to accept the challenge of it.

KEY WORDS

Ebola, Ebola hemorrhagic fever, awareness, perception, transmission of Ebola virusTitle.

INTRODUCTION

Ebola virus (EBOV, formerly designated Zaire Ebola virus) is one of five known viruses within the genus Ebola virus. Because of its high mortality rate, EBOV is also listed as a select agent, World Health Organization Risk Group 4 Pathogen (requiring Biosafety Level 4-equivalent containment (1). The first cases of filo virus haemorrhagic fever were reported in 1967 in Germany and the former Yugoslavia, and the causative agent was identified as Marburg virus. Then, it was discovered in 1976 near the Ebola River. The natural reservoir of Ebola virus remains unknown. However, researchers believe that the virus is animal-borne (bats) (2).

Ebola virus disease (EVD), formerly known as Ebola hemorrhagic fever, is a severe, often fatal illness in humans. It is a rare and deadly disease caused by infection with Ebola virus. It can cause disease in humans and nonhuman primates (monkeys, gorillas, and chimpanzees).This virus was spread more in 1989, 1994, 2000, 2010 and then the virus causing the west African outbreak belongs to the Zaire species in this year (2014) (3).

The virus is transmitted to people from wild animals and spreads in the human population through human-to-human transmission. The average EVD case fatality rate is around 50%. Case fatality rates have varied from 25% to 90% in past outbreaks. Community engagement is key to successfully controlling outbreaks, by applying a package of interventions, namely case management, surveillance and contact tracing, a good laboratory service, safe burials and social mobilization. Early supportive care with rehydration, symptomatic treatment improves survival. There is as yet no licensed treatment proven to neutralize the virus but a range of blood, immunological and drug therapies are under development. There are currently no licensed Ebola vaccines (4).

Very few studies have been conducted on this issue worldwide. In a cross sectional study conducted in Pakistan in 2002 on knowledge regarding viral hemorrhagic fever (VHF) (knowledge level regarding VHF management, spread and prevention) among healthcare personnel (doctors, nurses, laboratory technicians and janitors/orderlies) at two largest tertiary care hospitals of Karachi. The result of the study showed that Differences in knowledge of different groups is obvious but it is essential to raise the knowledge regarding VHF at all levels including the housekeeping staff. There is also a need for continuing medical education (CME) for all health care personnel for the emerging health problems in Pakistan (5). During outbreaks, public education to raise awareness and increase knowledge should be the primary focus of public health practitioners. Protective measures that individuals can take must be emphasized and information should be disseminated very efficiently and rapidly (6). Communication theory and techniques, aided by the electronic revolution, provide new opportunities and challenges for the effective transfer of laboratory, epidemiologic, surveillance and other public health data to the public (7). Another study results showed that majority (over 80%) of the respondent's exhibit positive attitudes towards Ebola disease prevention. However, misconceptions about the prevention and treatment of the disease exist among 20-42% of the respondents (8). In another study, it was calculated that infectious diseases associated with mass gatherings vary depending on the type and location of the mass gathering. The key features of the preparedness for the 2014 Hajj and Umra, review the recent impact of emerging viruses such as Ebola in West Africa and the Middle East respiratory syndrome corona virus (MERS-CoV) in affected countries, and highlight the updated requirements and the required vaccines (9). In another study, ost residents and students at our institution felt that outbreak preparedness was not adequately addressed in their respective curricula and almost all wanted more instruction. This training is particularly important for our students and residents (10).

In account of the recent outbreak of Ebola virus, it is important to anticipate and perceive knowledge regarding this among health care personnel. Therefore, the rational of the study is to determine the knowledge, attitude and practice pattern of our health care providers regarding this Ebola virus so that future planning to face the virus if came to our country.

METHODOLOGY

This cross sectional; non-probability purposive sampling technique was used in this study, conducted on doctors, nurses and Para-medical staff of all medical units, pediatrics and emergency department of Civil Hospital Karachi. For this purpose we have select 3 senior doctors (senior registrar to professor), 10 junior doctors (RMO and house officers) 2 staff nurses, 4 student nurses and 6 Para medical staff (dispenser, dresser, ward boy, sweeper) in each unit. Therefore overall 25 participants were selected from each unit. The questionnaire was filled through interview bases. The duration of study was six months after approval of synopsis. Our inclusion criteria were the Doctors and Para medical staff having emergency exposure and currently in practice and willing to participate in the study whereas our exclusion criteria was all those Medical personnel who were not available during survey or did not willing to participate in the study. Sample size was calculated through Open Epi online sample size calculator taking 57% doctors knew the common signs and symptoms of VHF, 5% Confidence limits and 95% Confidence Level, it was estimated 370. The data thus collected were analysed by using SPSS 21. Mean and S.D was calculated for quantitative data like age etc. Frequency and percentages were calculated for all the qualitative variables. Chi-square test was applied for statistically significance level of the data at $P = < 0.05$.

ETHICAL CONSIDERATION

This study involves human subjects. The ethical approval of the study was obtained from Institutional Review Board of Dow University of Health Sciences. Before filling the questionnaire informed written consent was taken from each study participant.

RESULTS

Out of 370 participants, 249 (67.3%) were ever heard about Ebola virus. Out of these, majority know the symptoms as fever, bleeding severe headache and muscle pain. Regarding appearance of these symptoms, only 53% of 249 participants have the correct knowledge. The way of transmit of the Ebola virus was known by 69.45% participants whereas 64.6% know the PCR is the best for lab diagnosis of this virus. Only 27.7% of health care providers know that it is spread in the community through animals and infected persons (Table I).

Very low percentage of the participants (32.4) was ready to handle the Ebola virus patients in the hospital; however majority (79.2%) were ready to get training on this issue (Table II).

Physicians were less ready for handling the Ebola virus patient as well as to get training materials for it as compared to nursing staff (P<0.001). Similarly year of service is directly correlated with the owning the responsibility of handling the patients and to get training for handling the Ebola virus as compared to nursing and other health care providers. (P<0.001)

Table I
Knowledge about Ebola Virus in Health Professionals

S#	Knowledge	Yes	No
1.	Have you ever heard about Ebola Virus?	249(67.3%)	121(32.7%)
2.	What are the symptoms of Ebola Virus?	Fever	225 (90.4%)
		Bleeding	183 (73.5%)
		Severe head ache	167 (67.1%)
		Muscle pain	163 (64.5%)
		Bruising	127(51.0%)
		Weakness	123 (49.4%)
		Abdominal pain	109(43.8%)
		Diarrhea	80(32.1%)
		Vomiting	76(30.5%)
3.	How does Ebola virus transmit?	1. Direct contact with blood of EBV patients	140(56.2%)
		2. Handling body fluids of EBV patients	173(69.5%)
		3. Contaminated needle syringe	88(35.3%)
		4. Handling infected wild life	64(25.7%)
		5. Contact with objects like bedding/clothing of EBV patients	104(41.8%)
4.	How Ebola virus spread in the community	1. Through animals	69(27.7%)
		2. Through person to person	161(64.7%)
		3. Through used items of a infected patient	125(50.2%)
		4. Through medical equipment used for the patient	76(30.5%)

Table II
Attitude of Participants Regarding Ebola Virus Handling

S#			Number (%)
1.	Are you ready to handle the patient in your hospital?	1. Yes	120(32.4%)
		2. No	140(37.8%)
		3. Not sure	74(20.0%)
		4. Will see the condition of the patient	36(9.7%)
2.	Are you ready to get any training/ material on this issue?	1. Yes	293(79.2%)
		2. No	32(8.6)
		3. We have no proper information	28(7.6%)
		4. Think that information providers are also having lacking information.	17(4.6%)

DISCUSSION

The Ebola virus is life threatens but by increase the proper knowledge and proper prevention procedures of health care providers it can be prevented or reduce the chances to spread it. In a study (8) it was concluded that majority (over 80%) of the respondent's exhibit positive attitudes towards Ebola disease prevention. In our study, Physicians were less ready for handling the Ebola virus patient as well as to get training materials for it as compared to nursing staff (P<0.001).

In the same study (8), it was mentioned that misconceptions about the prevention and treatment of the disease exist among 20-42% of the respondents. Our study results were also similar to it.

CONCLUSION

The Ebola virus is lives threaten but the knowledge and proper prevention procedures of health care providers was not up to the standard. Their behavior shows that our health care providers are not yet ready to accept the challenge of it. It is therefore essential to provide them basic knowledge regarding Identification of high-risk patients, symptoms and precautions to hand blood and body fluid of patients to reduce the risk of spread of such viral hemorrhagic fevers.

LIMITATIONS

This study has been conducted in single government hospital. Therefore the results of multiple hospitals, private as well as government may be different from it. Secondly as no such case have come to this city, therefore majority are not ready to see the effects of it to adopt proper prevention measures.

REFERENCES

1. From Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Ebola_virus
2. Sokol, D.K. (2006). Virulent epidemics and scope of healthcare workers' duty of care. *Emerg Infect Dis.*, 12, 1238-41. DOI: 10.1056/NEJMp1410301
3. Geisbert, T.W., Feldmann, H. (2011). Ebola haemorrhagic fever, *Lancet*. 377(9768), 849-862. DOI: 10.1016/S0140-6736(10)60667-8.
4. *Fact sheet* No. 103, Updated September 2014.
5. Lakhani, A., Mahmood, H., Laeeq, A., Mansoor, S., Lodhi, S., Majid, S., Kizilbash, Q. and Altaf, A. (2002). Viral Hemorrhagic Fever in Pakistan: Awareness among Health Care Personnel. *JPMA*, 52, 214-7.
6. Elisha, A. and Adegboro, B. (2004). ebola virus diseases (evd), *Afr. J. Cln. Exper. Microbiol.*, 15(3), 178-121. http://www.ajol.info/index.php/ajcem/article/download/107472/97340.
7. Freimuth, V., Linnan, H.W. and Potter, P. (2000). Communicating the Threat of Emerging Infections to the Public. *Emerging Infectious Diseases*, 6(4), 337-341.
8. *Study on Public Knowledge, Attitudes, and Practices Related to EVD Prevention and Medical Care in Sierra Leone*. Sierra Leone: UNICEF, FOCUS 1000 & Catholic Relief Services, 2014.
9. Al-Tawfiq, J.A. and Memish, Z.A. (2014). Mass gathering medicine: 2014 Hajj and Umra preparation as a leading example. *International Journal of Infectious Diseases*, 27, 26-31. DOI: http://dx.doi.org/10.1016/j.ijid. 2014.07.001.
10. May, L., Omron, R., Piller, M., Haile-Mariam, T. and Scott, J. (2006). Integrating Emerging Infections Education into Medical Education: An Innovative Approach, *Med. Educ. Online* [serial online] 12, 5, http://www.med-ed-online.org.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 123-128

RECENT TREND IN SPECIALTY CHOICES OF DOW MEDICAL COLLEGE STUDENTS, KARACHI

Syed Hussain Zaidi, Arif Ali, Hasham Naim, Salman Farrukh and **Haris Hamid**
Dow Medical College, Dow University of Health Sciences, Karachi, Pakistan
Email: arif_ali79@hotmail.com

ABSTRACT

Understanding regarding distribution of career selection in medicine is substantial for educators and policy-makers to plan the provision of specialists. This research study analyzed the career choices made by medical students and house officers in Dow Medical College, Karachi, Pakistan. A self-administered questionnaire was completely filled by 260 students. 15.51% Internal medicine, 11.84% cardiology, 9.39% general surgery and 10.2% pediatrics were the 4 highest ranked specialties, with internal medicine ranked top as both the 1st and 3rd choice. We also analyzed each specialty at first, second and third choices among male and female. It was identified that cardiology was the top ranked at the first choice among male while internal medicine was on the top among female students at first choice (16% versus 18.3%).Medical students chose a wider range of specialties. The results of this research study may be useful for Karachi Medical Council in order to sustain, support to develop interest of medical students in those medical specialties that are less likely chosen by them.

INTRODUCTION

Most of the countries around the world are experiencing shortage of doctors especially in developing countries (1, 2). Lack of physicians poses a serious threat to the delivery of basic health care facilities in already vulnerable societies (1).

Literature shows that there are many researches that have explained this subject and emphasized various reasons for low proportion of few medical specialties, for example family (3-5) medicine, radiology (6-8) and psychiatry (9, 10).

In the prevalent shortage of trained physicians it is important to know the changing trend, interests and expectations among new generation of medical students. Selecting a medical specialty can be confusing and complicated experience for medical students and house officers (11).

Few researches have been completed in order to guess the distribution of medical students in future among medical specialties (12) and factors account able for that distribution (13, 14).

There are many problems, factors and incentives such as desire to earn high income, family responsibilities, interest in one specific field, prestige related to certain specialties, years of training that influences an individual over deciding a particular career (1,2). Some studies in past have even suggested that carrier preference at beginning of medical school may be predictor of student's eventual career (15).

Literature exhibits that few researches have also been conducted in Pakistan for the purpose of exploration of specialty selection. For example a research study was conducted in a private institute in order to identify final year students' career preferences, that study based on sample size (16) nevertheless such kind of studies may guide while counseling students about their career and making policies (17). Another study (18) was conducted in Karachi, Pakistan in 2007 regarding specialty choices among medical students and house officers in 2 private and 2 public institutions.

The purpose of this research study was to determine the trends in specialty choices of students in 2015 from one of most prestigious and famous public institute Dow Medical College, Karachi.

METHODOLOGY

This cross-sectional study was conducted from May 2015 to December 2015 at Dow Medical College, Dow University of Health Sciences, Karachi, Pakistan. Only 4th and 5th year medical students and house officers were included in the study; first three years students were not included as they have limited clinical exposure. Foreign students and those who were not present at the time of data collection were also excluded. All study participants were selected though non probability, purposive sampling technique.

Questionnaire was composed by structured questions which included demographic characteristics of students and factors that influenced career choice among medical students. Instructions were given to the students regarding filling the questionnaire.

Ethical permission regarding conduct of this research was taken from Institutional Review Board, Dow University of Health Sciences. Students were informed about the eligibility criteria and study objectives before they signed the consent form.

The collected data were entered into and analyzed through Statistical Package for Social Sciences (SPSS) version 21. The data were presented in frequency and percentages for all qualitative variables. Means and standard deviations were calculated for all continuous variables.

RESULTS

A total of 260 medical students and house officers participated in this study. The mean age of study participants was 22.02±1.77 and majority 67.1% of the study participants were female. Out of 260 medical students, 15.51% Internal medicine, 11.84% cardiology, 9.39% general surgery and 10.2% pediatrics were the 4 highest ranked specialties, with internal medicine ranked top as both the 1st and 3rd choice. We also analyzed each specialty at first, second and third choices among male and female. It was identified that cardiology was the top ranked at the first choice among male while internal medicine was on the top among female students at first choice (16% versus 18.3%). Whereas, at second choice cardiology was again on the top among male and pediatrics was among female (17.1% versus 17.4%).

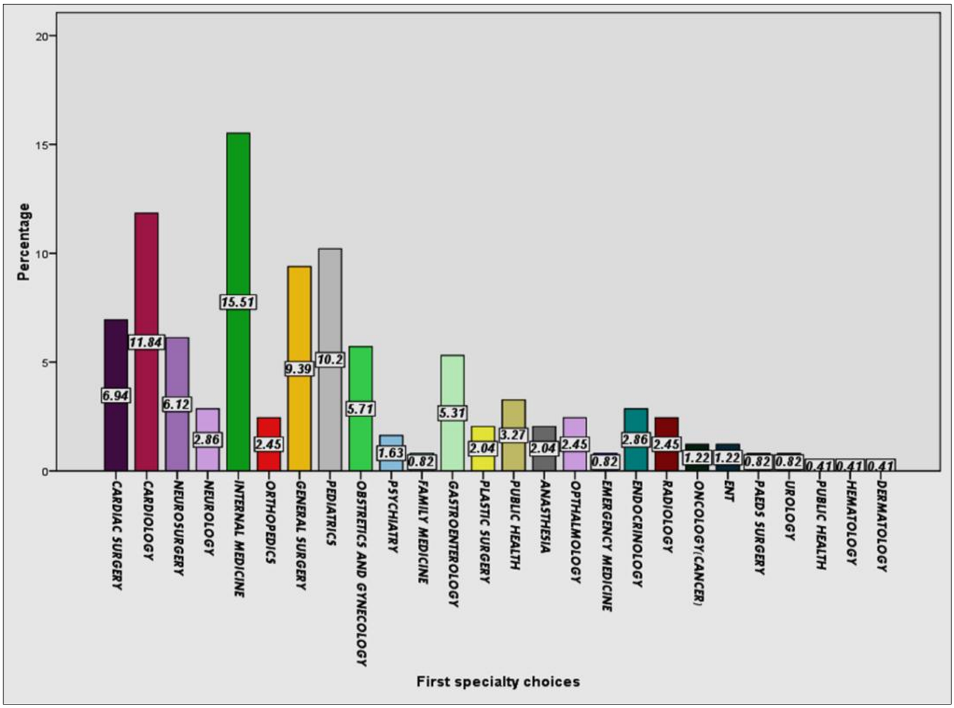


Fig. 1: Distribution of First Preference of Medical Specialty among Medical Students

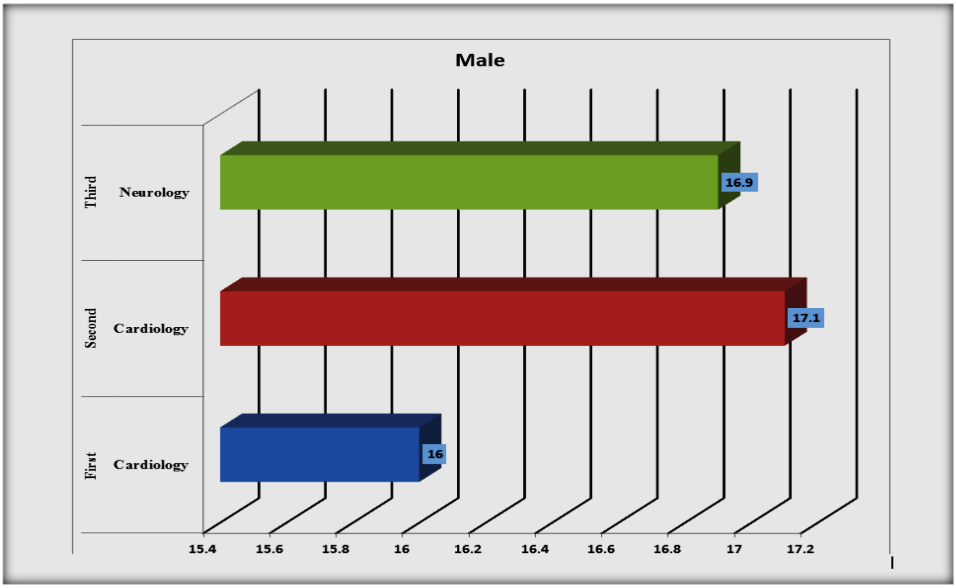


Fig. 2: Pattern of top Ranked First, Second and Third Preference among Male Students at First Ranked Choice

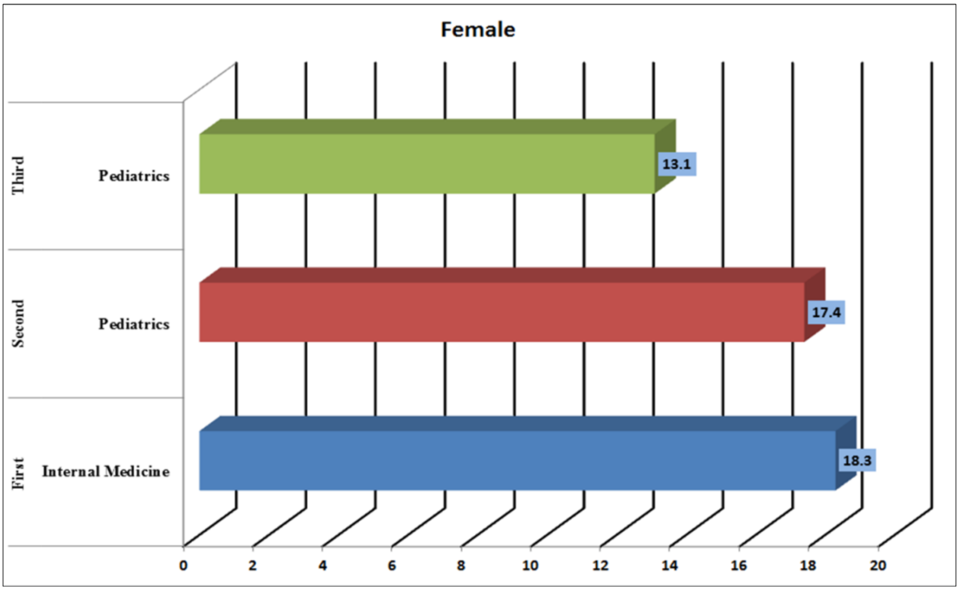


Fig. 3: Pattern of Top ranked First, Second and Third Preference among Female Students at First Ranked Choice

DISCUSSION

In our study 15.4% Internal medicine, 11.9% cardiology, 10% general surgery and 9.6% pediatrics were the 4 highest ranked of specialties. In other study completed in Pakistan, Internal medicine was the favorite with 20.5% of medical students reporting it as their 1st choice, followed by surgery (15.0%), pediatrics (13.5%), and obstetrics and gynecology (9.3%) (18).

Internal medicine ranked top as both the 1st and 3rd choice. Cardiology was the top ranked at the first choice among male while internal medicine was on the top among female students at first choice (16% versus 18.3%).

Aslam et al. showed that Internal medicine was also the top-ranked at 2nd choice specialty (17.2%), followed by pediatrics (12.8%), obstetrics and gynecology (9.3%), and surgery (9.0%)(18). In our study at second choice pediatrics ranked top with Internal medicine coming second and cardiology and general surgery at third and fourth place respectively.

For the 3rd choice specialty, internal medicine was also popular (11.2%) which was consistent with ours followed by pediatrics and general surgery.

An analysis of the top-10 list of preferred career choices in medicine revealed some interesting information. Internal medicine, cardiology, pediatrics and general surgery were the 4 highest ranked specialties by our study participants, which is consistent with the findings reported by Khan at el, from a medical university in Pakistan(16).

CONCLUSION

Medical students chose a wider range of specialties. The results of this research study may be useful for Karachi Medical Council in order to sustain, support to develop interest of medical students in those medical specialties that are less likely chosen by them.

REFERENCES

1. Puertas, E.B., Arósquipa, C., Gutiérrez, D. (2013). Factors that influence a career choice in primary care among medical students from high-, middle-, and low-income countries: a systematic review. *Revista Panamericana de Salud Pública*, 34(5), 351-8.
2. Knox, K.E., Getzin, A., Bergum, A., McBride, P., Rieselbach, R., Friedsam, D. (2008). Short report: Factors that affect specialty choice and career plans of Wisconsin's medical students. *Wisconsin Medical Journal (WMJ)*, 107(8), 369-73.
3. Campos-Outcalt, D., Senf, J., Pugno, P.A., McGaha, A.L. (2007). Family medicine specialty selection: a proposed research agenda. *Family Medicine-Kansas City*, 39(8), 585-589.
4. Starfield, B. (2007). Primary care in Canada: coming or going? *Healthcare Papers*, 8(2), 58-62; discussion 4-7.
5. McGaha, A.L., Garrett, E., Jobe, A.C., Nalin, P., Newton, W.P., Pugno, P.A. and Kahn, N.B. Jr. (2007). Responses to medical students' frequently asked questions about family medicine. *Am. Fam. Physician*, 76(1), 99-106.
6. Maj, L. and Baerlocher, M.O. (2006). Factors leading to radiology career selection: results of the 2004 National Physician Survey. *Canadian Association of Radiologists Journal*, 57(3), 175-178.
7. Kazerooni, E.A., Blane, C.E., Schlesinger, A.E. and Vydareny, K.H. (1997). Medical students' attitudes toward radiology: comparison of matriculating and graduating students. *Academic radiology*, 4(8), 601-7.
8. Ehara, S. (2007). Preclinical medical student education in radiology. *American Journal of Roentgenology*, 189(3), W166.
9. Goldacre, M.J., Turner, G., Fazel, S. and Lambert, T. (2005). Career choices for psychiatry: national surveys of graduates of 1974-2000 from UK medical schools. *The British Journal of Psychiatry*, 186(2), 158-64.
10. Wigney, T. and Parker, G. (2007). Medical student observations on a career in psychiatry. *Australian and New Zealand Journal of Psychiatry*, 41(9), 726-31.
11. Al-Fouzan, R., Al-Ajlan, S., Marwan, Y. and Al-Saleh, M. (2012). Factors affecting future specialty choice among medical students in Kuwait. *Medical Education Online*, 2012, 1-7.
12. Saigal, P., Takemura, Y., Nishiue, T. and Fetters, M.D. (2007). Factors considered by medical students when formulating their specialty preferences in Japan: findings from a qualitative study. *BMC Medical Education*, 7(1), 1.
13. Lawrence, J., Poole, P. and Diener, S. (2003). Critical factors in career decision making for women medical graduates. *Medical Education*, 37(4), 319-27.
14. Kassebaum, D.G., Szenas, P.L. and Schuchert, M.K. (1996). Determinants of the generalist career intentions of 1995 graduating medical students. *Academic Medicine*. 71(2), 198-209.

15. Wright, B., Scott, I., Woloschuk, W. and Brenneis, F. (2004). Career choice of new medical students at three Canadian universities: family medicine versus specialty medicine. *Canadian Medical Association Journal*. 170(13), 1920-4.
16. Khan, H., Taqui, A.M., Khawaja, M.R. and Fatmi, Z. (2007). Problem-based versus conventional curricula: influence on knowledge and attitudes of medical students towards health research. *PLoS One*. 2(7), e632.
17. Biggs, J. (2008). Postgraduate medical training in Pakistan: observations and recommendations. *J. Coll. Physicians Surg. Pak.*, 18(1), 58-63.
18. Aslam, M., Ali, A., Taj, T., Badar, N., Mirza, W., Ammar, A., Muzaffar, S. and Kauten, J.R. (2011). Specialty choices of medical students and house officers in Karachi, Pakistan. *Eastern Mediterranean Health Journal*, 17(1), 74-9.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 129-134

EVALUATION OF PERCEPTIONS OF HEALTH PROVIDERS REGARDING USE OF ELECTRONIC MEDICAL RECORDS IN A TERTIARY CARE HOSPITAL OF KARACHI

M. Zaka Uddin, Tayyaba Aamir, Saba Wajeeh
and **Masood Hussain Rao**
Dow University of Health Sciences, Karachi, Pakistan

ABSTRACT

Background:

The computer-based patient record is an essential technology for health care. The concept of electronic health record was generated in 1990s but due to lack of technological standards, difficulty in using the systems; and System cost implementation of such system was not widespread. The idea that the EMR is needed to transform the health system to improve quality and to enhance safety.

Objectives:

1. To determine the level of acceptance of electronic medical record in the physicians and other health care providers.
2. To evaluate the barriers/ misconceptions regarding the use of electronic medical records by physicians and healthcare officials.
3. To develop the awareness and need of electronic health record of each patients in the hospital.

Materials & Methodology:

A cross-sectional study was conducted in Memon Medical Hospital, Karachi Pakistan during 2015. A sample size of 80 was calculated. The study population was consultants, doctors and paramedics who are using this system atleast for therlast three months.

Results:

Most of the physicians and paramedics (64%) were using it and found it easier or slightly easier to use the electronic medical record whereas (22%) did not agree about it, and the remaining (14%) found difficult or slightly difficult (Graph 1).

Conclusion:

According to consultants, doctors & paramedics EMR system is quite easy to use, time saving, reduction of errors, data safety and security, timely retrieval of records and authenticity. However Reluctances were found in some consultants and few para medic staff due to complexity of infrastructure it cannot be implemented. EMR system implementation must include establishing connections to all the clinical data including linking to pharmacies, labs, and radiology with privacy assurance and accountability.

KEYWORDS

Electronic medical record, perception, misconception, barriers, utility of EMR.

INTRODUCTION

The computer-based patient record is an essential technology for health care (1). The concept of Electronic Health Record (EHRs) or Computerized Patient Records (CPRs) is not a new idea and it was initially generated in 1990s but due to lack of technological standards, difficulty in using the systems; and mostly System implementation cost of such system, it was not widespread in the world as it should be (2,3,4,). The idea that the EMR is needed to transform the health system to improve quality and to enhance safety (5). Pakistan has also recently expressed a great deal of interest in electronic medical health records in healthcare system (6, 7).

The rational of the study is that no study have been done on assessment of acceptance by the health care providers and reduction of errors and Perceived ability to deliver high-quality patient care through quick accessibility about past health conditions. This study will help in development of awareness in them and remove the misconceptions regarding safety and misuse of the data of the patients and provide a quick access of the patients past history and current lab and radiological findings. Further more if it is proven than this Electronic Health Record will be utilized in to National Health Records, research purposes, health policy making and in decision making for health care projects.

OBJECTIVES

1. To determine the level of acceptance of electronic medical record in the physicians and other health care providers.
2. To evaluate the barriers/ misconceptions regarding the use of electronic medical records by physicians and healthcare officials.
3. To develop the awareness and need of electronic health record of each patients in the hospital.

MATERIALS & METHODOLOGY

A cross-sectional study was conducted in Memon Medical Hospital, Karachi Pakistan during 2014. Regarding ethical approval of the study and participants, permission was taken from the institute and a writing consent was taken from the each individual participant. For this study the sample size of 80 was selected as their were only 80 health care providers in that setup. The study population was consultants, doctors and paramedics who are using this system atleast for the last three months. Face to face interaction with consultants, doctors & paramedics for getting information. Overall 80 questionnaires were filled by targeted population that is the response rate was 100 %. The was collected through linkert scale, as all the participants have to answer in significantly more difficult, slightly difficult, not sure, easier or slightly easier. The collected data was analyzed through SPSS version 17.0.

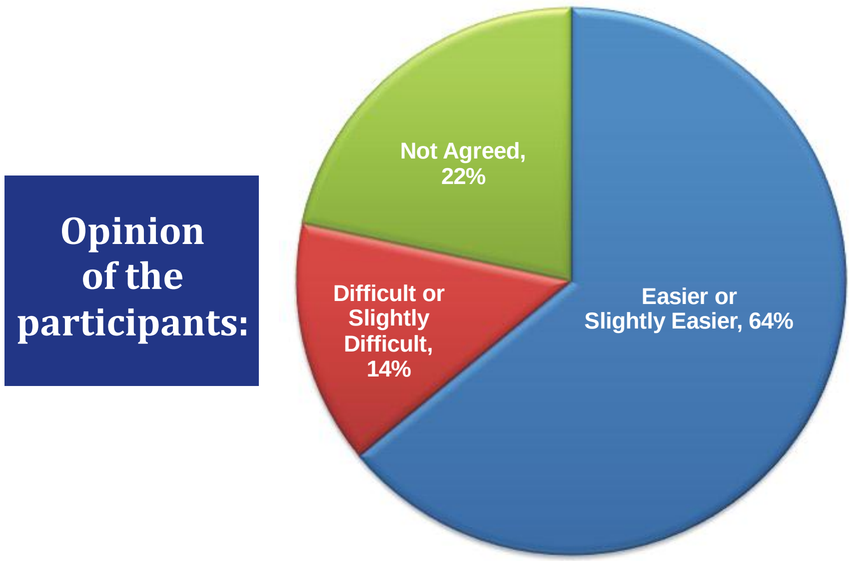
RESULTS

Majority of the health care providers participate in this study was found the system easier or slightly easier regarding review the patients problem, seeking specific. information without recall biases follow the results of a particular test, radiological information or to proper writing information to their patients (Table 1).

Table 1
Important Information get from the Health Providers (in %)

S#	Questions	significantly/ Slightly more Difficult	Not Sure	Easier or Slightly Easier
1.	To review the patients problems	24.2	11.0	64.8
2.	To seek out specific information from patient records	18.7	8.8	72.5
3.	To follow the results of a particular test of investigation over time	11.1	1.1	87.8
4.	To obtain results from new test or investigation	20.8	8.7	70.5
5.	To order treatment directly	13.2	15.4	71.4
6.	To give written general medical information to patients	11.0	18.7	70.3
7.	To give written individual information to patients	14.3	31.9	53.8

According to the analysis of the data, most of the physicians and paramedics (64%) were using it and found it easier or slightly easier to use the electronic medical record. However 22% of the physicians and paramedics did not agree about it, and the remaining (14%) found it difficult or slightly difficult (Graph 1).



The beneficial points of electronic medical record were easier in ordering or obtaining the laboratory test reports and radiological findings record directly. The main barriers/ misconceptions were security of the information, lack of training and storage facilities of the data.

DISCUSSION

According to our study, it is concluded that Electronic Medical Record (EMR) is highly useful in healthcare settings, for safety and security of patient's data; it can be retrieved anytime without any errors and wastage of time, which also builds trust of patients for organization. The same is proving by the other studies conducted previously in Pakistan and worldwide (8, 9, and 10).

Similarly, in our study, the main barriers /misconceptions were security of the information, lack of training and storage facilities of the data. This was also documented in other studies that their health care providers are also having the same barriers and misconception especially the confidentiality of the patient's records (8, 11 and 12).

CONCLUSION

Although it is not fully implemented in our country, but wherever implemented like Indus Hospital, Aga Khan University Hospital and Memon Medical Institute, gives for better performance after the implementation, so it is highly recommended by the consultant, doctors and paramedics. According to consultants, doctors & other health paramedic's staff, the electronic health system is quite good, easy to use, time saving, reduction of errors, data safety and security, timely retrieval of records and authenticity. However reluctances were found in some consultants and few Para medical staff due to complexity of infrastructure it cannot be implemented. EMR system implementation must include establishing connections to all the clinical data including linking to pharmacies, labs, and radiology with privacy assurance and accountability.

LIMITATION OF THE STUDY

As this study has been done in a single hospital, the result of multicenter and major big hospitals may be quite different from this study. Similarly in our setup, due to complexity of infrastructure it may not be implemented easily.

RECOMMENDATIONS

The Government of Pakistan should come forward to provide financial assistance to physicians in the implementation of EMRs in all the hospitals. At the same time physicians awareness should also be developed in the health care providers and steps should be taken to change their processes and frame of mind regarding EHR. It is also needed for policies and procedures for system access. Each employee needs training in the patient data privacy requirements established under the Health Insurance Portability and Accountability Act.

REFERENCES

1. Karahanna, E., Straub, D.W. and Chervany, N.L. (1999). Information technology adoption across time: a cross sectional comparison of pre-adoption and post-adoption beliefs. *MIS Quarterly*, 183-213.
2. Chiasson, M.W. and Davidson, E. (2004). Pushing the contextual envelope: developing and diffusing IS theory for health information systems research. *Information and Organization*, 14(3), 155-188.
3. Ilie, V., Courtney, J.F. and Van Slyke. C. (2007). Paper versus electronic: challenges associated with physicians' usage of electronic medical records. *40th Annual Hawaii International Conference on System Sciences*, HICSS 2007. DOI: 10.1109/HICSS.2007.424.
4. DesRoches, C.M., Campbell, E.G., Rao, S.R., Donelan, K., Ferris, T.G., Jha, A., Kaushal, R., Levy, D.E., Rosenbaum, S., Shields, A.E. and Blumenthal, D. (2008). Electronic health records in ambulatory care—A national survey of physicians. *N. Engl. J. Med.*, 359:50-60.
5. Simon, S.R., Kaushal, R., Cleary, P.D., Jenter, C.A., Volk, L.A., Orav, J. Burdick, E., Poon, E.G. and Bates, D.W. (2007). Physicians and electronic health records: a statewide survey. *Archives of Internal Medicine*, 167(5), 507-512.
6. Randeree, E. (2007). Exploring physician adoption of EMRs: a multi-case analysis. *Journal of medical systems*, 31(6), 489-496.
7. Meeks, D.W., Smith, M.W., Taylor, L., Sittig, D.F., Scott, J.M. and Singh, H. (2014). An analysis of electronic health record-related patient safety concerns. *Journal of the American Medical Informatics Association*, 21, 1053-1059.
8. Butt, F.S. (2014). Barriers in implementation of electronic medical records in Pakistan. *Science International*, (Lahore) 26(4), 1681-1686.
9. Noor, M.S. and Khan, K. (2012). Reluctance of US Doctors in Adopting EHR Technology. *Global Journal of Management & Business Research*, 12(23), 15-22.

10. Miller, R.H. and Sim, I. (2004). Physicians' use of electronic medical records: barriers and solutions. *Health Affairs*, 23(2), 116-126.
11. Kalogriopoulos N.A., Baran, J., Nimunkar, A.J., et al. (2009). Electronic medical record systems for developing countries: review. Conference proceedings: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2009, 1730-1733.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 135-142

EFFECT OF MACROECONOMIC VARIABLE ON EQUITY RETURNS- A SALOMAN BROTHER'S APPROACH

Syed Monis Jawed and **Muhammad Abdul Rehman**
Department of Statistics, University of Karachi, Karachi, Pakistan
Email: monis_jawed@hotmail.com

ABSTRACT

The purpose of this study is to assess the impact of macroeconomic variables on Karachi Stock Exchange (recently renamed as Pakistan Stock Exchange). For investigating this relationship we have applied Saloman's Brother Approach. This model used to assess the equity return through six macroeconomic factors- economic growth, business cycle, long term interest rate, short term interest rate, inflation and exchange rate. Saloman's Brother does not only suggest the macroeconomic fundamentals but they have also described the proxies that can depict the above mentioned macroeconomic fundamentals. Many previous studies have used this to explain the equity returns for different stock exchanges around the globe; but so far, this model has not been investigated for Pakistan. We are intending to explore this relationship via the usage of Vector Autoregression (VAR) Model. The time period considered in this study is from October 2003 to December 2012. It is found that under the Saloman's Brothers model framework, the economic growth directly affects the performance of stock market, whereas other macroeconomic variables included in the model indirectly imparts their impact on KSE-100 index. We conclude that KSE-100 index responds to macroeconomic factors included in Saloman Brothers model directly as well as indirectly, thus channels of transfer to KSE 100 index from macroeconomic factors can be determined on the basis of findings of this study.

KEYWORDS

Fundamental Analysis Saloman Brother's Vector Autoregression Equity Market

INTRODUCTION

It has long been a tradition amongst financial economists, analysts and practitioners to figure out the relationship between stock prices and various macroeconomic and financial factors. Such analysis is conventionally termed as fundamental analysis. Fundamental analysis enables us to predict the stock prices consequently it may shape the investor's decision.

There are various levels of conducting fundamental analysis. For instance, one may try to establish the relationship between stock prices prevailing in the country to the stock prices and other macroeconomic and financial factors that are prevailing in other countries. Another perspective of carrying out fundamental analysis is to explore the relationship between stock prices and macroeconomic factors prevailing in some economy. Similarly, another context in which fundamental analysis could be carried out is exploring the relationship between stock prices of a particular industry with the factors that govern the operating performance in a particular industry. Finally, fundamental analysis can be carried out by observing the relationship between stock prices and a firm's financial performance as revealed through its balance sheet.

In all the above mentioned approaches, the primary purpose for conducting fundamental analysis is to figure out the factors that determines stock performance. The present study also aims at figuring out various factors that may enable us in determining the level of stock returns. Starting with the work of King (1966), various multifactor models, such as Chen, Roll and Ross (1986) and Burmeister and McElory (1988) were developed to figure out determinants of stock price movements. However, in this study, we are carrying out fundamental analysis in the context of domestic economy. We have used the approach proposed by Sorensen et al. (1989). This approach was followed by Saloman Brother's- a famous brokerage house at Wall Street. Consequently, this approach is pronounced at Saloman Brother's model. This model proposes certain macroeconomic factors that could affect the security prices. The model not only identifies the macroeconomic factors that explain stock market prices, it also suggests proxies that can be used to depict the macroeconomic phenomenon.

In this study, we have used vector autoregresssion (VAR) model to implement Saloman Brothers model. VAR is a reduced form model and possesses an edge in exploring macroeconomic relations. The VAR model enables us to observe the interactions between various variables present in the model, thus, it also enable us in identifying the channels through which different factors affect the stock prices. Various studies have explored the relationship between fundamentals and stock prices in Pakistan, but to the best of our knowledge, thus far, no study has investigated the impact of fundamentals through Saloman Brother's approach. Thus, this study will shed new light on the existing body of empirical literature on fundamental analysis in the context of Karachi Stock Exchange. The organization of the study is as follows. The next sections the literature review, which would be followed by the empirical methodology deployed in this paper. The fourth section contains the empirical results, while the conclusion is presented in the last section.

2. LITERATURE REVIEW

Gunasekarage et al. (2004) studied the relationship between Colombo Stock Exchange and macroeconomic variables such as inflation measured by CPI, exchange rate measure by USD to SLR rates, interest rates measure by treasury bill rate and money supply measured by M2 monetary aggregate. They argue that assessing the impact of macroeconomic variables on stock markets can facilitate us in discounting any security and that would ultimately shape our investment decision. The study considered the monthly data for the time period of January 1985 to December 2001. The authors applied vector error correction model (VECM) for studying the relationships. The study found that inflation, money supply and interest rate were related to stock market while exchange rate does not impart any impact on stock market.

KARAGÖZ et al. (2009) studied the impact of macroeconomic variables on Istanbul Stock Exchange- Turkey's prime equity market. They have used five macroeconomic factors in their study- inflation given by CPI, money supply given by M1 monetary aggregate, exchange rate given by RDK (foreign currency basket), interest rate presented by 3 months interbank rate and real economic activity given by industrial production index. The sample period in the study is from December 1998 to December 2008. For studying the relationship, the study deployed Johansen Co-integration analysis. The results confirmed that apart from industrial production, all variables imparts significant impact on Istanbul Stock Exchange.

Hussainey and Ngoc (2009) investigated the relationship between Vietnamese stock prices and interest rates, industrial production and US stock prices. The considered time period in this study is from January 2001 to April 2008 and Vector Autoregression (VAR) methodology is deployed in the paper for developing the causal relationship. The results found that domestic macroeconomic factors exert much more pressure on Vitnamese Stock market as compare to the factors related to other economies.

Osamwonyi and Osagie (2012) studied the relationship between Nigerian Stock Market and macroeconomic factors such as inflation, interest rate, exchange rate, fiscal deficit and economic growth from 1975 to 2005.The authors used VECM for studying the relationship. The utility of VECM techniques becomes very high if a low frequency data, as in this study, is encountered. The results revealed that interest rate and inflation were not significant at 5%, whereas all the other variables were having highly significant impact. The authors conclude that it is very necessary for Nigerian policy makers to avoid increasing money supply rapidly as it is not going to enhance the investment activity in the context of Nigeria.

Vejzaic and Zarafat (2013) investigated the relationship between Malaysian stock market and certain macroeconomic factors. The included factors were inflation, interest rate, money supply and exchange rate. The time period of the study was September 2006 to September 2012. In order to explore the relationships between variables, the study used VECM. The study found that all variables except inflation are related to Malaysian stock market.

3. DATA AND METHODOLOGY

As we are applying the Saloman Brothers model in this article, hence we have used the factors and their proxies as suggested by Sorensen et al. (1989). In accordance to it, we are using KSE 100 index returns for stock market activity, 3 month KIBOR Rates returns for short term interest rates, 1 year KIBOR Rates returns for long term interest rates, 3 year KIBOR Rates returns for business cycle, Large Scale Manufacturing index for economic growth and USD to PKR rate for foreign exchange rate. The time period considered in this study is from January 2003 to December 2012. We have used monthly data for all the variables that are used in this study and at log levels, so that the variance in the model can be stabilized.

Multivariate time series methods are widely used by empirical economists and econometricians have focused a great deal of attention at refining and extending these techniques so that they are well suited for answering economic questions. Vector auto regressions (VARs) were introduced into empirical economics by Sims (1980), who demonstrated that VARs provide a frame- work for analyzing economic time series. The vector auto regression (VAR) model is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model to dynamic multivariate time series. The VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting. It often provides superior forecasts to those from univariate time series models Forecasts from VAR models are quite flexible because they can be made conditional on the potential future paths of specified variables in the model.

This is just a multiple time-series generalization of the AR model. The VAR model is easy to estimate because we can use the OLS method Vector Autoregression (VAR) model is an extension of univariate auto regression model to multivariate time series data. VAR model is a multi-equation system where all the variables are treated as endogenous. There is one equation for each variable as dependent variable. Right-hand side of each equation includes lagged values of all dependent variables in the system. VAR (p) model can be represented as follows.

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + ... + A_p y_{t-p} + e_t,$$

where c is a $k \times 1$ vector of constants (intercepts), A_i is a time-invariant $k \times k$ matrix and e_t is a $k \times 1$ vector of error terms.

Impulse response functions, or IRFs, measure the effects of a shock to an endogenous variable on itself or on another endogenous variable. IRF of a dynamic system is its output when presented with a brief input signal, called an impulse. More generally, an impulse response refers to the reaction of any dynamic system in response to some external change. A VAR is written in vector MA (∞) form as:

$$y_t = \mu + \varepsilon_t + \Psi 1 \varepsilon t-1 + \Psi_2 \varepsilon_{t-2} + \cdots.$$

Thus, the matrix Ψs has the interpretation $\partial y_{t+s}/\partial \varepsilon'_t = \Psi_s$. It's a change of a shock with respect to time.

Alike IRF, variance decomposition is also a technique which helps us in interpretation of VAR model. The variance decomposition indicates the amount of information each variable contributes to the other variables in the auto regression. It determines how much of the forecast error variance of each of the variables can be explained by exogenous shocks to the other variables.

$$V(Y) = E[V(Y / X)] + V[E(Y / X)]$$

The above equation can be represented in terms of ANOVA as

$$SST = SSR + SSE$$

The estimates of VAR obtained and further treatment of VAR estimates through IRF and variance decomposition would ultimately accomplish our analysis.

4. EMPIRICAL RESULTS

In order to apply Saloman Brothers approach through VAR model, we have to first of all verify the stationary properties of all the series that are used in this series. We have considered all the variables at log level. Table 4.1 below shows the results of ADF unit root test proposed by Dickey and Fuller (1981) statistics for all the series that are used in the analysis.

Table 4.1
ADF Statistics

Variables	Augmented Dicky Fuller Statistics
CPI	-4.0641
KSE	-3.968
FX	-3.2096
Business Cycle	-4.4034
Long Term Rate	-4.8818
Short Term Rate	-6.9558
LSM	-5.9923

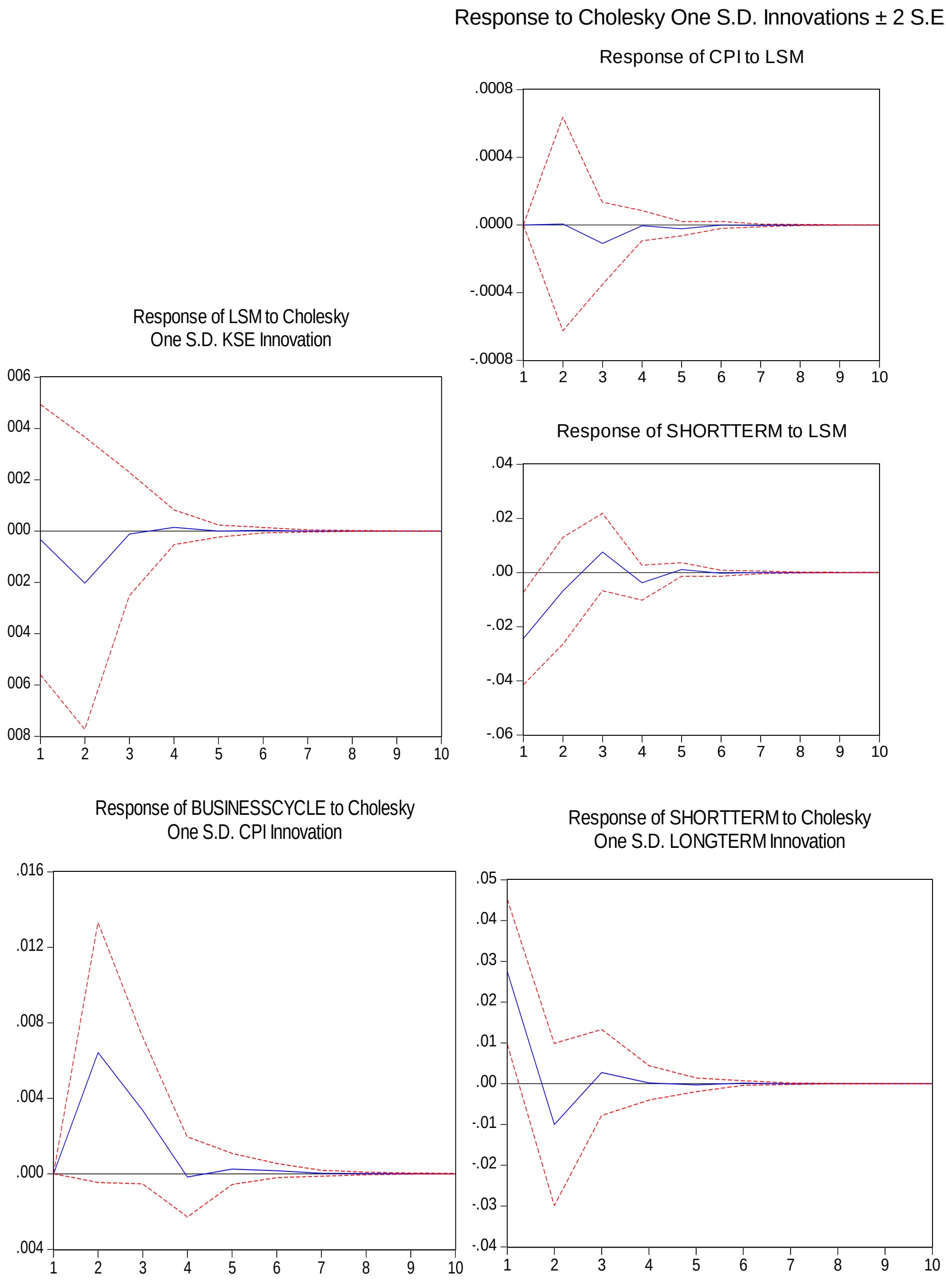
Since, all the variables are stationary at level; hence we can proceed for VAR analysis. VAR is basically a reduced form model and it does not require conditions such as cointegration necessarily to be satisfied. Table 4.2 below show the estimates of VAR model.

Table 4.2
VAR Estimates of Saloman Brothers Model

	FX(-1)	CPI(-1)	B.Cycle (-1)	Long Rate (-1)	Short Rate (-1)	KSE (-1)	LSM (-1)
FX	0.3144531*	0.1906658	-0.03379823	0.02732844	-0.00017827	-0.0203380	-0.0265602
CPI	0.1074903	0.1368377	-0.03193234**	0.02396781	0.003883216	0.00409479	0.00369168
Business cycle	0.4365777	1.9108874**	0.370063184	-0.3875624	0.134058077*	-0.0174906	-0.0390816
Long term rate	0.4772233	1.8277474	0.554922595	-0.49986001	0.156849048*	0.00625249	-0.0828693
Short term rate	0.6027976	1.2244438	-0.18817162	-0.01919349	-0.45391344*	-0.0728762	-0.6533644
KSE	0.4993309	-0.380598	-0.58584124	0.46528064	0.016964905	0.18536839	0.25048545**
LSM	-0.488065	-2.019216**	0.006996997	0.10295392	0.0476128***	-0.0665846**	-0.2075023**

We have found that Manufacturing sector have a positive and significant impact. Also KSE itself has a positive and significant impact. It is interesting to note that short term interest rate is positively and inflation negatively is linked with manufacturing index. Furthermore, business cycle is also significantly related with inflation. Similarly, long term interest rate is also linked with short term interest rate. Now if we construct a wholesome view regarding KSE on the basis of the results obtained, then we can say that business cycle imparts its impact on inflation and inflation later on responds to LSM which ultimately affects KSE. Long term interest rates are intuitively related with short term interest rates which also imparts its impact on KSE by influencing the LSM index. Thus, we can say that all the variables, presented in Saloman Brothers model, except foreign exchange rate, influences KSE- 100 Index, either directly or indirectly.

Based on the above causal relationships, we are now presenting IRF’s for the selected causal relationships. The causal relationships which are chosen here are based upon the channel through which they affect KSE-100 Index.



The above IRF's are showing that all the variables except business cycles are achieving their long run equilibrium within six months of the shocks been applied. However, it takes slightly high for business cycles (around eight months) to fully become stable.

In order to further sharpen our conclusions, we are presenting the variance decomposition estimates below in table 5.3.

Table 5.3
Variance Decomposition Results

Period	FX	CPI	Business cycle	Long term rate	Short term rate	KSE	LSM
1	0.05100	0.00255	0.001834	0.006608	0.000311	0.93768	0
2	0.04710	0.01429	0.007002	0.016196	0.000404	0.88075	0.03424
3	0.04681	0.01877	0.009269	0.017002	0.000498	0.87253	0.03510
4	0.04681	0.01879	0.009276	0.017049	0.000528	0.87243	0.03510
5	0.04681	0.01879	0.009292	0.017048	0.000533	0.87241	0.03510
6	0.04681	0.01879	0.009292	0.017048	0.000533	0.87241	0.03510
7	0.04681	0.01879	0.009292	0.017048	0.000533	0.87241	0.03510
8	0.04681	0.01879	0.009292	0.017048	0.000533	0.87241	0.03510
9	0.04681	0.01879	0.009292	0.017048	0.000533	0.87241	0.03510
10	0.04681	0.01879	0.009292	0.017048	0.000533	0.87241	0.03510

The variance decomposition estimates are suggesting even more stability in the system as compared to the IRF. It is showing that shocks even do not last beyond four or five time periods. Thus, we can conclude that the VAR estimates obtained in this study are by and large stable.

5. CONCLUSION

In this study, we have tried to verify an important portfolio management model-Salomon Brothers Model on KSE-100 index. There are six determinants of stock market growth in this model. However except of growth rate, which is represented here by large scale manufacturing index, none of the other variable is explaining the movements in KSE-100 index significantly. However, indirect causation is observed in this study between KSE and other macroeconomic factors. The results confirmed that KSE significantly responds to the developments in macroeconomic arena. The study made an important contribution in the empirical literature by identifying the channels through which macroeconomic information affects KSE 100 index.

REFERENCES

1. Burmeister, E. and McElroy, M.B. (1988). Joint estimation of factor sensitivities and risk premiums for the arbitrage pricing theory. *Journal of Finance*, 43, 721-33.
2. Chen, N., Roll, R. and Ross, S.A. (1986). Economic forces and the stock market. *Journal of Business*, 59, 383-403.
3. Dickey, D.A. and Fuller, W.A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49, 1057-1072.
4. Gunasekarage, G., Pisedtasalasai, A. and Power, D.M. (2004). Macroeconomic Influence on the Stock Market: Evidence from an Emerging Market in South Asia. *Journal of Emerging Market Finance*, 2, 285-304.
5. Hussainey, K. and Ngoc, L.K. (2009). The impact of macroeconomic indicators on Vietnamese stock prices. *The Journal of Risk Finance*, 10, 321-322.
6. KARAGÖZ, K., ERGÜN, S. and KARAGÖZ, M. (2009). Impact of Macroeconomic Factors on Stock Market: Evidence from Istanbul Stock Exchange. *International Symposium on Sustainable Development*, 9, 122-130
7. King, B.F. (1966). Market and industry factors in stock price behavior. *Journal of Business*. 39, 139-190.
8. Maysami, R.C., Howe, L.C. and Hamzah, M.A. (2004). Relationship between Macroeconomic Variables and Stock Market Indices: Cointegration Evidence from Stock Exchange of Singapore's All-S Sector Indices. *Journal Pengurusan*, 2, 47-77.
9. Osamwonyi, I.O. and Osagie, E.I.E. (2012). The Relationship between Macroeconomic Variables and Stock Market Index in Nigeria. *Journal of Economics,* 3, 55-63.
10. Sims, C. (1980). Macroeconomics and Reality. *Econometrica*, 48, 1-48.
11. Sorenson, E., Mezrich, J. and Chee, T. (1989). The Salomon Brothers U.S. Stock Risk Attribute Model, Salomon Brothers.
12. Vejzagic, V. and Zarafat, H. (2013). Relationship between Macroeconomic Variables and stock Market Index: Co-Integration Evidence From Ftse Bursa Malaysia Hijrah Shariah Index. *Asian Journal of Management Sciences & Education*, 2, 94-108.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 143-150

VOLATILITY SPILLOVERS ACROSS FINANCIAL MARKETS OF PAKISTAN

Syed Monis Jawed, Syed Yasir Alam and **Junaid Shamshad**
Department of Statistics, University of Karachi, Karachi, Pakistan
Email: monis_jawed@hotmail.com

ABSTRACT

Asset allocation decision is a major task in portfolio management. Effective asset allocation leads to successful diversification which eventually lowers down the exposure of risk in a certain portfolio. One of the most popular ways of doing diversification is to make investments across different financial markets or assets. In this paper too, we are following the above defined diversification technique. Most of the previous research work in this area is concerned with observing the relationships between or among the asset returns of different financial markets or assets. However, in this paper we have investigated the relationships among different financial markets by considering the risk factor among them. In this paper, we are considering the relationship among the risk factor of stock index, money market rate and gold prices in the context of Pakistan. The time period which is considered in this study is from January 2000 to December 2012. For figuring out the risk factor, volatility of the said markets is estimated in a multivariate framework. A dynamic conditional correlation (DCC) model, which comes under the category of multivariate GARCH models, is applied. The results revealed that the risk factor prevailing in money market is linked with the risk factor in stock and commodity market, thus in order make any portfolio diversify, money market assets should not be combine with stock or commodity market assets'.

KEYWORDS

Volatility Spillovers; Capital Market; Money Market; Hedging; Diversification.

1. INTRODUCTION

Financial Markets act as a channel of capital accumulation in any economy. These markets are the prime avenues for making investments. Thus, the public funds or national savings can be transferred to Government or producers in the economy via financial market. Putting another way around, financial institutions broadly and financial markets specifically enable the deficit sector of an economy to relate with the surplus sector of an economy.

Financial markets can be classified into various categories. The categorization can be made through various perspectives such as mode of trading, term or maturity and the mode of offering. These financial assets are usually characterized from the perspective of their designated markets. On the basis of maturity, financial assets are widely classified as money market securities and capital market securities, where the latter category refer to the securities having higher maturity and the former category is the vice versa.

Assets are generally allocated in different financial markets by portfolio managers and investment analysts. The prime reason for such allocation is based upon the principal of diversification. Asset and liability structure or liquidity needs is another important factor which induces investors to hold assets belonging to different financial markets. This practice has actually motivated many researchers in recent years to investigate the relationship between the two financial markets. Anderson et al (2003) noted that there exists a strong contempourous relationship between intra country financial markets. Similarly, Rigobon and Sack (2003) investigated the relationship between US stock, bond and money market. Similarly, Ehrmann et al. (2011) investigated the relationship between the money markets and stock markets of European Union region. Nevertheless, a large body of literature has investigated the relationship between stock market and macroeconomic factors in various contexts and money market dynamics are often found embedded in macroeconomic factors.

Financial assets are mostly described on the basis of two attributes- risk and return. Risk is actually refereed as the chances of achieving less than the expected return. Many treat risk as an opportunity cost of getting higher returns. It implies that investors would only divert their resources to any financial asset having excessive risk, if the prospects of earning returns are higher. For modeling risks in financial assets, autoregressive conditional hetroskedastock models (ARCH) proposed by Engle (1982) and later on extended by Bollerslev (1986) has remained very popular among financial econometricians during the last three decades. This study also figures out the risk factor using the multivariate extension of ARCH type model.

In this study, we are actually exploring that how the risk factor generated in one financial market can be transmitted to another financial market within the economy. Furthermore, we are interested in figuring out the capacity of gold as a hedge for investors of financial markets in Pakistan. To the best of our knowledge and belkief, no study thus fara has investigated the volatility spillovers for Pakistani financial markets with the inclusion of a hedge factor. Thus, this paper is filling a major gap in the empirical literature. The organization of this study is as follows. The next section covers the review of selected studies, which would be followed by the section describing the data and empirical methodology applied in this paper. The fourth section presents the empirical results computed and finally the last study is concluded in the last section.

2. LITERATURE REVIEW

Rigobon and Sack (2003) investigated the interaction between various financial markets in the context of US. They have considered money, equity and bond market in their study. For observing the volatility spillover amongst the markets; they have deployed structural GARCH model. The paper finds that there exists a contempourous relationship between financial markets of United States. They suggested that figuring out volatility spillovers is necessary for managing risk effectively.

Song (2009) investigated the volatility spillovers across equity markets of greater China region. The stock markets which are included in the study were Shanghai Stock Exchange, Shenzhen Stock Exchange, Hong Kong Stock Exchange, Singapore Exchange and Japan Exchange Group. The study found that stock markets within China depict greater volatility spillovers as compare to the volatility spillovers between stock markets of different countries.

Baur and Lucey (2010) explored the interactions between stock market and bond market of developed countries including US, UK and Germany. They further examined that if gold can proof to be a safe heaven or hedge for risks which emanate from investing in capital markets. The authors used asymmetric GARCH approach for exploring the relationship. The study found that gold can served as a safe haven for stock markets however for bond market this safe heaven characteristic of gold is not evident.

Ehrmann et al. (2011) studied the interaction between equity, bond, money and foreign exchange market of United States and European Union. The interactions between the markets were observed in multiple contexts. This study investigates the influence of domestic markets on each other and the influence of international financial markets- both asset and cross asset, on domestic financial markets. The results found that international markets though have power to explain the dynamics of local market, but markets can best been explained by domestic markets as they are all part of the similar macroeconomy.

Mighri and Mansouri (2013) studied the volatility spillovers US equity market and the equity markets of some sixteen developed and emerging countries. The main research objective of their study was to ascertain the financial contagion, that is, the intensity of volatility spillovers become higher in episodes of financial crisis. They considered time period in their study was from 2007 to 2013. The study shows positive evidences in favor of financial contagion, which means that equity markets around the world respond to the adverse developments in their economy sharply.

3. DATA AND METHODOLOGY

In this study we are measuring the capital market activity through KSE-100 Index, money market activity through KIBOR rates and gold prices for the hedge factor. The analysis has been made at monthly frequency and the time period which is considered here is from January 2000 to December 2012. We have used the log difference of all the variables here in order to assure the stationary properties of the variables used in this paper.

In the contemporary global finance, it appears imperative both for asset management and for risk analysis to gain a better understanding of the changing correlations between a large number of assets and even between different financial markets. Hence multivariate GARCH is a logical choice for observing the spillovers across different financial markets and assets. Many multivariate GARCH models such as VECH, Constant Conditional Correlations (CCC), BEKK, DCC and O-GARCH been developed for observing the volatility spillovers. In this paper, we are applying the DCC model proposed by Engle and Shephard (2001) and Engle (2002). The major advantage of using this technique over other multivariate GARCH models is the fact that it has least number of parameters as compared to other models.

Suppose we have n assets in the portfolio so the return vector is the column vector that is

$$r_t = (r_{1t}, r_{2t}, \dots, r_{nt})' \tag{1}$$

with the assumption that is $r_t|\mathcal{F}_{t-1}\sim N(0,H_t)$ means it has mean E[$r_t|\mathcal{F}_{t-1}$]=0 and variance that is $E[r_t' r_t|\mathcal{F}_{t-1}] = D_t R_t D_t = H_t$ which also implies that

$$r_t = H_t^{\frac{1}{2}} z_t \tag{2}$$

where

$$z_t = (z_{1t}, z_{2t}, \dots, z_{nt})' \sim N(0,1) \tag{3}$$

We can obtain $H_t^{\frac{1}{2}}$ used in equation 2 by Choleskey decomposition.

Since we know that in general $\frac{COV_{i,j}}{\delta_i \delta_j}$, in DCC model for n assets we can write it as

$$\frac{H_t}{D_t D_t} = R_t \tag{4}$$

or

$$H_t = D_t R_t D_t \tag{5}$$

where, H_t: time varying conditional co-variance matrix of the process r_t where

$[H_t]_{i,j} = h_{i,t} h_{j,t} \rho_{ij,t} i \neq j$, where 1 ≤ i, j ≤ n

and D_t is the time varying diagonal matrix of standard variation from univariate GARCH processes.

$$\begin{bmatrix} \sqrt{h_{1t}} & 0 & \cdots & 0 \\ 0 & \sqrt{h_{2t}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{h_{nt}} \end{bmatrix} \tag{6}$$

However R_t is the correlation matrix of the standardized disturbance that is.

$$\varepsilon_t = \frac{r_t}{D_t} \sim \mathrm{N}\,(0,R_t) \tag{7}$$

$$R_t = \begin{bmatrix} 1 & q_{12,t} & q_{13,} & \cdots & q_{1n,t} \\ q_{21,t} & 1 & & & q_{2n,t} \\ q_{31,t} & & 1 & & q_{3n,t} \\ \vdots & & & \ddots & \vdots \\ q_{n1,t} & q_{n2,t} & q_{3n,t} & \cdots & 1 \end{bmatrix} \tag{8}$$

Before analyzing further, we have to assure that $H_t$is positive definite covariance matrix. SinceH_t is in quadratic form and in order to ensure positiveH_t, R_t has to be positive definite. .Hafneer and Franses (2003) noted that in conditional correlation matrix all the elements have to be equal or less than 1. If we have both these point followed than R_t is decomposed into

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1} \tag{9}$$

where Q_t is a positive definite matrix defining the structure of the dynamics and Q_t^{*-1} is simply the diagonal inverted matrix with the square root of the diagonal elements of Q_t.

$$Q_t^{*-1} = \begin{bmatrix} 1/\sqrt{q_{11t}} & 0 & \dots & 0 \\ 0 & 1/\sqrt{q_{22t}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1/\sqrt{q_{nnt}} \end{bmatrix} \tag{10}$$

Now we can write the DCC model with the help of a GARCH(1,1) process of "Variance Targeting" approach of Engle and Mezrich (1996). By using GARCH(1,1) we can reduce not only the number of parameters as well as make the computation easier. To illustrate how its work we can use it as deviations from the unconditional variance from GARCH (1,1) model:

$$h_i - \bar{h} = w - \bar{h} + \alpha\left(e_{t-1}^2 - \bar{h}\right) + \beta\left(h_{t-1} - \bar{h}\right) + \alpha\bar{h} + \beta\bar{h} \tag{11}$$

By re-arranging we have

$$h_t = w - [1 - \alpha - \beta]\bar{h} + [1 - \alpha - \beta]\bar{h} + \alpha e_{t-1}^2 + \beta h_{t-1} \tag{12}$$

Now if $w = [1 - \alpha - \beta]\bar{h}$ then equation above can be written as

$$h_t = [1 - \alpha - \beta]\bar{h} + \alpha e_{t-1}^2 + \beta h_{t-1} \tag{13}$$

It is not only easier to compute but also say that un-conditional variance is $\bar{h} = \frac{w}{1-\alpha-\beta}$. This simply work if $\alpha + \beta < 1$ and if the weights are $\alpha > 0, \beta > 0$.

Now we can write DCC model with the help of a GARCH (1, 1) "targeting variance" approach that we have done above ,so the model will be

$$Q_t = [1 - \alpha - \beta]\bar{Q} + \alpha\varepsilon_{t-1}\varepsilon_{t-1}' + \beta Q_{t-1} \tag{14}$$

4. EMPIRICAL RESULTS

For pursuing the DCC nalysis, we are firstly observing the variances into a time invariant context. Table 1 below shows the correlations among the variables.

Table 1
Correlation Matrix

	KIBOR	STOCK	GOLD
KIBOR	1		
STOCK	0.02148007	1	
GOLD	-0.001165977	-0.06011840	1

We can observe that the correlations between kibor and stock is 0.02148007(2%), between kibor and gold is -0.001165977(0.01%),and between gold and stock is -0.060118400(6%). The results gives us the intuition that the commodity market in Pakistan is not yet linked with money market and capital market as the correlation signs are negative. As the correlation between capital market and money market is very low, therefore we cannot conclude that allocating assets amongst the two markets is not useful. In order to carry out the DCC analysis, we have to satisfy the stationary properties of the data. For this purpose we are using the ADF unit root test. Table 2 below contains its estimates.

Table 2
ADF Estimates

VARIABLES	STATS	P-VALUE
KIBOR	-5.5049	0.01
STOCK	-4.1561	0.01
GOLD	-5.2627	0.01

We found that all the variables are having the same p-value that is 0.01 and thus, we can reject the null hypothesis of unit root, which implies that the all the variables are stationary at highest level of significance. The next step in our analysis is to estimate the volatility through a univariate technique. For that purpose, we have used the most parsimonious volatility model- GARCH (1,). Table 3 below carries its estimates.

Table 3
GARCH (1,1) Estimates

Variables	ω	α	β	$\alpha+\beta$
KIBOR	0.0004919122	0.445565366	0.5544346330	0.999999999
KSE	0.000463341	0.111476038	0.832123535	0.943599573
GOLD	0.0004308041	0.1495645668	0.5529852466	0.702549813

The GARCH(1,1) parameters showing us that we have the persistency in the model because the all the three GARCH(1,1) models are having the values $\alpha+\beta<1$ and $\alpha>0$, $\beta>0$ which satisfies the axiom of positive definiteness for GARCH as well. We have obtained higher values of $\alpha+\beta$ for capital market and money market which is consistent with the fact that volatility in financial markets generally have long memory, however $\alpha+\beta$ factor is much lower for the hedge series. Now, we have sufficient inputs for computing the DCC estimates, the major objective of our study. The estimation would have immense importance for understanding the market dynamics due to the fact that correlation anticipation among different financial assets or markets can facilitate us in figuring out the price convergence or divergence between them as noted by Chevallier (2011). Table 4 contains the DCC estimates.

Table 4
DCC Estimates

Variables	KIBOR	STOCK	GOLD
ω	0.0005038083	0.0006999582	0.0004458368
α	0.5427029	0.069835714	0.149911649
β	0.5519457	0.8405710	0.5904607
DCC α	7.803864e-09		
DCC β	8.889324e-01		

We find out that in DCC model the variable kibor- the measure of money market- have the highest value of α indicating that the KIBOR is the variable which is creating large volatility spillovers to both the corresponding variables, that is, gold hedge and capital markets. Now when we look upon the β values of the variables we have found among all the variables the stock; indicator of capital markets, have the largest values of β, implying that among all the variable stock is the variable which is more dependent on its prior observations. The DCC α and β is simply to check the DCC constraint that is either the covariance matrix is positive definite or not that is having all the values < 1. By the help of obtained parameters, we can construct conditional correlation among all variables which may help us in our asset allocation decision. Figure 1 below contains DCC plots of the series which are considered in this study.

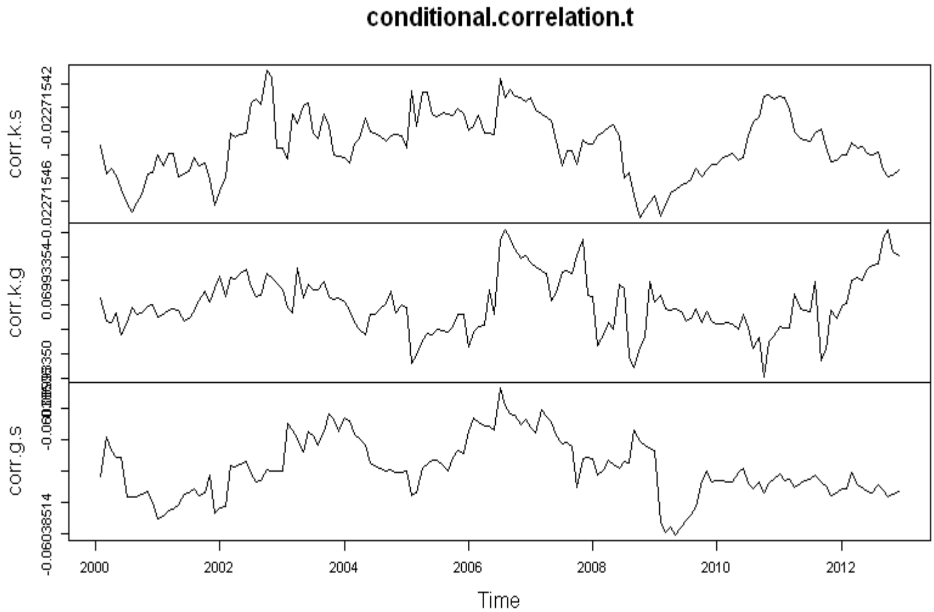


Figure 1: Dynamic Conditional Correlation between Assets

We can observe in the above plot that correlation between K (money market) and S (capital markets) ,and between K (money market) and G (hedge) is increasing overtime, which hints that such diversification might not deliver us in managing the risk factor in the portfolio. Whereas, holding a portfolio having with the combination of gold (hedge) and stock (capital market) may well diversify compare to the other portfolios as its conditional correlation is decreasing with respect to time.

5. CONCLUSION

In this technical report, we have analyzed three different markets and assess how the asset allocation could take place among them, keeping in view the risk factor. The DCC results reveals to us that volatility prevailing in interest rate has a great impact on the volatility present in the other two markets. The plot of DCC also shows that over the period of time the stock market and the gold hedge have become more connected with the money market in terms of risk. It implies that if any investor is making any portfolio with allocating his assets in money and gold hedge simultaneously or money and capital market simultaneously then he cannot achieve the benefits of diversification as risks increases at both places simultaneously. Thus, a better diversified portfolio can be formed by investing in stock and gold hedge. The analysis presented in this paper can further be extended by changing the proxies used for money and capital market and hedge.

REFERENCES

1. Andersen, T. Bollerslev, T. Christoffersen, P. and Diebold, F. (1998). Practical Volatility and Correlation Modeling for Financial Market Risk Management, in *Risks of Financial Institutions*, University of Chicago Press, 513-548.
2. Baur, D.G. and Lucey, B.M. (2010). Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold. *The Financial Review*, 45, 217-229.
3. Beach, S.L. and Orlov, A.G. (2007). An application of the Black-Litterman Model with EGARCH-Derived Views for International Portfolio Management. *Financial Markets and Portfolio Management*, 21, 147-166.
4. Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics,* 31, 307-327.
5. Chevallier, J. (2011). Anticipating correlations between EUAs and CERs: a Dynamic Conditional Correlation GARCH model. *Economics Bulletin*, 31, 255-272.
6. Ehrmann, M., Fratzscher, M. and Rigobon, R. (2011). Stocks, Bonds, Money Markets and Exchange Rates: Measuring International Financial Transmission. *Journal of Financial Econometrics,* 26, 948-974.
7. Engle, R.F. (2002). Dynamic conditional correlation - A simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, 20, 339-350.
8. Engel, R.F. and Mezrich, J. (1996). GARCH for groups, *Risk*, 9, 36-40.
9. Engle, R.F. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH, UCSD DP 2001-15.
10. Hafner, M.C. and Franses, H.P. (2003). A Generalized Dynamic Conditional Correlation Model for Many Asset Returns (Econometric Institute Report EI 2003-18), Erasmus University Rotterdam, Eraasmus School of Economics (ESE), Econometric Institute.
11. Mighri, Z. and Mansouri, F. (2013). Dynamic Conditional Correlation Analysis of Stock Market Contagion Evidence from the 2007-2010 Financial Crises. *International Journal of Economics and Financial Issues,* 3, 637-661.
12. Song, X. (2009). *Multivariate GARCH Models for the Greater China Stock Markets*, Singapore Management University, Singapore.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 151-160

IMPACT OF AGE, WEIGHT ON BLOOD PRESSURE WITH A SPOTLIGHT ON SEX DIFFERENCES

Shahla Siddiqui[1], Sara Wahab[2] and **Sumrat Rasheed[3]**
[1] Education and Literacy Department Sindh, Karachi, Pakistan
[2] Research Department, Jinnah Sindh Medical University, Karachi, Pakistan
[3] Department of Statistics, University of Karachi, Karachi, Pakistan

ABSTRACT

A present study explores the application of impact of age, gender and weight on blood pressure .A cross sectional study of 100 adult hospital patients suggests a linear relationship between age, weight and systolic blood pressure, from age 18 to 74 years. The data will be used to developed suitable predication model to determine blood pressure of 18 to 74 year with a focus on sex differences.

Arterial systolic blood pressure, weight, age values followed a normal distribution, predication models will be developed to predict blood pressure based upon age groups and weight with sex differences. In this paper we shall discuss the regression modeling for quantitative response outcomes. First step of study, the data of 100(50 male and 50 female) subject were collected. The data like age, weight, systolic blood pressure to the normally distributed.

Abbreviation:
Systolic blood pressure = SBP, Weight (wg), Gender = (female = 0, male = 1).

KEY WORDS

Systolic blood pressure, weight, age, gender, simple linear regression and multiple linear regression analysis.

1. INTRODUCTION

In order to investigate various variables impact blood pressure change. Simple and multiple linear regression analyses were used [1]. Multiple linear relationships occurred showing dependence of Systolic blood pressure on age, weight focus on gender. High blood pressure is a common problem that afflicts both men and women. Blood pressure is well known to increase with age[2-4], and age has been thought to be an independent cause of the increase[5], several studies, however, show that this hypothesis is only valid in populations with a high intake of salt and fatty acids or a large body weight by age [6-10].

A number of reports have elucidated the association between different characteristics (age, serum cholesterol levels, weight, Smoking habits etc) and Systolic blood pressure change with gender [11-17]. The majority of these reports conclude that attained weight and weight change is the strongest predicator of absolute blood pressure level and blood pressure change. Although most studies on cardiovascular threat factors focus on men, sex differences have been shown in the effects of changes in weight and blood pressure on cardiovascular mortality [18]. However, to our awareness, no study has draw attention to on sex differences in the effects of initial body weight and weight change on blood pressure change. In this paper we had an opportunity to investigate a blood pressure changes on different variables (age, weight) [1].

2. MATERIAL AND TECHNIQUE

The data will be used to developed suitable predication model to determine blood pressure of 18 to 74 year with a focus on sex differences. The technique was divided in to four step. (1) expressive (2) modification (3) Predictors (4) Predication.

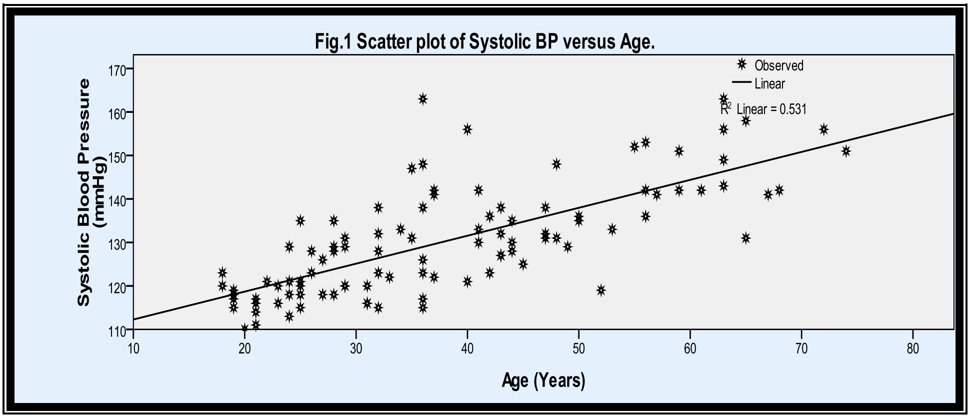
In this paper we shall discuss the regression modeling for quantitative response outcomes. First step of study, the data of 100(50 male and 50 female) subject were collected. The data like age, weight, systolic blood pressure to the normally distributed. The data measure were expressed in means and standard deviation ,it was clear from the data collected that the subjects studied were from a wide range of age starting form 18 years to 74 years in case of female subjects and 18 to 68 years in male[28]. Same interpretation was applicable to weight and systolic blood pressure also. Mean age of male subject was 38.50 years whereas in female it was 38.32 (Table-1a). The mean age of male was slightly greater. Similarly weight and systolic blood pressure was greater to female.

Second step of technique the data (N=100) on the age and the systolic BP were collected and we want to setup a linear regression model to predict SBP with age. Here we could, after checking the normality assumptions for both variables (age, BP), do a correlation =0.729, $p<0.001$ and a graphical scatter plot would be helpful (See Fig. 1a).

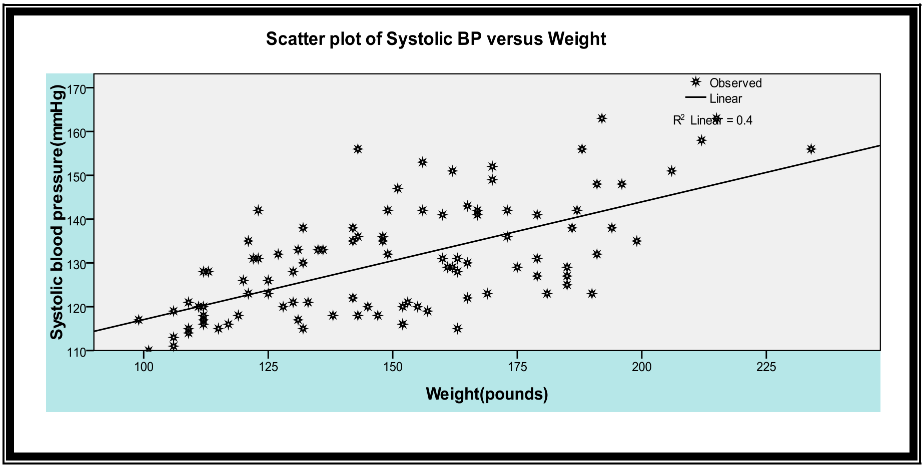
Third step of technique the data (N=100) on the weight and systolic BP were collected and we want to setup regression model to predict SBP with weight. Correlation =0.633, $p<0.001$) and a graphical scatter plot would be helpful (See Fig. 1b). Simple linear regression was used to compare the contribution of age and weight on systolic blood pressure. Age is more important predicator of systolic blood pressure. So we use multiple linear regression model is higher impact on blood pressure. Fourth step of technique we use a multiple regression model with both age and weight status correlation with systolic BP could be performed (Table-1b).

Table 1(a)
Descriptive Statistics

Gender		N	Range	Minimum	Maximum	Sum	Mean	Std. Deviation	Variance
Female=0	Age in year	50	56	18	74	1916	38.32	15.027	225.814
	Weight in given pounds	50	80	99	179	6478	129.56	20.242	409.721
	Systolic Pressure (mm Hg)	50	46	110	156	6479	129.58	12.259	150.289
Male=1	Age in year	50	50	18	68	1925	38.50	13.871	192.418
	Weight in given pounds	50	103	131	234	8529	170.58	23.036	530.657
	Systolic Pressure (mm Hg)	50	48	115	163	6573	131.46	13.139	172.621
Total	Age in year	100	56	18	74	3841	38.41	14.388	207.012
	Weight in given pounds	100	135	99	234	15007	150.07	29.839	890.349
	Systolic Pressure (mm Hg)	100	53	110	163	13052	130.52	12.677	160.717



(Fig.1a)



(Fig.1b)

Table 1(b)
Gender Wise Impact

		N	r	R^2
Female=0	Systolic Pressure (mm Hg) on Age	50	0.844	0.712
	Systolic Pressure (mm Hg) on Weight	50	0.859	0.738
	Systolic Pressure (mm Hg) on age and Weight	50	0.897	0.805
Male=1	Systolic Pressure (mm Hg) on Age	50	0.620	0.384
	Systolic Pressure (mm Hg) on Weight	50	0.763	0.582
	Systolic Pressure (mm Hg) on age and Weight	50	0.891	0.794
Total	Systolic Pressure (mm Hg) on Age	100	0.729	0.531
	Systolic Pressure (mm Hg) on Weight	100	0.633	0.401
	Systolic Pressure (mm Hg) on age and Weight	100	0.802	0.643

3. RESULT AND DISCUSSION

Overall, it appears that older adults have higher systolic blood pressure levels. Variability in blood pressure levels seems also to be greater for older patients, suggesting that a transformed data set may yield a better linear relationship. The relationship appears somewhat weaker when only the male patients are included, with older patients exhibiting greater variability in blood pressure readings (See Fig. 2a).

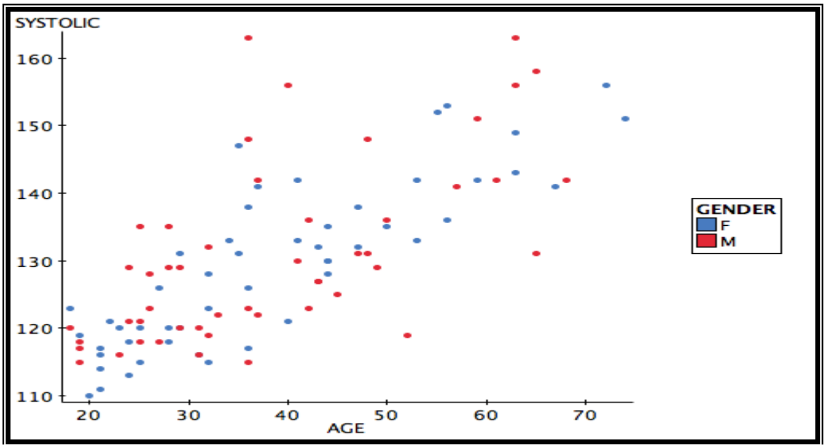


Fig. (2a) Age vs. Systolic by Gender

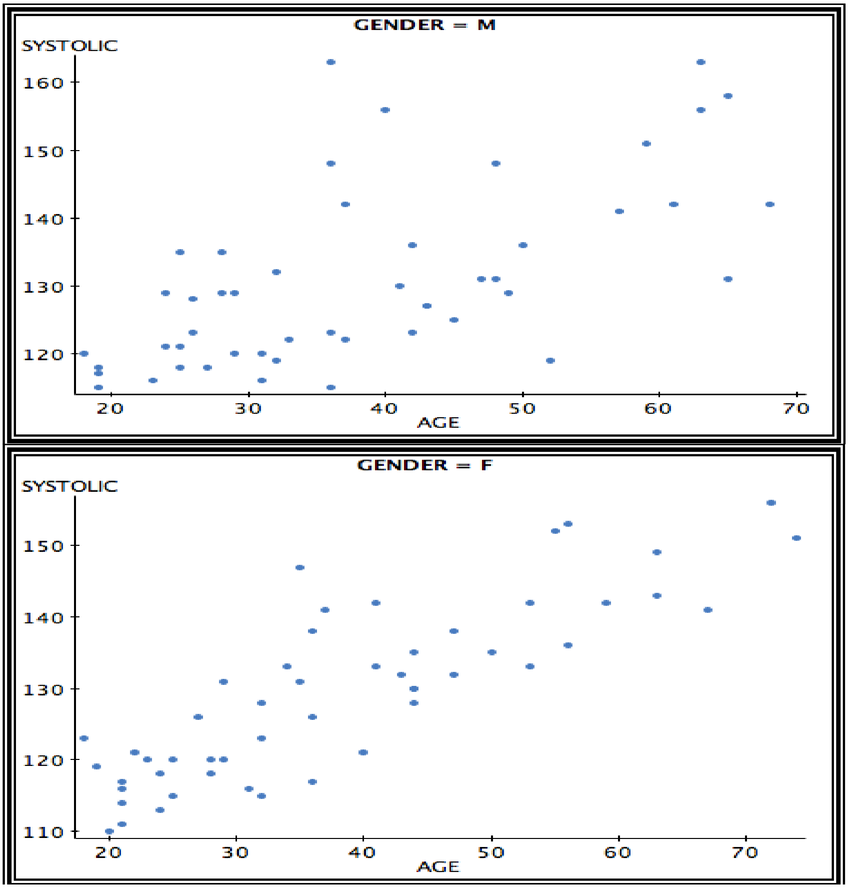


Fig. (2b) Age vs. Systolic (Male Female)

But, a stronger correlation exists among the female patients in the group. In this subset, the variance across the ages appears to be more consistent. Above are the calculated values of the age-systolic blood pressure correlation coefficient for male patients and for female patients (See Fig.2b and Table 1b). There's a moderately strong correlation between age and systolic blood pressure but how could we 'quantity 'this strength. We Apply simple linear regression model to relate SBP with age and weight well be (Table-2). The regression estimate (β_1) and constant (β_0) will be derived from the data (using method of least square [29]). Here the correlation between SBP and age is given($r=0.729$), $R^2=0.531$ which implies that only 53.3% of the systolic blood pressure is explained by the age of a person. So that only 71.2% of the SBP is explained of a female and 38.4% of the SBP is explained by the age of a male. Variability in blood pressure levels seems to be greater female patients. Also SBP and weight is given by ($r=0.6331$), $R^2=0.401$ which implies that only 40.1% of the systolic blood pressure is explained by the weight of a person. We analyze that risk factor age is most important variable as compare to weight. (Table-1b).

Table 2

Model		Coefficients	95.0% Confidence Interval for B	
		β	Lower Bound	Upper Bound
Total	(Constant)	105.849	100.893	110.81
	Age in year	0.642	0.521	0.763
	Systolic Pressure (mm Hg) on Age	$Y_{SBP} = 105.849 + 0.642_{age}$		
	(Constant)	90.181	80.091	100.271
	Weight in given pounds	0.269	0.203	0.335
	Systolic Pressure (mm Hg) on Weight	$Y_{SBP} = 90.181 + 0.269_{weight}$		
	(Constant)	87.644	79.804	95.485
	Age in year	0.490	.370	.609
	Weight in given pounds	0.160	.103	.218
	Systolic Pressure (mm Hg) on age and Weight	$Y_{SBP} = 87.644 + 0.490_{age} + 0.160_{weight}$		
Female=0	(Constant)	103.205	97.98	108.43
	Age in year	0.688	0.561	0.815
	Systolic Pressure (mm Hg) on Age	$Y_{SBP} = 103.205 + 0.628_{age}$		
	(Constant)	62.183	50.382	73.983
	Weight in given pounds	0.52	0.43	0.61
	Systolic Pressure (mm Hg) on Weight	$Y_{SBP} = 62.183 + 0.52_{weight}$		
	(Constant)	75.946	63.536	88.357
	Age in year	0.354	0.176	0.531
	Weight in given pounds	0.309	0.177	0.441
	Systolic Pressure (mm Hg) on age and Weight	$Y_{SBP} = 75.946 + 0.354_{age} + 0.309_{weight}$		
Male=1	(Constant)	108.84	100.027	117.653
	Age in year	0.588	0.372	0.803
	Systolic Pressure (mm Hg) on Age	$Y_{SBP} = 108.84 + 0.588_{age}$		
	(Constant)	57.187	38.788	75.587
	Weight in given pounds	0.435	0.329	0.542
	Systolic Pressure (mm Hg) on Weight	$Y_{SBP} = 57.187 + 0.435_{weight}$		
	(Constant)	60.302	43.673	76.93
	Age in year	0.318	0.137	0.499
	Weight in given pounds	0.345	0.237	0.454
	Systolic Pressure (mm Hg) on age and Weight	$Y_{SBP} = 60.302 + 0.318_{age} + 0.345_{weight}$		

Table 2 provides the quantification of the relationship between age and systolic BP. With every increase of one year in age, the systolic BP (on the average) increase by 0.642(95% CI 0.521 to 0.763) units<0.001.95% confident the mean value of age lies between this interval. Similarly we conclude that BP with increase every increase of one year in weight the SBP (on average) increase by 0.269(95% CI 0.2033 to 0.335). Higher weight and age both predict elevated blood pressure in adult hospital patients, but age was a more important predictor of systolic blood pressure than weight. We use multiple linear regression model was used to compare the contribution of age and weight on blood pressure.

There are significant moderate to high correlations between High with age and weight; we perform a multiple liner regression model. Correlation only enable us to check multi colinearity between any two variables; but sometimes a variable could be co-linear with a combination of others variables. If our interest is only to determine whether age affects systolic blood pressure after taking into the weight from (Table-2) we say that age is still statistically significantly affecting systolic BP (and the p-value of the weight is no interest) in gender.

We conclude that both age and weight are significant risk factors affecting the SBP. A weight has on the average 0.160 (95% CI 0.103 to 0.218) higher BP. A weigh has on the average 0.309 (95% CI 0.177 to 0.441) for female and average 0.345 (95% CI 0.237 to 0.454) male higher BP compared to a female (given the same age).

This paper presents additional finding of multiple linear regression relationships that existed in normotensive and hypertensive groups which rely on blood pressure as both dependent and independent variable. The equations provide information that aid in understand the relationships between the variables. We examined the linear dependency of blood pressure age weight on each other.

We found that age, gender and weight were positively correlated with blood pressure and there was a significant difference in blood pressure with respect to gender. The results of prediction model showed that in (Table-2).

REFERENCES

1. Wilsgaard, T., Schirmer, H. and Arnesen, E. (2000). Impact of body weight on blood pressure with a focus on sex differences: the Tromsø Study, 1986-1995. *Archives of Internal Medicine*, 160(18), 2847-2853.
2. Kotchen, J.M., McKean, H.E. and Kotchen, T.A. (1982). Blood pressure trends with aging. *Hypertension*, 4(5 Pt 2), III-128-III-134.
3. Flynn, M.A., Nolph, G.B., Baker, A.S. and Krause, G. (1992). Aging in humans: a continuous 20-year study of physiologic and dietary parameters. *Journal of the American College of Nutrition*, 11(6), 660-672.
4. Weinberger, M.H. and Fineberg, N.S. (1991). Sodium and volume sensitivity of blood pressure. Age and pressure change over time. *Hypertension*, 18(1), 67-71.
5. Lakatta, E.G. (1989). Arterial pressure and aging. *International Journal of Cardiology*, 25, S81-S89.

6. Appel, L.J., Moore, T.J., Obarzanek, E., Vollmer, W.M., Svetkey, L.P., Sacks, F.M. Bray, G.A., Vogt, T.M., Cutler, J.A., Windhauser, M.M. and Lin, P.H. (1997). A clinical trial of the effects of dietary patterns on blood pressure. *New England Journal of Medicine*, 336(16), 1117-1124.
7. Elliott, P., Stamler, J., Nichols, R., Dyer, A.R., Stamler, R., Kesteloot, H. and Marmot, M. (1996). Intersalt revisited: further analyses of 24 hour sodium excretion and blood pressure within and across populations. *Bmj*, 312(7041), 1249-1253.
8. Poulter, N.R., Khaw, K.T., Mugambi, M., Peart, W.S., Rose, G. and Sever, P. (1985). Blood pressure patterns in relation to age, weight and urinary electrolytes in three Kenyan communities. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 79(3), 389-392.
9. Huang, Z., Willett, W.C., Manson, J.E., Rosner, B., Stampfer, M.J., Speizer, F.E. and Colditz, G.A. (1998). Body weight, weight change, and risk for hypertension in women. *Annals of Internal Medicine*, 128(2), 81-88.
10. Pavan, L., Casiglia, E., Pauletto, P., Batista, S.L., Ginocchio, G., Kwankam, M.M., Biasin, R., Mazza, A., Puato, M., Russo, E. and Pessina, A.C. (1997). Blood pressure, serum cholesterol and nutritional state in Tanzania and in the Amazon: comparison with an Italian population. *Journal of Hypertension*, 15(10), 1083-1090.
11. Hubert, H.B., Eaker, E.D., Garrison, R.J. and Castelli, W.P. (1987). Life-style correlates of risk factor change in young adults: an eight-year study of coronary heart disease risk factors in the Framingham offspring. *American Journal of Epidemiology*, 125(5), 812-831.
12. Sparrow, D., Garvey, A.J., Rosner, B. and Thomas, H.E. (1982). Factors in predicting blood pressure change. *Circulation*, 65(4), 789-794.
13. Svärdsudd, K., Wedel, H. and Wilhelmsen, L. (1980). I. Factors associated with the initial blood pressure level and with the subsequent blood pressure increase in a longitudinal population study. *European Heart Journal*, 1(5), 345-354.
14. Sonne-Holm, S., Sørensen, T.I., Jensen, G. and Schnohr, P. (1989). Independent effects of weight change and attained body weight on prevalence of arterial hypertension in obese and non-obese men. *Bmj*, 299(6702), 767-770.
15. Meltzer, A.A., Mueller, W.H., Annegers, J.F., Grimes, B. and Albright, D.L. (1988). Weight history and hypertension. *Journal of Clinical Epidemiology*, 41(9), 867-874.
16. Sedgwick, A.W., Davidson, A.H., Taplin, R.E. and Thomas, D.W. (1983). Relationships between weight change and changes in blood pressure and serum lipids in men and women. *International Journal of Obesity*, 8(4), 343-353.
17. Borkan, G.A., Sparrow, D., Wisniewski, C. and Vokonas, P.S. (1986). Body weight and coronary disease risk: patterns of risk factor change associated with long-term weight change the normative aging study. *American Journal of Epidemiology*, 124(3), 410-419.
18. Selmer, R. and Tverdal, A. (1995). Body mass index and cardiovascular mortality at different levels of blood pressure: a prospective study of Norwegian men and women. *Journal of Epidemiology and Community Health*, 49(3), 265-270.
19. Njølstad, I., Arnesen, E. and Lund-Larsen, P.G. (1996). Body height, cardiovascular risk factors, and risk of stroke in middle-aged men and women a 14-year follow-up of the Finnmark study. *Circulation*, 94(11), 2877-2882.

20. Bønaa, K.H., Bjerve, K.S., Straume, B., Gram, I.T. and Thelle, D. (1990). Effect of eicosapentaenoic and docosahexaenoic acids on blood pressure in hypertension: a population-based intervention trial from the Tromsø study. *New England Journal of Medicine*, 322(12), 795-801.
21. Berchtold, P., Jörgens, V., Kemmer, F.W. and Berger, M. (1982). Obesity and hypertension: cardiovascular response of weight reduction. *Hypertension*, 4(5 Pt 2), III50-III55.
22. Rostrup, M. and Ekeberg, Ø. (1992). Awareness of high blood pressure influences on psychological and sympathetic responses. *Journal of Psychosomatic Research*, 36(2), 117-123.
23. Lerner, D.J. and Kannel, W.B. (1986). Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population. *American Heart Journal*, 111(2), 383-390.
24. Tverdal, A., Thelle, D., Stensvold, I., Leren, P. and Bjartveit, K. (1993). Mortality in relation to smoking history: 13 years' follow-up of 68,000 Norwegian men and women 35–49 years. *Journal of clinical epidemiology*, 46(5), 475-487.
25. Nyboe, J., Jensen, G., Appleyard, M. and Schnohr, P. (1991). Smoking and the risk of first acute myocardial infarction. *American Heart Journal*, 122(2), 438-447.
26. Daniels, S.R., Heiss, G., Davis, C.E., Hames, C.G. and Tyroler, H.A. (1988). Race and sex differences in the correlates of blood pressure change. *Hypertension*, 11(3), 249-255.
27. Deshmukh, T.R., Kuthe, A.M., Ingole, D.S. and Thakre, S.B. (2010). Prediction of femur bone geometry using anthropometric data of Indian population: a numerical approach. *International Journal of Medical Science*, 10(1), 12-18.
28. Kleinbaum, D.G., Kupper, F.L.L., Muller, K.E. and Nizam, A. (1998). *Applied Regression Analysis and other Multivariable Methods*. Pacific Grove, CA: Brooks / Cole.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 161-172

EVALUATION OF LIFESTYLE MODIFICATION OF TYPE 2 DIABETES MELLITUS (DM) RELATED KNOWLEDGE, ATTITUDE AND PRACTICES OF PAKISTANI PATIENTS

Kinza Tasadduq, Unza Shaikh, Areeba Ehsan, Bushra Zulfiqar, Aiman Fatima Naqvi and **Arif Ali**
Dow Medical College, Dow University of Health Sciences, Karachi, Pakistan

ABSTRACT

Introduction: The number of type-II diabetic patients continues to rise, primarily because of increases in the prevalence of sedentary lifestyle and lack of knowledge and practice regarding diabetes. The chronic nature of type-II Diabetes leads to devastating morbidities via the micro vascular and macro vascular damage that in the end leads to early mortalities. Type-II Diabetes Mellitus has both a genetic and an environmental etiology. Where the genetic cause cannot be controlled by human intervention, the environmental causes can be altered and negated to prevent diabetes or to lessen its morbidity. The key towards achieving optimum blood glucose seems to be an extensive and thorough approach to modify several risk factors all at a time.

Objective: To evaluate the knowledge, attitude and practice towards lifestyle modification in type2 diabetics.

Method: A cross-sectional study was done using the non-probability convenient sampling. The study was conducted between January 2015 and July 2015. Data was collected by a face to face interview employing a structured questionnaire from a total of 457 type-II diabetic patients attending outpatient departments (OPD) of Civil Hospital Karachi. Baseline characteristics of these patients were obtained after informed consent and their knowledge, attitude and practices regarding lifestyle modification were assessed.

Result: Among 457 participants more than half were males 62.8% and the mean age was 54 years with standard deviation of 10 years. Only 14.2% were reported to be current smokers and 10.1% being past smokers. More than half 60.8% had a positive family history of diabetes. BMI of 44.6% were in overweight range (BMI: >25 - <30). More than one-third 38.1% of the participants had controlled blood glucose levels. In this research 84.7% of the patients had good knowledge about their disease, 91.7% had a positive attitude towards their disease management while only 39.4% showed good practice regarding lifestyle modifications of diabetes.

Conclusion: Despite having good knowledge and positive attitudes about lifestyle modification of diabetes, they are not able to manage adequate practice of lifestyle modification. There is a critical need to develop diabetes educational programs that could help patients in diabetes management and support to change lifestyle modification.

1. INTRODUCTION

The prevalence of type 2 Diabetes Mellitus is increasing across the world [1], specially so in developing countries like Pakistan where changing lifestyle trends have not kept pace with the understanding of modifiable diseases like Diabetes. The difference in increase of prevalence of Diabetes between developed and developing countries further proves the above point. In developed countries, the increases in prevalence of Diabetes from 1995 to 2025 is estimated to be 27%, from 6.0% to 7.6%. In developing countries, the increases will be 48% from 3.3% to 4.9%. Another estimate of prevalence of diabetes suggests that more than 60% of the world's diabetics live in Asia, where most of the developing countries are to quote national data[2], the prevalence of diabetes in Pakistan is 6.8%[3]. By 2035, it is said that the number of people in Pakistan with Diabetes will rise to 12.8 million.

Type 2 Diabetes is a preventable disease and even when it has developed, it can be successfully managed by a combination of lifestyle modifications and pharmacological interventions.[4] The most important step in the management of Diabetes is patient education. Educating the patient is likely to be affective if the knowledge of the disease, attitude about it and practices that can lead to a healthier life are taught and enforced. Evidence suggests that self-care among patients has led to good diabetic control and prevention of other metabolic diseases. [5] Patients with good compliance to drugs but who failed to have good knowledge of the disease and failed to practice a healthy life style, had poor glycemic control. This shows that good knowledge, attitude and practice regarding the disease is just as important as taking medicine for it. Apparently, there exists a gap between knowledge and attitude towards diabetes between diabetic patients.[6] The patients may have good knowledge of diabetes, but poor attitude and fail to practice lifestyle modifications such as exercise, diet control and regular check-ups. Bad attitude and poor practices lead to uncontrolled diabetes with its own set of complications, which are usually known and feared by most of the un-practicing sufferers. Other metabolic diseases like hypertension, hypercholesterolemia and obesity may follow suit, or may be aggravated by diabetes, if already present. This is why diabetes is also called the "silent killer".

The aim of this study is to evaluate the lifestyle modifications of diabetics of Pakistan. We will also co-relate their demographic characteristics with their knowledge, attitude and practices of lifestyle modifications in diabetes.

2. MATERIAL AND METHODS

This was a cross-sectional study and subjects were selected through non-probability convenient sampling methods. Diabetic patients who fulfilled the inclusion criteria, attending the diabetic Out Patient Departments (OPD) of Civil Hospital Karachi and from January 2015 to July 2015 were included.

Inclusion Criteria:

- Patients giving consent for participation.
- Patients diagnosed with type 2 Diabetes.

Exclusion Criteria:

- Patients not giving consent
- Patients with gestational diabetes or any other form of diabetes
- Patients below 18 years of age

Study Tool: A questionnaire was used by the investigators to interview the study participants. It was also translated in Urdu to overcome language barrier. It consisted of four parts starting with demographics and diabetic characteristics. The weight and height measurements were taken via weight and height scales present at the Outpatient Departments. Participants' mode of management was determined as oral hypoglycemic, insulin, both or none. The diabetes control status was subjectively assessed from the patient's OPD record into controlled, partially controlled or uncontrolled. The family history of diabetes was also inquired along with comorbidities and complications of diabetes.

The second, third and fourth parts of the questionnaire template interrogated respectively about knowledge, attitude and practices of lifestyle modifications. Each part contained 8 close ended questions, each question was merited a score of '1' for correct answer and '0' for wrong.

Patient's knowledge, attitude and practice were calculated by adding their responses in terms of scores. Based on calculated scores the participants are divided into 2 groups. Scores below ≤50% were considered as poor score level while scores >50% were considered as good score level regarding knowledge, attitude and practices.

Sample Size: The Sample size was calculated using online openepi calculator, taking 7.4% of average knowledge[6] about benefits of exercise and weight loss for diabetic control, confidence limit of 2.4 % and confidence level of 95%. The calculated sample size was of 457 diabetic patients who were randomly included in the research.

Statistical Analysis: The collected responses were entered and analyzed by SPSS 21.0. Frequency and Percentages were calculated for all qualitative variables and Mean±SD were estimated for all quantitative variables. Association was checked through t-test and chi square test and significant level was considered at < 0.05.

Ethical Consideration: The participants were recruited to study interview after taking informed consent in Urdu language. The data collected was kept anonymous and confidential. The study took place after permission from the management of respective departments

3. RESULTS

General Demographics of the Participants

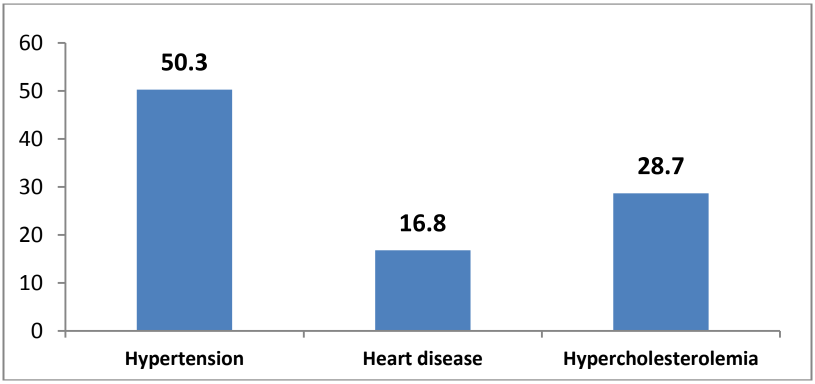
The demographic and diabetic characteristic of the participants is shown in Table 1 Out of the total, majority 287 (62.8%) of the participants were males and 37.2% were females. The mean age of the participants was 54 years and 380 (83.2%) were married.

The majority of the participants 204(44.6%) had BMI in the overweight range. Whereas the mean BMI was found to be 26.62±4.16 kg/m^2. The diabetes control status of the participants was found to be controlled in 38.1% and 32.2% had uncontrolled, while 29.8% had partially uncontrolled.

Table 1
General Demo Graphics

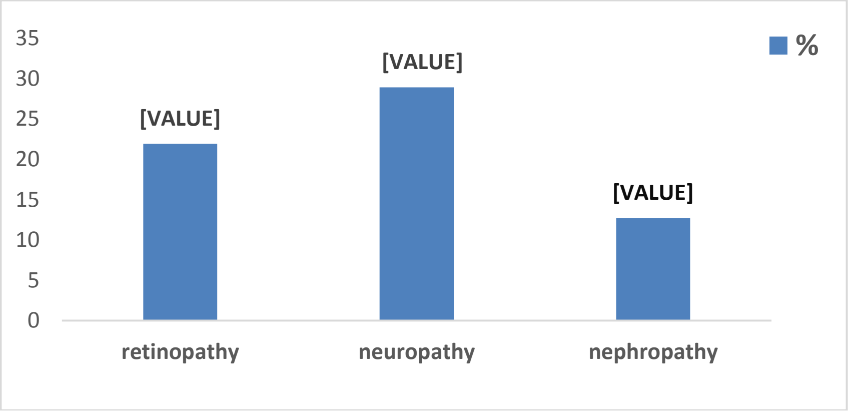
	n (Number)	% (percentage)
Age Groups (Years)		
Mean± SD	54.53±10.13	
35-45	104	22.8
46-55	156	34.1
56-65	150	32.8
>66	47	10.3
Gender		
Male	287	62.8
Female	170	37.2
Marital status		
Not currently married	77	16.8
Married	380	83.2
Weight (Kg) Mean± SD	72.77±14.006	
Height (m2) Mean± SD	1.6522±0.12476	
BMI		
Mean± SD	26.6215±4.16420	
Healthy Weight	172	37.6
Overweight	204	44.6
Obese	81	17.7
Diabetes duration (months)		
Mean± SD	85.45±71.981	
Less than 2.5 years	125	27.4
5 years	116	25.4
10 years	104	22.8
more than 10	112	24.5
Smoking status		
Current	65	14.2
Past	46	10.1
Non-smoker	346	75.7
Diabetes status		
uncontrolled	147	32.2
partially uncontrolled	136	29.8
controlled	174	38.1
Management by insulin		
No	335	73.3
Yes	122	26.7
Management by oral hypoglycemic		
No	131	28.7
Yes	326	71.3
Family history		
No	179	39.2
Yes	278	60.8

The frequency of associated co-morbid conditions like hypertension, heart disease and hypercholesterolemia was assessed, shown in graph 1.



Graph 1: Frequency of Co Morbidities in Type 2 Diabetic Patients

Frequency of complications associated with diabetes is shown in gable 2. 28.9% of the patients had neuropathy, while 21.9% and 12.7% had retinopathy and nephropathy respectively.



Graph 2: Frequency of Complications in type2 Diabetic Patients

Knowledge Assessment of the Participants

Assessment of knowledge of lifestyle modifications of diabetic patients was compared with the general demographic characteristics shows Table 2.

87.1% of the males, while 83.3% female participants had good knowledge about their disease and lifestyle modifications changes (p =0.278).

Table 2
Knowledge

Factors	Poor Knowledge		Good Knowledge		Chi square Test	P-value
	Frequency	Percentage	Frequency	Percentage		
Age (Years)					19.26	<0.001
35-45	30	42.9%	74	19.1%		
46-55	16	22.9%	140	36.2%		
56-65	18	25.7%	132	34.1%		
>66	6	8.6%	41	10.6%		
Gender					1.178	0.278
male	22	12.9%	148	87.1%		
female	48	16.7%	239	83.3%		
Marital status					22.253	<0.001
Single	12	27.3%	32	72.7%		
Married	48	12.6%	332	87.4%		
Separated	4	25.0%	12	75.0%		
Divorced	-	-	6	100.0%		
Widow	6	54.5%	5	45.5%		
BMI					1.209	0.546
Healthy Weight	30	17.4%	142	82.6%		
Overweight	30	14.7%	174	85.3%		
Obese	10	12.3%	71	87.7%		
Diabetes Duration(months)					14.087	0.003
less than 2.5 years	30	24.0%	95	76.0%		
5 years	20	17.2%	96	82.8%		
10 years	8	7.7%	96	92.3%		
more than 10	12	10.7%	100	89.3%		
Management by insulin					7.477	0.006
No	42	12.5%	293	87.5%		
Yes	28	23.0%	94	77.0%		
Management by oral hypoglycemic					11.750	0.001
No	32	24.4%	99	75.6%		
Yes	38	11.7%	288	88.3%		
Diabetes status					19.988	<0.001
uncontrolled	30	20.4%	117	79.6%		
partially uncontrolled	30	22.1%	106	77.9%		
Controlled	10	5.7%	164	94.3%		

ATTITUDE ASSESSMENT OF THE PARTICIPANTS

The attitude about lifestyle modifications assessed was compared with the general demographic characteristics and it showed the results shown in table 3. Our study showed that attitude of participants towards lifestyle modifications in all categories were majorly positive.

Table 3
Attitude

Factors	Negative Attitude		Positive Attitude		Chi square Test	P-value
	Frequency	Percentage	Frequency	Percentage		
Age (Years)					29.949	<0.001
35-45	20	19.2%	84	80.8%		
46-55	14	9.0%	142	91.0%		
56-65	-	-	150	100.0%		
>66	4	8.5%	43	91.5%		
Gender					30.922	<0.001
Male	30	17.6%	140	82.4%		
Female	8	2.8%	279	97.2%		
Marital status					77.096	<0.001
Single	18	40.9%	26	59.1%		
Married	16	4.2%	364	95.8%		
Separated	4	25.0%	12	75.0%		
Divorced	0	.0%	6	100.0%		
Widow	0	.0%	11	100.0%		
BMI					10.537	0.005
Healthy Weight	14	8.1%	158	91.9%		
Overweight	24	11.8%	180	88.2%		
Obese	0	.0%	81	100.0%		
Diabetes Duration(months)					23.519	<0.001
less than 2.5 years	18	14.4%	107	85.6%		
5 years	16	13.8%	100	86.2%		
10 years	4	3.8%	100	96.2%		
more than 10	0	.0%	112	100.0%		
Management by insulin					5.029	0.025
No	22	6.6%	313	93.4%		
Yes	16	13.1%	106	86.9%		
Management by oral hypoglycemic					1.355	0.244
No	14	10.7%	117	89.3%		
Yes	24	7.4%	302	92.6%		
Diabetes status					38.588	<0.001
uncontrolled	6	4.1%	141	95.9%		
partially uncontrolled	28	20.6%	108	79.4%		
Controlled	4	2.3%	170	97.7%		

PRACTICE ASSESSMENT OF THE PARTICIPANTS

The practice among participants for lifestyle modifications assessed was also compared with the general demographic characteristics and it showed the following results as shown in table 5.

Table 5
Practice

Factors	Bad practice		Good practice		Chi square Test	P-value
	Frequency	Percentage	Frequency	Percentage		
Age (Years)					38.220	<0.001
35-45	84	80.8%	20	19.2%		
46-55	78	50.0%	78	50.0%		
56-65	98	65.3%	52	34.7%		
>66	17	36.2%	30	63.8%		
Gender					4.783	0.029
male	92	54.1%	78	45.9%		
female	185	64.5%	102	35.5%		
Marital status					13.622	0.009
single	30	68.2%	14	31.8%		
married	222	58.4%	158	41.6%		
separated	8	50.0%	8	50.0%		
divorced	6	100.0%	0	.0%		
widow	11	100.0%	0	.0%		
BMI					26.025	<0.001
Healthy Weight	112	65.1%	60	34.9%		
Overweight	100	49.0%	104	51.0%		
Obese	65	80.2%	16	19.8%		
Diabetes Duration (months)					22.136	<0.001
less than 2.5 years	71	56.8%	54	43.2%		
5 years	85	73.3%	31	26.7%		
10 years	46	44.2%	58	55.8%		
more than 10	75	67.0%	37	33.0%		
Management by insulin					0.436	0.509
No	200	59.7%	135	40.3%		
Yes	77	63.1%	45	36.9%		
Management by oral hypoglycemic					4.851	0.028
No	69	52.7%	62	47.3%		
Yes	208	63.8%	118	36.2%		
Diabetes status					12.139	0.002
uncontrolled	96	65.3%	51	34.7%		
partially uncontrolled	93	68.4%	43	31.6%		
Controlled	88	50.6%	86	49.4%		

The global scores of knowledge, attitude and practice were assessed. Majority, which is 84.7% of the participants had good knowledge regarding their disease and lifestyle modifications for it. Most of them, 91.7% had a positive attitude but practice was lacking. Only 39.4% showed good practice regarding lifestyle modifications.

4. DISCUSSION

Diabetes mellitus is a major public health problem. A diabetic patient should have a working knowledge of diabetes such as signs and symptoms, management etc. for a better patient compliance.[8]

This study was conducted with the aim of assessing the sociodemographic profile of patients attending a diabetes clinic in a tertiary care hospital and their knowledge, attitude and practices regarding the disease.[9]

In our study the data showed that prevalence of type 2 diabetes is higher in men. Another study revealed that the prevalence of diabetes is higher in women when compared to men and higher in people of lower SES when measured by education and or income.[10]

Our data showed that mean BMI of patients were 26.6215±4.16420 which is higher than normal, another study showed that the metabolic syndrome is associated with a 5-fold increase in the incidence of T2DM and abdominal obesity is major criteria for diagnosis of metabolic syndrome.[11]

Cigarette smoking has been firmly linked with increased risk of type 2 Diabetes Mellitus. Increased insulin resistance occurs in smokers with or without diabetes.[12] The smoking reduced insulin mediated glucose uptake by 10% to 40% in men who smoked compared with non-smoking men.[13] Our study showed that among the diabetic population that we conducted the study in, 14.2% were active smokers.

Regular annual screening for diabetes complications allows treatable diseases to be identified. Patient's lack of knowledge affects their ability to manage the complications. Out of the three complications which we included in our study, neuropathy was found to be the most common complication in 28.9% of patients. In a study conducted in Kolkata 2010 they stated eye complication to be the most common complication in diabetic patients.[14] While in a study done in Sudan high prevalence of neuropathy was found (36.7%).[15]

We assessed various co morbidities of diabetes such as hypertension, cardiovascular diseases and patients with high cholesterol levels or any other. We found that hypertension was prevalent in 230 (50.3%), CVD in 77 (16.8%), high cholesterol levels in 131 (28.7%). Also in a study similar results were seen and hypertension was associated in (52.8%), high cholesterol in (44.8%), and CVD in (36.8%) of patients.[15] Another study revealed hypertension as most prevalent in diabetic patients (77%) and second most prevalent hyperlipidemia in (58.5%) of patients.[16]

Our study analysis comprised of four age groups among which the highest prevalence of diabetes was in the age group of 46- 55 years (34.1%) while it was closely followed by the age group of 56-65 years which had 32.8% of diabetic population. The Chennai Urban Rural Epidemiology Study (CURES) showed that the prevalence of diabetes was highest in the age group of 60-69 years (33.6%) while the age group 50-59 years had 29.7% with 20.4% prevalence in 40-49 years of age. In our study we compared demographics with knowledge, attitude and practices among type 2 diabetic patients. Our data showed that 42.9% diabetic patients of age group 35-45 years had poor knowledge while 36.2% of age group 46-55 years had good knowledge.[17]

Among the 287 male participants, 87.1% had good knowledge, while out of 170 female participants, 83.3% had good knowledge. This showed that in our study males had a comparatively higher knowledge of lifestyle modification for diabetes control. A recent study also reported higher total knowledge scores for males (17.8%) as compared to females (15.24%).[18]

In our study we compared good and poor knowledge among individuals with duration of diabetes we concluded that greater number of individuals, with longer duration of diabetes which is greater than 10 years in our research, had good knowledge that is 89.3%.

For the management of diabetes (26.7%) of patients use insulin, (71.3%) of patients use oral hypoglycemic, among insulin users 77.0% individuals had good knowledge, while among oral hypoglycemic users 88.3% individuals had good knowledge, while a study conducted in Oman in 2013 showed that (25.5%) of patients use insulin, (57.5%) of patients use oral hypoglycemic drugs and (6.6%) of patients use no medication. This reveals that diabetic patients mostly rely on drugs to control their disease.[19]

Diabetes status of patients in our study was found controlled in (38.1%) among them 94.3% had good knowledge, while in a study conducted in Malaysia, the status of patients with controlled diabetes was (30.8%).[20]

We also compared the data for negative and positive attitudes among men and women and practices. (28%) of the respondents with good body weight had good practices in a study conducted in Kenya and our research showed the data of (34.9%) had good practices. [21] Patients are not willing to give much time to perform or engage in healthy lifestyle. Higher level of positive attitude among women (97.29) as compared to men (82.4%). It showed that there was less negative attitude among both gender give much time to perform or engage in healthy lifestyle. Community health education interventions for diabetes need to take into account the disparity and uniqueness which exist between gender, age and different groups.

5. COMMENTS AND CONCLUSION

We conclude that despite having good knowledge and positive attitudes, patients are not able to manage their disease as shown by their uncontrolled diabetes status. The plausible factors could be their high BMI, associated comorbidities, lack of exercise, not following a modified diet plan and negligence regarding self-care examinations. This proves that the high prevalence of Diabetes in our country is largely due to lack of self-motivation and social support of good practices rather than poor knowledge or attitude about the disease. Thus for Diabetes control programs to succeed, there is a need for a self-empowerment approach which motivâtes the patient to utilize the knowledge and good attitude by practicing it.

LIMITATIONS: Our study set up was confined to Civil Hospital Karachi. More than half of the collected information was based on subjective assessment. The diabetes control status of patients couldn't be confirmed by laboratory investigation.

6. ACKNOWLEDGEMENT

- Dr. Khalil Ali Pathan, Incharge Diabetic Out Patient Department, Civil Hospital Karachi.
- Dr. Shaheen, Incharge Medical Out Patient Department, Civil Hospital, Karachi.

REFERENCES

1. Shah, V., Kamdar, P.K. and Shah, N. (2009). Assessing the knowledge, attitudes and practice of type 2 diabetics among patients of Saurashtra region, Gujarat. *Int. J. Diab. Dev. Ctries.*, 29(3), 118-122.
2. Nanditha, A., Ma, R.C.W., Ramachandran, A., Snehalatha, C., Chan, J.C.N., Chia, K.S., Shaw, J.E. and Zimmet, P.Z. (2016). Diabetes in Asia and the Pacific: Implications for the Global Epidemic. *Diabetes Care*, 39(3), 472-485.
3. Asma, A. and Qing, Q. (2004). Lifestyle intervention eases battle with diabetes. *Diabetes Voice*, 59, 29-31.
4. Serrano-gil, M. and Jacob, S. (2010). Engaging and Empowering Patients to Manage their Type 2 Diabetes, Part I: a Knowledge, Attitude, and Practice Gap? *Adv. Ther.*, 27(6), 665-680.
5. Kushwaha, A.S., Kumari, S. and Kushwaha, N. (2016). Self-care in diabetes: a study amongst diabetics in an urban community. *International Journal of Community Medicine and Public Health*, 3(1), 293-298.
6. Ng, S.H., Chan, K.H., Lian, Z.Y., Chuah, Y.H. and Waseem, A.N. (2012). Reality vs Illusion: Knowledge, Attitude and Practice among diabetic patients. *International Journal of Collaborative Research on internal medicine and Public health*, 4(5), 709-718.
7. Okonta, H.I., Ikonbele, J.B. and Ogunbanjo, G.A. (2014). Knowledge, attitude and practice regarding lifestyle modification in type 2 diabetic patients. *Afr. J. Prm. Health Care Fam. Med.*, 6(1), 655.
8. Al-bimani, Z., Khan, S. and David, P. (2015). Evaluation of T2DM related knowledge and practices of Omani patients. *Saudi Pharmaceutical Journal*, 23(1), 22-27.
9. Mukhopadhyay, P., Paul, B., Sengupta, N. and Majumder, R. (2010). Perceptions and practices of type 2 diabetics: A cross- sectional study in a tertiary care hospital in Kolkata. *Int. J. Diab. Dev. Ctries.*, 30(3), 143-149.
10. Bennett, N.R., Bidulescu, A., Ferguson, T.S., Francis, D.K., Hambleton, I.R., Harris, E.N. et al. (2015). Disparities in diabetes mellitus among Caribbean populations: a scoping review. *International Journal for Equity in Health*. 2015, 14:23. DOI: 10.1186/s12939-015-0149-z
11. Obeidat, A.A., Ahmad, M.N., Haddad, F.H. and Azzeh, F.S. (2015). Alarming high prevalence of metabolic syndrome among Jordanian adults. *Pak. J. Med. Sci.*, 31(6), 1377-1382.

12. Shi, L., Shu, X-O., Li, H., Cai, H., Liu, Q., Zheng, W., Xiang, Y-B. and Villegas, R. (2013). Physical Activity, Smoking, and Alcohol Consumption in Association with Incidence of Type 2 Diabetes among Middle-Aged and Elderly Chinese Men. Public Library Of science, 8(11), e77919. DOI: 10.1371/journal.pone.0077919
13. Chang, S.A. (2012). Smoking and Type 2 Diabetes Mellitus. *Diabetes and Metabolism Journal*, 36(6), 399-403.
14. Mukhopadhyay, P., Paul, B., Sengupta, N. and Majumder, R. (2010). Perceptions and practices of type 2 diabetics: A cross- sectional study in a tertiary care hospital in Kolkata. *Int. J. Diab. Dev. Ctries*, 30(3), 143-149.
15. Agyemang, C. and Bos, M. (2013). Prevalence and complications of diabetes mellitus in Northern Africa, a systematic review. *BMC Public Health*, 13, 1-7.
16. Schafer, I. and Hansen, H. (2012). The influence of age, gender and socio-economic status on multimorbidity patterns in primary care first results from the multicare cohort study. *BMC Health Services Research*, 12, 1-15.
17. Mohan, V., Deepa, M., Deepa, R., Shanthirani, C.S. and Farooq, S. (2006). Secular trends in the prevalence of diabetes and impaired glucose tolerance in urban South India--the Chennai Urban Rural Epidemiology Study (CURES-17). *Diabetologia*, 49(6), 1175-1178.
18. Al-Maskari, F., El-Sadig, M., Al-Kaabi, J.M., Afandi, B., Nagelkerke, N. and Yeatts, K.B. (2013). Knowledge, Attitude and Practices of Diabetic Patients in the United Arab Emirates. *Public Library of Science*, 8(1). E52857.
DOI: 10.1371/journal.pone.0052857
19. Al-Bimani, Z., Khan, S. and David, P. (2015). Evaluation of T2DM related knowledge and practices of Omani patients. *Saudi Pharmaceutical Journal*, 23(1), 22-27.
20. Hui Ng, S., Chan, K.H., Lian, Z.Y., Chuah, Y.H., Waseem, A.N. and Kadirvelu, A. (2012). Reality vs Illusion: Knowledge, Attitude and Practice among Diabetic Patients. *International Journal of Collaborative Research on Internal Medicine & Public Health*, 4(5), 723-732.
21. Maina, W., Ndegwa, Z., Njenga, E. and Muchemi, E. (2010). Knowledge, attitude and practices related to diabetes among community members in four provinces in Kenya: a cross sectional study. *Pan African Medical Journal*, 7, 685-692.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 173-178

CLASSIFICATION OF MENIRAN (PHYLLANTHUS NIRURI) IN JAVA ISLAND BASED ON FTIR SPECTRA

Siti Mahmuda, Farit Mochamad Afendi and **Mohamad Rafi**
Department of Statistics, Bogor Agricultural University, Indonesia
Email: sitimahmuda24@gmail.com

ABSTRACT

Indonesia is home to many medicinal plants, one of them is the *Meniran* (*Phyllanthusniruri*). The plant grows in several region of Indonesia under various conditions. It is possible that the variation of cause's different FTIR spectra within the species, hence resulting in different contain of bioactive as well. This study attempted to classify *Meniran* (*Phyllanthusniruri*) in Java Island based on FTIR spectra. The principle component analysis (PCA) and cluster analysis were used to get the spectra classification from different places of origin spread Java Island (West Java, Central Java and East Java). For the purpose of this study, slope variation, scatter effects and baseline drifts were removed with preprocessed IR spectra (baseline and derivatization). The results show that FTIR spectra can be used to classify *Meniran* (*Phyllanthusniruri*) from places of origin (Java Island).

KEYWORDS

Classification, Cluster, FTIR, Meniran, Origin, PCA, *Phyllanthusniruri*.

1. INTRODUCTION

Background

Indonesia has many potential plants frequently used as medicine by the local people. Traditional medicine has been known since a long time ago. Hence, many research on medical plants have been conducted (Haris, 2011). *Meniran* (*Phyllanthusniruri*) is an important medicinal plant. The plant is widely used for the treatment of hepatic disease, oedema, dropsical condition, and urinary troubles (Paithankar *et al.,* 2011).

There are two types of *Meniran*in Indonesia, namely *Phyllanthusniruri* and *Phyllanthusurinaria*. Their distinction can be found on the stem. The *Meniran* (*Phyllanthusniruri*) grows well in damp places and lowland area until the altitude of 1000 metres above sea level (Sulaksana *et al*. 2004). In Indonesia, the *Meniran* (*Phyllanthusniruri*) is widely spread because there are some local names attached to these plants, such as *sikolop* (Sumatra), *memeniran* (Java), *sidukunganak* (Sulawesi), and *belalangbabiji* (Maluku) (Kardinan *et al.,* 2004).

The active chemical component of a plant system is a mixture of compounds that are very widely, depending on many factors. These factors include the environment in where it grows, soil nutrients, climate, altitude, quality of seeds, cultivation technology (including lighting), age of the plant during harvest, post-harvest processing and storage

Therefore, differences in geographical origin of medicinal plants allow for differences in the characteristics of that lead to inconsistency in efficacy (Dalimartha, 2008).

Fourier transform infrared (FTIR) spectroscopy was developed in order to overcome the limitations encountered by dispersive instruments. The main difficulty was the slow scanning process. A method for measuring all of the infrared frequencies simultaneously. FTIR can be used to identify unknown materials, determine the quality or consistency of a sample and measure the amount of components in a mixture. This study attempted to classify *Meniran* from places of origin based on FTIR spectra. Absorbent number of FTIR spectra will be produced in wave number range of 400-4000 cm^{-1}. In preprocessed, the wave number (400-4000 cm^{-1}) can be selected to get significant peaks of spectral (Thermo Nicolet, 2001).

Spectral data pre-treatment is an important step before subjecting the FTIR spectra data for multivariate analysis. It is necessary to perform this step in order to minimize the effect of light scattering from baseline variation, occurrence of systematic noise, etc. (Chen, 2008). For the purpose of this study, slope variation, scatter effects and baseline drifts were removed with preprocessed IR spectra (baseline and derivatization).

Principal component analysis is essentially serves to explain the structure of variance covariance via a linear combination of variables. The analysis was conducted to reduce the dimensions of the data and to interpret the result. Utilizing cluster analysis we search for patterns in a data set by grouping the (multivariate) observations into clusters. The goal is to find an optimal grouping for which the observations or objects are similar within each cluster, but dissmilar between clusters (Johnson *et al.*, 1992).

2. DATA AND METHOD

Data

This study used FTIR spectra data of *Meniran* (*Phyllanthusniruri*). Origins of *Meniran* (*Phyllanthusniruri*) samples were West Java (Bogor, Cianjur, Sukabumi, Garut), Central Java (Karanganyar, Semarang), and East Java (Malang, Probolinggo). Total number of samples are 24 samples with detail as in the table below:

Table 2.1
Samples of Meniran (Phyllanthusniruri)

Provinces	Distiricts / Cities	Number of Samples
West Java	Bogor (BGR)	3
	Cianjur (CJR)	3
	Sukabumi (SKB)	4
	Garut (GRT)	2
Central Java	Karanganyar (KRG)	4
	Semarang (SMG)	2
East Java	Malang (MLG)	3
	Probolinggo (PBR)	3

FTIR spectra data consist of wave number and absorbent each samples.

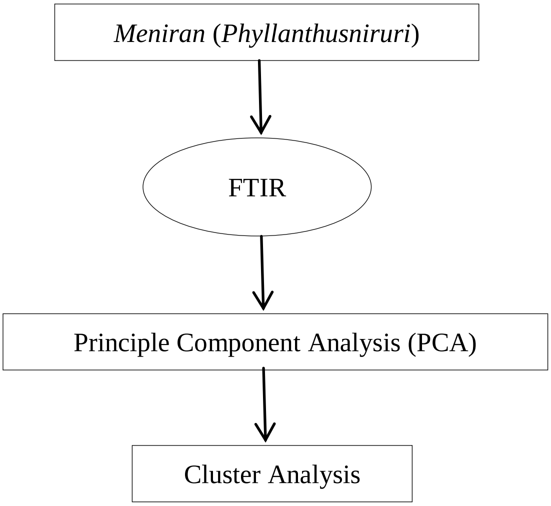


Fig. 1: Research Procedures

3. RESULT AND DISCUSSION

Pre-Processing

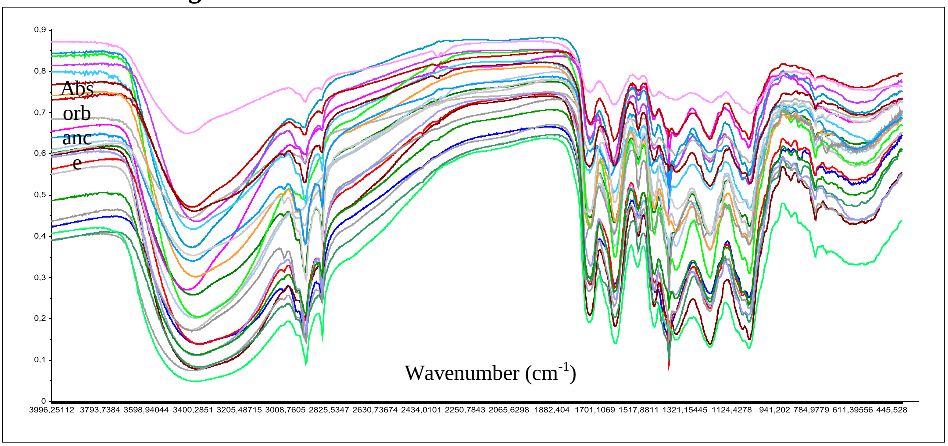


Fig. 2: FTIR spectra of 24 *Meniran* (*Phyllanthusniruri*) Samples.

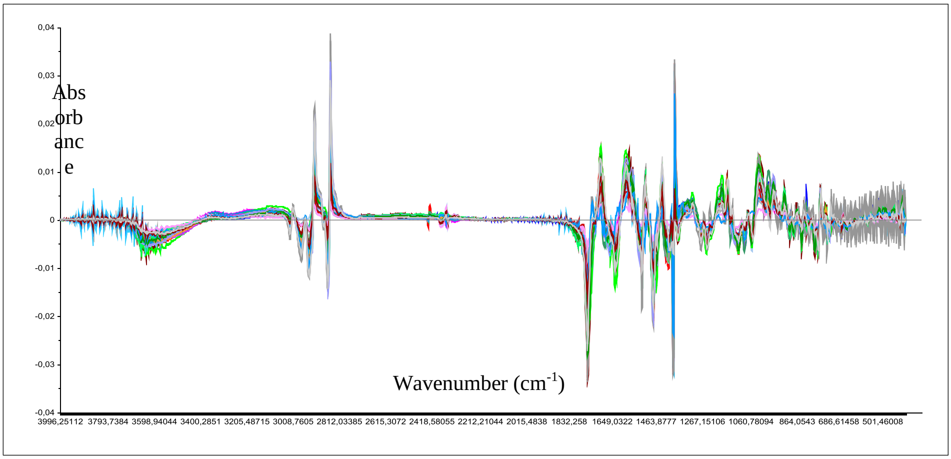


Fig. 3: Baseline and Derivative FTIR Spectra of 24 *Meniran* (*Phyllanthusniruri*) Samples.

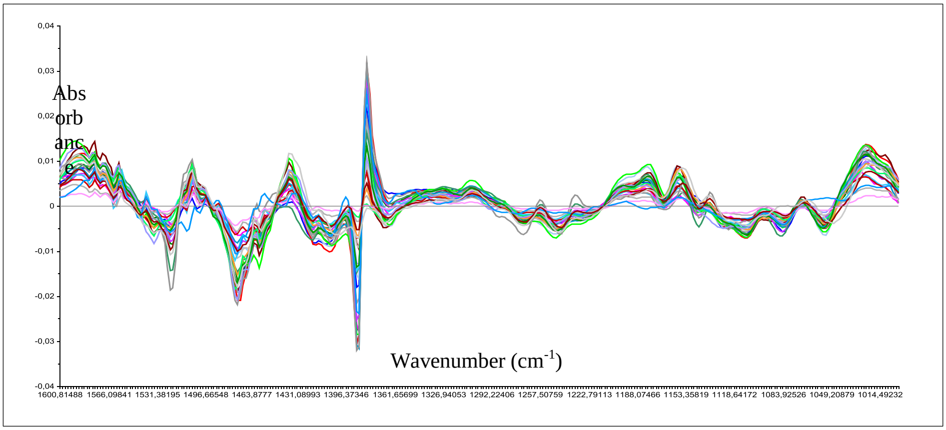


Fig. 4: FTIR Spectra of 24 *Meniran* (*Phyllanthusniruri*) Samples at wave number 1600-1000 cm^{-1}

Before the analysis process, it is necessary to do some treatment to remove the slope variation, scatter effects and baseline drifts at FTIR spectra. The treatments is called pre-processing FTIR spectra. In Figure 2 depicts spectra of 24 samples of *Meniran* (*Phyllanthusniruri*), while the result of preprocessed with baseline and derivatization is showed at Figure 3. The selected regions of the spectra is at wavelength of 1600-1000 cm^{-1} (312 wave number) to get the significant peaks of spectral. Hence, Regional absorbance at that wavelength (Figure 4) is analyzed.

Principal Component Analysis (PCA)

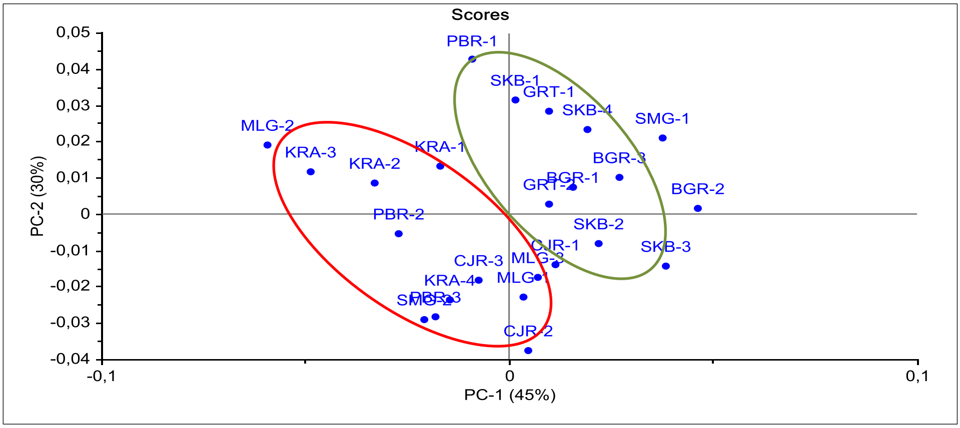


Fig. 5: Principle Component Analysis (PCA) Plot

A large data with complex correlation is not easy to be classified without technical intermediaries. Therefore, the Principal Component Analysis (PCA) is used as an intermediary analysis to reduce the dimensions of the data. The spectrum at wavelength 1600-1000 cm-1 was analyzed using principal component analysis (PCA). Two main components are able to explain variance (75% with (PC1 =45% and PC2 =30%). Further, Figure 5 shows The *Meniran* (*Phyllanthusniruri*) can be classified in two groups; (1) West Java (2) Central Java and East Java.

Cluster Analysis

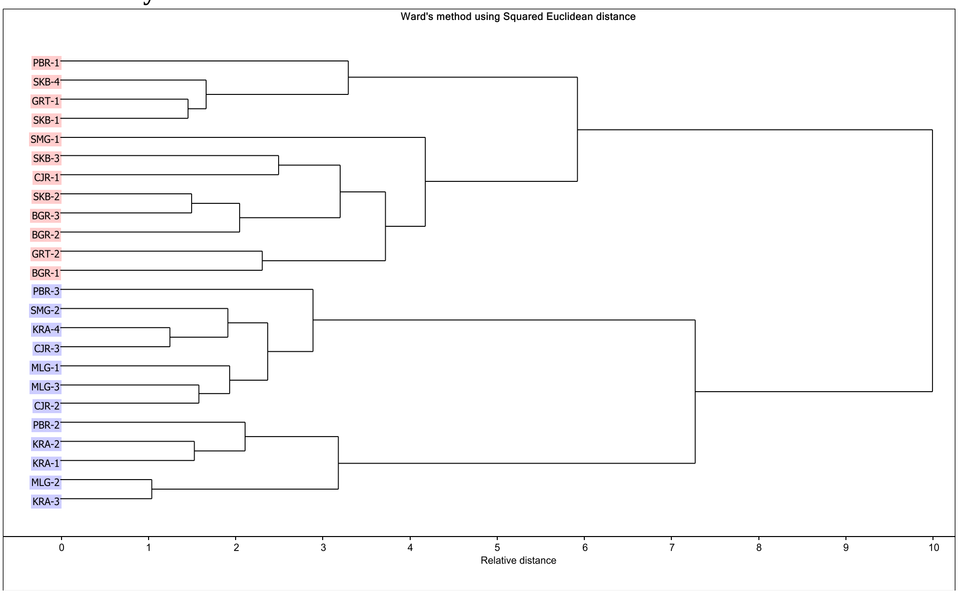


Fig. 6: Dendogram

Two main components were taken as the basis for classification. The formed cluster (Figure 6) shows characteristic differences in *Meniran (Phyllanthus niruri)*. In this study, Cluster Analysis using ward method with Euclidean distance. This study also found that *Meniran* (*Phyllanthus niruri*) can be distinguished by its origin in Java Island. The difference is shown by the formation of two groups, the first group is *Meniran* (*Phyllanthus niruri*)originating from West Java and the second group is meniran originating from East Java and Central Java.

4. CONCLUSION

The *Meniran* (*Phyllanthusniruri*) in Java Island can be classified in two groups based on FTIR spectra utilizing Principle Component Analysis (PCA) and Cluster Analysis. First group is West Java and the second group is Central Java and East Java. The groups indicate that there are differences of the *Meniran* (*Phyllanthusniruri*) in each province in Java Island. This could be caused by differences in altitude, weather and soil type in each regional group.

REFERENCES

1. Dalimartha, dr Setiawan. (2005). *Tanaman Obat di Lingkungan Sekitar Kita*. Jakarta: Puspa Swara, Anggota IKAPI.
2. Haris (2011). *Obattradisional*. Universitas Sanata Dharma, Yogyakarta.
3. Johnson, R.A. and Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*, second edition. Prentice-Hall International, Inc. New Jersey.
4. Kardinan (2004). *Meniran Penambah Daya Tahan Tubuh Alami Edisi 1*. Agro Media Pustaka, Jakarta.
5. Paithankar et al. (2011). *Phyllanthusniruri: A magic Herb*. Vidya bharati college of Pharmacy, Amravati.
6. Sulaksana et al. (2004). *Meniran*. Budi Daya & Pemanfaatanuntuk Obat. Penebar Swadaya, Jakarta.
7. Thermo Nicolet (2001). *Introduction to Fourier Transform Infrared Spectrometry*. A Thermo Electron business, USA.
8. Chen, Y., Xie, M.Y., Yan, Y., Zhu, S.B., Nie, S.P., Li, C., Wang, Y.X. and Gong, X.F. (2008). Discrimination of Ganoderma lucidum according to geographical origin with near infrared diffuse reflectance spectroscopy and pattern recognition techniques. *Anal. Chim. Acta*., 618, 121-130.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 179-186

EVOLUTION OF CORPORATE GOVERNANCE PRACTICES AND CONVENTIONAL BANKS PROFITABILITY

Nadeem Iqbal[1], **Nur Hilal A. Syahrir**[2] and **Abdus Subhan**[3]

[1] Faculty of Management Sciences, Ghazi University
DG Khan, Pakistan
Email: drnadeemiqbal1@gmail.com

[2] Bogor Agricultural University, Indonesia

[3] Indus International Institute DG Khan, Pakistan

ABSTRACT

This study has been conducted to discover the good governance practices influence on profit of conventional banking system. Prime aim is to discover the application of these practices in the corporation for betterment. This research based on mixture of primary and secondary data. Sample for primary data collection for study is different conventional banks of Pakistan. Observation method and structured questionnaires is used to obtain the data from respondents (banks employees). Obtained data is analyzed through statistical software SPSS. Regression analysis is done to check out the effects of corporate governance and its determinants on profitability of conventional financial institution. So findings possess that banks profit tend to be enhance with the espousal of ethical practices in corporate culture. Study concluded that if financial institution behave ethical manner, socially obliged and invest for well being of society after that they will get and enhance profit from social capital in addition to be considered as socially responsible corporations by adopting regulatory system aptly.

KEYWORDS

Corporate Governance, profitability, Ethical practices, regulatory system, performance measures.

1. INTRODUCTION

Corporate governance is known as combination of parameters, rules, regulation and business practices that lead the relationship among management and all stakeholders. This business procedure or practice involves fair, translucent and efficiency in administration which make them able to get the desired and planned results. Corporate governance operates and controls a corporation in systematic way although companies stimulate and attract the financial and human capital. Corporate governance system is embraced a broad array of business practices and institution since accounting laws and standard which concern the monetary confession and disclosure towards managerial compensation as well as composition of corporate board.

First time In Pakistan corporate governance code was disseminated in March 2002. Before this it was noted that current business school was emphasizing on the development of strong corporate governance. Pakistan institute of corporate governance was developed to put attention on training of faculty of SECP and stock exchange. In corporate world, that institute was shaped to develop an overpass or connection among corporation and its regulation system. In extensive intellect and broad view corporate governance is the complementary and corresponding set of lawful, social and financial institution that defend the interest of all stakeholders of corporation.

In Anglo American system, major shareholders known as owners and corporate manager works as agent so corporate governance conception assumes a fundamental stress among shareholders and corporate managers. Both of them put effort for their own interest so as corporation set regulatory bodies in the form of Directors who are the followers of stewardship theory. The establishment and founding of SEC security and Exchange Commission of Pakistan is represented a vital highlight for the progress of dictatorial structure of capital marketplace. SEC accomplishes operational activities and be successful for CLA Corporate Law Authority known as government department affixed to Ministry of Finance. In the beginning it was concerned with regulatory of capital market and corporate sector. Commission has been categorized into different sections like as insurance, securities, HR and Administration and company Law division and all of that work for effective management.

Good corporate governance proves more supportive for the development of economic sustainability through maximizing corporation’s performance and productivity as well as access to outside or exterior capital. It reduces liability and exposure of financial crises, reinforcing property rights and contributes to development of capital market. Corporate governance concerns to the linkage between major and minority shareholders, BOD’s, top management and all other stakeholders. East Asian Financial crises capture the serious consideration and notice to the vitality of corporate governance in developing countries.

In 1999, OECD (organization for Economic Co-operation and Development) has developed a series of principles for corporate governance that have been proved core pattern to assess the country CG pact and arrangements. It provides composition by which company goals are set, means of accomplishing all those objectives and monitoring performance and quality standards. General approaches to corporate governance to defend and protect rights of investors. Initially delegates power to investors and shareholders through legal way and focus for minorities shareholders’ rights and legal prohibition beside administrative self decision and dealing. The next major approach relates to concentrated ownership by major investors by matching considerable and significant cash flows and control rights. Large investors mostly rely on legal system and minimize the agency cost and issues as they do not have require many rights as minor shareholders needs to protect own rights, So Corporate governance is commonly exercised via major investors or shareholders.

Effective corporate governance leads to efficiency and high growth of companies, minimize risk and stabilized economic conditions by getting investors confidence in strategic way. Apart since all these financial development and modifications, corporate governance code has been executed to all listed firms of Pakistan in 2002, prime aim is to persuade good governance in whole dynamic business market.

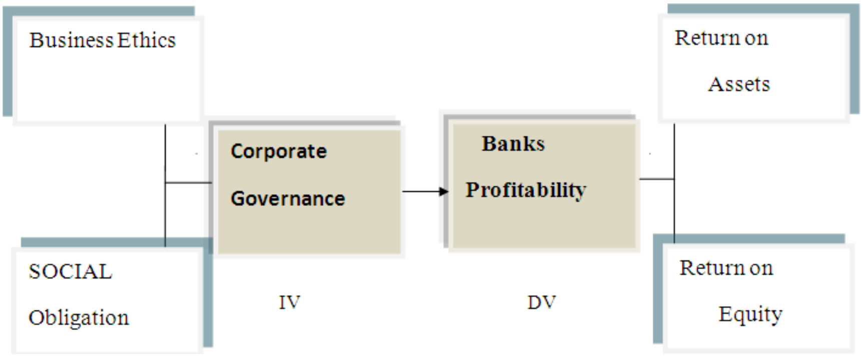
The aim of this research paper is to find out the link among corporate governance and corporate performance for financial institutions. So in this study we try to recognize the relationship among corporate governance determinants and practices with conventional banking regulatory and profitability in dynamic competitive market.

2. LITERATURE REVIEW

This research paper is going to review the previous studies on corporate governance practices in various financial institutions of different developing and developed countries. Corporate governance makes sure that companies get the responsibility to manage, control and monitor all matters in systematic way by preserving the interest of stakeholders (Berghe, Mirsiglia, 2005). The facets of corporate governance deal with initiating the process by which core values are achieved and ethics known as behavioral guide. (Sullivan and Shkolnikov; 2006) Berenbeim, Graafland (2006)

According to Mallin (2007) corporate governance is a mode through which companies administer and oversee their businesses by following its practices and reforms strategies and maximize profitability. Ineffective corporate governance mechanism in corporations prove as major factor to the current financial crises. Jamali, Safieddine, Rabbath (2008).argued that ideas and suggestions for governance variables as corporate social responsibility is the advancement and evolution since benevolent activities toward applicable tactic for getting the trust of community and major clients at all. In accordance to Marcinkowska, (2012) corporate governance determines all factors that have significant influence on institutional procedures such as employing and appointing controllers and regulatory systems for organizing activities. The customers are the co-initiator of assessment and value thus corporate governance practices and social responsibility of corporations is incorporated similar to others firm capital and possessions that give strength to shareholders as well as all stake holders. In accordance to Iqbal (2015), study finding explored strong connectivity among fiscal performance of banking sectors of Pakistan and application of corporate governance. It covers up basically many gauges involving effectiveness and independence of board and audit committee for regulatory system. Haider et al. (2015) explains that corporate governance practices and governance determinants and financial performance of Islamic financial institutions. Study revealed that significance link and connectivity among large board size and profitability because board protect the rights of major shareholders. They idealize the resource dependent theory which mainly emphasize on accountability, profitability and transparency for achieving corporate goals.

3. CONCEPTUAL FRAMEWORK



4. DETERMINANTS OF GOVERNANCE & PROFITABILITY AND DEVELOPMENT OF HYPOTHESES

Corporate governance is used as independent variable measured through its component which is Ethical practices applicable in corporation. Banks profitability is measured through two determinants that is Return on equity and Return on Assets. Research Hypotheses is developed after review previous studies regarding corporate governance and financial performance of financial institutions and test by descriptive research device.

H1: There is a positive link among corporate governance Practices and conventional banks profit.

H2: There is a positive link among Ethical practices (business ethics & social obligations) and firm financial performance.

5. RESEARCH METHODOLOGY

For this study, Descriptive Research design is adopted that include Questionnaires, observation method, individual interviews and case study but convenience method is structured questionnaires by using lickert scale method. Primary data is obtained by distributing questionnaires among bank's employees by hand and some through mails in different branches of conventional banks in Punjab. Response from electronic mode is less significance rather than by own visits and individual interviews from employees.

6. DATA ANALYSIS& RESULTS

Statistical analysis is conducted by the help of statistical instruments SPSS. Various tools and techniques are availed for testing research hypothesis. Regression analysis is done to interpret the obtained data. To get confirmation the significance of governance variables and profitability measures through firm performance, ROA and ROE.

ANOVA B

Model		Sum of Squares	Df		Mean Square	F	Sig.
1	Regression	220.324		4	55.081	2922.064	.000a
	Residual	1.791		95	.019		
	Total	222.115		99			

a. Predictors: (Constant), ETHICS, CSR
b. Dependent Variable: ROA

ANOVA B

Model		Sum of Squares	Df		Mean Square	F	Sig.
1	Regression	32013.365		4	8003.341	169.526	.000a
	Residual	4484.956		95	47.210		
	Total	36498.321		99			

a. Predictors: (Constant), Business Ethics, Social Obligations
b. Dependent Variable: ROE

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.996a	.992	.992	.13730

a. Predictors: (Constant), Business Ethics, Social Obligation

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.937a	.877	.872	6.87096

a. Predictors: (Constant), Business Ethics, Social Obligation

Regression Results

IV	DV	β coefficient	p-value
BUSINESS ETHICS	ROA	0.008	0.046
BUSINESS ETHICS	ROE	0.039	0.038

Regression Results

IV	DV	β coefficient	p-value
SOCIAL OBLIGATIONS	ROA	0.020	0.043
SOCIAL OBLIGATIONS	ROE	0.086	0.045

7. RESULTS

In this study analysis shows that the goodness of research model, in accordance to ANOVA test results significance value is .000<. .005 and R SQUARE results =.992 which implies 99% variation come into dependent variable measures ROA is due to applying ethical practices in the financial institution R-SQUARE value for ROE .87 which explains by CG determinants. Social obligation of corporate Coefficient value is 2% and 8% which possess that if firms invest 1% for social work then it will become cause increase near about 8% increment in profitability. Regression results shows p value is significant at .04 for ROA and .03 for ROE for both predictors as social obligation and ethics. Business ethics coefficient results reveals no more change in profitability measures as in results of investing in ethical practices. So statistical results are quite differ from previous study and discussed data in this research.

8. CONCLUSION& DISCUSSION

By applying different statistical tool to interpret the obtained data and to test the research hypothesis, it is identified positive influence of governance practices on banks profitability. First hypothesis states positive relationship among governance practices and conventional banks profitability so statistically proved the significance of link among both variables. Profitability ratio is enhanced little bit way when banking system behave themselves as ethical manners. By performing these codes of conduct and principles, performance and profitability of financial institutions will lead to the development of corporations. Second hypothesis is analyzed statistically, all components of ethical practices like as business ethics and social obligations are positively associated with firm financial performance. If corporations behave ethically and maintain business decorum and socially obliged to perform for well being of society thus they will not consider as materialistic entity that only work for profit maximization. Since corporations act as socially responsible and enhance their social contacts with clients and customers subsequently they achieve their goals ultimately.

9. DIRECTION FOR FUTURE RESEARCH

Current study has revealed the significance influence and relationship among corporate governance and profitability of banks. Thus here researcher emphasize on governance variables or determinants like as business ethics and social obligations of the corporations so others ethical practices must have to be discussed and evaluated in terms of financial performance indicators. Corporate governance code should be essential for adoption for every firm because adequate application of this code will prove more opportunistic ways for getting maximum profitability.

REFERENCES

1. Berghe, L.V. den, and Louche, C. (2005). The link between corporate governance and corporate social responsibility in insurance. *Geneva Papers on Risk & Insurance*, 30(3), 425-442.
2. Berenbeim, R.E. (2006). Defining an organization's ethics brand.(In The Conference Board. Paper delivered to the *Ethics and Compliance Conference: Putting It All Together*, 12 May. New York. 501-504.
3. Graafland, J., Mazereeuw, C. and Yahia, A. (2006). Islam and socially responsible business conduct: an empirical study of Dutch entrepreneurs. *Business Ethics: A European Review*, 15(4), 390-406
4. Jamali, D., Safieddine, A.M. and Rabbath, A. (2008). Corporate governance and corporate social responsibility synergies and inter relation ships. *High field Southampton*, Blackwell Publishing Ltd., 16(5), 443-459.
5. Mallin, C.A. (2007). *Corporate Governance*. (2nd ed.). New York, Oxford Press.
6. Marcinkowska, M. (2012). Corporate Governance in Banks: Problems and Remedies. *Financial Assets and Investing FAI*, 2012-2, 47-67.
7. Marsiglia, E. and Falautano, I. (2005). Corporate social responsibility and sustainability challenges for a Bancassurance company. *The Geneva Papers, Palgrave Macmillan Journals*, 30(3), 485-497.
8. Sullivan, J.D. and Shkolnikov, A. (2006). Business ethics: the essential component of corporate governance. *Economic Forum Feature Service*, 1-16. Jan 12.
9. Iqbal, N. Qazi, R. and Khan, N. (2015). Impact of Corporate Governance on Firm Financial Performance in Financial Sector of Pakistan.
10. Haider, N., Khan, N., Iqbal, N. (2015). Impact of Corporate Governance on Firm Financial Performance in Islamic Financial Institution. *International Letters of Social and Humanistic Sciences*, 51, 106-110.

Analysis of Financial Institutions

Company Name	Year	CEO Duality	Audit Committee	Board Independent	ROA %	ROE %
Meezan Bank Ltd	2009	Yes	Yes	Yes	13.9	16
	2010	Yes	Yes	Yes	9	11
	2011	Yes	Yes	Yes	14	12
	2012	Yes	Yes	Yes	13	11.56
Bank Islami	2009	No	Yes	Yes	14.57	13.69
	2010	No	Yes	Yes	12.5	13.01
	2011	No	Yes	Yes	11.278	14
	2012	No	Yes	Yes	13.32	14.56
State Life Insurance Pakistan	2009	No	Yes	No	9.89	9.46
	2010	No	Yes	No	9.09	9
	2011	No	Yes	No	8.67	9.78
	2012	No	Yes	No	8.82	9.97
New Jubliee Life Insurance	2009	No	Yes	Yes	0	0
	2010	No	Yes	Yes	2.3	4.5
	2011	No	Yes	Yes	5.2	5.98
	2012	No	Yes	Yes	5.8	6.2

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 187-194

PREDICTING TOTAL FERTILITY RATES (TFR) OF PAKISTAN USING ARIMA MODELS

Hira Fatima Waseem[1] and **Farah Yasmeen**[2]
Department of Statistics, University of Karachi, Karachi, Pakistan
Email: hira.waseem@live.com and riazfarah@yahoo.com

ABSTRACT

Fertility is one of the main demographic features of any population. Nowadays, decline in fertility has been a primary determinant of population aging and projected levels of fertility have important implications on the age structure of future populations, including the pace of population aging. The total fertility rate (TFR) is one of the key components in population projections. The aim of this study is to predict total fertility rates of Pakistan by using Box-Jenkins Autoregressive integrated Moving Average (ARIMA) methodology. We fit ARIMA model to the secondary data of total fertility rates from 1984 to 2007, obtained from Pakistan Demographic Surveys (PDS). The forecast accuracy of ARIMA models can be measured in terms of root mean square of error (RMSE), mean absolute percentage error. (MAPE), mean percentage error (MPE), where Akaike's information criteria and graphical techniques autocorrelation function (ACF) and partial autocorrelation function (PACF) plots can be used to determine the best ARIMA models. We also predict total fertility rates (TFR) of Pakistan for the next fifteen years using the best ARIMA model. The predicted values of fertility will help the government in planning children services and also give an idea to allocate future resources.

KEYWORDS

TFR, Box-Jenkins methodology, ARIMA, prediction.

1. INTRODUCTION

Fertility is one of the main demographic features of any population (Nasir et al., 2010). Fertility decline has been a primary determinant of population aging and projected levels of fertility have important implications on the age structure of future populations, including the pace of population aging. The total fertility rate (TFR) is one of the key components in population projections; it is the average number of children a woman would bear if she survived through the end of the reproductive age span (15-49) (Alkema et al., 2011).

Pakistan is the world's sixth most populous country. During 1950–2011, fertility has declined more slowly in Pakistan than in most other Asian countries. Demographers have been observed decline in fertility after 1990s in Pakistan (Akhtar et al., 2009). According to Pakistan Demographic Survey (PDS) during 1984-2007, the total fertility rate of Pakistan was came down from 6.95 to 3.7 births; (PDS, 2007) the estimates imply a decline of 3.25 births (46.76%) as compared to the year 1984. The drop in TFR is also partly due to postponement of marriage by women and knowledge of fertility control and use of contraceptives by Shitan and Ng (2015), Soon (1992) and Brown (2004).

The primary concerned for demographers is to model fertility curves to study the reproductive pattern. Several statistical models have been proposed to model fertility includes moving average, exponential smoothing, and ARIMA. These linear models have been the main research focuses and applied tools during the past few decades. The widely used time series models is the autoregressive integrated moving average (ARIMA) model because of simplicity of this model by Chi (2009) and Zhang (2003). The popularity of the ARIMA model is due to its statistical properties as well as the well-known Box–Jenkins methodology in the model building process (Adhikari & Agrawal, 2013). Basically, ARIMA model uses time series data to predict future trends and its model accuracy can also be measured in terms of mean absolute percentage error (MAPE).

ARIMA models have been applied by Saboia (1977), McDonald (1979), Land and Cantor (1983), McNown and Rogers (1989), Knudsen et al. (1993) and Carter (1996) in modeling and forecasting fertility rate as well as other demographic variables.[5,11-16] Recently, Hyndman and Booth (2008),[17] and Hyndman and Shang (2009)[18] were also used the ARIMA model for forecast fertility. The most convenient way in terms of data and calculation requirements is to directly forecast total fertility rate (TFR). Existing literature is lacking the ARIMA approach to forecast fertility of Pakistan.

Thus, the main objectives of this paper are:

- To study the pattern of the fertility rates of Pakistan for available years.
- To estimate the TFR for missing years using interpolation spline.
- To model the trends of TFR by ARIMA.
- To predict fertility with 80% prediction intervals for next fifteen years.

This paper is divided into five sections. Section 1 starts with introduction of the fertility in Pakistan and ARIMA model approach for demographic forecasting. In section 2, Box and Jenkins ARIMA model are discussed. In section 3, we have described source of data of fertility rates and estimation of fertility rates for missing years, the application of ARIMA models to fertility rates are discussed. We discuss the statistical results in section 4 and then final conclusion to overall analysis in section 5.

2. BOX AND JENKINS AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODEL

The steps of Box–Jenkins methodology include model identification, parameter estimation and diagnostic checking.

2.1 Steps of Box–Jenkins

i) **Model Identification:** At first, we plot the autocorrelation function and partial autocorrelation function of data to identify the order of the ARIMA model for making the time series stationary.

ii) **Parameter Estimation:** After model specification, we estimate the model parameters. The parameters are estimated such that an overall measure of errors is minimized.

iii) **Diagnostic Checking:** At last, we check the assumption of model is satisfied by plotting the residuals of model.

This three-step model building process is typically repeated several times until a satisfactory ARIMA model is finally selected. The final selected model can then be used for prediction purposes.

2.2 ARIMA (p,d,q) Model

This model is a generalized form of ARMA model by Box and Jenkins. In ARIMA (p,d,q) model where p, d, and q are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.

The ARIMA (p,d.q) is defined as:

$$\boldsymbol{Y_t = \delta + \emptyset_1 Y_{t-1} + \cdots\cdots\cdots + \emptyset_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots\cdots\cdots + \theta_q Z_{t-q}} \quad (1)$$

where $\{Z_t\}$ is a sequence of uncorrelated random variables with $\{\boldsymbol{Z_t}\}\sim \boldsymbol{WN}\ (\boldsymbol{0}, \boldsymbol{\sigma^2})$. The time series $\{Y_t\}$ is an ARIMA (p,d,q) process if its causal ARMA (p,q) process defined by

$$\emptyset(\boldsymbol{B})\boldsymbol{Y_t} = \boldsymbol{\theta}(\boldsymbol{B})\boldsymbol{Z_t} \quad (2)$$

where $\emptyset(\boldsymbol{Z}) = \boldsymbol{1} - \emptyset_1 \boldsymbol{Z} - \cdots - \emptyset_p \boldsymbol{Z^p}$, $\boldsymbol{\theta}(\boldsymbol{Z}) = \boldsymbol{1} + \boldsymbol{\theta_1 Z} + \cdots + \boldsymbol{\theta_q Z^q}$ and B is the backward shift operator

In ARIMA, the future value of a variable is assumed to be a linear function of several past observations and random errors. To evaluate the forecasting performance of the models, post sample forecast accuracy criterion such as mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) are used. The MAE, RMSE and MAPE are defined as

$$\text{Mean Absolute Error, } MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$$

$$\text{Root Mean Square Error, } RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$$

$$\text{Mean Absolute Percentage Error, } MAPE = \frac{\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right|}{n} \times 100\%$$

where $\boldsymbol{yi}$ is the observed value, $\boldsymbol{\hat{y}\imath}$ is the predicted value and *n* is the number of predicted values. The model with the smallest values of MAE, RMSE and MAPE will be the most appropriate model for forecasting.

3. MATERIALS AND METHODS

3.1 Source of Data

Pakistan demographic survey is the most reliable source of population dynamics data in Pakistan. A Secondary data of total fertility rates (TFR) of Pakistan during 1984-2007 has been taken from the Pakistan Demographic Surveys. These data are available for 1984-1986, 1988-1992, 1995-1997, 1999-2001, 2003 and 2005-2007. Total fertility rates are missing for rest of the years, so we estimate that using interpolating splines in R Software 3.1.2 version as shown in table 1.

3.2 Fitting ARIMA Models to TFR of Pakistan

At first, we plot TFR of Pakistan for available years in figure 1. We difference the fertility data for making the series stationary which was shown in figure 2. We plot ACF and PACF of fertility to identifying the order of ARIMA model which was shown in figure 3 and 4. We also identify best fit ARIMA model on the basis of minimum value of Akakie's Information Criteria (AIC) which was shown in table 2. However, in R software, the "auto.arima" command is also identified best fit ARIMA model. The parameters of best fit ARIMA model as shown in table 3. The forecast accuracy of ARIMA models can be measured in terms of root mean square of error (RMSE), mean absolute percentage error. (MAPE) and mean percentage error (MPE) as shown in table 4. Residuals of best fit ARIMA model was plot in figure 5. We predict total fertility rates (TFR) of Pakistan for the next fifteen years along with 80% prediction interval using the best ARIMA model as shown in figure 6 and table 5. All statistical analysis was performed in R version 3.2.1.

4. RESULTS

Table 1 shows the TFR of Pakistan from 1984-2007 and TFR for missing years (1987, 1993, 1994, 1998, 2002 and 2004) are estimated by using spline function in R.

Table 1
Total Fertility Rates (TFR) of Pakistan

Year	TFR	Year	TFR
1984	6.95	**1996**	5.30
1985	7.01	**1997**	5.05
1986	6.94	**1998***	4.79
1987*	6.71	**1999**	4.54
1988	6.49	**2000**	4.34
1989	6.40	**2001**	4.12
1990	6.21	**2002***	3.99
1991	5.99	**2003**	3.93
1992	5.84	**2004***	3.86
1993*	5.79	**2005**	3.79
1994*	5.74	**2006**	3.74
1995	5.59	**2007**	3.70

* **Estimated: 1987, 1993, 1994, 1998, 2002, 2004 (Using spline in R software)**

Figure 1 shows a plot of TFR from 1984-2007 which clearly indicate that the TFR has been declining slowly from 1984 to 1990 and then declining was slightly increased after 1990s and then decreased till 2007. Figure 2 shows the first difference series of TFR for achieving the stationary.

Figure 1: Fertility Trend of Pakistan

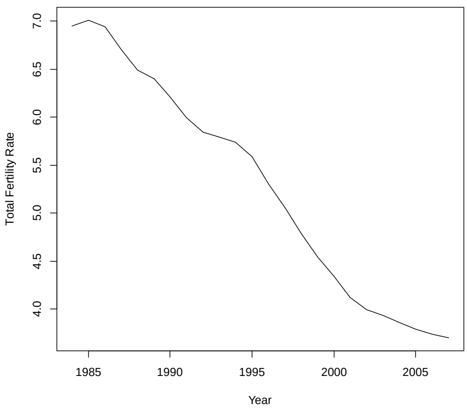
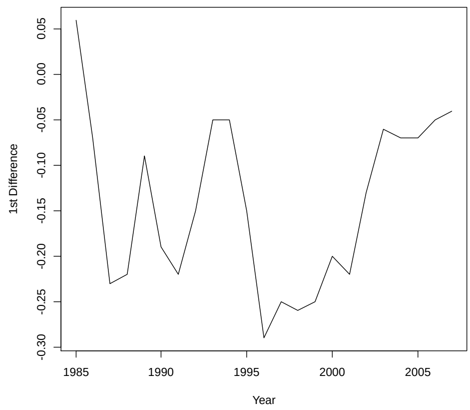


Figure 2: Fertility Trend of Pakistan After 1st Difference



In figure 3 and 4, the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) are shown of fertility data. The plot of sample ACF shows that the autocorrelation function exhibits a slow decay as the number of lag increases indicating that a trend is contained in the data and hence the series is non-stationary. Since the time series for TFR $\{Y_t: t=1,2,-----,24\}$ is not stationary, transformation is needed to eliminate the trend component. The time series, $\{Y_t\}$ is differenced at lag-1 to obtain stationary and shown in figure 2.

Figure 3: Sample ACF

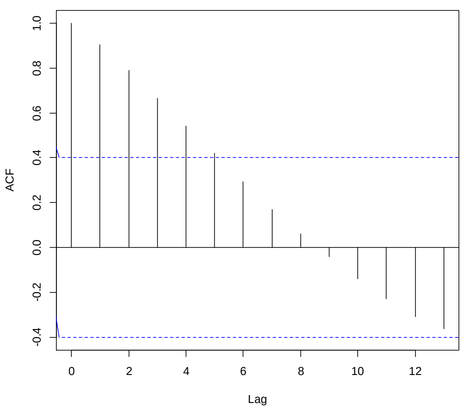
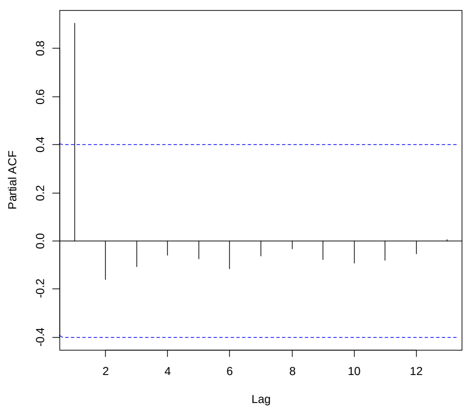


Figure 4: Sample PACF



Plausible models for fertility series can be observed from the sample ACF and sample PACF in figure 3 and 4. Furthermore, different models of ARIMA are evaluated on the basis of minimum AIC value as shown in table 2. Table 3 shows estimated parameters of best selected ARIMA (0,1,2) with drift model.

Table 2
Comparison of Different ARIMA Models

Fit Statistics	(0,1,2)	(2,1,0)	(1,1,0)	(2,1,1)
AIC	-50.47	-48.43	-48.804	-48.62

Table 3
Best Selected ARIMA Model

ARIMA (0,1.2) with drift	Estimate	SE	T	Sig.
Constant	-0.1309	0.0318	-4.11635	3.88E-05
MA1	1.073	0.2087	5.141351	2.77E-07
MA2	0.3962	0.1816	2.181718	2.91E-02

The appropriateness of ARIMA model was observed on the basis of residuals plot and after that model was considered for forecasting purpose. From figure 5, it can be seen that the residuals of ARIMA (0,1,2) model with drift show stationary, so this model can be used for prediction. Forecast accuracy of best selected ARIMA model is presented in table 4. Figure 6 shows forecast pattern of TFR of Pakistan for next fifteen years (2008 to 2022) with 80% prediction interval. Predicted values of TFR of Pakistan are also represented in table 5.

Table 4
Forecast Accuracy of Best Selected ARIMA Model

RMSE	MPE	MAPE
0.06107	-0.0223	0.91696

Figure 5: Residuals from ARIMA(0,1,2) with Drift

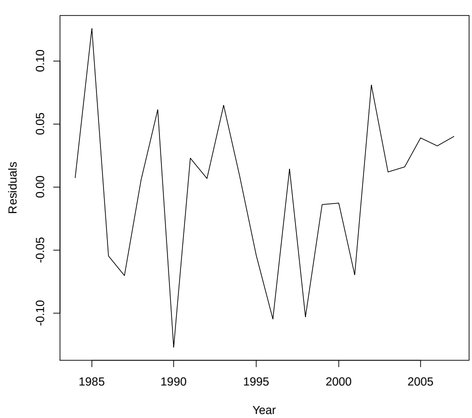


Figure 6:Forecast From ARIMA (0,1,2) with Drift

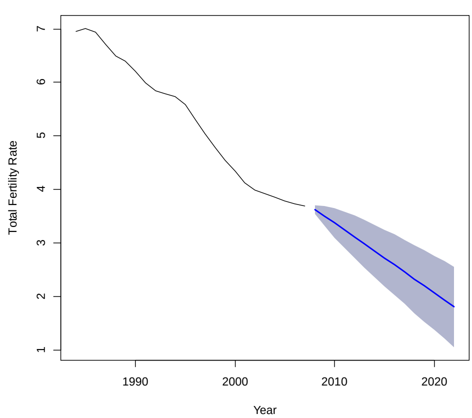


Table 5
Forecast TFR of Pakistan with 80% Prediction Interval

Year	Pakistan		
	Point	80% Prediction	
		Lower	Upper
2008	3.63	3.54	3.71
2009	3.51	3.32	3.70
2010	3.38	3.10	3.66
2011	3.25	2.91	3.59
2012	3.12	2.72	3.51
2013	2.99	2.54	3.43
2014	2.86	2.37	3.34
2015	2.72	2.20	3.25
2016	2.59	2.03	3.16
2017	2.46	1.86	3.06
2018	2.33	1.70	2.97
2019	2.20	1.54	2.87
2020	2.07	1.37	2.76
2021	1.94	1.22	2.66
2022	1.81	1.06	2.56

5. COMMENTS AND CONCLUSION

In this study, we applied different ARIMA models and we found that the ARIMA (0,1,2) with drift model is an appropriate model for forecasting TFR of Pakistan. Based on this model, the TFR of Pakistan is projected to decline and expected to be approximately 1.81 (average number of children per women) for the year 2022. Predicted values of fertility shows that fertility will be decline but other socioeconomic factors affect the fertility as well in future. To our knowledge, we first used TFR data of Pakistan for prediction of future TFR by using ARIMA models.

Basically demographic changes affect all areas of human activity economically, socially, culturally and even politically. Predicted value of fertility will help the government in planning children services and also give an idea to allocate future resources. However, the government of Pakistan has also been started different contraceptive measures to decrease the fertility of the country.

REFERENCES

1. Nasir, J.A., Tahir, M.H. and Riaz, M. (2010). Measuring and modeling the fertility profile of indigenous people in Pakistan: A study of the Arians. *Pakistan Journal of Commerce and Social Sciences*, 4(2), 132-146.
2. Alkema, L., Raftery, A.E., Gerland, P., Clark, S.J., Pelletier, F., Buettner, T. and Heilig, G.K. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography*, 48(3), 815-839.
3. Akhtar, J.A.N.M. and Tahir, M.H. (2009). Reproductivity and age-specific fertility rates in Pakistan after 1981. *Pakistan Journal of Statistics*, 25(3), 251-263.
4. Government of Pakistan (2007). *Population Demographic Survey-2007*. Federal Bureau of Statistics, Statistics Division, Islamabad, Pakistan.
5. Shitan, M. and Ng, Y.L. (2015). Forecasting the total fertility rate in Malaysia. *Pakistan Journal of Statistics*, 31(5), 547-556.
6. Soon, L. (1992). Determinants of Fertility in Malaysia: How Much Do We Know? *Journal of Southeast Asian Studies*, 23(1), 112-132.
7. Brown. R.L. (2004). Baby boom and baby bust-fertility rates and why they vary, *Contingencies January/February*, 17-20.
8. Chi, G. (2009). Can knowledge improve population forecasts at subcounty levels? *Demography*, 46(2), 405-427.
9. Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
10. Adhikari, R. and Agrawal, R.K. (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*.
11. Saboia, J.L.M. (1977). Autoregressive Integrated Moving Average (ARIMA) Models for Birth Forecasting. *Journal of the American Statistical Association*, 72(358), 264-270.
12. McDonald, J. (1979). A Time Series Approach to Forecasting Australian Total Live-Births. *Demography*, 16(4), 575-601.
13. Land, K.C. and Cantor, D. (1983). ARIMA Models of Seasonal Variation in U.S. Birth and Death Rates. *Demography*, 20(4), 541-568.
14. McNown, R. and Rogers, A. (1989). Forecasting Mortality: A Parameterized Time Series Approach. *Demography*, 26(4), 645-660.
15. Knudsen, C., McNown, R. and Rogeres, A. (1993). Forecasting Fertility: An Application of Time Series Methods to Parameterized Model Schedules. *Social Science Research*, 22, 1-23.
16. Carter, L.R. (1996). Forecasting U.S. Mortality: A comparison of box-jenkinsarima and structural time series models. *Sociological Quarterly*, 37(1), 127-140.
17. Hyndman, R.J. and Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24(3), 323-342.
18. Hyndman, R.J. and Shang, H.L. (2009). Forecasting functional time series (with discussion). *Journal of the Korean Statistical Society*, 38(3), 199-221.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 195-202

TIME SERIES ANALYSIS AND FORECASTING OF WATER RESERVOIR IN PAKISTAN

Zakia Batool[§] and **M. Yaseen**
Department of Math & Stat., University of Agriculture Faisalabad, Pakistan
[§]Corresponding author's email: zakia.batool16@gmail.com

ABSTRACT

In water resource management daily flow to the reservoir is most important factor, in this study the daily flow to the reservoir as well as upstream has been forecast by different models. Autoregressive Integrated Moving Average (ARIMA) and Autoregressive Moving Average (ARMA) models has been used to forecast daily flow of different dams and linked canals in Pakistan. Application in practical science ARIMA model plays an important role. Both ARIMA and ARMA model has been used to compare the capability of autoregressive forecast of daily dam reservoir inflow. Forecast accuracy of Tarbela Dam reservoir inflow has been increased if we increase the number of parameter in ARMA and ARIMA models. To forecast Dam reservoir inflow polynomial for ARMA and ARIMA models was derived up to four and six parameter. Root mean square error concludes that ARIMA model can be used to forecast the level of water for different rivers in Pakistan with less error than ARMA model.

KEYWORDS

Auto-regressive models, Dam Reservoir, Stream flow Forecasting.

1. INTRODUCTION

Hydrological processes are influenced by not only deterministic, but also stochastic factors (Wang et al., 2007). Hydrological data has been used to grab the stochastics process which is very important factor for time series stationarity. In analysis of time series trended data can be removed by the use of ARIMA model approach that converts non stationary data into stationary. Environmental factors like downstream, reservoir, upstream are important in hydrological and metrological department to study the trend and variation of these factors in time series analysis. Water resources plays and important role in economic development. Communities of water resource management, it is known to be conflict shortage of water issues, Water management maximizing efficiency forecast based on Stream flow is fundamental. ARMA and ARIMA models have been used to forecast daily inflow to reservoir. Autoregressive Moving Average model (ARMA) is a stochastic approach for presenting time series modelling. This enforces us to achieve stationarity on Autoregressive (AR) and inevitability on Moving Average (MA) parameters. There are two main reasons to fitted stochastic models to hydrological data: Generating sequence for synthetic data to forecast one or more time period ahead on the other hand deterministic models are used to forecast short time flow such as hourly. Deterministic models have great advantage then stochastic models even if deterministic model parameter cannot be measured or have no physical clarification; so for forecasts probability limits are obtained, it gives an advantage to use in stochastic modelling. In water resource management system stochastic models are used to evaluate the performance of future. Time series modelling for hydrological data numerous stochastic models have been suggested. These contain ARMA models, disaggregation models, and concept of pattern recognition based models. Mostly modelling for time series data fall inside the context of multivariate ARMA models. Box-Jenkins models also known as ARIMA has been widely used time series models to forecast hydrological events, weather prediction and predicting traffic flow. ARIMA models are commonly used to predict univariate time series (Halim, *et al*., 2007). According to Box and Jenkins (1976) aim of ARIMA modelling is to construct the most suitable model for observed data. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used to examine the behavior of system under time and also help to select the suitable parameters for the model. ARIMA model can be develop through the use of three stage modelling procedure Identification of parameters , Estimation and Diagnostic checking purposed by Box and Jenkins (1976). Aim of this study is to develop forecast model for the daily dam reservoir inflow using ARMA and ARIMA models.

2. MATERIAL AND METHODS

Data of daily inflow to the reservoir for Tarbela dam has been used to forecast for future. The data have been taken from Water and Power Development Authority (WAPDA). Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) have been used to forecast Tarbela dam reservoir. Data of daily inflow to the reservoir from 2006 to 2012 has been selected.

In statistics, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average or (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series. The ARIMA model is applied in some cases where data show evidence of non stationarity, where an initial differencing can be applied to remove the non stationarity. The model is generally referred to as an ARIMA (p,d,q) model where p, d, and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. ARIMA models form an important part of the Box-Jenkins approach to time-series modelling.

A process, $\{x_t\}$ is said to be ARIMA (p,d,q) if $\nabla^d x_t = (1-\mathrm{B})^d x_t$ is ARMA (p,q).In general, we will write the model as

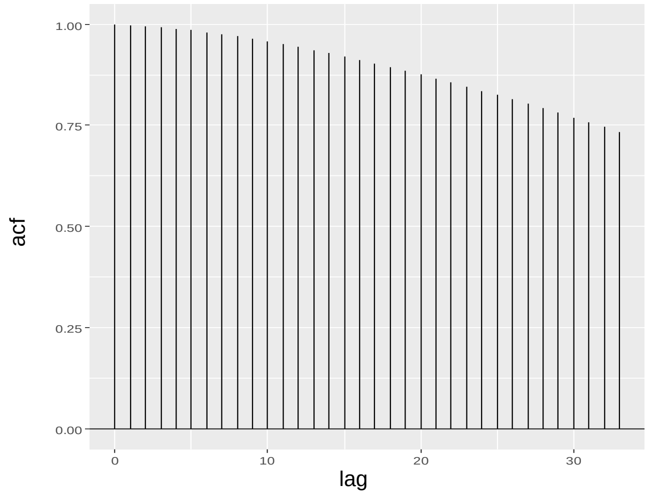
$$\phi(B)(1-B)^d x_t = \theta(B)w_t, w_t \sim WN(0,\sigma^2)$$

WN stands for white noise. Here, we define the backshift operator by $B^k x_t = x_{t-k}$ and the autoregressive operator and moving average operator are defined as follows:

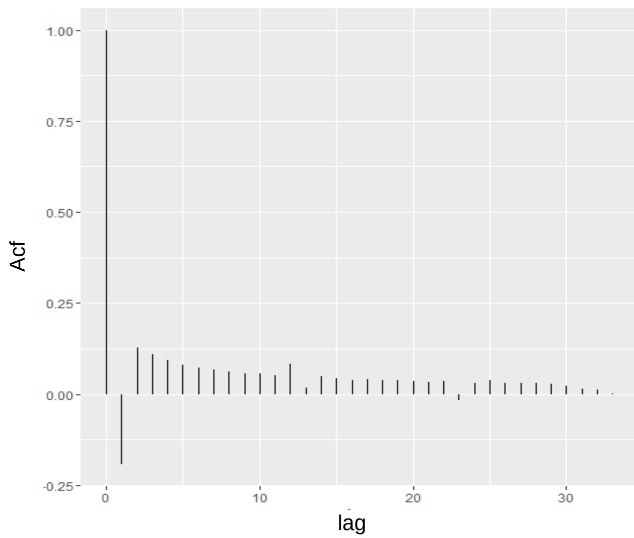
$$\phi(\mathrm{B}) = 1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

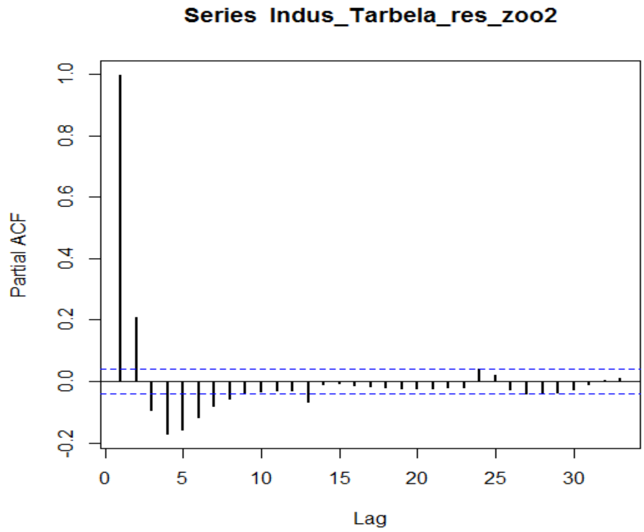
$$\theta(\mathrm{B}) = 1+\theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$

$\phi(B) \neq 0$ for $|B| \leq 1$, the process $\{x_t\}$ is stationary if and only if d=0, in which case it reduces to an ARMA(p,q) process. ACF and PACF plots are used to determine the parameter of ARMA and ARIMA models (Cryer and Kung-Sik, 2008; Mohammadi *et al.*, 2006). These plots are represented in following figures in which PACF and ACF shows at vertical line and delay times shows at horizontal line.



Also ACF and PACF plots are used to identify the stationarity in the data, for stationary series ACF drops quickly from zero, exponentially decay shows the non-statinarity in the data. If the time series is not stationary then make it stationary by using differencing.





So the series become stationary and see that ACF shows the order or MA component which is one because only one spike is outside the confidence interval. Plot of ACF shows the amount of delay 1 are high. So choosing one autoregressive parameter is sufficient (Karamouz and Araghinejad, 2012), but in order to better forecast for the future we have to increase the number of parameter. PACF shows that the order of autoregressive parameter is 2. Next step is model diagnostics that is concerned with testing the goodness of fit of a model and, if the fit is poor, suggesting appropriate adjustments. Analysis of residuals from the fitted model is used to check the model adequacy. Akaike Information Criteria (ACI) Bayes Information Criteria to select the best model. Estimate different model for different parameter then select the model which have minimum AIC or BIC.

3. CRITERION TO SELECT THE BEST STRUCTURE OF ARMA AND ARIMA MODELS

In order to select the best structure between ARMA and ARIMA the root mean square error and the mean bias error were used as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n} (Q_{ci} - Q_{oi})^2 / n}$$

where, RSME is the root mean square error, i is the number of days, Q_{ci} is the computational discharge in day i, the Q_{oi} observational discharge in a day i, and n is the total number observations in the data.

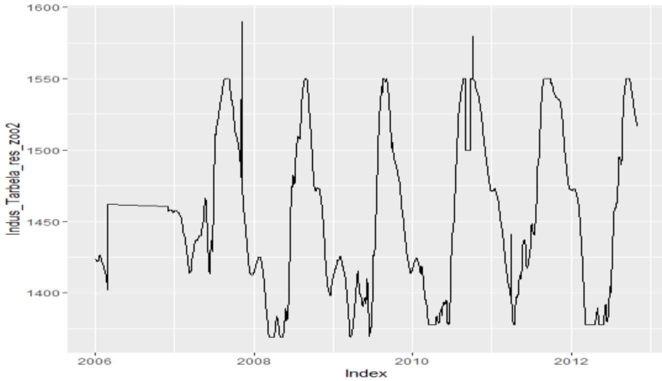
4. RESULTS

Results from forecasting the daily inflow to the reservoir shows that ARIMA (2, 1, 1) model is best

5. DISCUSSION

Measurements of the daily average water level in the Rivers of Pakistan were used in this study to build a forecasting model using ARIMA and ARMA. To observe the pattern of the data set ACE and PACF plots are used. For forecasting and water-level modelling an ARIMA model was found appropriate. ARIMA model gives better performance because of its importance in time series stationarity. In above figure ACF and PCF plots are used to check the stationary of the data. The graph shows that the data for the Tarbela reservoir was become stationary after taking differencing.

General behavior of the series



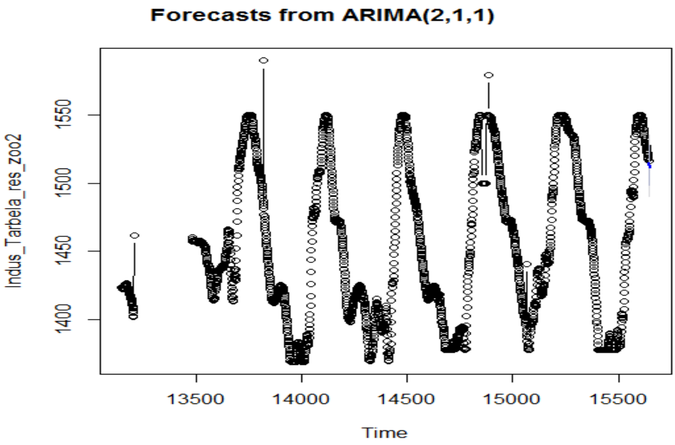
Results from forecasting daily flow to reservoir for different year:

Best structure (ARMA)	Best structure (ARIMA)
ARMA(1,0)	ARIMA(1,1,1)
ARMA(2,0)	ARIMA(2,1,1)
ARMA(4,0)	ARIMA(2,1,2)
ARMA(2,1)	ARIMA(1,1,2)

Using Aikek information criteria selection of best model is made. AIC shows that ARMA(1,0) and ARIMA(2,1,1) are best model for the Tarbela Dam reservoir.

Model	Best structure in considering ACF and PACF	RSME
ARMA	ARMA(1,0)	20.12
ARIMA	ARIMA(2,1,1)	18.75

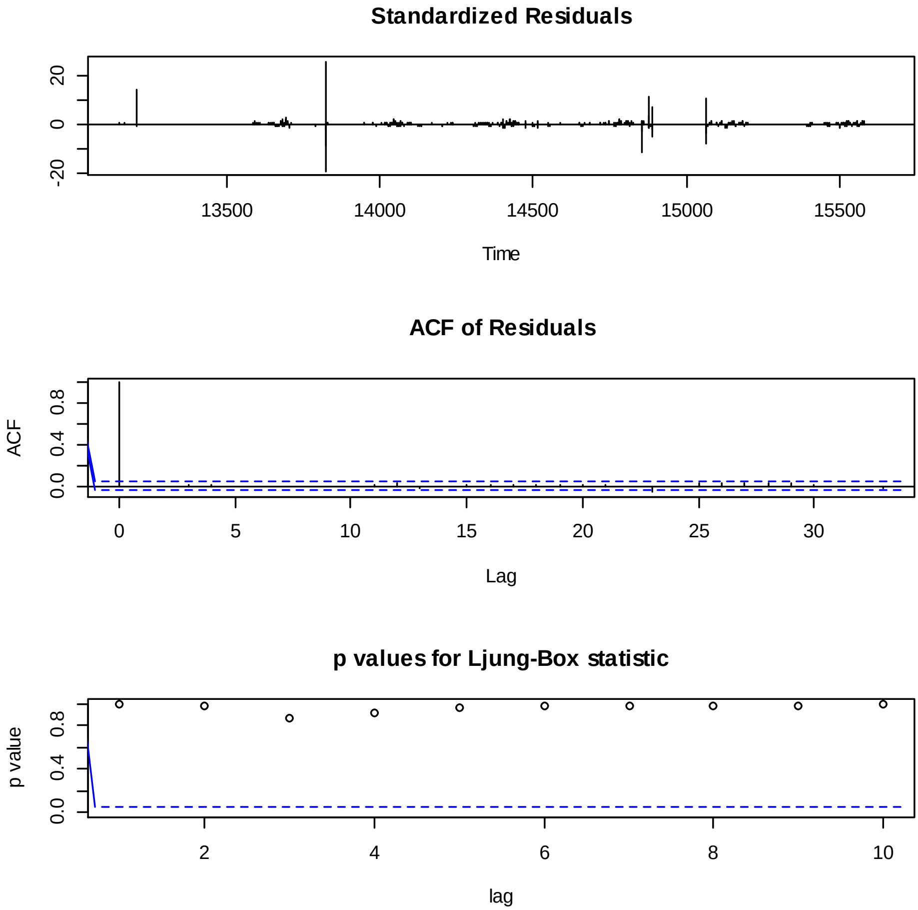
It is not possible to show all the results of different best structure by using ARMA and ARIMA model. RSME shows that ARIMA(2,1,1) is best model for future forecast of Tarbela Dam reservoir inflow, from all the models between ARMA and ARIMA model.
Forecast based on ARIMA model



Forecast points

95% Limits			
Point	**Forecast**	**Lower**	**Upper**
15644	1515.327	1504.684	1525.970
15646	1514.798	1501.531	1528.066
15647	1514.281	1498.703	1529.858
15648	1513.795	1493.189	1533.474
15649	1512.893	1490.517	1535.268
15650	1512.476	1487.890	1537.062
15651	1512.080	1485.299	1538.861
15652	1511.704	1482.744	1540.665

Above study shows that ARIMA (2, 1, 1) is better model to forecast future than ARMA because ARIMA contains stationarity in the model. Diagnostic checking for the above forecast model is given in the figure.



Three diagnostic tools in one display show in the above figure, standardized Residuals, ACF of residual, and P-value of Ljung- Box test statistics. Here the ACF of residual shows that ARIMA (2, 1, 1) is best model for forecasting. To increase forecast accuracy we should increase the parameter in the model.

6. CONCLUSION

In this paper, ARMA and ARIMA models are used to compare in Forecasting daily reservoir inflow at Tarbela Dam. Forecast accuracy was determined by increasing the number of Autoregressive and Moving Average parameters in the models. ARIMA model are found to be better than ARMA because of its importance in time series stationarity. For comparing ARMA and ARIMA we use root mean square error that shows ARIMA is better than ARMA. ARIMA model could be used for forecasting daily inflow to the reservoir. If we adopt appropriate procedures, if another dam is built to store water in the reservoir, there will be sufficient water to produce Hydro power.

REFERENCES

1. Ahmad, S., Khan, I.H. and Parade, B.P. (2001). Performance of stochastic approaches for forecasting river water quality. *Water Res.*, 35, 4261-4266.
2. Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, California, USA.
3. Cryer, J.D. and Chan, K.S. (2008). *Time Series Analysis: With Application in R*. 2nd Edn., Springer, New York.
4. Halim, S., Bisono, I.N. Melissa, and Thia, C. (2007).Automatic Seasonal Auto Regressive Moving Average Models and Unit Root Test Detection, *IEEE International Conference on Industrial Engineering and Engineering Management*, Singapore. Pp. 1129-1133.
5. Karamouz, M. and Araghinejad, S. (2012). *Advance Hydrology*. 1st Edn., Amirkabir University of Technology Press, Tehran.
6. Mohammadi, K., Eslami, H.R. and Kahawita, R. (2006). Parameter estimation of an ARMA model for river flow forecasting using goal programming. *J. Hydrology*, 331, 293-299.
7. Otache Y.M., Sadeeq, M.A. and Ahaneku, I.E. (2011). ARMA Modelling of Benue River Flow Dynamics: Comparative Study of PAR Model. *Open Journal of Modern Hydrology*, 1, 1-9.
8. Srikanthan, R., McMohan, T.A. and Irish, J.L. (1983). Time series analysis of annual flows of Australian streams. *J. Hydrology*, 266(1-4), 213-22.
9. Toth, E., Brath, A. and Montanari, A. (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. *J. Hydrol.*, 239, 132-147.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 203-214

AN EMPIRICAL INVESTIGATION OF ECONOMIC VARIABLES OF PAKISTAN BY USING AUTOREGRESSIVE DISTRIBUTED LAG MODEL

Hira Zafar, Muhammad Arif and **Muhammad Yaseen**
Department of Mathematics & Statistics,
University of Agriculture, Faisalabad, Pakistan
Email: hirastat@hotmail.com
arif_uaf@hotmail.com
myaseen208@gmail.com

ABSTRACT

In this research two main techniques of time series analysis, Co-integration and Autoregressive Distributed Lag (ARDL) model were discussed with the advantages of one technique over the other. This study used to investigate the relationship between five economic variables of Pakistan by using monthly data for the period 1975 to 2015. ARDL model was used to study the relationship between Exchange Rate on Consumer Price Index, Money Supply, Imports and Exports. This work inspects the Short Run (SR) dynamic and Long Run (LR) relationship with the aid of autoregressive distributed lag (ARDL) model**.** First apply Augmented Dicky-Fuller (ADF) test and Phillips-Peron (PP) Test for checking the stationary of variables. The second step was ARDL-Bound Test and it concluded that variables have long run relation with each other. After ARDL-Bound Test next step was selection of lag length for ARDL model. The diagnostic test inspected that model is good to fit. The ARDL model exposed significant and non-significant long run relation. The coefficient of ECM (–1) indicates how much of the disequilibrium in the short-run was fixed (eliminated) in the long-run.

KEY WORDS

ARDL, Co-integration, Exchange Rate, Consumer Price Index, Money Supply, Imports, Exports.

1. INTRODUCTION

The intention of this research is to establish the relation between five different economic variables; Consumer Price Index (CPI), Money Supply (MS), Exchange Rate (ER), Imports (IMP) and Exports (EXP). Data collected from State Bank of Pakistan (SBP) over the period 1975 to 2015.

There are many techniques which deal time series variables, here two main methods are discussed and compared. The first one is Co-integration and second is Autoregressive Distributed Lag (ARDL) Model.

When Integrated term used it means that, if a time series is non-stationary and has to be differenced at order one to make it stationary, then this time series is familiar as integrated at order one and denoted by I(1) and if a time series is stationary without taking any difference then this time series is integrated at order zero and denoted by I(0).

Granger (1981) introduced Co-integration technique. The concept behind this technique is that, if two variables are integrated at same order say (d) and a combination between them which is linear in nature is integrated of any order less than (d), then these two variables are said to be co-integrated.

The basic use of this technique is finding the long run relation between economic variables. If two variables are co-integrated then the long run relationship must exist (Khan and Ali, 1997). Co-integration technology is the fundamental variables of the same in terms of orders that are attached (Bakhat and Wurzburg, 2013).

In time series analysis, many times on the same order of variables are not added and in this situation co-integration technique is not reliable. Autoregressive Distributed Lag (ARDL) model overcomes this problem and also have many advantages as compared to co-integration method.

The (ARDL) model is established by Pesaran and Pesaran (1997) and Pesaran et al. (2001). The ARDL is becoming more useful method because of its many advantages. The first ARDL is a single equation structure based model. The model estimates both long and short run parameters simultaneously and explain the relationship between multiple variables (Alam and Ahmad, 2011).

Second ARDL models combine all the variables on the same order, that property is not required. It can be used whether variables are stationary at different order. Third this technique is reasonably stronger in small or finite samples (Naiya and Manap, 2013). Fort Turning out the way they model, as with different length surpassed sets various variables (Nchor et al., 2015).

By comparing the advantages of ARDL model with co-integration method, it is clear that ARDL model is superior to investigate the relationship between multiple time series.

2. MATERIAL AND METHODOLOGY

This empirical work employs monthly time series data on are Consumer Price Index (CPI), Money Supply (MS), Exchange Rate (ER), Imports (IMP) and Exports (EXP) for the period of 1975 to 2015. Data collected from State Bank of Pakistan (SBP). This study examine the long run and short run relationships between these economic variables. The Autoregressive Distributed Lag model is used for establishing the relationship between economic variables.

3. AUTOREGRESSIVE DISTRIBUTED LAG MODEL (ARDL)

The ARDL model is preferable to the other cointegration methods. One reason is that the model does not involve the pre- testing variables, which means the test on existence of relationship between variables in level can be applicable irrespective of whether the underlying repressors are purely I(0), purely I(1) or fractionally cointegrated (Pesaran and Pesaran 1997). Furthermore, the Error Correction Model (ECM) can be derived from ARDL model by a simple linear transformation (Banerjeeetal. 1993). ECM integrated short run adjustments with long - run equilibrium without losing the long run information. Another reason is the small sample properties of ARDL method (such as this study) are more robust and performs better to that of Johansen and Juselius's (1990) cointegration method (Pesaran and Shin 1999).

4. ERROR CORRECTION MODEL

The error correction model (also referred as short run model) has co-integrated restrictions incorporated into the specification, so that it is designed for use with non-stationary series are co-integrated. It restricts the long run behavior of endogenous variables to converge to their co-integrating relationship via the error correction term while allowing a short-run dynamics.

If there a long run relationship between different variables exists then an error correction process is also taking place. Error correction model indicates the speed of adjustment towards the long-run equilibrium after a short run shock Engle and Granger (1987).

5. STATIONARITY AND NON-STATIONARITY OF TIME SERIES DATA

The first step in time series analysis is to study the stationery property of the variables. The first step in conducting a time series analysis is therefore, conducting unit roots tests to determine the unit roots properties of the variables. There are various unit root tests, but this study Augmented Dickey-Fuller (ADF) unit root test is used. While the ARDL approach allows both I(1) and I(0) series, it is still important to check the possibility that any of the series are not I(2) Banerjee, et al., (2003).

6. BOUNDS TESTING

Next step is to use the bounds testing approach to investigate the long run relationship of the variables. Initially, bounds testing was employed through the ordinary least squares (OLS) method and computed the F-statistics for the joint significance of the lagged levels of variables to compare with the critical values provided in Table of (Pesaran et al., 2001).

7. LAG LENGTH CRITERIA

The topic of order determination has attracted considerable attention in the literature of time series and in those areas of research which are closely related to time series analysis such as econometrics and statistics. It is rarely the case that the 'true' order of a process is known. To overcome this difficulty several order selection criteria had been proposed in the literature but we don't have any criterion which could be considered as the best criterion in all situations. The lag order for ARDL model is selected by using Schwarz information criterion and Akaike information criterion (Shahbaz and Islam, 2011).

8. ARDL MODELING

The ARDL model for five economic variables is:

$$\begin{aligned}\mathrm{ER_t} = {} & \alpha_0 + \delta_1\mathrm{ER_{t-1}} + \cdots + \delta_k\mathrm{ER_{t-k}} + \gamma_1\mathrm{MS_t} + \gamma_2\mathrm{MS_{t-1}} + \cdots + \gamma_k\mathrm{MS_{t-k}} + \\ & \tau_1\mathrm{CPI_t} + \tau_2\mathrm{CPI_{t-1}} + \cdots + \tau_k\mathrm{CPI_{t-k}} + \eta_1\mathrm{IMP_t} + \eta_2\mathrm{IMP_{t-1}} + \cdots + \\ & \eta_k\mathrm{IMP_{t-k}} + \theta_1\mathrm{EXP_t} + \theta_2\mathrm{EXP_{t-1}} + \cdots + \theta_k\mathrm{EXP_{t-k}} + \varepsilon_t \end{aligned} \tag{1}$$

where

α_0 = Intercept
ER = Exchange Rate
MS = Money Supply
CPI = Consumer Price Index
IMP = Imports
EXP = Exports
ε_t = white noise term

$\delta_{1,\cdots,k}$,$\gamma_{1,2,\cdots,k}$,$\tau_{1,2,\cdots,k}$,$\eta_{1,2,\cdots,k}$ and $\theta_{1,2,\cdots,k}$ are short run Exchange, Money Supply Consumer Price, Imports and Exports elasticities of the related period respectively. In order to determine the long run elasticities it is assumed that in the long run

$$\mathrm{ER}^* = \mathrm{ER_t} = \mathrm{ER_{t-1}} = \mathrm{ER_{t-2}} = \cdots = \mathrm{ER_{t-k}} \tag{2}$$

$$\mathrm{MS}^* = \mathrm{MS_t} = \mathrm{MS_{t-1}} = \mathrm{MS_{t-2}} = \cdots = \mathrm{MS_{t-k}} \tag{3}$$

$$\mathrm{CPI}^* = \mathrm{CPI_t} = \mathrm{CPI_{t-1}} = \mathrm{CPI_{t-2}} = \cdots = \mathrm{CPI_{t-k}} \tag{4}$$

$$\mathrm{IMP}^* = \mathrm{IMP_t} = \mathrm{IMP_{t-1}} = \mathrm{IMP_{t-2}} = \cdots = \mathrm{IMP_{t-k}} \tag{5}$$

$$\mathrm{EXP}^* = \mathrm{EXP_t} = \mathrm{EXP_{t-1}} = \mathrm{EXP_{t-2}} = \cdots = \mathrm{EXP_{t-k}} \tag{6}$$

Here assumed that variables have attained some stable position values and are no longer changing.

So substitution these into Equation (1) yields:

$$\begin{aligned}\mathrm{ER_t^*} = {} & \alpha_0 + \delta_1\mathrm{ER_t^*} + \cdots + \delta_k\mathrm{ER_t^*} + \gamma_1\mathrm{MS_t^*} + \gamma_2\mathrm{MS_t^*} + \cdots + \gamma_k\mathrm{MS_t^*} \\ & +\tau_1\mathrm{CPI_t^*} + \tau_2\mathrm{CPI_t^*} + \cdots + \tau_k\mathrm{CPI_t^*} + \eta_1\mathrm{IMP_t^*} + \eta_2\mathrm{IMP_t^*} + \cdots \\ & +\eta_k\mathrm{IMP_t^*} + \theta_1\mathrm{EXP_t^*} + \theta_2\mathrm{EXP_t^*} + \cdots + \theta_k\mathrm{EXP_t^*} \end{aligned} \tag{7}$$

and re-arranging Equation (7) gives:

$$\begin{aligned}(1 - \delta_1 - \cdots - \delta_k)\mathrm{ER_t^*} = {} & \alpha_0 + \left(\gamma_1 + \gamma_2 + \cdots + \gamma_k\right)\mathrm{MS_t^*} \\ & +(\tau_1 + \tau_2 + \cdots + \tau_k)\mathrm{CPI_t^*} + \left(\eta_1 + \eta_2 + \cdots + \eta_k\right)\mathrm{IMP_t^*} \\ & +(\theta_1 + \theta_2 + \cdots + \theta_k)\mathrm{EXP_t^*} \end{aligned} \tag{8}$$

and rearranging further gives:

$$\begin{aligned}\mathrm{ER_t^*} = {} & \frac{\alpha_0}{(1-\delta_1 - \cdots - \delta_k)} + \frac{(\gamma_1+\gamma_2+\cdots+\gamma_k)}{(1-\delta_1 - \cdots - \delta_k)}\mathrm{MS_t^*} + \frac{(\tau_1+\tau_2+\cdots+\tau_k)}{(1-\delta_1 - \cdots - \delta_k)}\mathrm{CPI_t^*} \\ & + \frac{(\eta_1+\eta_2+\cdots+\eta_k)}{(1-\delta_1 - \cdots - \delta_k)}\mathrm{IMP_t^*} + \frac{(\theta_1+\theta_2+\cdots+\theta_k)}{(1-\delta_1 - \cdots - \delta_k)}\mathrm{EXP_t^*} \end{aligned} \tag{9}$$

where

$\frac{(\gamma_1+\gamma_2+\cdots+\gamma_k)}{(1-\delta_1-\cdots-\delta_k)}$ long run Money Supply elasticitie,

$\frac{(\tau_1+\tau_2+\cdots+\tau_k)}{(1-\delta_1-\cdots-\delta_k)}$ long run Consumer Priceelasticitie,

$\frac{(\eta_1+\eta_2+\cdots+\eta_k)}{(1-\delta_1-\cdots-\delta_k)}$ long run Imports elasticities and

$\frac{(\theta_1+\theta_2+\cdots+\theta_k)}{(1-\delta_1-\cdots-\delta_k)}$ long run Exports elasticity

The next step of the methodology is to check the goodness of fit for model. R^2, Adjusted R^2, Wald F-statistic used for goodness of fit (Atif et al., 2010)

9. RESULTS

The first step in any time series empirical analysis is to test for the presence of unit root in order to avoid the problem of spurious regression. Akaike information criterion (AIC) is used to determine the lag length while testing the stationarity of all variables. At first zero is selected as maximum lag. Following table presents the results of the test of series for the unit root using ADF and PP test.

Null Hypothesis = Variables has a unit root

Table 1
ADF Test at Level

Variables	ADF Test Statistics	P-value	Conclusion
ER	6.766	1.0000	Non-Stationary
CPI	-0.7831	0.3766	Non-Stationary
MS	13.9854	1.0000	Non-Stationary
IMP	-0.1349	0.6368	Non-Stationary
EXP	-0.2552	0.5939	Non-Stationary

Table 2
PP Test at Level

Variables	PP Test Statistics	P-value	Conclusion
ER	4.5935	1.0000	Non-Stationary
CPI	-0.7799	0.3780	Non-Stationary
MS	34.3186	1.0000	Non-Stationary
IMP	0.3755	0.7923	Non-Stationary
EXP	1.0055	0.9174	Non-Stationary

where ER = Exchange Rate
CPI = Consumer Price Index
MS = Money Supply
IMP = Imports
EXP = Export

Table 3
ADF Test at 1st Difference

Variables	ADF Test Statistics	P-value	Conclusion
ER	-12.8614	0.0000	Stationary
CPI	-21.9909	0.0000	Stationary
MS	-20.1445	0.0000	Stationary
IMP	-33.1062	0.0000	Stationary
EXP	-36.7808	0.0000	Stationary

Table 4
PP Test at 1st Difference

Variables	PP Test Statistics	P-value	Conclusion
ER	-13.1061	0.0000	Stationary
CPI	-21.9911	0.0000	Stationary
MS	-25.5096	00000	Stationary
IMP	-41.8596	0.0001	Stationary
EXP	-46.7862	0.0001	Stationary

According to table (1) the P-value are (1.0000, 0.3766, 1.0000, 0.6368, 0.5939) respectively to all variables are greater than $\alpha = 0.05$ so that null hypothesis cannot be rejected. It can be concluded that these variables have unit roots (non-stationary behavior) and these results compare by PP test table (2) which also showed that all variables are non-stable at level. Again apply these two test for all five variables at first difference.

The outcomes indicated in table 4.9 and 4.10 by using PP and ADF technique all five variables converted to stationary after taking first difference and the integrating order is one and ER ~ I(1), CPI ~ I(1), MS ~ I(1), IMP ~ I(1), EXP ~ I(1). So ARDL model can be used because none of the variable is not I(2).

Estimate this model by using OLS in Eviews-8

$$\begin{aligned}\Delta ER_T = {} & \alpha_0 + \beta_1 ER_{t-1} + \beta_2 MS_{t-1} + \beta_3 CPI_{t-1} + \beta_4 IMP_{t-1} + \beta_5 EXP_{t-1} \\ & + \textstyle\sum_{i=1}^{p} \delta_i \, \Delta ER_{t-i} + \sum_{i=1}^{p} \gamma_i \Delta MS_{t-i} + \sum_{i=1}^{p} \tau_i \Delta CPI_{t-i} \\ & + \textstyle\sum_{i=1}^{p} \eta_i \Delta IMP_{t-1} + \sum_{i=1}^{p} \theta_i \Delta EXP_{t-1} \qquad (10)\end{aligned}$$

Now by using Wald Test Bound Testing for ARDL Co-integration test the null hypothesis that is $H_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$ and alternative hypothesis $H_{01} \neq \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4 \neq \beta_5$ OR H_0 = There is no long - run relationship and H_1 = There is long - run relationship

For this purpose use Eviews 8 and option restricted Wald test was used.

Table 5
Wald Test Bound Testing for ARDL Co-integration

Wald Test		
Null Hypothesis:	C(2)=C(3)=C(4)=C(5)=0	
F-statistic	12.8395	p-value = 0.0000
Chi-square	51.3438	p-value = 0.0000

Both the p-values is less than $\alpha = 0.05$, so H_0 is rejected. The result shows that there is long term relation between Exchange Rate and Imports, Money Supply, Exports.
The presence of long run relation has been established through the Bound Test. Now the next step is testing, for each variable to determine the appropriate number of lags.

Lag	**AIC**
0	2.059521
1	1.851866
2	1.855913
3	1.855522
4	1.851532
5	1.855804
6	1.853225
7	1.854768
8	1.856004
9	1.841566*
10	1.842671

ARDL (7,0,0,7,0) selected based on Akaike Information Criterion

Dependent variable is Exchange Rate

472 observations used for estimation from 1975 M10 to 2015 M1

Table 6
Autoregressive Distributed Lag Estimates

Regressors	Coefficient	p-value
ER(-1)	1.3847	0.000
ER(-2)	-0.43174	0.000
ER(-3)	0.16377	0.045
ER(-4)	-0.21611	0.007
ER(-5)	0.071330	0.375
ER(-6)	-0.11500	0.153
ER(-7)	0.15514	0.002
CPI	0.8415E-3	0.089
MS	-0.2702E-6	0.000
IMP	0.6830E-3	0.001
IMP(-1)	0.2533E-3	0.197
IMP(-2)	-0.4662E-3	0.018
IMP(-3)	-0.5933E-3	0.004
IMP(-4)	0.9855E-3	0.000
IMP(-5)	-0.1889E-4	0.925
IMP(-6)	.4848E-3	0.018
IMP(-7)	-.3663E-3	0.043
EXP	-.7742E-3	0.003
INPT	-0.49347	0.003

Before interpret the result apply diagnostic checking for that ARDL model to check the validation of model and then interpret the results of ARDL model.

Table 7
Diagnostic Tests

R-Squared	0.99963
R-Bar-Squared	0.99961
S.E. of Regression	0.56317
F-stat. F(18, 453)	67336.8[.000]
DW-statistic	2.0066
A: Serial Correlation	
LM –Version	CHSQ (12) = 19.1187[.086]
F-Version	F(12, 441) = 1.5514[.103]
B: Functional Form	
LM –Version	CHSQ (1) = .95967[.327]
F-Version	F(1, 452)= .92088[.338]

A = Lagrange multiplier test of residual serial correlation
B = Ramsey's RESET test using the square of the fitted values

Table (7) shows the results of diagnostic checking for the model ARDL (7,0,0,7,0). For this model

R^2, $\overline{R^2}$ show high values which show goodness of model, F-test is significant at 1% which specifies goodness of fit. The calculated value of Durbin Watson is equals to two and Lagrange multiplier test of residual serial correlation, Ramsey's RESET test using the square of the fitted values show significant results. So the ARDL (7,0,0,7,0) model fulfill all the criteria of goodness of a model.

Now ARDL (7,0,0,7,0) model is used to establish the long and short run of variables. In this model Exchange Rate (ER) is dependent variable and Imports (IMP), Money Supply (MS), Exports (EXP) and Consumer Price Index (CPI) are independent variables.

ER and CPI has insignificant relation. ER has negative but significant relation with IMP. ER has negative but significant relation with EXP. ER has negative but significant relation with MS.

Table 8
Long Run Coefficients using the ARDL Approach

Regressor	Coefficient	P-value
CPI	-.069442	0.078
MS	.2230E-4	0.000
IMP	-.079388	0.000
EXP	.063893	0.000
INPT	40.7236	0.001

The results of table (8) indicates that Exchange Rate has significant long term relation with Imports, Money Supply and Exports but insignificant relation with Consumer Price Indices. The relation establish these variables according to table (7) exist from long period of time and to check which point of adjustment equilibrium exist Error Correction Model is apply.

Table 9
Error Correction Representation for the Selected ARDL Model

Regressors	Coefficient	p-value
dER 1	.37260	0.000
dER 2	-.059133	0.243
dER 3	.10464	0.038
dER 4	-.11147	0.025
dER 5	-.040142	0.433
dER 6	-.15514	0.002
dCPI	.8415E-3	0.089
dMS	-.2702E-6	0.000
dIMP	.6830E-3	0.001
dIMP 1	-.2570E-4	0.912
dIMP 2	-.4919E-3	0.039
dIMP 3	-.0010852	0.000
dIMP 4	-.9967E-4	0.664
dIMP 5	-.1186E-3	0.571
dIMP 6	.3663E-3	0.043
dEXP	-.7742E-3	0.003
dINPT	-.49347	0.003
ecm(-1)	-.012117	0.003

R-Squared	0.33134
R-Bar-Squared	0.30477
S.E. of Regression	0.56317
F-stat. F(18, 453)	13.2045[.000]
DW-statistic	2.0066

Model to examine the short-run dynamics, ECM is utilized. ECM term value is represented by ECT (-1) Long-term variables in the system to get back to the states on the basis of which the error correction terms. The short-run, the rate of loss of balance that shows the relationship in the long-run disequilibrium. Precisely, ECM correct disequilibrium and it is after the estimated coefficient statistically significant negative and that should be later considered balance between being long-run relationship variables.

Giving to outcomes in table (9) short time elasticities of imports, money supply and exports are -.0010852, -.2702E-6, and -.7742E-3 respectively. It is substance stating that the elasticities are considerable lower than long term elasticities. ECM is attained from long term relation and error terms from ECM (-1) is one part lag value. The quantity of ECM(-1) specifies how considerably the disequilibrium in the short term was be eliminated in the long term. As expected ECM term has been originate negative and statistically significant. The quantity of the ECM recommends that modification procedure is reasonably fast and 12% of last year’s disequilibrium in Exchange Rate from equilibrium route was modified in the current year.

10. COMMENTS AND CONCLUSION

The study has used the ARDL Approach to co- integration and the error correction model (ECM). The study examined the long and short run relationship. First apply Augmented Dicky- Fuller (ADF) test and Phillips-Peron (PP) Test for checking the stationary of variables which The main objective of the study was to analyze the impact of Exchange Rate on Consumer Price Index, Money Supply, Imports and Exports from the period 1975 to 2015 for the monthly data of Pakistan by consuming Auto Regressive Distributed Lag methodology established on bounds test technique suggested by Pesaran et al. (2001). ARDL methodology has be presented as supplementary powerful technique to discover the long term relation and also short term elasticities and yields reliable estimates of long term coefficients that are asymptotically normal irrespective of whether the underlying variables are I(0) or I(1).

Time series have been verified to inspect econometric difficulty of unit root which was examined by using ADF Test and Phillip Parren test the outcomes shows that the variables are not stationary at level, so apply these test again. At the first difference variables are stationary.

The second step of ARDL approach was Wald statistics for ARDL-Bound Test. By using ARDL-Bound Test it was concluded that Exchange Rate has long run relation with other four economic variable. After ARDL-Bound Test next step was selection of lag length for ARDL model. Minimum value of AIC is equals to lag 9. By using Microfit 4.1 ARDL (7,0,0,7,0). The diagnostic test R-square, Adjusted-R-square, DW-statistic and Lagrange Multiplier test of residual serial correlation, Ramsey's RESET test using the square of the fitted values, indicates that model is good. The ARDL model shows that Exchange Rate has significant relation with Exports, money supply and Imports but insignificant relation with CPI.

Long term Coefficients using the ARDL Approach indicates that Exchange Rate has significant long term relation with Imports, Money Supply and Exports but insignificant relation with Consumer Price Indices. The relation establish these variables according to ARDL exist from long period of time and to check which point of adjustment equilibrium exist Error Correction Model is apply.

The quantity of ECM(-1) specifies how considerably the disequilibrium in the short term was be eliminated in the long term. As expected ECM term has been originate negative and statistically significant. The quantity of the ECM recommends that modification procedure is reasonably fast and 12% of last year’s disequilibrium in Exchange Rate from equilibrium route was modified in the current year.

11. ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful.

Alhamdulillah, Bounteous praises to Allah for the strengths and His blessing in completing this research paper. My all praises are due to the HOLY PROPHET MUHAMMAD (P.B.U.H.), who came as the light of knowledge for all the seekers. Special appreciation goes to my supervisor, Muhammad Arif, for his supervision and constant support. Not forgotten, my appreciation to my co-supervisor, Dr. Muhammad

Yaseen for his support and knowledge regarding this topic. Last but not least, my deepest gratitude goes to my beloved parents; Mr. Zafar Hussain and Mrs. Nargis Ara and also to my brother Usama Zafar for their endless love, prayers and encouragement.

REFERENCES

1. Alam, S. and Ahmad, Q.M. (2011). Exchange rate volatility and Pakistan's bilateral imports from major sources: an application of ARDL approach. *International Journal of Economics and Finance*, 3, 245-254.
2. Atif, R.M., Jadoon, A. and Zaman, K. (2010). Trade liberalisation, financial development and economic growth: Evidence from Pakistan (1980-2009). *Journal of International Academic Research*, 10, 30-37.
3. Bakhat, M. and Wurzburg, K. (2013). Co-integration of oil and commodity prices: A comprehensive approach. *Economics for energy working paper* WP fa05.
4. Banerjee, A., Dolado, J., Galbraith, J., and Hendry, D. (2003). *Co-integration, Error Correction and the Econometric Analysis of Non-stationary Data: Advanced Text in Econometrics*. New York: Oxford university press.
5. Banerjee, A., Dolado, J., Galbraith, J.W. and Hendry, D.F. (1993). *Cointegration, Error-Correction, and the Econometrics Analysis of Non-stationary*. Data Oxford University Press, Oxford.
6. Engle, R.F. and Granger, C.W.J. (1987). Cointegration and Error-Correction: Representation. *Estimation and Testing, Econometrica*, 35, 251-276.
7. Granger, C.W. (1981). Some properties of time series data and their use in econometric model specification. *Journal of econometrics*, 16, 121-130.
8. http://www.sbp.org.pk/ecodata/index2.asp
9. Johansen, S. and Juselius, K. (1990). Maximum Likelihood and Estimation and inference and Co integration with Application to the Demand for Money. Oxford, *Bulletin of Econometric and Statistics*, 52, 170-207.
10. Khan, A.H. and Ali, S.S. (1997). The demand for money in Pakistan: An application of cointegration and error-correction modeling. *Savings and Development*, 49-62.
11. Naiya, I.I. and Manap, T.A.A. (2013). Structural transformation, poverty and inequality in Nigeria: An ARDL bound testing technique. *International Journal of Economics and Finance*, 5, 141-152.
12. Nchor, D., Darkwah, S.A. and Lubos Strelec (2015). Inflation, exchange rates and interest rates in Ghana: An autoregressive distributed lag model. *International Journal of Scientific and Research Publications*, 5, 1-8.
13. Pesaran, M.H. and Pesaran, B. (1997). *Working with Microfit 4.0: Interactive econometric analysis*, Oxford University Press, United Kingdom.
14. Pesaran, M.H. and Shin, Y. (1999). An Autoregressive Distributed Lag Modeling Approach to Cointegration Analysis, In Strom, S. (Ed): *Econometrics and Economic Theory in 20th Century*: *The Ragnar Frisch Centennial Symposium*, Chapter 11, Cambridge University Press.
15. Pesaran, M.H., Shin, Y. and Smith R.J. (2001). Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics, 16, 289-326.
16. Shahbaz, M. and Islam, F. (2011). Financial development and income inequality in Pakistan: An Application of ARDL Approach. *Munich Personal RePEc Archive*.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 215-220

IMPACT OF VOCATIONAL AND TECHNICAL EDUCATION TOWARDS WOMEN EMPOWERMENT IN PAKISTAN

Farhat Jahan
Khadim Ali Shah Bukhari Institute of Technology, Karachi, Pakistan
Email: farhatresearcher@gmail.com

ABSTRACT

This study investigated and focused on the impact of vocational and technical education towards women empowerment in Pakistan. Specific objective was to explore on the status of women can be change by providing them vocational and technical education. Research shows that vocational and technical educations are closely related with the women empowerment for their economic growth and are capable to create employment for their income generation. Vocational and technical educations are designed to develop the skills, ability and knowledge necessary for employment, income support and development of women empowerment. The acquisition of these vocational and technical skills will greatly help to empower women economically. Ardiansyah et al. (2011) defined that empowerment is basically not an only enhancing confidence levels. It also improves women skills to participate in political, social and economical life on international level. Ojobo (2008) explained women empowerment is also influenced by education which improves women's status and living standard. Geethanjali and Prabhakar (2012) defined women empowerment as "Empowerment is a process of awareness and capacity building leading to greater participation, to greater decision making power and control and transformative action. Empowerment of women signifies harnessing women power by conscientising their tremendous potential and encouraging them to work towards attaining a dignified and satisfying way of life through confidence and competence as person with self-respect, rights and responsibilities the progress of any nation is inevitably linked with social and economic plight of women in that particular country." Mustafa et al. (2005) explores that the need of vocational training and education is increasing day by day which generated the results as trained individuals and employments.

Targeted population was women who live underserved areas in Karachi and different areas of Pakistan. Data was collected from the sample size of 150. For analysis of data statistical tools ANOVA, REGRESSION ANALYSIS was used. The results were highly significant. This study was limited to women of Karachi and different cities of Pakistan. It had limitations due to time and budget constraints and some of these limitations.

KEYWORDS

Women Empowerment, Employment, Economic Development, Pakistan.

Jel Classification: J16, E24, O1

1. INTRODUCTION

This study revealed that status of women was alarming in Islamic countries, particularly in Pakistan. Our total population consists of Women comprises of more than 50 percent of Pakistan's total population. Rate of poverty level is very high in different region of Pakistan.

There are several root causes which directly influence on women empowerment in Pakistan are rape, acid throwing, honor killings, forced marriages, forced prostitution, several women are unaware from their legal rights due to lack of education and knowledge. This western society is male dominating society so women are not encouraged. There is less respect of women than a man. There is no proper guideline for a woman to lead a life.

It is emphasized that there was a need of vocational and technical education for enhancement of economic and social empowerment of women and helpful for the country for their intellectual capital.

2. LITERATURE REVIEW

Women Empowered by Education

"There are two powers in the world; one is the sword and other is the pen.

There is great competition and rivalry between the two. There is third power stronger than both, that is the power of women."(Quaid e Azam speech at Islamia college for Women March 25, 1940).

Ojobo (2008) defines education is a core stone which initiates any nation towards rapid development all around the world. Education also advances women's status and empower in terms of Knowledge 'skills and self confidence. Ojobo (2008) also focused that the literacy rate of women is significantly lower than a men. Investment on women education would be able to produce literate generation in the world. Women education also reduces poverty rate in the developing countries. In short, Education is an accelerator for women Empowerment. Shekh (2000) also explains that when a women and girls are educated so it will be fruitful factor for the country government, donors and practitioners for the reduction in poverty level in the country. Khan (2010) recommended that to get education is the basic right of a woman. She can be empowered by education and easily get the paid job. Government of Pakistan, women political leaders and social activist should highlight the importance of education.

Vocational and Technical Education for Sustainable Development

There are seventeen most important sustainable development goals. Among these goals No Poverty, Quaity education and Gender equality .Vocational and technical education can play a better role to achieve these goals.

(Alhaji, 2008) explained that without education sustainable development cannot be achieved. Vocational and technical education has a great impact to transform the economic development process. Govt. should organize vocational and technical education programs for poverty alleviation.

Role of Vocational Training and Education

Mustafa et al. (2005) explores that the need of vocational training and education is increasing day by day which generated the results as trained individuals and employments. Training and education are meridian constituents which are directly linked with the economic growth and are capable to create employment produced by the technological changes and globalization.

NJOGU et al. explained that "The Nairobi Forward Looking Strategies for the advancement of Women (1985) had reiterated the need for women to be given the opportunity to reach their full potential. The meeting affirmed social and economic development should be encouraged to secure the participation of women as equal partners with men in the field of work, equal access to all position of employment, equal pay for work equal value and equal opportunities for education and vocational training."

Status of HRD and Vocational Training Programme in Pakistan

Mustafa et al (2005) concluded that HRD and vocational training are not in better condition in Pakistan as compared to its neighboring and other developing countries like Asia and South Asia. The GDP rate on education in Malaysia is about 8% while Pakistan spending less than 2% which is lower than South Asian countries. The situation of Pakistan is very different there is no harmony between jobs demanded and supply of skilled and trained workers.

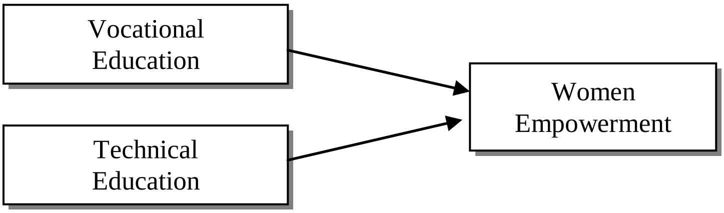
Technical Education and Vocational Training in Pakistan

Mustafa et al (2005) concluded that Technical education and Vocational training programmes have three categories in Pakistan i.e. Prevocational or school level (TVET), higher engineering and technology education. The status of Technical education and vocational training in Pakistan seems very fair that Government takes interest to promote technical education and vocational training among rural youths and unskilled women to improve their job and livelihood.

Quality of Technical Training and the Job Market in Pakistan

Mustafa et al (2005) explains that in Pakistan there are numerous technical institutions which provide quality educations to unskilled labour youth and women. These qualities of technical education produced skilled labour, generated income, raise rate of investment (ROI) and improve economic growth of Pakistan.

3. THEORETICAL FRAMEWORK



4. RESULTS AND FINDINGS:

Multiple Regression Analysis

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	**Tec Edu, Voc Edu[b]**	.	Enter

a. All requested variables entered.

b. Dependent Variable: Womenemp

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.794[a]	.630	.625	1.32447	.630	125.356	2	147	.000

a. Predictors: (Constant), Tec Edu, Voc Edu
b. Dependent Variable: Womenemp

Predictors:
(Constant), Tec Edu, Voc Edu

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	439.804	2	219.902	125.356	.000[a]
	Residual	257.871	147	1.754		
	Total	697.675	149			

a. Predictors: (Constant), Tec Edu, Voc Edu
b. Dependent Variable: Womenemp

5. EXPLANATION

- Regression result shows that F sig value is less than 0 .05 so over all model is significant.
- The value of Adjusted R square = 0.625 which shows 62.5% variation between dependent and independent variables.

6. CONCLUSION AND COMMENTS

After data analysis we are assured that our respondents believe that women can be empowered, confident, self esteemed and skilled. Women can be an important instrument in our industry.

After data analysis we came to the conclusion poverty with high literacy rate is very common in specific areas of Pakistan especially in rural areas .Women can be empower through vocational skills education programs .Vocational education programs act as weapon against discrimination issues, illiteracy, poverty, over depending on men.

Vocational education provides a better chance for women and girls climbing on a social ladder of developing faster. It also qualifies women and girls to fulfill certain economic, political and cultural functions and t the same time improves their socio economic status.

After data analysis we are also assured by our respondents that technical education provide a world class skills with knowledge, income generation opportunities and financial stability for women and girls.

As we all know about the 17 sustainable development goals i.e. No Poverty, Quality education and Gender Equality. Vocational and technical education can play a better role to achieve these goals.

7. ACKNOWLEDGMENT

I am thankful to Almighty Allah who has given me this great opportunity to write about this subject “Impact of vocational and technical education on women empowerment in Pakistan.

At the time of preparing this research paper I have gone through different research papers. I am actually focusing the topic women empowerment because I want to achieve the goals to empower women. I am thankful to female respondents from many vocational and technical institutes whose response supported me to accomplish this task.

REFERENCES

1. Ardieansyah, Abd Hair, A. Rozman, M.Y. and Buang, A. (2011). The imperative of training for women empowerment statistical evidence from Indonesia. *World Applied Sciences Journal* (Special Issue of Human Dimensions of Development), 13, 39-45.
2. Geetanjali, R. and Prabhakar, K. (2012). NGOs role in empowering women through SHGs in Chittoor District Andhra Pradesh. *International Journal of Social Science*, 1(6), 299-305.

3. Ibrahim, H. (2008). Revitalizing Technical and Vocational Training for Poverty Eradication and Sustainable Development through Agricultural Education. *African Research Review*, 2(1), 152-161.
4. Mustafa, U., Abbas, k. and Saeed, A. (2005). Enhancing Vocational Training for Economic Growth in Pakistan. The Pakistan Development Review 44(4), 567-584.
5. Khan, T.M., Mann, A.A., Zafar, M.I., Hashmi, M. and Akhtar, S. (2010). Determinants of Women Empowerment: A Case Study from District Rawalpindi, Punjab, Pakistan. *Pakistan Journal of Science*, 62(1).
6. Njogu, K. and Orchardson-Mazrui, E. (2008). Culture, Gender Inequality and Women rights in the Great Lakes, Chapter One, pps. 1-39, in *Culture, Performance and Identity: Paths of Communication in Kenya*, Twaweza Communications, Nairobi, Kenya: ISBN: 9966-7244-1-9.
7. Njogu, K., Orchardson-Mazrui, E. (2005). Can Culture Contribute to the Empowerment of Women? (UNESCO Research Paper: Available on UNESCO WEBPAGE).
8. Ojobo, J.A. (2008). Education: A catalyst for Women Empowerment in Nigeria. *Ethiop. J. Educ. & Sc*., 4(1), 93-108.
9. Shekh, M.N.I. (2000). Educating Women through NGO Programmes: Contradictions of Culture and Gender in Rural Bangladesh.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 221-228

SIZE DISTORTION FOR LINEAR REGRESSION MODEL IN THE PRESENCE OF HETEROSCEDASTIC AND NON-NORMAL ERROR TERMS

Haris Khurram[1] and **Muhammad Aslam**[2]
Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan.
Email: [1]hariskhurram2@gmail.com; [2]aslamasadi@bzu.edu.pk

ABSTRACT

The assumptions of error terms in linear regression models are much considerable before adopting any estimation procedure. The error terms are usually exposed to be heteroscedastic in many practical situations and this problem results in inefficient estimation and incorrect inference. One of the major consequences of heteroscedasticity is the size distortion as the usual tests tend to over-reject the true null hypotheses. The available literature suggests many methods to cope with this situation. In addition to the issue of heteroscedasticity, it is also possible for the error terms that they may not be normally distributed. However, the available literature does not address the heteroscedasticity and non-normality and the same time. The present work focuses on the same. The main aim of the work is to assess size distortion of the usual tests when the errors terms are both heteroscedastic and non-normally distributed. With the help of Monte Carlo simulations, different level of heteroscedasticity has been generated and different distributions of error have been studied, namely, t, uniform, exponential and normal distribution. Normal distribution of error has been taken for comparing the results obtained from the stated non-normal distributions. After application of the ordinary least squares, a comparison of null rejection rate is made for the usual t and F-tests in order to assess the size distortion.

KEYWORDS

Heteroscedasticity; Non-normal errors; Null rejection rate; Power curves; Size distortion.

1. INTRODUCTION

For the efficient estimation of parameters of a linear regression model, we have to meet with some of underlying assumptions of the classical linear regression model (CLRM). When all the assumptions of the CLRM are fulfilled then the best way is to use the ordinary least squares (OLS) estimator (OLSE) because it is the best linear unbiased estimator (BLUE). One of the important assumptions of the CLRM is that the error term should be homoscedastic. The consequences of violation of this assumption are that the OLSE does not remain efficient and its covariance matrix estimator becomes biased and inconsistent. A rich literature is available upon this issue and this literature suggests some robust estimators in order to draw correct inference in the presence of heteroscedasticity; see, for more details, White (1980), MacKinnon and White (1985), Long and Ervin

(2000), Cribari-Neto et al. (2007), Hayes and Cai (2007), Aslam et al. (2013) and Aslam (2014) among many others.

Another important assumption is that error term should be normal. The advantage of this assumption is that the OLS and maximum likelihood (ML) estimates provide the same results (Zechauser and Thompson, 1970; Curran and West, 1996; Goldfeld and Quandt, 1981; Shalabh et al. 2006).

If the error terms is non-normal, then asymptotic distribution of the ML estimator normal. Whatever the distribution of error term is, the central limit theorem (CLT) also ensures for large sample size, residual behave like a normal sample (Gnanadesikan 1997; Weisberg, 2005).

However, for small samples, non-normality affects the size of the test statistics. Literature provides detailed discussion about the problem of non-normality of error terms and also provides many robust methods to sort out the problem of heteroscedasticity. But, the distortion in size of test was not addressed when the error term is non-normal as well as heteroscedastic. The aim of present paper is to assess the size distortion of test statistic while testing the coefficients of linear regression model in the presence of heteroscedastic and non-normal errors.

2. LINEAR REGRESSION MODEL WITH HETEROSCEDASTIC AND NON-NORMAL ERRORS

Consider the following regression model with heteroscedastic and non-normal error:

$$y_i = \beta_0 + \beta_1 x_i + u_i \; ; \; i=1,2,\dots,n. \tag{1}$$

where $u_i = \epsilon_i \sigma_i$ and Var (u_i) = σ_i^2 which shows that error term is heteroscedastic (see Aslam, 2014 for such model) and $\epsilon_i$is unobservable error whose distribution is not necessarily to be normal.

3. SIZE OF TEST

Let $\tau = h(\beta)$ be the function of parameter vectorβ, $\hat{\tau}$ its estimate and $s(\hat{\tau})$ its asymptotic standard error.

Suppose, we are interested in testing

H_0: $\tau = \tau_o$ against H_1: $\tau = \tau_1$,

where τ_o is some specified value. Consider the Studentized t-statistic,

$$t_n = \frac{\hat{\tau} - \tau}{s(\hat{\tau})}$$

It is easy to show that

$$t_n \xrightarrow{d} N(0,1).$$

Let $Z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution. That is, if $Z \sim N(0,1)$, then $P(Z > Z_{\alpha/2}) = \alpha/2$ and $P(|Z| > Z_{\alpha/2}) = \alpha$. A test of asymptotic significance α rejects null hypothesis if $|t_n| > Z_{\alpha/2}$.

Hence α is also called size of test.

4. THE MONTE CARLO EXPERIMENT

In our work we used the same scheme as used by Cribari-Neto (2007), Aslam et al. (2013) and many others in their work. Under this scheme the model of interest is the same as shown in Eq. (1). Let β_0=β_1=1 in the said model. For simulation, the explanatory variable x_i is generated from uniform distribution over interval (0,1) and then kept fixed and ϵ_i is drawn randomly under four different data generating process. In first processϵ_i is assumed to be distributed as standard normal. This process is done to compare the results of non-normal with the results of normal error terms. In the second process, ϵ_i is assumed to be distributed as standard *t*. In the third process ϵ_i is assumed to be distributed as standard exponential while in the fourth process ϵ_i is uniformly distributed.

The variance of error term is generated as:

$$\sigma_i^2 = \exp\{\gamma x_i + \gamma x_i^2\},$$

where γ is a constant to vary level of heteroscedasticity. For different value of γ; γ=0, 0.8, 1.72 and 2.33, there are different degrees of heteroscedasticity λ; λ=1, 5, 30 and 100, respectively. The degree of heteroscedasticity is measured by $\lambda = \frac{\max(\sigma_i^2)}{\min(\sigma_i^2)}$. When γ=0, it means that errors are homoscedastic that is, under homoscedasticity, $\lambda = 1$. Otherwise, $\lambda > 1$. The sample size *n* is set to be 25, 50, 100 and 200, and the number of simulations is 5000 for each sample size (see also Cribari-Neto, 2004, Aslam et al., 2013 and Aslam, 2014 for more details). For computational purpose we use R language.

For the present study, we discuss the following results:

1. Average coverage and average length of confidence intervals.
2. Null rejection rate (NRR) for t-tests.
3. NRR for F-tests.

5. EMPIRICAL RESULTS

Table 1 and Fig. 1 and 2 show the coverage percentage and average length of confidence intervals of the slope (β_1) of while considering the model, given in Eq. (1). The measures are computed under different level of heteroscedasticity (λ), different error distributions and varying sample size *n*, as discussed earlier.

When the error terms are homoscedastic (i.e. λ = 1), the coverage percentage is closer to the nominal coverage of 95% for all the error distributions, normal, t, exponential and uniform. Expectedly, this coverage improves with increase in sample size. As the degree of heteroscedasticity increases, the empirical coverage tends to decrease, namely for the uniform distribution. For instance, if we consider severe heteroscedasticity (when λ =100 for *n* = 25), the empirical coverage percentages are 64.56, 66.34, 61.16 and 62.44, respectively when the distribution of error terms is normal, t, exponential and uniform.

The results show considerable coverage distortion. However, the results improve slightly with the increase sample size. Generally, the results favour normal and t distributions while the other the tests based on other two error distributions perform badly. Specially, if the error terms are heteroscedastic and uniformly distributed then one can expect poor coverage of the interval estimates. Same pattern of results can be found while considering the average lengths. Average length tends to become larger with the departure form normality, increase in the degree of heteroscedasticity and decrease in sample size. The above stated results can also be depicted from Fig. 1 and 2.

Table 1
Empirical Coverage (%) and Average Length of for Interval Estimates of β_1

λ	Distribution	*n*=25		*n*=50		*n*=100		*n*=200	
		Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
1	Normal	93.54	3.07	94.66	1.75	94.50	1.29	94.84	0.93
	t	92.88	3.04	94.88	1.72	94.70	1.29	95.34	0.93
	Exponential	93.28	2.97	94.30	1.72	94.50	1.28	94.98	0.93
	Uniform	92.74	8.95	94.36	5.06	95.06	3.73	95.14	2.69
5	Normal	85.18	4.18	92.70	2.65	91.78	1.90	92.10	1.38
	t	86.20	4.11	92.82	2.60	91.80	1.91	92.84	1.38
	Exponential	84.96	4.01	91.78	2.60	90.92	1.901	92.76	1.38
	Uniform	83.76	12.14	91.50	7.66	92.28	5.53	92.48	4.00
30	Normal	70.88	6.97	88.80	4.99	86.92	3.49	87.92	2.54
	t	73.92	6.78	88.70	4.87	87.38	3.50	87.92	2.53
	Exponential	69.40	6.57	86.44	4.85	85.76	3.46	88.72	2.53
	Uniform	68.98	20.35	87.18	14.46	87.00	10.18	87.82	7.35
100	Normal	64.56	10.78	86.66	8.11	83.56	5.61	85.22	4.07
	t	66.34	10.39	86.56	7.89	85.16	5.61	85.52	4.05
	Exponential	61.16	10.00	84.08	7.84	82.80	5.53	85.94	4.04
	Uniform	62.44	31.66	85.14	23.51	84.24	16.35	85.34	11.78

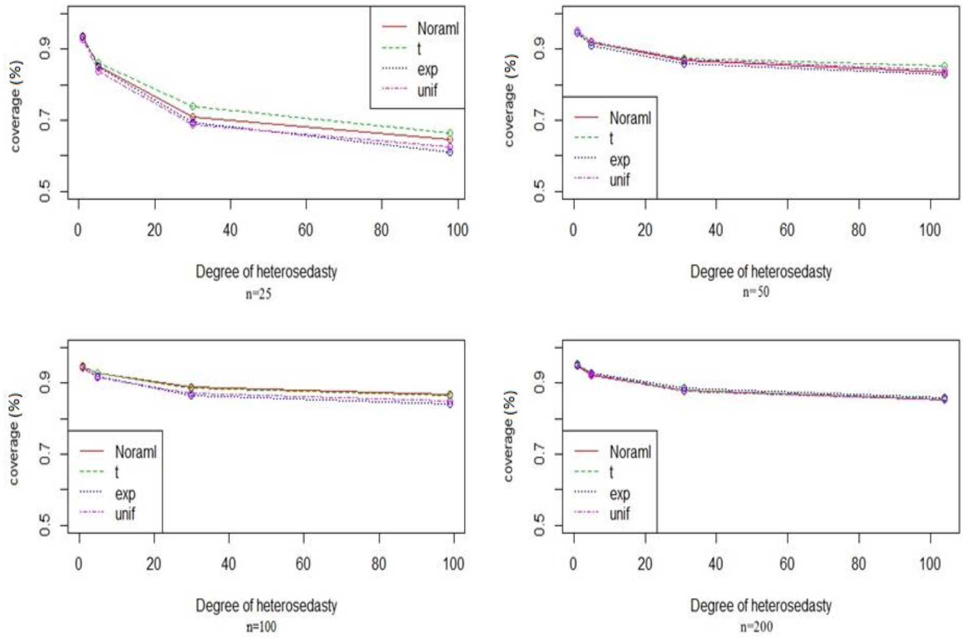


Fig. 1: Coverage of Confidence Interval for β_1

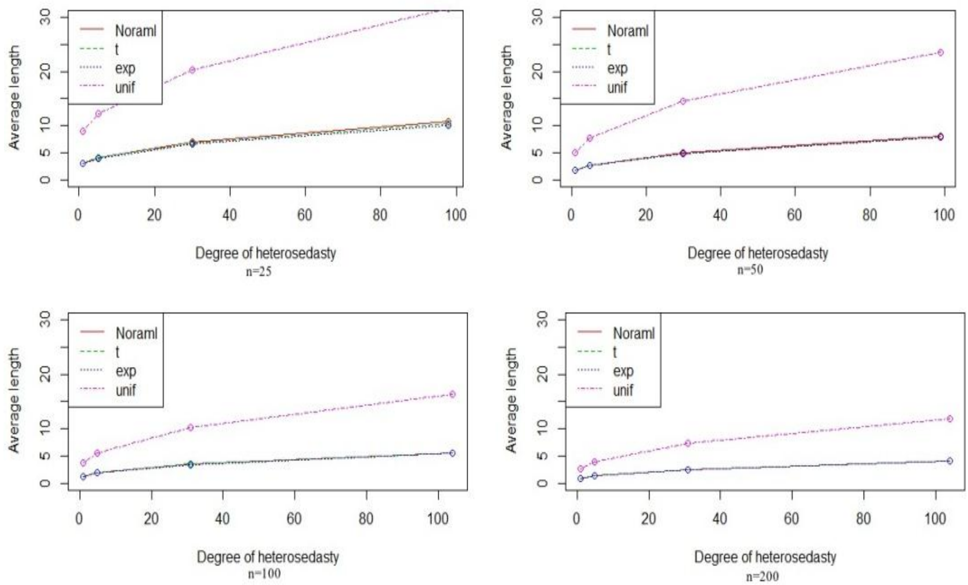


Fig. 2: Average Length of Confidence Interval for β_1

Table 2 shows the estimated NRR of t-test statistics for both β_0 and β_1 at 5% level of significance. For discussion, we focus on β_1 only. In case of homoscedastic errors, it is found that the estimated NRR for the t-test is approximately 5%, no matter what is the distribution of the error term.

Table 2
Estimated NRR of t-tests at 5% Level of Significance

λ	Distribution	n=25	n=50	n=100	n=200	n=25	n=50	n=100	n=200
		β_0				β_1			
1	Normal	5.68	4.96	4.96	5.30	5.72	4.74	4.66	5.32
	t	4.54	5.20	5.00	5.10	4.50	5.20	5.02	5.12
	Exponential	6.68	5.20	5.44	5.40	4.74	4.66	4.70	5.26
	Uniform	5.38	5.06	5.34	5.36	5.44	5.22	4.72	5.12
5	Normal	5.02	1.86	1.84	2.22	11.00	6.98	7.68	7.98
	t	3.72	1.74	2.14	2.20	11.00	6.98	7.68	7.98
	Exponential	4.62	1.76	2.00	2.52	12.26	7.58	7.90	7.52
	Uniform	4.86	1.88	2.04	2.10	12.90	7.56	7.36	7.48
30	Normal	5.96	0.52	0.74	0.94	25.90	10.62	12.22	12.10
	t	4.18	0.56	0.72	0.94	22.28	10.32	11.80	12.10
	Exponential	5.38	1.22	1.24	1.08	26.96	12.60	13.08	11.84
	Uniform	6.62	0.76	0.72	1.00	26.32	11.48	12.08	11.50
100	Normal	7.18	0.36	0.58	0.72	33.14	12.30	15.14	15.08
	t	5.34	0.32	0.60	0.76	36.22	14.52	17.02	15.34
	Exponential	8.86	1.42	1.18	1.16	34.90	15.26	16.04	15.54
	Uniform	9.14	0.58	0.70	0.92	33.14	13.12	14.98	13.90

When λ increases, t-tests distorts the NRR. These tests become considerably liberal and tend to over-reject the true null hypothesis. This size distortion becomes worse with the increase in degree of heteroscedasticity but gets improved for larger samples. Interestingly, it is found that the error distribution does not play some considerable role in this size distortion as all the tests perform identically when all the stated error distributions are considered. This size distortion varies from six times to three times when the sample size increases from 25 to 200, respectively.

Table 3 shows the estimated NRR of F-tests when we are interested to test β_0=β_1=1. A great size distortion is noted with small samples and severe degree of heteroscedasticity. Performance of the F-tests are almost identical and they are less liberal when the error are assumed to have normal and t distributions. Among the other two error distributions, the size distortion for uniform distribution is more serious as compared to that for the exponential distribution. The results get improve with increase in sample size. Generally, the same pattern of size distortion is observed as that is noted for individual coefficient testing.

Table 3
Estimated NRR of F-tests at 5% Level of Significance

λ	Distribution	n=25	n=50	n=100	n=200
1	Normal	6.78	5.62	5.72	4.90
	t	6.58	5.28	5.40	4.96
	Exponential	9.54	6.94	6.30	5.26
	Uniform	6.46	6.34	5.42	4.84
5	Normal	13.00	7.84	8.66	7.36
	t	11.58	7.52	7.96	7.32
	Exponential	15.42	9.18	9.44	7.28
	Uniform	13.84	9.06	8.20	7.36
30	Normal	24.30	11.66	12.92	11.78
	t	21.88	11.28	11.84	11.58
	Exponential	26.72	14.12	14.24	12.02
	Uniform	26.42	13.50	12.80	11.78
100	Normal	30.36	13.62	15.60	13.82
	t	28.28	13.22	14.06	13.84
	Exponential	33.58	16.30	16.78	14.44
	Uniform	32.42	14.96	14.94	14.14

6. COMMENTS AND CONCLUSION

The results conclude that when error terms are non-normal and as well as heteroscedastic the major reason for size distortion of usual t- and F-tests is the degree of heteroscedasticity only. Non-normality does not cause considerable size distortion. So if the error term is non-normal and heteroscedastic the major concern should lie in handling of heteroscedasticity. This can be accomplished by adopting some robust methods as suggested by Aslam et al. (2013) and many other authors.

REFERENCES

1. Aslam, M. (2014). Using Heteroscedasticity-Consistent Standard Errors for the Linear Regression Model with Correlated Regressors. *Communications in Statistics-Simulation and Computation*, 43(10), 2353-2373.
2. Aslam, M., Riaz, T. and Altaf, S. (2013). Efficient estimation and robust inference of linear regression models in the presence of heteroscedastic errors and high leverage points. *Communications in Statistics-Simulation and Computation*, 42(10), 2223-2238.
3. Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis*, 45, 215-233.
4. Cribari-Neto, F., Souza, T.C. and Vasconcellos, K.L.P. (2007). Inference under heteroskedasticity and leveraged data. *Communications in Statistics, Theory and Methods*, 36, 1877-1888. Errata: (2008) 37, 3329-3330.

5. Curran, P.J., West, S.G. and Finch, J.F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological methods*, 1(1), 16.
6. Gnanadesikan, R. (1997). *Methods for Statistical Analysis of Multivariate Data*, Second Edition. New York: Wiley.
7. Goldfeld, S.M. and Quandt, R.E. (1981). Econometric modelling with non-normal disturbances. *Journal of Econometrics*, 17(2), 141-155.
8. Hayes, A.F. and Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression*: An introduction and software implementation. Behavior Research Methods*, 39(4), 709-722.
9. Long, J.S. and Ervin, L.H. (2000). Using heteroskedasticity consistent standard errors in the linear regression model. *The American Statistician* 54, 217-224.
10. MacKinnon, J.G. and White, H. (1985). Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29, 305-25.
11. Weisberg, S. (2005). *Applied linear regression* (Vol. 528). John Wiley & Sons.
12. White, H. (1980). A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity. *Econometrica*, 48, 817-838.
13. Zeckhauser, R. and Thompson, M. (1970). Linear regression with non-normal error terms. *The Review of Economics and Statistics*, 52(3), 280-286.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 229-238

ASSOCIATION OF MUSCULOSKELETAL PAIN WITH HEAVY BAG PACKS AMONG SCHOOL CHILDREN

Syed Mustansir Hussain Zaidi, Fahim Ahmed Ansari, Hira Fatima Waseem, Sania Fahim and **Muhammad Irfan**
Department of Statistics, Liaquat National Hospital, Karachi, Pakistan
Email: mustansir.zaidi@lnh.edu.pk
fahim.ahmed@lnh.edu.pk
hira.waseem@live.com
sania.amin@lnh.edu.pk
irfan.zafar@lnh.edu.pk

ABSTRACT

School bag carried by school children as daily load becomes a health problem such as musculoskeletal pain. Most common example of musculoskeletal pain is low back pain. Back pain is seen to occur more in healthy children. Back pain in children is much more likely to have a serious underlying disorder. Now a day's student is often seen with heavy school bags and it is observed that it is influencing their health in this transitional period of life to adulthood over many years. Many students carry school backpacks that exceed 10 percent to 15 percent of their body weight, which puts them at risk for back pain and related disorders. Carrying backpacks increases the risk of back pain and possibly the risk of back pathology.

The aim of the study is to investigate the association of musculoskeletal pain with heavy bag packs among school children. A cross-sectional study will be conducted among school children having age of 9-16 years. Data will be collected using a questionnaire regarding gender, side of back pain, time to arrive to school, type of bag, mode of transport, posture of carrying bag, pain intensity level and measure weight of student with and without bag and their height. Data will be analyzed using statistical package for social sciences (SPSS). Mean and standard deviation will be computed for quantitative variable and frequency and percentage will be calculated for qualitative variables. Chi-square and logistic regression will be applied to see the association of musculoskeletal pain with independent variables. Preventive measures and appropriate guidelines with regard to safe load carriage in school children are therefore needed to protect this age group.

KEYWORDS

Back pain, bag pack, load carriage, musculoskeletal.

1. INTRODUCTION

Musculoskeletal pain is a known outcome of repetitive strain, overuse, and work-related musculoskeletal disorders.[1] These injuries can cause a pain in bones, joints, muscles, or surrounding structures.[2,3] Musculoskeletal disorder should be considered as

3 distinct entities that are the neck, upper back and lower back pain being more prominent in younger children. Hence, carrying heavy school bags by school children could cause a wide spectrum of pain related to musculoskeletal disorders and postural dysfunctions.[4] Most common example of musculoskeletal pain is low back pain.[2]

Musculoskeletal symptoms in school children are multi-factorial in origin; but the carriage of heavy school bag is the suspected factor.[5-10] Recent studies have shown that prevalence of musculoskeletal problems is increasing in school children and adolescents.[11-14] Research has shown that the percentage of children who suffer from back pain varies between 11% and 52.1%.[15] Consequently, it is worthwhile to examine the factors which contribute to pain among children. One factor which might contribute to pain in children is overloaded schoolbags. There is evidence that heavy backpacks carried by students can cause altered gait and bad posture.[16-18] The American Occupational Therapy Association, the American Academy of Orthopaedic Surgeons and the International Chiropractic Pediatric Association suggest that the load should not be more than 10%.[19] Limon et al. indicated that 30% to 54% of students carried backpack greater than 15% of their body weight.[20] Studies reported that the lifetime incidence of low back pain (LBP) in children was 30% to 50%.[11-14,21] Girls are at higher risk of back problems due to their short stature and thin body mass.[22]

Several studies have recommended weight of the bag packs should be reduced because if they exceed proportionally from the limits set for children than it can cause discomfort, and tiredness during backpack carrying and the duration of time spent bearing the backpack on the shoulders are parameters associated with back pain.[23] When the school children walk with heavy bag packs on their back and shoulder it is possible that they may change their posture, they lean forward and raise their heads.[24] According to study, the school bag weight and the duration for carrying them have an adverse effect on cervical and shoulder region that can lead to a forward head posture.[25] The aim of the study is to investigate the association of musculoskeletal pain with heavy bag packs among school children.

This paper is divided into five sections. Section 1 starts with introduction of musculoskeletal pain and factors affecting on it. In section 2, methodology of data collection is discussed. We discuss the statistical results in section 3 and then final conclusion to overall analysis is discussed in section 4.

2. MATERIAL AND METHOD

A cross-sectional study was carried out among school children whose age was 9 to 16 years of either gender. Data was taken from five different schools of Karachi from September 2015 to December 2015. Ethical approval was obtained from the principal of Schools. Inclusion criteria were healthy children without any acute or chronic condition, no history of fractures of upper and lower extremities within last one year. The exclusion were children with sensory or motor neurological manifestations, postural deformity as scoliosis or hypnosis, any cause of balance disturbance as vertigo, musculoskeletal disorders as leg length discrepancy, any cause leading to inability of the children to stand upright on both feet.

By taking prevalence of musculoskeletal pain=57.3% on the basis of literature, margin of error(d)=5%, the calculated sample size is 376 with the help of WHO software for sample size calculation taking 95% confidence interval.

Questionnaires were used and care was taken to simplify the questions as much as possible. Questions regardent demography, body weight, schoolbag weight, type of schoolbag, posture while carrying bag, feeling about their bag weight and the outcome variable was musculoskeletal pain was taken by principal investigator. However, musculoskeletal pain was defined as pain in the low back region and measured on visual analogue scale (VAS) score. The children weights were measured on weight machine and height was measured by using measurement scale. The weight was first measured when carrying the school bag and then without the school bag and the difference between the two weights were recorded as the weight of the school bag.

Data was analyzed by using statistical package for social sciences (SPSS) version 21. Mean and standard deviation were computed for quantitative variable and frequency and percentage were calculated for qualitative variables. Chi-square test was applied to see the association of musculoskeletal pain with categorical variables considering p-value < 0.05 as significant.

3. RESULTS

A total of 376 children were taken in which 26 children were excluded due to incomplete responses of respondents. Analysis was performed on 350 children in which 154(44%) was boys and 196(56%) was girls. Overall mean age of children was 12.49±2.40 years, mean BMI of children was 18.35±1.74 kg/m^2 and mean weight of bag was 3.98±1.23 kg as shown in Table 1. The musculoskeletal pain was assessed in 269(76.9%) students. BMI was further categorize which shows 168(48%) was under weight, 182(52%) having normal weight while no obese children was found in our data. Mode of transport was assessed which showed that school van was used by 144(41.1%) students, 101(28.9%) using bikes for reaching school, 90(25.7%) was coming school by walking while only 15(4.3%) of students using private carriage for transportation. Most of the students having pain in lower back 118(33.7%) as shown in table 2. We also found that 46.3% of students bags were 10-15% of their weight and most of the students feel that they have heavy bag packs (56.3%), type of bag was shown that 254(72.6%) students using double strip bag while 96(27.4%) using single strip bag. 194 (55.4%) of students takes less than 5 minutes to reach class from school gate. while most of students 171(48.9%) always keep bag away from body as shown in table 2. Association of musculoskeletal pain with gender, mode of transport, bag weight, body weight ratio and characteristics regarding usage of school bag found significant at $p < 0.05$ as shown in table 3.

Table 1
Descriptive Statistics of Demographic Data

Quantitative Variables	Mean± SD
Age(Years)	12.49±2.40
Body Weight(kg)	42.76±10.22
Height(cm)	151.25±12.42
BMI(kg/m^2)	18.35±1.74
Bag Weight(Kg)	3.98±1.23
Bag Weight/Body Weight	11.84±4.70

Table 2
Frequency Distribution of Demography and Characteristics of Musculoskeletal Pain

Variables		N (%)
Gender	Male	154(44%)
	Female	196(56%)
BMI Categories	Under Weight	168(48%)
	Normal Weight	182(52%)
Mode of Transport	School Bus	144(41.1%)
	Bike	101(28.9%)
	Walking	90(25.7%)
	Private Carriage	15(4.3%)
Side of Back Pain	Lower back	118(33.7%)
	Upper Back	49(14%)
	Shoulder	94(26.9%)
	Neck	66(18.9%)
	Lower back And Shoulder	23(6.6%)
Bag Weight to Body Weight Ratio	<10%	109(31.1%)
	10-15%	162(46.3%)
	>15%	79(22.6%)
Feeling about bag Weight	Light Weight	20(5.7%)
	Average Weight	133(38%)
	Heavy Weight	197(56.3%)
Type of Bag	Single Strip	96(27.4%)
	Double Strip	254(72.6%)
Time to Carry Bag	<5 Min	194(55.4%)
	5-10 Min	114(32.6%)
	>10 Min	42(12%)
Posture While Carrying Bag	Always Bag Away From Body	171(48.9%)
	Some Time Bag Away From Body	131(37.4%)
	Always Bent over From Body	48(13.7%)

Table 3
Association of Musculoskeletal pain with Gender, Mode of transport and bag Weight/Body Weight Ratio and Characteristics of Carrying School Bag

Qualitative Variables		Musculoskeletal pain Yes N(%)	Musculoskeletal pain No N(%)	P-value
Gender	Male	73(47.4)	81(52.6)	0.000
	Female	196(100)	0(0)	
Mode of Transport	School Bus	138(95.8)	6(4.2)	0.025
	Walking	98(97)	3(3.0)	
	Bike	31(34.4)	59(65.6)	
	Private carriage(Rikshaw, Car etc.)	2(13.3)	13(86.7)	
Bag weight/ Body weight	<10%	85(78)	24(22)	0.001
	10%-15%	135(83.3)	27(16.7)	
	>15%	49(62)	30(38)	
Feeling About school bag	Light Weight	2(10)	18(90)	0.023
	Normal Weight	76(57.1)	57(42.9)	
	Heavy Weight	191(97)	6(3)	
Type of Bag	Single Strip	90(93.8)	6(6.3)	0.002
	Double Strip	179(70.5)	75(29.5)	
Duration of walking with school bag	<5 Min	188(96.9)	6(3.1)	0.000
	5-10 Min	79(69.3)	35(30.7)	
	>10 Min	2(4.8)	40(95.2)	
Posture of Carrying Bag	Always Bag Away From Body	165(96.5)	6(3.5)	0.001
	Some Time Bag Away From Body	102(77.9)	29(22.1)	
	Always Bent over From Body	2(4.2)	46(95.8)	

Chi-Square applied
P-value <0.05 considered as Significant
* Not Significant at 0.05 level.

4. DISCUSSION AND CONCLUSION

Most of the studies reported that more than half of the students have pain while carrying school bags.[27-29] Studies evaluated pain that occurs while carrying schoolbags, as opposed to back pain experienced during daily life activities. It is clear that schoolbag weight is not an independent contributor to back pain development in schoolchildren. In contrast to schoolbag weight, the time spent carrying the schoolbag and fatigue were significant contributors to pain, which coincides with previous studies.[30,31]

Studies clearly suggest that schoolbag weight itself does not contribute to pain. Instead, it appears that schoolbag weight has an impact when the bag is carried for a long duration or when the student feels fatigued. The duration of carrying the schoolbag causes the trunk to lean forward and influences cervical and shoulder posture.[26,32,33]

Studies demonstrate that girls are more likely to report back pain when they carry their school bags.[34-37] Schoolbag carriage causes the trunk to lean forward lean when walking to school.[38] However, this effect is not instantaneous but it is repeated each time the child walks to school or returns home.

In a study, a total of 580 students who complained about pain in the neck, shoulder or back region was 19 males and 7 females.[39] Several studies shows that females carry heavier school bags compared to males.[39-41] Studies show that the recommended school bag weight is 10 – 15% to the body weight.[41,42] The mean schoolbag weight as a percentage of body weight was 8.5% however about 30.8% of the children had bags which were more than the recommended limit of 10%.[43] most of the studies report rates between 15%- 50%.[44]

Low back pain children also experienced pain in the upper body involving the neck, shoulders and upper back. Pain in these areas is associated with carrying heavy loads Carrying a heavy school bag for long periods of time could result in repetitive stress injuries to the growing body. This follows the shifting of the child’s centre of gravity in the direction of the load when carrying a backpack. Bag pack loads exceeding 10% of body weight have been shown to increase energy consumption, increase trunk forward lean.[45]

The results of study indicate that heavy bag pack has significant risk factor for producing musculoskeletal pain. Preventive measures and appropriate guidelines with regard to safe load carriage in schoolchildren are therefore needed to protect this age group.

5. RECOMMENDATIONS

To further reduce on the effects of bag weight schools ought to provide students with lockers for storage of their scholastic materials. Schools should also have fully functional libraries where students can sit, read and borrow text books instead of carrying them daily in their bag. Lunch packs and water bottles as additional contributors to the school bag weight. Parents should be urged to select school bags and items which are made of light-weight materials. Remind and help their children to pack school bags every day according to the timetable. Regularly ask your child if their backpack is causing fatigue or pain. To prevent low back pain and associated muscle fatigue, muscle strain, and other serious back injuries, many experts recommend limiting the school bag load to 10% to 15% of body weight.

REFERENCES

1. Andrews, D.F. and Hampel, F.R. (2015). *Robust estimates of location: survey and advances*. Princeton University Press.
2. Beaton, A.E. and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics,* 16(2), 147-185.
3. Birch, J.B. and Myers, R.H. (1982). Robust analysis of covariance. *Biometrics,* 699-713.
4. David, H.A. and Nagaraja, H.N. (1970). Order statistics. *John Wiley & Sons*, Inc.
5. Iqbal, Z. and Ahmad, M. (2015). Some Properties of Generalized Log Pearson Type VII Distribution. *Pak. J. Statist.*, 31(2), 171-186.
6. Azuan, M., Zailina, H., Shamsul, B.M.T., Asyiqin, M.A., Azhar, M.N. and Aizat, I.S. (2010). Neck, upper back and lower back pain and associated risk factors among primary school children. *Journal of Applied Sciences,* 10(5), 431-435.
7. Craig, K.D. (2010). Global year against musculoskeletal pain. Pain research & management: *Journal of the Canadian Pain Society*, 15(5), 284.
8. Ramprasad, M., Alias, J. and Raghuveer, A.K. (2010). Effect of backpack weight on postural angles in preadolescent children. *Indian Pediatrics,* 47(7), 575-580.
9. Brackley, H.M., Stevenson, J.M. and Selinger, J.C. (2009). Effect of backpack load placement on posture and spinal curvature in prepubescent children. *Work,* 32(3), 351-60.
10. Mayank, M., Singh, U. and Quddus, N. (2007). Effect of backpack loading on cervical and shoulder posture in Indian school children. Indian *Journal of Physiotherapy and Occupational Therapy-An International Journal,* 1(2), 3-12.
11. Mackenzie, W.G., Sampath, J.S., Kruse, R.W. and Sheir-Neiss, G.J. (2003). Backpacks in children. *Clinical Orthopaedics and Related Research,* 409, 78-84.
12. Chansirinukor, W., Wilson, D., Grimmer, K. and Dansie, B. (2001). Effects of backpacks on students: measurement of cervical and shoulder posture. *Australian Journal of Physiotherapy,* 47(2), 110-116.
13. Hong, Y. and Cheung, C.K. (2003). Gait and posture responses to backpack load during level walking in children. *Gait & Posture,* 17(1), 28-33.
14. Pascoe, D.D., Pascoe, D.E., Wang, Y.T., Shim, D.M. and Kim, C.K. (1997). Influence of carrying book bags on gait cycle and posture of youths. *Ergonomics,* 40(6), 631-640.
15. Grimmer, K.A., Williams, M.T. and Gill, T.K. (1999). The associations between adolescent head-on-neck posture, backpack weight, and anthropometric features. *Spine,* 24(21), 2262.
16. Van Gent, C., Dols, J.J., Carolien, M., Sing, R.A.H. and de Vet, H.C. (2003). The weight of schoolbags and the occurrence of neck, shoulder, and back pain in young adolescents. *Spine,* 28(9), 916-921.

17. Panicker, R.K. and Sandesh, T.S. (2014). Prevalence of Musculoskeletal Pain in School Going Adolescents Using School Bags: A Co-relational Research. *International Journal of Therapies and Rehabilitation Research,* 3(4), 1.
18. Sheir-Neiss, G.I., Kruse, R.W., Rahman, T., Jacobson, L.P. and Pelli, J.A. (2003). The association of backpack use and back pain in adolescents. *Spine,* 28(9), 922-930.
19. Koley, S. and Kaur, N. (2010). An association of backpack weight and postural habits in school going children of Amritsar, Punjab, India. *The Anthropologist,* 12, 107-111.
20. Cavallo, C.M., Hlavaty, T.M. and Tamase, M.G.M. (2003). A pilot study for the development of a primary prevention program: What is the average weight of a fourth grader's backpack? *Work*, 20(2), 137-158.
21. Balague, F., Troussier, B. and Salminen, J.J. (1999). Non-specific low back pain in children and adolescents: risk factors. *European Spine Journal,* 8(6), 429-438.
22. Balamurugan, J. (2014). School bags and musculoskeletal pain among elementary school children in Chennai city. *International Journal of Medical Science and Clinical Invention*, 1, 302-309.
23. Wiersema, B.M., Wall, E.J. and Foad, S.L. (2003). Acute backpack injuries in children. *Pediatrics,* 111(1), 163-166.
24. Lockhart, R.A., Jacobs, K. and Orsmond, G. (2004). Middle school children's participation in activities and the effects of pain from backpack use on participation. *Work*, 22(3), 155-168.
25. Barkhordari, A., Ehrampoush, M. H., Barkhordari, M., Derakhshi, F. and Mirzaii, M. (2013). Assessment of School Backpack Weight and Other Characteristics in Elementary Schools, Yazd-Iran. *The Journals of Community Health Research*, 2(1), 2-7.
26. Kjaer, P., Wedderkopp, N., Korsholm, L. and Leboeuf-Yde, C. (2011). Prevalence and tracking of back pain from childhood to adolescence. *Bio Medical Central Musculoskeletal Disorders,* 12(1), 1.
27. Mwaka, E.S., Munabi, I.G., Buwembo, W., Kukkiriza, J. and Ochieng, J. (2014). Musculoskeletal pain and school bag use: a cross-sectional study among Ugandan pupils. *Bio Medical Central Research Notes,* 7(1), 222.
28. Rodríguez-Oviedo, P., Ruano-Ravina, A., Pérez-Ríos, M., García, F. B., Gómez-Fernández, D., Fernández-Alonso, A. and Turiso, J. (2012). School children's backpacks, back pain and back pathologies. *Archives of Disease in Childhood, Archdischild*-2011.
29. Papadopoulou, D., Malliou, P., Kofotolis, N., Emmanouilidou, M.I. and Kellis, E. (2013). The association between grade, gender, physical activity, and back pain among children carrying schoolbags. *Archives of Exercise in Health and Disease,* 4(1), 234-242.
30. Forjuoh, S.N., Schuchmann, J.A. and Lane, B.L. (2004). Correlates of heavy backpack use by elementary school children. *Public Health,* 118(7), 532-535.
31. Moore, M.J., White, G.L. and Moore, D.L. (2007). Association of relative backpack weight with reported pain, pain sites, medical utilization, and lost school time in children and adolescents. *Journal of School Health,* 77(5), 232-239.
32. Whittfield, J., Legg, S.J. and Hedderley, D.I. (2005). Schoolbag weight and musculoskeletal symptoms in New Zealand secondary schools. *Applied Ergonomics,* 36(2), 193-198.

33. Chiang, H.Y., Jacobs, K. and Orsmond, G. (2006). Gender-age environmental associates of middle school students' low back pain. *Work*, 26(1), 19-28.
34. Navuluri, N. and Navuluri, R.B. (2006). Study on the relationship between backpack use and back and neck pain among adolescents. *Nursing & Health Sciences*, 8(4), 208-215.
35. Chansirinukor, W., Wilson, D., Grimmer, K. and Dansie, B. (2001). Effects of backpacks on students: measurement of cervical and shoulder posture. *Australian Journal of Physiotherapy*, 47(2), 110-116.
36. Korovessis, P., Koureas, G., Zacharatos, S. and Papazisis, Z. (2005). Backpacks, back pain, sagittal spinal curves and trunk alignment in adolescents: a logistic and multinomial logistic analysis. *Spine*, 30(2), 247-255.
37. Moore, M.J., White, G.L. and Moore, D.L. (2007). Association of relative backpack weight with reported pain, pain sites, medical utilization, and lost school time in children and adolescents. *Journal of School Health*, 77(5), 232-239.
38. Navuluri, N. and Navuluri, R.B. (2006). Study on the relationship between backpack use and back and neck pain among adolescents. *Nursing & Health Sciences*, 8(4), 208-215.
39. Puckree, T., Silal, S.P. and Lin, J. (2004). School bag carriage and pain in school children. *Disability and Rehabilitation*, 26(1), 54-59.
40. Siambanes, D., Martinez, J.W., Butler, E.W. and Haider, T. (2004). Influence of school backpacks on adolescent back pain. *Journal of Pediatric Orthopaedics*, 24(2), 211-217.
41. Whittfield, J., Legg, S.J. and Hedderley, D.I. (2005). Schoolbag weight and musculoskeletal symptoms in New Zealand secondary schools. *Applied Ergonomics*, 36(2), 193-198.
42. Hong, Y. and Cheung, C.K. (2003). Gait and posture responses to backpack load during level walking in children. *Gait & Posture*, 17(1), 28-33.
43. Panicker, R.K. and Sandesh, T.S. (2014). Prevalence of Musculoskeletal Pain in School Going Adolescents Using School Bags: A Co-relational Research. *International Journal of Therapies and Rehabilitation Research*, 3(4), 1.
44. Pascoe, D.D., Pascoe, D.E., Wang, Y.T., Shim, D.M. and Kim, C.K. (1997). Influence of carrying book bags on gait cycle and posture of youths. *Ergonomics*, 40(6), 631-640.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 239-246

SAMPLE SIZE ESTIMATION OF DIAGNOSTIC TEST STUDIES IN HEALTH SCIENCES

Syed Mustansir Hussain Zaidi, Hira Fatima Waseem, Fahim Ahmed Ansari, Muhammad Irfan and **Sania Fahim**
Department of Statistics, Liaquat National Hospital, Karachi, Pakistan
Email: mustansir.zaidi@lnh.edu.pk
fahim.ahmed@lnh.edu.pk
fatima.waseem@lnh.ed.pk
sania.amin@lnh.edu.pk
irfan.zafar@lnh.edu.pk

ABSTRACT

Sensitivity and specificity are terms used to evaluate a clinical test. Sensitivity and specificity measure inherent validity of a diagnostic test against a gold standard. Adequate sample size is precisely estimate the validity of a diagnostic test. They are dependent on the prevalence of the disease in the population of interest. The sensitivity and specificity of a quantitative test are dependent on the cut-off value above or below which the test is positive. The purpose of the study is to provide concept of sample size calculations of diagnostic test accuracy in various conditions and test outcomes. We will use formula in MS excel for sample size calculation of diagnostic test which include sensitivity, specificity, absolute precision, known prevalence of disease, and 95% confidence level. This could be easily used to determine the sample size for estimating the sensitivity or specificity of a diagnostic test. This would help out the medical professionals for designing diagnostic test studies of adequate sample is chosen based on statistical principles in order to guarantee the reliability of study.

KEYWORDS

Sample size, sensitivity, specificity, diagnostic test.

1. INTRODUCTION

In diagnostic studies of health sciences which the test yields dichotomized outcome (positive or negative results), the accuracy is evaluated by sensitivity and specificity.[1] Sensitivity and specificity are two components that measure the inherent validity of a diagnostic test compared to the gold standard.[2] These two measures are not influenced by prior probability of disease (or prevalence) in population.[1,3-5]

The gold standard may be another test without errors but a more expensive diagnostic method or invasive method. It can be the combination of tests that may be available in

The sensitivity indicates the proportion of diseased subject with positive test result and specificity determines the proportion of non-diseased subject with negative test results. When a gold standard is available, sensitivity and specificity can be estimated directly.[6-8] For example, diagnosis of pancreatic carcinoma can be confirmed only by laprotomy for alive or by autopsy for dead patients. Sometimes the gold standard is expensive, less widely available, more invasive, riskier, and takes more time to produce results. Such issues compel researchers to develop new diagnostic methods as surrogate to the gold standard.[2]

For implementing new diagnostic method in clinical practice, an adequate sample size is needed to ensure that the study will yield estimate of the sensitivity and specificity with acceptable precision, smaller sample size produces imprecise estimate, and unduly large sample is wastage of resources especially when the new method is expensive. Furthermore, the prevalence of disease was included in the sample size formula by Buderer, because the sample size without considering the prevalence would be adequate either for sensitivity or for specificity but not for both.[1,2]

Buderer provides the sample size tables for sensitivity and specificity but they are only for the 10% precision level. A study was conducted by Bochmann to assess the sample size calculation in diagnostic accuracy articles published in 2005 and found only 1 out of 40 studies reporting the sample size calculation before initiating the study.[1,9,10] This may be due to reluctance in using a mathematical formula or computer software. Malhotra and Indrayan[11] argued that the sample size without considering prevalence would be adequate for sensitivity or specificity alone but not for both while Obuchowski[12] addressed that this is because of unknown true disease status at time of sampling from target population. Charley et al.[13] have provided monogram for estimation of sensitivity and specificity with too many lines and curves make complexity in reading. Malhotra and Indrayan[11] presented a table of sample size calculation based on Borderer formula only for estimating sensitivity and specificity but not for testing. Obuchowski[12] also provided a review of sample size formula for a various diagnostic accuracy but did not provide practical tables for calculating sample sizes.

Several other authors also developed methods for sample size calculation in diagnostic medicine[14-17] because of complexity of their methods for clinician and the lack of availability of software their methods were not used frequently in clinical practices. In practice, researchers generally decide a sample size for validating a new diagnostic test arbitrarily or at their convenience or use the previous literature.

The objective of the study is to determine the sample size in diagnostic test studies of health sciences because adequate sample size is precisely estimate the validity of a diagnostic test in health sciences.

This paper is divided into five sections. Section 1 starts with introduction of sample size calculations for diagnostic health studies conceptually. In section 2 we addressed some conceptual definitions used in diagnostic studies. In section 3, the formula of sample size based on statistical principles with respect to study purpose and accuracy of interest (sensitivity/specificity) is discussed. In section 4, a view of the critical elements of calculated the required sample size for various situations and the calculated sample size were tabulated for practical convenience of clinician in diagnostic test evaluation. We discuss the final conclusion to overall analysis in section 5.

2. SOME BASIC DEFINITIONS OF DIAGNOSTIC TEST STUDIES

Diagnostic Accuracy: Diagnostic accuracy test study provides evidence on how well a test correctly identifies or rules out disease

"How well new test discriminates between certain two conditions of interest i.e. health and disease"

2.1 Measures of Diagnostic Accuracy

1. **Sensitivity *(True Positive - TP)*:** Sensitivity refers to the new test's ability to correctly detect patients who do have the disease. The percentage of sick people who are correctly identified as having the disease. Actually, sensitivity is defined as the probability of getting a positive test result in subjects with the disease. Hence, it relates to the potential of a test to recognize subjects with the disease.
   $$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \times 100$$
2. **Specificity (True Negative - TN):** Specificity relates to the new test's ability to correctly detect patients who do not have the disease. The percentage of healthy people who are correctly identified as not having the disease. Specificity represents the probability of a negative test result in a subject without the disease. Therefore, we can postulate that specificity relates to the aspect of diagnostic accuracy that describes the test ability to recognize subjects without the disease, i.e. to exclude the condition of interest.
   $$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \times 100$$
3. **Positive Predicative Values *(False Positive - FP):*** A proportion of patients with positive test of disease. Result in total subjects with positive result.
   $$\text{Positive Predictive Value} = \frac{\text{TP}}{\text{TP+FP}} \times 100$$
4. **Negative Predicative Values *(False Negative - FN):*** A proportion of patients without disease with a negative test result in total subjects with negative test result.
   $$\text{Negative Predictive Value} = \frac{\text{TN}}{\text{TN+FN}} \times 100$$

2.2 Gold Standard

The gold standard is the best single test (or a combination of tests) that is considered the current preferred method of diagnosing a particular disease (X). All other methods of diagnosing X, including any new test, need to be compared against this ′gold′ standard. The gold standard is different for different diseases.

2.3 Validity

It is the extent to which a test measures what it is supposed to measure; in other words, it is the accuracy of the test. Validity is measured by sensitivity and specificity. These terms, as well as other jargon, are best illustrated using a conventional two- by-two (2 x 2) table.

The information obtained by comparing a new diagnostic test with the gold standard is conventionally summarized in a two-by-two table.

	Gold Standard	
	Positive	**Negative**
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

2.4 Measurement of Diagnostic Effectiveness (Accuracy)

Another global measure of diagnostic accuracy is so called diagnostic accuracy (effectiveness), expressed as a proportion of correctly classified subjects (TP+TN) among all subjects (TP+TN+FP+FN). Diagnostic accuracy is affected by the disease prevalence. With the same sensitivity and specificity, diagnostic accuracy of a particular test increases as the disease prevalence decreases. This data, however, should be handled with care. In fact, this does not mean that the test is better if we apply it in a population with low disease prevalence. It only means that in absolute number the test gives more correctly classified subjects. This percentage of correctly classified subjects should always be weighed considering other measures of diagnostic accuracy, especially predictive values. Only then a complete assessment of the test contribution and validity could be made.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP}+\text{TN} + \text{FP}+\text{FN}} \times 100.$$

3. ESSENTIAL ELEMENTS FOR SAMPLE SIZE CALCULATION

Based on statistical principle, as a general rule of sample size calculation for proportions, since sensitivity (or specificity) is a proportion, it is intuitively appealing that the four essential elements are required for calculation of sample size in estimating sensitivity (or specificity):

- ❖ A pre-determined value of sensitivity (or specificity) that is available from previous published studies.
- ❖ The confidence level (1-α) for statistical judgment where α is the probability of type I error.
- ❖ The precision of estimates of sensitivity (or specificity) i.e. the maximum difference between estimated sensitivity (or specificity) and the true value.
- ❖ Additionally, the prevalence of disease in population is needed to be known and considered in sample size calculation.

Prevalence: It represents the number(%) of cases of a disease that are present in a particular population at a given time

$$\mathbf{P\ (\%) =} \frac{\textbf{Number of Cases of Disease}}{\textbf{Total No. of Population}}$$

3.1 Sample Size Formula for Sensitivity And Specificity

Buderer’s formula is used for sample size calculation in diagnostic accuracy studies at the required absolute precision level for sensitivity and specificity.

$$\text{Sample size } (n) \text{ based on sensitivity} = \frac{Z^2_{1-\alpha/2} \times S_N \times (1-S_N)}{L^2 \times Prevalence}, \text{ and}$$

$$\text{sample size } (n) \text{ based on specificity} = \frac{Z^2_{1-\alpha/2} \times S_P \times (1-S_P)}{L^2 \times (1-Prevalence)},$$

where n = required sample size,
S_N = anticipated sensitivity,
S_p = anticipated specificity,
α = size of the critical region ($1-\alpha$ is the confidence level),
$z_{1-\alpha/2}$ = standard normal deviate corresponding to the specified size of the critical region (α), and
L = absolute precision desired on either side (half-width of the confidence interval) of sensitivity or specificity.

3.2 MS Excel Sample Size Calculator for Diagnostic Test Studies:

This calculator is based on Buderer's formula for sensitivity and specificity of diagnostic health studies. This was created in MS Excel and downloads this calculator from link: http://www.kck.usm.my/ppsg/statistical_resources/samplesize_forsensitivity_specificitystudiesLinNaing.xls (*Written by Dr Lin Naing in 2004)*

4. EXAMPLES FOR SAMPLE SIZE CALCULATIONS IN DIAGNOSTIC HEALTH STUDIES

First we briefly described some examples of relevant biomedical informatics research for sample size calculation.

Example-1: **To determine the diagnostic accuracy of serial C-reactive protein (CRP) in detection of neonatal sepsis taking blood culture as gold standard.**

The information for sample size calculation:

Sensitivity=70%,[18] Specificity=72.3%,[18]
Prevalence=26.6%,[19] Desired Precision=13%,
95% confidence level

Sample Size (n)=181 patients

Example-2: Accuracy of MRCP in Diagnosis of Choledocolithiasis taking ERCP as Gold Standard.

The information for sample size calculation:

Sensitivity=85%,[20] Specificity=93%,[20]
Prevalence=20%,[21] Desired Precision=5%,
95% confidence level

Sample Size(n)=386 patients

Example-3: Accuracy of CT Scan in Evaluation of Gallbladder Carcinoma taking Histopathology as Gold Standard.

The information for sample size calculation:

Sensitivity=82.5%,[22] Specificity=75.9,[22]
Prevalence=9.1%,[23] Desired Precision=10%,
95% confidence level

Sample Size (n)=616 patients

Sample Size Ideas:

- The formula based on sensitivity and specificity yield a similar sample size at prevalence of 0.5.
- With low prevalence, the required sample size based on sensitivity is much higher than that of specificity
- While the prevalence of diseased becomes more than 0.50, the sample size based on sensitivity lower than that of specificity.

Case where Sensitivity is important: But sometimes the researcher is interested more in sensitivity than specificity in that case, the final sample size would be based on the sensitivity only.

Example-4: Sensitivity of CT Scan in Patients with Suspected Invasive Fungal Sinusitis Taking Histopathology as Gold Standard.

The information for sample size calculation:

Sensitivity=62%,[24]
Prevalence=40%,[25] Desired Precision=10%,
95% confidence level

Sample Size (n) = 228 patients

Example-5: Specificity of CT Scan in Patients with Suspected Invasive Fungal Sinusitis Taking Histopathology as Gold Standard.

The information for sample size calculation:

Specificity=97%,[24]
Prevalence=40%,[25] Desired Precision=1%,
95% confidence level

Sample Size (n) =1864 patients

5. CONCLUSION AND COMMENTS

Sample size estimation is key component in performing effective comparative studies. An understanding of the concepts of power, sample size, and type I and II errors will help the researcher and the critical reader of the medical literature. When evaluating a new diagnostic test, it is generally wise to assume that the sensitivity and specificity of the reference test are not precisely known, and to use available methods to estimate them as well. Unfortunately, sample sizes calculations are rarely reported by clinical investigators for diagnostic studies and few clinicians are aware of them. Researchers often decide about the sample size arbitrary either for their conveniences or from the previous literature.

The final sample size depends on the interest of the researcher. If sensitivity and specificity are equally important for the study, determine the sample size for both sensitivity and specificity, separately. The final sample size of the study would be the larger of these two. But sometimes the researcher is interested more in sensitivity than specificity. In that case, the final sample size would be based on the sensitivity only.

In this study, we used MS Excel calculator which is easy to use and can facilitate the medical professionals for designing studies for diagnostic accuracy, with adequate sample size in order to the reliability of study.

REFERENCES

1. Hajian-Tilaki, K. (2014). Sample size estimation in diagnostic test studies of biomedical informatics. *Journal of biomedical informatics*, 48, 193-204.
2. Malhotra, R.K. and Indrayan, A. (2010). A simple nomogram for sample size for estimating sensitivity and specificity of medical tests. *Indian Journal of Ophthalmology*, 58(6), 519-522.
3. Swets, J.A. (1979). ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology*, 14(2), 109-121.
4. Lasko, T.A., Bhagwat, J.G., Zou, K.H. and Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics*, 38(5), 404-415.
5. Hanley, J.A. (1989). Receiver operating characteristic (ROC) methodology: the state of the art. *Crit. Rev. Diagn. Imaging*, 29(3), 307-35.
6. Kumar, R. and Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics*, 48(4), 277-287.
7. Zou, K.H., O'Malley, A.J. and Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115(5), 654-657.
8. Enøe, C., Georgiadis, M.P. and Johnson, W.O. (2000). Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Preventive veterinary medicine*, 45(1), 61-81.
9. Aphinyanaphongs, Y., Tsamardinos, I., Statnikov, A., Hardin, D. and Aliferis, C.F. (2005). Text categorization models for high-quality article retrieval in internal medicine. *Journal of the American Medical Informatics Association*, 12(2), 207-216.
10. Bochmann, F., Johnson, Z. and Azuara-Blanco, A. (2007). Sample size in studies on diagnostic accuracy in ophthalmology: a literature survey. *British Journal of Ophthalmology*, 91(7), 898-900.
11. Malhotra, R.K. and Indrayan, A. (2010). A simple nomogram for sample size for estimating sensitivity and specificity of medical tests. *Indian Journal of Ophthalmology*, 58(6), 519-522.
12. Obuchowski, N.A. (1998). Sample size calculations in studies of test accuracy. *Statistical Methods in Medical Research*, 7(4), 371-392.

13. Carley, S., Dosman, S., Jones, S.R. and Harrison, M. (2005). Simple nomograms to calculate sample size in diagnostic studies. *Emergency Medicine Journal*, 22(3), 180-181.
14. Fosgate, G.T. (2009). Practical sample size calculations for surveillance and diagnostic investigations. *Journal of Veterinary Diagnostic Investigation*, 21(1), 3-14.
15. Li, J. and Fine, J. (2004). On sample size for sensitivity and specificity in prospective diagnostic accuracy studies. *Statistics in Medicine*, 23(16), 2537-2550.
16. Steinberg, D.M., Fine, J. and Chappell, R. (2009). Sample size for positive and negative predictive value in diagnostic research using case–control designs. *Biostatistics*, 10(1), 94-105.
17. Kumar, R. and Indrayan, A. (2002). A nomogram for single-stage cluster-sample surveys in a community for estimation of a prevalence rate. *International Journal of Epidemiology*, 31(2), 463-467.
18. Himayun, M., Ahmad, S. and Rasool, A. (2009). Role of C-Reactive protein in early onset neonatal sepsis. *Journal of Pediatric Neonatol*, 11(2).
19. Nuntnarumit, P., Pinkaew, O. and Kitiwanwanich, S. (2002). Predictive values of serial C-reactive protein in neonatal sepsis. *Journal of the Medical Association of Thailand= Chotmaihetthangphaet*, 85, S1151-8.
20. Verma, D., Kapadia, A., Eisen, G.M. and Adler, D.G. (2006). EUS vs MRCP for detection of choledocholithiasis. *Gastrointestinal Endoscopy*, *64*(2), 248-254.
21. Shadan, A., Malik, G.M., Kamili, M.M.A., Umar, K., Showkat, H., Willayat, A. and Suhail, M. (2011). Role of mrcp in the evaluation of suspected biliary and pancreatic disease. *JK–Practitioner*, 16(1-2), 20-25.
22. Vijayakumar, A., Vijayakumar, A., Patil, V., Mallikarjuna, M.N. and Shivaswamy, B.S. (2012). Early diagnosis of gallbladder carcinoma: an algorithm approach. *ISRN Radiology*, 2013.
23. Naqvi, S.Q.H., Mangi, I.H., Dahri, F.J., Khaskheli, Q.A. and Akhund, A.A. (2005). Frequency of carcinoma of gall bladder in patients with cholelithiasis. *Gomal Journal of Medical Sciences*, 3(2).
24. Dhong, H.J., Jung, J.Y. and Park, J.H. (2000). Diagnostic accuracy in sinus fungus balls: CT scan and operative findings. *American Journal of Rhinology*, 14(4), 227-231.
25. Groppo, E.R., El-Sayed, I.H., Aiken, A.H. and Glastonbury, C.M. (2011). Computed tomography and magnetic resonance imaging characteristics of acute invasive fungal sinusitis. *Archives of Otolaryngology–Head & Neck Surgery*, 137(10), 1005-1010.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 247-256

FERTILITY FORECASTING IN URBAN AND RURAL AREAS OF PAKISTAN: AN APPLICATION OF PRODUCT-RATIO FUNCTIONAL MODEL

Farah Yasmeen and **Hira Fatima Waseem**
University of Karachi, Karachi, Pakistan
Email: riazfarah@yahoo.com and hira.waseem@live.com

ABSTRACT

This paper is an application of coherent functional models of Hyndman, Booth and Yasmen (2013) on the disaggregation of fertility forecasts by regions. In this paper, we intend to obtain coherent forecasts of age-specific fertility rates of Pakistan for the two broader regions, namely, Urban and Rural areas. We found on the basis of previous studies, that fertility rates in rural areas are higher than those of urban areas. Here, we can assume that the future fertility rates of rural areas will remain higher than those of urban areas, for all age groups. Hence, the fertility forecasts thus obtained are assumed to be *coherent* or more realistic with the historical fertility rates.

At first, we describe the concept of coherence in the context of age-specific fertility and discuss some problems with using independent functional time series models, separately for urban and rural regions. Then, we apply the coherent functional model to the age-specific fertility data. An empirical comparison of the independent and coherent models based on the fertility rates has been made. The purpose here is to see the performance of coherent forecasting models for the fertility rates of these two regions of Pakistan and to find the fertility forecasts which are more realistic.

KEY WORDS

Age-specific fertility rates, rural-urban fertility gap, fertility, fertility levels, functional time series, forecasting.

1. INTRODUCTION

Demographers always interested to study the fertility during the last six decades. Fertility levels and differentials are directly impact to population growth. Fertility requires more attention than other demographic indicators because changes in fertility rates have important implications on social, economic and political situation for the persons living in a country. History of Pakistan also indicates that rural populations are more fertile than urban populations. Demographers always insist to model and forecast the reproductively pattern. Studies applied different mathematical models to study fertility include Nasir et al. (2009), Hyndman and Ullah(2007), Erbas et al. (2007).

Yasmeen and Waseem (2014) applied Functional Time Series (FTS) models to forecast the age-specific fertility of urban and rural areas fertility. Their forecasting pattern showed that urban fertility is quite lower than the rural one due to lack of educational facilities, early marriages as well as less awareness about family planning in rural area of Pakistan. Forecasting pattern from Independent FTS Model on ASFR of urban verses rural areas are almost divergent in the long term.

Non-divergent forecasts for subpopulations within a larger population have been labeled "coherent" by Li and Lee (2005). Coherent forecasting based on product-ratio functional model seeks to ensure that the forecasts for related subgroups maintain certain structural relationships based on extensive historical observations. However, demographers interested to obtain resultant forecast will exhibit relationship on the basis of historical observations of subgroups. To obtain exhibit relationship, coherent based on product and ratio functional model have been used by Hyndman et al. (2013) for forecasting mortality of Sweden and Australia, Nikitovic (2011) used the same approach for forecasting fertility Serbian region-wise (Central Serbia and Vojvodina).

In this paper, we apply coherent based on product-ration functional time series models to the age-specific fertility rates of urban and rural areas of Pakistan to obtain non-divergent forecast in the long term. The main objectives of this paper are:

- To apply product-ratio functional model to fertility rates of urban and rural areas of Pakistan.
- To observe components of product fertility rates and ratio fertility rates of two regions
- To obtain coherent forecasts based on product-ratio functional model of fertility curve for the next twenty years.

This paper is divided into five sections. Section 1 starts with introduction of the fertility of urban and rural areas of Pakistan and different statistical models for forecasting fertility are discussed. In section 2, product-ratio functional model are discussed. In section 3, source of data of fertility rates and steps of product-ratio functional model are discussed. In section 4, the application of product-ratio functional model to fertility rates are discussed. Final comments and conclusion is given in section 5.

2. COHERENT BASED ON PRODUCT-RATIO FUNCTIONAL MODEL

Coherent based on product-ratio functional model that provides forecasts of two or more groups within a given population by simultaneously modeling the geometric mean of the age-specific fertility rates and the ratio of the group-specific rates to the geometric mean of the fertility rates of the population as a whole.

Suppose $f_{t,r}(x)$ denote the age-specific fertility rate for rural women of age x in year t, t = 1…, n. We model the log fertility rate, $y_{t,r}(x) = \log[f_{t,r}(x)]$, and assume that there is an underlying smooth function that we are observing with error. Thus,

$$y_{t,r}(x) = \log[f_{t,r}(x)] + \sigma_{t,r}(x_i)\,\epsilon_{t,w,i} \tag{1}$$

where

x_i = the centre of age group $i, (i = 1, \ldots.., p)$,

$\varepsilon_{t,r,i}$ = independent and identically distributed standard normal random variable and

$\sigma_{t,r}(x_i)$ = allows the amount of noise to vary with age 'x'.

Similar notation is used for urban fertility rates.

Product-Ratio Method for Rural and Urban

The product component represents the overall effect of the whole rates irrespective of urban and rural groups. The ratio component shows the square root of the ratio of the fertility rates of urban to the fertility rates of rural.

Square roots of the products $p_t(x)$ and ratios $r_t(x)$ of the smoothed rates for rural and urban regions are as follows:

$$p_t(x) = \sqrt{f_{t,r}(x)\, f_{t.u}(x)} \tag{2}$$

$$r_t(x) = \sqrt{f_{t,r}(x) / f_{t.u}(x)} \tag{3}$$

We model products and ratios of the smoothed rates rather than the original sex-specific fertility rates because the benefit of this approach is that the product and ratio behave generally independent of each other, provided that the subgroups have approximately equal variances.

Components of Product-Ratio Functional Models

We use functional time series models (Hyndman and Ullah 2007) for components of $p_t(x)$ and $r_t(x)$. Thus,

$$\log\left[p_t(x)\right] = \mu_p(x) + \sum_{k=1}^{k} \alpha_{t.k}\phi_k(x) + e_t(x) \tag{4}$$

$$\log\left[r_t(x)\right] = \mu_r(x) + \sum_{l=1}^{l} \beta_{t.l}\omega_l(x) + h_t(x) \tag{5}$$

where $\left\{\phi_k(x)\right\}$ and $\left\{\omega_l(x)\right\}$ = principal components obtained from decomposing

$\left\{p_t(x)\right\}$ and $\left\{r_t(x)\right\}$. $\alpha_{t,k}$ & $\beta_{t.l}$ = corresponding principal component scores, $\mu_p(x)$ = mean of the set of curves $\left\{p_t(x)\right\}$, $\mu_r(x)$ = mean of the set of curves $\left\{r_t(x)\right\}$. and $e_t(x)$ & $h_t(x)$ = error terms have zero mean and are serially uncorrelated.

The models avoid the problem of the functions $\{\varphi_k(x)\}$ and $\{\omega_k(x)\}$ changing over time.

Forecasting of Product-Ratio Functional Models

The product model coefficients, $\{\alpha_{t,1}, \alpha_{t,2,} \ldots \ldots \ldots \ldots, \alpha_{t,k}\}$, forecast by using probably nonstationary autoregressive integrated moving average (ARIMA) models without restriction. When fitting ARIMA models, we use the automatic model selection algorithm given by Hyndman and Khandakar (2008)[8] to select the appropriate model orders.

The ratio model coefficients, $\{\beta_{t,1,j.}\beta_{t,2.j}, \ldots \ldots \ldots \ldots, \beta_{t,l\,.j}\}$, j = 1, 2,..., J, forecast by using autoregressive fractionally integrated moving-average (ARFIMA)(p,d,q) process. The stationarity requirement ensures that the forecasts will be coherent. We use the algorithm of Hyndman and Khandakar (2008)[8] to choose the orders p and q, In the stationary ARFIMA models, $-0.5 < d < .05$.

3. MATERIAL AND METHODS

3.1 Source of Data

The secondary data of age-specific fertility rates (ASFR), obtained from Pakistan Demographic Surveys from 1984 to 2007 are used. These data are available for 1984-1986, 1988-1992, 1995-1997, 1999-2001, 2003 and 2005-2007. Missing values of data in between the years 1984 to 2007 are estimated by interpolating spline function in R Software 3.1.2 version.

3.2 Fitting Coherent FTS Models

Firstly, we smooth ASFR for the years 1984-2007 of both regions by using weighted regression B spline and consider them as functional observation. Secondly, the product component $p_t(x)$ and ratio component $r_t(x)$ of coherent functional model are obtained by using equation (2) and (3) as shown in figure 1 and 2 which represent smoothed log fertility rates. In third step, we apply functional time series model to the product and ratio components. We use ARIMA models to forecast the product component coefficient and forecasting curve as shown in figure 3. For forecasting the ratio coefficients, we use stationary ARFIMA (p, d, q) $(-0.5 < d < 0.5)$ models and obtain forecasts as shown in figure 4. The twenty years forecast based on coherent functional model of urban and rural fertility as shown in figure 5 and 6.

4. RESULTS

Figure 1 shows product component of smoothed log fertility rates for women of both groups (urban verses rural) increased with age till age 27 years, and then they declined. This decline was sharper in the age groups of 25 to 35 years and relatively slower in the earlier and older ages.

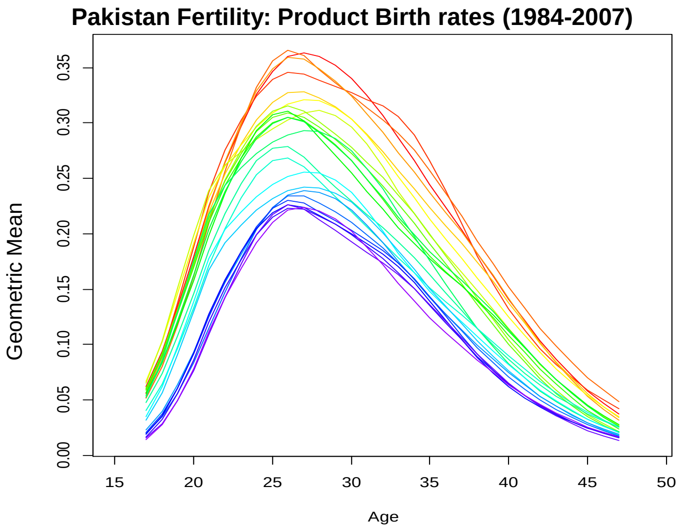


Fig. 1: Product Component of Fertility Rates of Pakistan (Urban and Rural Areas)

Figure 2 shows ratio component of smoothed log fertility and represent graphically that urban to rural fertility is less than 1 which suggests that the urban women show lower fertility as compared to rural women for all ages during 1984-2007. Also, the ratio is decreasing for the women aged below 25 years and over 35 years, where the ratio of women between ages of 25 to 35 years is relatively stable, It means that the major differences in the fertility rates of urban and rural populations occur in younger and older women and relatively stable in middle age-group women.

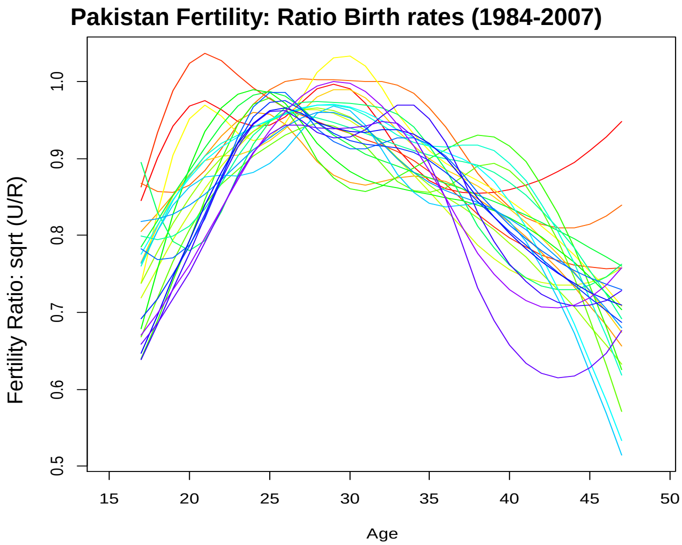


Fig. 2: Ratio Component of Fertility Rates of Pakistan (Urban and Rural Areas)

Figures 3 and 4 shows that basis functions and respective time series coefficients obtained by applying the functional principal components decomposition to the product and ratio curves, along with twenty-year forecasts of the time series coefficients. These plots represent the various sources of variation in the two components.

Figure 3 shows that the mean function of the product model is an increasing till age 27 years and then it is declining. The first basis function shows that ASFR for the two groups of women were decreasing for younger women. Similarly, the second basis function is a contrast between very young (15 years) and very old (45 years). Its respective coefficient shows that the rates for older women were initially decreasing from 1984 till 2000 and then they started to increase slightly from year 2000. These rates are expected to decline slowly in the next 20-years. Figure 4 shows that the mean function of the ratio component represent which represents the urban to rural fertility ratio increases with age till age 25 years and then decreasing. The first basis function of the ratio component shows that most of the variation in the urban to rural fertility-ratio is in the age groups of women aged 17 years and under, and above 45 years and the ratio of their fertility rates is decreasing since 1984. While the second basis function represent ages 40 years and above, with corresponding rates are expected to decrease in the next 20-years. The twenty-year forecasts of the first ratio coefficient show that urban to rural fertility ratio may remain constant in the future. This implies that the forecasts for the entire fertility curves for urban and rural women will not diverge in the long run. The other basis functions and coefficients have little effect on the ratio forecast, as the forecast values of the coefficients are nearly zero.

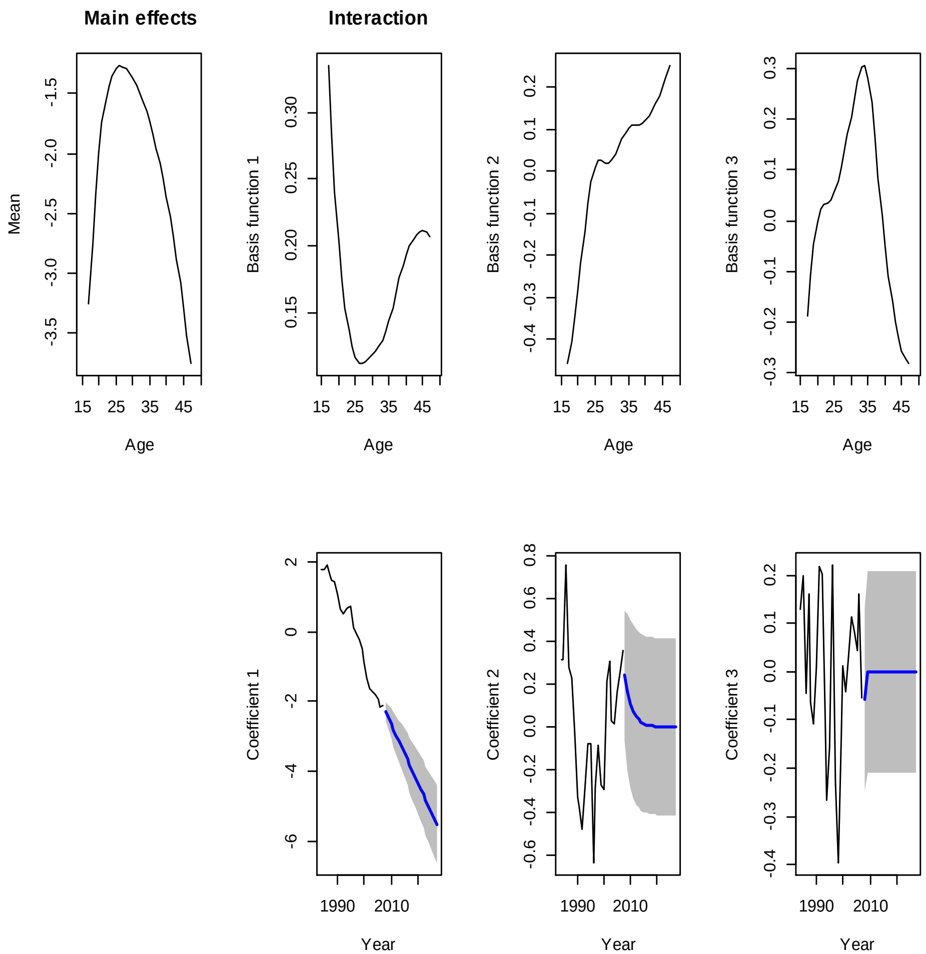


Fig. 3: Forecast for the First Three Coefficients of FTS Model for Product Component

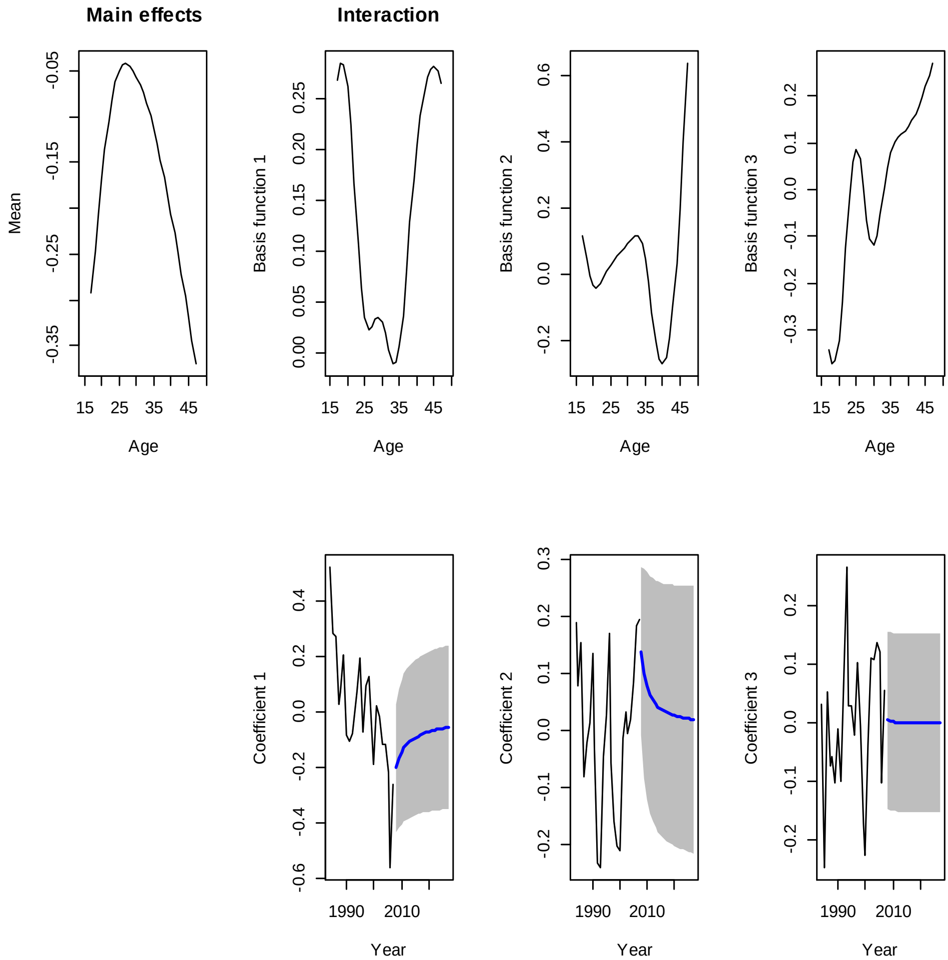


Figure 4: Forecast for the First Three Coefficients of FTS Model for Ratio Component

From figures 3 and 4, we also observed that the coefficients of product models are non-stationary, whereas the coefficients of ratio models are stationary with narrow prediction intervals. Hence, the stationary condition makes the forecasts of the two groups to be non-divergent.

Figures 5 and 6 shows twenty-year forecasts of age-specific fertility for urban and rural areas of Pakistan. These graphs show a decline in fertility rates for both urban and rural women over the next twenty years. This decline is steeper among urban women of middle age (25-35 years), and relatively slower among younger women (under 25 years of age) and older women (above 35 years). However, for rural areas, future fertility rates are expected to decline substantially for women aged between 22 and 38 years, and relatively slowly for women of other ages.

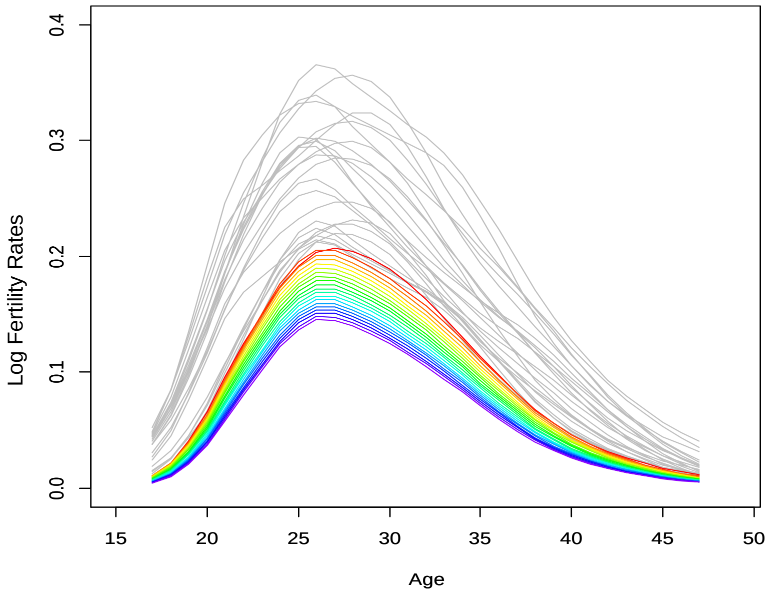


Fig. 5: Twenty Year Fertility Forecasts of Pakistan Urban Area using Coherent Functional Models

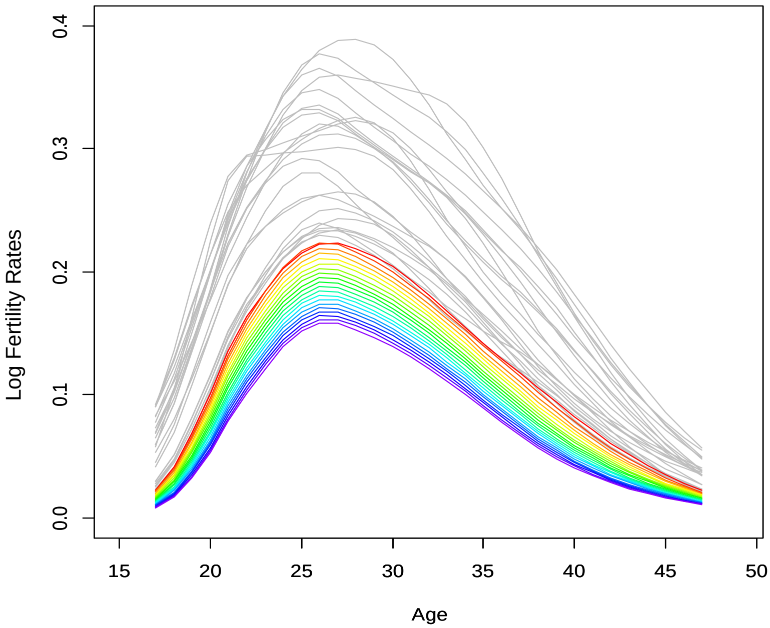


Fig. 6: Fertility Forecasts of Pakistan Rural Region Using Coherent Functional Models

5. COMMENTS AND CONCLUSION

We have observed that rural women have higher fertility rates than urban women (Sathar and Casterline, 1998, Yasmeen, Waseem and Mahmood (2014) in all age-groups from 1984 till 2007 in Pakistan. Transition in fertility has been observed after 1990s in Pakistan. Fertility differences in urban verses rural women of Pakistan due to differences in their socio-economic conditions, literacy rate, level of education and age at marriage.

In this paper, we label fertility forecasts as coherent when the forecast ASFR for any two subgroups (urban verses rural) converge to a set of appropriate constants. We applied product-ratio functional model to forecast fertility rates of urban verses rural areas to converge forecast in the long term. Our result showed that differences of future fertility rates are expected to be maintained over the next twenty years. We obtained 20-years prediction plot for the ASFR of Pakistan. These plots exhibit that the future fertility rates will decline for all ages, but that the decline will be greatest for the middle age women, and relatively slower for the youngest and oldest women. The results obtained here are consistent with other studies findings that the overall fertility rates are decreasing. by Yasmeen and Mahmood (2012), Yasmeen, Waseem and Mahmood (2014). This study forecasting result will help the government to allocate family planning services, contraceptive awareness program and educational facilities should be implementing in rural areas of Pakistan because major part of the population of Pakistan is still inhabited in rural areas.

REFERENCES

1. Hyndman, R.J., Booth, H. and Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50(1), 261-283.
2. Nasir J.A., Akhtar, M. and Tahir, M.H. (2009). Reproductivity and age-specific fertility rates in Pakistan after 1981. *Pak. J. Statist.*, 25(3), 251-263.
3. Hyndman, R.J. and Ullah, M.S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics and Data Analysis*, 51(10), 4942-4956.
4. Erbas, B., Hyndman, R.J and Gertig, D.M. (2007). Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine*, 26, 458-470.
5. Yasmeen, F., Waseem, H.F. and Mahmood, Z. (2014). An FDA approach to forecast age-specific fertility rates of Pakistan region-wise. *Computer Science,* 1(6), 341-348.
6. Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42(3), 575-594.
7. Nikitović, V. (2011). Functional data analysis in forecasting Serbian fertility. *Stanovnistvo*, 49(2), 73-89.
8. Hyndman, R.J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1-22.
9. Government of Pakistan (2007). *Population Demographic Survey-2007*, Federal Bureau of Statistics, Statistics Division, Islamabad, Pakistan.
10. Mosk, C. (1980). Rural-Urban Fertility Differences and the Fertility Transition, *JSTOR: Population Studies*. (34)1, 77-90.

11. Peristera, P. and Kostaki, A. (2007). Modelling fertility in modern populations. *Demographic Researc*h, 16, 141-194.
12. Sathar, Z.A. and Casterline, J.B (1998). The onset of fertility transition in Pakistan. *Population and Development Review*, 24, 773-796.
13. Sathar, Z.A. and Kazi, S. (1990). Women, work and reproduction in Karachi. *International Family Planning Perspectives*, 16, 66-80.
14. Yasmeen, F. and Mahmood, Z. (2012). Forecasting age-specific fertility rates of Pakistan using functional time series models, *Proc. of 12th Islamic Countries Conference on Statistical Sciences (ICCS-12)*, Doha, Qatar, 23, 193-198.
15. Hakim, M.A. (2000). Fertility Transition and Differentials in Pakistan, Pakistan's Population Issues in the 21st Century. *Conference Proceedings, 24-26 October Karachi. Islamabad: Population Association of Pakistan,* 611-633.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 257-278

DEVELOPING A SENSITIVE NON-PARAMETRIC CONTROL CHART UNDER THE REPETITIVE SAMPLING SCHEME

Nasrullah Khan[1], Muhammad Azam[1], Liaquat Ahmad[1], Muhammad Aslam[2] and **Munir Ahmad[3]**

[1] University of Veterinary and Animal Sciences, Lahore, Pakistan
Email: nas_shan1@hotmail.com; mazam72@yahoo.com
liaquatahmad@uvas.edu.pk

[2] Deptt. of Statistics, Faculty of Sciences, King Abdulaziz University
Jeddah, Saudi Arabia. Email: aslam_ravian@hotmail.com

[3] National College of Business Administration and Economics
Lahore, Pakistan. Email: drmunir@ncbae.edu.pk

ABSTRACT

A control chart is the most extensively used graphical tool for monitoring the manufacturing process. The early detection of the out-of-control process is the major concern of any control process under usage. For this purpose the improved control chart for different sampling scheme have been developed. In this paper, we will propose three non-parametric control charts using the repetitive sampling. The proposed control charts are more sensitive to detect early shift in the process. A simulation study is given to show the efficiency of proposed charts. A real data example is given to illustrate the proposed control charts.

KEYWORDS

Non-parametric statistic, Average run length, Monte Carlo simulation, Repetitive sampling scheme

1. INTRODUCTION

Control chart is an important tool of statistical process control for monitoring the production process for any unusual change in the process. Walter A. Shewhart introduced, during 1920's, the idea of graphical display of the numerical values on the chart to improve the quality of the products. A lot of literature since that has been published for improving the monitoring of the production process and thus to reduce the loss (Montgomery, 2013).

It is common practice in control chart literature that whenever the interested quality characteristic is measured in continuous scale, say X, then it is assumed that the underlying distribution follows the normal distribution. If X is assumed as approximately normally distributed then the changes in the non-random causes may affect the parameters of the distribution (Stoumbos & Reynolds, 2000). A distribution free or the non-parametric technique provide a robust alternative to the researchers whenever the distribution of underlying phenomena is unknown or the usual normality condition do not met (A. Mukherjee, Chakraborti & Graham, 2012). The prime edge in applying the non-parametric technique is the flexibility obtained from no need to assume the distributional form of the probability distribution. It is worth mentioning here that non-parametric technique does not mean no parameters at all (Das, 2009). The X-bar and dispersion charts for the non-parametric scenario have been studied by many authors including (Abbasi, Miller, & Riaz, 2013; Das, 2008). Likewise the EWMA and CUSUM control charts for non-parametric methods have been studied by many researchers including (Graham, Chakraborti, & Human, 2011; S.-Y. Li, Tang, & Ng, 2010; Amitava Mukherjee & Chakraborti, 2012; Su-Fen, Wen-Chi, Tzee-Ming, Chi-Chin, & Smiley, 2011; S.-F. Yang, Lin, & Cheng, 2011; S. F. Yang & Cheng, 2011).

The repetitive group sampling (RGS) scheme was introduced by (Sherman, 1965) which has attracted the attention of many control chart experts for its efficiency over the single sampling scheme. Sherman (1965) claimed that the RGS scheme is intermediate in efficiency between single sampling scheme and the sequential probability ration sampling scheme. Balamurali, Park, Jun, Kim, and Lee (2005) used the RGS scheme for developing the acceptance sampling plan and proved that RGS scheme is efficient in average sample number as compared to the single and double sampling schemes. The RGS scheme has been studied by many authors including (Ahmad, Aslam, & Azam, 2015; Ahmad, Aslam, & Azhar, 2015; Ahmad, Aslam, & Jun, 2014, 2015; Muhammad Aslam, Azam, & Jun, 2013; Muhammad Aslam, Lio, & Jun, 2013; M. Aslam, Niaki, Rasool, & Fallahnezhad, 2012; Muhammad Aslam, Wu, Jun, Azam, & Itay, 2013; Muhammad Aslam, Yen, & Jun, 2011; Balamurali & Jun, 2006; Lee, Aslam, & Jun, 2012; Yen, Chang, & Aslam, 2015).

The ARL is the most popular measure, among the quality control researchers, used for comparing the performance of the designed control chart schemes. If LR be the random variable relating to the number of samples falls within control limits till an out-of-control signal is received then the probability distribution of this random variable LR is said to the run length distribution. Thus ARL can be defined as the average of the run length distribution. The in-control ARL is denoted by ARL0 while the out-of-control ARL is denoted by ARL1.

The rest of the paper is organized as: the design of the proposed control chart is elaborated in Section 2. Section 3 explains the performance evaluation of the proposed control chart. In Section 4 a comparison of the proposed control chart with the existing control charts. In Section 5 an application of the proposed chart is presented through examples and finally Section 6 gives a summary and concluding remark.

2. DESIGN OF THE PROPOSED CONTROL CHARTS

Let $\mathbf{X}$ be a quality characteristic under study with the mean $\boldsymbol{\mu}$, we obtain the random sample of size $\mathbf{n}$ and computed a transformed random variable $\mathbf{Y_i = X_i - \mu}$ for all $\mathbf{i's}$. Let $\mathbf{p = P(Y_i > 0)}$ the objective is to monitor the location parameter by ensuring the stability in $\mathbf{p}$. We take $\mathbf{I}$ as indicator function takes value 1 for positive $\mathbf{Y_i}$ and zero for elsewhere, we introduce a new variable $\mathbf{M = \sum_{i=1}^{n} I_i}$ which can be transformed as

$$\mathbf{Z = Sin^{-1}\left(\sqrt{M/n}\right)}$$

The random variable $\mathbf{M}$ follows the binomial distribution with parameters $\mathbf{n}$ and $\mathbf{p_0}$ and the transformed variable $\mathbf{Z}$ follows the approximately normal distribution with mean $\mathbf{Sin^{-1}}\left(\sqrt{\mathbf{p_0}}\right)$ and variance $\frac{\mathbf{1}}{\mathbf{4n}}$ (S.-F. Yang et al., 2011).

The NPS Control chart:

$$\mathbf{M_t} = \sum_{i=1}^{n} \mathbf{I_i} \tag{1}$$

The mean and variance for $\mathbf{W_t}$ will be as

$$\mathbf{E(M_t) = np_0} \tag{2}$$

$$\mathbf{Var(M_t) = np_0(1 - p_0)} \tag{3}$$

The control limits for the repetitive sampling scheme are

$$\mathbf{LCL_1 = np_0 - k_1\sqrt{np_0(1 - p_0)}} \tag{4}$$

$$\mathbf{LCL_2 = np_0 - k_2\sqrt{np_0(1 - p_0)}} \tag{5}$$

$$\mathbf{UCL_2 = np_0 + k_2\sqrt{np_0(1 - p_0)}} \tag{6}$$

$$\mathbf{UCL_1 = np_0 + k_1\sqrt{np_0(1 - p_0)}} \tag{7}$$

The probability of in-control under repetitive scheme is

$$\mathbf{P_a = P(LCL_2 \le M_t \le UCL_2)}$$
$$\mathbf{P(LCL_2 \le M_t \le UCL_2) = 2.\Phi(k_2) - 1}$$

and

$$\mathbf{P_{rep} = P(LCL_1 \le M_t \le LCL_2) + P(UCL_2 \le M_t \le UCL_1)}$$
$$\mathbf{P(LCL_1 \le M_t \le LCL_2) = \Phi(k_1) - \Phi(k_2)}$$
$$\mathbf{P(UCL_2 \le M_t \le UCL_1) = \Phi(k_1) - \Phi(k_2)}$$

Thus

$$\mathbf{P_{in} = \frac{P_a}{1 - P_{rep}}} \tag{8}$$

The average run length for in control process is

$$\mathbf{ARL_0 = \frac{1}{1 - P_{in}}} \tag{9}$$

For the shifted process we assume that our location parameter is shifted due to that the mean is shifted from $\mathbf{p_0}$ to $\mathbf{p_1 = cp_0}$ thus the shifted process has the mean and variance respectively as

$$\mathbf{E(M_t) = np_1} \tag{10}$$

$$\mathbf{Var(M_t) = np_1(1 - p_1)} \tag{12}$$

where $\mathbf{p_1 = cp_0}$

The probability of in control for shifted process is computed as

$$\mathbf{P_{a1} = P(LCL_2 \le M_t \le UCL_2 | p = p_1)}$$

$$\mathbf{P_{a1}} = \mathbf{\Phi}\left(\frac{\mathbf{UCL_2 - np_1}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right) - \mathbf{\Phi}\left(\frac{\mathbf{LCL_2 - np_1}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right)$$

$$\mathbf{P_{a1}} = \mathbf{\Phi}\left(\frac{\mathbf{np_0(1-c) + k_2\sqrt{np_0(1-p_0)}}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right) - \mathbf{\Phi}\left(\frac{\mathbf{np_0(1-c) - k_2\sqrt{np_0(1-p_0)}}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right)$$

and

$$\mathbf{P_{rep1}} = \mathbf{P(LCL_1 \le M_t \le LCL_2 | p = p_1) + P(UCL_2 \le M_t \le UCL_1 | p = p_1)}$$

$$\mathbf{P(LCL_1 \le M_t \le LCL_2 | p = p_1)} = \mathbf{\Phi}\left(\frac{\mathbf{LCL_2 - np_1}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right) - \mathbf{\Phi}\left(\frac{\mathbf{LCL_1 - np_1}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right)$$

$$\mathbf{P(LCL_1 \le M_t \le LCL_2 | p = p_1)} = \mathbf{\Phi}\left(\frac{\mathbf{np_0(1-c) - k_2\sqrt{np_0(1-p_0)}}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right) - \mathbf{\Phi}\left(\frac{\mathbf{np_0(1-c) - k_1\sqrt{np_0(1-p_0)}}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right)$$

Also

$$\mathbf{P(UCL_2 \le M_t \le UCL_1 | p = p_1)} = \mathbf{\Phi}\left(\frac{\mathbf{UCL_2 - np_1}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right) - \mathbf{\Phi}\left(\frac{\mathbf{UCL_1 - np_1}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right)$$

$$\mathbf{P(UCL_2 \le M_t \le UCL_1 | p = p_1)} = \mathbf{\Phi}\left(\frac{\mathbf{np_0(1-c) + k_1\sqrt{np_0(1-p_0)}}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right) - \mathbf{\Phi}\left(\frac{\mathbf{np_0(1-c) + k_2\sqrt{np_0(1-p_0)}}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right)$$

Thus

$$\begin{aligned}\mathbf{P_{rep1}} = {} & \mathbf{\Phi}\left(\frac{\mathbf{np_0(1-c) - k_2\sqrt{np_0(1-p_0)}}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right) - \mathbf{\Phi}\left(\frac{\mathbf{np_0(1-c) - k_1\sqrt{np_0(1-p_0)}}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right) \\ & + \mathbf{\Phi}\left(\frac{\mathbf{np_0(1-c) + k_1\sqrt{np_0(1-p_0)}}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right) - \mathbf{\Phi}\left(\frac{\mathbf{np_0(1-c) + k_2\sqrt{np_0(1-p_0)}}}{\sqrt{\mathbf{np_1(1-p_1)}}}\right)\end{aligned}$$

Thus the probability of in-control for the shifted process is as under

$$P_{in1} = \frac{P_{a1}}{1 - P_{rep1}}$$

Thus the average run length for the shifted process is

$$ARL_1 = \frac{1}{1 - P_{in1}} \tag{13}$$

Table 1
Average Run Length Values of the NPS Chart under RGS Scheme when $r_0 = 370$ and $p_0 = 0.50$

Shift	r_0=370; p_0=0.5; k_1=3.0108; k_2=2.069					
	n					
	5	10	20	30	50	100
	Arls					
0.05	191.19	1.00	1.00	1.00	1.00	1.00
0.10	51.17	1.11	1.00	1.00	1.00	1.00
0.30	36.38	4.18	1.12	1.01	1.00	1.00
0.50	60.89	15.45	3.28	1.58	1.06	1.00
0.60	86.84	29.52	7.80	3.41	1.48	1.02
0.70	130.76	58.32	20.39	9.87	3.72	1.33
0.75	163.13	83.54	34.05	18.00	7.25	2.10
0.80	204.40	121.17	58.30	34.15	15.49	4.50
0.85	254.29	176.56	102.27	67.49	35.79	12.34
0.90	307.82	251.96	180.73	137.61	88.75	40.58
0.95	352.23	332.02	297.17	268.21	222.93	152.37
0.99	369.26	368.33	366.49	364.66	361.05	352.31
1.00	370.00	370.00	370.00	370.00	370.00	370.00
1.01	369.26	368.33	366.49	364.66	361.05	352.31
1.05	352.23	332.02	297.17	268.21	222.93	152.37
1.10	307.82	251.96	180.73	137.61	88.75	40.58
1.15	254.29	176.56	102.27	67.49	35.79	12.34
1.20	204.40	121.17	58.30	34.15	15.49	4.50
1.25	163.13	83.54	34.05	18.00	7.25	2.10
1.30	130.76	58.32	20.39	9.87	3.72	1.33
1.35	105.89	41.26	12.48	5.65	2.17	1.09
1.40	86.84	29.52	7.80	3.41	1.48	1.02
1.50	60.89	15.45	3.28	1.58	1.06	1.00
1.70	36.38	4.18	1.12	1.01	1.00	1.00
1.90	51.17	1.11	1.00	1.00	1.00	1.00
1.95	191.19	1.00	1.00	1.00	1.00	1.00

Table 2
Average Run Length Values of the NPS Chart under RGS Scheme when $r_0 = 300$ and $p_0 = 0.50$

Shift	r_0=300; p_0=0.5; k_1=2.9426; $k_2$2.2155					
	n					
	5	10	20	30	50	100
	Arls					
0.05	141.13	1.01	1.00	1.00	1.00	1.00
0.10	42.35	1.18	1.00	1.00	1.00	1.00
0.30	31.45	4.26	1.17	1.01	1.00	1.00
0.50	52.08	14.25	3.37	1.67	1.08	1.00
0.60	73.64	26.31	7.53	3.48	1.56	1.03
0.70	109.76	50.52	18.58	9.38	3.77	1.39
0.75	136.12	71.45	30.26	16.52	7.03	2.19
0.80	169.44	102.35	50.65	30.38	14.35	4.49
0.85	209.30	147.29	87.03	58.35	31.79	11.59
0.90	251.56	207.56	150.75	115.92	75.98	35.86
0.95	286.22	270.50	243.25	220.46	184.56	127.94
0.99	299.42	298.71	297.28	295.86	293.07	286.29
1.00	300.00	300.00	300.00	300.00	300.00	300.00
1.01	299.42	298.71	297.28	295.86	293.07	286.29
1.05	286.22	270.50	243.25	220.46	184.56	127.94
1.10	251.56	207.56	150.75	115.92	75.98	35.86
1.15	209.30	147.29	87.03	58.35	31.79	11.59
1.20	169.44	102.35	50.65	30.38	14.35	4.49
1.25	136.12	71.45	30.26	16.52	7.03	2.19
1.30	109.76	50.52	18.58	9.38	3.77	1.39
1.35	89.36	36.23	11.69	5.57	2.27	1.12
1.40	73.64	26.31	7.53	3.48	1.56	1.03
1.50	52.08	14.25	3.37	1.67	1.08	1.00
1.70	31.45	4.26	1.17	1.01	1.00	1.00
1.90	42.35	1.18	1.00	1.00	1.00	1.00
1.95	141.13	1.01	1.00	1.00	1.00	1.00

Table 3
Average Run Length Values of the NPS Chart under Single Sampling Scheme when $r_0 = 370$ and $p_0 = 0.50$

Shift	r_0=370; p_0=0.5; k=2.999673					
	n					
	5	10	20	30	50	100
	Arls					
0.05	395.70	1.98	1.00	1.00	1.00	1.00
0.10	85.02	2.76	1.01	1.00	1.00	1.00
0.30	44.86	7.38	1.75	1.14	1.00	1.00
0.50	67.11	19.72	5.29	2.62	1.37	1.01
0.60	92.42	34.25	10.73	5.30	2.35	1.16
0.70	135.56	63.31	24.33	12.90	5.60	2.00
0.75	167.37	88.47	38.47	21.69	9.84	3.30
0.80	207.90	125.78	63.03	38.47	18.93	6.50
0.85	256.84	180.44	106.94	72.19	40.08	15.47
0.90	309.25	254.52	184.50	141.93	93.41	44.93
0.95	352.65	332.90	298.81	270.45	226.00	156.50
0.99	369.27	368.37	366.57	364.79	361.27	352.73
1.00	370.00	370.00	370.00	370.00	370.00	370.00
1.01	369.27	368.37	366.57	364.79	361.27	352.73
1.05	352.65	332.90	298.81	270.45	226.00	156.50
1.10	309.25	254.52	184.50	141.93	93.41	44.93
1.15	256.84	180.44	106.94	72.19	40.08	15.47
1.20	207.90	125.78	63.03	38.47	18.93	6.50
1.25	167.37	88.47	38.47	21.69	9.84	3.30
1.30	135.56	63.31	24.33	12.90	5.60	2.00
1.35	111.12	46.16	15.91	8.07	3.48	1.42
1.40	92.42	34.25	10.73	5.30	2.35	1.16
1.50	67.11	19.72	5.29	2.62	1.37	1.01
1.70	44.86	7.38	1.75	1.14	1.00	1.00
1.90	85.02	2.76	1.01	1.00	1.00	1.00
1.95	395.70	1.98	1.00	1.00	1.00	1.00

The NPAS Control Chart:

We introduce a new variable $\mathbf{M} = \sum_{i=1}^{n} \mathbf{I_i}$ which can be transformed as

$$\mathbf{Z_t} = \mathbf{Sin^{-1}}\left(\sqrt{\mathbf{M}/\mathbf{n}}\right) \tag{14}$$

The mean and variance for $\mathbf{W_t}$ will be as

$$\mathbf{E(Z_t) = sin^{-1}\sqrt{p_0}} \tag{15}$$

$$\mathbf{Var(Z_t)} = \frac{\mathbf{p_0(1-p_0)}}{\mathbf{n}} \tag{16}$$

The control limits for the RGS scheme are

$$\mathbf{LCL_1 = \sin^{-1}\sqrt{p_0} - k_1\sqrt{\frac{p_0(1-p_0)}{n}}} \tag{17}$$

$$\mathbf{LCL_2 = \sin^{-1}\sqrt{p_0} - k_2\sqrt{\frac{p_0(1-p_0)}{n}}} \tag{18}$$

$$\mathbf{UCL_2 = \sin^{-1}\sqrt{p_0} + k_2\sqrt{\frac{p_0(1-p_0)}{n}}} \tag{19}$$

$$\mathbf{UCL_1 = \sin^{-1}\sqrt{p_0} + k_1\sqrt{\frac{p_0(1-p_0)}{n}}} \tag{20}$$

The probability of acceptance under RGS scheme is

$$\mathbf{P_a = P(LCL_2 \le Z_t \le UCL_2)}$$
$$\mathbf{P(LCL_2 \le Z_t \le UCL_2) = 2.\Phi(k_2) - 1}$$

and

$$\mathbf{P_{rep} = P(LCL_1 \le Z_t \le LCL_2) + P(UCL_2 \le Z_t \le UCL_1)}$$
$$\mathbf{P(LCL_1 \le Z_t \le LCL_2) = \Phi(k_1) - \Phi(k_2)}$$
$$\mathbf{P(UCL_2 \le Z_t \le UCL_1) = \Phi(k_1) - \Phi(k_2)}$$

Thus

$$\mathbf{P_{in} = \frac{P_a}{1 - P_{rep}}}$$

The average run length for in control process is

$$\mathbf{ARL_0 = \frac{1}{1-P_{in}}} \tag{21}$$

For the shifted process we assume that our location parameter is shifted due to that the mean is shifted from $\mathbf{p_0}$ to $\mathbf{p_1 = cp_0}$ thus the shifted process has the mean and variance respectively as

$$\mathbf{E(Z_t) = \sin^{-1}\sqrt{p_1}} \tag{22}$$

$$\mathbf{Var(Z_t) = \frac{p_1(1-p_1)}{n}} \tag{23}$$

where $\mathbf{p_1 = cp_0}$

The probability of in control for shifted process is computed as

$$\mathbf{P_{a1} = P(LCL_2 \le Z_t \le UCL_2 | p = p_1)}$$

$$\mathbf{P(LCL_2 \le Z_t \le UCL_2 | p = p_1) = \Phi\left(\frac{UCL_2 - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) - \Phi\left(\frac{LCL_2 - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right)}$$

$$\mathbf{P(LCL_2 \le Z_t \le UCL_2 | p = p_1)} = \mathbf{\Phi}\left(\frac{\sin^{-1}\sqrt{p_0} + k_2\sqrt{\frac{p_0(1-p_0)}{n}} - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) - \mathbf{\Phi}\left(\frac{\sin^{-1}\sqrt{p_0} - k_2\sqrt{\frac{p_0(1-p_0)}{n}} - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right)$$

and

$$\mathbf{P_{rep1} = P(LCL_1 \le Z_t \le LCL_2 | p = p_1) + P(UCL_2 \le Z_t \le UCL_1 | p = p_1)}$$

$$\mathbf{P(LCL_1 \le Z_t \le LCL_2 | p = p_1)} = \mathbf{\Phi}\left(\frac{LCL_2 - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) - \mathbf{\Phi}\left(\frac{LCL_1 - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right)$$

$$\mathbf{P(LCL_1 \le Z_t \le LCL_2 | p = p_1)} = \mathbf{\Phi}\left(\frac{\sin^{-1}\sqrt{p_0} - k_2\sqrt{\frac{p_0(1-p_0)}{n}} - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right)$$

Also

$$\mathbf{P(UCL_2 \le Z_t \le UCL_1 | p = p_1)} = \mathbf{\Phi}\left(\frac{UCL_1 - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) - \mathbf{\Phi}\left(\frac{UCL_2 - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right)$$

$$\mathbf{P(UCL_2 \le Z_t \le UCL_1 | p = p_1)} = \mathbf{\Phi}\left(\frac{\sin^{-1}\sqrt{p_0} + k_1\sqrt{\frac{p_0(1-p_0)}{n}} - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) - \mathbf{\Phi}\left(\frac{\sin^{-1}\sqrt{p_0} + k_2\sqrt{\frac{p_0(1-p_0)}{n}} - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right)$$

Thus

$$\mathbf{P_{rep1}} = \Phi\left(\frac{\sin^{-1}\sqrt{p_0} - k_2\sqrt{\frac{p_0(1-p_0)}{n}} - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) - \Phi\left(\frac{\sin^{-1}\sqrt{p_0} - k_1\sqrt{\frac{p_0(1-p_0)}{n}} - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) + \Phi\left(\frac{\sin^{-1}\sqrt{p_0} + k_1\sqrt{\frac{p_0(1-p_0)}{n}} - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) - \Phi\left(\frac{\sin^{-1}\sqrt{p_0} + k_2\sqrt{\frac{p_0(1-p_0)}{n}} - \sin^{-1}\sqrt{p_1}}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right)$$

So the probability of in control for the shifted process is

$$\mathbf{P_{in1}} = \frac{\mathbf{P_{a1}}}{1 - \mathbf{P_{rep1}}}$$

Thus the average run length for the shifted process is

$$\mathbf{ARL_1} = \frac{1}{1 - \mathbf{P_{in1}}} \tag{24}$$

Table 4
Average Run Length Values of the NPAS Chart under RGS Scheme when $r_0 = 370$ and $p_0 = 0.50$

Shift	r_0=370; p_0=0.5; k_1=3.0088; k_2 = 2.1409					
	N					
	5	10	20	30	50	100
	Arls					
0.05	1.07	1.00	1.00	1.00	1.00	1.00
0.10	2.64	1.00	1.00	1.00	1.00	1.00
0.30	20.29	2.52	1.04	1.00	1.00	1.00
0.50	52.39	13.05	2.85	1.46	1.05	1.00
0.60	81.05	27.26	7.21	3.20	1.44	1.02
0.70	127.41	56.58	19.78	9.61	3.67	1.33
0.75	160.90	82.20	33.53	17.77	7.21	2.12
0.80	203.19	120.33	57.95	34.02	15.52	4.56
0.85	253.86	176.24	102.18	67.52	35.95	12.52
0.90	307.81	252.01	180.90	137.85	89.08	40.94
0.95	352.27	332.11	297.35	268.46	223.28	152.85
0.99	369.26	368.33	366.50	364.68	361.08	352.37
1.00	370.00	370.00	370.00	370.00	370.00	370.00
1.01	369.26	368.33	366.50	364.68	361.08	352.37
1.05	352.27	332.11	297.35	268.46	223.28	152.85
1.10	307.81	252.01	180.90	137.85	89.08	40.94
1.15	253.86	176.24	102.18	67.52	35.95	12.52
1.20	203.19	120.33	57.95	34.02	15.52	4.56
1.25	160.90	82.20	33.53	17.77	7.21	2.12
1.30	127.41	56.58	19.78	9.61	3.67	1.33
1.35	101.34	39.21	11.86	5.40	2.12	1.09
1.40	81.05	27.26	7.21	3.20	1.44	1.02
1.50	52.39	13.05	2.85	1.46	1.05	1.00
1.70	20.29	2.52	1.04	1.00	1.00	1.00
1.90	2.64	1.00	1.00	1.00	1.00	1.00
1.95	1.07	1.00	1.00	1.00	1.00	1.00

Table 5
Average Run Length Values of the NPAS Chart under RGS Scheme when $r_0 = 300$ and $p_0 = 0.50$

Shift	r_0=300; p_0=0.5; k_1=2.9461; k_2 = 2.0771					
	n					
	5	10	20	30	50	100
	Arls					
0.05	1.03	1.00	1.00	1.00	1.00	1.00
0.10	2.06	1.00	1.00	1.00	1.00	1.00
0.30	16.28	2.23	1.03	1.00	1.00	1.00
0.50	43.22	11.08	2.57	1.39	1.04	1.00
0.60	67.13	23.04	6.29	2.88	1.38	1.01
0.70	105.48	47.54	16.95	8.37	3.30	1.29
0.75	132.94	68.80	28.52	15.30	6.35	1.97
0.80	167.32	100.19	48.90	28.99	13.43	4.08
0.85	208.12	145.77	85.47	56.92	30.65	10.90
0.90	251.08	206.76	149.66	114.74	74.79	34.86
0.95	286.12	270.28	242.85	219.93	183.85	127.01
0.99	299.42	298.70	297.26	295.84	293.02	286.20
1.00	300.00	300.00	300.00	300.00	300.00	300.00
1.01	299.42	298.70	297.26	295.84	293.02	286.20
1.05	286.12	270.28	242.85	219.93	183.85	127.01
1.10	251.08	206.76	149.66	114.74	74.79	34.86
1.15	208.12	145.77	85.47	56.92	30.65	10.90
1.20	167.32	100.19	48.90	28.99	13.43	4.08
1.25	132.94	68.80	28.52	15.30	6.35	1.97
1.30	105.48	47.54	16.95	8.37	3.30	1.29
1.35	83.97	33.05	10.24	4.77	1.96	1.07
1.40	67.13	23.04	6.29	2.88	1.38	1.01
1.50	43.22	11.08	2.57	1.39	1.04	1.00
1.70	16.28	2.23	1.03	1.00	1.00	1.00
1.90	2.06	1.00	1.00	1.00	1.00	1.00
1.95	1.03	1.00	1.00	1.00	1.00	1.00

Table 6
Average Run Length Values of the NPAS chart under Single Sampling Scheme when $r_0 = 370$ and $p_0 = 0.50$

Shift	r_0=370; p_0=0.5; k=2.999673					
	n					
	5	10	20	30	50	100
	Arls					
0.05	3.79	1.00	1.00	1.00	1.00	1.00
0.10	7.83	1.12	1.00	1.00	1.00	1.00
0.30	26.21	4.51	1.34	1.04	1.00	1.00
0.50	57.63	16.57	4.47	2.28	1.26	1.00
0.60	85.86	31.31	9.71	4.81	2.17	1.13
0.70	131.58	60.91	23.22	12.27	5.32	1.92
0.75	164.59	86.48	37.38	21.01	9.51	3.20
0.80	206.22	124.33	62.06	37.80	18.56	6.37
0.85	256.06	179.59	106.22	71.62	39.71	15.29
0.90	309.03	254.21	184.15	141.59	93.12	44.75
0.95	352.63	332.87	298.76	270.38	225.92	156.41
0.99	369.27	368.37	366.57	364.78	361.26	352.72
1.00	370.00	370.00	370.00	370.00	370.00	370.00
1.01	369.27	368.37	366.57	364.78	361.26	352.72
1.05	352.63	332.87	298.76	270.38	225.92	156.41
1.10	309.03	254.21	184.15	141.59	93.12	44.75
1.15	256.06	179.59	106.22	71.62	39.71	15.29
1.20	206.22	124.33	62.06	37.80	18.56	6.37
1.25	164.59	86.48	37.38	21.01	9.51	3.20
1.30	131.58	60.91	23.22	12.27	5.32	1.92
1.35	105.87	43.44	14.83	7.51	3.25	1.37
1.40	85.86	31.31	9.71	4.81	2.17	1.13
1.50	57.63	16.57	4.47	2.28	1.26	1.00
1.70	26.21	4.51	1.34	1.04	1.00	1.00
1.90	7.83	1.12	1.00	1.00	1.00	1.00
1.95	3.79	1.00	1.00	1.00	1.00	1.00

The NPSS Control Chart:

The NPSS type control chart is developed by using the moving average of $\mathbf{M_t}$ as

$$\mathbf{Q_t = mean(M_t, M_{t-1})} \tag{25}$$

The mean and variance of the above estimator are

The mean and variance for $\mathbf{W_t}$ will be as

$$\mathbf{E(Q_t) = np_0} \tag{26}$$

$$\mathbf{Var(Q_t) = np_0(1 - p_0)/2} \tag{27}$$

The control limits for the RGS sampling scheme are

$$\mathbf{LCL_1 = np_0 - k_1\sqrt{np_0(1-p_0)/2}} \tag{28}$$

$$\mathbf{LCL_2 = np_0 - k_2\sqrt{np_0(1-p_0)/2}} \tag{29}$$

$$\mathbf{UCL_2 = np_0 + k_2\sqrt{np_0(1-p_0)/2}} \tag{30}$$

$$\mathbf{UCL_1 = np_0 + k_1\sqrt{np_0(1-p_0)/2}} \tag{31}$$

The probability of acceptance under RGS scheme is

$$\mathbf{P_a = P(LCL_2 \le Q_t \le UCL_2)}$$
$$\mathbf{P(LCL_2 \le Q_t \le UCL_2) = 2.\Phi(k_2) - 1}$$

and

$$\mathbf{P_{rep} = P(LCL_1 \le Q_t \le LCL_2) + P(UCL_2 \le Q_t \le UCL_1)}$$
$$\mathbf{P(LCL_1 \le Q_t \le LCL_2) = \Phi(k_1) - \Phi(k_2)}$$
$$\mathbf{P(UCL_2 \le Q_t \le UCL_1) = \Phi(k_1) - \Phi(k_2)}$$

Thus

$$\mathbf{P_{in} = \frac{P_a}{1 - P_{rep}}}$$

The average run length for in control process is

$$\mathbf{ARL_0 = \frac{1}{1 - P_{in}}} \tag{32}$$

For the shifted process we assume that our location parameter is shifted due to that the mean is shifted from $\mathbf{p_0}$ to $\mathbf{p_1 = cp_0}$ thus the shifted process has the mean and variance respectively as

$$\mathbf{E(Q_t) = np_1} \tag{33}$$

$$\mathbf{Var(Q_t) = np_1(1-p_1)/2} \tag{34}$$

where $\mathbf{p_1 = cp_0}$

The probability of in control for shifted process is computed as

$$\mathbf{P_{a1} = P(LCL_2 \le Q_t \le UCL_2 | p = p_1)}$$

$$\mathbf{P(LCL_2 \le Q_t \le UCL_2 | p = p_1) = \Phi\left(\frac{UCL_2 - np_1}{\sqrt{np_1(1-p_1)/2}}\right) - \Phi\left(\frac{LCL_2 - np_1}{\sqrt{np_1(1-p_1)/2}}\right)}$$

$$\mathbf{P(LCL_2 \le Q_t \le UCL_2 | p = p_1) = \Phi\left(\frac{np_0(1-c) + k_2\sqrt{np_0(1-p_0)/2}}{\sqrt{np_1(1-p_1)/2}}\right) - \Phi\left(\frac{np_0(1-c) - k_2\sqrt{np_0(1-p_0)/2}}{\sqrt{np_1(1-p_1)/2}}\right)}$$

and

$$\mathbf{P_{rep1} = P(LCL_1 \le Q_t \le LCL_2 | p = p_1) + P(UCL_2 \le Q_t \le UCL_1 | p = p_1)}$$

$$\mathbf{P(LCL_1 \leq Q_t \leq LCL_2 | p = p_1)} = \mathbf{\Phi\left(\frac{LCL_2 - np_1}{\sqrt{np_1(1-p_1)/2}}\right) - \Phi\left(\frac{LCL_1 - np_1}{\sqrt{np_1(1-p_1)/2}}\right)}$$

$$\mathbf{P(LCL_1 \leq Q_t \leq LCL_2 | p = p_1)} = \mathbf{\Phi\left(\frac{np_0(1-c) - k_2\sqrt{np_0(1-p_0)/2}}{\sqrt{np_1(1-p_1)/2}}\right) - \Phi\left(\frac{np_0(1-c) - k_1\sqrt{np_0(1-p_0)/2}}{\sqrt{np_1(1-p_1)/2}}\right)}$$

Also

$$\mathbf{P(UCL_2 \leq Q_t \leq UCL_1 | p = p_1)} = \mathbf{\Phi\left(\frac{UCL_2 - np_1}{\sqrt{np_1(1-p_1)/2}}\right) - \Phi\left(\frac{UCL_1 - np_1}{\sqrt{np_1(1-p_1)/2}}\right)}$$

$$\mathbf{P(UCL_2 \leq Q_t \leq UCL_1 | p = p_1)} = \mathbf{\Phi\left(\frac{np_0(1-c) + k_1\sqrt{np_0(1-p_0)/2}}{\sqrt{np_1(1-p_1)/2}}\right) - \Phi\left(\frac{np_0(1-c) + k_2\sqrt{np_0(1-p_0)/2}}{\sqrt{np_1(1-p_1)/2}}\right)}$$

Thus

$$\mathbf{P_{rep1}} = \mathbf{\Phi\left(\frac{np_0(1-c) - k_2\sqrt{np_0(1-p_0)/2}}{\sqrt{np_1(1-p_1)/2}}\right) - \Phi\left(\frac{np_0(1-c) - k_1\sqrt{\frac{np_0(1-p_0)}{2}}}{\sqrt{\frac{np_1(1-p_1)}{2}}}\right) + \Phi\left(\frac{np_0(1-c) + k_1\sqrt{np_0(1-p_0)/2}}{\sqrt{np_1(1-p_1)/2}}\right) - \Phi\left(\frac{np_0(1-c) + k_2\sqrt{np_0(1-p_0)/2}}{\sqrt{np_1(1-p_1)/2}}\right)}$$

Thus the probability of in-control process for shifted process is

$$\mathbf{P_{in1}} = \mathbf{\frac{P_{a1}}{1 - P_{rep1}}}$$

Thus the average run length for the shifted process is

$$\mathbf{ARL_1} = \mathbf{\frac{1}{1 - P_{in1}}} \tag{36}$$

Table 7
Average Run Length Values of the NPSS Chart under RGS scheme when $r_0 = 370$ and $p_0 = 0.50$

Shift	r_0=370; p_0=0.5; k_1=3.0081; k_2 = 2.1695					
	n					
	5	10	20	30	50	100
	Arls					
0.05	1.01	1.00	1.00	1.00	1.00	1.00
0.10	1.17	1.00	1.00	1.00	1.00	1.00
0.30	4.58	1.16	1.00	1.00	1.00	1.00
0.50	16.12	3.52	1.22	1.03	1.00	1.00
0.60	30.32	8.21	2.15	1.29	1.03	1.00
0.70	59.21	21.02	6.01	2.84	1.39	1.02
0.75	84.44	34.79	11.36	5.43	2.22	1.11
0.80	122.03	59.13	22.85	11.79	4.75	1.59
0.85	177.30	103.13	48.74	28.33	12.81	3.81
0.90	252.46	181.45	109.82	74.66	41.32	15.21
0.95	332.19	297.49	244.31	205.57	153.15	87.47
0.99	368.34	366.50	362.88	359.32	352.39	336.05
1.00	370.00	370.00	370.00	370.00	370.00	370.00
1.01	368.34	366.50	362.88	359.32	352.39	336.05
1.05	332.19	297.49	244.31	205.57	153.15	87.47
1.10	252.46	181.45	109.82	74.66	41.32	15.21
1.15	177.30	103.13	48.74	28.33	12.81	3.81
1.20	122.03	59.13	22.85	11.79	4.75	1.59
1.25	84.44	34.79	11.36	5.43	2.22	1.11
1.30	59.21	21.02	6.01	2.84	1.39	1.02
1.35	42.11	12.99	3.43	1.75	1.11	1.00
1.40	30.32	8.21	2.15	1.29	1.03	1.00
1.50	16.12	3.52	1.22	1.03	1.00	1.00
1.70	4.58	1.16	1.00	1.00	1.00	1.00
1.90	1.17	1.00	1.00	1.00	1.00	1.00
1.95	1.01	1.00	1.00	1.00	1.00	1.00

Table 8
Average Run Length Values of the NPSS Chart under RGS Scheme when $r_0 = 300$ and $p_0 = 0.50$

Shift	r_0=300; p_0=0.5; k_1=2.9461; k_2= 2.0771					
	n					
	5	10	20	30	50	100
	Arls					
0.05	1.00	1.00	1.00	1.00	1.00	1.00
0.10	1.12	1.00	1.00	1.00	1.00	1.00
0.30	3.94	1.13	1.00	1.00	1.00	1.00
0.50	13.73	3.17	1.19	1.03	1.00	1.00
0.60	25.70	7.20	2.01	1.25	1.02	1.00
0.70	49.86	18.09	5.36	2.62	1.34	1.01
0.75	70.78	29.69	9.95	4.87	2.08	1.10
0.80	101.72	50.02	19.72	10.34	4.28	1.52
0.85	146.76	86.39	41.46	24.37	11.22	3.47
0.90	207.20	150.23	91.96	63.02	35.30	13.28
0.95	270.37	243.02	200.77	169.73	127.37	73.65
0.99	298.70	297.27	294.44	291.66	286.23	273.42
1.00	300.00	300.00	300.00	300.00	300.00	300.00
1.01	298.70	297.27	294.44	291.66	286.23	273.42
1.05	270.37	243.02	200.77	169.73	127.37	73.65
1.10	207.20	150.23	91.96	63.02	35.30	13.28
1.15	146.76	86.39	41.46	24.37	11.22	3.47
1.20	101.72	50.02	19.72	10.34	4.28	1.52
1.25	70.78	29.69	9.95	4.87	2.08	1.10
1.30	49.86	18.09	5.36	2.62	1.34	1.01
1.35	35.59	11.28	3.12	1.66	1.10	1.00
1.40	25.70	7.20	2.01	1.25	1.02	1.00
1.50	13.73	3.17	1.19	1.03	1.00	1.00
1.70	3.94	1.13	1.00	1.00	1.00	1.00
1.90	1.12	1.00	1.00	1.00	1.00	1.00
1.95	1.00	1.00	1.00	1.00	1.00	1.00

Table 9
Average Run Length Values of the NPSS Chart under Single Sampling Scheme when $r_0 = 370$ and $p_0 = 0.50$

Shift	r_0=370; p_0=0.5; k=2.999673					
	n					
	5	10	20	30	50	100
	Arls					
0.05	1.98	1.00	1.00	1.00	1.00	1.00
0.10	2.76	1.01	1.00	1.00	1.00	1.00
0.30	7.38	1.75	1.02	1.00	1.00	1.00
0.50	19.72	5.29	1.74	1.19	1.01	1.00
0.60	34.25	10.73	3.29	1.84	1.16	1.00
0.70	63.31	24.33	8.07	4.18	2.00	1.11
0.75	88.47	38.47	14.00	7.35	3.30	1.41
0.80	125.78	63.03	26.09	14.41	6.50	2.32
0.85	180.44	106.94	52.49	31.71	15.47	5.34
0.90	254.52	184.50	113.54	78.49	44.93	18.01
0.95	332.90	298.81	246.49	208.30	156.50	91.27
0.99	368.37	366.57	363.02	359.53	352.73	336.69
1.00	370.00	370.00	370.00	370.00	370.00	370.00
1.01	368.37	366.57	363.02	359.53	352.73	336.69
1.05	332.90	298.81	246.49	208.30	156.50	91.27
1.10	254.52	184.50	113.54	78.49	44.93	18.01
1.15	180.44	106.94	52.49	31.71	15.47	5.34
1.20	125.78	63.03	26.09	14.41	6.50	2.32
1.25	88.47	38.47	14.00	7.35	3.30	1.41
1.30	63.31	24.33	8.07	4.18	2.00	1.11
1.35	46.16	15.91	4.98	2.64	1.42	1.02
1.40	34.25	10.73	3.29	1.84	1.16	1.00
1.50	19.72	5.29	1.74	1.19	1.01	1.00
1.70	7.38	1.75	1.02	1.00	1.00	1.00
1.90	2.76	1.01	1.00	1.00	1.00	1.00
1.95	1.98	1.00	1.00	1.00	1.00	1.00

3. THE PERFORMANCE EVALUATION OF THE PROPOSED CONTROL CHART

The performance of control chart is judged by calculating the widely and commonly accepted technique of the ARL. The ARL may be defined as the average number of samples before the process signals an out-of-control observation (Z. H. Li, Zou, Gong, &

Wang, 2014). Crowder (1987) presented a simple method for the calculation of the ARL and the standard deviation of the ARL for the EWMA control charts. Usually the ARL values of the in-control process should be larger which are denoted by ARL0 and the performance of the chart is evaluated for the smaller number of the out-of-control process, which is denoted by ARL1. The ARL0 and the ARL1 equations have been derived for the proposed RGS scheme control chart in Eq. (9) and (13) for the NPS chart, Eq. (21) and (24) for the NPAS chart and Eq. (33) and (36) for the NPSS chart. Using these equations ARL values have been estimated for 10, 000 replications under the explained criteria for Monte Carlo simulation technique using R-language program. A pair of the control chart coefficients k1 and k2 have been estimated for different chart settings. Without losing the similarity the ARL0 and the ARL1 values have been estimated for the existing control charts of NPS, NPAS and NPSS single sampling scheme charts.

4. COMPARISON OF THE PERFORMANCE OF PROPOSED WITH THE EXISTING CHARTS

In this section ARL1 comparison of the proposed RGS scheme with the single sampling scheme of NPS, NPAS and NPSS charts is discussed. Tables 1 through 9 are generated for the ARL values for different possible settings of the processes. These tables have been generated for the process setting of RGS parameter, sample size $\mathbf{n = 20}$, ARL0 =370 and 300, p0 = 0.05 under different shift levels ranging from 0.05 to 1.95. Table 1 and 2 describe the ARL values for proposed RGS scheme NPS chart for the process setting of ARL0 = 370 and 300, p0 = 0.05. Table 4 and 5 describe the ARL values for proposed RGS scheme NPAS chart for the process setting of ARL0 =370 and 300, p0 = 0.05. Table 7 and 8 describe the ARL values for proposed RGS scheme NPSS chart for the process setting of ARL0 =370 and 300, p0 = 0.05. Table 3, 6 and 9 describe the ARL values for single sampling scheme for NPS chart, NPAS chart and NPSS chart respectively for the process setting of ARL0 = 370 and p0 = 0.05.

The better chart is considered as one which shows the smaller ARL1 values, means the ability to quickly detect the out-of-control process. This can be observed in all the generated tables from 1 through 9. For example, in Table 1 the ARL1 value is estimated as 10 for a shift of 1.05 for the proposed chart while from Table 3 this value is 18 for the same settings. The same pattern of performance can be observed in other tables.

5. AN APPLICATION OF PROPOSED CHART THROUGH EXAMPLES

In this section we describe the application of the proposed scheme by using the simulated data. The first 20 observations are generated from in-control process assuming n = 20 and p0 = 0.5, next 25 observations are generated from the shifted process p1 = c*p0 where c = 1.1 from the Table 1 ARL1 is 57 the shift is not detected by the single sampling schemes whereas the shift is detected for the NPS chart, NPAS chart and NPSS control chart. The posting of estimators for these proposed control charts are presented in Fig. 1 through Fig. 3. It can be observed that none of these control charts are able to detect the shift of the production process whereas the proposed control charts show the comparatively better edge for quickly detecting the out-of-control process.

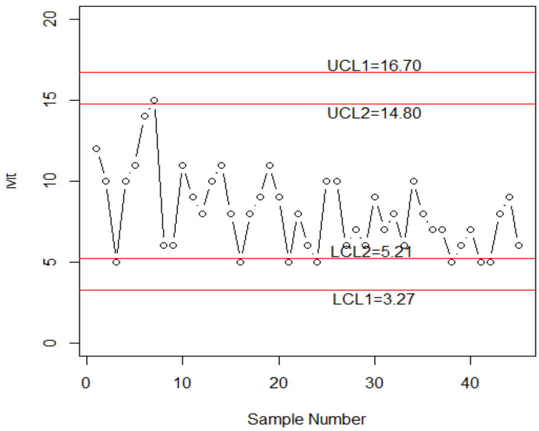


Fig. 1: Control Chart for the Proposed RGS Scheme for the NPS

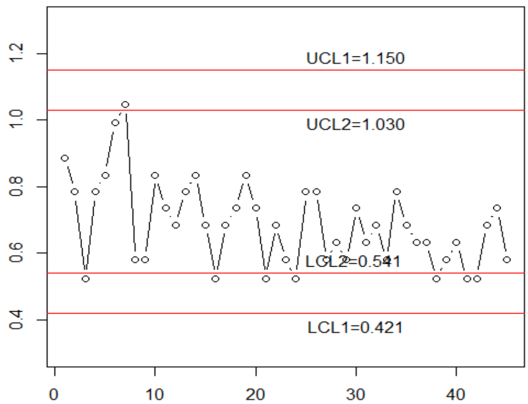


Fig. 2: Control Chart for the Proposed RGS Scheme for the NPAS

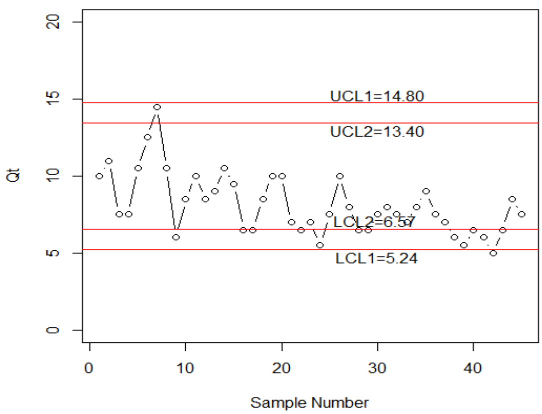


Fig. 3: Control Chart for the Proposed RGS Scheme for the NPSS

6. COMMENTS AND CONCLUSION

In this paper a RGS scheme has been introduced in the control chart literature for the distribution-free data. The performance of the proposed control chart is studied and compared with the existing control chart schemes of NPS chart, NPAS chart and the NPSS chart. The control chart coefficients of paired limits have been estimated for different process settings. It has been observed that the proposed design of monitoring the process is efficient in detecting the out-of-control process quickly. An example has been presented to explain the use of the proposed scheme.

REFERENCES

1. Abbasi, S.A., Miller, A. and Riaz, M. (2013). Nonparametric Progressive Mean Control Chart for Monitoring Process Target. *Quality and Reliability Engineering International,* 29(7), 1069-1080.
2. Ahmad, L., Aslam, M. and Azam, M. (2015). Developing variables repetitive group sampling control chart using regression estimator. *Proceeding of 13th International Conference on Statistical Sciences*, Peshawar, Pakistan, 129-136.
3. Ahmad, L., Aslam, M. and Azhar, H. (2015). Repetitive Group Acceptance Sampling Plan for Exponentiated Weibull Distribution. *Proceeding of 13th International Conference on Statistical Sciences*, Peshawar, Pakistan, 167-178.
4. Ahmad, L., Aslam, M. and Jun, C.-H. (2014). Designing of X-bar control charts based on process capability index using repetitive sampling. *Transactions of the Institute of Measurement and Control,* 36(3), 367-374.
5. Ahmad, L., Aslam, M. and Jun, C.-H. (2015). The Design of a new Repetitive sampling Control Chart Based on Process Capability Index. [first published on February 23, 2015]. *Transactions of the Institute of Measurement and Control*. DOI: 10.1177/0142331215571120.
6. Aslam, M., Azam, M. and Jun, C.-H. (2013). A mixed repetitive sampling plan based on process capability index. *Applied Mathematical Modelling,* 37(24), 10027-10035.
7. Aslam, M., Lio, Y.L. and Jun, C.-H. (2013). Repetitive acceptance sampling plans for burr type XII percentiles. *The International Journal of Advanced Manufacturing Technology,* 68(1), 495-507.
8. Aslam, M., Niaki, S.T.A., Rasool, M. and Fallahnezhad, M.S. (2012). Decision rule of repetitive acceptance sampling plans assuring percentile life. *Scientia Iranica. Transaction E, Industrial Engineering* 19(3), 879-884.
9. Aslam, M., Wu, C.W., Jun, C.H., Azam, M. and Itay, N. (2013). Developing a variables repetitive group sampling plan based on process capability index Cpk with unknown mean and variance. *Journal of Statistical Computation and Simulation,* 83(8), 1507-1517.
10. Aslam, M., Yen, C.H. and Jun, C.H. (2011). Variable repetitive group sampling plans with process loss consideration. *Journal of Statistical Computation and Simulation,* 81(11), 1417-1432.
11. Balamurali, S. and Jun, C.H. (2006). Repetitive Group Sampling Procedure for Variables Inspection. *Journal of Applied Statistics,* 33(3), 327-338.

12. Balamurali, S., Park, H., Jun, C.H., Kim, K.J. and Lee, J. (2005). Designing of Variables Repetitive Group Sampling Plan Involving Minimum Average Sample Number. *Communications in Statistics-Simulation and Computation,* 34(3), 799-809.
13. Crowder, S.V. (1987). A simple method for studying run-length distribution of exponentially weighted moving average charts. *Technometrics*, 29, 401-407.
14. Das, N. (2008). A Note on the Efficiency of Nonparametric Control Chart for Monitoring Process Variability. *Economic Quality Control,* 23(1), 85-93.
15. Das, N. (2009). A comparison study of three non-parametric control charts to detect shift in location parameters. *The International Journal of Advanced Manufacturing Technology*, 41(7-8), 799-807.
16. Graham, M.A., Chakraborti, S. and Human, S.W. (2011). A Nonparametric EWMA Sign Chart for Location Based on Individual Measurements. *Quality Engineering*, 23(3), 227-241.
17. Lee, S.-H., Aslam, M. and Jun, C.-H. (2012). Repetitive Group Sampling Plans With Double Specification Limits. *Pak. J. Statist.*, 28(1), 41-57.
18. Li, S.-Y., Tang, L.-C. and Ng, S.-H. (2010). Nonparametric CUSUM and EWMA control charts for detecting mean shifts. *Journal of Quality Technology,* 42(2), 209.
19. Li, Z.H., Zou, C.L., Gong, Z. and Wang, Z.J. (2014). The computation of average run length and average time to signal: an overview. *Journal of Statistical Computation and Simulation*, 84(8), 1779-1802.
20. Montgomery, D.C. (2013). *Design and Analysis of Experiments*: John Wiley & Sons, Inc.
21. Mukherjee, A. and Chakraborti, S. (2012). A Distribution-free Control Chart for the Joint Monitoring of Location and Scale. *Quality and Reliability Engineering International,* 28(3), 335-352.
22. Mukherjee, A., Chakraborti, S. and Graham, M.A. (2012). Distribution-free exponentially weighted moving average control charts for monitoring unknown location. *Computational Statistics & Data Analysis,* 56(8), 2539-2561.
23. Sherman, R.E. (1965). Design and Evaluation of a Repetitive Group Sampling Plan. *Technometrics,* 7(1), 11-21.
24. Stoumbos, Z.G. B. and Reynolds, M.R. (2000). Robustness to non-normality and autocorrelation of individuals control charts. *Journal of Statistical Computation and Simulation,* 66(2), 145-187.
25. Su-Fen, Y., Wen-Chi, T., Tzee-Ming, H., Chi-Chin, Y. and Smiley, C. (2011). Monitoring process mean with a new EWMA control chart. *Production*, 21(2), 217-222.
26. Yang, S.-F., Lin, J.-S. and Cheng, S.W. (2011). A new nonparametric EWMA sign control chart. *Expert Systems with Applications,* 38(5), 6239-6243.
27. Yang, S.F. and Cheng, S.W. (2011). A new non-parametric CUSUM mean chart. *Quality and Reliability Engineering International*, 27(7), 867-875.
28. Yen, C.-H., Chang, C.-H. and Aslam, M. (2015). Repetitive variable acceptance sampling plan for one-sided specification. *Journal of Statistical Computation and Simulation,* 85(6), 1102-1116.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 279-284

FREQUENCY AND AWARENESS OF THALASSEMIA IN FAMILIES WITH COUSIN MARRIAGES: A STUDY FROM KARACHI, PAKISTAN

Faizan-ul-Haq, Muhammad Muizzuddin§, Muhammad Mannan Ali Khan, Sundas Sajid, Asma Sarfaraz, Namerah Nasir, Aneeba Nazim, Ghuncha Kamran, Bushra Maqsood and **Amna Ahmed**
Dow Medical College, Dow University of Health Sciences, Karachi, Pakistan
§Email: muhammadmuizzuddin0@gmail.com

ABSTRACT

Objectives:

- To assess the frequency of thalassemia in families with cousin marriages.
- To assess the awareness of thalassemia in affected parents.

Methodology: This cross-sectional study was conducted at Afzal memorial thalassemia foundation, Omair Sana Foundation and Husaaini Blood Bank, Karachi from January to December 2015. A total of 100 sample size were recruited. All thalassemic people having consanguineous marriages were included in the study... An informed verbal consent was taken from the parents. A pilot study was data was entered and analyzed using SPSS version 20 with frequencies, percentages, confidence interval and p value of 0.05 was considered as statistically significant.

Result: 100 questionnaires were filled by the parents of thalasemic patients. (48% males and 52% females patients). The study group includes patients from mean age group of 10 years. According to data collected 56% ($p = 0.00$) patients had the prevalence of thalassemia in their families with 28% ($p = 0.00$) had deaths in their families due to thalassemia. 92% of parents with cousin marriages had no awareness about thalassemia before their thalassemic child and 63% ($p = 0.00$) were not even aware about treatment for thalassemia. Regarding screening test for thalassemia 62% ($p = 0.021$) were aware of it and 98% ($p = 0.00$) agreed to go through this before marriage/during pregnancy. When they were asked if they would have aborted this child if they would have known about the thalassemic child 66% ($p = 0.00$) parents said yes. If patients siblings get neglected or not 61% (0.036) parents denied the neglection of their other children. Thalassemia is one lifelong disease 91% ($p = 0.00$) were aware of it.

Conclusion: The study concludes that cousin marriage among families who possess thalassemic trait is one of the major causes of thalassemia. Awareness of parents of the patients was found inadequate.

KEY WORDS

Thalassemia, cousin marriages, awareness, screening test, blood, transfusion, disability, genetics.

1. INTRODUCTION

Thalassemia is one of the most common inherited diseases in Pakistan. According to few experts approximately 5000 children are diagnosed with this disease every year and according to some sources there are more than 50,000 thalassemic patients registered in treatment centres all over the country. Thalessemia is an inherited haemoglobinopathy that affects people throughout the world. It is an autosomal recessive disorder. It occurs most often among the people of Greek, Italian, Southern Asia and African descent. Thalessemia is a clinically heterogenous disorder resulting from different genetic lesions that variably impair globin chain synthesis. Thalessemic patients have defect in their red blood cells that affect the ability of these cells to produce the normal hemoglobin. Hemoglobin is a protein that carries oxygen to tissues. In this disorder, RBCs production is normal but the size of RBCs are small and they are more fragile. So, this causes a chronic lifelong microcytic anemia and this is the reason, these patients are frequently treated with transfusions.

Consanguinity is a deeply rooted social trend and cousin marriage is one of the major cause in increasing the number of patients presenting with thalessemia but unfortunately in our society in spite of knowing the consequences of cousin marriages they don’t even go for screening before marriage. Consanguinity-associated deaths were consistently higher in the neonatal, infant and childhood periods. If two partners who are carriers of thalassemia get married, based on the probability, their children will have severe thalassemia (25%), will be healthy (25%) and be carriers of thalassemia (50%).[7] Awareness and knowledge among affected families plays an important role to change the attitude of people towards family extension and reduce the chances of Thallesemia to run in family. There is no known cure but prevention is possible, practicable and successfully achieved in some countries.

2. METHODOLOGY

This cross-sectional study was conducted at Afzal memorial thalassemia foundation, Omair Sana Foundation and Husaaini Blood Bank, Karachi from January to December 2015. A total of 100 sample size were recruited in this study design. All thalassemic people having consanguineous marriages were included in the study. Parents of patients having other blood disorders like hereditary spherocytosis, sickle cell disease and others were excluded. The information was collected by interviewing thalassemic parents or their close relatives. An informed verbal consent was taken from the parents before the start of interview. The collected information includes diagnosis i.e. thalassemia, age and method of diagnosis, along with family history of the disease. A pilot study was conducted to assess the authenticity of questionnaire. Structured questionnaire was devised to collect the data from patients and then data was entered and analyzed using SPSS version 16 with frequencies, percentages, confidence interval and p value of 0.05 was considered as statistically significant.

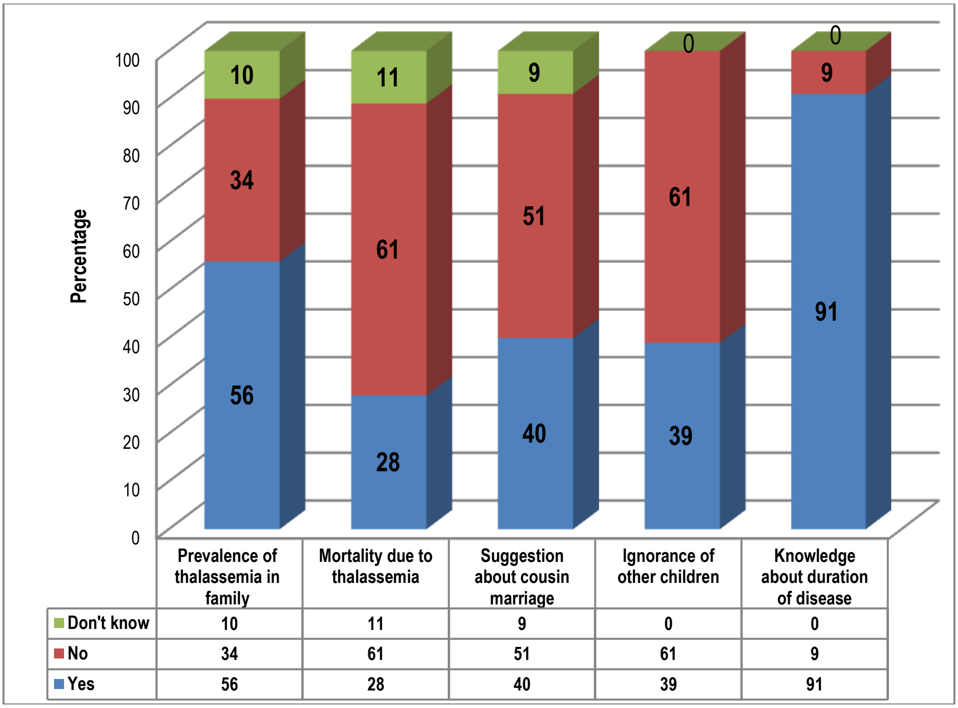
3. RESULT

In the study there was a slight preponderance of female which accounted for 52% of the patients and males 48%. The study group includes patients from mean age group 10 years. According to data collected, majority of the patients were from East and Central region, Father of 51% of the patients were labourers, 12% were drivers and 16% were jobless. When the parents of the patients were asked about the reason of their marriage in cousins, 82% (p-value 0.00) gave the reason of traditional values, while in 5% cases the marriage in cousins was by force. 50% (p-value 0.00) of parents were thalassemia minor, 6% were thalassemia major, while 44% doesn't know about their thalassemic status. 51% and 48% of patients got diagnosed through doctors and medical reports respectively. 56% (p = 0.00) patients had positive extended family history of thalassemia in their families while 28% (p = 0.00) had deaths in their families due to thalassemia.

92% of parents with cousin marriages had no awareness about thalassemia before birth of their thalassemic child. 92% (p-value 0.00) of parents first came to know about the disease after having thalassemic child, while 4% were aware of it before marriage and 4% came to know about it after marriage.

92% (p –value 0.00) of parents got awareness through Doctor/Nurse and 8% got awareness through TV. In 51% of subjects both, husband and wife had thalassemia minor, while 44% subjects were unaware of their thalassemia status. In our study 63% (p = 0.012) of parents of the patients were aware about treatment for thalassemia. Regarding screening test for thalassemia 62% (p = 0.021) of parents were aware of it and 98% (p = 0.00) agreed to go through this before marriage/during pregnancy. When they were asked if they would have aborted this child if they would have known about the thalassemic child 66% (p = 0.00) parents said yes, 23% of parents said No and 11% of parents gave no suggestion about it. When the parents were questioned about if there other off springs get neglected because of the diseased child or not, 61% (0.036) parents denied the neglecting of their other children. On the question about duration of disease, 91% (p = 0.00) were aware of that thalassemia is one lifelong disease. 51% (p – value 0.00) of the parents were against cousin marriages now, 40% were still in favor of cousin marriages in their next generation, while 9% gave no suggestion.

50% (P-value 0.00) of patients visit the transfusion centre twice a month while 36% visit once a month. In 80% (p-value 0.00)of cases total expenditure per month was found to be between Rs. 3000 to 5000, which is a huge burden for poor.



4. DISCUSSION

Thalassemia is one of the most common inherited diseases in Pakistan. There are about 80,000 children in Pakistan suffering from thalassemia major, the most common fatal genetic disease in the world and additional over 5,000 children with thalassemia born annually which is considered to be a very high rate for any inherited disorder and according to some sources there are more than 50,000 thalassemic patients registered in treatment centres all over the country.

Thalessemia is an inherited haemoglobinopathy that affects people throughout the world. It is an auto somal recessive disorder of blood. Consanguineous marriage in those families who possess thalassemiac trait is thought to be one the main causes of spread of thalassemia into the off spings. Marriages between first cousins and close family members is a largely ignored subject and one which many do not want to face or even talk about.

According to the research this indicates that there is a more than a 50 percent higher risk of inherited disorders like thalassaemia, deafness, blindness, mental disorders, diabetes and muscular disorders in children born to first cousins. Though more common in first-cousin partners, marriages within the same community also carry a higher risk, more than 80 percent of all parents in Pakistan are first cousins, seven percent are related by blood, about six percent belong to the same caste and only about four percent marry outsiders. In many countries, first-cousin marriages are forbidden and children of these marriages stand at 20 percent risk of having genetic problems.

Another important article was published by Rebecca Le fort in The Daily Telegraph on Feb 12. She carried out research in the Pakistani community in the United Kingdom. She found that more than 700 children were born with genetic diseases every year to parents who were first cousins. These children were ten times more likely to inherit disorders than the general population. These included infant mortality, birth defects, learning difficulties, blindness, hearing problems and metabolic disorders. Unfortunately, despite mounting evidence, the Pakistani community was not willing to acknowledge the risk, take preventative measures or, sometimes, even discuss the issue.

Pakistani people have a strong cultural inclination towards consanguineous marriage. Consanguineous relatives are defined within various degrees, according to the likelihood of their sharing genetic predisposition. The probability of homozygosity rate in consanguineous marriages is higher than that of unrelated parents; may be because of the limited gene pool and thus more expression of recessive alleles.

Thus, keeping in view the whole study it can be deduced that thalassemia is a hereditary lifetime disease and it can have lethal complications. However, despite knowing all of the major risks many people still continue practicing consanguineous marriages without screening, following the tradition running in family. The need to control its utmost awareness and strict imposition of laws related to screening before marriages should be carried in each individual.

5. COMMENTS AND CONCLUSION

The study concludes that cousin marriage among families who possess thalassemic trait is one of the causes of thalassemia. In our study it is appreciated that majority of the patients had a trend of cousin marriages in their families, most of the parents were not aware about the disease before having a thalassemic child, majority of the parents knew about screening test of thalassemia and they also had a positive opinion about doing the test before marriage. It is also found that most of the parents opted to abort the pregnancy when they were asked about what will they do if thalassemia is detected in the fetus. The level of awareness of parents of thalessemic patients was found inadequate, general public, parents and families of the patients needs to be educated in this regard. And means of electronic, social and print media can be used to spread the awareness in every corner of the country which is the demand and necessity of the time, so our forthcoming generation can be saved.

6. ACKNOWLEDGEMENT

I would specially like to mention those names who shared a big hand in the completion of our research work, these are Prof. Nazeer Khan, Ifrah Tanweer, Bisma Noor, Fatima Akhlaq, Mahzareen Ishrat, Imtiaz (Omair Sana Foundation), Management of Afzal Memorial Foundation.

REFERENCES

1. Hussaini, R.1. (1998). The impact of consanguinity and inbreeding on perinatal mortality in Karachi, Pakistan. *PPE* (9805711), 12(4), 370-82. http://www.ncbi.nlm.nih.gov/pubmed/9805711
2. Muhammad, A., Abdul, A., Abdul, Q. and Muhammad, K. (2010). Beta thalassemia major: The Moroccan experience. *RPH&E* (2141-2316), 2(2), 25-28. http://www.academicjournals.org/article/article1379342646_Agouzal%20et%20al.pd
3. Mausumi, B. (2015). A Study on Knowledge, Attitude and Practice about Thalassemia among General Population in Outpatient Department at a Tertiary Care Hospital of Kolkata. *JPM&HH*, 1(1), 5-12.
4. Ishaq, F. Abid, H., Kokab, F., Akhtar, A. and Mahmood, S. (2012). Awareness Among Parents of β-Thalassemia Major Patients, Regarding Prenatal Diagnosis and Premarital Screening. *JCP&S*. 22(4), 218-221. http://jcpsp.pk/archive/2012/Apr2012/05.pdf
5. Tazeen, M., Adil A.M., Ujala, N., Safwan, R.M. and Jovaria, M. (2013). Frequency of β-Thalassemia Trait in Families of Thalassemia Major Patients, Lahore. *JAMCA*, 25(3-4), 58-60. http://www.ayubmed.edu.pk/JAMC/25-3/Tazeen.pdf
6. Bushra, M., Tahira, I., Amer, J. and Faqir, M. (2013). Effect of β-Thalassemia on Hematological and Biochemical Profiles of Female Patients. *PJL&SS*. (1727-4915). 11(1), 25-28. http://www.pjlss.edu.pk/pdf_files/2013_1/25-28.pdf
7. Denic, S., Aden, B., Nagelkerle, N. and Essa, A.A. (2013). β-Thalassemia in Abu Dhabi: consanguinity and tribal stratification are major factors explaining the high prevalence of the disease. Hemoglobin. 37(4), 351-8. http://www.ncbi.nlm.nih.gov/pubmed/23600619
8. Zafar, Z., Qamruz, Z., Basharat, H., Saima. W. and Muhammad, I. (2015). Knowledge, attitude and practices with relevance to thalassemia. *J. Med. Sci.*, (Peshawar, Print) 23(2), 109-112. www.jmedsci.com/admin/uploadpic/JMS-15-April2015-Vol23No2.pdf

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 285-288

IMPACT OF CORPORATE SOCIAL RESPONSIBILITY ON PROFITABILITY OF ISLAMIC BANKING

Rashda Qazi and **Nadeem Iqbal**
Ghazi University, D.G. Khan, Pakistan
Email: rashida_qazi@yahoo.com; drnadeemiqbal1@gmail.com

ABSTRACT

The purpose of this paper is to compare the level, impact of corporate social responsibility (CSR) among Islamic and Conventional Banking system in relationship with earning per share (EPS), return on asset (ROA), return on equity (ROE). An empirical analysis is conducted, based on the annual reports of 4 banks for the year 2010-2011. To verify the relationship between EPS, ROA, ROE and CSR regression models are used. The results shows that there is lack of CSR in Pakistan and the regression model shows that there is positive relationship between profitability (EPS, ROA, ROE) and CSR practices.

1. INTRODUCTION

Today's organizations face different issues due to technological development and globalization in order to deal with uncertain situation in everyday life. At start of 21st century, there are many corporate scandals faced by corporations such as Enron case and world com. Due to these, large enterprises and different agencies are not stable. Furthermore, if there will be enhanced productivity of machinery and automation, it leads to environmental wipeout and increase in resources consumption, which give more attention to corporate social responsibility. Corporate social responsibility (CSR) refers to the ethical conduct of a corporation, which is responsible for the general stake holders, and not just its stockholders. The concept of CSR arose in the 1970s, but it has come to maturity this century. The idea of CSR in the U.S. and European countries is becoming matured. It has been considered to improve corporate image& value. Given the strong high correlation between all aspects of business management and corporate profit, corporations are becoming increasingly active in gaining consumers' trust and making a lasting impression. Given these inter-relationships, corporations that implement CSR are better tool to establish a positive image.

The success of a company highly dependent on its relationships with its key stakeholders and their satisfaction level (Elkington (1995), showing "good corporate governance" has become critical. Aware of and interested in companies" CSR efforts, consumers weigh them in their decision-making process (Arli and Lasmono, 2009). They are more likely to provide greater support to companies that engage in social and environmental causes and reward them by purchasing their products (Olsen and Hill, 2006). CSR is therefore most effective at improving 7g consumers "attitude towards the company, enhancing consumer loyalty as well as downsizing the level of consumer skepticism, i.e., reducing consumers" concerns and doubts regarding the company's products and services (Pirsch et al., 2007).

In today's business environment, competition is no longer solely limited to manufacturing better and higher quality goods. There are many other critical factors that will determine a company's success and make it stand out. One such factor is corporate social responsibility. As a number of academics have argued, maintaining and enhancing corporate social responsibility can have a strong impact on customer value and customer loyalty and create a differential competitive advantage (e.g. Jeremy Galbreath, 2008; Maohua Li, 2009). A company reputation is closely tied up to the emotional belief of its various stakeholders (Chang, 2006). In the case of Thailand, as with most developing countries, CSR practice is most appreciated both for its social and environmental impacts, particularly programs promoting education, and as a tool designed to increase corporate transparency toward all the concerned stakeholders.

2. LITERATURE REVIEW

Many researchers have worked over the corporate social responsibility. Corporate social responsibility explains the ethical responsibility or obligation of the corporation towards all stakeholders. The link between CSR, profitability of the firm and growth of the corporation has been studied in depth by numerous researchers but all has contradiction in their opinions. Some of the researchers said that there is a positive link between CSR, firms' financial performance and long term firms' growth and other explained the negative link between CSR and short term profitability of the firm.

Lin et al. (2009) examined the relationship between CSR and firms' profitability in the context of Taiwan. Researchers have taken 1000 Taiwan companies from 2002 to 2004 in order to run regression analysis. Researchers have concluded that there is a positive link between CSR and short term financial performance but it is not significant. They have also concluded that there is no fiscal advantage to firms in investing money for the purpose of CSR.

Malik and Nadeem (2014) studied the relationship between CSR and Firms' financial performance of Banks in Pakistan. Data was collected from different financial reports of 2008-2012 and regression model was used to show the impact of CSR n EPS, ROA and ROE of banks. Researchers have concluded that the foresaid variables have significant relationship among them which means as companies prefer CSR practices; they earn more profit than other banks. Fombrun and Shanley (1990) determined that a firm's CSR Practices attract investors and better qualified personnel, lowers the cost of capital and enhances its competitive ability. Herremans et al. (1993) found that companies with better CSR outperform those with poorer reputations and provide investors with higher stock market returns. Firms with good reputation also command higher prices, generate more employee loyalty and greater productivity, have bargaining power with their suppliers, more stable revenues, and are less exposed to crises (Fombrun, 1996).

The conceptual framework of this research has been adapted from Carroll's (1979/1991) Bhattacharya and Sen's (2004), and Oskamp's (1997) studies. The purpose of this study is to determine the influence of SCG's various CSR programs (economic, legal, ethical, and philanthropic) on corporate financial performance. It is based on the assumption that stakeholders will almost always favor a company that acts responsibly and has set up operational policies beneficial to the local community and society at large. It also assumes that, as a result, these stakeholders will feel strongly and positively about the company.

3. METHODOLOGY AND MATERIAL

Population of study includes all Islamic banks that are operating and performing business activities in Pakistan. Secondary Sources used for data collection like as annual reports. Results are quantified through using Statistical tools. Regression Analysis is done to explore the relationship between corporate social responsibility and financial performance.

Model Summary

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	23943.811	6130.964		3.905	.030
	CSR	.972	4.830	.115	.201	.853

The regression analysis shows the positive relationship between CSR and Net Profit of the Islamic banks and results shows that there is positive link among CSR and profitability. Data shows that CSR contribute in the profitability positively. Data and table shows positive link and relationship among CSR and CFP for Islamic banking in the country.

REFERENCES

1. Aqueveque, C. (2005). Marketing and market development signaling corporate values: Consumers' suspicious minds. *Corporate Governance*, 5(3), 70-81.
2. Araña, J. and León, C. (2005). *Consumers' valuation of corporate social responsibility: A stated preference approach*. Unpublished manuscript, University of Las Palmas de Gran Canaria, Canarias, Spain.
3. Arendt, S. and Brettel, M. (2010). Understanding the Influence of CSR on Corporate Identity, Image, and Firm Performance. *Management Decision*, 48(10), 1469-1492.
4. Barone, M., Miyazaki, A. and Taylor, K. (2000). The influence of cause-related marketing on consumer choice: Does one good turn deserve another? *Journal of Academy Science*, 28(2), 248-262.
5. Becchetti, L. and Rosati, F. (2007). Global social preferences and the demand for socially responsible products: Empirical evidence from a pilot study on fair trade consumers. Working paper, *CEIS Tor Vergata Centre for International Studies on Economic Growth*, Rome, Italy.
6. Belk, R., Devinney, T. and Louviere, J. (2005).Consumer ethics across cultures. Consumption, Markets and Culture, 8(3), 275-289.

7. Berens, G. (2004). Corporate branding: The development of corporate associations and their influence on stakeholder reaction. Unpublished master's thesis, Erasmus University Rotterdam, Rotterdam, Netherlands.
8. Calder, B., Phillips, L. and Tybout, A. (1982). The concept of external validity. *Journal of Consumer Research*, 9(3), 240-244.
9. Campbell, D. and Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.
10. Carrigan, M. and Atalla, A. (2001). The myth of the ethical consumer—Do ethics matter in purchase behavior? *Journal of Consumer Marketing*, 18(7), 560-577.
11. Carrigan, M., Szmigin, I. and Wright, J. (2004). Shopping for a better world? An interpretative study of the potential for ethical consumption within the older market. *Journal of Consumer Marketing*, 21(6), 401-417.
12. Carroll, A.B. (1979). A three dimensional conceptual model of corporate performance. *The Academy of Management Review*, 4(4), 497-505.
13. Carroll, A.B. (1991). The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders. *Business Horizons*, 34(4), 39-48.
14. Carroll, A.B. (1979). A Three-dimensional Conceptual Model of Corporate Performance. *Academy of Management Review*, 4(4), 497-505.
15. Chatpol, S. (2009). A Study on How CSR of PTT Influence Purchasing Intention of PTT Users in Bangkok. MBA Thesis, Assumption University of Thailand.
16. Lantos, G.P. (2002). The Ethicality of Altruistic Corporate Social Responsibility. *Journal of Consumer Marketing*, 19(2/3), 205-228.
17. Lin, C.H., Yang, H.L. and Liou, D.Y. (2009). The impact of corporate social responsibility on financial performance: Evidence from business in Taiwan. *Journal of Technology in Society*, 31(1), 56-63.
18. Malik M.S. and Nadeem M. (2014). Impact of corporate social responsibility on the financial performance of banks in Pakistan. *International Letters of Social and Humanistic Sciences*, 21, 9-19.
19. Mazzanti, M. (2003). Discrete choice models and valuation experiments. *Journal of Economic Studies*, 30(5/6), 584-601.
20. McFadden, D. (2001). Economic choices. *American Economic Review*, 91(3), 351-378.
21. McKinsey (2006). The McKinsey global survey of business executives: Business and society. *The McKinsey Quarterly*, 1-5.
22. Oskamp, S. (1991). *Attitudes and Opinions*, 2nd edition. Englewood Cliffs, NJ: Prentice Hall, p. 7.
23. Walters, C.G. and Paul, G.W. (1970). *Consumer behavior: An integrated framework*, (3rd ed.), New York, NY.: Richard D. Irwin, Inc.
24. Whetten, D.A. and Mackey, A. (2002). A Social Actor Conception of Organizational Identity and Its Implications for The Study of Organizational Reputation. *Business and Society*, 41(4), p.393.
25. Yamane, T. (1973). *Statistics: An Introductory Analysis*, 3rd edition. New York: Harper & Row.
26. Zikmund, W.G. (2003). *Business Research Methods*, 7th edition. Mason, Ohio USA: Thompson – South-Western.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 289-310

ROLE OF PRECIOUS METALS IN DIVERSIFYING AND HEDGING STOCK MARKET RISK

Uzair Hamid and **Javed Iqbal**
Department of Mathematical Sciences
Institute of Business Administration, Karachi, Pakistan
Email: uzair.hamid@khi.iba.edu.pk
jiqbal@iba.edu.pk

ABSTRACT

This paper aims at providing empirical evidence of the potential of precious metals i.e. gold, silver and platinum in diversifying risk of investing in emerging stock markets. Following the approach similar to the Hillier, Draper, & Faff (2006), we investigate whether a decrease in stock market index returns is associated with a significant increase in the precious metal returns for the US, China (Mainland), Russia, Brazil, Hong Kong (China), India and Pakistan using daily data for the period of 1995 to 2014.

The diversification potential is investigated in both the normal conditions as well as extreme stock market conditions characterized by excess volatility or lower returns. Keeping in view the empirical characteristics of estimated volatility and returns we made a modification in Hillier, Draper & Faff (2006) methodology of classifying adverse stock market conditions. The results show that the precious metals do not provide a reliable diversification benefit in the emerging markets. This is in contrast to the US case where we found that gold and silver act as hedge in investing stock market risk.

1. INTRODUCTION

In today's world, capital formation has no boundaries and hence international investors actively invest in different countries of the world in both the direct and portfolio investment. As identified by UN Conference on Trade and Development (2014), foreign direct investment (FDI) in emerging economies has risen to $484 billion in 2014. International investment also encourages investors of both home and foreign countries to take risk which is safe, specifically when markets are "co-integrated". If the stock markets of both countries share common factors it can negatively affect international investment. As the financial markets are becoming integrated, they are becoming vulnerable to financial shocks arising in any one country. For instance, as Nashier (2015) points out when there was financial crisis in the USA (2007-08), investors diverted their investment to other markets and products as they provide them better opportunities. Vermeulen (2013) concluded that foreign equity holdings and stock market correlations were not related before financial crisis but had a negative relationship during the recent global financial crisis. Owing to the vulnerability of investing in stock markets, investors are seeking alternative instrument that can provide hedge against stock market risk. Thus emphasis has been placed on the importance of precious metals that can act hedging asset when there are adverse conditions in the stock markets (Daskalaki & Skiadopoulos, 2011). Within precious metals, gold has become a safe haven and even a hedging tool (Gencer & Kilic, 2013). Various studies have been conducted in the recent past e.g. by Jaffe (1989) and Chua, Sick, & Woodward (1990) that analyse the importance of gold as a diversification of financial portfolios. Hedging ability of other precious metals has also been studied as it is related to a low level of correlations with inflation level as well as the equity index (Conover, Jensen, Johnson, & Mercer, 2009).

Due to their beauty and scarcity, precious metals are highly valued. Resistance towards corrosion and rust, reflectivity, conductivity and malleability are other unique traits that are directly related to these metals. Gold is one of the oldest and reliable precious metals ever used. It has been in use is for past hundreds of years as an important financial asset. Gold is known for its durability, divisibility and for standardizing values (Capie, Mills, & Wood, 2005). When people are in crisis or they are short of cash they sometimes sell their gold. A large number of people in the world even buy gold for saving or to increase their wealth. Banks invest in gold for increase in their assets just like investor buy gold and sell them whenever there is an increase in the gold prices to gain profit. Gold has also been used as the basis of monetary system until the collapse of Bretton Woods system in 1971. Gold has changed the way it is being used, and today it is also used to hedge against inflation and exchange rate risk (Capie, Mills, & Wood, 2005).

Platinum is a highly valuable and rare metal that has very distinguished features that do make the metal extremely valuable and in demand across a wide variety of potential consumers. Due to its unique physical and chemical properties, platinum has been used in commercial and industrial applications. Due to its core traits and attributes, metal has become highly precious and valuable (World Platinum Investment Council, 2015). Physical form of platinum including platinum-backed ETFs have been bought by investors in USA, Canada, Europe, Africa and even some parts of Asia.

As compared to other highly acclaimed precious and highly valuable metals, silver is comparatively scarce. However, it is still the most economical and value for money product that has earned a reputation of a highly valuable metal. Due to its unique traits and attributes, silver is used for a wide variety of operations such as technology, decoration, medicine, photography etc. This has made it indispensable within the commodities market (The Silver Institute, 2015). Industry, business investment and jewellery and are the three most important uses of silver. As a result, these three highly crucial areas are comprised of more than 95% of annual demand in silver. Only 167.0 million ounces of silver was used for the purpose of jewellery, 487.4 million ounces for industrial applications and 101.3 million ounces were used for the purpose of making coins and medals in the year 2010 (The Silver Institute, 2015). Due to the fact that market of silver is much smaller and has lower liquidity as compared to gold, there is a possibility that silver prices may fluctuate at a much higher rate in comparison to gold. It also means that with respect to dollars, investors may be able to receive a much larger and better returns in silver as compared to gold (CPM Group, 2014). Figure 1illustrates the growth the prices of precious metals over the sample period considered in the present study. It appears that the gold prices has shown quite smoother increase up to 2012. The gold price shows a declining trend in the later periods. Both silver and platinum are observed to have much more volatile history than he gold prices.

There has been emerging interest in gold specifically in times of economic recession and economic instability. It has been believed to be a medium for exchange, a standard value (Ibrahim & Baharom, 2011). According to Capie, Mills, & Wood (2005); Hillier, Draper, & Faff (2006); and Baur & Lucey (2010), gold is considered a safe haven for investors and it can also be used in the future to diversify the risk portfolio of a company. After analysing returns of the US, UK and German stock markets as well as bond markets, Baur and Lucey (2010) found that gold can be considered as a safe investment. According to Hillier, Draper, & Faff (2006) the hedging property of precious metals make them ideal investment instruments specifically in times of stock market uncertainty. Smith (2002) found that weak negative correlation was found between stock markets and gold prices in short term. Wang, Lee, & Thi (2011) found from the examination of the long term and short term relationship between oil, gold, foreign currency rate and stock market analysis of Japan, Germany, USA and China that apart from the USA, a long-term correlation exists between oil, gold, exchange rates and stock prices.

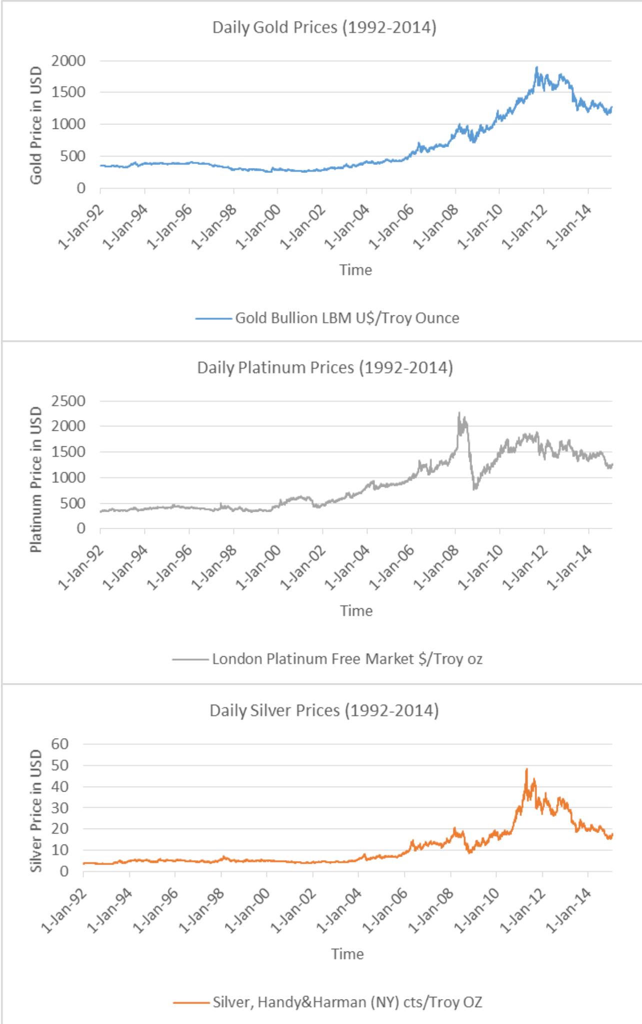


Fig. 1: Daily Prices of Precious Metals from 1992 to 2014

There are numerous stock exchanges throughout the world. However, only 50 stock exchanges are presently active. According to Claessens & Schmukler (2007), some of the well-known markets include NASDQ, NYSE, London Stock Exchange, Singapore Stock Exchange, National Stock Exchange of India, Brazil Stock Exchange, Shanghai and Shenzhen Stock Exchanges in China, Moscow Exchange in Russia. One of the most agile and most profitable stock markets is the Chinese market. However, stock market of China has also become highly active for various financial purposes. It is a conceived fact that China has become an economic hub and is a great contributor to the global economy. Although the stock market of China is not really an open market but investors in China are well aware of market conditions of other economies and have become active players in the global financial market (Chen, Blenman, Bin, & Chen, 2011). As of 2014, the stock market of China had a market capitalization of almost $4.2 trillion and this is primarily due to a strong regulatory financial system (Gismatullin & Nishizawa, 2014).

The scenario of Russia is of high interest. Russia was once a market leader in the Soviet region and was forced to create a stock market when it was required to have a transformation from a merely planned system to a highly developed and robust market economy. As of 2005, the stock market of Russia had become one of the strongest emerging markets in the world and has a market capitalization of more than $600 billion.

When compared to Brazil and South Africa, the stock market of Russia is highly volatile due to lack of strong governance and political and legal controls in the country. Brazil is a country that has one of the highest and most fragile financial risks. This does prove that investors, financial institutions, credit rating agencies etc. have to be consistently aware of the country's financial condition.

After the financial crisis in 2008, Brazil was able to attract more investors from across the world. This can also be attributed to the fact that the higher authorities in Brazil do have the know-how for managing financial crisis that have been detrimental for the economy. It is an irony that Brazil has been unable to take any benefit from various financial opportunities present across the globe. The stock market of Brazil has suffered to a huge extent due to a consistent level of hyper-inflation and also due to social, political and economic instability. However, the Plano Real was one of the biggest achievements in the country's stock market and has achieved high growth since 2003.

The biggest stock exchange in Latin America as well as one of the biggest stock exchanges in the world is Bovespa. The most important index of Bovespa is Ibovespa. Ibovespa helps in evaluating more than 82% of either operations in the share market or the overall value of stock exchange. Although the return on Bovespa has been highly profitable in recent years but the realistic performance can only be gauged by comparing it with other highly performing and emerging stock markets, for instance the BRICs.

In the fast growing region of Asia, Hong Kong can be stated as one of the most highly industrialized and also a highly developed region. Because it provides easy access towards finance for international as well as local investors and also due to the fact that it has lead towards swift development of Asia, Hong Kong can be recognized as one of the most promising financial centres in the world. According to Agarwal, Liu and Rhee (2007), Hong Kong is a well-known and fast growing market economy with very few barriers to economic growth and it also provides transparency for flow of information within various stock markets in Asia and this has also allowed it to be internationally recognized for trading and business development purpose.

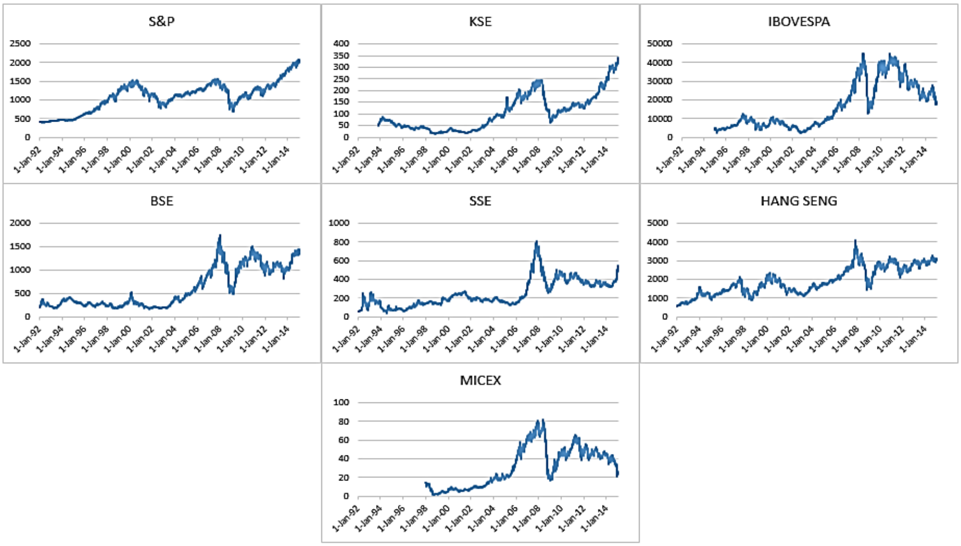


Fig. 2: Stock Market Returns in Dollar terms (adjusted for exchange rate)

It is a fact that Hong Kong has strong social and economic ties with China (Cheung & Sami, 2000). Thus, the stock market of Hong Kong has become the epicentre for investors who want to invest in China. HANG SENG Index (HSI) has been used as a free floating stock index. Apart from that, HSI has also been used as a key part for determining the performance of the stock market of Hong Kong. According to Lee, Shie, & Chang (2012), HSI has been highly pivotal because it has been able to represent more than 60% of overall market capitalization of HKSE. Apart from that, the index also performs and records performances and even monitors day to day stock price movement of the most profitable companies in the HKSE. Figure 2 presents the time series plots of the market indices considered in this paper.

Analysis of the linkage of precious metal return with the stock markets in developed economies, including US and of developing economies including BRIC (Brazil, Russia, India, and China), Hong Kong and Pakistan is the main aim of this paper. Prior research has been done regarding the hedging and diversification properties of precious metals but most of them are restricted to gold only. This paper intends to improve upon previous researches by examining the potential of precious metals e.g. gold, silver and platinum in diversification of risk of investment in stock market portfolios. The paper tests whether gold can hedge risk of investing in stock portfolios using daily data for the period of 1995 to 2014. Following Hillier, Draper, & Faff (2006), the diversification benefit of precious metal is investigated in the framework of the GARCH models. In the first model the mean equation links precious metal returns to stock market index return. The second model allows a more informative analysis of whether the precious metal returns have significantly negative relationship with stock index returns during the normal stock market conditions and during the extremely adverse conditions characterized by either high volatility or extremely low returns. Following Baur and Lucey (2010) and Baur and

McDermott (2010) the extreme market conditions are characterized as those periods when the volatility exceeds the top 5% quantile.

Following this introductory section the paper is structured as follows: Literature review is presented in section 2. Section 3 discusses the modelling framework and methodology used in this paper. Findings and results are discussed in section 4 and section 5 concludes the research findings.

2. LITERATURE REVIEW

Before proceeding towards subsequent analysis, it is important to build a theoretical argument. Literature has identified three types of asset categories that can be considered. They are hedge, diversifier, and safe haven. Baur & Lucey **(2010)** have differentiated these three asset categories as follows:

Hedge Asset – On average, this asset is uncorrelated or negatively correlated with another asset or portfolio.

Safe Haven Asset – This asset is uncorrelated or negatively correlated with another asset or portfolio in adverse market conditions.

Diversifier Asset – On average, this asset is positively correlated (but not perfectly) with another asset or portfolio.

From a business and literal perspective, assets that are preferred by investors during period of uncertainty to secure their capital investment are known as safe haven **(Lucey & Li, 2015)**. For quite a long time, a large group of assets have been considered as safe havens. Hedge and diversifier assets cannot provide returns in extreme market conditions as the requirement of correlation property is on average only **(Baur & Lucey, 2010)**.

During the 1980's and 1990's, the most common safe havens were the US assets and the US dollar **(Lucey & Li, 2015)**. Other assets which can be considered as safe haven are the Swiss Franc, gold, German government bonds, the Yen, the Euro, the New Taiwan dollar, the Singapore dollar, money market funds, the commercial market, Russian assets, equity and real estate **(Lucey & Li, 2015)**.

There has been an intensive level of study in the last few years over the importance of gold as a safe haven asset against stock market. The safe haven properties of assets such as platinum, palladium, gold and silver, were examined by Lucey & Li **(2015)**. Researchers analysed why they have been considered so important over the time. To analyse the cross-market correlation dynamics, Engle's **(2002)** dynamic conditional correlation model was applied.

Citak **(1999)** was among the first to discuss about gold as a safe haven in the research. Since the mid 1980's, gold has been considered as a safe haven, as stated by Citak **(1999)**. The speculation of silver, the inflation panic in USA and the second oil shock in 1979 were the core reasons for the safe haven issue. Even in stock market crashes, gold has been considered as safe haven in various countries as a short term solution **(Baur & Lucey, 2010)**. However, all the big markets in the world can be viewed from a broader perspective in this regard. For European stock markets and USA gold may be considered as a safe haven but can't be considered safe for countries like Japan, Canada and

Australia and for developing countries such as Brazil, Russia, India and China (BRIC) **(Baur & McDermott, 2010)**. According to Coudert & Raymond **(2011)** gold may not be productive in the long run but they can be considered as a safe haven in the short term. Gold can be considered as a monetary asset as well as a safe haven specifically against exchange rate fluctuations in the UK and USA **(Ciner, Gurdgiev, & Lucey, 2013)**. As highlighted by Jaffe **(1989)**, Worthington & Pahlavani **(2007)**, Pullen, Benson, & Faff **(2014)**, gold stocks and gold derivatives can also work as hedge and safe haven.

According to Agyei-ampomah et al. **(2014)** and Lucey & MacKenzie **(2013)**, the safe haven potential of precious metal assets such as platinum, palladium and silver has been rarely discussed as a hedge, diversifier or safe haven as compared to gold and only recently has their importance been highlighted as risk diversifier. As stated by Batten et al. **(2010)** and Batten et al. **(2015)**, the relationship between silver, gold, palladium and platinum are different from each other.

There is a varying level of volatility and spill over between the metals, although the metals have a varying degree of supply and demand fundamentals between them and they even portray interesting interconnections between them. The volatility link between the precious metals may not always be constant. There is a strong level of correlation between precious metals specifically in their last decade **(Sensoy, 2013)**. Silver has a unidirectional volatility shift contagion on platinum and palladium while gold has this effect on all the precious four metals. Precious metals cannot be considered as a single or unilateral asset as they are too distinct for that **(Batten, Ciner, & Lucey, 2010)**. Only monetary variable such as inflation, growth rate in money supply and interest rate are affected due to gold’s volatility in stringent and tough economic conditions. Palladium and platinum are responsive to financial as well as monetary variables and they are also related to S&P 500 returns while nothing affects silver in this regard. There is certainly a cross-metal interaction between convenience yields **(Chng & Foster, 2012)**. The return and volatility of other precious metals is affected by the yielding potential of gold and silver and it is also linked with any change in the condition of global economy. There are some limited spillovers that are shared by gold and silver and for this reason palladium and platinum sometimes appear to be isolated from silver and gold **(Batten, Ciner, & Lucey, 2015)**.

Smith **(2001)** examined the relationship between price of gold and US stock. The study was carried out for US stock indices only (Dow Jones Industrial Average, NASDAQ, S&P 500 Composite Stock Price Index, and Russell 3000 Index). Where appropriate, daily, weekly and monthly data was used, ranging 1991 to 2001. Small or negative short-run correlation was found in between returns on US stock price and gold. No co-integration was found between gold and stock price for US. However, in long run, Granger causality indicated that unidirectional causality exists. US stock returns granger causes London gold price being set in morning and evening.

Hillier et al. (2006) extended the study by including other precious metals. Hillier et al. (2006) comprehensively analysed the relationship between gold, silver, and platinum with capital markets for both, US and Non-US. For US stocks, S&P 500 Index was used, whereas for Non-US stocks, MSCI Europe, Australasia, and Far East (EAFE) Index were used. Apart from incorporating additional commodities and Non-US markets, researchers also introduced GARCH methodology. Daily data from 1976 to 2004 concluded that commodities (Platinum, Gold and Silver) should be included in investment portfolios as the diversify risk. Hedging capability has been witnessed at times of stock market being volatile and abnormal. Hillier et al. **(Hillier, Draper, & Faff, 2006)** also ascertained that portfolios containing gold perform better than portfolios having financial assets only.

Gilmore et al., **(2009)** moved further in incorporating VAR, VECM, and variance decomposition to examine the dynamics involved in gold prices, stocking prices of gold mining companies, and whether or not if they have a co-movement with stock market prices. London Gold Bullion prices were evaluated against two gold company stock price indices (GOX and HUI) and three stock market indices representing large, mid and small capitalization stocks are employed (The S&P 500 (LCAP); S&P midcap 400 (MCAP); and S&P smallcap 600 (SCAP). Weekly data was taken of 11 years (1996 to 2007). Gilmore et al., (2009) were able to find correlation among these variables. A vector error-correction model revealed that both gold and large-cap stock prices adjusted to disturbances to restore the long-term relationship between the variables. Moreover, short-term unidirectional causal relationships were running from large-cap stock prices to gold mining company stock prices and from gold mining company stock prices to gold prices.

Do et al., **(2009)** used GARCH for ASEAN emerging stock markets to study the effects of international gold market on stock exchange volatility. Although, many empirical works have been carried out on stock exchange volatility over the world, only few researches on gold return volatility have been done i.e., influences of macro-economic variables on gold returns and volatility **(Tully & Lucey, 2007)**, and response of gold returns and volatility to public information arrival **(Kutan & Aksoy, 2004)**. Daily closing data (2000-2008) revealed that gold could be a substitute commodity for stocks in Vietnam and the Philippines, while it could be a complement for stocks in Indonesia, Thailand and Malaysia.

Baur & Lucey **(2010)** analysed stocks, bonds, and golds via GARCH and VAR model to ascertain if gold is a hedge or a safe haven. As stated by Baur & Lucey **(2010)**, gold is a hedge if it is uncorrelated with stocks or bonds on average, or it is a safe haven if it is uncorrelated with stocks and bonds in a market crash. Constant and time-varying relations between U.S., U.K. and German stock and bond returns and gold returns was studied from 1995 to 2005. Results show that gold is a safe haven for stocks. However, gold is generally not a safe haven for bonds in any market. Gold only functions as a safe haven for a limited time, around 15 trading days. In the longer run, gold is not a safe haven, that is, investors that hold gold more than 15 trading days after an extreme negative shock lose money with their gold investment. This finding suggests that investors buy gold on days of extreme negative returns and sell it when market participants regain confidence and volatility is lower.

Baur & McDermott **(2010)** also undertook a multi-country analysis of emerging and developing countries about gold being a safe haven or not. The research examined the role of gold in the global financial system. Baur & McDermott **(2010)** tested the hypothesis that gold represents a safe haven against stocks of major emerging and developing countries. GARCH model was run on daily, weekly and monthly continuously compounded stock returns of a sub-set of the 53 constituents of a world index from 1970 to 2009. It was concluded that gold is both a hedge and a safe haven for major European stock markets and the US but not for Australia, Canada, Japan and large emerging markets such as the BRIC countries. Looking at specific crisis periods, we find that gold was a strong safe haven for most developed markets during the peak of the recent financial crisis.

So far, Baur **(2013)** has also been successful in analysing gold from a very different perspective. Baur **(2013)** studied recurring annual events potentially introducing seasonality into gold prices. Monthly data of 30 years (1980 to 2010) run through Asymmetric GARCH model that analysed returns of gold bullion spot prices (London fixing), COMEX gold future prices, Silver cash prices (London fixing), GSCI commodity index returns, GSCI Precious Metal index returns, MSCI Emerging Market index return, MSCI World index return, and S&P500. As per the results, September and November were the only months with positive and statistically significant gold price changes. This "autumn effect" holds unconditionally and conditional on several risk factors. The anomaly can be explained with hedging demand by investors in anticipation of the "Halloween effect" in the stock market, wedding season gold jewellery demand in India and negative investor sentiment due to shorter daylight time. Lastly, the autumn effect can also be characterized by a higher unconditional and conditional volatility than in other seasons.

Probit model has also been used by Mulyadi & Anwar, **(2012)** to conduct a comparison between the stock investment and gold investment. Based on previous research, it is concluded that gold is a good portfolio diversifier, a hedge against stock and safe haven in extreme stock market condition **(Baur & Lucey, 2010)**. As an investment instrument, stock was exposed to macroeconomic risks and global stock market risks. Weekly data from 1997 to 2011 showed that the gold investment is quite safe for the investors and could be categorized as safe haven. This conclusion is also supported by the previous research that identified gold as a good portfolio diversifier and a hedge against stocks as well as a safe haven in extreme stock market conditions.

The Istanbul Stock Exchange (ISE) sector and their volatility in the market return were further analysed by Gencer & Kilic, **(2013)**. They used a combined gold and oil volatilities and return impact in order to conduct the analysis. The CCC M-GARCH multivariate model has been applied to present multidimensional interactions regarding the volatility processes and return of the specific variables at hand. For each model, it was found that there are close to unity and significant oil GARCH effects present. Hence, oil prices are regarded as major source for the volatility of the portfolio. In magnitude, the oil GARCH parameters are followed by the gold GARCH effects. This aspect also shows that the portfolio variability is significantly affected by the gold prices. The three sector indices which include commercial sectors, main metals and holding have a negative correlation with gold according to Gencer & Kilic, **(2013)**.

3. METHODOLOGY

We follow the approach similar to Hillier et al. **(2006)**. A market model-type regression for assessing the basic diversification properties of precious metals is specified as:

$$R_{pm,t} = \alpha_{pm} + \beta_{pm} R_{m,t} + e_{pm,t} \tag{1}$$

where,

$R_{pm,t}$ = Return on precious metal at time t

$R_{m,t}$ = Return on stock market index at time t

Equation (1) constitutes the mean equation of the GARCH model (Engle, 1982, Bollerslev, 1986). The variance equation is specified as the (z_t is a student t random variable).

$$e_{pm,t} = z_t \sigma_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \sigma_{t-i}^2 + \sum_{j=}^{q} \beta_j e_{pm,t-j}^2 \tag{2}$$

In most of the case we have employed a GARCH (1,1) as the volatility equation as this was found to be most adequate specification.

If regression slope coefficient (β_{pm}) is negative and significant, an asset will be termed as a diversifying asset. Contrary to this, an asset will not be a diversifying asset if the β_{pm} (slope coefficient of regression / elasticity) is positive and significant. However, the equation presented above is devoid of stock market conditions and is unable to capture all the effects. In order to capture conditional diversifying features, equation 1 is modified by including two interaction terms. These terms represent the product of stock market return and dummy variables. Here VD_t assumes value 1 when the volatility estimated by GARCH model exceeds the 95% quantile and zero otherwise. Similarly RD_t is a dummy variables which equals 1 when the market return is below 5% quantile and zero otherwise. Thus both dummy variables represent extremely adverse conditions for investors in the stock market. The purpose of adding these terms is to be able to ascertain whether the diversification potential of precious metals is different in different stock market conditions.

$$R_{pm,t} = \alpha_{pm} + \beta_{pm} R_{m,t} + \gamma R_{m,t} \times VD_t + \delta R_{m,t} \times RD_t + \varepsilon_{pm,t} \tag{3}$$

The volatility equation in this case is also the same as given in equation (2). As identified by Hillier et al., **(2006)** equation 3 can be used to elaborate upon the conditional diversifying features of precious metals. Precious metals can be assessed for this diversifying benefits in good and bad times of the economy.

Equation 3 can be categorised in 4 cases. Firstly, the effect of relatively normal market conditions and high market returns is captured through β_{pm}, which depicts good and stable position for any investor. Secondly, $\beta_{pm} + \gamma$ represents the effect on precious metal returns during highly volatile market. Thirdly, $\beta_{pm} + \delta$ is the effect on precious metal returns of extremely poor performance period of stock market as characterized by worst 5% returns. Finally $\beta_{pm} + \gamma + \delta$ represents the case when both market is highly volatile and the market performance is worst. These cases have been summarized in table 1.

	Volatility	
	Stable	Unstable
Equity Returns	**Less than 95% quantile**	**More than 95% quantile**
Good (More than 5% quantile)	β_{pm}	$\beta_{pm} + \gamma$
Bad (less than 5% quantile)	$\beta_{pm} + \delta$	$\beta_{pm} + \gamma + \delta$

Hillier et al. **(2006)** classified the adverse conditions of volatility and stock returns based on mean and standard deviation. For example in the Hillier et al. **(2006)** case, the dummy variable VD_t represents the periods when volatility is two standard deviations above its average level. Similarly RD_t represented a period when the market returns were two standard deviations below the average return. Thus Hillier et al. **(2006)** implicitly assumed a symmetric normal distribution of estimated volatility and returns. However empirically both of these assumptions are far from reality. To illustrate this fact Fig 3 plots the non-parametric Kernel density curves fitted to the time varying standard deviations estimated by the GARCH models for the markets considered in our study. It is clearly seen that the empirical distributions of the estimated volatility are highly positively skewed. In such skewed distributions mean and standard deviations are not adequate summary measures. We therefore classify the market conditions based on quantiles. Similar quantile based measures are also used by Baur and Lucey (2010) and Baur and McDermott (2010) to classify adverse market conditions.

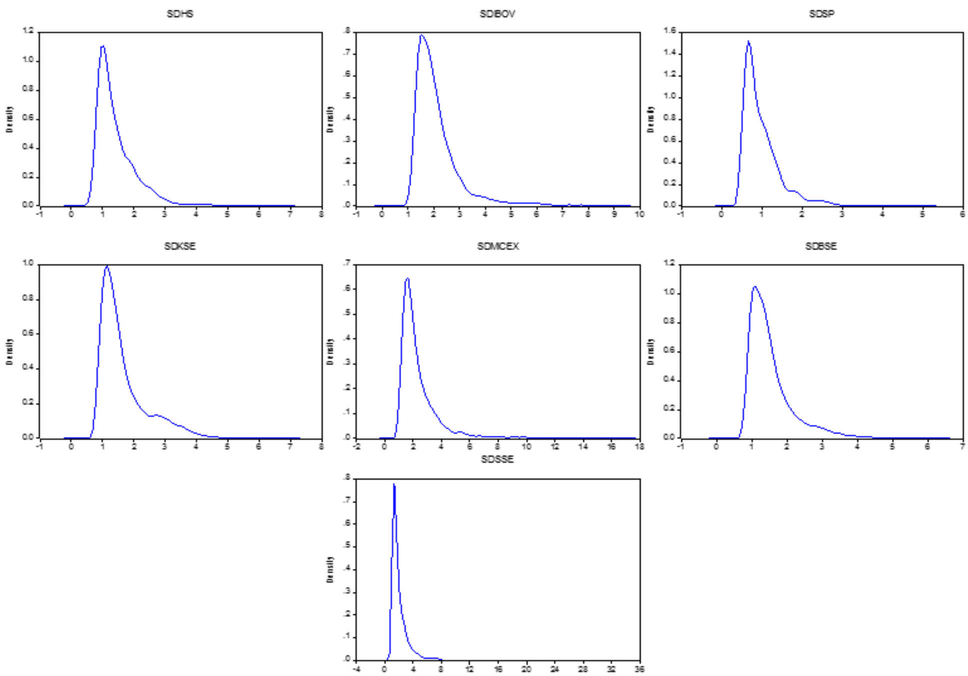


Fig. 3: Kernel Density Curves of the Distributions of Stock Market Volatility

The daily data for gold, silver and platinum have been extracted from DataStream International. The precious metal data used are the Gold bullion LBM U$/Troy Ounce, Silver Cents/Troy Ounce (converted into the US dollar), and London Platinum Free Market US $/Troy Ounce. S&P 500 Composite price index has been used as a proxy to exhibit perspective of investors for US stock market returns. Bombay Stock Exchange (BSE 100) represents of Indian stock market, IBOVESPA 50 is used to depict Brazil, MICEX 50 to showcase Russian stock market and Shanghai Stock Exchange (SSE) and HANG SENG 48 (HSI) are used to represent China's Mainland and China's Hong Kong markets respectively. Karachi Stock Exchange (KSE 100 Index) has been used for Pakistan. In the DataStream the local stock market indices are expressed in local currencies. The local indices have been converted to the US dollar based on the exchange rate of the particular date.

The metal price and stock market indices are converted into continuously compounded returns. The US market has been included in this research to identify if there has been any change in the diversification potential of the precious metals in the US using more recent data. The BRIC nations have giant economies and are centre of attention by international investors. As compared to the BRIC, Pakistan is smaller economy and a smaller stock market. Stock market of Pakistan is a typical emerging market since it has high level of returns, higher volatility, market concentration is high (Iqbal, 2012; Khwaja and Mian, 2005). After adjusting for exchange rate (dollar denomination), sample periods for S&P, BSE, SSE, and HSI ranges from 2-Jan-1992 to 19-Jan-2015. The sample periods for KSE, IBOVESPA and MICEX range over the periods (26-Oct-1993 to 19-Jan-2015), (3-Jan-1995 to 19-Jan-2015) and (2-Jan-1998 to 19-Jan-2015) respectively.

Table 2 presents basic descriptive statistics of the variables used in this study over the sample period involved. Mean daily returns have been positive and within the range of 0.01 percent to 0.04 percent. In precious metals, silver attains the highest (approximately 7.6 percent a year). Mean daily return of 0.025%. For stock markets, highest mean daily return of 0.037 percent has been achieved by the Shanghai Stock Exchange (approximately 11.1 percent a year). The Russian stock market represented by MICEX has been the most volatile series with a standard deviation of 2.901 percent whereas the US market represented by the S&P-500 index is found to be the least volatile market. Among the precious metals gold is associated with the lowest volatility of about 1%. It is noted that all the precious metal and stock index returns are found to be non-normal with excess kurtosis. The JB test easily rejects the normality of these series.

Table 1
Basic Descriptive Statistics of Precious Metals and Stock Index Returns

	RGOLD	RPLAT	RSILVER	S&P	KSE	IBOVESPA	MICEX	BSE	SSE	HSI
Mean	0.021	0.022	0.025	0.026	0.034	0.024	0.012	0.032	0.037	0.028
Median	0.000	0.000	0.000	0.026	0.034	0.081	0.071	0.065	0.000	0.023
Maximum	7.382	11.728	13.665	10.957	12.751	23.603	28.931	16.640	71.915	17.273
Minimum	-10.162	-17.277	-12.982	-9.470	-13.214	-17.238	-33.375	-13.072	-40.534	-14.696
Std. Dev.	1.001	1.369	1.772	1.137	1.766	2.377	2.901	1.676	2.417	1.608
Skewness	-0.319	-0.649	-0.516	-0.249	-0.363	-0.005	-0.520	-0.035	4.440	-0.046
Kurtosis	10.770	13.784	9.178	12.368	8.101	10.381	20.909	10.028	158.888	12.387
Annualized Mean Return (300 working days)	6.4	6.6	7.6	7.9	10.3	7.3	3.6	9.6	11.1	8.5
Jarque-Bera	15,229.2	29,561.6	9,829.2	22,050.0	6,128.2	11,873.1	59,631.5	12,375.6	6,108,144.0	22,078.0
Probability	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Observations	6,013	6,013	6,013	6,013	5,540	5,230	4,447	6,013	6,013	6,013

4. RESULTS AND DISCUSSION

Table 3 presents the correlation between pairs of precious metals (gold, platinum and silver) and stock indices of USA, Pakistan, India, China – Shanghai, China – Hong Kong, and Brazil. Stock indices have been adjusted for exchange rate for each country to arrive at dollar denominated stock index. Panel A covers full sample periods where sub periods have been created in Panel B and C to show any changes in correlation structure over the two sub periods. It is observed that only gold has negative correlation with S&P in all the periods. Correlation of Pakistan (KSE) with precious metals is approximately zero (Panel A) and consistent over the two sub periods. Correlation for India (BSE) with precious metals has shown weak positive correlation. Same result are observed for Brazil and China. For full sample period, the lowest correlation (-0.03) can be seen for gold and S&P, whereas, highest correlation (0.17) exists between platinum and Hong Kong. Average correlation of stocks markets and precious metals is 0.08.

Lastly, gold and silver returns have shown strong correlation (0.66) whereas the correlation of platinum/silver and gold/platinum averages around 0.40. Over the period of time, correlations between the returns of these precious metals have increased from the average of 0.33 (1995-2004) to 0.56 (2005-2014). It is observed that correlations are positive but far from perfect diversification benefit can be realized by a portfolio consisting of investment in stock markets and precious metals.

Table 2
Correlation Structure of Gold, Silver, Platinum, S&P, KSE, BSE, SSE, HANG SENG, IBOV, MICEX

	Gold	Silver	Platinum	USA	Pakistan	India	Shanghai	Hong Kong	Brazil	Russia
A. Full Sample Period (1995-2014)										
Gold	1.00									
Silver	0.66	1.00								
Platinum	0.41	0.38	1.00							
USA	-0.03	0.06	0.07	1.00						
Pakistan	0.02	0.02	0.02	-0.01	1.00					
India	0.08	0.14	0.14	0.18	0.08	1.00				
Shanghai	0.04	0.05	0.08	0.02	0.04	0.14	1.00			
Hong Kong	0.06	0.13	0.17	0.18	0.09	0.39	0.23	1.00		
Brazil	0.04	0.11	0.09	0.48	0.01	0.19	0.06	0.22	1.00	
Russia*	0.08	0.18	0.15	0.18	0.06	0.26	0.09	0.29	0.22	1.00
B. 1995 - 2004										
Gold	1.00									
Silver	0.50	1.00								
Platinum	0.26	0.23	1.00							
USA	-0.07	-0.04	0.01	1.00						
Pakistan	0.02	0.03	0.00	-0.01	1.00					
India	0.00	0.02	0.00	0.06	0.07	1.00				
Shanghai	0.02	0.00	-0.01	-0.04	0.05	0.04	1.00			
Hong Kong	0.04	0.05	0.04	0.12	0.10	0.21	0.04	1.00		
Brazil	-0.01	0.03	0.01	0.40	0.00	0.08	-0.02	0.13	1.00	
Russia*	0.02	0.09	0.00	0.07	0.07	0.14	0.02	0.20	0.09	1.00
C. 2005 - 2014										
Gold	1.00									
Silver	0.72	1.00								
Platinum	0.50	0.47	1.00							
USA	-0.01	0.12	0.11	1.00						
Pakistan	0.02	0.01	0.04	-0.01	1.00					
India	0.14	0.21	0.27	0.27	0.09	1.00				
Shanghai	0.06	0.09	0.17	0.07	0.04	0.26	1.00			
Hong Kong	0.08	0.18	0.30	0.24	0.08	0.58	0.46	1.00		
Brazil	0.08	0.18	0.19	0.59	0.03	0.33	0.18	0.35	1.00	
Russia	0.13	0.26	0.31	0.31	0.03	0.38	0.16	0.40	0.41	1.00

* Russia (1998 onwards)

Table 4 report the estimated parameters of the market model regressions of eq (1) where returns of gold, silver and platinum have been regressed on the returns on S&P (USA), KSE (Pakistan), IBOVESPA (Brazil), MICEX (Russia), BSE (India), SSE (China - Shanghai), and HANG SENG (China – Hong Kong).

Table 4
Basic Diversification Properties of Precious Metal Returns

		Gold	Platinum	Silver
S&P	Constant	0.0120	0.0260**	0.0307**
	Market Return	-0.0387***	0.0430***	-0.0187
KSE	Constant	0.0102	0.0261**	0.0300*
	Market Return	0.0088*	0.0164**	0.0058
Brazil	Constant	0.0092	0.0217	0.0311*
	Market Return	0.0038	0.0304***	0.0193***
Russia	Constant	0.0389***	0.0496***	0.0383**
	Market Return	0.0055	0.0477***	0.0425***
India	Constant	0.0099	0.0255**	0.0285*
	Market Return	0.0065	0.0677***	0.0458***
China-Shanghai	Constant	0.0097	0.0273**	0.0296*
	Market Return	0.0004	0.0072	0.0030
China-Hong Kong	Constant	0.0093	0.0243**	0.0274*
	Market Return	0.0110**	0.0764***	0.0600***
* Significant at 10%				
** Significant at 5%				
*** Significant at 1%				

It is observed that only the US stock market has negative relationship with gold and silver returns. However, only gold is found to hedge stock market investment risk significantly in the US. The elasticity of the platinum in the US is significantly positive. As the correlation between platinum and stock index is very small positive number, as per the definition of Baur and McDermott (2010), it appears that platinum can diversify the risk of a portfolio if it is held with US stocks in a portfolio. It is also observed that no precious metal acts as a hedge in reducing stock market risk in any of the emerging markets under study.

This result is consistent with Hillier et al. (2006) who found positive precious metal elasticities with non-US stocks. Results are in harmony with the research of Smith (The price of gold and stock price indices for the United States, 2001) on US Stock and gold price (1991-2001) depicting small or negative short run correlation. The insignificant elasticity of gold with stock indices of BRIC countries indicate that some diversification benefit can be obtained by including gold in stock portfolios during the normal market conditions. The findings also strengthen the claim of Baur & McDermott (Is gold a safe haven? International evidence, 2010) who found negative coefficient for US and European stock markets but positive coefficients for BRIC nations.

In situation where stock market is performing poorly or is highly volatile, investors would definitely be looking for hedging and diversification opportunities. In this paper, “high volatility” (i.e. when market volatility exceeds 95% quantile) and “poor market performance” (i.e. when market returns is less than 5% quantile) have been expressed in quantile terms. These definitions have been used to produce conditional analysis that has been presented in Table 5.

The results for the market beta coefficient is similar to the table 4 as it is observed that evidence of gold and silver acting as a hedge against stock market is found only in the US case. In BRIC countries the betas are positive but insignificant (except Brazil where it is significant). Thus some diversifying benefits of gold can be realized in these markets during normal times. In all the countries platinum returns are positive and significant with stock market. Thus platinum is found to be more like a pro-cyclic metal whose prices are higher during good business conditions. To a lesser extent silver also has similar characteristics.

The coefficient of gold for the volatility dummy are found to be negative and significant only in case of the US and Russia. Thus according to the definition of Baur and Lucey (2010), gold is found to be a safe haven asset against adverse movements of stock markets in these counties. The coefficient of volatility dummy with gold is insignificant in other countries. Thus in these countries gold cannot act as hedging risk of stock investment but gold can provide diversification benefit in adverse conditions of stocks. It is observed that only in case of Russia platinum is acting as a safe haven as the volatility dummy carries a negative and significant coefficient for Russia. In contrast silver is found to act as a safe haven asset in US, Brazil and Russia as during high market volatility periods silver appears to compensate losses occurring in stock markets of these countries. Other stock markets do not any hedging asset in high market volatility periods. Hillier et al., (Do precious metals shine? An investment perspective, 2006) found gold, silver and platinum as ideal investment in US only when stock market uncertainty was high. In this research, we can see that platinum is no more a hedging asset in times of high market volatility for US. However, the phenomenon do exist in case of Russia where all 3 metals perform as diversifying assets at times of high market volatility. Only in case of Russia we find evidence of precious metals being strong safe haven.

The coefficients of the return dummy variables are found to be mostly insignificant for all the countries except for the US. Hillier et al., (Do precious metals shine? An investment perspective, 2006) also concluded that the estimated coefficients on the return dummy was insignificant for all the three metals which points to the limitation in the diversifying role of precious metals when the stock market perform poorly.

Table 3
Diversification Properties of Precious Metals in Periods of High Volatility and Poor Performance

		Gold	Platinum	Silver
S&P	Constant	0.0128	0.0239*	0.0409***
	Market Return	-0.0351***	0.0500***	-0.0343**
	Volatility Dummy	-0.0712***	-0.0053	-0.0729*
	Market Return Dummy	0.0109	-0.024	0.0870***
KSE	Constant	0.0103	0.02548*	0.0322*
	Market Return	0.009	0.01764*	0.0009
	Volatility Dummy	-0.0022	-0.0019	0.0069
	Market Return Dummy	0.0004	-0.0027	0.0093
Brazil	Constant	0.0068	0.0186	0.0262
	Market Return	0.0010**	0.03827***	0.0333***
	Volatility Dummy	-0.0112	-0.0199	-0.0404**
	Market Return Dummy	-0.0102	-0.0132	-0.0173
Russia	Constant	0.0410***	0.04637***	0.0390**
	Market Return	0.0093	0.06453***	0.0542***
	Volatility Dummy	-0.0180**	-0.0481***	-0.0425***
	Market Return Dummy	0.0073	-0.0099	0.005
India	Constant	0.012	0.0223*	0.0292
	Market Return	0.0021	0.0746***	0.0468
	Volatility Dummy	0.0011	-0.0072	-0.015
	Market Return Dummy	0.0156	-0.0236	-0.0056
China - Shanghai	Constant	0.0105	0.0259**	0.0303*
	Market Return	0.0001	0.0119**	0.003
	Volatility Dummy	-0.0004	-0.012	-0.0016
	Market Return Dummy	0.0019	-0.0035	0.0021
China - Hong Kong	Constant	0.0101	0.0272**	0.0328**
	Market Return	0.0128*	0.0715***	0.0483***
	Volatility Dummy	-0.0206	-0.0072	0.0112
	Market Return Dummy	0.0055	0.0208	0.0378*
* Significant at 10%				
** Significant at 5%				
*** Significant at 1%				

5. CONCLUSION

In the contemporary world, globalization plays a vital role in the development of the economy. The stock market movements are a major concern for the investors, governments, academicians, economists and economic policy makers. However stock markets are quite risky and are subject to crashes and crises. It is important to consider alternative investments which can either hedge stock market risk or provide a diversification benefit if hold together with stock portfolios. This paper examined the potential of precious metals i.e. gold, silver and platinum in hedging and diversification of risk of investment in stock market portfolios in the US, BRIC countries, Hong Kong and Pakistan using daily data for the period of 1995 to 2014. Following Hillier et al. (2006) two models were estimated. In the first model, the GARCH model was estimated with the mean equation that links precious metal returns to stock market index return. The second model allows a more informative analysis of whether the precious metal returns had significantly negative relationship with stock index returns during the normal stock market conditions and during the extremely adverse condition characterized by either high volatility or extremely low returns.

This research extended the analysis of Hillier et al. (2006) in two important ways. First whereas Hillier et al. (2006) included S&P 500 index and a composite index of Europe, Far East and Australia, this study investigated the investment potential of precious metals for several important emerging markets including China, India, Brazil and some smaller markets e.g. Pakistan. Secondly Hillier et al. (2006) characterized adverse stock market condition as the period when the market volatility is greater than two standard deviation above the mean volatility level. However, it is generally observed that the volatility estimated by the GARCH model has a highly positively skewed behaviour so mean and standard deviations are not very useful summary measures for the volatility series. Instead, this paper has used a quantiles based measure as employed in Baur & Lucey (2010) and Baur & McDermott (2010) where the extreme market conditions are characterized as the volatility falling in the top quantiles of its distribution e.g. top 5%.

Research indicates that over the period of time, correlations between the returns of these precious metals have increased from the average of 0.33 (1995-2004) to 0.56 (2005-2014). Correlations are positive but not close to being perfect thus providing opportunities for investment diversification.

It is found that in the normal market conditions, gold acts as a hedge against stock market risk only in the case of USA. In other markets during normal times, the precious metals have no or positive relationship with stocks. Thus there is some diversification benefit of holding a portfolio of precious metal with stocks.

It is found that during adverse market condition characterized by high market volatility gold acts as safe haven in US and Russia. In case of Russia silver and platinum are observed to be safe haven assets. Other economies do not possess such safe haven asset in high market volatility periods. In addition to this, under poor market performance (when market returns are at the lowest 5%), no precious metal acts as safe haven asset.

6. ACKNOWLEDGEMENT

Parents, teachers, colleagues, and mentors are undoubtedly a blessing. This research would not have been attainable if I were alone. Firstly, I would like to thank my supervisor, Dr. Javed Iqbal, for his continuous feedback, extensive knowledge, and motivation that made me complete this research. His intellectual advice, encouragement, and frequent meetings had been the guiding path. He has been there to listen, discuss, and resolve issues surrounding my research.

I feel privileged in having studied under finest instructors, namely Dr. Mohammed Nishat, Dr. Adnan Haider, Dr. Naved Ahmad, Dr. Farooq Pasha, Sadiqul Huda and Asim

Jahangir. I am also grateful to Institute of Business Administration in general and Mr. Muhammad Anwar (Head Librarian) in particular for providing a conducive research environment. All these people have supported me in completing this research. Without them it would have been very difficult or even impossible to materialize my final step towards degree completion.

During my study at IBA, I have formed strong friendship bond with Danish us Salam, Hasaan Anwar Mohiuddin, Pirzada Osama Qasmi, Muhammad Arsalan Jamil, and Ali Sajad. These guys have helped me get through the graduate program by their late-evening conversations that acted as stress relievers, early morning breakfasts on weekends that fuelled energy for the coming week, and those car-pooling rides filled with discussions about everything and anything. Due to them, I will cherish my graduate experience forever.

My immediate family had an important role to play in providing emotional, financial and spiritual support during my research. Love, care, prayers and a firm belief of my parents in me are worthy of deepest appreciation. Lastly, my fiancé, for being supportive, and ensuring that I stick to my target of completing research this year.

I would like to express my gratitude and thank you everyone who has been an important contributor to this research.

REFERENCES

1. Agarwal, S., Liu, C. and Rhee, G. (2007). Where does price discovery occur for stocks traded in multiple markets? Evidence from Hong Kong and London. *Journal of International Money and Finance,* 26(1), 46-63.
2. Agyei-Ampomah, S., Gounopoulos, D. and Mazouz, K. (2014). Does gold offer a better protection against losses in sovereign debt bonds than other metals? *Journal of Banking & Finance*, 40, 507-521.
3. Batten, J., Ciner, C. and Lucey, B. (2010). The macroeconomic determinants of volatility in precious metals markets. *Resources Policy*, 35(2), 65-71.
4. Batten, J., Ciner, C. and Lucey, B. (2015). Which precious metals spill over on which, when and why? Some evidence. *Applied Economics Letters*, 22(6), 466-473.
5. Baur, D. G. and Lucey, B. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Financial Review*, 45(2), 217-229.
6. Baur, D. (2013). The autumn effect of gold. *Research in International Business and Finance*, 27(1), 1-11.
7. Baur, D. and McDermott, T. (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance,* 34(8), 1886-1898.
8. Capie, F., Mills, T. and Wood, G. (2005). Gold as a hedge against the dollar. *Journal of International Financial Markets, Institutions and Money*, 15(4), 343-352.
9. Chen, D.-H., Blenman, L., Bin, F.-S. and Chen, J. (2011). The effects of open market reforms on the behavior of China’s stock prices. *Journal of Chinese Economics and Finance*, 1, 41-53.
10. Cheung, D. and Sami, H. (2000). Price and trading volume reaction: the case of Hong Kong companies’ earnings announcements. *Journal of International Accounting, Auditing and Taxation*, 9(1), 19-42.

11. Chng, M. and Foster, G. (2012). The implied convenience yield of precious metals: safe haven versus industrial usage. *Review of Futures Markets*, 20, 349-94.
12. Chua, J., Sick, G. and Woodward, R. (1990). Diversifying with gold stocks. *Financial Analysts Journal*, 46(4), 76-79.
13. Ciner, C., Gurdgiev, C. and Lucey, B. (2013). Hedges and safe havens: An examination of stocks, bonds, gold, oil and exchange rates. *International Review of Financial Analysis*, 29, 202-211.
14. Citak, S. (1999). World Gold Markets, Istanbul Gold Exchange and Gold Risk Management. *Istanbul Stock Exchange Review*, 3(12), 51-88.
15. Claessens, S. and Schmukler, S. (2007). International financial integration through equity markets: Which firms from which countries go global? *Journal of International Money and Finance*, 26(5), 788-813.
16. Conover, M., Jensen, G., Johnson, R. and Mercer, J. (2009). Can precious metals make your portfolio shine? *Journal of Investing*, 18(1), 75-86.
17. Coudert, V. and Raymond, H. (2011). Gold and financial assets: are there any safe havens in bear markets. *Economics Bulletin*, 31(2), 1613-1622.
18. CPM Group. (2014). *Silver Investment Demand.* New York: The Silver Institute.
19. Daskalaki, C. and Skiadopoulos, G. (2011). Should investors include commodities in their portfolios after all? New evidence. *Journal of Banking & Finance*, 35(10), 2606-2626.
20. Do, Q., Mcaleer, M. and Sriboonchitta, S. (2009). Effects of international gold market on stock exchange volatility: evidence from asean emerging stock markets. *Economics Bulletin*, 29(2), 599-610.
21. Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3), 339-350.
22. Gencer, G. and Kilic, E. (2013). Conditional Correlations and Volatility Links Among Gold, Oil and Istanbul Stock Exchange Sector Returns. *International Journal of Economics and Financial Issues*, 4(1), 170-82.
23. Gilmore, C., McManus, G., Sharma, R. and Tezel, A. (2009). The dynamics of gold prices, gold mining stock prices and stock market prices comovements. *Research in Applied Economics*, *1* (1), 1-19.
24. Gismatullin, E. and Nishizawa, K. (2014). *Shanghai-Hong Kong Link to Start in a Week as China Opens*. Retrieved June 8, 2015, from http://www.bloomberg.com/news/articles/2014-11-10/shanghai-hong-kong-stock-exchange-link-will-start-in-one-week
25. Hillier, D., Draper, P. and Faff, R. (2006). Do precious metals shine? An investment perspective. *Financial Analysts Journal*, 62(2), 98-106.
26. Ibrahim, M. and Baharom, A. (2011). The role of gold in financial investment: a Malaysian perspective. *International Center for Education in Islamic Finance*.
27. Jaffe, J. (1989). Gold and gold stocks as investments for institutional portfolios. *Financial Analysts Journal*, 45(2), 53-59.
28. Kutan, A. and Aksoy, T. (2004). Public information arrival and gold market returns in emerging markets: Evidence from the Istanbul Gold Exchange. *Scientific Journal of Administrative Development*, 2(1), 13-26.

29. Lee, C., Shie, F. and Chang, C. (2012). How close a relationship does a capital market have with other such markets? The case of Taiwan from the Asian financial crisis. *Pacific-Basin Finance Journal*, 20(3), 349-362.
30. Lucey, B. and Li, S. (2015). What precious metals act as safe havens, and when? Some US evidence. *Applied Economics Letters*, 22(1), 35-45.
31. Lucey, B. and MacKenzie, D. (2013, August 22). *Correlation Behaviour of iShares Silver Trust During the 2010 Flash Crash.* Retrieved April 13, 2014, from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2313667
32. Mulyadi, M. and Anwar, Y. (2012). Gold versus stock investment: An econometric analysis. *International Journal of Development and Sustainability*, 1(1), 1-7.
33. Nashier, T. (2015). Financial integration between BRICS and developed stock markets. *International Journal of Business and Management Invention*, 4(1), 65-71.
34. Pullen, T., Benson, K. and Faff, R. (2014). A comparative analysis of the investment characteristics of alternative gold assets. *Abacus*, 50(1), 76-92.
35. Sensoy, A. (2013). Dynamic relationship between precious metals. *Resources Policy*, 38(4), 504-511.
36. Smith, G. (2002). Tests of the random walk hypothesis for London gold prices. *Applied Economics Letters*, 9(10), 671-674.
37. Smith, G. (2001). The price of gold and stock price indices for the United States. *The World Gold Council*, 8, 1-16.
38. The Silver Institute. (2015). *Silver Essentials | The Silver Institute*. Retrieved May 20, 2015, from https://www.silverinstitute.org/site/silver-essentials/
39. Tully, E. and Lucey, B. (2007). A power GARCH examination of the gold market. *Research in International Business and Finance*, 21(2), 316-325.
40. Vermeulen, R. (2013). International diversification during the financial crisis: A blessing for equity investors? *Journal of International Money and Finance*, 35, 104-123.
41. Wang, K.-M., Lee, Y.-M. and Thi, T.-B. (2011). Time and place where gold acts as an inflation hedge: An application of long-run and short-run threshold model. *Economic Modelling*, 28(3), 806-819.
42. World Platinum Investment Council. (2015). *World Platinum Investment Council - Supply & Demand*. Retrieved May 20, 2015, from http://platinuminvestment.com/supply-and-demand
43. Worthington, A. and Pahlavani, M. (2007). Gold investment as an inflationary hedge: Cointegration evidence with allowance for endogenous structural breaks. *Applied Financial Economics Letters*, 3(4), 259-262.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 311-322

WAVELET CHARACTERIZATION OF DISCONTINUITIES IN ASTROPHYSICAL SIGNALS

M. Ayub Khan Yousuf Zai[1§], **Khusro Mian**[2]
and **M. Rashid Kamal Ansari**[3]

[1] University of Karachi, Karachi, Pakistan.
[2] Sir Syed University of Engineering and Technology, Karachi, Pakistan
[§] Corresponding author's Email: ayubzai@yahoo.com

ABSTRACT

In this communication we have utilized Wavelet analysis to detect the exact instant when a signal changes. The sites of the change have been identified with their amplitude. This clearly suggests the presence of high frequency information i.e. a sudden change or discontinuity. We have determined that the spikes approximation level a_1 to a_5 of perturbed, unperturbed & mean concentration and parcel velocity of ionospheric F_2 & E_S layers respectively at Pakistan air space. It has been found out that during the spikes of mean concentration & parcel velocity the unperturbed concentration and parcel velocity is very smooth. This process shows the corresponding perturbed concentration & parcel velocity while the spikes of mean concentration & parcel velocity at that instant show the strength of the signal. Similarly, at the instant when the spikes exist in unperturbed concentration & parcel velocity there is comparatively very smooth behavior obtained for the data set of mean concentration & parcel velocity. This clearly emphasizes that solar flare activity is responsible for perturbed concentration & parcel velocity.

KEY WORDS

Astrophysical Signals, Perturbation of concentration, parcel Velocity, Spikes detection.

1. INTRODUCTION

It is an established fact that the ionosphere efficiency is changing because of natural phenomena activities. In this situation one can well conceive that hazard of extinction of various communication and surveillance system because of ionospheric plasma turbulence (IPT). Wavelet transformation is often referred to as the most effective technique for signal processing as a preferable alternative to Fourier analysis, used when signals are random and comprised of fluctuations of different scales. It is mathematical microscopes. We have analyzed IPT in upper atmospheric plasma at Pakistan air space by spectral analysis, fast Fourier transformation and wavelet transformation. We have planned an idea to find out the unseen periodicities in singular systems. The influence of sunspot on day after day communication coordination of ionosphere creates turbulence in the form of Perturbed Electron concentration (PEC) and Perturbed Parcel Velocity (PPV). Literature review shows that the regularity of astrophysical signal follows the 22-year consecutive solar flare and sun spot cyclic. This regularity and magnetic disturbances vary with the solar activity [1-3].

We have investigated to examine the periodicities in the time of astrophysical signals. This is one of the important aspects of astrophysical signals. The random behavior of sun generates solar activity, which often associated with astrophysical signals. Turbulence is a highly unstable state of fluid flows, where by fluids, continuously movable and deformable liquids and plasma gas considered being fluids, when the scale of observation is much larger than the molecular mean free path.

Wavelet analysis is a method used to study non stationary processes with their data sets. Also it is utilized to quantify both signal of variability non stationary unification between plasma turbulence and solar activity. The wavelet transformation because use it measures the local scaling field, is the appropriate tool for verifying these different interpretation in relating the power low scaling of the energy spectrum to the shape of possible singularities. Wavelet theory is useful for identifying the discontinuities in the plasma layers. First through wavelet decomposition, we have identified the spikes in the data of IPT and observed the behavior of unperturbed & perturbed electron concentration and as well as unperturbed & perturbed parcel velocity at particular instant. After de-noising we have measured the amplitude of the dissimilarity in the data [4-6].

2. IONOSPHERIC PLASMA WAVE TURBULENCE

It plays an important role in ionosphere. It is the event that plasma and electromagnetic wave variation randomly exhibit. The electrons will exceed and fluctuate around their equilibrium positions with a characteristic frequency known as plasma frequency. This fluctuation is so fast that the massive ions do not have time to respond to the oscillating field and may be considered fixed. When the oscillation amplitude is small, the waveform is generally sinusoidal and there is only one component. The density of plasma wave is a sinusoidal oscillating quantity and thus can be represented as in Cartesian coordinates follows:

$$N = \overline{N} \bullet e^{i(k.\gamma - \omega t)}, k\gamma = k_1 x + k_2 x_2 + k_3 x_3, \tag{2.1}$$

Here $\overline{N}$ is a constant defining the amplitude of the plasma wave and k is the propagation constant. When one observes that waves have grown to larger steady amplitudes then linear theory does not exit and some nonlinear effect is limiting the amplitude. It has been explained that the amplitude at saturation is rather small and thus a wave experiences a number of changes when its amplitude becomes large. These change its shape from sine wave to triangular waveform. Thus Fourier components are at other frequencies or wave numbers are generated. A large wave can trap particles in its potential troughs thus changing the properties of the medium in which it propagates. Taking the case of nonlinear Landau damping, if plasma is so strongly excited that a continuous spectrum of frequencies is present, it is in the state of turbulence. This state must be described statistically as in the case of ordinary fluid hydrodynamics. Plasma wave turbulence leads to anomalous resistively in which the electrons are slowed down by collision with random electric field fluctuations. This effect is used for ohmic heating of plasma. The turbulence means crowd of vortices in nonlinear interaction. The structure of these vortices on the mechanism of their production by nonlinear instabilities in shear layers on the long range collective dynamics. Turbulence is highly unstable state of fluid flows like plasma in the present study [5-6].

3. REYNOLDS NUMBER

Turbulence is characterized by the Reynolds number, which is the ratio of the nonlinear inertial forces, responsible for the flow instability, to the linear dissipative damping. Another type of measurement of astrophysical turbulence is the wavelet Reynolds number, $\tilde{R}e(x,r)$, at each position x and scale r, defined as follows:

$$\tilde{R}e(x,r) = \frac{\bar{u}(x,r)r}{\upsilon} \tag{3.1}$$

where υ is the kinematics viscosity of the fluid and $\bar{u}$ is the mean square value of the velocity field contribution at position x and scale r defined as

$$\bar{u}(x,r) = \left(\frac{1}{C_\psi} \sum_{i=1}^{3} \left| \bar{u}_i(x,r) \right|^2 \right)^{1/2} \tag{3.2}$$

With the constant

$$C_\psi = \int_{IR^2} \left| \hat{\psi}(k) \right|^2 \frac{d^2k}{|k|^2} \tag{3.3}$$

We have focused on fully developed turbulence namely the limit of very large Reynolds number, which corresponds to either very large velocities, strong advection for flows encountered in hydraulics and naval engineering. Reynolds number, are of the order of aeronautics engines, airplanes, shuttle, meteorology, coherent, oceanography and astrophysics as shows in table I [3-5].

Table 1
Characterization of Reynolds Number in Solar Space

S#	Aspects	Reynolds Numbers
1	Aeronautics Engines	10^2 to 10^6
2	Airplanes	10^2 to 10^6
3	Shuttle	10^2 to 10^6
4	Meteorology	10^6 to 10^8
5	Coherent	10^7
6	Oceanography	10^8 to 10^{12}
7	Astrophysics	Larger than 10^{12}

4. FOURIER ANALYSIS

Fourier Analysis (FA) is mathematical tool to transform time signal into frequency domain, Fourier analysis loss time information in signal; it is drawback of Fourier analysis. Mathematical we can manifest $x(t) = (t + T)$

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt \tag{4.1}$$

Consider the Sine signals $f(t) = Sin(\omega_0 t) = Sin\left(\frac{2\pi}{T} t\right)$ which solves to

$$F(\omega) = \frac{2\pi}{\omega^2 - 4\pi^2}\left(e^{-j(+\infty)\omega} - e^{-j(-\infty)\omega}\right) \tag{4.2}$$

The amplitude spectrum of $F(\omega)$ is depicted in Fig. 4.1.

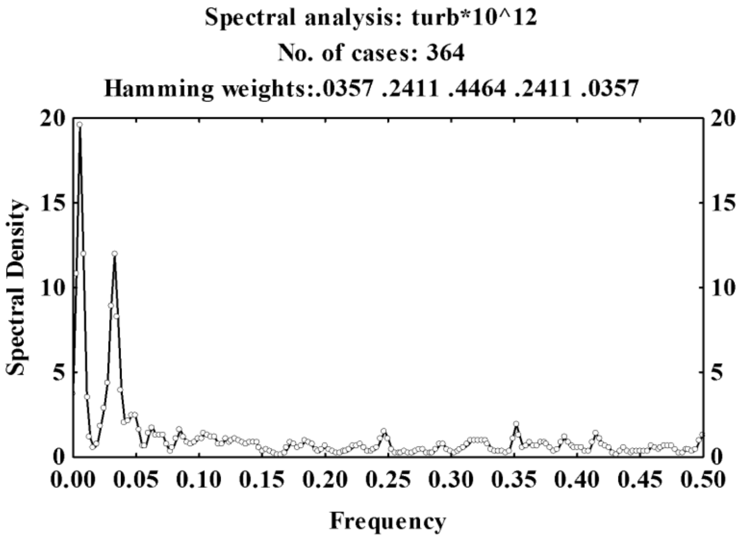


Fig. 4.1: The Amplitude Spectrum of $F(\omega)$

We have needed an infinite number of Sine waves to reconstruct original signal perfectly, or at least a large number of Sine waves to get close to the original.

Fourier analysis,

$$f(t) = \overline{f(t)} + \sum_P [A_P Cos(\omega_P t) + B_P Sin(\omega_P t)] \tag{4.3}$$

where $\overline{f(t)}$ is the mean value of the function, A_P, B_P are the coefficient of trigonometric function of $Sin(\omega_P t)$, $Cos(\omega_P t)$ and the angular frequency ω_P are integer p = 1, 2, 3… multiples of the total length of the time series.

We have presented the plasma turbulence in the form of astrophysical signals as depicted in Fig 4.1 and listed their parameters in Table 2. *Fourier analysis is one of the famous methods for identifying periodic components.*

Table 2
Illustration of IPT as Frequency (F), Period (P), Coefficient Cosine, Coefficient Sine and Period-o-Gram (Pgm)

F	P	cos	Sin	Pgm
0.0		0.003	0.0	0.002
0.002	364	-0.045	0.082	1.607
0.005	182	-0.437	-0.186	41.20
0.008	121	-0.088	-0.092	2.987
0.01	91	-0.071	0.099	2.734

5. WAVELETS APPROACH

It has been examined that wavelet analysis exhibits low and high frequency information for local behavior and also shows the exact location in time of discontinuity, *mathematically we can express*

$$C(a,b) = \int_{-\infty}^{+\infty} f(t)\psi(a,b,t)dt$$

(5.1)

where a, b are scale, position respectively. Wavelet characterization is a hottest approach of exploring turbulence in upper atmospheric plasma at Pakistan air space; we have presented histogram, cumulative histogram and statistics shown in Fig 5.1 and tables 5.1-5.2

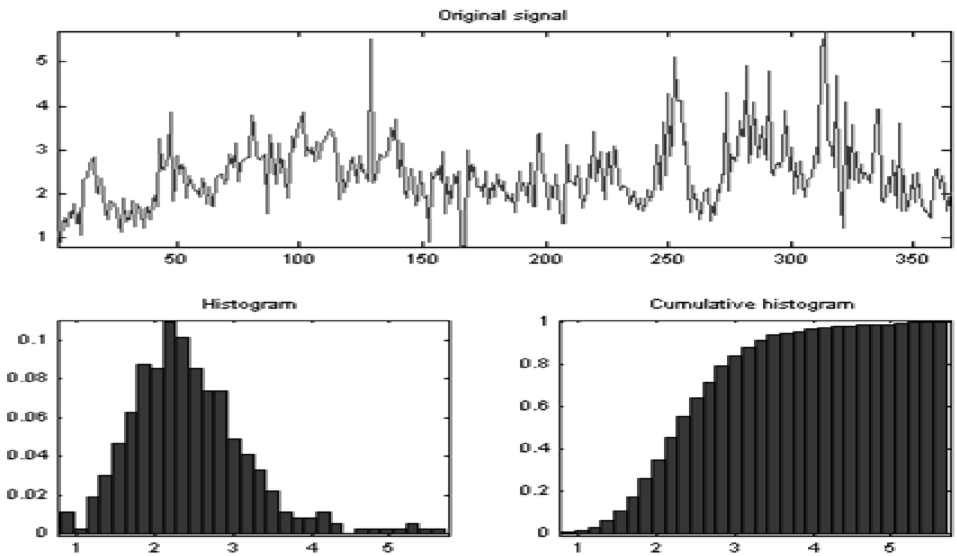


Fig. 5.1: Original Signal, Histogram & Cumulative Histogram Plot of Plasma Turbulence

Wavelet analysis decompose actual signal into different signals to be analyzed into principal and residual part. By mathematical:

$$s = a_j + \sum_j d_j \tag{5.2}$$

where s, a_j and d_j , are the signal, principal part j level and residual part j level.

The dyadic scale is $a = 2^j$ for level 5 the resolution is given by 1/a, or 2^{-j}. We have carried out wavelet analysis of turbulence in Fig. 5.2

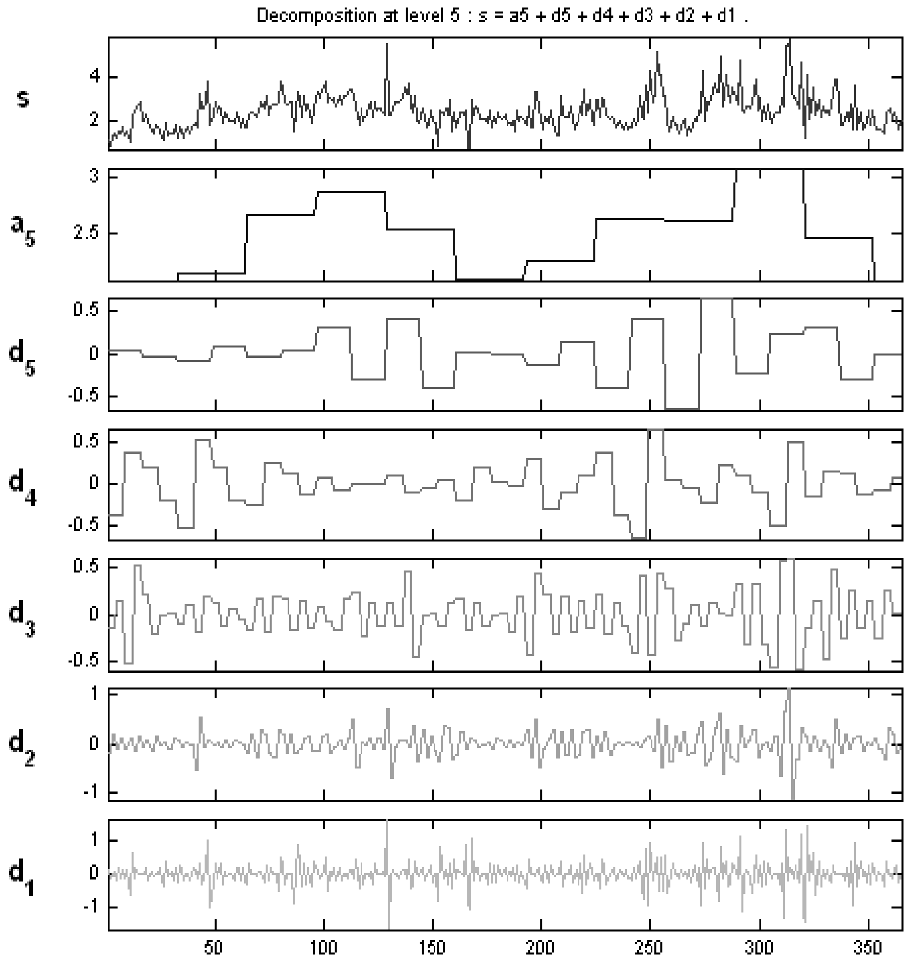


Fig. 5.2: Decomposition at Level 5, of Plasma Turbulence

The variation is presented in the form of different resolution at level-5 of wavelet type in the details and approximated part the cyclic variation is also presented at different level.

$$
\begin{aligned}
s &= a_1 + d_1 , \\
s &= a_2 + d_2 + d_1 , \\
s &= a_3 + d_3 + d_2 + d_1 , \\
s &= a_4 + d_4 + d_3 + d_2 + d_1 , \\
s &= a_5 + d_5 + d_4 + d_3 + d_2 + d_1
\end{aligned} \tag{5.3}
$$

We have constructed model IPT data. In the detailed and approximated parts at the lowest resolution several peaks are appeared. The inequality of wavelet in continuous time series function:

$$
f(t) = \left\{ \begin{array}{c} 1; 0 \le t < (1/2)t_0 \\ -1; (1/2)t_0 \le t < t_0 \\ 0; otherwise \end{array} \right\} \tag{5.4}
$$

The IPT data are non-linear and local minimum in this period, which shows a positive skew ness of frequency histogram [6-7].

6. WAVELET AND ITS 1-D CHARACTERISTIC

Wavelet analysis and its one-dimension characteristic have been observed that the astrophysical signals live in the space with noise. Model circumstances are such that this noise may diminish to certain levels, it damages the signal and it must be detached in array to improve the preferred signals and proceed with additional data analysis. This noise removal takes place in the original signal or in a transform domain. The signal is taken in some v_{j} , state and then mathematical decomposition as

$$
l^2(R) = v_j \oplus (\underset{j \ge j_0}{\oplus} w_j) , \tag{6.1}
$$

Multi resolution emphasizing the normalization factor $a^{-1/2}$ normalization for small scales or high frequency containing the singularities of the signal and the choice $a^{-1/2}$ makes the transform unitary l^2 norm is interpreted

$$
v_j = v_{j_0} \oplus (\bigoplus_{j=j_0}^{j=1} w_j) , \tag{6.2}
$$

Fig. 5.2 shows decomposition of signal, approximation and details order 5,

$$
v_0 = v_{-5} \oplus w_{-5} \oplus w_{-4} \oplus w_{-3} \oplus w_{-2} \oplus w_{-1} \tag{6.3}
$$

As we just saw, appropriate filters generate orthogonal wavelets bases. The signal s lives in v_0 and it is decomposed into its approximation $a_5 \in v_{-5}$ and the increasingly finer details $d_j \in w_{-j}, j = 1,2,3,4,5.$ in Fig. 6.1 - 4.8. We located, from first to last, the raw signal s, the approximation at level 5, a_5 and details from the coarsest level d_5 to the finest level d_1.

Fig. 6.1 Fig. 6.2 depict the upper approximation level 1 & 2 and lower panel details level 1&2 and of plasma turbulence

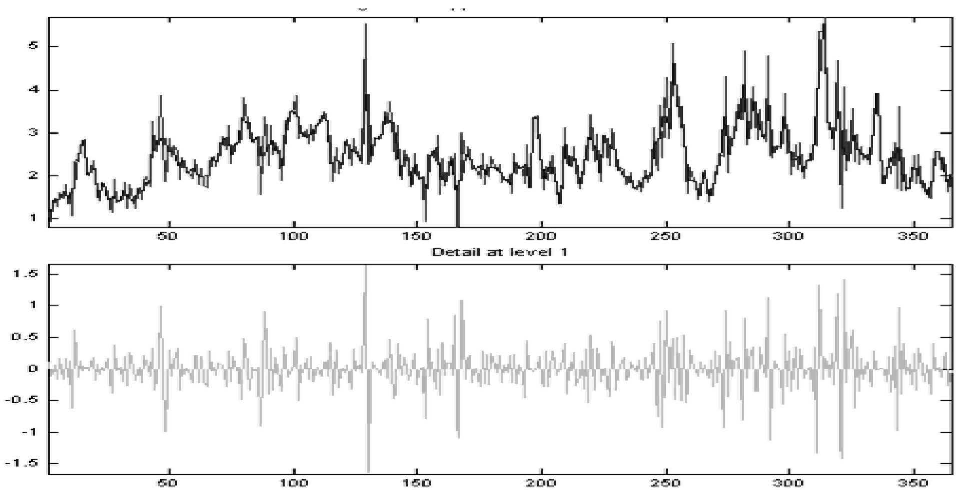


Fig. 6.1

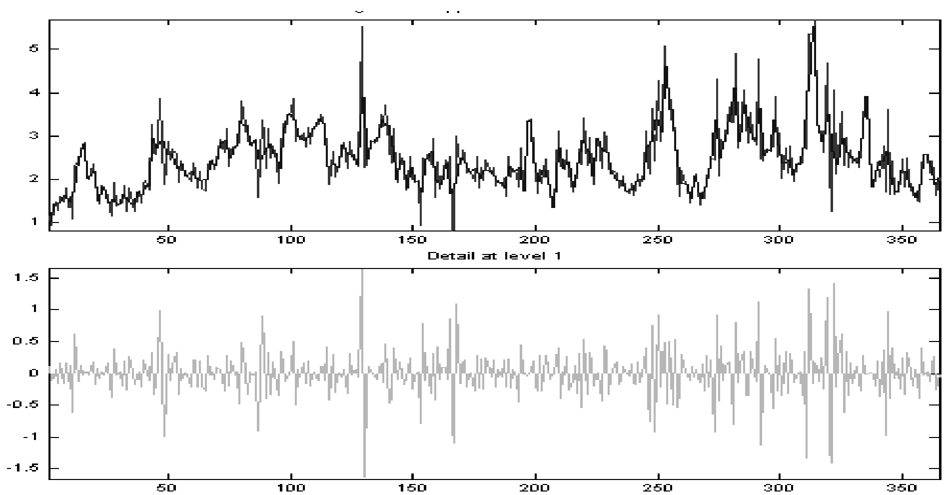


Fig. 6.2

Figures 6.3. and 6.4. display upper approximation level 3 & 4 and lower panel details level 3&4 of plasma turbulence

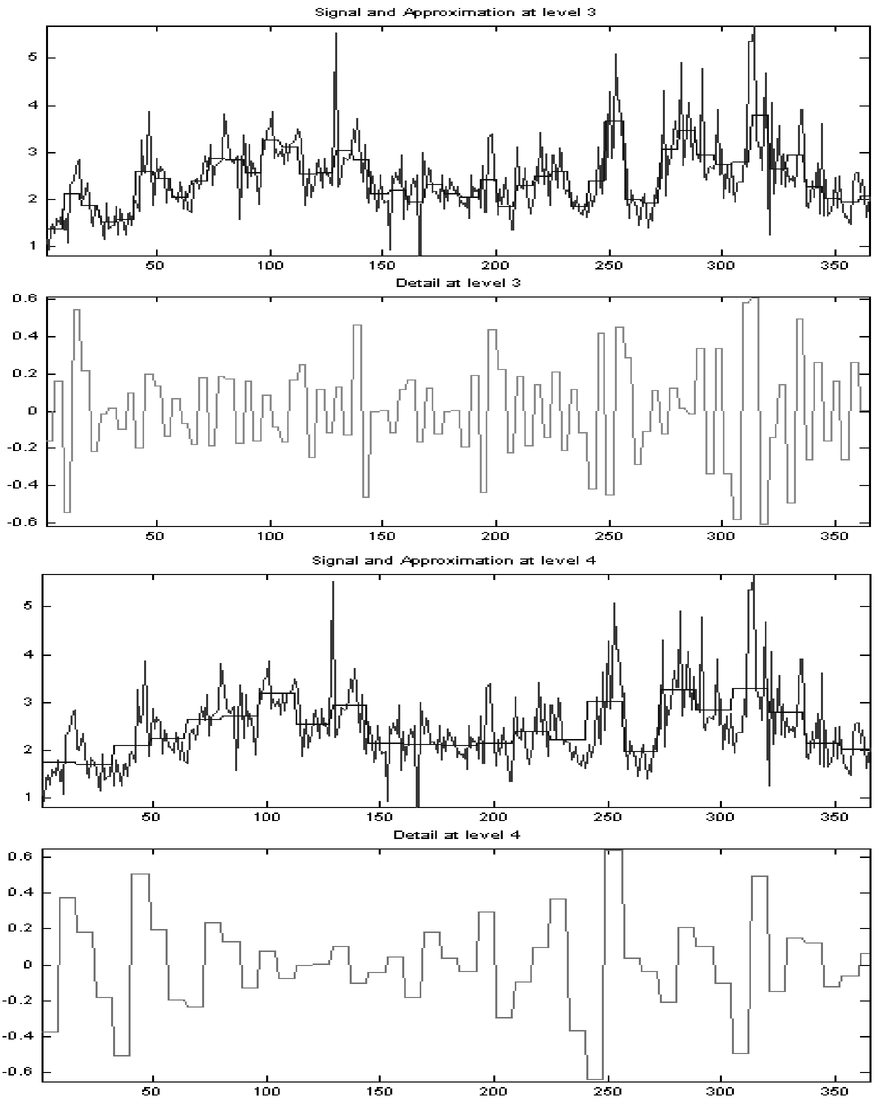


Figure 6.5 shows Upper panel approximation level 5 and lower panel details level 5, of plasma turbulence

We can thus state that d_1 contains the components of the signals of periods between d_1 to d_5. The study makes it achievable to path possible outliers, which are detected to the very high values of d_1 around the position 125 and 325. We can distinguish the details d_1 and d_2 measurement and state noises, which gives in details oscillating to zero. In detail d_3 also have noise at the position 300 to 325. We can not distinguish periods in the details d_4 and d_5.

7. INVESTIGATING AND RECOGNIZING DISCONTINUITIES IN IPT

Electromagnetic waves are reflected at the bounding surface between two media with different refractive indices. There is then a discontinuity in the plasma turbulence. It is now of interest to ask weather reflection can occur at a surface where the IPT is continues, but the gradient of the turbulence is discontinuous. We have presented an idea to find discontinuities in different processes quantifying the astrophysical plasma turbulence that was analyzed by wavelet transformation. A decomposition of the signal $x(t)$ onto a family of basic functions in the time scale plane $h_{t,a}(u)$, t and a, are the translation and the scale variables.

$$WT_x(t,f) = \int x(u) h_{t,f}(u) du \,, \; h_{t,f}(u) = \left| \frac{f}{f_0} \right|^{1/2} h\left[(u-t)\frac{f}{f_0} \right] \tag{7.1}$$

The decomposition result in time components, each one computed for different scales. Leakage as due to discontinuities in the time domain result in high frequency components and spread of the spectrum as predicted from the Gibbs effect [8-10].

Wavelet can identify the accurate movement, when a signal changes. The sporadic signal of IPT consists of slow Sin wave suddenly followed by medium Sin wave. The first & second level detail d_1 and d_2 illustrate the discontinuities some piece evidently, as the break contains the high-frequency part. The discontinuity is localized very precisely, merely little area just about time = 140 & 150 (positions) contains in d_1 and d_2 respectively as depicted in Fig. 7.1

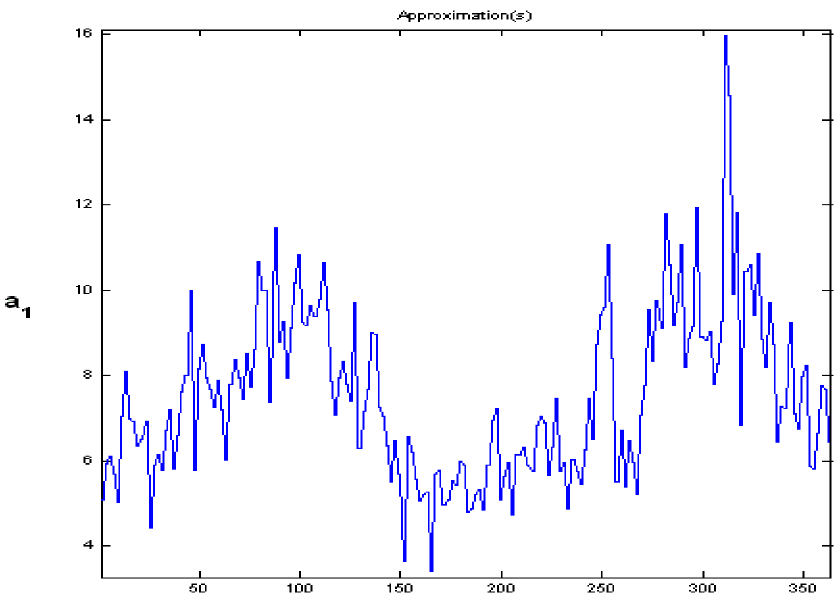


Fig. 7.1

Details d_3 and d_4 have the medium sine wave. The slow Sin wave is obviously distant in approximation a_5, from which the higher frequency filtered. The deterministic fraction of the signal may experience sudden changes as depicted in Fig. 7.1

Wavelet analysis can be utilized to detect the exact instant when a signal changes. The sites of the change have been identified with their amplitude. This suggests the presence of high frequency information i.e. a sudden change or discontinuity. We determined that the spikes approximation level a_1 to a_5 of perturbed, unperturbed & mean concentration and parcel velocity of ionospheric F_2 & E_S layers. This shows the corresponding perturbed concentration & perturbed parcel velocity while the spikes of mean concentration & mean parcel velocity at that instant shows the strength of the signal. Similarly, at the instant when the spikes exist in unperturbed concentration & parcel velocity. There is comparatively very smooth behavior obtained for the data set of mean concentration & parcel velocity. This clearly emphasizes that solar flare activity is responsible for perturbed concentration & parcel velocity. The presence of high-frequency information i.e. a sudden change or discontinuity with its position and amplitude are illustrated in Table 7.1 [11-12].

Table 7.1
Wavelet decomposition results shows statistic of spikes
For Unperturbed, Perturbed & Mean concentration for F_2 layer
P= Position , and A= Amplitude

Unperturbed		**Mean**		**Perturbed**		**Amplitude difference b/w Unperturbed & Mean**	**Amplitude difference b/w perturbed & Mean**
P	**A**	**P**	**A**	**P**	**A**		
50	207.6	50	203.9	50	3.5	-3.7	200.4
100	207	100	204.1	100	3.4	-2.9	200.7
150	207.8	150	204.4	150	3.4	-3.4	201
200	208.3	200	204.4	200	3.8	-3.9	200.6
250	209.3	250	204.9	250	4.8	-4.4	200.1
300	206.6	300	203.8	300	2.8	-2.8	201
350	206.5	350	203.3	350	2.7	-3.2	200.6

8. COMMENTS AND CONCLUSION

Plasma turbulence was analyzed using Fourier analysis, distinguishing astrophysical signals, wavelet and its 1-D characteristic. We have also explained investigating and recognizing discontinuities in IPT, continuous wavelet transformation and long term signals manifestation. 1-D wavelet packet, de-noising astrophysical signals. This work was based on the data analysis of unperturbed perturbed parcel velocity, unperturbed perturbed electron concentration and plasma turbulence for the same period.

We have found the behavior of ionosphere in response to parcel velocity and electron concentration. Solar activity, sunspot, 11 years solar cyclic, solar winds and coronal mass ejection are all the consequences of plasma turbulence. Wavelet models were constructed to establish the nature of unperturbed perturbed parcel velocity, unperturbed & perturbed electron concentration and plasma turbulence. All of these models are selected after diagnostic techniques and their applications.

7. ACKNOWLEDGEMENT

The authors would like to thank the organizers of 14th International Conference on Statistical Sciences for providing opportunity to deliver this piece of information before this learned gathering.

SUPARCO technical staff is highly appreciated for providing Data for ionospheric variables.

REFERENCES

1. Makridakis, S. (1983). Forcasting Methods and application, 2nd Ed, 1983, John Wiley and Sons, Inc. Canada.
2. Yousuf Zai, M.A.K., Ansari, M.R.K. and Jawaid Iqbal, J. (2012). Modeling and Simulating the Effects of Ozone Layer Depletion on Climate Change at Pakistan Air Space. *Proceedings of the Pakistan Academy of Sciences,* 49(2), 145-152.
3. Namgaladze, A.A., Zolotov, O.V., Zakharenkova, I.E., Shagimuratov, I.I. and Martynenko, O.V. (2009). Ionospheric total electron content variations observed before earthquakes: Possible physical mechanism and modeling. *Proceedings of the MSTU*, 12, 308-315. http://goo.gl/A8cLx
4. Zinke-Allmang, Martin, (2003). *Physics for life sciences*, 2nd Ed, Nelson, a division of Thomson Limited Canada
5. Zhanga, S.R. Fukaoa, S., Oliverc, W.L. and Otsukaa, Y. (1999). The height of the maximum ionospheric electron density over the MU radar. *Journal of Atmospheric and Solar-Terrestrial Physics*, 61, 1367-1383.
6. Radicella, S.M. and Zhang, M.L. (1995). The Improved DGR analytical model of electron density height profile and total electron content in the ionosphere. *ANNALI DI GEOFISICA*, XXXVIII, 35-41.
7. Altaisky, M.V. (2005). *Wavelets theory applications implementation*, Universities Press India ppw-152.
8. Frazier, M.W. (1999). An Introduction to Wavelets through Linear Algebra, Springer Veriag New York.
9. Misiti, M., Misiti, Y. Oppenheim, G. and Poggi, J.M. (2007). Wavelets and their Applications. ISTE, Ltd. USA, p.13-23.
10. J.C. Van Den Berg, (2007). Wavelets in PHYSICS, Cambridge University Press UK.
11. Horton, W. Plasma Turbulence Institute for Fusion Studies, The University of Texas, Austin, Texas 78712, Globalstar LP, P.O. Box 640670, 3200 Zanker Road, San Jose, CA 951640670.
12. Hong, D. (2005). *Real Analysis with an Introduction to Wavelets and Applications*, P-123-213, Elsevier Inc. India.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 323-334

A FRACTIONAL NUMERICAL SOLUTION AND STABILITY ANALYSIS OF FACE BOOK USERS MATHEMATICAL MODEL

M. Khalid§ and **Fareeha Sami Khan**
Department of Mathematical Sciences, Federal Urdu University
for Arts, Sciences and Technology, Karachi, Pakistan
§Email: khalidsiddiqui@fuuast.edu.pk

ABSTRACT

Everything has its own impact, either positive or negative. The same idea can be applied to internet usage. In this research, the usage of Facebook is addressed. Nowadays, Facebook usage has increased tremendously as compared to that of other social applications. Facebook is used by almost everyone e.g. kids, teenagers, youngsters, adults. This social interaction via Facebook has escalated into a second habit to the extent of levels of addiction which has begun to take serious toll on our lives, our happiness, our relationships, and our goals. A distinct approach to study Facebook users is being presented through a mathematical model primarily dealing with four kinds of users. The global stability of this mathematical model will be discussed under Lyapunov stability theory. Confirmation of this model will be substantiated through numerical solution of integer order as well as of its fractional order. Graphical solutions can be used to easily predict the potential behavior of Facebook users. Its equilibrium points will be discussed mathematically.

1. INTRODUCTION

Usage of Facebook has been increased rampantly in comparison to other social applications. Facebook users exist in every age group. This source of social interaction is grown into a habit and then it matured into addiction at the cost our lives, happiness, relationships and goals [1]. What is addiction, neurologically? It is consistent craving for pleasure and functionally speaking, in the centre of brain there is a component named "Nucleus accumbens". This component senses and translates various humanly joys, like deliciousness of food, power of money, pleasure of beauty etc. So when someone uses Facebook, this part of brain gets activated and began issuing pleasure signals to that human, the same manner as a drug addict gets pleasure while taking drugs and gets oblivion of the realities of life temporarily. Almost every type of professional utilizes Facebook for diverse purposes e.g. **researchers** use them for scientific researches, accesses to the research documents, to discuss ideas. **Poets and Writers** use them for promotion of their work, books, sharing their writings and expand their poetry etc. **Women** use this platform for fashion, cooking, articles about kids, techniques to make their household better. In short, innumerous subjects and pages have plethora of viewers and fans. Many researchers who study the utilization of internet, have shown consensus in their studies that Facebook is the most used social networking site world over as compared to Twitter, Myspace, Linkedin or another social networking site [3]. Till December 2015, 1.59 billion users have been recorded of Facebook. Fractional differential equations are the best way to analyze such type of models as they have the capability to observe the tiniest change in the system. It's not the first time such a mathematical model is being constructed. Abdullah [2] has used fractional differential equations to explain the Michaelis–Menten reaction in a 2-d region containing obstacles. Craiem et al. [4] used Fractional calculus to apply on arterial viscoelasticity. Smoking, which is a worldwide problem causing serious issues for lives, is also been studied via fractional order system of differential equations [5].

This work is orchestrated among five sections. The first section is about a few important definitions and theorems of fractional calculus. Second is related to the formation of fractional mathematical model and its description. Third is about the stability analysis of the fractional mathematical model with Lyapunov direct method for fractional systems. Then state space representation and analysis along with its numerical simulation is provided in fourth section. Lastly the conclusion and discussion over this paper are presented in section five.

2. IMPORTANT RESULTS

Some results that will be used in this work are presented here [6].

Def 1: Riemann-Liouville Fractional Derivatives:

The Riemann-Liouville definition of fractional derivative of the order α in the following form is

$$ {}_aD_t^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \frac{d^n}{dt^n} \int_a^t \frac{f(\tau)}{(t-\tau)^{\alpha+1}} d\tau \text{ for } a \in R,\ \alpha < 0 \tag{1} $$

where Γ is the Euler Gamma function.

Def 2: Caputo Fractional Derivative:

The Caputo Fractional Derivative is written as

$$ {}_aD_t^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \int_a^t \frac{f^{(n)}(\tau)}{(t-\tau)^{\alpha-n+1}} d\tau \text{ for } n-1 < \alpha \le n \tag{2} $$

Def 3: The autonomous system

$$ D^{\alpha} x = Ax \text{ with } x(0) = x_o \text{ is said to be stable iff } \forall x_o, \exists A, \forall t \ge 0, \|x(t)\| \le A \tag{3} $$

and asymptotically stable iff $\lim\limits_{t \to \infty} \|x(t)\| = 0$

Theorem 1:

The following asymptotic equivalents for $E_{\alpha}^{j}(\lambda, t)$ as $t \to \infty$ for $\left| \arg(\lambda) \le \frac{\alpha\pi}{2} \right|$

$$ E_{\alpha}^{j}(\lambda, t) \sim \frac{1}{\alpha(j-1)!} \left\{ \left(\frac{d}{d\sigma} \right)^{j-1} e^{\frac{\sigma t}{\alpha}} \right\}_{\sigma=\lambda} \tag{4} $$

It has the structure of a polynomial of degree $j-1$ in t, multiplied by $e^{\frac{\sigma t}{\alpha}}$.

for $\left|\arg(\lambda) \le \frac{\alpha\pi}{2}\right|$, $\quad E_{\alpha}^{j}(\lambda, t) \sim \frac{1}{\Gamma(1-\alpha)}(-\lambda)^{-j} t^{-\alpha}$ which decays slowly towards 0.

Theorem 2:

Let the origin $x = 0 \in D \subset R^n$ be an equilibrium point for $D^{\alpha} x = f(x)$. Let $V : D \to R$ be a continuously differentiable function such that $V(0) = 0$, $V(x) > 0$, $\forall x \in D \setminus \{0\}$, $D^{\alpha} x \le 0 \ \forall x \in D$ then $x = 0$ is stable. Moreover, if $D^{\alpha} x < 0 \ \forall x \in D \setminus \{0\}$ then $x = 0$ is asymptotically stable. Its proof can be found in [7].

3. MATHEMATICAL MODEL

A new fractional mathematical model is assumed here with four differential equations. $H(t)$: Ultra High Active Users: This label is considered for such cases that are highly addicted such as they check their timeline once in an hour or daily upload their updates, pictures, activities etc. $S(t)$: A Moderately High i.e. a bit less active than highly active ones: In this category such cases are considered that check their timeline daily and update their activities once a week. $L(t)$: Normal Users: In these cases such people are considered that check their timeline very often, once in a week or thrice of a month. $T(t)$: Recovered i.e. from high usage to normal or nominal usage, but can get addicted again. Typically in this category, falls the cases who had once got addicted and been given psychological or medical help. Having been cured, there is a fair likelihood that they can get addicted within short span again as they start company with highly active users. This Facebook fever is so widespread in our society that anyone can get easily addicted to it.

Here $N(t)$ is the number of total population. Whereas this total population is the sum of all above sub categories i.e.

$$N(t) = H(t) + S(t) + L(t) + T(t) \tag{5}$$

The following system of fractional differential equation has been obtained

$$\begin{aligned}
\frac{d^{\alpha} H}{dt^{\alpha}} &= -\delta H + bHT + eHS + fHL \\
\frac{d^{\alpha} S}{dt^{\alpha}} &= -\delta S - eHS + gST \\
\frac{d^{\alpha} L}{dt^{\alpha}} &= \delta + \delta L - jLT - fHL \\
\frac{d^{\alpha} T}{dt^{\alpha}} &= \delta T - bHT + jLT - gST
\end{aligned} \tag{6}$$

where contact rates between subclass Facebook users is given as:

Users Contact Rate = Total number of users interaction per unit time $\times$ Probability of addiction. These parameters are all positive b, e, f, g and j. Also δ: the per capita birth and death rates.

The considered subclasses have a fixed ratio of population entering into each subclass and that ratio is given by $p : 1-p$. This system has three most Facebook affected users i.e. $H(t)$, $S(t)$ and $T(t)$ and there is only one subclass which acts normal i.e. $L(t)$. Mathematically, the addiction free steady state can be written as $H = S = T = 0 \Rightarrow L = N$. From this an equilibrium point can be evaluated that is $(0,0,1,0)$. By putting the fractional derivative term equal to zero in (6), another equilibrium point can be obtained as

$$\frac{\delta}{fg + ej}(g - j, -b + f + j, -b - e + g, f + e). \tag{7}$$

4. STABILITY ANALYSIS

In this paper Lyapunov second method or Lyapunov direct method is being used to analyze the stability of this system. According to Lyapunov stability analysis, the system to be asymptotically stable it should be bounded as fractional-order extension of Lyapunov direct method [8]. To show that, let

$$\frac{d^{\alpha}H}{dt^{\alpha}} + \frac{d^{\alpha}S}{dt^{\alpha}} + \frac{d^{\alpha}L}{dt^{\alpha}} + \frac{d^{\alpha}T}{dt^{\alpha}} = \delta - \delta(H + S + L + T) \tag{8}$$

or this implies that

$$\frac{d^{\alpha}H}{dt^{\alpha}} + \frac{d^{\alpha}S}{dt^{\alpha}} + \frac{d^{\alpha}L}{dt^{\alpha}} + \frac{d^{\alpha}T}{dt^{\alpha}} \le \delta - \delta(H + S + L + T) \tag{9}$$

As we know that $N(t) = H(t) + S(t) + L(t) + T(t)$, therefore, $N^{\alpha}(t) \le \delta - \delta(N(t))$. The initial value problem $F^{\alpha} = \delta - \delta(F)$ with $F(0) = N(0)$, on solving with Laplace transform has the general solution $F(x) = 1 - E_{\alpha}(-\delta x^{\alpha}) + \Gamma(1+\alpha)F(0)$ where $E_{\alpha}(-\delta x^{\alpha})$ denotes the Generalized Mittag- Leffler function [6]. If $F(0) > 0$ this implies that $\lim_{t\to\infty} F(x) = 1$. Therefore $N(x) \le F(x)$ which implies that $\lim_{t\to\infty} \sup N(x) \le 1$. Thus the considered system is bounded such as $\{(H, S, L, T) : H + S + L + T \le 1, H \ge 0, S \ge 0, L \ge 0, T \ge 0\}$. This implies that the system is Mittag Leffler stable [9]. Consider the Lyapunov Function of this system as

$$V(t) = \lambda_1 H^2(t) + \lambda_2 S^2(t) + \lambda_3 L^2(t) + \lambda_4 T^2(t)$$
$$D_t^{\alpha} V(t) = D_t^{\alpha}\left(\lambda_1 H^2(t) + \lambda_2 S^2(t) + \lambda_3 L^2(t) + \lambda_4 T^2(t)\right) \tag{10}$$

On solving

$$\begin{aligned} D_t^{\alpha} V(t) = {} & (bHT + eHS + fHL - \delta H)^2 \lambda_1 + (-\delta S - eHS + gST)^2 \lambda_2 \\ & + (\delta + \delta L - jLT - fHLH)^2 \lambda_3 + (\delta T - bHT + jLT - gST)^2 \lambda_4 \end{aligned} \tag{11}$$

which clearly shows that $D_t^{\alpha}V(t)>0$. Then according to Theorem 2 this system is unstable around origin. Assume the Fractional order system of differential equations is

$$\begin{aligned}
\frac{d^{\alpha}H}{dt^{\alpha}} &= A(x) = -\delta H + bHT + eHS + fHL \\
\frac{d^{\alpha}S}{dt^{\alpha}} &= B(x) = -\delta S - eHS + gST \\
\frac{d^{\alpha}L}{dt^{\alpha}} &= C(x) = \delta + \delta L - jLT - fHL \\
\frac{d^{\alpha}T}{dt^{\alpha}} &= P(x) = \delta T - bHT + jLT - gST
\end{aligned} \tag{12}$$

where $X=(H,S,L,T)$. On disturbing the equilibrium point by adding positive terms η_1,η_2,η_3 and η_4 such that $H(t)-\hat{H}=\eta_1(t)$, $S(t)-\hat{S}=\eta_2(t)$, $L(t)-\hat{L}=\eta_3(t)$, $T(t)-\hat{T}=\eta_4(t)$. Also assume $\hat{X}=(\hat{H},\hat{S},\hat{L},\hat{T})$. Then it gives

$$\begin{aligned}
D^{\alpha}(\eta_1(t)) &= A(\hat{X}) + \frac{\partial A(\hat{X})}{\partial H}\eta_1(t) + \frac{\partial A(\hat{X})}{\partial S}\eta_2(t) + \frac{\partial A(\hat{X})}{\partial L}\eta_3(t) + \frac{\partial A(\hat{X})}{\partial T}\eta_4(t) \\
D^{\alpha}(\eta_1(t)) &= B(\hat{X}) + \frac{\partial B(\hat{X})}{\partial H}\eta_1(t) + \frac{\partial B(\hat{X})}{\partial S}\eta_2(t) + \frac{\partial B(\hat{X})}{\partial L}\eta_3(t) + \frac{\partial B(\hat{X})}{\partial T}\eta_4(t) \\
D^{\alpha}(\eta_1(t)) &= C(\hat{X}) + \frac{\partial C(\hat{X})}{\partial H}\eta_1(t) + \frac{\partial C(\hat{X})}{\partial S}\eta_2(t) + \frac{\partial C(\hat{X})}{\partial L}\eta_3(t) + \frac{\partial C(\hat{X})}{\partial T}\eta_4(t) \\
D^{\alpha}(\eta_1(t)) &= P(\hat{X}) + \frac{\partial P(\hat{X})}{\partial H}\eta_1(t) + \frac{\partial P(\hat{X})}{\partial S}\eta_2(t) + \frac{\partial P(\hat{X})}{\partial L}\eta_3(t) + \frac{\partial P(\hat{X})}{\partial T}\eta_4(t)
\end{aligned} \tag{13}$$

Here using the fact that $A(\hat{X})=B(\hat{X})=C(\hat{X})=P(\hat{X})=0$ and obtained a linearized system about the equilibrium system that is

$$D^{\alpha}(V)=\hat{J}V \tag{14}$$

where $V=(\eta_1(t),\eta_2(t),\eta_3(t),\eta_4(t))$ and $\hat{J}$ is the fractional Jacobian matrix evaluated at the equilibrium. We have $Q^{-1}\hat{J}Q=R$ where R is the diagonal matrix for the Jacobian matrix $\hat{J}$ given as

$$\hat{J}=\begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix} \tag{15}$$

where $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are the Eigen values and Q is the eigenvectors of $\hat{J}$. This gives

$$\begin{cases} D^{\alpha}(\varepsilon_1) = \lambda_1 \varepsilon_1 \\ D^{\alpha}(\varepsilon_2) = \lambda_2 \varepsilon_2 \\ D^{\alpha}(\varepsilon_3) = \lambda_3 \varepsilon_3 \\ D^{\alpha}(\varepsilon_4) = \lambda_4 \varepsilon_4 \end{cases} \tag{16}$$

Here $\varepsilon = [\varepsilon_1 \quad \varepsilon_2 \quad \varepsilon_3 \quad \varepsilon_4]^t$ also its solutions are given by Mittag Leffler function as

$$\varepsilon_i(t) = \sum_{n=i}^{\infty} \frac{t^{n\alpha}\lambda_i^n}{\Gamma(n\alpha+1)} \varepsilon_i(0) = E_{\alpha}(\lambda_i t^{\alpha})\varepsilon_i(0), \; i = 0,\ldots,4 \tag{17}$$

Def 3 and theorem 1 implies that if $|2\arg(\lambda_i) > \alpha\pi|$, $i = 1,2,3,4$ then $\varepsilon_i(t)$ are decreasing and therefore are also $\eta_i(t)$ decreasing. Thus, let the solution $(\eta_1(t), \eta_2(t), \eta_3(t), \eta_4(t))$ of (14) exists. If the solution of (14) is increasing then $(\hat{H}, \hat{S}, \hat{L}, \hat{T})$ is unstable and if $(\eta_1(t), \eta_2(t), \eta_3(t), \eta_4(t))$ is decreasing, then the equilibrium point $(\hat{H}, \hat{S}, \hat{L}, \hat{T})$ is locally asymptotically stable.

5. STATE SPACE ANALYSIS

Consider system (6) with the initial conditions H (0)>0, S (0)>0, L (0)>0, T (0)>0. By converting this system into State Space model it becomes

$$D^{\alpha}\vec{F}(t) = A\vec{F}(t) + B\vec{U} \text{ and } Y(t) = C\vec{F}(t) \tag{18}$$

with $\vec{F}(0) = f_o$, where $\vec{F}(t) = [H(t) \quad S(t) \quad L(t) \quad T(t)]^t \in R^n$, $\vec{U}(t) \in R^{inp}$ and $Y(t) \in R^{out}$ are state, input and output vector of the system and $A \in R^{n\times n}$, $B \in R^{n\times inp}$ and $C \in R^{n\times out}$. Also (A, B) is controllable and $f_o = [H(0) \quad S(0) \quad L(0) \quad T(0)]^t$. If expanded it becomes

$$\begin{bmatrix} D^{\alpha}H(t) \\ D^{\alpha}S(t) \\ D^{\alpha}L(t) \\ D^{\alpha}T(t) \end{bmatrix} = \begin{bmatrix} -\delta & eH & fH & bH \\ -eS & -\delta & 0 & gS \\ -fL & \delta/S & \delta & -jL \\ -bT & -gT & jT & \delta \end{bmatrix} \begin{bmatrix} H \\ S \\ L \\ T \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \vec{F} \tag{19}$$

It is well established in literature [10] that the concepts established for integer order space state models also hold for fractional systems. So, the system (19) is controllable if the rank of the controllability matrix $C = [BABA^2B..A^{n-1}B]$ is equal to n. Also this system is asymptotically stable as $|2\arg(\text{eigen values}(A)) > \alpha\pi|$. And therefore both techniques obtain the same result.

6. NUMERICAL SIMULATIONS

To verify our findings, data has been obtained from a research project based on Pew Research Centre [3]. From this, the parameters are assumed that $\delta = 0.05$, $b = 0.03$, $e = 0.02$, $f = 0.1$, $j = 0.2$, $g = 0.3$. The initial conditions obtained are $H(0) = 0.52$, $S(0) = 0.15$, $L(0) = 0.28$, $T(0) = 0.06$. The calculated eigen values are given as $\lambda_1 = -0.792503$, $\lambda_2 = -2.28108\,i$, $\lambda_3 = 2.28108\,i$, $\lambda_4 = 0.792503$. According to the study in [11] these eigen values shows that $|2\arg(\text{eigen values}(\lambda_i))| > \alpha\pi$ so it's an unstable node and for complex eigen values its an unstable focus. Based on these calculations figures obtained are

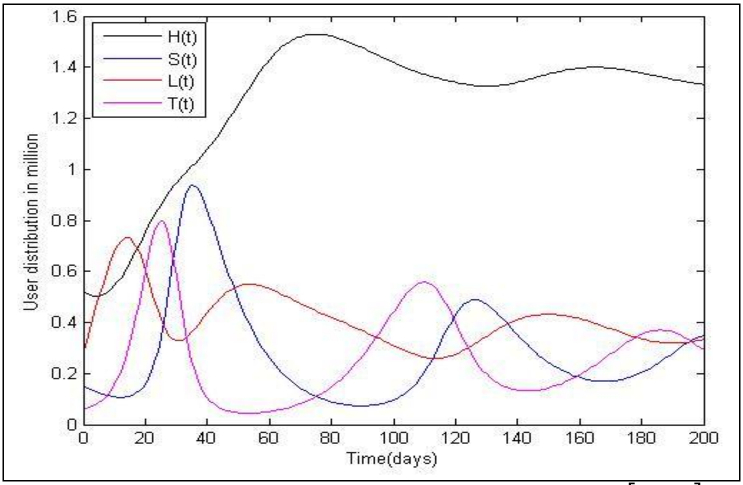


Figure 1: Behavior of Whole System at $\alpha = 1$ for $t \in [0,200]$

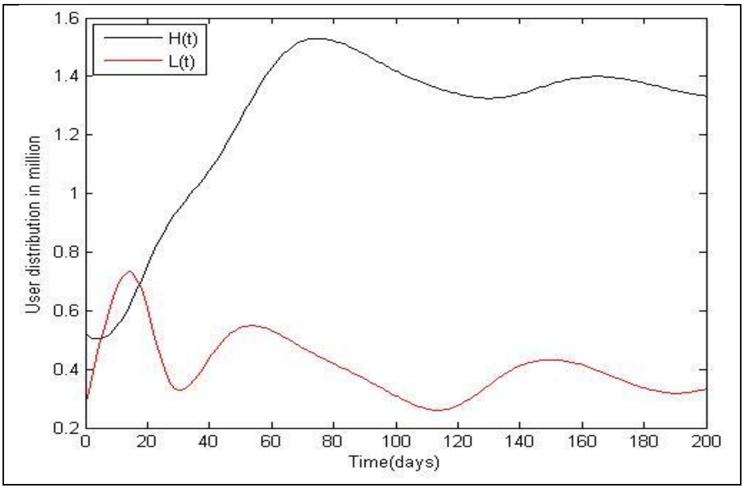


Figure 2: Interaction of Normal Users with Highly Active Users at $\alpha = 1$

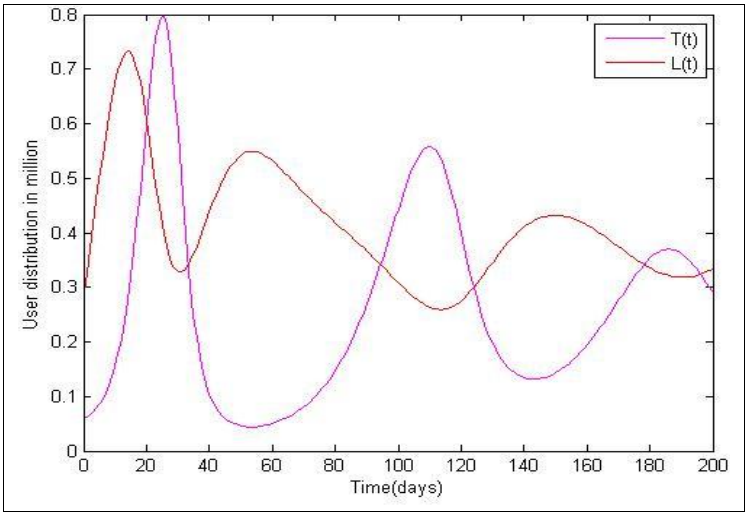


Figure 3: Interaction of Normal Users with Recovered Users at $\alpha = 1$

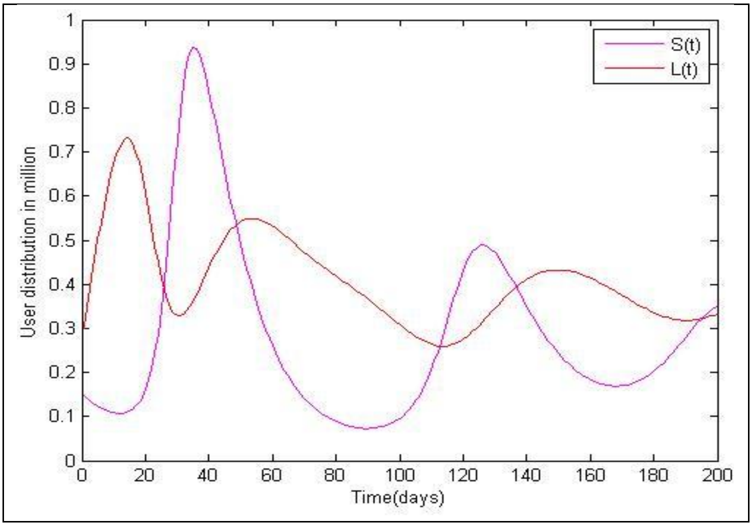


Figure 4: Interaction of Normal Users with Moderate Active Users at $\alpha = 1$

In figure 1, 2, 3 and 4 the behavior of Facebook users is shown at $\alpha = 1$. It is clearly evident that behavior of $H(t)$ is abruptly increasing and can't be stopped without interference. Whereas, other variables $S(t)$, $L(t)$ S and $T(t)$ can influence each other and can alter the amount of population with each others behavior. They have a positive domain to act upon, which means this addiction can't take their lives and vanish them from the system but there are other factors of life which get disturbed due to excess use of this social networking. Also excessive use of Facebook can be seen especially at weekends whereas in the mid of week, due to pressure of work or other activities Facebook usage gets slower.

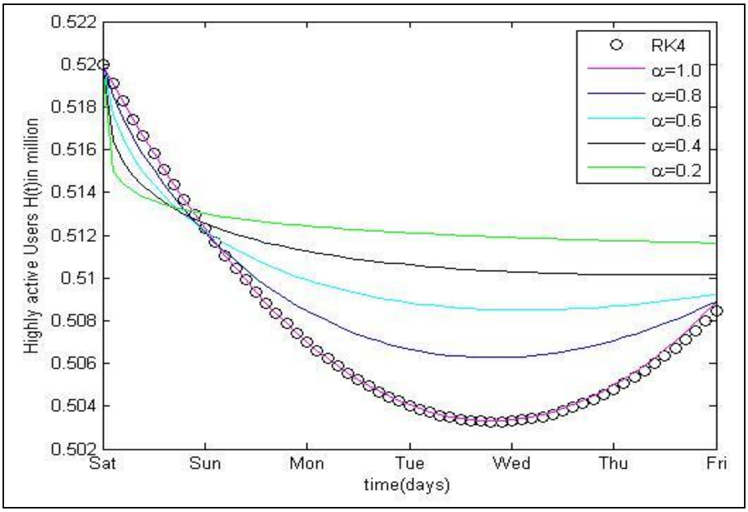


Figure 5: Behavior of Ultra High Active Users at Different Value of α and RK4 Results

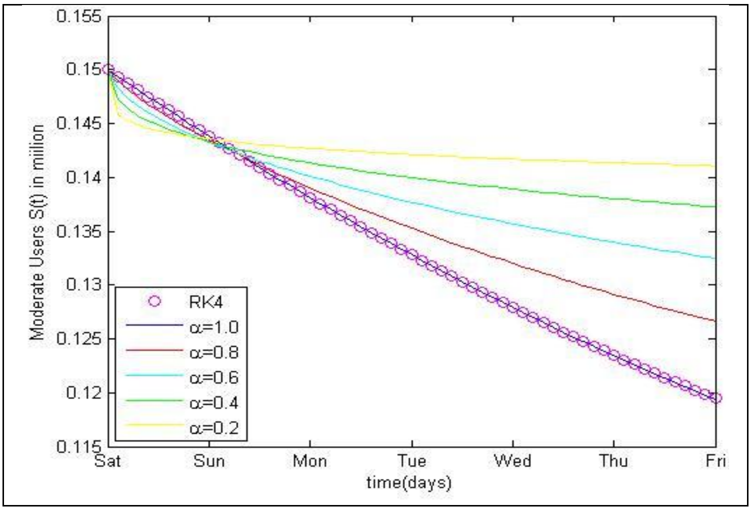


Figure 6: Behavior of Moderate Active Users at Different Value of α and RK4 Results

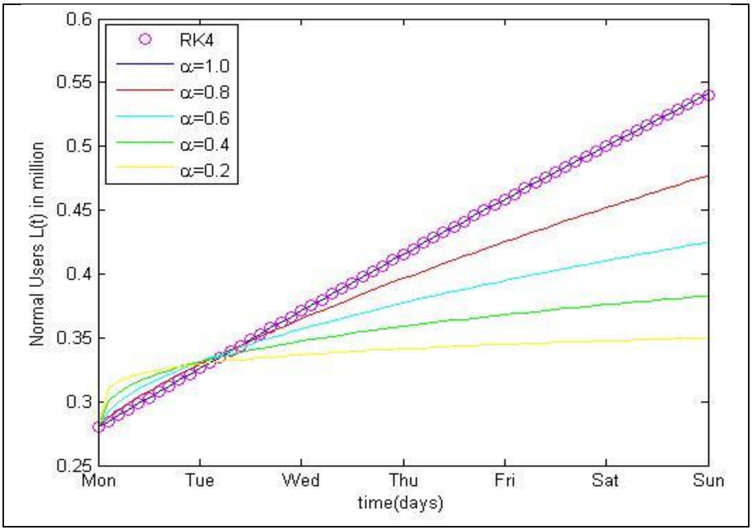


Figure 7: Behavior of Normal Users at Different Values of α and RK4 Results

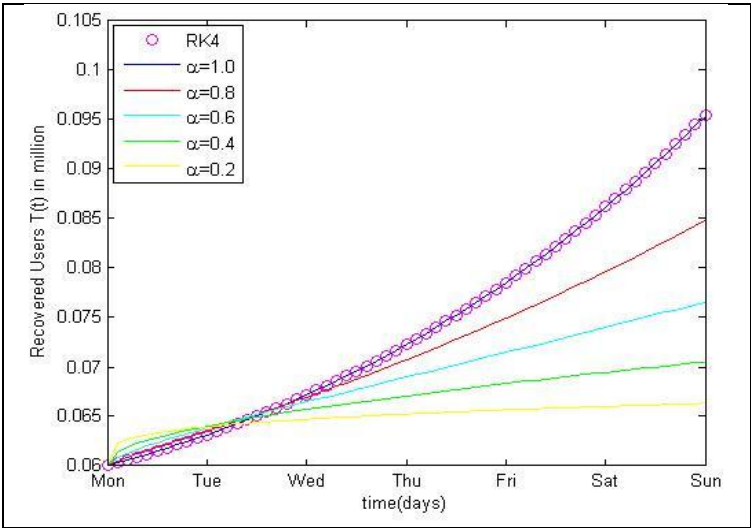


Figure 8: Behavior of Recovered Users at Different Values of α and RK4 Results

Figure 5-8 shows the comparison of RK4 method by a numerical technique known as Perturbation Iteration Algorithm at $\alpha = 1$. Also the behavior at different fractional values is also shown. The results obtained here are completely in coherence with the stability results.

6. CONCLUSION AND DISCUSSION

In this paper, a fractional mathematical model is constructed for describing the behaviors of Facebook users. Four types of users have been considered such as highly active users; less active users, normal users and recovered users who had sought some physiological or medical treatment. Two equilibrium points have been obtained $(0,0,1,0)$ and Eq. (7). Its Lyapunov stability has been discussed. Which shows that due to the behavior of highly active users the whole system gets unstable equilibriums and if this subclass isn't being interfered by some external variable its number will keep rising and making system more unstable. Lyapunov is considered due to its phenomenon that whether the fractional system has stable or unstable solution it gives the analysis of its stability far better than other techniques. To understand this system's eigen values and stability its space-state representation is also provided. From which four eigen values have been obtained $\lambda_1 = -0.792503$, $\lambda_2 = -2.28108\,\mathrm{i}$, $\lambda_3 = 2.28108\,\mathrm{i}$, $\lambda_4 = 0.792503$. For Complex eigen values, its stable focus if these eigen values satisfy $2\left|\arg\left(eign\ D^{\alpha} f(\bar{x})\right)\right| > \alpha\pi$. And its unstable focus if these eigen values satisfy $\left|2\arg\left(eign\ D^{\alpha} f(\bar{x})\right)\right| < \alpha\pi$. For real eigen values, asymptotically stable if all the eigen values satisfy $\left|2\arg\left(eign\ D^{\alpha} f(\bar{x})\right)\right| > \alpha\pi$, stable node if $\left|2\arg\left(eign\ D^{\alpha} f(\bar{x})\right)\right| = \alpha\pi$ or unstable node if all the eigen values satisfy $\left|2\arg\left(eign\ D^{\alpha} f(\bar{x})\right)\right| < \alpha\pi$. Also it could be a saddle point if some eigen values satisfy $\left|2\arg\left(eign\ D^{\alpha} f(\bar{x})\right)\right| > \alpha\pi$ and some satisfy $\left|2\arg\left(eign\ D^{\alpha} f(\bar{x})\right)\right| < \alpha\pi$.

REFERENCES

1. https://blog.bufferapp.com/psychology-of-facebook
2. Abdullah, F.A. (2011) Using fractional differential equations to model the Michael is Menten reaction in a 2D region containing obstacles, *Science Asia*, 37 (1), 75-78.
3. Keith N.H., Lauren, S.G., Lee, R. and Kristen, P. The Pew Research Center's Internet and American Life Project, Washington, D.C.
4. Craiem, D.O., Rojo, F.J., Atienza, J.M., Guinea, G.V. and Armentano, R.L. (2008) Fractional calculus applied to model arterial viscoelasticity, *Latin American Applied Research*, 38(2), 141-145.
5. Vedat, S.E., Gul, Z. and Shaher, M. (2012) A numeric analytic method for approximating a giving up smoking model containing fractional derivatives, *Computers and Mathematics with Applications*, 64, 3065-3074.
6. Ivo, P. (2011) *Fractional-order non linear systems, Modeling analysis and simulation*, Higher Education Press, Springer-Verlag, Beijing.
7. Matignon, D. (1996). Stability results for fractional differential equations with applications to control processing, *Computational Engineering in Systems Applications., Lille, France*, 2, 963-968.

8. Bandyopadhyay, B. and Kamal, S. (2015). *Stabilization and Control of Fractional Order Systems: A Sliding Mode Approach*, Lecture Notes in Electrical Engineering, Springer International, Switzerland.
9. Norelys, A.C. Manuel, A., Duarte, M. and Javier, A.G. (2014). Lyapunov functions for fractional order systems. *Commun. Nonlinear Sci. Numer. Simulat.*, 19, 2951-2957.
10. Chen, Y. and Vinagre, B.M. (2010). *Fractional Order Systems and Controls, Fundamentals and Applications*, Springer, Berlin.
11. Fatemeh, K., Gholam, H.E. and Mahmoud, B. (2014). On Stability of Equilibrium Points in Nonlinear Fractional Differential Equations and Fractional Hamiltonian Systems, *Complexity*, 21, 93-99.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 335-346

EFFECTIVENESS OF LEADERSHIP STYLE ON EMPLOYEE PERFORMANCE

Hammad Waseem, Muhammad Ibrahim Shamsi, Sadiq Ali, Zohaib Aziz and **Rafique Abid**
National University of Computer & Emerging Sciences – FAST, Karachi, Pakistan
Email: k142185@nu.edu.pk
ibrahim.shamsi@nu.edu.pk
k142268@nu.edu.pk
k142253@nu.edu.pk
k142199@nu.edu.pk

ABSTRACT

Effective leadership has always played an important role in the growth and better performance of the organization. The managers usually get stuck while thinking about different styles of leadership to follow. To solve this problem, we have focused on two successively used leadership styles which includes transformational and transactional. Several studies highlighted that transformational leadership style leads to employee motivation, whereas the transactional leadership style achieves productivity by giving rewards and punishments to the employees. Our hypotheses was that transformational leadership style has positive impact on employee performance in such a way that it shifts employee performance towards the ability to do assigned works efficiently and transactional leadership style has negative impact on employee performance because in this style, there is a pressure on employee to achieve goals successfully and before deadlines. Two hundred employees and employers from various software houses of Karachi city were targeted as the respondents. A multifactor leadership questionnaire was constructed to measure the leadership style – independent variable, and the employee performance – dependent variable. Regression analysis was performed and proposed two models using SPSS. The results of statistical tests were concluded in a way which leadership style is most effective in terms of productivity and job satisfaction. The final results showed that transformational leadership style is better for employee's performance.

KEYWORDS

Transformational leadership style, transactional leadership style, regression.

1. INTRODUCTION

Ever since the beginning of industrialization and the corporate era, employers have tried to come up with the best techniques to maximize productivity in a way that is also beneficial for their employees. The Gallup Organization, which has polled millions of employees and tens of thousands of managers, has found that the single most important variable in employee productivity and loyalty isn't pay or benefits or workplace environment; it's the quality of the relationship between employees and their direct supervisors. Within the last decade and a half, exceptional leaders who infuse ideological values and moral purpose into organizations and who have extraordinary effects on their followers and organizations have captured the attention of leadership scholars.

Several effective leadership styles have emerged in the last few decades; chief among them are transactional leadership style – which achieves productivity by giving rewards and punishments - and transformational leadership style - which gives importance to employees' motivation. Transactional leadership focuses on results, conforms to the existing structure of an organization and measures success according to that organization's system of rewards and penalties. Transactional leaders have formal authority and positions of responsibility in an organization. This type of leader is responsible for maintaining routine by managing individual performance and facilitating group performance. The main components of transformational leadership are; idealized influence and inspirational motivation, which serving as a charismatic role model and articulating a vision of the future that can be shared. Individualized consideration involves the leader paying attention to individual differences. However, since both styles work very differently, there is a lot of debate going on about which works better and why.

Our objective was to find out which leadership style is most effective in terms of productivity and job satisfaction so as to help managers adopt the best leadership approach for their employees and company.

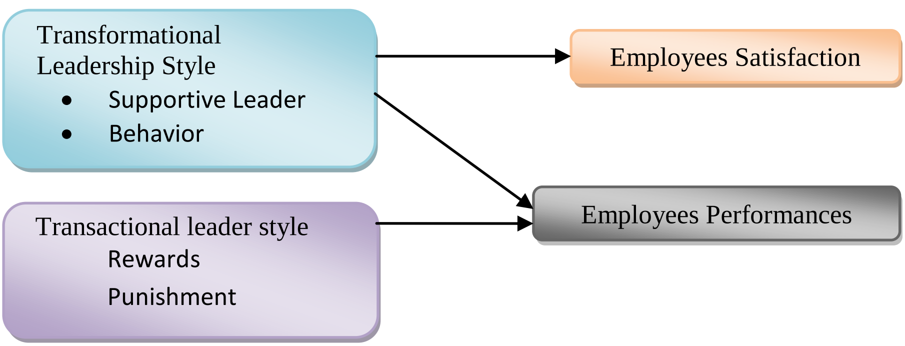
2. HYPOTHESIS

Based on the stated theoretical and conceptual framework, following hypotheses are developed and tested in the research paper:

H1: Transformational leadership style has positive impact on employee performance in such a way that it shifts employee performance towards the ability to do assigned works efficiently.

H2: Transactional leader style has negative impact on employee performance because in this style there is a pressure on employee to achieve goals successfully and before time deadline.

THEORETICAL AND CONCEPTUAL FRAMEWORK



3. LITERATURE REVIEW

Leadership can be defined as delivering the vision and motivation to a group of people to help them work together towards achieving the same goals. It also includes understanding what each individual in the team is capable of doing and effectively motivating each person to contribute individually their best toward achieving the group goal.

Burns (1978) defined leadership as leaders inducing followers to act for certain goals that represent the values and motivations, the wants and needs, the aspirations and expectations of both leaders and followers. Leadership styles in relation with employee performance have been closely monitored by various researchers in the last two decades. Adnan Riaz (2010) studied the impact of transformational and transactional leadership style on job and career satisfaction and found that job success is more dependent on leadership style. Gholamreza Jandaghi and Hasan ZareiMatin (2009) compared the impact of transformational leadership in successful and unsuccessful companies. They concluded that successful companies implement transformational leadership style more, as compared to less successful companies. However, our research was mainly focused on employees of different companies rather than the companies themselves. K. Rukmani (2010) conducted a study on the impact of transactional and transformational leadership style on organizational effectiveness. The results showed that managers are often using transformational leadership style. Ali Taj and Zahra Abdolvahabi (2010) shed light on the way coaches tend to bring the best out in their players in terms of motivation and performance. Their research shows that transformational leadership behaviors (individualized consideration, intellectual stimulation, inspirational motivation, and idealized influence) of coaches had more effect on increase of power motive than transformational leadership characteristics (extraversion, neuroticism, openness to experience, agreeableness, conscientiousness). One the other hand, transactional leadership styles of coaches did not have significant correlation with three scales of motives related (power, achievement and fair of failure).

In late 20th century, employers had begun to realize that to survive in the ever changing environment, they needed a change in the way their managers lead their employees. Julian Barling and Tom Weber (1996) evaluated the training programs of transformational leadership theory on different bank branch managers using outcome criteria and found it to be effective for manager’s behaviors and financial performances. With the advent of globalization, extended spans of control and advanced communication technology, managers are expected to lead hundreds of employees on a daily basis, including those with whom face-to-face communication is not possible. Here a question arises that in the global marketplace, are there common leadership behaviors being used universally or it deviates across countries and cooperate cultures? Karen Boehnke and Nick Bontis (2002) answer this question as the transformational leadership style is more useful than other leadership styles for reaching needs of those whom you are working with and for creating exceptional performances.

4. METHODOLOGY

We have collected our data by designing a kind of Multifactor Leadership Questionnaire (MLQ) to measure our dependent variable (i.e. Employee Performance) and independent variables (i.e. Transformational and Transactional Leadership Style). We have divided a Questionnaire in 4 sections in which we have section A, B, C, and D for Transactional Leadership, Transformational Leadership, Employee Performance and Demographic Information respectively. Every section (except Demographics) contains 5 questions with 5 point Likert scaling (Strongly Agree to Strongly Disagree). We filled out 200 questionnaires from software houses and startups at Karachi, Pakistan (e.g. The Nest I/O, Kay GeeS, Tech Logics, Gaditek, GFK Etilize, Folio3, Systems Ltd, Avanza) through visiting and/or circulating Google forms online. All subjects were fully informed as to the nature of the investigation and were assured that all data collected on them would be kept completely confidential. After analysis of Demographic Information, we found that we have collected our data from 26% females and 74% males in which mostly are young (i.e. 72 % respondents are 16 – 35 years old) and people with moderate experience (i.e. 34.5% respondents have worked 2 – 4 years in their current position) and have managerial positions (i.e. 56% respondents have supervised 1 – 15 employees).To analyze the data, we have used linear regression (simple and multiple). In statistics, linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression. We have also used correlation and ANOVA (analysis of variance) between the variables. A correlation is a single number that describes the degree of relationship between two variables. ANOVA is a statistical analysis tool that separates the total variability found within a data set. The ANOVA test for regression is used to determine the impact independent variables have on the dependent variable in a regression analysis.

R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination. It is the percentage of the response variable variation that is explained by a linear model. R-squared is always between 0 and 100% where 0% indicates that the model explains none of the variability of the response data around its mean and 100% indicates that the model explains all the variability of the response data around its mean.

A major difference between R-squared and the adjusted R-squared is that R-squared supposes that every independent variable in the model explains the variation in the dependent variable. It gives the percentage of explained variation as if all independent variables in the model affect the dependent variable, whereas the adjusted R-squared gives the percentage of variation explained by only those independent variables that in reality affect the dependent variable.

The standard error of the estimate is a measure of the accuracy of predictions made with a regression line.

$$\sum\left(Y-Y'\right)^2$$

The sum of squares represents a measure of variation or deviation from the mean. It is calculated as a summation of the squares of the differences from the mean.

Unstandardized coefficients indicate the average change in independent variable associated with 1-unit change in the dependent variable standardized coefficient. In statistics, standardized coefficients or beta coefficients are the estimates resulting from a regression analysis that have been standardized so that the variances of dependent and independent variables are 1. Therefore, standardized coefficients refer to how many standard deviations a dependent variable will change, per standard deviation increase in the predictor variable.

5. RESULTS AND DISCUSSION

Once the data was collected, aforementioned techniques were used to calculate the results.

A. TRANSACTIONAL MODEL

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.366[a]	.134	.127	.49065

a. Predictors: (Constant), Transactional
b. Dependent Variable: Performance

ANOVA

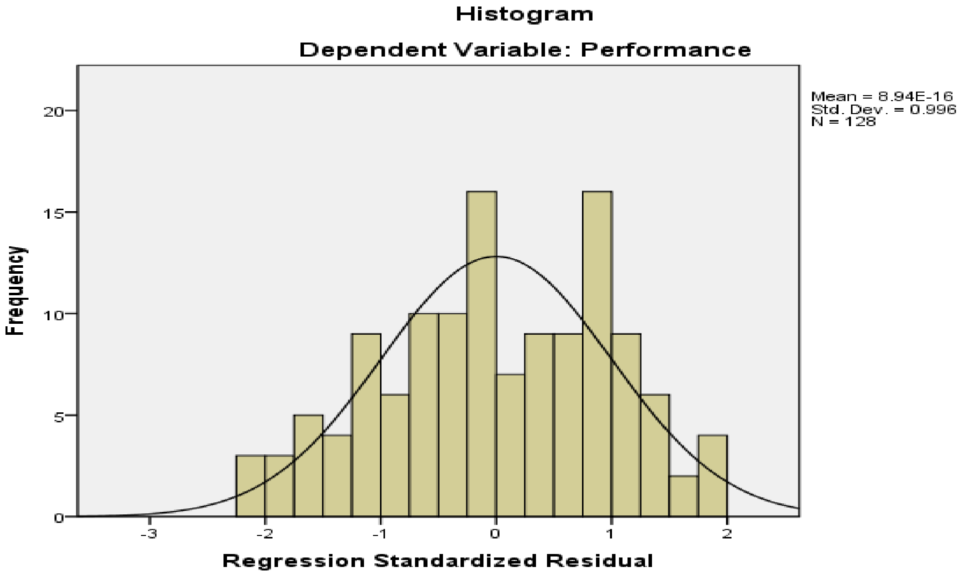
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.697	2	4.697	19.509	.000[a]
	Residual	30.333	198	.241		
	Total	35.030	200	s		

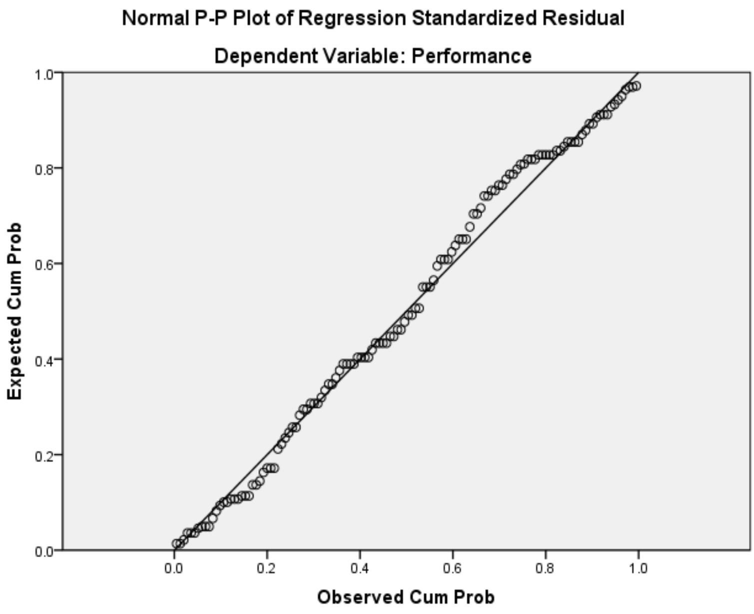
a. Predictors: (Constant), Transactional
b. Dependent Variable: Performance

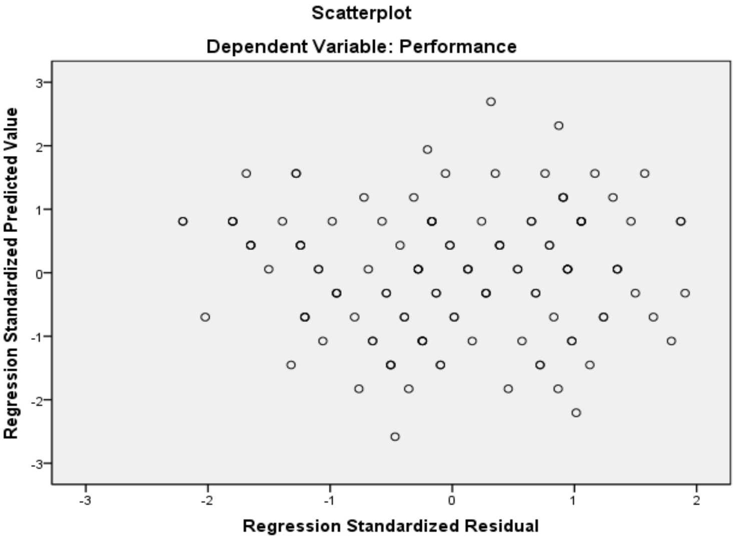
Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.633	.296		8.891	.000
	Transactional	.362	.082	.366	4.417	.000

a. Dependent Variable: Performance


Histogram
Dependent Variable: Performance
Mean = 8.94E-16
Std. Dev. = 0.996
N = 128
Frequency
Regression Standardized Residual


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: Performance
Expected Cum Prob
Observed Cum Prob



Correlations

		Transactional	Transformational	Performance
Transactional	Pearson Correlation	1	.168	.366**
	Sig. (2-tailed)		.058	.000
	N	200	200	200
Transformational	Pearson Correlation	.168	1	.423**
	Sig. (2-tailed)	.058		.000
	N	200	200	200
Performance	Pearson Correlation	.366**	.423**	1
	Sig. (2-tailed)	.000	.000	
	N	200	200	200

**. Correlation is significant at the 0.01 level (2-tailed).

This shows that the relation between performance and transactional leadership is 36.6% which depicts a moderate relationship. The relation between transformational and performance is much stronger i.e. 42.3% which proves our hypothesis that employee performance is effected more by transformational leadership style rather than transactional leadership style. However, all the models are significant.

This also proves our supposition that transactional and transformational leadership styles are independent of each other.

MODEL:

Performance = 2.633 + .362(Transactional Leadership)
which depicts a positive relationship.
The R-squared is 0.134 i.e. the predictive ability of this model is 13.4%.

B. TRANSFORMATIONAL MODEL

The R-squared for transformational model is 17.9%, almost 5% greater than the transactional model.

MODEL:

Performance = 2.327 + .446 (Transformational Leadership)
which shows a positive relationship.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.423[a]	.179	.172	.47780

a. Predictors: (Constant), Transformational
b. Dependent Variable: Performance

ANOVA

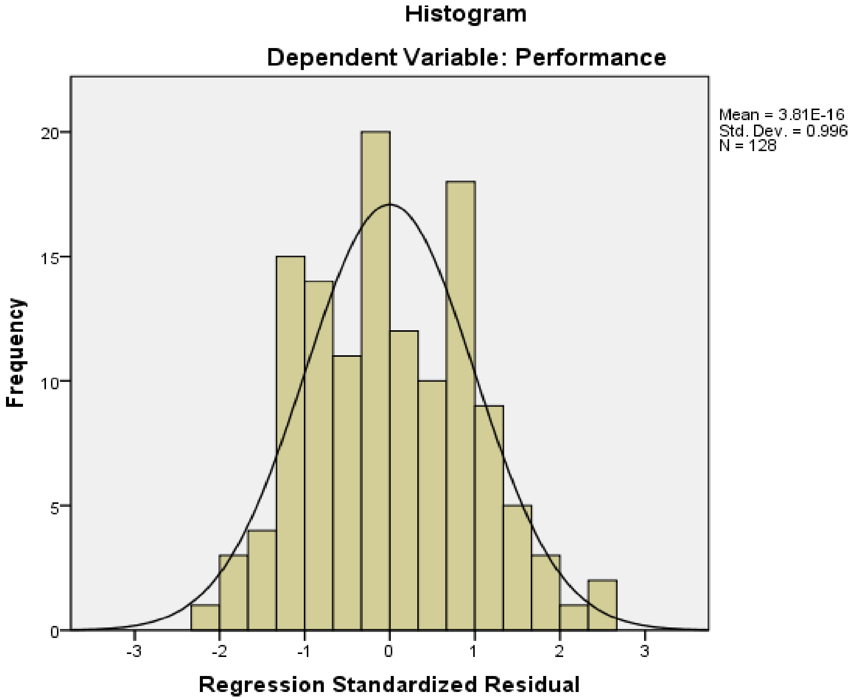
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.264	2	6.264	27.439	.000[a]
	Residual	28.765	198	.228		
	Total	35.030	200			

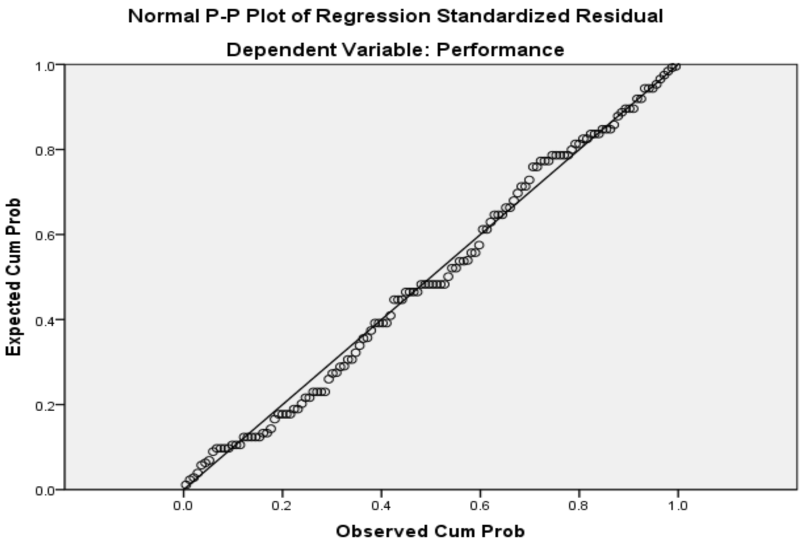
a. Predictors: (Constant), Transformational
b. Dependent Variable: Performance

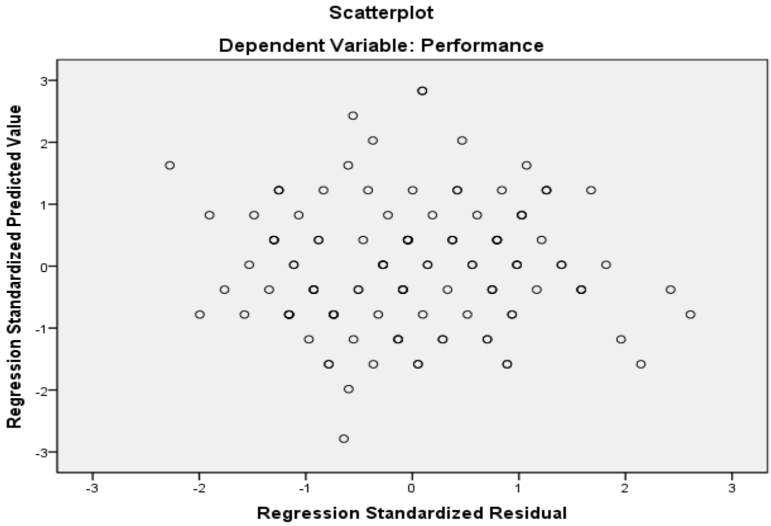
Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.327	.308		7.550	.000
	Transformational	.446	.085	.423	5.238	.000

a. Dependent Variable: Performance


Histogram
Dependent Variable: Performance
Mean = 3.81E-16
Std. Dev. = 0.996
N = 128
Frequency
0
5
10
15
20
-3
-2
-1
0
1
2
3
Regression Standardized Residual


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: Performance
Expected Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0
Observed Cum Prob



OVERALL MODEL
Multiple Linear Regressions

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.518[a]	.268	.257	.45277

a. Predictors: (Constant), Transactional, Transformational
b. Dependent Variable: Performance

As it can be observed, R-Squared has significantly increased. For transactional leadership, R-Squared was 13% and for transformational it was 18% but when they are combined to form a multiple regression model, the predictive ability increases to almost 27%.

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9.405	2	4.702	22.939	.000[a]
	Residual	25.625	125	.205		
	Total	35.030	127			

a. Predictors: (Constant), Transactional, Transformational
b. Dependent Variable: Performance

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.447	.369		3.925	.000
	Transformational	.392	.082	.372	4.792	.000
	Transactional	.301	.077	.304	3.914	.000

a. Dependent Variable: Performance

6. CONCLUSION

We can conclude from our study that transformational leadership style has an edge on transactional leadership style to improve employee performance. This verifies our hypothesis that transformational leadership is a better approach than transactional leadership for motivating employees and improve their productivity. However, it is also shown that for the best results, a combination of both leadership styles works ideally.

REFERENCES

1. Riaz, A. and Haider, M.H. (2010). Role of transformational and transactional leadership on job satisfaction and career satisfaction. *BEH - Business and Economic Horizons*, 1(1), 29-38
2. Taj, A., Abdolvahabi, Z., Naghavi, V., Rahmati, H. and Naini, S.S. (2010). The Relationship between the Coach's Transformational and Transactional Leadership Styles. *World Applied Sciences Journal*, 10(1), 9-18.
3. Jandaghi, G., Matin, H.Z. and Farjami, A. (2009). Comparing Transformational Leadership in Successful and Unsuccessful Companies. *The Journal of International Social Research*, 2(6) 357-370.
4. Kelloway, E.K., Barling, J., Kelley, E., Comtois, J. and Gatien, B. (2002). Remote transformational leadership. *Leadership & Organization Development Journal*, 24(3), 163-171.
5. EranVigoda-Gadot (2006). Leadership style, organizational politics, and employees' performance: An empirical examination of two competing models. *Emerald Insight*, 36(5), 661-683.
6. Jandaghi, G. Matin, H.Z. and Farjami, A. (2009). Comparing transformational leadership in successful and unsuccessful companies. *African Journal of Business Management*, 3(7), 272-280.
7. Barling, J., Weber, T. and Kelloway, E.K. (1996). Effects of Transformational Leadership Training on Attitudinal and Financial Outcomes: A Field Experiment, *Journal of Applied Psychology*, 81(6), 827-832.
8. Boehnke, K., Bontis, N., DiStefano, J.J. and DiStefano, A.C. (2002). Transformational leadership: an examination of cross-national differences and similarities. *Leadership & Organization Development Journal*, 24(1), 5-15.
9. Rukmani, K. (2010). Effect of Leadership Styles on Organizational Effectiveness. *European Journal of Social Sciences*, 15(3), 365-370.

10. Gupta, V. and Krishnan, V.R. (2004). Impact of Socialization on Transformational Leadership: Role of Leader Member Exchange. *South Asian Journal of Management*, 11(3), 7-20.
11. David M. Lane (2012). Introduction to Linear Regression [.html]. Retrieved from http://onlinestatbook.com/2/regression/intro.html
12. Encyclopedia Britannica (2010) Regression and correlation analysis [.html] Retrieved from http://abyss.uoregon.edu/~js/glossary/correlation.html
13. Steven B. Achelis (2000). Correlation Analysis [.html] Retrieved from http://www.metastock.com/customer/resources/taaz/?c=3&p=44
14. Statistics Lectures (2010) *Introduction to ANOVA*. Retrieved from https://www.youtube.com/watch?v=qV-WoquC4dA

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 347-354

THE EFFECTS OF ILLITERACY AS A MAIN CAUSE OF CORRUPTION

Meghna Jadav, Syeda Mahrukh and **Muhammad Ibrahim Shamsi**
National University of Computer & Emerging Sciences – FAST, Karachi, Pakistan
Email: k142089@nu.edu.pk
k142054@nu.edu.pk
ibrahim.shamsi@nu.edu.pk

ABSTRACT

Corruption is the seed which is rooted deeply in our country. It is now becoming more common as the public prefers easy and unethical ways to get their work done without having to work hard. Corruption is taking place on large and small scales as well. Corruption can be a result of a number of things but the main reason can be identified as illiteracy. Illiteracy is becoming a common problem in our country. Our public does not give enough importance to education. A reason might be insufficient help from the government or due to lack of enough resources and poverty. Unfortunately, this has a negative effect on the nation as a whole. This research would include the questions about how importantly the educational ratio of the nation affects its people's mentality and how that effect the corruption involved, mainly in the name of economical welfare and political decisions. This research topic emphasizes on answering the question of how the eradication of illiteracy can help in eradicating corruption from the society. The purpose for this study is identifying the thinking of people which leads to corruption in the country. We analyzed the concepts built about corruption in the minds of the public of our country. We also identified the difference of mentality of students of different ages and different educational statuses. We have used a questionnaire and interviews in our survey to collect the data from 200 respondents, to find the factors which are the result of illiteracy and become the reason of corruption. The findings proposed a solution which might help in eradicating illiteracy, and mainly the ideas which sow the seeds of corruption in people's minds.

KEY WORDS

Illiteracy, Corruption.

1. INTRODUCTION

The topic of our research is "Illiteracy as the main cause of corruption". Illiteracy is one of the major problems of Pakistan and is the root of majority of the problems faced by the country. Education plays an important role in the development of a country, but corruption negatively affects this development. However, both these things are somehow related to one another. To investigate whether illiteracy and uneducated people are one of the main causes of corruption, we carried this research. There are a lot of causes of illiteracy such as inflation, expensive education system, low income etc, and there might be a lot of causes of corruption, but what we will be discussing in our research is that how is Illiteracy one of the major causes of corruption. And will decrease in the rate of Illiteracy in Pakistan decrease the rate of corruption or not.

2. LITERATURE REVIEW

Illiteracy has been defined by numerous scholars at different times, Selfe (2001) stated that literacy is not referred to only as the ability to reading and writing but also the ability to comprehend and learn about the technology. So the concept of illiteracy now does not only include that a specified number of years of education have not been undertaken, but also the learning of a person on different levels and platforms. It has also been said by (Scribner, 1984) "…the single most compelling fact about literacy is that it is a social achievement; individuals in societies without writing systems do not become literate. Literacy is an outcome of cultural transmission; the individual child or adult does not extract the meaning of written symbols through personal interaction with the physical objects that embody them" Another concept that we see in a recent study by (Perry, 2012)), it says that literacy can be stated as a set of skills owned by an individual that can be applied in any situation, stating that a person with less or no skills may be considered an illiterate. If we statistically speak of the situation Pakistan with respect to literacy, the literacy ratio is below average. One of the latest statistics recorded by Aamir Latif (2011), we see that the literacy ratio of Pakistan is only 46% and only 26% of girls are literate. There are about 163000 schools in Pakistan but only 40,000 cater girls. This makes Pakistan one of the most illiterate countries standing on 160th position, out of all countries of the world, with respect to the literacy ratio.

(Dridi, 2014) says that corruption is related to public decisions and can also be referred to as the violation of public office for their personal benefit. It may also be defined as a social pathology. It destroys the nation just like a terminal disease effects a person, and is a social network phenomena. Pak Hung Mo (2000) said that "…corruption works like piece-rate pay for bureaucrats; it induces a more efficient provision of government services, and also provides a leeway for entrepreneurs to bypass inefficient regulations." (Mauro, 1995) the author has analyzed a newly assembled data set consisting of subjective indices of corruption, the efficiency of the judicial system, and various categories of political stability for a cross section of countries. About the statistics of Pakistan, Corruption Perceptions Index (2015), by Transparency International, reported that Pakistan is the 117th least corrupt nation out of 175 countries. Corruption rank in Pakistan averaged 107.90 from 1995 until 2015, reaching an all time high of 144 in 2005 and a record low of 39 in 1995. The complete statistics show that Pakistan was ranked the 142nd least corrupt nation in 2006, then became 138th in 2007, 134th in 2008, then going up to 139th least corrupt in 2009, 143rd least corrupt in 2010, then declining to 134th position in 2011, then becoming 139th least corrupt nation in 2012, 127th in 2013, 126th in 2014 and 117th in 2015.

(Scribner, 1984), has stated literacy as adaptation, power, and a state of grace. She has explained the roles which literacy plays in promoting these factors, however she has not linked it directly with corruption and how illiteracy effects corruption. In another article, (Bormuth, 1973) has analyzed the concept of literacy in order to identify measurement problems associated with specifying each of these parameters, and to describe literacy assessment procedures now available for dealing with measurement problems. The principal focus of the paper is on the development of models for identifying performance criteria that can serve as the goal of instructional programs and of the research and development programs that lead to them. The five parameters discussed are the classes of literacy behaviors, the level of performance that serves as the criterion of literate performance, the kinds of reading tasks on which the behaviors are tested, the proportion of the reading tasks that serves as the criterion of literacy on some corpus of reading tasks, and certain characteristics of the propel tested, such as the levels of aptitude and perseverance represented within it. Corruption has also been the focus of a few authors. Corruption is stated by Dridi, as the violation of public office. Scott states that corruption is acting against the laws pertaining to it, or acting against what the public opinion deems integrity, or, again, acting against general interest. However, these articles do not link illiteracy to corruption. This paper will include how illiteracy has become a cause of corruption. This study has analysis about the effect of illiteracy on corruption using factors like income, educational resources, educational systems, inflation etc. These factors included, this study has not been reported earlier in Pakistan.

3. MODEL FRAMEWORK

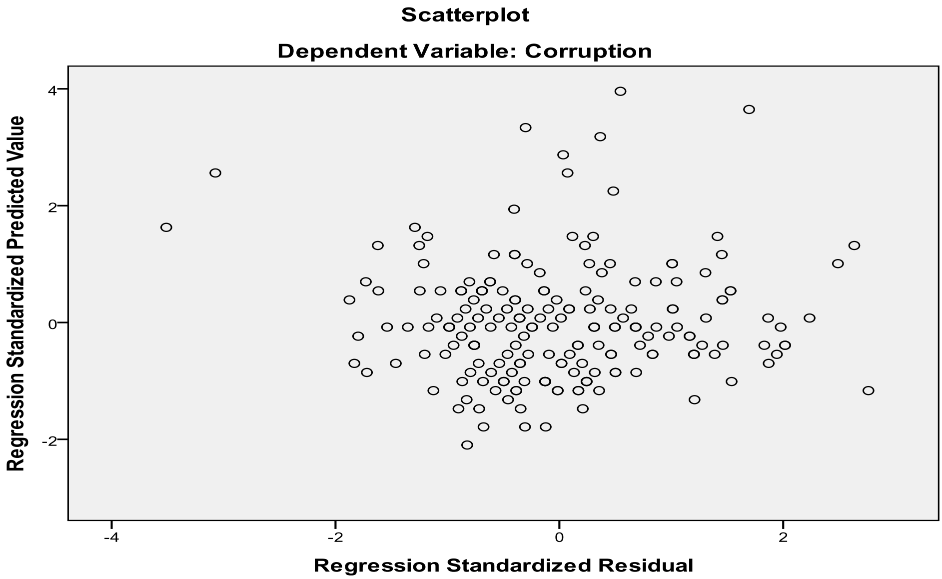
Illiteracy and corruption are closely related with Illiteracy being the cause and corruption being the effect. The factors affecting the rate of Illiteracy in Pakistan include inflation, expensive education system, ghost institutes low income, child labor etc. And the factors affecting the rate of corruption in Pakistan include low income, greed for money, unemployment, illiteracy etc. However, the causes for both, Illiteracy and corruption are similar. Illiteracy is the most important reason for unemployment, and unemployment results in low or no income. When a person does not get enough money to feed his family, he then uses corrupt means to earn. Moreover, illiterate people instead of educating their children and due to the increasing inflation, make their children work so that they can bring extra income for the house. However, literate people realize the importance of education.

4. METHODOLOGY

In this research, the data has been collected from 200 people of different age groups, educational backgrounds, and occupations. The data was collected by asking these people a few questions related to the topic, linking the factors which cause illiteracy to the causes of corruption. The questionnaire has been filled by people in hardcopy and also in softcopy. The data collected was analyzed using the application of SPSS.

5. COMMENTS AND CONCLUSION

This research has been analyzed by using the correlation and regression methods in SPSS. The result shows the following graphs



The scatter plot helps us to see that are variables have a linear relation between them.

The correlation coefficient can range from -1 to +1, with -1 indicating a total negative correlation, and a +1 indicating a total positive correlation, and 0 indicating no correlation at all. Correlation coefficient can tell us the extent to which we can predict value of one variable given a value of the other variable. .586 tells how tightly the points are around the imaginary line on the scatter plot, the higher the value the closer they are to the line, the lower the value, the further the points would be.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.586[a]	.343	.340	.54026

a. Predictors: (Constant), Illiteracy
b. Dependent Variable: Corruption

This table shows the value of correlation in the "R" column which is .586. and in the "R Square" column it shows that 34.3% of the dependent variable can be explained by the independent variable.

Correlations

		Corruption	Illiteracy
Pearson Correlation	Corruption	1.000	.586
	Illiteracy	.586	1.000
Sig. (1-tailed)	Corruption	.	.000
	Illiteracy	.000	.
N	Corruption	200	200
	Illiteracy	200	200

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	30.236	1	30.236	103.588	.000[a]
	Residual	57.793	198	.292		
	Total	88.029	199			

a. Predictors: (Constant), Illiteracy
b. Dependent Variable: Corruption

This table shows us that the regression model statistically significantly predicts the outcome variable (i.e., it is a good fit for the data), because in the "Regression" row the "Sig" column shows the value is less than 0.05.

Coefficients[a]

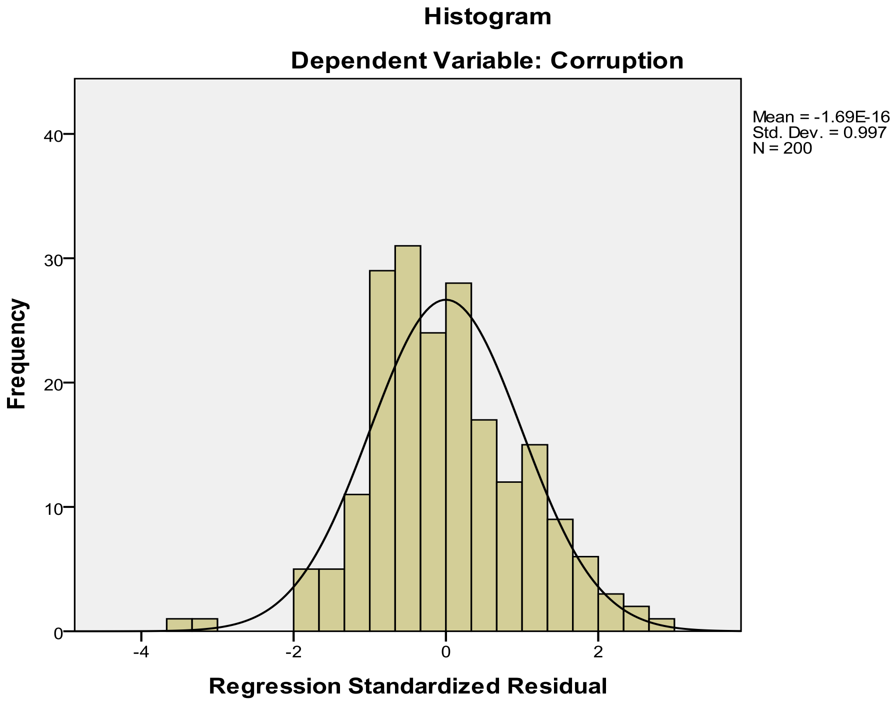
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.719	.156		4.595	.000
	Illiteracy	.726	.071	.586	10.178	.000

The coefficients table tells us with that we can predict corruption from illiteracy. As well as by looking at the "Sig" column we can determine whether illiteracy contributes statistically significantly to the model, because it is again less than 0.05. To get the equation we use the values of the "B" column. Using the general equation:

i.e. dependent variable = b0 + b1 (independent variable),

we form the regression equation :

Corruption = 0.719 + 0.726(illiteracy)



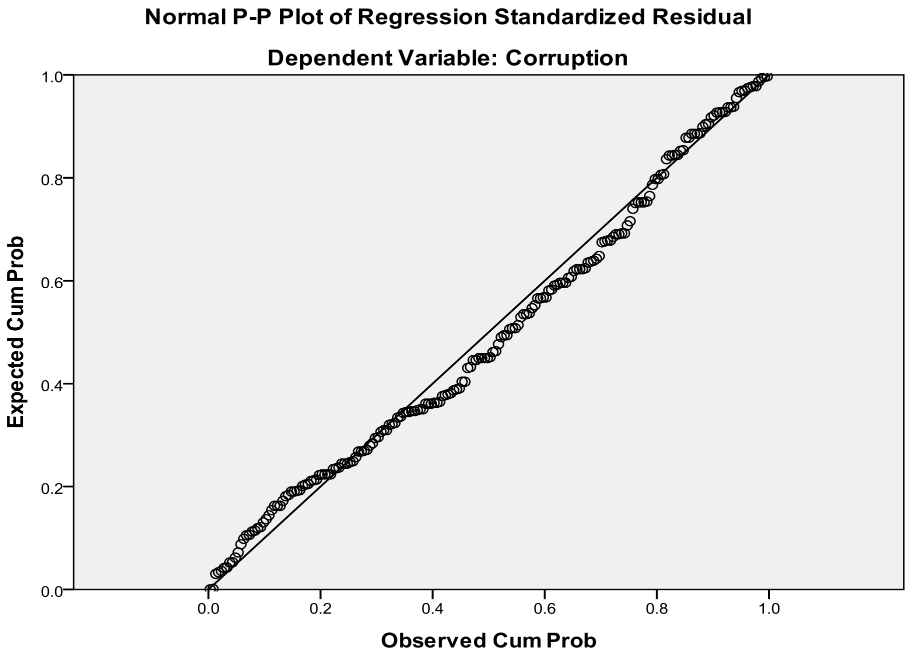
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.4449	3.8051	2.2625	.38979	200
Residual	-1.89732	1.49198	.00000	.53890	200
Std. Predicted Value	-2.097	3.957	.000	1.000	200
Std. Residual	-3.512	2.762	.000	.997	200
a. Dependent Variable: Corruption					

This frequency histogram and the table show us the frequency, mean, standard deviation, residual, predicted value and the standard predicted value.

The mean of the predicted values is 2.2625, the standard deviation of the predicted value is .38979, and the residual values lie between -1.8973 to 1.49198, calculated using the general formula i.e. Residual = Observed value – Predicted Value.

The predicted value is calculated using the regression equation, while the observed value depends on our observations.



In this graph we can see that the points are very close to the line, which shows that our predicted values are accurate, the further the points from the line, the less accurate our values would be.

With the results that we get, we may conclude that we can calculate corruption with the help of illiteracy, to an extent; hence corruption is dependent on illiteracy. If we keep the factors such as income, educational institutes, educational systems etc. in our mind and with the help of our government try to reduce illiteracy from our nation, corruption may gradually start to decrease. And there might come a time that we may have a highly literate population and very little corruption in our country.

REFERENCES

1. Bormuth, J.R. (1973). Reading Literacy: Its Definition and Assessment. *Reading Research Quarterly*, 9(1), 7-66. http://doi.org/10.2307/747227
2. Dridi, M. (2014). Corruption Within Education Sector: A Typology of Consequences. Online at http://mpra.ub.uni-muenchen.de/46874/
3. Gardiner, J. (2002). Defining corruption. Political corruption: *Concepts and contexts*, 3, 25-40.
4. Harman, D. (2002). Illiteracy: An Overview. *Harvard Educational Review*, 40, 226-230.
5. Hillerich, R. (1976). Toward an Assessable Definition of Literacy. *The English Journal*. 65(2), 50-55.

6. Huston, S.J. (2010). Measuring financial literacy. *Journal of Consumer Affairs*, 44(2), 296-316.
7. Kurer, O. (2005). Corruption: An alternative approach to its definition and measurement. *Political Studies*, 53(1), 222-239.
8. Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 681-712.
9. Perry, K.H. (2012). What Is Literacy?--A Critical Overview of Sociocultural Perspectives. *Journal of Language and Literacy Education*, 8(1), 50-71.
10. Scribner, S. (1984). Literacy in three metaphors. *American Journal of Education*, 93(1), 6-21.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 355-362

THE IMPACT OF LOGISTICS SUPPORT ON BUILDING TRUST IN ONLINE SHOPPING

Adila, Suffah Aftab and **Muhammad Ibrahim Shamsi**
National University of Computer & Emerging Sciences – FAST, Karachi, Pakistan
Email: k142075@nu.edu.pk
k142040@nu.edu.pk
ibrahim.shamsi@nu.edu.pk

ABSTRACT

Shopping is one of those things which had evolved over time. As world became globalized, means of shopping also started to change. Pakistan being a developing country has not particularly adopted the concept of online shopping. The most important thing which behaves as a hurdle in the growth of online customers is “trust”. Due to lack of trust people do not prefer to shop online. Three logistics supports have been discussed in this study that highly influenced trust in online shopping – Product quality, Transaction security and Timely delivery. The data was generated from a questionnaire surveyed from 500 online customers. The multiple linear regression model was developed to predict the influence of the logistics support on building trust. The study concluded that the concept of online shopping could be common among people to save their time and effort, if focused on the quality logistics supports through building the trust.

KEY WORDS

Logistics support, Product quality, Transaction security, Timely delivery, Trust, MLR.

1. INTRODUCTION

Over the past few years E-commerce has become more popular among people. E-commerce gives the opportunity to the individuals to develop their own business and grow in that business. E-commerce lets its customers gain benefit from a wide variety of products, without being time bound. Today most of the people use internet as a medium to gain knowledge start a business or do those tasks which would save their time if done through electronic media. Online shopping is one of those concepts which had evolved overtime and with the passage of time this concept started to become more common among people. Today most of the people do shopping online, saving their time and effort.

Pakistan being a developing country has not completely adopted the idea of online shopping as people are very much concerned about the security of their personal information and more importantly the quality of product being provided. The major concern for the people not engaging that much in online shopping is trust; people do not have trust on online stores because they cannot personally verify the quality of the product or security of their information.

In this paper we will identify three factors including product quality, transaction security and timely delivery that have an influence on building trust in online shopping. The main focus of this research is on online customers, we have divided online customers into three categories according to their age group. First is the group of online customers that are below 20 years, second are those online customers that are between 20-30 years of age and lastly those online customers that are of 30 years or above. Based on the survey of 197 respondents it is found that it is very difficult for the people to build trust on online stores as they have many concerns which needs to be dealt with. The three factors have an impact on building trust in online shopping.

2. LITERATURE REVIEW

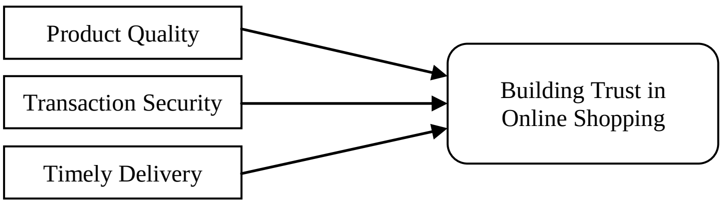
Transaction security can be defined as the customer's belief that the information they are providing for online transactions would be safe. Transaction security plays an important role on building trust towards online transactions. Customer would not trust an online store until they have a doubt about their important information such as credit card number not being kept confidential and safe. Customers are concerned about the security of the information they are providing and most of them hesitate to provide their personal information as they themselves cannot verify the security of the information being provided while doing online shopping. (Kwek Choon Ling, 2010). Statistics also showed that when respondents were asked about their online transactions it was found that about 24% of the respondents spent Rs. 4000 in 6 months, 36% respondents had spent about Rs. 5,000-9,000, 24% respondents had spent between Rs. 10,000-14,000 and 5% respondents had spent Rs. 20,000 and above in purchasing products online (Hooria Adnan, 2014). Product Quality can be defined as including the elements that satisfies the needs of the customers and make them free from defects. Product quality means the raw materials, technologies and skills used to design a product that meets customer's requirements. These requirements include shape, size, colour, dimension, appearance and strength of the product. Product quality is an important factor for the customers to build trust in online shopping. When customers purchase products online and pay high prices for it they expect to get the best quality product. If customers are not satisfied with the product they get then it influences their trust towards that online store (Gaurav Akrani, 2016). Statistics conducted during a research showed that in Pakistan 59 respondents preferred to purchase accessories, 29 respondents preferred to purchase electronic devices, 1 respondent preferred to purchase books and 5 respondent purchased tickets (Hooria Adnan, 2014). Timely delivery of the product can be defined as delivering the products safely and in time to the customers. Online stores often deliver products to the customers through shipment where some of the products gets damaged, company needs to control the timely delivery of the product because if products are not delivered in time then the customer would not prefer to shop from that online store again (wikinvest, 2009)

Trust can be defined as the willingness of the customer to accept the weakness in online Transaction, product quality and timely delivery of the product and believing the fact that everything good will happen in the future. Trust is belief of an individual that whatever will happen would be up to his/her expectations. Trust can be made on a person, product or even a business. (Kwek Choon Ling, 2010).Trust in online shopping is effected by the fact that online shopping is available all the time with more variety, better deals and price comparisons in less time and cost. (Adnan, 2014). Another study shows that trust has a positive relationship with online shopping approval. (samra chaudary, 2014).Trust also combines with online payment risk (transaction security) that customers hesitate to give their information while shopping online which include their debit/credit card number or their bank account number (cha, 2013). There are some security issues which hinders the data provided by the customers from being protected. These issues are phishing and online frauds. Phishing attacks can be very dangerous for the customers doing online transactions as an email is sent to the customers informing them about their account being expired and that they need to update their account. Customers believe that they would protect the information being provided in this way whereas in actual their information would not be kept confidential instead the information that the customer is providing would be used for some false action. (Ravi Das, 2016).Online frauds are also one of the hindrances in transaction security. An online fraud includes all the emails, websites and advertisements that use the customer's information for false actions. Some solutions have already been provided by the online business organizations to reduce the security issues. Encryption approach can be used to secure the provided data by converting the plain text into cipher text which cannot be read by anyone else except the sender and the receiver. Secure electronic transaction (SET) is also used for securing the data provided during online transaction. It requires some specifications of credit card for the safety of data while doing online shopping. (Pradnya B. Rane, 2012). A study discuss the importance of quality in trust on online shopping as their sample supported their hypothesis of quality orientation which tells that the high quality products delivered by the seller is very important and my expectations about quality from the seller is extraordinary. (Kwek Choon Ling, 2010). Some of the factors of product quality are design, reliability, safety and storage. If the product is not reliable or if the product is defective then the customer would not purchase products from that online store. Companies have solved this problem by keeping a quality control over raw material, machinery and finished products. Companies also make sure that the finished product has all the elements to satisfy the customer's needs. (Gaurav Akrani, 2016). Timely delivery also serves as a factor for trust in online shopping as mentioned in a study that trustworthiness that a product will be delivered in time increases customer satisfaction and their intention to adopt online shopping. (Koyuncu, 2004).

Trust has always been dependent on quality of the product, transaction security and timely delivery of the product. Different researches has shown relationship of quality orientation, perceived risk in online shopping, ease of use on trust to shop online. The study conducted in 2003 has evaluate trust in online shopping from two different aspects technology oriented and trust oriented which showed that risk and ease of use have a direct impact on trust. (Hans van der Heijden, 2003). Another study showed that online customers usually assess the formation of website and how secure the website is which shows that security is most important factor to the online consumers (Gurvinder S Shergill, 2005). (Kwek Choon Ling, 2010) The research conducted in 2010 showed that quality awareness, brand consciousness and online purchase experience influence s customer's intentions for online shopping. Most studies discuss the acceptance of online shopping but mostly people accept online shopping but they themselves does not prefer to do online shopping. The three most important factor of hesitation to shop online is timely delivery, quality of the product and transaction security which most of the studies does not evaluate together although it causes a massive impact to shop online. So this research handles all the three factors together in a balanced manner with appropriate reference to online shopping which will cause better results to target those customers that are online shoppers which tell that how online stores can build trust according to customers and how their logistic support effects trust to do online shopping.

3. MODEL FRAMEWORK

E-commerce have changed the way people used to purchase products, earlier people used to go to shops, pick out the product which was reasonable and of good quality. The traditional method of doing shopping used to consume a lot of time. Today people have started to get themselves aware to the concept of online stores. Trust of a customer is very important for the growth of business, without trust a company cannot flourish and achieve its target. Three factors play an important role in developing trust. If quality of the product is not upto the expectation then customer would not trust that online store, similarly if customers have any kind of doubt that the information that they are providing for transaction is being used for unfair means then also the customer would not develop trust on that online store, if products are not delivered in time or if the products delivered are damaged then it would be not according to the expectation of the customer and the customer would not trust that online store.



This study investigates that how all the logistics influence the trust of Pakistani customers towards online shopping and how these factors influence the trust of people with different age groups. The main objective of this research is to identify that which logistic support is important on building trust in online shopping? Based on the objective of this research we have developed the following hypothesis:

H1: There is a positive impact of Timely Delivery on building trust in online shopping.

H2: There is a positive impact of Transaction Security on building trust in online shopping.

H3: There is a positive impact of Product Quality on building trust in online shopping.

4. METHODOLOGY

This paper targets the online customers of Pakistan. Statistics of 2014 shows that 50% of the people in Pakistan having age 21-29 do online shopping, 30% of the people below 21 years of age and 20% of the people above 30 years of age are indulged in online shopping. Earlier very small numbers of people were indulged in online shopping but with time people have started to accept the idea of online shopping. Data was collected by conducting surveys from online customers of different age group. A web based survey was also conducted and in total 500 questionnaires were filled, out of these 500 questionnaires only 197 questionnaires were usable.

A pilot survey was conducted with five people to evaluate if there could be any improvements made in the questionnaire; after the pilot survey was conducted it was found that there was a need of rephrasing 1-2 questions. In the finalized questionnaire those improvements were made and people did not face any difficulty while filling out the questionnaire.

Method used for showing the results of the survey was the multiple regression analysis. It is a technique used for determining the value of dependent variable i.e. trust from the values of independent variable i.e. product quality, transaction security and timely delivery.(Explorable, 2008).

5. RESULTS AND DISCUSSION

A Survey was carried out from online customers to determine that which logistic support influences developing trust the most. From the survey which was conducted 121 usable questionnaires were filled by female and 76 usable questionnaires were filled by male. The main focus of this research is to determine the impact of logistic support on online customers with different age group. It was found that 81 respondents were below 20 years of age, 103 respondents were between 20-30 years of age and 13 respondents were above 30 years of age. The concept of online shopping varies with people having different occupation; according to the survey that was conducted 158 respondents were students, 32 respondents were employees and 7 respondents were others having different occupation.

	YES	**NO**
AMAZON	50	147
E-BAY	14	183
GOOGLE PRODUCT SEARCH	20	177
DARAZ	96	101
OLX	48	149
OTHER	13	184
CASH ON DELIVERY	166	31
CREDIT CARD	38	159
OTHER	4	193

While doing online shopping it is very important for a customer to choose from which online store to shop from, when people were asked about the online store they would prefer to shop from following were the responses made, the one thing which can stop people from purchasing products online is the payment method, in general people hesitate to provide their personal information while purchasing products online. When people were asked about their preference to pay for the online purchased products following were the responses made, most of the people preferred to pay cash on delivery rather than by credit card the clear reason being that people are concerned about their information being used for unfair means and not being kept confidential.

Correlations

		Trust	Product Quality	Transaction Security	Timely Delivery
Pearson Correlation	Trust	1.000	.202	.065	.424
	Product Quality	.202	1.000	.111	.338
	Transaction Security	.065	.111	1.000	.207
	Timely Delivery	.424	.338	.207	1.000
Sig. (1-tailed)	Trust	.	.002	.182	.000
	Product Quality	.002	.	.060	.000
	Transaction Security	.182	.060	.	.002
	Timely Delivery	.000	.000	.002	.
N	Trust	197	197	197	197
	Product Quality	197	197	197	197
	Transaction Security	197	197	197	197
	Timely Delivery	197	197	197	197

Pearson correlation coefficient measures the strength between variables and relationships. The correlation matrix given above indicates that all the relationships are positively correlated, as the coefficient value of all the variables are close to 1. The strongest relationship is between trust and timely delivery of the product which is r = .42, it means that if products are delivered in time and without any damage the customers would build more trust on that online store. Product quality and timely delivery of the product also have a strong relationship which is r = .33.
(David W. Stockburger, 1996).

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.430[a]	.185	.172	.47584	.185	14.557	3	193	.000	1.659

a. Predictors: (Constant), Timely Delivery, Transaction Security, Product Quality
b. Dependent Variable: Trust

The model summary shows that R-square is .17 which means that the predictors can tell 17% about the change in dependent variable trust.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9.889	3	3.296	14.557	.000[a]
	Residual	43.700	193	.226		
	Total	53.589	196			

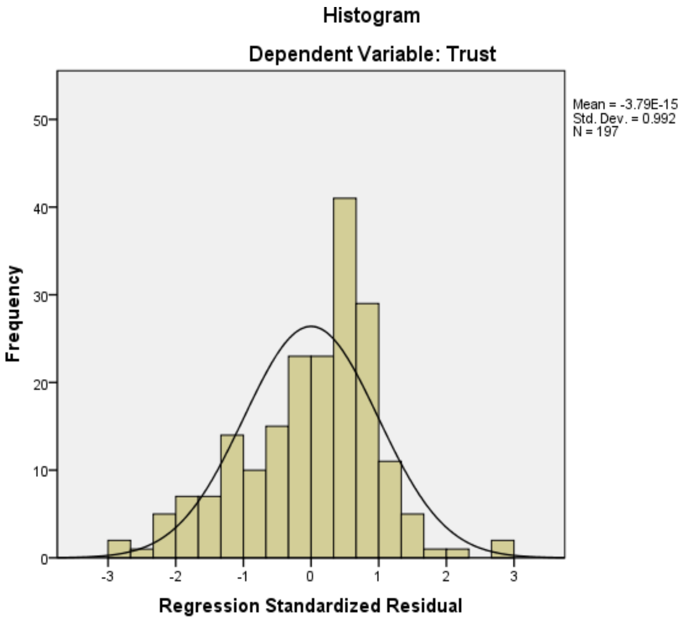
a. Predictors: (Constant), Timely Delivery, Transaction Security, Product Quality
b. Dependent Variable: Trust

The analysis of the variance table or ANOVA table is shown above. It can be seen from the above table that the regression sum of squares is the variation attributed to the relationship between dependent and independent variable. Each sum of square has a degree of frequency associated with it. Regression df shows the number of independent variables i.e. product quality, transaction security and timely delivery. Residual df is the difference between regression df and total df.

Coefficients[a]

Model		Unstand-ardized Coefficients		Standar-dized Coeff-icients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	1.525	.228		6.703	.000	1.076	1.974					
	Product Quality	.069	.072	.067	.966	.335	-.072	.211	.202	.069	.063	.884	1.131
	Transaction Security	-.030	.075	-.027	-.400	.690	-.177	.118	.065	-.029	-.026	.955	1.047
	Timely Delivery	.435	.075	.407	5.799	.000	.287	.583	.424	.385	.377	.857	1.167

a. Dependent Variable: Trust



6. COMMENTS AND CONCLUSION

In this research it was analyzed that all the factors have impact on building trust in online shopping. The findings of this research showed that specifications of a product are very important for online customers. Customers would not shop from that online store again that has provided them defective product. In this way product quality plays an important role in developing trust. Similarly transaction security also has an impact on building trust as findings showed that customers are very much concerned about their information not being secured and hesitate to pay for their purchased products through credit card. Since the concept of online shopping has started to gain grounds in Pakistan over a past few years people are still getting used to this idea and are trying to get as much benefit from it as they can. The last factor i.e. timely delivery of the product has the strongest relationship with trust.

REFERENCES

1. Adnan, H. (2014). An analysis of the factors affecting online purchasing behaviours of pakistani consumers. *International journal of Marketing Studies,* 6 , 16.
2. Hong, I.B. and Cha, H.S. (2013). The mediating role of consumer trust in an online merchant in predicting purchase intention. *International Journal of Information Management*, 33(6), 927-939.
3. Stockburger, D.W. (1996). *Introductory Statistics: Concepts, Models, and Applications*. Missouri: David W. Stockburger.
4. *Explorable*. (2008). Retrieved from https://explorable.com/
5. Shergill, G.S. and Chen, Z. (2005). Web-Based Shopping: Consumers' attitudes Towards Online Shopping in New Zealand. *Journal of Electronic Commerce Research*, 6(2), 78.
6. Van der Heijden, H., Verhagen, T. and Creemers, M. (2003). Understanding online purchase intentions: contributions from technology and trust perspectives. *European Journal of Information Systems*, 12(1), 41-48.
7. Adnan, H. (2014). An Analysis of the Factors Affecting Online Purchasing Behavior of Pakistani Consumers. *International Journal of Marketing Studies*, 6(5), 133.
8. Koyuncu, C. and Bhattacharya, G. (2004). The impacts of quickness, price, payment risk, and delivery issues on on-line shopping. *The Journal of Socio-Economics*, 33(2), 241-251.
9. Ling, K.C., Chai, L.T. and Piew, T.H. (2010). The effects of shopping orientations, online trust and prior online purchase experience toward customers' online purchase intention. *International Business Research*, 3(3), 63.
10. Rane, P.B. and Meshram, B.B. (2012). Transaction Security for E-commerce Application. *International Journal of Electronics and Computer Science Engineering*. Available Online at: www. ijecse.org ISSN-2277-1956. pp, 1720-1726.
11. Das, Ravi. (2012). *Threats to E-Commerce Servers*. Technology Executives Club. Web. 3 Feb. 2012. http://www.technologyexecutivesclub.com/Articles/security/artThreatstoEcommerceServers.php.
12. Chaudary, S., Rehman, M.A. and Nisar, S. (2014). Factors Influencing the Acceptance of Online Shopping in Pakistan. *The Lahore Journal of Bussiness*, 23.
13. Wikinvest (2009, March 31). Retrieved from wikinvest: http://www.wikinvest.com/stock/Lifetime_Brands_(LCUT)/Companyrsquos_Ability_Deliver_Products_Customers_Timely_Manner_Satisfy

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 363-372

ON THE LABOR RIGHTS THE EFFECTS OF CORRUPTION IN CONSTRUCTION BUSINESS

Muhammad Wahaj Khan, Muhammad Abdul Waheed Haris, Asif Amin and **Deepak Kumar**
National University of Computer & Emerging Sciences – FAST, Karachi, Pakistan
Email: k142026@nu.edu.pk
waheed.haris@nu.edu.pk
k142103@nu.edu.pk
k142164@nu.edu.pk

ABSTRACT

Human, is consider the wisest creature in the entire universe. The modern era divided the people equally and distributed the people with some rights. In every aspect of life the society standardized some status and sort of rights and laws on which the independence and freedom of the person occur. But what happen when these rights are usurp by the greedy behavior of human and make the life of the innocent in danger. Unfortunately, the similar situation is now facing by the labor society of Pakistan due to the greedy and corrupt behavior of their supervisors. The "Labor" sector of the country is also one of them, that play a vital role in the society development, but in Pakistan the situation is even worst. The problem is due to the corruption in the governmental level it spread its roots to the every major aspect of development even it covers the construction business in which many workforce of labor working on the daily wages and due to corruption, they don't get their due share and sometime it may cost their own life. This study mainly focused on the effects of corruption on the labor rights in the construction business. The wages and safety measures have been considered as the influential factors in measuring labor rights. We have targeted 150 labor class people working in different construction sites in Karachi. The data was generated through a multifactor rating scale questionnaire and analyzed using SPSS. Regression analysis and correlation were used to conclude the results. The findings concluded that there is a negative relationship between the usage of labor rights and involvement of corruption in it.

KEYWORDS

Labor Safety, Labor Rights, Wages, Linear regression, correlation, corruption.

1. INTRODUCTION

The objective of our paper is to prove the presence of corruption in the labor rights in the construction business, because we strongly believe that the labor are the most respectful sector for any country and their rights should be fully implemented.

Background: Human, is considered the wisest creature in the entire universe. In today's modernize the world, people have distributed themselves; but these distributions of the people are based on some rules and regulations; that rule represent person authorities and independence in their respective aspect of life. But, what happen? When these rights are usurp by the greedy behavior of human and make the life of the innocent in danger. Unfortunately, the similar situation is now facing by the labor society of Pakistan because of the greedy and corrupt behavior of their supervisors. The “Labor” sector of the country is one of the most important sectors for every country; because the economic condition and development of every country depend on their shoulders. And we should have to admire the labor a right ordinances that have been passed in many countries of the world; and also the depiction of their implementation seems to be great in the European countries, but in Pakistan the situation is even worst. The problem is due to the corruption in the governmental level, it spread its roots to the every major aspect of development even it covers the construction business in which many workforce of labor working on the daily wages and due to corruption, they don’t get their due share and sometime it may cost their own life. Thus, now due to the increasing importance of this issue, we decided to cover the impact of corruption on the most common factors in the labor rights which totally neglected in the past decades. These factors are:

1) Labor Wages
2) Labor Safety

Our main hypothesis is get the answer of the questions that, is there is corruption factor affects the labor rights in construction business, is there is corruption in the wages of labor in construction business and is there is corruption in the labor safety in construction business. This paper will prove the affects of corruption in labor rights and also answered all our hypothesis questions.

Outline: This paper has been divided into parts in order to give complete information to a reader regarding the corruption affects on the labor rights in construction business. Part 1 contains the Introduction and objective of the research, Part 2 contains the literature review, Part 3 contains the Methodology adopted to get the data, Part 4 contains the Results and Findings which are obtained from the data using the linear regression model.

2. LITERATURE REVIEW

Corruption:

Definition of Corruption:

Corruption is defined as; for the sake of private benefits the person misuses his authorities (Javaid, 2010). This greedy behavior of human can easily be seen as common in the under-developed countries. Pakistan being a developing country is critically under the influence of the corrupt, almost every sector of Pakistan is immersed in the corruption, including the political and nepotism sector. In fact, we can see the environment of corruption with full color (Chene & Marie, 2010). And with the passage of the effects of corruption is increasing.

As prescribed by the World Bank Report (WB, 1998) there are four causes that increased the chances of Corruptness in the developing country. In global perspectives we can say that the corruption is based on the behavior of greed, of man, because he wants to earn more with limited resources, secondly rents of monopoly increased the chance of corruption, similar the unavailability of the proper laws and weak political and civil bodies are responsible for the corruption (WB, 1998).

Theories on Corruption:

The literature on the corruption completely depicts its image as an inhuman behavior. The corruption as an independent variable was studied by (Javaid, 2010), in which the survey organized by the NACS and NAB concluded that the percentage of corruption in Pakistan is rapidly increasing. Similarly, corruption as the dependent variable was studied by the (Khan, Masood, Riaz, & Dubnick, 2004) in which they considered that the impacts of corruption can be reduced through a proper concern and monitoring process. The new policy order (GOP, 2002) 2002 was passed in Pakistan in order to pursue and rectify the corruption status.

Labor Rights:

The empirical literature collected on the labor rights has been molded into the theoretical perspectives, especially where the Labor rights are considered to be the exogenous intervention and a line of interference in the market forces (Deakin, 2009). As we move forward and consider the labor rights as an effective role playing in the society development and economical growth, then we come to the fact that these all labor force and rights are controlled by the select group of the people and consider as an external imperfection source (Deakin, 2009). The inefficiencies in the wages regulation and rules in the labor rights create a new depression for the labor demands, and ultimately it affects the labor rights by forcibly tends to make the labor to work for a lower wages and also neglects many of the other labor rights (Minford, 1985).

Labor Rights in Developing Countries and Ordinance:

Labor right in the under-developing countries are likely to be said under the great influence of the economic globalization, and it can be said that all the labor rights are likely to be mixed in these developing countries (Gallagher, 2005). Every country has set its rules and regulation for the labor, no matter which sector of the government it belongs (Mosley, 2006). Ordinances have been passed on the labor rights in order to meet the current requirements of the economic globalization. These ordinances revisit and reviewed every year and amendments bring with it. But the question which always arises is about the implementation of these ordinances on the labor. Because in most of the under-developed countries, there are most likely to have groups that control the labor sector and neglect the entire ordinance related to the labor rights (Pakistan, 2010).

Pakistan is also one of the countries that are under the influence of supremacy group mafia’s. According to the employers and labor bureau (Government, 2010) the labor force of Pakistan consists of approximately 37.15 million in which 47% in agriculture sector, 10.50% in manufacturing and the other 42.50 % associated with the other professions (Government, 2010). Thus, the Pakistani government has passed many ordinance form the 1965 to 2008. In the labor ordinance of Pakistan there are many rights have been enclosed, in which the most important is related to the “wages” and “safety” of the labor. The Payment of Wages act 1936, Minimum Wages Ordinance 1961 and Hazardous occupation rules 1978 (Government, 2010) are of greater importance, because these three ordinances are the most commonly ignored in the labor sector not only of Pakistan but also in many other countries in fact based on their ordinance.

Labor Rights Situation in Pakistan:

Labor rights in Pakistan are not well-settled because of the group supremacy system. In the Labor rights the two most affected factors are the labor wages and Labor safety. In Pakistan the system of wage distribution has been described in the ordinance and there is always a profound implication of wages on the labor utilization (AHMED, 1985). The wages of Labor are constantly divided on the basis of the sector either it is a formal or informal sector. But after the research of (AHMED, 1985) it can be concluded that the real wages of the labor are always lower than the nominal wages as prescribed by the ordinance of Pakistan (AHMED, 1985).

- **Labor Wages in Construction Sector:**

  The construction industry is the backbone of all the development activities and is estimated to employ over a million workers. Since the sector is witnessing a quick expansion, health, safety and occupational hazards in this industry are likely to pose new challenges and issues. In order to guard against occupational hazards and to provide safe working conditions for those employed in this vital sector of the economy, the Government shall enact suitable legislation to be definite health and safety of construction workers and to provide benefits obtainable to other formal sector workers. (Government, 2010)

  In the construction sector of Pakistan, the research conducted by NESPAK and the federal bureau of Statistics (Statistics, 1980) the labor wages were not so frequently increased and there was a declined of the labor wages in the construction sector from the year of 1970 to 1980 (AHMED, 1985). The wage factor in the labor rights is also highly effective by the attitude of the supervisor in the construction business. The supervisor strict and inhuman behavior forced labor to work for a lower wages and they consider the casual labor as a free labor (Gazdar, January 2004). The labors working in the construction sector are forced to get lower wages and due to which the corruption factor increase (Gazdar, January 2004).

- **Labor Safety in Construction Sector:**

  Labor safety is the most important factor in the labor rights and for the implementation of it the Government of Pakistan has passed in 2010 labor laws ordinance in which it is said that the hazards and safety protection should be given to the labor working in any sector. According to research conducted for 6 months in the year of 2010 the rate of sharp injuries per person in labor is 3.7 and this increases significantly year by year (Janjua, Khan, & Mahmood, 2010).

  Labor safety rights are getting more and more concern not only in Pakistan but also in the other developing countries. A research conducted on the Labor safety in the year of 2003 stated that the most of the problem arises due to the improper gadgets and equipments (Muhammad & Tam, 2009). The most commonly noted problems are the vibrating of the body, shift work, awkward postural requirements, etc., and among these the most fatal and dangerous is falling from the height due to the unavailability of the proper safety equipments.

Research Gap Analysis

The empirical literature collected on the corruption and labor rights is quite accurate on their own, but not significantly satisfied, because there is always a flaw of and a gap between them. For instance the study on the corruption affects researched by the (Javaid, 2010) is only related to the sectors that are politically under the influence of Government of Pakistan and there is no research analysis in it that represents the effects of corruption in the labor sector or in the construction business. Similarly the literature reviewed in the article of corruption published by the World Bank (WB, 1998) only stated the corruption and its causes in the developing countries. So, the gap in all these important corruption researches is their impact on the labor rights and also in the construction business.

Similarly; the literature on the labor rights are also not very clear. The article reviewed in case of labor rights is not completely discussing the rights of labor, for instance the article (Gazdar, January 2004) only show that the poverty is responsible for the decrease in the labor force and many other similar papers are stuck to only one problem i.e. the spoliation of the labor rights due to the factors of poverty, group mafia and etc.

These all research paper does not completely covers the two most important rights of labor, i.e. labor wages and labor safety. Our research analysis is based on the fact of that the Government of Pakistan has passed ordinance on the labor laws, which are completely affected due to the corruption. Our research problem is to identify and explore the effects of corruption on the labor rights in the construction business. The variables taken in our research as independent are the corruption and the dependent variable is the labor rights, in order to explore the effects of corruption, we decided to work on the most important sector of every developing country i.e. the construction business. The construction business clearly depicts the condition of labor rights as mentioned in the ordinance of labor laws 2010. Thus, by conducting a brief survey of the labor in construction sites we measure the effects of corruption, and bridging the gap which is not has been explored in all the related literature.

Methodology

Study Design and Participants:

We conducted the surveys in some important areas of Karachi. We surveyed through those areas mostly where we were sure about the presence of whistle blowing. The areas we covered for the surveys included Karimabad, Malir, and Gulshan-e-Iqbal. Generally the construction sites we visited have around 30 to 40 labor per area working there and busy in the construction of usually commercial buildings. Most of the supervisors we have met show there misbehaving attitude, that was the clear indication that there was a suspicious whistle blowing in that areas. We surveyed through 150 to 160 labors.

Questionnaire Design:

In order to validate the hypothesis of our research we designed the questionnaire in such a way that it measured the most important factors of labor rights with the effect of corruption. We used rating scale measurement technique, in which we defined the scale of 1 to 5, and marked each rating with a response action with for 1 being strongly agree to 5 being strongly disagree. For instance, for 1 we used strongly agree, for 2 we used agree, for 3 we used neutral, for 4 we used disagree and for 5 we used strongly disagree. The questionnaire also contained the demographic information of the personal including person's name, age and gender. We targeted to survey 150 to 160 labor in the construction sector.

Data Analysis:

We entered the collected survey data into IBM SPSS, and first calculate the age factors of the investigated people. Then using a likert scale analysis, we feed the data of our questionnaire with demographic information and then questions in the form of A1-to-A8 for labor rights estimation and from B1-to-B8 for corruption impact rights estimation. Then we computed the mean of labor rights as depended variable and corruption as independent variable. After the measurement of mean variables, we applied Linear Regression Model to analyze the affects of corruption on labor rights.

3. RESULTS AND FINDINGS

We surveyed through the 150 labor working in the construction business of different area, in which the average age of the labors is in between 20 to 29.

AGE

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
Below 19	12	8.0	8.0	8.0
20-29	57	38.0	38.0	46.0
30-39	49	32.7	32.7	78.7
Above 40	32	21.3	21.3	100.0
Total	150	100.0	100.0	

For the calculation, we used the corruption as an independent variable and the labor rights as a dependent variable. The first table shows the summary of our selected model in which the value of R^2 is (0.123) which is 12.3%, this shows that here in linear regression model there is 12.3% variance in the data. Hence the assumption for linear regression model of having independence of observations is true.

Table 1
Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.351[a]	.123	.117	.41300
a. Predictors: (Constant), Corruption				
b. Dependent Variable: LaborRights				

Now in the next table of ANOVA analysis, the value of F with regression shows that there are null hypotheses exist so there is now linear relationship between the two variables. But in our model of F-test having the value of F = 20.796 which is very significant and shows that the assumption can easily be made that there is a linear relationship between the two variables i.e. corruption and the labor Rights. Hence the assumption for equality of variance in linear regression is true and validates the effect of corruption in labor rights.

Table 2
ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.547	1	3.547	20.796	.000[b]
	Residual	25.245	148	.171		
	Total	28.792	149			
a. Dependent Variable: Labor Rights						
b. Predictors: (Constant), Corruption						

Now, as me move forward the table 3 it shows the regression coefficients, the intercepts used and the in the same way in our model the significance of our coefficients and intercepts. Hence we can say that according to the table 3 the negative value of corruption displayed that in every sample of 150 labors working in the construction business there is a negative impact of the corruption in the labor rights with the significant value of 0 or null. Hence the regression equation will be:

LaborRights = slope x corruption + intercept

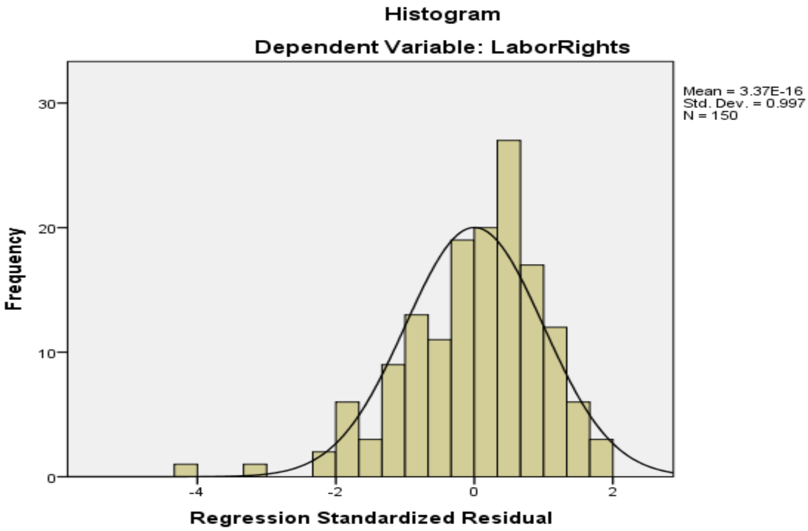
So, now LaborRights = -0.481 x corruption + 4.290.

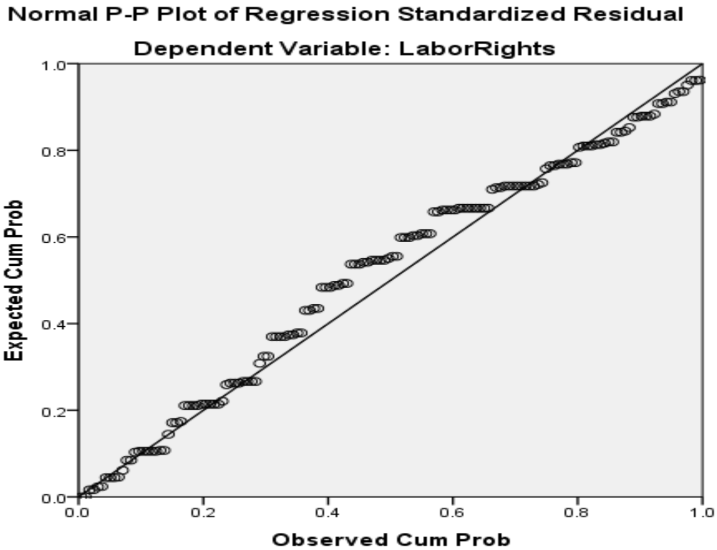
If the labor mark the corruption question 1 = 3 then according to the model the value is -0.481 x 3 + 4.290 = 2.847 i.e. $\lfloor 2.847 \rfloor \sim 2$ so if the person has been asked question 3 in corruption section, that corruption affects safety measures and gadgets then his answer would be "agree" because 2 refers to agree. So the result is accurate and clearly depicts the correctness of the model of linear regression.

Table 3
Coefficients[a]

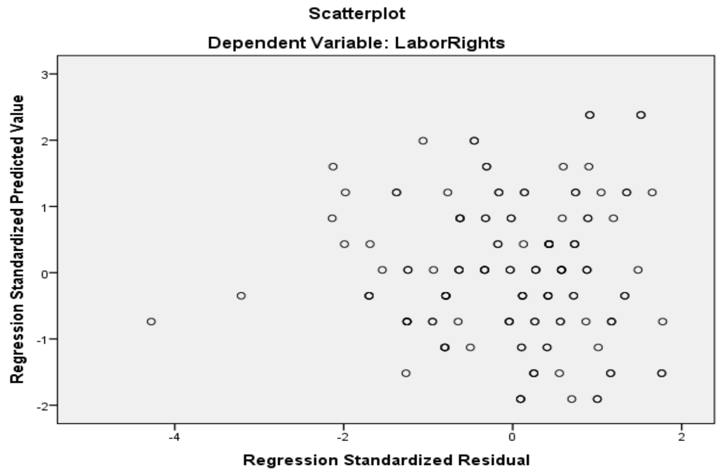
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.290	.202		21.224	.000
	Corruption	-.481	.106	-.351	-4.560	.000
a. Dependent Variable: LaborRights						

Now, the last assumption of linear regression model, in our model the plot show the residual approximate a normal distribution can easily be observed in the histogram. Hence the assumption for homoscedasticity for linear regression is true.





The Q-Q plot between Z*PRED and Z*PRESID display that there is no tendency in terms of error in our linear regression analysis, the factor of corruption is involved in the labor rights. Thus our hypothesis that there is corruption in the labor rights in construction business is validating.



Correlation analysis also shows the negative value, it means that when the corruption factor increases then the negative impact will be generated on the labor rights.

Descriptive Statistics

	Mean	Std. Deviation	N
Corruption	1.8883	.32062	150
Labor Rights	3.3808	.43958	150

Correlations

		Corruption	Labor Rights
Corruption	Pearson Correlation	1	-.351**
	Sig. (2-tailed)		.000
	N	150	150
Labor Rights	Pearson Correlation	-.351**	1
	Sig. (2-tailed)	.000	
	N	150	150
**. Correlation is significant at the 0.01 level (2-tailed).			

ACKNOWLEDGEMENT

I would like to thank our supervisor Mr. Abdul Waheed Haris and Mr. Ibrahim Shamsi, because of their guidance, encouragement and stimulating suggestion we have got enough potential to compile this report. I would also thankful and wants to give my special gratitude to my team mates Asif Amin and Deepak Kumar who helped and contributed a lot to in the making of this report. And last but not least a very special thanks to the labor sector who give us their precious and golden time to cover their rights situation.

REFERENCES

1. Irfan, M., Ahmed, M.A. and Khan, M.F. (1985). Real Wages in Pakistan: Structure and Trends, 1970-84. *The Pakistan Development Review*, 24(3/4), 423-440.
2. Chêne, M. (2008). *Overview of corruption in Pakistan*. Transparency International.
3. Deakin, S. (2009). The Evidence-Based Case for Labour Regulation. *Regulating Decent Work Conference,* Geneva: ILO.
4. Gallagher, M.E. (2005). Contagious Capitalism: Globalization and the Politics of Labor in China. *Princeton University Press*.
5. Gazdar, H. (2004). Labour markets and poverty in Pakistan: Institutional Arrangements and Policy. ILO/UNDP.
6. Government of Pakistan (2002). New Policy Order. *Government of Pakistan (Punjab)*.
7. Government of Pakistan (2010). Pakistan Employment Laws. *The Employer's and Labour Bureau*. Government of Pakistan.
8. Janjua, N.Z., Khan, M.I. and Mahmood, B. (2010). Sharp Injuries and their Determinants among Health Care Workers at First-Level Care Facilities in Sindh Province, Pakistan. *Tropical Medicine & International Health*, 1244-1251.
9. Javaid, U. (2010). Corruption and its deep impact on good governance in Pakistan. *Pakistan Economic and Social Review*, 123-134.
10. Khan, M., Kakakhel, N.A. and Dubnick, M. (2004). Prosecuting Corruption: The Case of Pakistan. In *Ethics Forum, American Society for Public Administration* (26-27).
11. Minford (1985). *Unemployment: Cause and Cures*. Oxford.
12. Mosley, L. (2006). *Racing to the Bottom or Climbing to the Top*? Chapel Hill, NC 27599-3265.
13. Muhammad, S. and Tam, V.W. (2009). National Culture and Safe Work Behaviour of Construction Workers in Pakistan. *Safety Science* DOI: 10.1016/J.Ssci.2008.01.003.
14. Government of Pakistan (2010). The Employer's and Labour Bureau of Pakistan.
15. NESPAK and Federal Bureau of Statistics (1980).
16. World Bank (1998). World Bank Report on Corruption and Development. *World Bank*.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 373-380

IMPACT OF OPERATING SYSTEM ATTRIBUTES ON CUSTOMER SATISFACTION

Muhammad Zubair Hasan, Muhammad Jamil Usmani, Hamza Godil and **Shariq Masood**
National University of Computer & Emerging Sciences – FAST, Karachi, Pakistan
Email: k142257@nu.edu.pk
m.jamil@nu.edu.pk
k142187@nu.edu.pk
k142131@nu.edu.pk

ABSTRACT

Over the past we have seen that many operating systems have failed to do business in market due to various factors. For example, Vista failed due to lack of stability and Windows 8 failed due to GUI. Hence, there is a need to find the attributes which cause an operating system to do sufficient business in market. It is a need of a computer scientist as a developer to perfectly understand the requirements of the customers. In this study we are trying to analyze the needs of customer related to the operating system. To do so, we are trying to cover all of the most major factors which cause the customer dissatisfaction, so that the future operating systems can be more innovative and updated to modern technology along with covering all the aspects of customer needs. To guide the OS industry to produce such OS which will get approval of customer. Factor analysis was performed to extract the dimensions in the nine attributes User Interface, reliability, Glitches, bugs, processing speed, security, user friendly, size of OS, interact with I/O devices. Almost 250 computer scientist were selected to answer these questions. The result concluded that GUI, Speed and Stability are the major concern of the customer and due to weakness in these three attributes many operating systems in the past have failed. Hence, in order to produce an operating system which will win users approval these factors should be kept in mind by the developers.

KEYWORDS

User Interface, reliability, Glitches, bugs, processing speed, security, user friendly, size of OS, interact with I/O devices, Stability, Factor Analysis,

1. INTRODUCTION

Computer was the most trending invention of late 1900's and even in this era computers are the most availed technology. For those who have even the slightest knowledge of computers are aware of the word Operating Systems and are aware that Operating Systems are the essentials part of this technology. From 1980 up till now many operating systems have been introduced under the hood of brands like Windows, Apple and Linux. Many of those Operating Systems like, Windows 98, Windows XP and Windows 7 were great hits and did astonishing sales in market, but there were fails to like, Windows 2000, Windows Vista and Windows 8. Interestingly all of those failures were faced with in just last 15 years; hence in present there is a need of finding the key areas of operating system which are causing it to fail to satisfy its users.

The objective of this research is to find the key areas of Operating System which are important to customers and to find what developers are doing wrong in those areas and also a solution to change those wrongs to rights. Hence, to do so the researchers of this paper have formulated a strategy to focus on 8 attributes (key areas) of OS which are the most important to a Computer Scientists. Hence the sample space of this research is Computer Scientists (Either Students, Teachers or Employs) and the sample size is 250. More over following are the research questions to this research:

1. Which sort of **User Interface** is more popular among users and how to innovate it with modern technology?
2. To what extend **Glitches** and **bugs** effects OS's sales
3. How much **processing speed** matters to customers?
4. How much concerned users are related to OS **security**?
5. How much **user friendly** the successful OS were?
6. How **size** of OS effect its other attributes and its market?
7. How users want OS to interact with **I/O devices and** what problems user face with I/O devices?
8. What is the level of importance of the above mentioned attributes?

2. LITERATURE REVIEW

Operating System

Operating System is an interface and resource allocator between the resources of the computer and applications (Loidl, 2012). The job of an Operating System is to provide user programs with a better, simpler, cleaner, model of the computer and to manage the software and hardware component of Computer (Tanenbaum, 2009).

For the purpose of our research, researcher has taken few attributes of OS in consideration.

User Interface: User interface is a part of an OS, which helps connecting user with OS itself. It is considered to be an essential part for computing. A good user interface of an operating system means more ease in use for the user. (Ben Shneiderman, 2008). Graphical Interface, Gesture Interface and Metro Interface are discussed in this research paper.

Bugs: A bug in the terminology of Computer Science is defined as a coding or designing error which causes a program to crash or to produce unexpected results. Recent studies have shown that bugs are becoming very common in large software like Operating Systems due to replicated code and are causing a lot of run time issues (Li, 2006).

Glitches: Glitches in computer terminology have more than one meaning, however, in general it is considered as a suddenly arose problem in the functionality of the running system. Glitch can be of hardware, software or network (Sipser, 2001).

Size of Operating System: In modern operating systems there is a concept of updates due to which the OS is rapidly growing in size. If such operating system is on an hardware that lacks in memory resources than there are great chances that the working capability of OS is affected (Shinder, 2007).

Processing Speed: Operating System is a complex program made for handling process. Processing speed of OS is an important aspect and to make an OS run properly, a process scheduling algorithm whit not only utilize the hardware of OS properly but also avoids starving (Silberschatz, 2009).

I/O Interfacing: I/O interfacing is a mean of connecting peripheral devices with Operating System. Management of I/O devices so important task that, entire subsystems are used to perform it (Silberschatz, 2009).

Security: Security of an Operating System is a mean to control the access of the user other than super user. (Loidl, 2012). OS are complex programs and usually have security issues, so malwares and virus takes advantages of these loopholes to perform computer crimes. (Loidl, 2012)

User Friendliness: The level of User friendliness of a system is the level of ease of using the system via user. A user friendly interface consist of four major attributes, windows, icons, menus that are easy to understand , systems that are straightforward and easy to use (Perry, 2002).

Customer Satisfaction: (Shepherd, 2015)Shed light on the comparison between Windows and Linux. He shared some basic features and differences between both operating systems providing pros and cons regarding each attribute. For instance, let us consider the software compatibility for both. Most of the programs are predominantly written for windows, many of them are also available for Linux, but still there is a vast collection of software that are not available for Linux. On the other hand, in terms of security, Linux has a stronger hand over the security than Windows has. Linux protects the data efficiently which attracts security conscious individuals. If we talk about the user friendliness, Windows is clear-cut above the competition.

(Brashear, 2015) a survey was conducted regarding the use of Linux. Majority of the users use Linux on non-server computers and it is used majorly for fun. Among all the versions of Linux, Arch-Linux is the most powerful one followed by Ubuntu. Linux is mostly used on virtual machines and laptops whereas on server computers, window is more popular. Those server computers that use Linux majorly run Debian and Ubuntu on them. In terms of graphics, Gnome is preferred whereas Unity is the most disliked graphical environment. Like windows, Intel is the most preferred processor to run Linux and NVIDIA leads in terms of graphical use of Linux.

(Havanstein, 2008) stated that although Windows Vista was a newer version than Windows XP, users still preferred Windows XP over Windows Vista. (McCracken, 2012) stated that as per recent surveys windows 7 is preferred by users.

In the above books and articles there are few missing aspects which are the focus of this paper. The books and articles under the Operating System title have not discussed the intensity of effect the stated attributes have over customers. These cited books and articles just focus on the problems and their solutions. As for articles and books under the heading of Customer satisfaction target not all the above mentioned attributes in Operating System heading, but more importantly they are just targeting a particular Operating System model like, Windows XP, Windows Vista, Windows 7. However, in this research, the researcher is targeting on Operating System as whole instead of any particular model.

3. METHODOLOGY

It was one of the descriptive studies including both quantitative as well as qualitative technique. For making background and highlighting the different views of the different researchers, research papers, articles are used for the references, which highlights how operating system attributes impacting the customer satisfaction and to support that background the statistical tools and tests are used as well. The field experiment was conducted in the non-contrived environment with excessive researcher interference in the working setting, the unit of analysis was collected from the primary data collection through Questionnaire made on Likert scale and that is considered to be employees working in different organizations in the field of computer science, students who are studying computer science and teacher for the same faculty, are targeted, where the time horizon was cross-sectional in the nature because the researcher has a limited time to complete the study. The data was tested with the help of SPSS and Regression to reach some concrete conclusion.

All the manager and officer level employees working in different organization and are related to perform their duties in computer science, so that the students and teachers of different universities related to the faculty of computer science.

Approximately 250 sample size has taken including employees, students and teachers of different universities related to computer science faculty were selected randomly.

- Primary data has been collected by survey.
- Primary source is employees, Students and teachers of Computer Science
- Primary data has been collected by questionnaire. The questionnaire was sent to main email domain of respective companies' including different levels of employees. The reminder data were collected by personal surveys. Data was collected and feed in SPSS customized database for analysis and result conclusion.

Secondary resources like websites, published journals, previous research papers, and Pakistan researcher papers publications.

As per the requirement of data collection and scale use in the questionnaire standard tools use to analyze and process the data.

The standard statistical tools used for the Hypothesis Analysis

- Frequency Distribution
- Multiple Linear Regression

For processing and analyzing the data SPSS is used.

4. ANALYSIS AND DISCUSSION

The survey conducted by the researcher provides the following statistical information regarding these three factors i.e. occupation, age and gender.

Table 4.1.1
Occupation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Student	204	81.6	81.6	81.6
	Teacher	26	10.4	10.4	92.0
	Employee	20	8.0	8.0	100.0
	Total	250	100.0	100.0	

Table 4.1.2
Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Below 20	75	30.0	30.0	30.0
	20-30	160	64.0	64.0	94.0
	30+	15	6.0	6.0	100.0
	Total	250	100.0	100.0	

Table 4.1.3
Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	31	12.4	12.4	12.4
	Male	219	87.6	87.6	100.0
	Total	250	100.0	100.0	

The following table shows the relationship between the dependent variable customer satisfaction and independent variable User interface, size of OS, Processing speed, I/O interface, security, user friendliness, bugs and Glitches now as we can see that Customer satisfaction having 0.196 with user interface which means there is a weak relationship between Customer satisfaction and user interface where as there is a good relationship between User interface, size of OS, Processing speed, I/O interface, security, user friendliness, bugs and Glitches with customer satisfaction, means individually they are making the relationship with their dependent variable.

Furthermore, if we see the significance of the same table the dependent variable is showing the significant relation with all the independent variables, means all these factors has an impact with customer satisfaction.

Table 4.2.1.1

Correlations

		Customer Satisfaction	UserInterface	SizeoOS	Processing Speed	IOInterface	Security	User Friendliness	Bugs	Glitches
Pearson Correlation	CustomerSatisfaction	1.000	.196	.386	.508	.398	.138	.392	.401	.441
	UserInterface	.196	1.000	.145	.200	.289	-.022	.103	.207	.254
	SizeoOS	.386	.145	1.000	.430	.250	.119	.194	.339	.335
	ProcessingSpeed	.508	.200	.430	1.000	.395	.122	.283	.382	.380
	IOInterface	.398	.289	.250	.395	1.000	.245	.222	.204	.211
	Security	.138	-.022	.119	.122	.245	1.000	.165	.286	.046
	UserFriendliness	.392	.103	.194	.283	.222	.165	1.000	.195	.227
	Bugs	.401	.207	.339	.382	.204	.286	.195	1.000	.390
	Glitches	.441	.254	.335	.380	.211	.046	.227	.390	1.000
Sig. (1-tailed)	CustomerSatisfaction	.	.001	.000	.000	.000	.014	.000	.000	.000
	UserInterface	.001	.	.011	.001	.000	.363	.053	.000	.000
	SizeoOS	.000	.011	.	.000	.000	.030	.001	.000	.000
	ProcessingSpeed	.000	.001	.000	.	.000	.027	.000	.000	.000
	IOInterface	.000	.000	.000	.000	.	.000	.000	.001	.000
	Security	.014	.363	.030	.027	.000	.	.005	.000	.235
	UserFriendliness	.000	.053	.001	.000	.000	.005	.	.001	.000
	Bugs	.000	.000	.000	.000	.001	.000	.001	.	.000
	Glitches	.000	.000	.000	.000	.000	.235	.000	.000	.
N	CustomerSatisfaction	250	250	250	250	250	250	250	250	250
	UserInterface	250	250	250	250	250	250	250	250	250
	SizeoOS	250	250	250	250	250	250	250	250	250
	ProcessingSpeed	250	250	250	250	250	250	250	250	250
	IOInterface	250	250	250	250	250	250	250	250	250
	Security	250	250	250	250	250	250	250	250	250
	UserFriendliness	250	250	250	250	250	250	250	250	250
	Bugs	250	250	250	250	250	250	250	250	250
	Glitches	250	250	250	250	250	250	250	250	250

In the following table the value of R is coefficient of correlation between the variables, and it is showing the value 0.660 which means it has a strong correlation between the variables. Whereas the adjusted R square s showing the value 0.436 which means that dependent variable is explained by all the independent variables and having the variation of 0.436 in dependent variable due to the independent variables.

Furthermore, the adjusted R square is showing the reliability of R square, the value of adjusted r square is 0.417 as compare to r square the change is approx. 0.02 which is less than 0.1 so it is concluded that the results of R square is reliable.

Table 4.2.2.1

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.660[a]	.436	.417	.38929	.436	23.303	8	241	.000

a. Predictors: (Constant), Glitches, Security, User Interface, User Friendliness, Size of OS, IO Interface, Bugs, Processing Speed

In the following table it showing that our F value is 23.303 which is greater than 10, so it is concluded that our model is statistically significant and the model is good fit. Furthermore, the sig value is less than 0.05 which means there is a significant relationship between the dependent and independent variables.

Table 4.2.3.1
ANOVA

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	28.253	8	3.532	23.303	.000[a]
	Residual	36.524	241	.152		
	Total	64.776	249			

a. Predictors: (Constant), Glitches, Security, User Interface, User Friendliness, Size of OS, IO Interface, Bugs, Processing Speed
b. Dependent Variable: Customer Satisfaction

In the following table we can express the multi linear regression in the form

$$Y = a + b_1 X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6 X_6 + b_7X_7 + b_8 X_8$$

$$Y = 1.074 - 0.009X_1 + 0.076X_2 + 0.169X_3 + 0.140X_4 - 0.022X_5 + 0.151X_6 + 0.104X_7 + 0.120X_8$$

The values of user interface, size of OS, processing time, I/O interfacing, security, user friendliness, bugs and glitches are **-0.009, 0.076, 0.169, 0.140, -0.022, 0.151, 0.104, 0.120** respectively. 1.074 is the value of **'a'** showing the constant. The impact on dependent variable will be positive if each independent variable is increased by one unit, because all the independent variables are positive except user interface and security.

This means if all the independent variables are eliminated except user interface and security, then there will be negative influence on dependent i.e. if the user interface and security are included in operating system alone then they will have negative impact on customer satisfaction. Although it has been observed that user interface and security effect the dependent variable only when taken with all other independent variables.

Table 4.2.4.1
Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
	B	Std. Error	Beta			Zero order	Partial	Part
(Constant)	1.074	.238		4.518	.000			
User Interface	-.009	.032	-.014	-.269	.788	.196	-.017	-.013
Size of OS	.076	.041	.102	1.838	.067	.386	.118	.089
Processing Speed	.169	.048	.211	3.504	.001	.508	.220	.169
IO Interface	.140	.042	.185	3.299	.001	.398	.208	.160
Security	-.022	.039	-.030	-.563	.574	.138	-.036	-.027
User Friendliness	.151	.038	.206	3.998	.000	.392	.249	.193
Bugs	.104	.041	.145	2.512	.013	.401	.160	.121
Glitches	.120	.036	.189	3.349	.001	.441	.211	.162

a. Dependent Variable: Customer Satisfaction

5. COMMENTS AND CONCLUSION

After analysis of the conducted survey from 250 Computer Scientists, in which majority were students, the researchers have reached to the conclusion that 5 out of 8 attributes are significant, while other 2 are less important. Processing Speed on the top of the list of the attributes, while Glitches stand second, Bugs third, followed by I/O Interfacing, User Friendliness, Size of OS, User Interface and Security. In a nutshell, user is keener towards an operating system which runs his/her application with ease rather than being keen towards an Operating System which has high security and good interfacing.

6. ACKNOWLEDGEMENT

We want to show our sincere gratitude to all those who made this study possible. Our first and foremost humble gratitude to “ALLAH” the Almighty for giving us the courage to remain dedicated to make this research paper. Then we would like to express my thankfulness to my course supervisor **Mr. Ibrahim Shamsi** and our project supervisor **Mr. Muhammad Jamil Usmani**. We are very thankful to our supervisor for investing his precious time to discuss and criticize this study in depth, and explained the meaning of different concepts and how to think when it comes to problem discussions and theoretical discussions.

REFERENCES

1. Ben Shneiderman, C.P. (2008). *Designing User Interface.* Pearson Education.
2. Brashear (2015). /r/ distribution survey. *Fighting the System*.
3. Havanstein, H. (2008). Developers Prefer XP Over Vista, Survey Shows. *Computer World*.
4. Li, Z. (2006). CP-Miner. *IEEE*, 176-192.
5. Loidl, H.-W. (2012). *Operating System Security.*
6. McCracken (2012). The State of Windows 7 Satisfaction. *technologezir*.
7. Perry, T. (2002). Of mice and menus: designing the user-friendly interface. *IEEE Spectrum*.
8. Shepherd, A. (2015). Windows Vs Linux: what's the best operating system? *ITPRO*.
9. Shinder, D. (2007). 10 Things that can go wrong when you upgrade your operating system. *10 Things*.
10. Silberschatz, A. (2009). *Operating System Concepts.*
11. Sipser (2001). *Computing Fundamentals.*
12. Tanenbaum, A. (2009). *Modern Operating Systems, Third Edition.* Pearson Education Inc.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 381-392

LOSS GIVEN DEFAULT (LGD) & EXPOSURE AT DEFAULT (EAD) IN THE VIEW OF BASEL III

Sohaib Ahmed, Syed Wajid Ali and **M. Toobab Zubair**
Department of Actuarial Science and Risk Management,
University of Karachi, Karachi, Pakistan
Email: isohaibahmed@gmail.com

ABSTRACT

The New Basel Accord allows internationally active banking organizations to calculate their credit risk estimation techniques using an internal rating based (IRB) approach, subject to supervisory review. One of the modeling components is loss given default (LGD), the credit loss incurred if an obligor of the bank defaults. The parameter loss given default (LGD) of loans plays a crucial role for risk-based decision making of banks including risk-adjusted pricing. Depending on the quality of the estimation of LGDs, banks can gain significant competitive advantage. For bank loans, the estimation is usually based on discounted recovery cash flows, leading to workout LGDs. In this paper, we reveal several problems that may occur when modeling workout LGDs, leading to LGD estimates. The flexibility to determine LGD values tailored to a bank's portfolio will likely be a motivation for a bank to want to move from the foundation to the advanced IRB approach. This paper gives the detail description of "Credit Risk Estimation Techniques" and proposes a portfolio credit risk model with loss given default (LGD) in the context of credit portfolio losses, which allows for a reasonable economic interpretation and can easily be applied to real data. We build up a precise mathematical framework and stress some general important issues when modeling dependent LGD, whereas EAD is taken as gross exposure in the event of obligor default, typically in 3 months.

1. INTRODUCTION

1.1 Brief Context

A Bank is a monetary organization, which is included in acquiring and loaning cash. Banks take client stores consequently to pay clients a yearly premium installment. The bank then uses the greater part of these stores to loan to different clients for an assortment of advances. The distinction between the two loan costs is adequately the overall revenue for banks. Banks assume a vital part in the economy for offering an administration for individuals wishing to spare. Banks likewise assume an essential part in offering account to organizations who wish to contribute and grow. These credits and business speculation are critical for empowering financial development.

With the exception of the extremely wealthy, very few people buy their homes in all-cash transactions. Most of us need a mortgage, or some form of credit, to make such a large purchase. In fact, many people use credit in the form of credit cards to pay for everyday items. The world as we know it would not run smoothly without credit and banks to issue it.

1.2 Formulation of Basel

The Basel Committee on Banking Supervision is an advisory group of managing an account supervisory power, which was set up by the national bank governors of the Group of Ten (G10) nations in 1975. It comprises of senior agents of bank supervisory powers and national banks from Belgium, Canada, France, Germany, Italy, Japan, Luxembourg, the Netherlands, Sweden, Switzerland, the United Kingdom and the United States.

The board is one of a few advisory groups set up under the support of the Bank for International Settlements (BIS). The Basel Committee meets frequently four times each year. It has around thirty specialized working gatherings and teams, which likewise meet routinely. The ten nations are each spoken to by their national bank furthermore by the power with formal obligation regarding the prudential supervision of managing an account business where this is not the National bank.

The Committee does not have any formal supranational supervisory power, and its decisions do not have legitimate power. Maybe, it figures wide supervisory norms and rules and prescribes articulations of best practice in the desire that individual powers will find a way to actualize them through point-by-point game plans that are most appropriate to their own national frameworks.

Lately, the Basel Accord has experienced different changes. The first accord distributed in 1988 is generally known as Basel I, and the new capital accord distributed in 2004 is normally known as Basel II.

In an effort to continue to strengthen the risk management frameworks of banking organizations and foster stability in the financial sector, the Basel Committee for Banking Supervision (BCBS) introduced, in December 2010, *Basel III:* A global regulatory framework for more resilient banks and banking systems.

1.3 Estimators of Credit risk under Basel III

Managing credit risk is a complex multidimensional problem and as a result, there are a number of different approaches in use, some of which are quantitative while others involve qualitative judgments. The key element is to understand the behavior and predict the likelihood of particular credits defaulting on their obligations. The Basel framework allows banks to assess their credit risk by using their own estimates of Loss Given Default (LGD).

There are various approaches for the estimation of credit risk, so let us take a review of these approaches that what is the difference of estimating credit risk among them.

1.3.1 Standardized

Measure credit risk as per altered risk weights taking into account outside credit evaluations (appraisals)

Least complex capital figuring's; minimum separation in required capital in the middle of more secure and riskier credits.

Generally, most elevated capital weights.

1.3.2 Foundation IRB

Measure credit risk utilizing advanced equations; inside decided inputs of probability of default (PD) and inputs altered by controllers of loss given default (LGD), exposure at default (EAD) and maturity (M).

More risk touchy capital prerequisites, more separation in required capital in the middle of more secure and riskier credits.

1.3.3 Advanced IRB

Measure credit risk using sophisticated formulas and internally determined inputs of PD, LGD, EAD and M.

Most risk sensitive (although not always lowest) capital requirements, most differentiation in required capital between safer and riskier credits.

Under Basel II and Basel III, banks have strong incentive to move to IRB status by improving risk management systems, thereby reducing required total regulatory capital.

One of the central pillars of the new Basel-III regulatory framework is the concept of risk-based capital requirements. Under the "Advanced" variant of the IRB (internal-ratings-based) approach, the amount of capital that a bank will have to hold against a given exposure will be a function of the estimated credit risk of that exposure. Estimated credit risk in turn is taken to be predetermined function of four parameters: probability of default (PD); loss given default (LGD); exposure at default (EAD); and maturity (M). Banks operating under the "Advanced" variant of the IRB approach will be responsible for providing all four of these parameters themselves, based on their own internal models.

Whereas, in "Foundation" variant of the IRB approach was only responsible for providing the PD parameter, with the other three parameters to be set externally, by the Basel committee.

1.4 Objective of the study

We will focus on the estimation of LGD. Historically, a lot of focus has been devoted to the estimation of PD while, LGD has received less attention and has sometimes been treated as constant. For a sophisticated risk management, LGD clearly needs to be assessed in more detail.

The estimation of LGD is preferably conducted by using historical loss data, but for certain portfolios and counterparties, there is a shortage of such data due to the high quality of the assets and the low number of historical defaults. Portfolios of this kind are often referred to as Low Default Portfolios (LDPs). LDPs include portfolios with exposures to banks, sovereigns and highly rated corporations.

The purpose of this research is to study quantitative models for estimation of LGD and empirically evaluate how these models work on LDPs in order to find a model that can be used in practice. For the model to be useful in practice, it must produce reasonable and justifiable values despite little default data. While the models are based solely on quantitative factors also qualitative considerations are taken into account when the models are constructed.

The benefit of a better credit risk assessment is twofold. First, it gives banks a better control over the risks they are facing and can be a support for business decisions. Secondly, internal models typically results in lower risk measures and thereby lower capital requirements. Since capital is costly, this is a direct benefit for a bank. On a more general level, society as a whole benefits from sound financial institutions with good credit risk assessments.

We will research on a financial data of corporate sector because the data of corporate sector can easily be collected and can give us better results. We want a model, which is standardized and fulfilling the assumptions of Basel III.

The outcome of the study will constitute of two parts, the first being an overview of the academic progress in this area and the second an evaluation of models from a practical perspective. The direct beneficiaries of the thesis are banks and financial institutions with the need to assess their credit risk exposure as explained in Basel III accord.

2. METHODS OF ESTIMATING LGD & EAD

A number of rating techniques and methodologies have evolved over time. The methodologies range from a spectrum of purely expert/professional judgment taking into account only qualitative factors, to a sophisticated statistical model based methodology solely taking into account the quantitative factors. Although the degree of subjectivity becomes lesser with the movement on the spectrum towards statistical methods, yet neither of the two extremes is advisable. An ideal internal risk rating system is based on both quantitative and qualitative factors concluding the decision based on many different attributes, involving the human judgment.

Banks are allowed to receive any of the methodologies/techniques keeping in perspective their size, multifaceted nature of operations and customer base. The methodologies/techniques should be adaptable to suit present and future risk profile of the bank, the foreseen level of expansion and modernity in loaning exercises. In any case, whatever the strategy utilized, the aftereffect of the assessment should be fit as a fiddle that gives important data, which can be further utilized for viable credit risk estimation and measurement of the credit exposure at an individual level and at a portfolio level. (Xiao Yao, Jonathan Crook and Galina Andreeva, 2014).

A rating methodology may be used based on asset class/product lines, e.g. corporate loans/consumer finance or based on line of credit, e.g. for over draft/running finance etc. Within each asset class, a bank may utilize multiple rating methodologies/systems. However, these methodologies should be able to be integrated in overall risk management system i.e. the ratings developed by different methods should be comparable with each other. When multiple systems are used, it is required that, the rationale for assigning a borrower to a rating system must be documented and applied in a manner that best reflects the level of risk of the borrower.

The internal risk rating system should be integrated with other systems of the banks such as portfolio monitoring, loan loss reserves analysis for provisioning, pricing of the loan, internal capital planning and return on capital analysis. Banks should not use separate rating systems for lending purposes, risk quantification and capital allocation.

2.1 Scope of Ratings

The internal risk rating is based on a two tier rating system.

1. **An obligor rating** based on the risk of borrower default and representing the probability of default by a borrower or group in repaying its obligation in the normal course of business and that can be easily mapped to a default probability bucket.
2. **A facility rating**, taking into account transaction specific factors, and determining the loss parameters in case of default and representing loss severity of principal and/or interest on any business credit facility. As we have focused in this research on facility rating

2.2 Loss Given Default LGD

The New Basel Capital Accord is designed to better align regulatory capital with the underlying risk in a bank's credit portfolio. It allows banks to compute their credit risk capital in two ways: a revised standardized approach based on the original 1988 Capital Accord, and two versions of an internal ratings based (IRB) approach whereby banks are permitted to develop and use their own internal risk ratings. The IRB approach is based on four key parameters used to estimate credit risks

1. *PD* The probability of default of a borrower over a one-year horizon
2. *LGD* The loss given default (or 1 minus recovery) as a percentage of exposure at default
3. *EAD* Exposure at default (an amount, not a percentage)
4. *M* Maturity

For a given maturity, these parameters are used to estimate two types of expected loss (*EL*).

Expected loss as an amount:

$$EL = PD \times LGD \times EAD, \qquad \text{(i)}$$

2.2.1 Measurement and Estimation of LGD

LGD is generally characterized as the ratio of losses to exposure at default, however not surprisingly; the overlooked details are the main problem. Once a default event has occurred, loss given default incorporates three sorts of losses:

a) The loss of principal
b) The carrying costs of non-performing loans, e.g. interest income foregone
c) Workout expenses (collections, legal, etc.)

There are broadly three ways of measuring *LGD* for an instrument:

A. Market *LGD*: Observed from market prices of defaulted bonds or marketable loans soon after the actual default event.
B. Workout *LGD*: The set of estimated cash flows resulting from the workout and/or collection process, properly discounted, and the estimated exposure.
C. Implied Market *LGD*: *LGD*s derived from risky (but *not* defaulted) bond prices using a theoretical asset-pricing model.

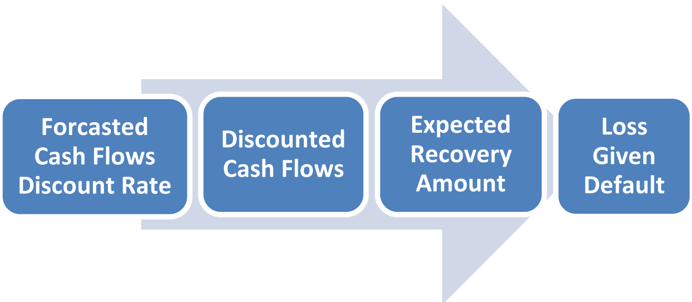
We will examine work out LGD in detail.

2.2.2 Workout LGD

LGD observed throughout a workout is more entangled than the straight forwardly watched market LGD. Consideration should be paid to the timing of the money streams from the bothered asset. Measuring this timing will affect downstream gauges of acknowledged LGD. The money streams ought to be marked down, yet it is in no way, shape clear which rebate rate to apply. For case, the obligation rebuilding could bring about the issuance of dangerous resources, for example, value or warrants, or less hazardous ones, for example, notes, bonds or even money. On a basic level, the right rate would be for an advantage of comparable danger. Essentially, once the obligor has defaulted, the bank is a financial specialist in a defaulted resource and ought to esteem it in like manner, perhaps at the bank's obstacle rate. Improper competitors incorporate the coupon rate (set ex bet of default, so too low) and the danger free (or Treasury) rate.

Workout LGD calculates the LGD based on the actual cash flows that can be recovered from the firm by the workout process, once the firm has defaulted. The Workout LGD methodology involves prediction of the future cash flows that can be recovered from the company, after the company has defaulted on its payments. It takes into account all cash flows from the distressed asset linked to the recovery. (Naeem Siddiqui, 2012).

2.3 Estimation Framework



Computation of workout LGD can be shown in Table 3.1 based on our model for LGD:

where,

EAD is Exposure at Default

τ Is the default time

m and T are the start and finish point of the workout process respectively

PV(R(t)) is the recoveries during the workout process

PV(C(t)) is the cost during the work out process

Though this formula looks simple, it is difficult to calculate. This is because there are subjective decisions involved such as timing of cash flows and discount rate to be considered (risk-free rate or a bank's hurdle rate). Furthermore, recoveries are not always in the form of cash, they can be illiquid securities or ones with no secondary market, leaving their values unclear. This means that, until all the recovered claims are sold –

which could take a long time – an accurate estimation of LGD cannot be calculated. However, their expected values can be used for calculation.

Although the calculation process is difficult, workout LGD is considered to precisely reflect the losses post default. Banks generally use the workout approach to calculate LGD for illiquid loans and market LGD approach for observable market prices. (Anthony Van Berkel and Naeem Siddiqui, 2012)

2.4 Exposure at Default EAD

An EAD model estimates the maximum amount that could be lost (assuming no recovery) if a default occurs. There are two different cases in estimating the EAD, depending on the permission of an off-balance sheet exposure. In the first case, there is only an on-balance sheet exposure, which means that the obligor is not allowed to increase the exposure. In this case, the EAD is equal to the on-balance sheet amount. In the second case, the obligor can increase its exposure with the off-balance sheet amount. Whereas, we have taken Exposure at Default (EAD) in our data as outstanding amount at default on 90 days past due, plus markup of the bank, which is the sum of KIBOR and Bank's return on the loan. (Edward Tong, Christophe Mues, Iain Brown, Lyn Thomas, 2015)

3. RESULTS OF THE ANALYSIS

3.1 Background

We have considered 100 different customers of corporate sector who have taken loan from the bank to fulfill their financial obligations, but they had defaulted 90 days past due. One year data is taken for estimating EAD & LGD. The variables in table 4.1, are used to calculate the Exposure at Default (EAD) & the Loss Given Default (LGD) are as follows;

- **Facility,** Although there are many facilities but we have only considered "Funded Facilities" as Cash Finance, Term Finance, Demand Finance and Running Finance.
- **Collateral,** The collaterals are against their corresponding facility as Fixed Asset (moveable & immoveable), Government Securities (PIBs & T-Bills), Documentary Collaterals, Mutual Funds and Stocks etc.
- **EAD,** Exposure at Default
- **Market Value,** The current value of collaterals in the market to liquidate
- **Forced Sale Value,** Amount recovered from their collaterals, which is often less than market value
- **Discount Factor,** which is used in calculating the present value of the recovered amount & the direct/indirect cost at the time of liquidation
- **Net Recovery,** Discounted recovered amount minus the discounted cost
- **LGD,** loss given default

(The concept of discounting of recovery and cost is extracted from Berkel, A.V. and Siddiqui, N. (2012) research).

Table 3.1 shows the EAD's & LGD's of 20 individual customers, but we have estimated EAD's & LGD's of 100 individuals.

Table 3.1
Workout LGD

S#	Facility	Collateral	Outstanding Amount at Default		CALCULATIONS					LGD
					RECOVERY AFTER DEFAULT			Cost	Net Recovery	
			Out. Amount	EAD	Market Value	Amount Recovered Force Sale Value	Dis. Amount			
1	CASH FINANCE	Fixed Assets (immoveable)	11,000,000	12,155,000	9,000,000	8,934,000	7,724,559	105,000	7,633,773	37%
2	TERM FINANCE	Fixed Assets (moveable)	15,000,000	16,575,000	17,500,000	17,000,000	14,494,701	265,000	14,268,754	14%
3	TERM FINANCE	Govt Securities	92,000,000	101,660,000	87,500,000	85,420,000	78,227,333	250,000	77,998,384	23%
4	TERM FINANCE	Documentary Collaterals	83,000,000	91,715,000	74,520,000	73,500,000	66,853,514	287,400	66,592,103	27%
5	DEMAND FINANCE	Mutual Funds	17,000,000	18,785,000	12,900,000	12,500,000	11,307,525	95,000	11,221,587	40%
6	TERM FINANCE	Mutual Funds	78,000,000	86,190,000	71,853,000	71,250,000	65,864,678	150,000	65,726,016	24%
7	DEMAND FINANCE	Stocks	66,000,000	72,930,000	65,212,500	64,000,000	56,639,580	145,000	56,511,256	23%
8	RUNNING FINANCE	Fixed Assets (immoveable)	76,000,000	83,980,000	73,569,000	72,000,000	64,227,739	95,000	64,142,994	24%
9	CASH FINANCE	Documentary Collaterals	34,000,000	37,570,000	24,987,400	23,500,000	21,492,361	250,000	21,263,719	43%
10	DEMAND FINANCE	Fixed Assets (moveable)	58,000,000	64,090,000	59,984,500	59,000,000	52,372,188	350,000	52,061,505	19%

Table 3.1 shows the Workout of EAD's & LGD's of 20 individual customers, but we have estimated EAD's & LGD's of 100 individuals.

3.2 Facility Rating Corresponding to their LGD's

Table 3.2
Facility Rating Corresponding to their LGD's

FACILITIES	LGD's	Rating
CASH FINANCE	53%	D
DEMAND FINANCE	48%	D
RUNNING FINANCE	37%	C
TERM FINANCE	44%	D

The Table 3.2 is based on facility rating criteria according to State Bank of Pakistan guidelines.

The model develop in this report can be used to find the individual's LGD and also gives the result of the overall portfolio's LGD of each facility, as we have worked on the funded facilities to find their LGD's and their facility ratings so this table shows the facility rating of the particular bank whose data was taken under modeling.

Our result shows that Cash Finance, Demand Finance & Term Finance facilities are under the rating of "D" because their LGD's lies between 40%-60%, whereas the Running Finance facility is under the rating of "C" as it lie between 20%-40%.

4. COMMENTS & CONCLUSION

The New Basel Capital Accord (Basel Committee on Banking Supervision) has been created with an objective to better adjust regulatory capital with the underlying risk in a bank's credit portfolio. The new Accord requires international banks to develop and use internal risk models for calculating credit risk capital requirement. It allows banks to compute their regulatory capital in two ways: (1) using a revised standardized approach based on the 1998 Capital Accord, which uses regulatory ratings for risk weighting assets, or (2) using an internal rating based (IRB) approach where banks are permitted to develop and use their own internal risk ratings.

The IRB approach is based on three key parameters used to estimate credit risk: PD – the probability of default of a borrower over a one-year horizon, LGD – the loss given default, the credit loss incurred if a counterparty of the bank defaults and EAD – exposure at default. These parameters are used to estimate the expected loss, which is a product of PD, LGD and EAD. Banks need to understand LGD, its components and various associated issues. This research contributes to propose a methodology to estimate loss given default.

In this paper, we analyzed several aspects of the economic loss. Particularly, we have focused on an appropriate discount factor and timing of recovery rates to identify the efficient recovery period. Discounting methods are applied to test empirically the determinants of recovery rates. We found that the main drivers are a relative value of collateral, facility name, amount recovered, various costs, loan size and a year of the loan origination.

From a policy perspective, our paper provides evidence that workout LGD is a viable option in credit risk estimation despite various methodological difficulties. In this study, we try to provide a reasonable detail of various issues to be tackled and proposed methodological alternatives how cope with these issues.

Table 3.2 shows the rating of each facility according to their LGD. Facilities i.e. Cash Finance, Demand Finance, Running Finance & Term Finance are assigned rating according to the buckets described by SBP Circular guidelines.

5. ACKNOWLEDGEMENTS

We, first express profound gratitude to Almighty Allah (S.W.T.) who made every one of us and shower boundless leniencies on us. Secondly, we would like to thank our guardians who gave us continuing backing for the duration of our lifetime. After them, we would also like to thank our instructors who gave us gigantic information amid our studies. In spite of the fact that all instructors has actuated colossal jolt in us, we need to name few of them-without their important direction we can never have the capacity to made this technical report. Among those teachers, Chairperson Muhammad Aslam, Dr. Mudassir uddin, Dr. Estisham Hussain, Dr. Farah Yasmin & Syed Monis Jawed has provided us valuable guidance and supervision.

It is our pleasure to acknowledge the guidance and support of our technical report supervisor, Syed Wajid Ali, for his endless patience, encouragement, insight, guidance for giving us an opportunity to work under his supervision and inspiring us to succeed.

Our most profound appreciation goes to the greater part of our loved ones for their ethical bolster, comprehension and offer some assistance which allowing us to stay centered and keeps up a decent personal satisfaction amid this procedure.

REFERENCES

1. Altman, E. (2008). *Default Recovery Rates and LGD in Credit Risk Modeling and Practice. Advances in Credit Risk Modelling and Corporate Bankruptcy Prediction*, (Stewart Jones ed.), Cambridge University Press.
2. Arsova, A., Haralampieva, M. and Tsvetanova, T. (2011). Comparison of regression models for LGD estimation. *Credit Scoring and Credit Control*, XII, 24-26.
3. Yang, B.H. and Tkachenko, M. (2012). Modeling of EAD and LGD: Empirical Approaches and Technical Implementation. Journal of Credit Risk, 8(2).
4. Bellotti, T. and Crook, J. (2012). Loss Given Default Models Incorporating Macroeconomic Variables for Credit Cards. *International Journal of Forecasting*, 28, 171-182.
5. Bonini, S. and Caivano, G. (2014). *Development of a LGD Model Basel2 Compliant: A Case Study. In Mathematical and Statistical Methods for Actuarial Sciences and Finance* (45-48). Springer International Publishing.

6. Siddiqi N. (2006). *Credit Risk Scorecards; Developing and Implementing Intelligent Credit Scoring*. John Wiley & Sons, Inc.
7. Brown, I. (2011). *Regression Model Development for Credit Card Exposure at Default (EAD)* using SAS/STAT® and Enterprise Miner™ 5.3”. University of Southampton, Southampton, UK.
8. Berkel, A.V. and Siddiqui, N. (2012). *Building Loss Given Default Using Weight of Evidence Bins* in SAS® Enterprise Miner™. Paper 141-2012, SAS Institute Inc.
9. Chalupka, R. and Kopecsni, J. (2009). Modeling Bank Loan LGD of Corporate and SME Segments: A Case Study. Charles University in Prague, Faculty of Social Sciences, *Czech Journal of Economics and Finance*, 59(4).
10. Dermine, J. and Neto De Carvalho, C. (2006). Bank Loan Losses-Given-Default: A Case Study. *Journal of Banking and Finance*, 30, 1219-1243.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 393-402

A REMARK ON HAZARD CURVES OF BETA EXPONENTIATED WEIBULL DISTRIBUTION

Sharqa Hashmi[1] and **Ahmed Zogo Memon**[1,2]
[1] CSAS, University of the Punjab, Lahore Pakistan.
Email: sharqa1972@yahoo.com
[2] Institute of Management Sciences, Lahore
Email: azogo.memon@hotmail.com

ABSTRACT

The five–parameter Beta Exponentiated Weibull Distribution (BEWD) is a versatile distribution in modelling lifetime of various industrial products. Its hazard rate assumes decreasing, increasing and bathtub behaviours. In this paper we explore restrictions on the shape parameters that broadly characterize the unique patterns of BEWD hazard curves.

KEY WORDS

Hazard rate (Failure rate), Bathtub shaped hazard rate, Upside down bathtub shaped hazard rate.

1. INTRODUCTION

Weibull distribution has gained popularity due to its importance in modelling lifetime of industrial products, breaking strength of material and wind speed. It has useful applications in survival analysis, industrial engineering, biological research, reliability engineering and in extreme value theory. When modelling monotone (increasing or decreasing) hazard rates, this distribution is often one of the options but it does not describe the bathtub (U-shaped) and the unimodal failure rates that commonly arise in reliability and biological studies. Since the hazard rate has a major role in survival analysis, actuarial science and reliability engineering, hazard rate can assume the bathtub pattern, various other distributions have been evolved. A generalization of Weibull distribution is proposed by using the idea of Eugene et al. (2002) and Jones (2004), named the Beta Weibull (BW) distribution by Gauss M. Cordeiro et al. (2008). They provide different mathematical properties as well as expression for rth moment. Some other distributional properties of the Beta Weibull distribution (BW) was studied by Lee et al. (2007) providing its applications to censored data. The new family is also appropriate for modelling data that indicate non-monotonic hazard rates. But these models do not present much practicability for its use. To cope with this limitation, a new class of distributions, namely Beta Exponentiated Weibull distribution has been proposed in literature by Singla et al. (2012) and Gauss M. Cordeiro et al. (2013), the properties of which were also investigated by Sharqa (2013).

The Beta Exponentiated Weibull (BEW) distribution has been derived, including its various properties, by Cordeiro et al (2013) on using the concept given by Eugene et al.

(2002). According to it, if G denotes the cumulative distribution function of a random variable then the generalized class of distribution function can be defined as:

$$F(x) = \frac{1}{\beta(a,b)} \int_0^{G(x)} w^{a-1}(1-w)^{b-1} dw, \qquad a > 0, b > 0$$

where β(a,b) denotes beta function. The p.d.f. f(x) of BEWD is thus:

$$f(x) = \frac{\alpha\gamma\lambda^{\gamma} e^{-(\lambda x)^{\gamma}} x^{\gamma-1}}{\beta(a,b)} \left[1 - e^{-(\lambda x)^{\gamma}}\right]^{\alpha a-1} \left[1 - \left\{1 - e^{-(\lambda x)^{\gamma}}\right\}^{\alpha}\right]^{b-1}, \qquad x > 0 \qquad (1)$$

Its cdf under the condition of non-negativity of the parameters can be expressed as

$$F(x) = \frac{1}{\beta(a,b)} \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(b)}{j!\,(a+j)\Gamma(b-j)} \left(1 - e^{-(\lambda x)^{\gamma}}\right)^{\alpha(a+j)} \qquad (2)$$

The support of BEWD is based on a five dimensional parameter space. A family of this distribution is created when its parameters assumes various nonnegative values. The BEWD distributions provide different shapes of their hazard rate function depending on what the values of these parameters are. A member of this family is likely to bear an exponential, semi-exponential, skewed or an even a symmetrical pdf. It is vital to have an idea of the effects of parameters on the shapes of the hazard curves because in reliability analysis the guidelines for using a probability model can also come from the shape of a sample hazard curve. It is in this context that this paper is broadly exploring subfamilies of the parameter space that have different probability density shapes and hazard curves. Because of a larger number of parameters, the use of method by Glaser R. E. (1980) does not offer help to understand the geometry of hazard curves, and so we develop a different approach to seek the guidelines for their shapes. Such information is useful for applications of this distribution in understanding the life of industrial products.

2. LEMMAS

To investigate BEWD we state the following lemmas which are not difficult to prove.

2.1 Lemma 1: For v>0 and m>0,

$$\lim_{x \to 0} \left(x^{v-1} \left(1 - e^{-x^{v}}\right)^{m} \right) = \begin{cases} 0 & v(m+1) > 1 \\ 1 & v(m+1) = 1 \\ \infty & v(m+1) < 1 \end{cases} \qquad (3)$$

2.2 Lemma 2

$$\lim_{x \to \infty} \frac{x^{v-1} e^{-x^{v}}}{\left(1 - e^{-x^{v}}\right)} = 0, \qquad \textit{for } v > 0$$

2.3 Lemma 3

For n>0, v>0

$$\lim_{x\to\infty}\frac{x^{v-1}e^{-x^{v}}}{\left(1-\left(1-e^{-x^{v}}\right)^{n}\right)}=\begin{cases}0 & v<1\\ \dfrac{1}{n} & v=1\\ \infty & v>1\end{cases} \tag{4}$$

3. SHAPE OF BEWD

BEWD acquires various shapes depending on values of its parameters. In this section we explore restrictions on these parameters so that a broad distinction in their shapes can be made. We find the following limits of its pdf f(x)

3.1 limit of f(x) at origin

$$\lim_{x\to 0} f(x)=\lim_{x\to 0}\frac{\alpha\ \gamma\ \lambda^{\gamma}e^{-(\lambda x)^{\gamma}}}{\beta\left(a,b\right)}\left[x^{\gamma-1}\left(1-e^{-(\lambda x)^{\gamma}}\right)^{\alpha a-1}\right]\left[1-\left\{1-e^{-(\lambda x)^{\gamma}}\right\}^{\alpha}\right]^{b-1}$$

On rearranging the terms and using Lemma 1, we have

$$\lim_{x\to 0} f(x)=\begin{cases}0 & \gamma(\alpha a)>1\\ \dfrac{\alpha\ \gamma\ \lambda^{\gamma}}{\beta\left(a,b\right)} & \gamma(\alpha a)=1\\ \infty & \gamma(\alpha a)<1\end{cases} \tag{5}$$

Clearly three subfamilies of this distribution emerge with different shape behaviours. These subfamilies are generated by the hyperspaces:

i) $\left\{(\alpha,\lambda,\gamma,a,b);\ \gamma\alpha a>1\right\}$

ii) $\left\{(\alpha,\lambda,\gamma,a,b);\ \gamma\alpha a=1\right\}$

iii) $\left\{(\alpha,\lambda,\gamma,a,b);\ \gamma\alpha a<1\right\}$

The first and third subfamilies consist of the distributions with pdf curves that rise above the x-axis and descend from infinity at (or near) the origin.

The second subfamily merits further investigation as the f(x) of its each member under the restriction $\gamma\alpha a=1$ has a curve that begins at (or near) x=0 from a nonzero value of f(x) and may trend up or below, or be parallel to x axis. The nature of its trend can be determined from

$$\lim_{x->0} f'(x)=\lim_{x->0} f(x)\left(\begin{array}{l}-\lambda^{\gamma}\gamma(x)^{\gamma-1}+\dfrac{(\gamma-1)}{x}+\dfrac{(\alpha a-1)\lambda^{\gamma}\gamma x^{\gamma-1}}{(1-e^{-(\lambda x)^{\gamma}})}e^{-(\lambda x)^{\gamma}}\\ -\dfrac{(b-1)\alpha\lambda^{\gamma}\gamma}{(1-(1-e^{-(\lambda x)^{\gamma}})^{\alpha})}x^{\gamma-1}e^{-(\lambda x)^{\gamma}}(1-e^{-(\lambda x)^{\gamma}})^{\alpha-1}\end{array}\right) \tag{6a}$$

For this subfamily, its lim of (f′(x) / f(x)) at x=0 is:

-b if α=1
-1 α>1 orb=1
-∞ α<1 and b>1 (6b)
∞ α<1 and b<1

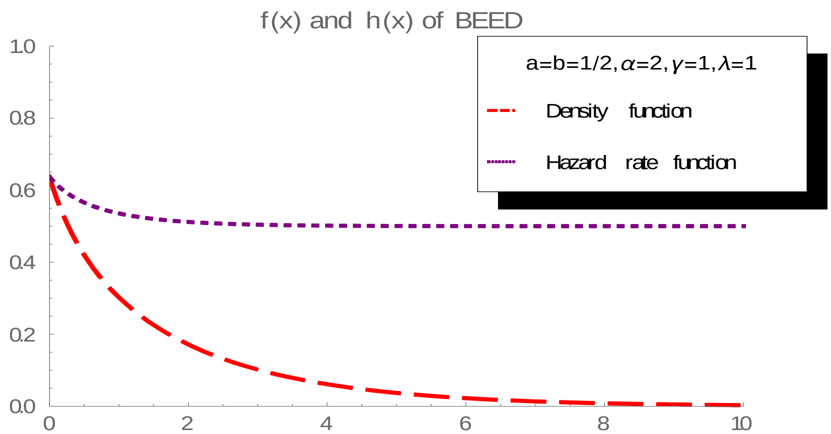
when γ=1.

For γ<1 the limit is -∞, and

for γ>1 the limit is

- ∞ if α<1 and b>1 (6c)
0 α≥1 or b=1; or if α<1 and b<1

For example if a=b=0.5, γ=1, α=2, λ=1 the BEW density curve descends at the origin from the point $\frac{\alpha\,\gamma\,\lambda^{\gamma}}{\beta(a,b)} = 0.63662$ on the vertical axis since $\lim\limits_{x->0} f'(x) = -1 < 0$ by Eq. (6b).



3.2 The Limit of f(x) at Infinity

By Lemma 2 we have,

$$\lim_{x\to\infty} f(x) = 0 \tag{7}$$

for all the three subfamilies of BEWD.

4. HAZARD RATE FUNCTION OF BEWD

Hazard Rate Function of BEW Distribution: Glaser (1980) gives sufficient conditions to characterize hazard rate function of a given distribution as bathtub shaped (BT), increasing failure rate (IFR), upside-down bathtub (UBT), or decreasing failure rate (DFR). Generally his method is helpful in dealing with distributions involving one parameter but BEWD has a complicated pdf with five parameters and as such we develop another approach.

The hazard rate is a non-negative function of x, which may or may not be monotonic. It is a ratio of f(x) to the survival function (1-F(x)).

Using Eqs. (1) and (2), we have

$$h(x)=\frac{\dfrac{\alpha\gamma\lambda^{\gamma}e^{-(\lambda x)^{\gamma}}x^{\gamma-1}}{\beta(a,b)}\left[1-e^{-(\lambda x)^{\gamma}}\right]^{\alpha a-1}\left[1-\left\{1-e^{-(\lambda x)^{\gamma}}\right\}^{\alpha}\right]^{b-1}}{1-\dfrac{\Gamma(b)}{\beta(a,b)}\sum_{j=0}^{\infty}\dfrac{(-1)^{j}}{j!(a+j)\Gamma(b-j)}\left[1-e^{-(\lambda x)^{\gamma}}\right]^{\alpha(a+j)}},\quad x>0 \tag{8}$$

4.1 The Shape of HR Curve

To see the shape or behaviour of hazard rate function, we consider the parameter subspaces of BEWD that produce different shapes both at origin and at infinity.

4.1.1 The Behaviour of HR at the Origin

Since F(x) =0 at origin, the hazard curve has the initial characteristics similar to that of f(x).

$$\lim_{x\to 0} h(x)=\lim_{x\to 0} f(x)=\begin{cases}0 & \gamma\alpha a>1\\ \dfrac{\alpha\ \gamma\ \lambda^{\gamma}}{\beta(a,b)} & \gamma\alpha a=1\\ \infty & \gamma\alpha a<1\end{cases} \tag{9}$$

4.1.2 The Behaviour of HR When x Approaches Infinity

In limit at infinity, h(x) is

$$\lim_{x\to\infty} h(x)=\lim_{x\to\infty}\frac{-f'(x)}{f(x)}=\begin{cases}0 & when\ \gamma<1\\ b\lambda & when\ \gamma=1\\ \infty & when\ \gamma>1\end{cases} \tag{10}$$

(without loss of generality we may take λ=1).

To find the direction of h(x) at x=0 for the second subfamily of BEWD distributions, it can be shown that

$$\begin{aligned}\lim h'(x) &= \lim\,[(h(x))^2+f'(x)]\\ &= (\lim f(x))^2+ø\end{aligned}$$

where ø = lim f(x) times (10a)

-b	if α=1
-1	if α >1, or b=1
-∞	if α<1 and b>1
∞	if α<1 and b<1

when γ=1.

The lim $h'(x)$ is $-\infty$ for γ<1. And for γ>1 the limit is

$-\infty$ if α<1 and b>1 (10b)

0 $\alpha\geq1$ or b=1; or if α<1 and b<1

4.2 Classes of Hazard rate Curves Relating to BEWD

According to the previous section there exist nine combinations of parametric conditions relating to limits of h(x) at 0 and infinity. Using Eqs. (9) and (10), we can expect the following possible hazard rate curves: (see graphs in Appendix 2a).

Parametric Conditions		**Possible shape of hazard curve**
$\gamma\alpha a<1$	$\gamma<1$	Monotonically decreasing
	$\gamma=1$	Partial bathtub
	$\gamma>1$	Bathtub
$\gamma\alpha a=1$	$\gamma<1$	Monotonically decreasing
	$\gamma=1$	Partial bathtub / constant rate when each parameter =1
	$\gamma>1$	Monotonic increasing
$\gamma\alpha a>1$	$\gamma<1$	Upside down bathtub (modal)
	$\gamma=1$	Monotonically increasing / partial upside down bathtub
	$\gamma>1$	Monotonically increasing

The above table highlights the flexibility of BEWD family for applications in the study of industrial products. For example, in the case an exponential distribution since all parameters have the value 1, the height of its hazard curve remains uniform. We provide the following examples (for more details see Appendix 2b).

Examples:

i) In the case of exponential distribution all parameters of BEWD assume the value 1. Here lim h(x) at x=0 is 1 which is the same as lim h(x) as $x\to\infty$. Also, $h'(x)=0$ when $x\to0$, or ∞. So that the hazard rate of this distribution is uniform.

ii) For the reduced Weibull distribution with a= b= α = λ=1 let us consider two restrictions on γ.
For γ< 1 lim h(x) at x=0 is ∞, and lim h(x) =0 as $x\to\infty$.The hazard function displays a decreasing pattern.
For γ> 1, lim h(x) at x=0 is 0, and lim h(x) = ∞ as $x\to\infty$, so that the hazard function is an increasing one.

iii) Suppose we consider the BEWD with γ=2 and a= b= λ =1, α < 1/2, that is, of the third type for which $a\alpha\gamma$ <1. Here, lim h(x) at both ends = ∞, so that a bath tub appears. Alternately, it can be shown that lim $h'(x)$ < 0and >0 as x approaches 0 and infinity respectively.

iv) Suppose γ= λ=1 and a=b=1/2, α =2 and so the distribution is a member of the second subfamily with $a\alpha\gamma$=1. By Eqs. (9), (10) the functions f(x) and h(x) have the same limit $2/\pi$ at x=0; and their limits are respectively 0 and 1/2 as x approaches infinity

5. CONCLUSIONS

The five-parameter BEWD is a generalized distribution in modelling lifetime of various industrial products. Its family displays density curves which are widely heterogeneous in their shapes. In this paper we explore the conditions on BEWD′s shape parameters that provide a class of nearly homogeneous density curves. We investigate the shapes of hazard curves of the subfamilies of BEWD. And present the parametric conditions under which the associated hazard curve of a distribution assumes decreasing, increasing or a bathtub behaviour.

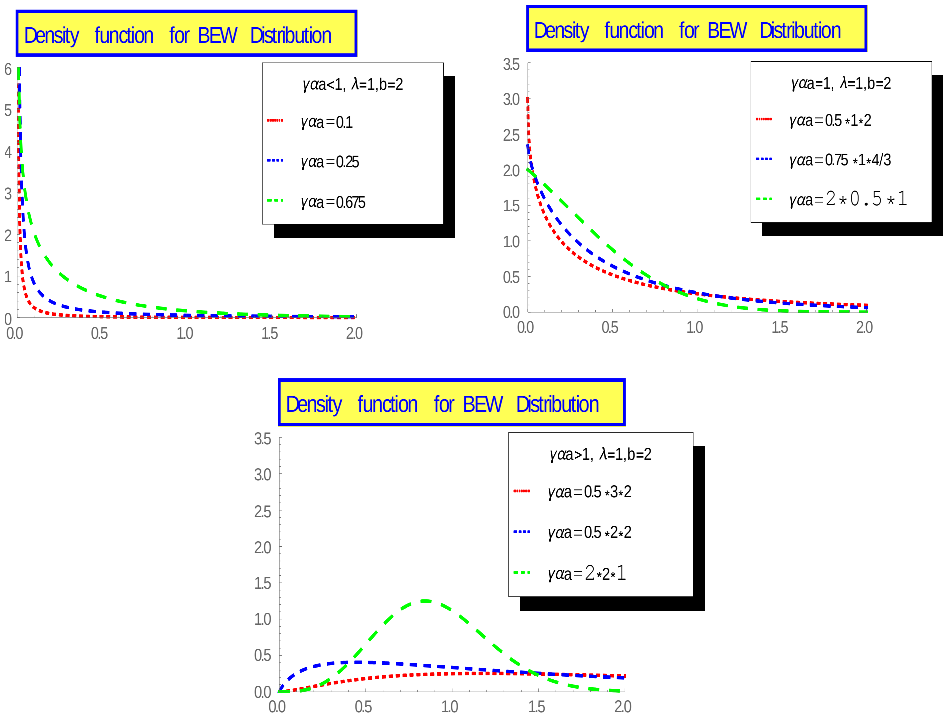
REFERENCES

1. Carrasco, J.M., Ortega, E.M. and Cordeiro, G.M. (2008). A generalized modified Weibull distribution for lifetime modeling. *Computational Statistics & Data Analysis*, 53(2), 450-462.
2. Cordeiro, G.M., Simas, A.B. and Stosic, B.D. (2008). *Explicit expressions for moments of the beta Weibull distribution*. arXiv preprint arXiv:0809.1860.
3. Cordeiro, G.M., Gomes, A.E., da-Silva, C.Q. and Ortega, E.M. (2013). The beta exponentiated Weibull distribution. *Journal of Statistical Computation and Simulation*, 83(1), 114-138.
4. Cordeiro, G.M., Simas, A.B. and Stošic, B.D. (2011). Closed form expressions for moments of the beta Weibull distribution. *Anais da Academia Brasileira de Ciências*, 83(2), 357-373.
5. Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and methods*, 31(4), 497-512.
6. Glaser, R.E. (1980). Bathtub and related failure rate characterizations. *Journal of the American Statistical Association*, 75(371), 667-672.
7. Gupta, P.L. and Gupta, R.C. (2000). The monotonicity of the reliability measures of the beta distribution. *Applied Mathematics Letters*, 13(5), 5-9.
8. Gupta, R. C., & Warren, R. (2001). Determination of change points of non-monotonic failure rates. *Communications in Statistics-Theory and Methods*, 30(8-9), 1903-1920.
9. Lee, C., Famoye, F. and Olumolade, O. (2007). Beta-Weibull distribution: some properties and applications to censored data. *Journal of modern applied statistical methods*, 6(1), 17.
10. McDonald, J.B. and Richards, D.O. (1988). Hazard rates and generalized beta distributions. In *10th Advances in Reliability Technology Symposium* (378-386). Springer Netherlands.
11. Mudholkar, G.S., Srivastava, D.K. and Freimer, M. (1995). The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. *Technometrics*, 37(4), 436-445.
12. Mudholkar, G.S. and Hutson, A.D. (1996). The exponentiated Weibull family: some properties and a flood data application. *Communications in Statistics–Theory and Methods*, 25(12), 3059-3083.
13. Nadarajah, S. and Kotz, S. (2006). The beta exponential distribution. *Reliability Engineering & System Safety*, 91(6), 689-697.
14. Nassar, M.M. and Eissa, F.H. (2003). On the exponentiated Weibull distribution. *Communications in Statistics-Theory and Methods*, 32(7), 1317-1336.

15. Rajarshi, S. and Rajarshi, M.B. (1988). Bathtub distributions: A review. *Communications in Statistics-Theory and Methods*, 17(8), 2597-2621.
16. Silva, G.O., Ortega, E.M. and Cordeiro, G.M. (2010). The beta modified Weibull distribution. *Lifetime Data Analysis*, 16(3), 409-430.
17. Singla, N., Jain, K. and Sharma, S.K. (2012). The beta generalized Weibull distribution: properties and applications. *Reliability Engineering & System Safety*, 102, 5-15.

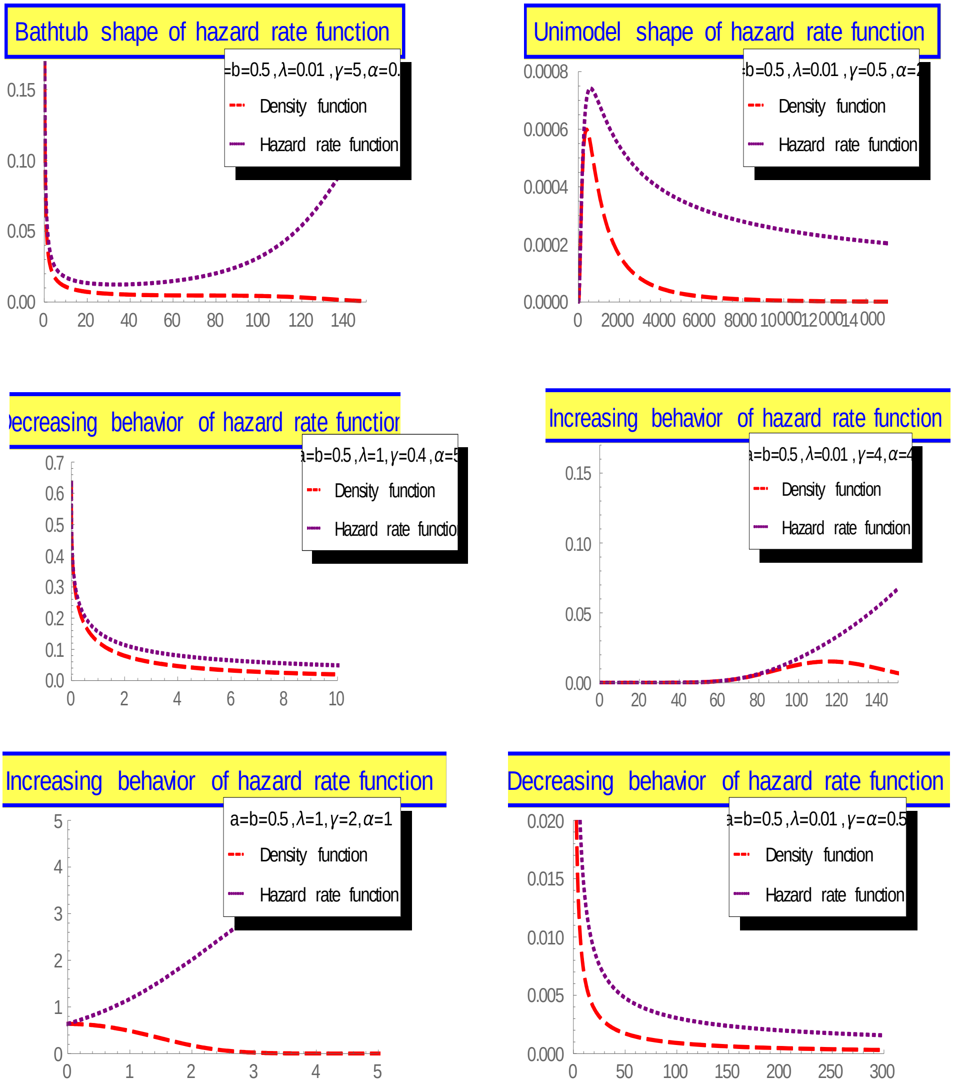
APPENDIX 1:
Shapes of BEW Density Functions (Three Sunfamilies)

Pdf Curves of BEWD under Three Conditions



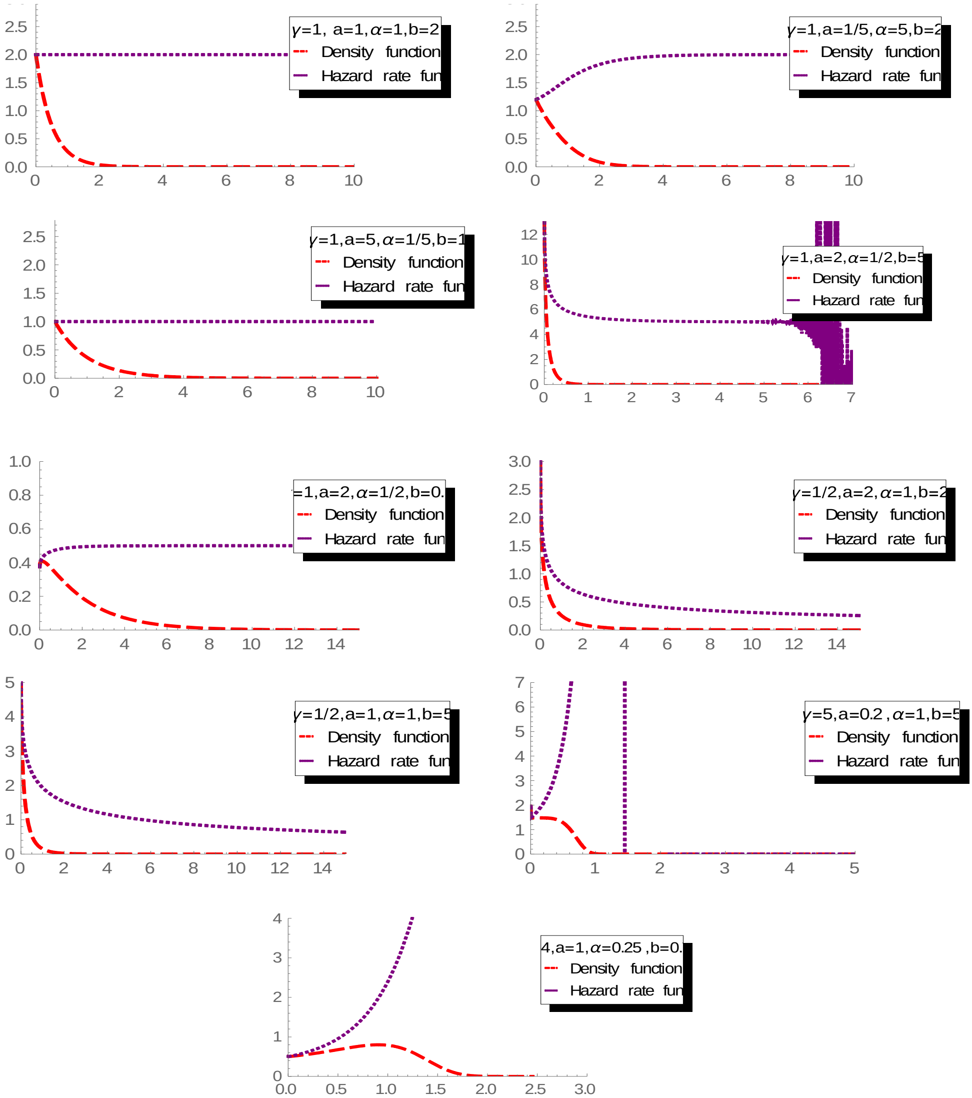
APPENDIX 2A:
Hazard Curves of BEWD Family

Forms of Hazard Rate Function of BEWD



APPENDIX 2B:
Hazard Curves of BEWD Family (Uniform)

Graphs of f(x) and HR under Second Type of BEWE Family



Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 403-412

STUDY THE WAVELET INTERPRETATION OF RESONANCE FOR SEISMICITY PATTERNS AT COASTAL REGIONS OF PAKISTAN

Syed M. Haroon Rashid and **M. Ayub Khan Yousufzai**[§]
Institute of Space & Planetary Astrophysics, University of Karachi, Karachi, Pakistan
[§]Corresponding author Email: ayubzai@yahoo.com

ABSTRACT

Estimation of Power Spectra (Spectral Windows) is used for analytical research work on seismic waves propagation in the coastal region. Their transient periods are detectable through wavelet application and interpretation reflects the impact, their intensity, attenuation process in the coastal environment. Though the seismicity pattern is an indicator of Body and surface waves propagation with emission of heat due to geothermal gradient of the region but shock waves are dominant with vibrational mode of particles.

Earthquake ground motion which is transient motion and energy dissipation due to quality factor, loss factor and specific damping capacity are used for evaluating the nonlinear response of the system. However, there are number of physical phenomena that result in non deterministic manner where future instantaneous values cannot be written in a deterministic sense. Such phenomena have unpredictability of their value at any future time which is also explained by the random vibration. The degree of randomness in the process depends upon the understanding of parametric values of the variable associated with the phenomena and the ability to control them. It has been said that phenomena who's out come at a future instant can be predicted as they are classified as non deterministic factorial Brownian motion.

The response of a system to a random excitation is also random phenomena. Because of the complexity involved, the description of random functions in terms of time does not appear as a particularly meaningful approach, and a new approach of analysis must be adopted. Random vibrations are met rather frequently in nature and may be characterized as vibratory processes in which the vibration particles undergo irregular motion cycles that never repeat themselves exactly.

To obtain a complete description of the vibration, an infinitely long time record is theoretically necessary. This is overcome by using these methods which based on statistical mechanics and wavelet analysis involves concepts such as amplitude probability distributions and probability densities and continuous vibration frequency spectra in term of mean square spectral densities.

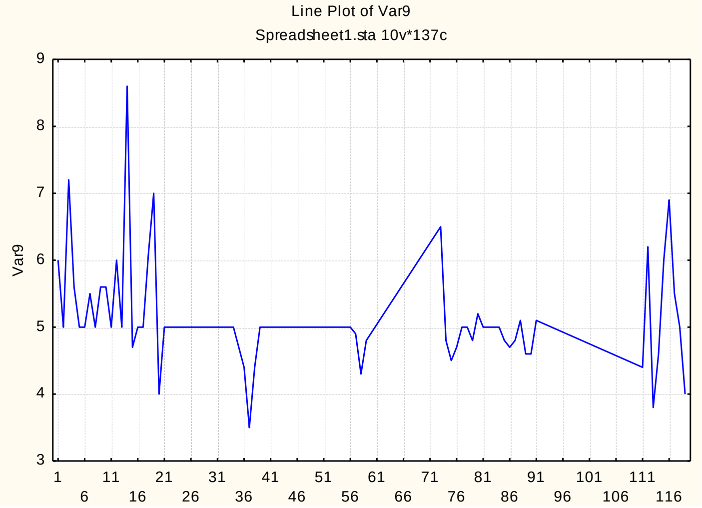
1. INTRODUCTION

Wavelets application for interpreting Seismicity patterns of resonance at the coastal regions of Pakistan is a statistical based analysis which is synchronized with emission of seismic waves propagation on the surface of earth. Geological tectonic setting of coastal regions is consisting a very complex lithological features which show the fluctuation of patterns not only due to the saturation of water but also variability of temperature because of conduction, convection and radiation processes.

In order to solve the problem on temperature fluctuations in the seismicity pattern either naturally production of shock waves or conducted experimental activities. Their exposure represent analytical view only the maximum responses of a number of different structures. However, the response of structures is of great importance in earthquake engineering. The response spectrum has proven to be a useful tool for characterization of strong ground motion. To conduct wavelet analysis for Energy due to strain release prove attenuation of travelling seismic waves underneath the surface and surface of the ground as well. There are wide selection of wavelet basis functions, including several boundaries correction methods like wavelet packet transform and entropy-based wavelet packet.

Wavelet Spectrum provides some results about a wavelet decomposition of a stationary random process which is parallel to a time localized Fourier spectral representation. The wavelet based counterparts of spectral tools are most useful in applications to non stationary time series. Wavelet methods are used in the theory of random processes. Applications of wavelets to stationary time series and non stationary processes are used for modeling. An analysis of time series can be performed in the time domain as well as in the frequency domain. Wavelets also generate novel random processes and functions, suitable for multi scale modeling.

Applications of wavelet in smoothing of periodogram are a task of different nature related to the nonparametric procedures. Wavelets provides additional insight in the analysis of time series via scale analysis. In multi scale analysis often Fourier- based tools for exploring time series have their counter parts wavelet spectra, wavelet periodogram and scalogram for determining the entire hidden periodicities. Wavelet spectrum measures the local power in the variance-covariance decomposition of the process at a certain scale and a time location.



To estimate the evolutionary wavelet spectrum by means of a wavelet periodogram or scalogram; the squared coefficients from a discrete or stationary wavelet transformation is conducted by wavelet analysis of seismic signals which execute polynomial regression.

The Fourier amplitude spectrum and the closely related power spectral density, combined with the phase spectrum for detail analysis.

2. GENERATION RANDOM DENSITIES

Describe a ground motion and response spectrum provides additional information on its potential effects on structures. For modeling, it is of interest to generate random densities that satisfy some constraints. Most typical are requirements on the smoothness, symmetry, unimodality, and skewness. And Transformation of random densities to constrained ones. Wavelets in statistics are used mostly in regression and density estimation problems, virtually any procedure involving orthogonal series decompositions can be waveletized for calculation efficiency. Applications of wavelets in the shape analysis are bridging the areas of statistical modeling, statistical theory of shapes, and image processing. Implementing wavelet and filter bank design for analysis are crucial approach to recognize nonlinearity with vanishing moment.

In addition to the design tools, the toolbox provides code for wavelet applications for both univariate and bivariate denoising, and code for processing of SAR images.

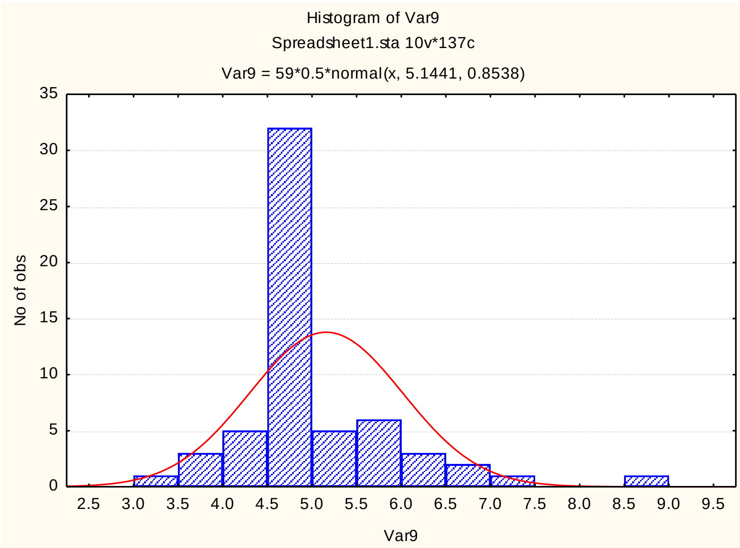


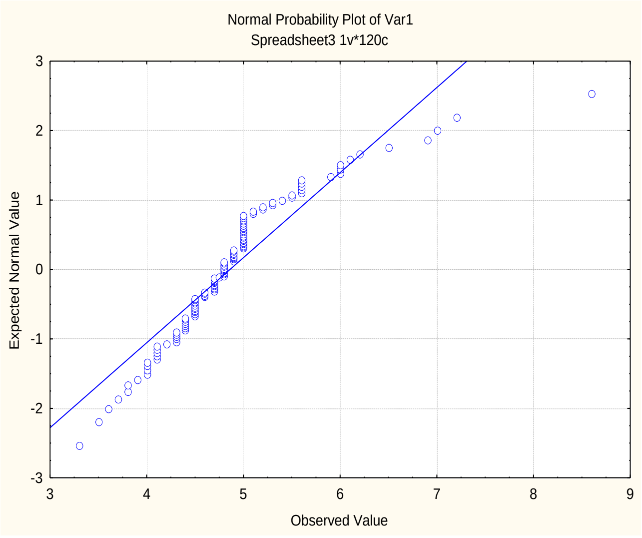
Plate tectonic movement and seismic energies relations, Seismic waves and magnitudes, acceleration and intensities, Earth quakes – Main, Fore and after shocks. Seismic waves attenuations, structures and Source of shock waves develop different scales of magnitudes for earthquake risk prone regions.

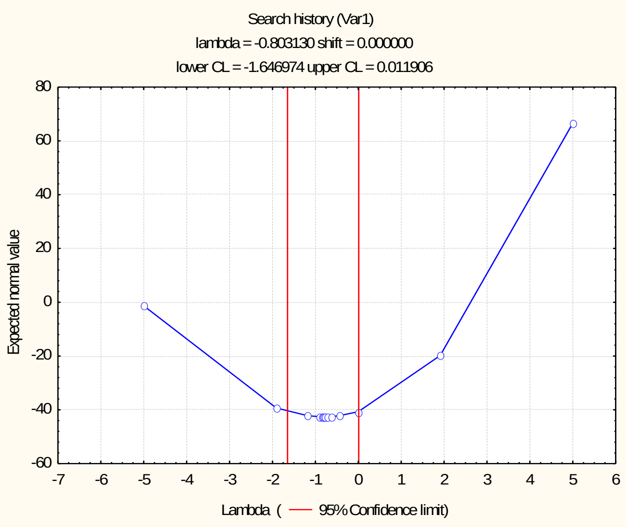
3. VIBRATIONAL PROCESS

Instability of faults and plate boundaries of Arabian sea and Sind- Baluchistan coast, are interprated by expolatory data analysis of vibrational duration where Seismic waves induced vibrational process. Mechanism of estimating energies induced vibrational process, Seismic radiation from focal depth of coastal regions show correlation of acceleration with epicentral distance. Aspects of shock waves Dynamics and seismic waves vibrational increment with reference to subduction zones.

Dynamic of plate tectonics at coastal region and contribution of tectonics in strain with energies release develop by contribution of fault growth increment during seismic wave's propagation.

Spectral Analysis and wavelet characteristics of fluctuating dynamics of seismic waves, Periodogram Analysis and Estimation of Sequential Periodicities, Estimation of Power Spectra (Spectral Windows) reflect the significant behavior of the regions.



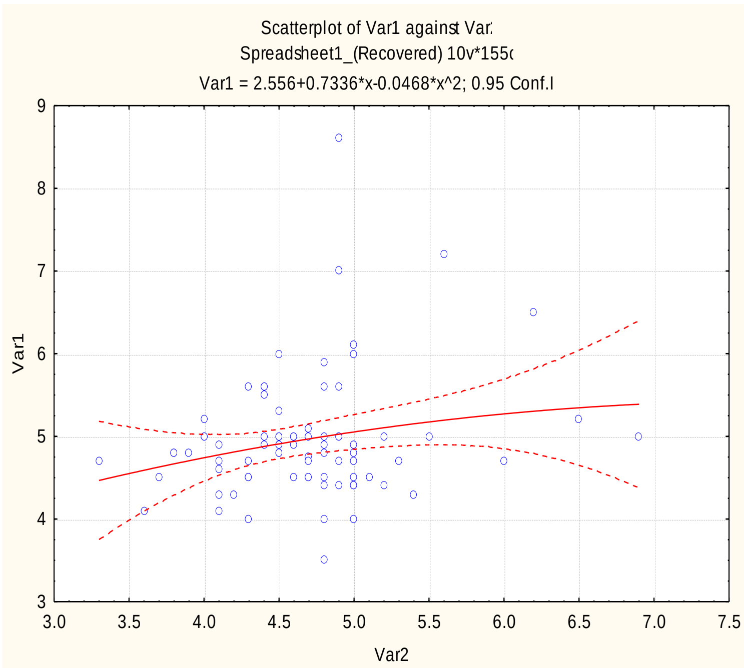


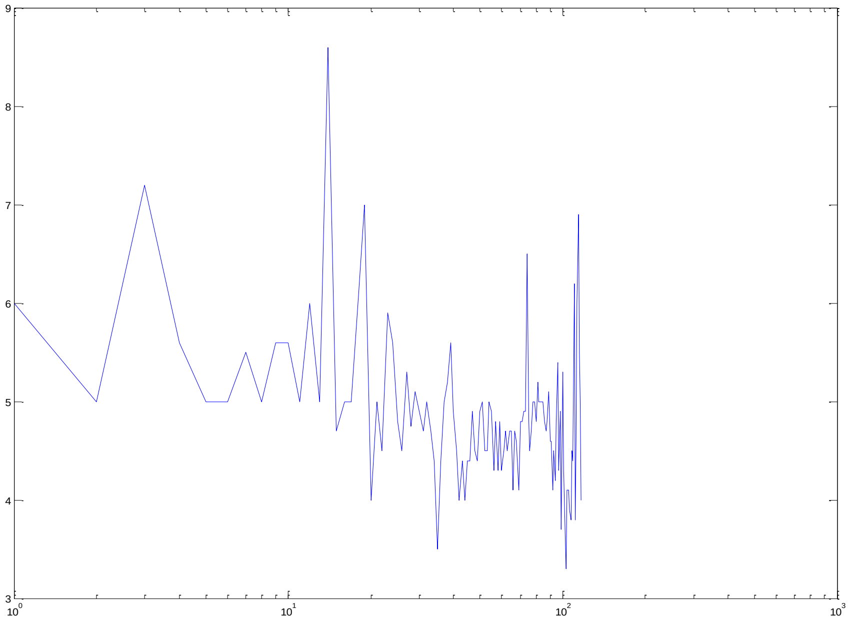
The properties of the material elements include their elasticity modulus, strengths, Poisson ratios, gravities, penetration abilities and temperature features.

For all these parameters, the system takes the average values of the elements. Ultimate compression strain after failure, when the compression strain of the element surpasses this value, the element stiffness will increase and become a contact element. This element will be activated again to act as a compressed material after the phase change it plays the role of delivering stresses to its neighboring elements。

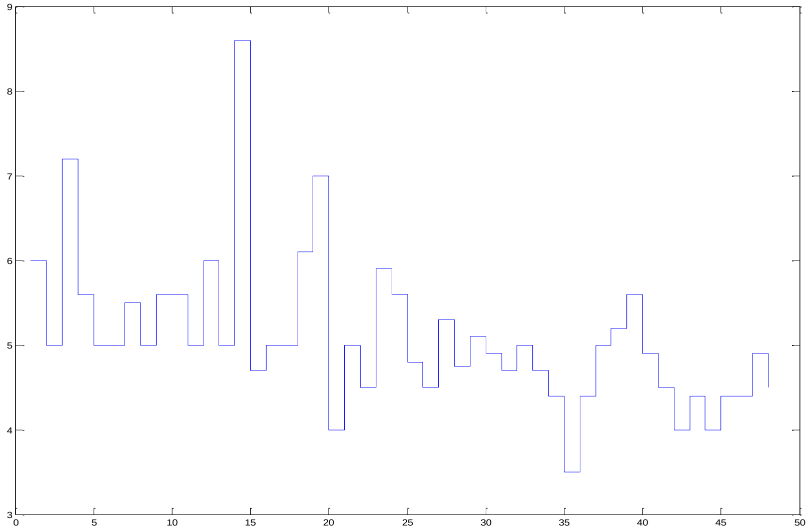
The Damage Coefficient is the increased multiple of the seepage rate once the basic element has been damaged; the Separation Coefficient is the increased multiple of the seepage rate when the element is completely destroyed; and the Coupling Coefficient represents the stress effect on the material seepage.

In the graphics type, users will input data for the graphics type and displaying sizes to generate after simulation.

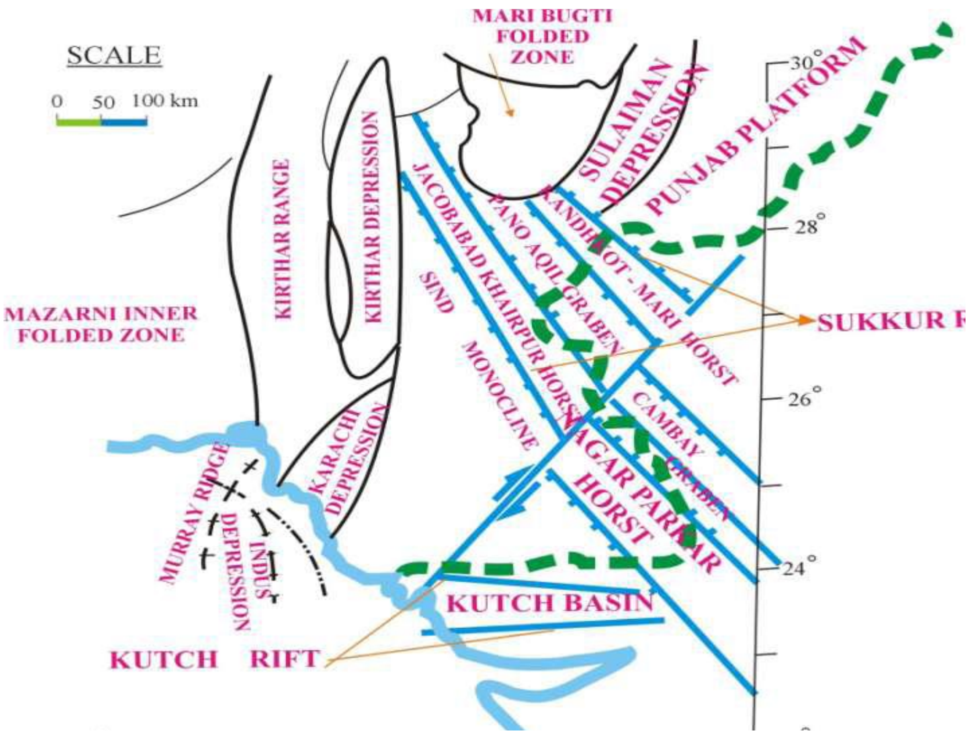




In the cited above graphic presentation shows the frequencies in term of logarithmic behavior of resonance in different time periods.



Application of Haar Wavelet for interpreting non-linear periodic resonance.



Regional map shows different tectonical setting for resonance fluctuation.

Simulation of the rock failure process, such as the effects of stress redistribution due to initial failure during shocks; Simulation of the ways that the acoustic emission arises during the development of the rock failure; and the relations of emission frequency and rock burst degree during rock failure.

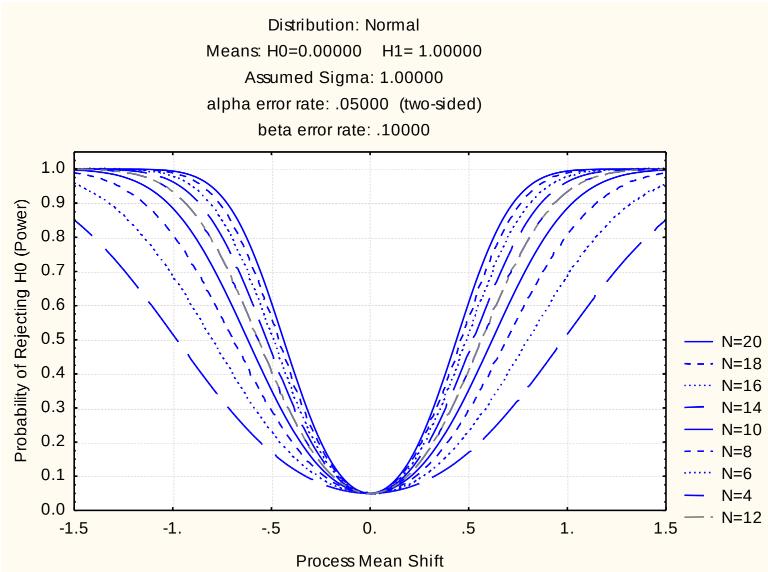
By using some specialized graphical tools, one can consider the micro-defects of the simulated materials and also the macro-fractures, e.g. the joints and other visible fractures. The Weibull distribution, the normal distribution, the uniform distribution, and a number of statistical methods have been incorporated into the system.

4. ELEMENTAL MECHANICAL STRENGTHS

In order to allow wavelet analysis of the unequal distributions of the elemental mechanical strengths and elasticity moduli of materials, enabling further research on the non-linearity of rock deformation.

Ultimate compression strain. After failure, when the compression strain of the element surpasses this value, the element stiffness will increase and become a contact element. This element will be activated again to act as a compressed material after the phase change it plays the role of delivering stresses to its neighboring elements.

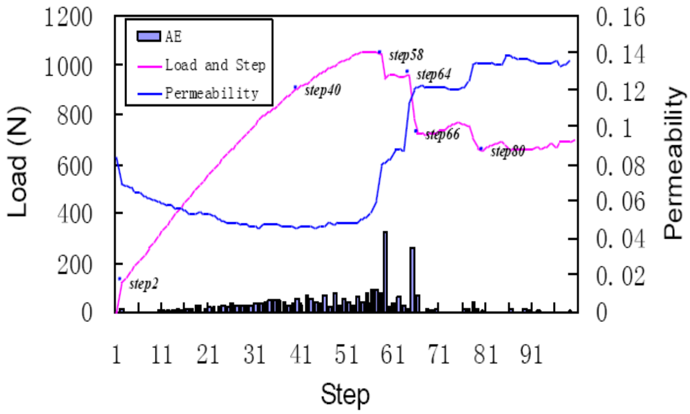
Maximum tensile strain and Ultimate tensile strain. After failure, when the tensile strain of the element surpasses this value, it becomes an air element and is called a null element: it is a failed element that cannot be recovered.



Here alpha error rate and beta error rate control the phase change points with assumed sigma.

5. MATERIAL TEMPERATURE

- Choosing “Temperature Known” will set the material temperature to a certain value. “Initial Temperature” is the initial temperature set at the calculation of Instant heat conduction. By choosing “Increasing Temperature” or “Decreasing Temperature”, the program can conduct operations on temperature changes of the “temperature-known” model.
- The accuracy control is for users to set the computing accuracy; e.g. if the accuracy setting is 0.001, it means that the system will quit running the current step when number of the damaged elements reaches 1/1000 of the total number of elements.
- The advanced parameter settings are mainly for adjustment of strengths, elasticity moduli, and gravities of the “null” and elements after being damaged.
- When the extensive deformation of an element has reached the amount defined by the separation coefficient, this element is considered as being completely separated.



Numerically obtained stress-strain, acoustic emission (AE) and permeability evolution.

- The diagram is used for the interpretation of the elastic modulus, the acoustic emission, the shear stress, the maximum and minimum stresses, permeability, pore pressure distribution, and the distribution of the seepage coefficient.

6. CONCLUSION AND COMMENTS

Coastal regions have a dominant complex features because of strata variability during period of deposition under oceanic environment. Not only the stratified formations but also the deposition of unconsolidated material that is the mixture of fossilized shells, organic components and crystallized minerals, all together play an important role during propagation of seismic waves. Hydrodynamical features provide a more complicated ground for travelling waves than a dry and unsaturated water containing strata. Exposure of coastal surface under oceanic environment gives a fluctuating temperature pattern, which develop a confined condition for energy release and dissipation.

7. ACKNOWLEDGEMENT

Authors would like to express their appreciation for organizing 14th International Conference on Statistical Sciences 14-16th March 2016 at Jinnah Sindh Medical University Karachi, such a grand conference and to provide opportunity to present this piece of information before the august gathering.

REFERENCES

1. Powell, C.M. (1979). A speculative tectonic history of Pakistan and surroundings; some constraints from the Indian ocean: Geodynamics of Pakistan, *Geol. Survey*, (Ouetta).
2. Quittmeyer, R.C., Farah, A.B.U.L. and Jacob, K.H. (1979). The seismicity of Pakistan and its relation to surface faults. *Geodynamics of Pakistan*, 271-284.
3. West, W.D. (1935). Preliminary geological report on the Baluchistan (Quetta) earthquake of May 31st, 1935, *Rec. Geol. Surv. India*, 69, Part 2, 203-240.

4. Heron, A.M. (1911). The Baluchistan earthquake of the 21st October 1909. Rec. Geol. Surv. India, 41, part 1, 22-35.
5. West, W.D. (1934). The Baluchistan (Quetta) earthquake of August 25th and 27th, 1931, *Mem. Geol. Surv. India*, 67, Part 1, 1-82
6. Fitz-Gibbon, C.T. and Morris, L.L. (1997). *How to Design a Program Evaluation*. Beverly Hills, CA: Sage Publication.
7. Esrey, Steven, Jean Gough, et al. (1988). *Ecological Sanitation*. Stockholm, Sweden: Swedish International Development Authority (SIDA).
8. Hansen, R.D. and Soong, T.T. (2001). *Seismic Design with Supplemental Energy Dissipation Devices*, Earth Quake Engineering research Institute, Berkeley, CA.
9. Soong, T.T. and Spencer, B.F. (2000). Active, semi-active and hybrid control of structures. Paper *World Conference on Earthquake Engineering*, Auckland, Elsevier, Amsterdam.
10. PAHO (1981). *A Guide to Emergency Health Management After Natural Disaster*, Scientific Publication No. 407, Pan American Health Organization, Regional Office of the World Health Organization, 525 Twenty-third Street, NW, Washington, DC 20037, USA.
11. PAHO (1982). *Epidemiological Surveillance After Natural Disaster*, *Scientific* Publication No. 420, Pan American Health Organization, Regional Office of the World Health Organization, 525 Twenty-third Street, NW, Washington, DC 20037, USA.
12. Yamahara, H. (1969). *Ground motion during earthquake and incidence loss in seismic wave*. Nippon Kenchiku Gakkai Ronbun Hokoku-shu, No 165.
13. Kawashima, K., Aizawa, K. and Takahashi, K. (1983). *Estimation of peak earthquake motion and seismic response acceleration spectrum Part 5. Development of empirical formula with various damping ratio*. Domoku Kenkyisho Shiryo, No 2001.
14. Kinoshita Shigeo et al. (1986). Estimation of average amplification properties of short period S wave in deposited layer. *Jishin*, 39, 67-80.
15. Goodman, R.E. (translated by K. Akai, T. Kawamoto, Y. Onishi) (1978). Geotechnical Engineering of Discontinuous Rocks. *Morikita Shuppan*, 250-269.
16. Toki, K. and Miura, F. (1985). Simulation of a fault rupture mechanism by a two-dimensional finite element method. *J. Phys. Earth*, 33, 485-511.
17. Brani Vidakovic (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc.
18. Coburn, A. and Spence, R. (2002). *Earthquake Protection*, 2nd edition, John Wiley & Sons, Inc.
19. Katayama, T. et al. (2000). *Dynamic Analysis and Earthquake Resistant design*, Vol. Methods of dynamic analysis, Oxford & IBH Publishing Co.
20. Borradaile, G. (2003). *Statistics of Earth Science Data, their distribution in time, space and orientation*, Springer.
21. Lliboutry, L. (2000). *Quantative Geophysics and Geology*, Springer, Chichester, UK.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 413-418

INVESTIGATING THE INFLUENCE OF COSMIC RAYS ON THE BIOSPHERE FOR PAKISTAN AIR SPACE

Faisal Nawaz[1] and **M. Ayub Khan Yousuf Zai**[2]
[1] Mathematical Science Research Centre, Federal Urdu University of Arts, Sciences and Technology, Karachi, Pakistan
[2] Solar—Terrestrial and Atmospheric Research Wing, Department of Applied Physics, University of Karachi, Pakistan.

ABSTRACT

In the present study monthly data of Cosmic ray intensity obtained by spatial interpolation method (krigging method) and Cloud cover for different regions at different time scale from the period 1984-2006 have been studied to investigate the effects of cosmic ray intensity on the Cloud cover. This work stems from observed correlation between cosmic ray intensity and the cloud amount for the period under study. It has been found from our investigation that the cosmic ray and cloud cover have positive correlation but for some regions this may be the negative. This sort of study confirms about our visualization relating the space and the atmospheric variables.

KEYWORDS

Cosmic rays intensity, Cloud Cover, Climate Change, linear transfer method.

1. INTRODUCTION

In fact, all the radiation that astronomers collect to study the skies and their effects on other objects are example of electromagnetic radiation. These radiations are called stellar radiation or stellar emission because they are emitted by stars (A star is a massive, luminous balls of plasma that is held together by the gravity). Plasma is an ionized gas (when we talk about ionized gas it should keep it mind that it is that which is short of one or more than one electron and become electrically charged.). Stellar radiation of all kinds travel freely through a vacuum at the speed close to the speed of light which is the universal constant which has the values of 1,86,000 miles a second or 300,000 km per seconds .Sources have been suggested ranging all the w ay from the terrestrial to the cosmological. Here we shall investigate what constraints may be placed on the nature of stellar emission. High energy cosmic rays are known as "Galactic cosmic rays" from interstellar space. The discovery of cosmic rays were done by in 1912 by "Hess" when he flew in a balloon to altitude of about 5 km and he discovered that instead of decreasing in value of ionization of air increases with altitude. He gave the explanation that a radiation having high penetrating power enters the atmosphere from above and he called it the cosmic rays and also awarded the noble prize in 1936.

There are two types of cosmic rays the "**Primary**" and the "**Secondary**". The former are those which come from **[1-5]** the outer space while the later are the remnants of the Primary when they break up in the atmosphere All clouds are formed in the Troposphere.

The geographical and climatic condition of Pakistan is as. Latitude is between 24° N to 37°N and longitudes 60°E to 75°E Latitude is between 24° N to 37°N Longitudes 60°E to 75°E covering total area of 796,096 km^2. Pakistani has sharing its countries as follows. India in the East, India in the East, China to North, Iran to West Coast of Arabian Sea to its South. Afghanistan in North West Pakistan has four seasons which makes the climate of Pakistan. They are summer, winter, out am and the spring and some precious things which we call gift of the season. Spring is the season normally comes in April and end in May. The weather is normal and fruits and birds grow. After this the summer season starts which normally begins in June and prolong 4 to 5 months which ends in September. Sun has its maximum range in this season.

An important among other is the rigidity of cosmic rays which is used to measures its ability for bending in a magnetic field. Cosmic rays also travel close to the speed of light and having energy > (GeV). Cosmic rays are highly influenced by the sun's radiation or Sun's conditions.

However, we have to see still the correlation between the space and atmospheric variables as the uncertainty between the variables is not sure **[6, 7]**.

2. METHODOLOGY FOR ANALYSIS

In this communication data have been utilized for the period from January 1984 to December 2006m. This set of data for cosmic ray intensity for Pakistan Air-Space is calculated using Spatial interpolation Krigging approach This method involves the inverse distance formula which is known spatial interpolation form **[8]**. Cosmic ray data were recorded at Beijing super neutron monitor (18-MN-64) situated in the edge of Beijing (40.08N, 116.260E), with the given data set of China and other neighboring countries like Russia, Climax we implemented the technique.

Current study is based on exploratory data analysis (EDA) the purpose of this technique is to compute parameters using variables **[9, 10]**.

3. RESULTS AND DISCUSSIONS

The monthly averages of Cosmic rays intensity on the Cloud cover illustrate the dependence of it for different time and region of Pakistan Air-Space. The reason of it is the close correlation for land regions of the country. However, relatively small changes in the intensity play a significant role on the biosphere merely the production of large quantities of nitrogen oxide which act as a catalyst to destroy the earth's atmosphere.

The model equation of given data points (n =276) taken as:

$$Y = 90.69009 - 0.0454505\, x \tag{1}$$

where "y" is the cloud cover at 0000 UTC (Karachi region) and "x" is the cosmic ray intensity, also

b_0 = 90.69009 (parameter value of intercept), b_1 = - 0.0454505 (parameter value of slope), since it is with negative sign indicates negative correlation of cosmic ray intensity and cloud cover. The R^2 is 30.79% and adjusted R^2 is 30.5%. which shows that only 31%

of cloud cover is explained by the linear regression with the cosmic ray intensity for this region of Pakistan?

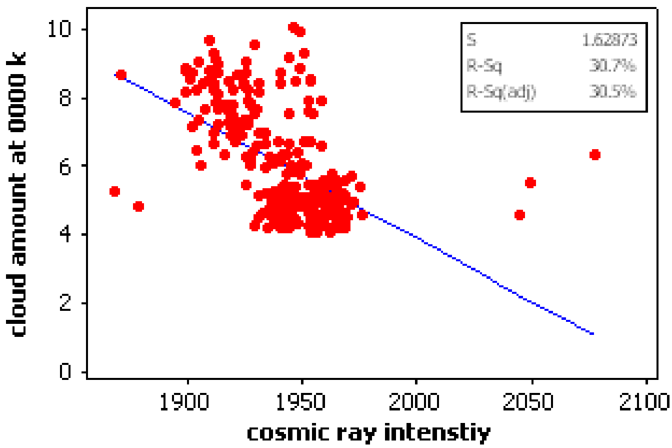


Fig. 1: Monthly mean variation of Cloud cover and the Cosmic ray intensity for Karachi region at 0000 UTC

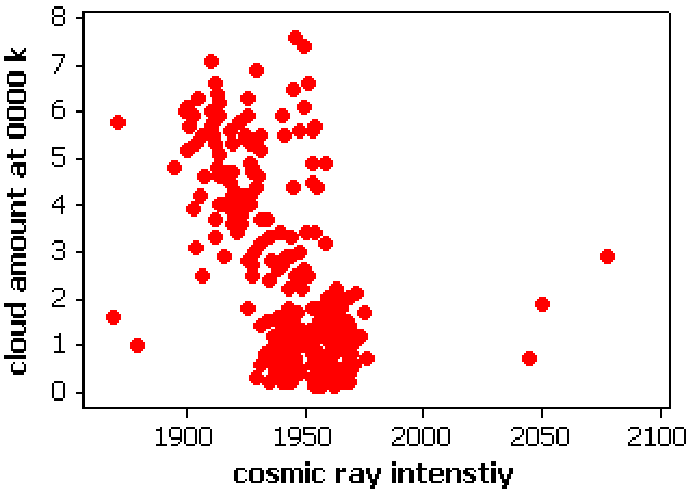


Fig. 2: Scattered Plot between Cloud cover for Karachi region at 0000 UTC and Cosmic ray intensity for Pakistan Air-Space for period 1984-2006

The regression coefficient $b_1 = -0.0454505$ is significantly considerable as the P-value = 0.0000 at 95% level of significance. The actual reliability of the linear model is based on the F-Statistics having 118.1966. The standard error of the above model is 1.633709 indicating the suitable fit for forecasting the cloud cover data. Hence the above model is a meaning full linear model.

Figure 2, shows the general pattern (scattered plot) between the atmospheric and the space parameter. The correlation coefficient between them is week negative but still reliable for forecasting cloud cover from the regression model.

Through this study, we first find the cosmic ray intensity for Pakistan Air-Space by the spatial interpolation technique (Krigging method), then we correlate it with the cloud cover for Karachi region at 0000 UTC.

In this work, we have compared the cosmic ray intensity for Pakistan Air-Space and China Air-Space with the help of time series modeling which is shown by Figure 3, the comparison shows that for the data of two countries the overall trend shows much similarities.

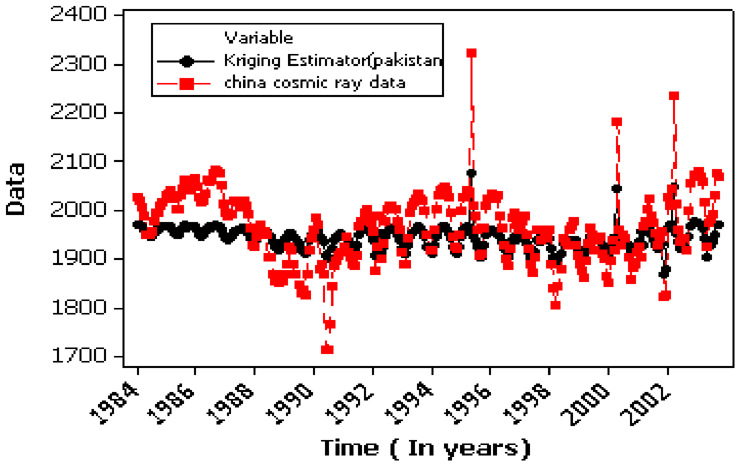


Figure 3: Monthly comparison between Pakistan Cosmic ray intensity data obtained by spatial interpolation method (Krigging method) and the China cosmic ray intensity data obtained by world data centre (WDC) over the period January 1984 to December 2006.

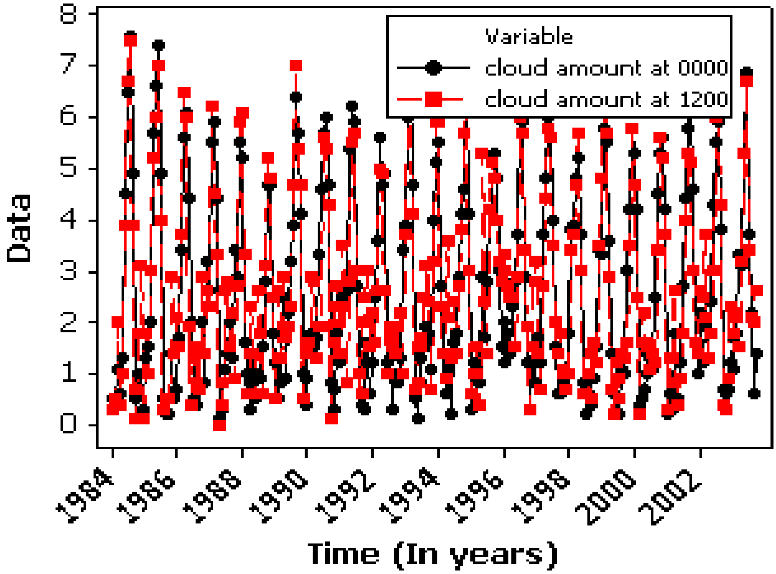


Figure 4: Monthly total cloud cover comparison between Karachi region at 00000 UTC and 1200 UTC. The correlation coefficient between them is 0.892. The correlation coefficient between them is calculated from the period January 1984 to December 2006

4. CONCLUSION

Using Cosmic ray of Pakistan by Krigging method and that for China by World Data Centre and the Cloud cover data for Karachi at 0000 UTC obtained by Pakistan Metrological Department (MET), we have found that model equation (1) give linearly relation between Cosmic ray data and Cloud cover having the "R^2" equals 30.7% which indicates that only 31% explained by linearly regression model. The overall trend of the cosmic ray data for two nations shows much similar trend for the given period under study. Evidences of our quantitative work give the increasing pattern with respect to time.

5. ACKNOWLEDGMENT

I am very thankful to my Supervisor Professor Dr. Muhammad Ayub Khan Yousuf Zai, Department of Applied Physics, University of Karachi, for his constant and untiring efforts in guiding me up to this research work. I further wish to Department of Space and Astrophysics (ISPA) especially the in charge of the department Dr. Jawwad Iqbal as giving me the opportunity to use the Computational lab to perform my research work. Among all, I don't want to forget the Pakistan Metrological department to provide the cloud cover data for Karachi region at 0000 UTC and at 1200 UTC.

REFERENCES

1. Herman, J.R. and Goldberg, R.A. (1978). *Sun, weather, and climate*, NASA. SP-426, Washington, DC, 1-294.
2. Rind, D. and Overpeck, J. (1993). Hypothesized causes of decade-to-century-scale climate variability: climate model results. *Quaternary Science Reviews*, 12(6), 357-374.
3. Phillips, A.C. (2013). *The physics of stars*. John Wiley & Sons.
4. Mukherjee, S. (2008). Cosmic Influence on the Sun-Earth Environment. *Sensors*, 8(12), 7736-7752.
5. Svensmark, H., and Friis-Christensen, E. (1997). Variation of cosmic ray flux and global cloud coverage—a missing link in solar-climate relationships. *Journal of Atmospheric and Solar-Terrestrial Physics*, 59(11), 1225-1232.
6. Sun, B. and Bradley, R.S. (2002). Solar influences on cosmic rays and cloud formation: A Reassessment. *Journal of Geophysical Research*: *Atmospheres*, 107(D14).
7. Draper, N.R., Smith, H. and Pownell, E. (1966). *Applied regression analysis* (Vol. 3). New York: Wiley.
8. Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (2008). *Forecasting methods and applications*. John Wiley & Sons.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 419-426

ON A DECILE-BASED SIA-ESTIMATOR OF THE SHAPE PARAMETER OF THE LOG-LOGISTIC DISTRIBUTION

Kessica Xavier and **Saleha Naghmi Habibullah**
Kinnaird College for Women, Lahore, Pakistan
Email: salehahabibullah@gmail.com

ABSTRACT

Habibullah and Fatima (2015) introduced the terminology Self-Inverse at A (SIA) to represent the distribution of the random variable X for which the distribution of X/A is identical to the distribution of A/X where A is the median of the distribution of X. The SIA property is able to provide estimators of distribution parameters that are *more efficient* than well-known estimators leading to probability models that fit relevant real-life data-sets *with greater accuracy* than that which is achievable through estimators that are generally used. The log-logistic distribution (also known as the Fisk Distribution) is characterized by two parameters, shape and scale, and finds applications in a variety of areas including economics, survival analysis, hydrology and networking. In this paper, we utilize the fact that the log-logistic distribution is a member of the class of SIA distributions to develop an SIA-estimator of the shape parameter based on the first and the ninth deciles. Through Monte Carlo simulation, we show that this estimator is *more efficient* than the non-SIA estimator obtained by utilizing the first decile or the ninth decile. We fit the distribution to a data-set picked up from the literature, the objective being to demonstrate that the newly derived estimator provides a better fit than either of the two that are not based on the SIA property. Given that the log-logistic distribution has applications in important areas such as hydrology, economics and cancer studies, the importance of accurate modeling cannot be over-emphasized.

1. INTRODUCTION

The Log-Logistic distribution is essentially used in different data sets referring to precipitation and stream flow rates. It also help in the distributions related to the field of economics i.e. wealth and income, mortality rate from cancer following treatment or diagnosis, the importance of accurate modeling is self-evident.

Seshadri (1965), Saunders (1974), and Saunders (2007), have discussed continuous random variables Y the reciprocals of which possess exactly the same distributions as do the original random variables. Habibullah et al. (2010) adopted the term Strictly Closed Under Inversion (SCUI) for such random variables. Habibullah and Saunders (2011) introduce the concept of Self-Inverse distributions, and Habibullah and Fatima (2015) adopt the nomenclature SIA for this class of distributions. The median of any SIA distribution lies at A (also called the point of reciprocity) and, for any SIA random variable Y, the distribution of Y/A is identical to the distribution of A/Y.

SIA distributions seem to possess great potential for providing estimators of population parameters that are more efficient than well-known estimators leading to probability models that fit relevant real-life data-sets *more accurately* than that which is achievable through estimators obtained by their well-known counterparts. In this paper, we utilize the SIA property to propose an estimator of the shape parameter of the log-logistic distribution based on the first and ninth deciles along with the median. Through a simulation study, we show that the coefficient of variation of the sampling distribution of the newly derived estimator is *smaller* than that of (a) the sampling distribution of the estimator obtained by utilizing the first decile and the median but not the ninth decile, and (b) the sampling distribution of the estimator obtained by employing the median and the ninth decile but not the first decile. Using each of the three estimators, we fit the log-logistic distribution to a data-set pertaining to survival times, the objective being to demonstrate that the newly derived estimator provides a better fitting model than either of the two that are not based on the SIA property.

2. SIA-DISTRIBUTIONS

Habibullah and Saunders (2011) introduce the term 'Self-Inversion' for any random variable Y with median a for which Y/a is distributed as a/Y, $a \neq 1$, asserting that any such variable can be regarded as being strictly closed under inversion in the *generalized* sense, or "generalized SCUI", so that, by letting $a = 1$, SCUI becomes its special case.

Habibullah and Fatima (2015) introduced the terminology Self-Inverse at A abbreviated as SIA to represent the distribution of the random variable X for which the distribution of X/A is identical to the distribution of A/X where A is the median of the distribution of X.

For the case of sampling from SIA distributions, Habibullah and Fatima (2015) present a new estimator of the distribution mean derived through the utilization of the SIA property, and show that, when employing an SIA distribution for modeling real data, the newly derived estimator provides a *better-fitting* probability model than the one achievable by employing the ordinary sample mean.

3. THE LOG-LOGISTIC DISTRIBUTION

The log-logistic distribution is also known as the Fisk Distribution is a continuous probability distribution for a positive random variable. The PDF of the distribution is as follows:

$$f(\mathrm{x}) = \frac{\frac{\beta}{\mathrm{A}}\left(\frac{x}{\mathrm{A}}\right)^{\beta-1}}{\left(1+\left(\frac{x}{\mathrm{A}}\right)^{\beta}\right)^{2}}, 0 < \mathrm{x} < \infty,\ \mathrm{A} > 0,\ \beta > 0 \tag{3.1}$$

where A is the median as well as the scale parameter, and β is the shape parameter. The cumulative distribution function (CDF) is given as

$$F(x;\alpha,\beta) = \frac{x^{\beta}}{\mathrm{A}^{\beta}+x^{\beta}} \quad \textit{where}\ \ x > 0,\ \mathrm{A},\ \beta > 0 \tag{3.2}$$

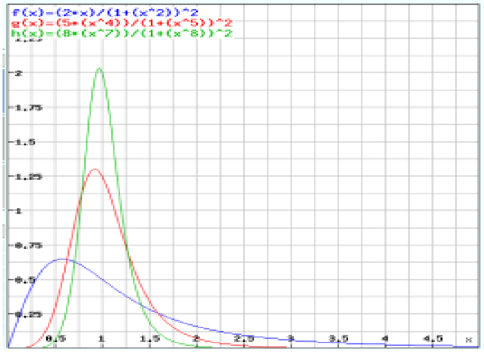


Figure 3.1: PDF of Log-Logistic distribution when A=1 **and** $\beta = 2,5,8$ ***drawn using rechneronline software***

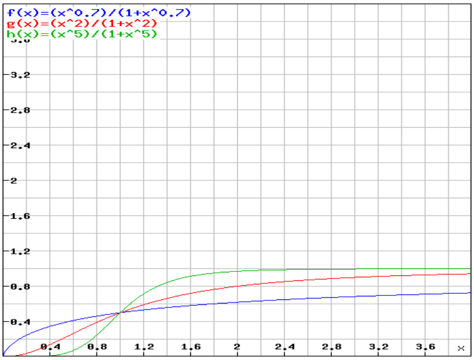


Figure 3.2: CDF of Log-Logistic distribution when A=1 **and** $\beta = 2,5,8$ ***drawn using rechneronline software***

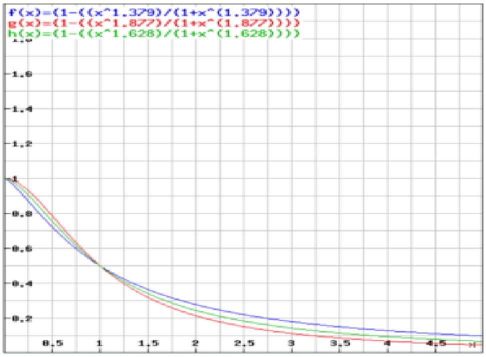


Figure 3.3: Survival function of Log-Logistic distribution when A=1 **and** $\beta = 2,5,8$ **drawn using rechneronline software**

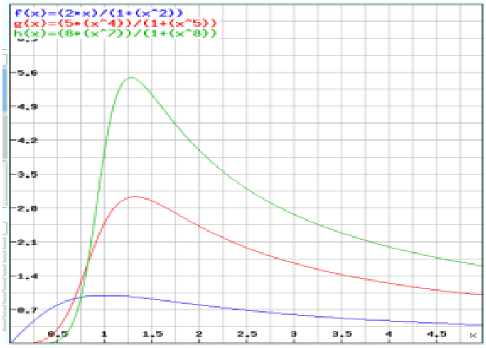


Figure 3.4: Hazard rate of Log-Logistic distribution when $A=1$ **and** $\beta = 2,5,8$ **drawn using rechneronline software**

Figures 3.1 to 3.4 have been reproduced from Xavier and Habibullah (2015). Figures 3.1 and 3.2 present the shapes of the PDF and the CDF, respectively, for the various values of the shape parameter i.e. β. Figures 3.3 and 3.4 present the shapes of the survival function and the hazard function, respectively.

A few of the basic properties of the log-logistic distribution are as follows:

The k^{th} raw moment exist only when k> β, when it is given by

$$E(X^k) = \mathrm{A}^k \mathrm{B}(1-k/\beta, 1+k/\beta) = \mathrm{A}^k \frac{k\pi/\beta}{\sin(k\pi/\beta)}$$

where $\mathrm{B}(.,.)$ implies the beta function. The expressions for mean, variance, skewness and kurtosis can also be derived from this function. The mean is

$$E(\mathrm{X}) = \frac{\mathrm{A}\pi}{\beta \sin\left(\pi/\beta\right)}, \beta > 1$$

and the variance is

$$\mathrm{Var}(\mathrm{X}) = \mathrm{A}^2 \left(\frac{2\pi}{\beta \sin\left(2\pi/\beta\right)} - \frac{\pi^2}{\beta^2 \sin^2\left(\pi/\beta\right)} \right), \beta > 2$$

It is easy to show that when β tends to infinity the mean tends to A, the variance and the skewness tends to zero and the coefficient of kurtosis tends to 6/5.

The inverse cumulative distribution function (quantile function) is given by:

$$F^{-1}(p; \mathrm{A}, \beta) = \mathrm{A} \left(\frac{p}{1-p} \right)^{1/\beta}$$

Here, A is the median, the first or lower decile is $9^{-1/\beta}A$ and the upper decile is $9^{1/\beta}A$.

4. ESTIMATION OF PARAMETERS

It is well-known that one of the simplest methods of estimation is the method of moments. Similar to moments, the sample quantiles can also be equated to the corresponding population quantiles in order to estimate the distribution parameters. Employing this method, we estimate the shape parameter β using (a) the first decile and the median of the sample and (b) the median and the ninth decile of the sample. The respective estimators are given in eq. 4.1 and eq. 4.2.

$$\hat{\beta}_1 = \frac{\ln 9}{\ln\hat{A} - \ln\hat{D}_1} \tag{4.1}$$

$$\hat{\beta}_2 = \frac{\ln 9}{\ln\hat{D}_9 - \ln\hat{A}} \tag{4.2}$$

On the other hand, utilizing the SIA property, we propose the following estimator of β :

$$\hat{\beta}_{SIA} = \frac{\ln 9}{2}\left[\frac{1}{\ln\hat{A} - \ln\hat{D}_1} + \frac{1}{\ln\hat{D}_9 - \ln\hat{A}}\right] \tag{4.3}$$

where $\hat{A}$ is the sample median, $\hat{D}_1$ the 1stdecile of the sample and $\hat{D}_9$ the 9th decile of the sample data.

5. MONTE CARLO SIMULATION

In this section, a simulation study has been conducted to determine that the SIA-estimator is more efficient than the Non-SIA estimator based on the first decile and the median of the sample as well as the Non-SIA estimator based on the median and the ninth decile of the sample.

Table 5.1

Table of Means, Variances and Coefficients of Variation of (i) the Estimator obtained through the First Decile and the Median, (ii) the Estimator obtained through the Median and the Ninth Decile, and (iii) the SIA-Estimator of the Shape Parameter of the Log-Logistic Distribution for β=0.2 & β=0.9

		Estimator based on D_1 and Median	Estimator based on Median and D_9	SIA-Estimator	Estimator based on D_1 and Median	Estimator based on Median and D_9	SIA-Estimator
n ↓	$\beta \rightarrow$	β =0.2	β =0.2	β =0.2	β =0.9	β =0.9	β =0.9
99	Mean	0.2098656	0.2070075	0.2084366	0.942102	0.9358334	0.9389677
	Variance	0.001061184	0.0009657106	0.0004716507	0.02129853	0.01965642	0.009505671
	Coefficient of Variation	0.1552223	0.1501196	0.1041925	0.1549091	0.1498144	0.1038343
75	Mean	0.2134473	0.2095492	0.2114983	0.9529686	0.9486008	0.9507847
	Variance	0.001371922	0.001383356	0.0006385619	0.02768456	0.02788117	0.01292516
	Coefficient of Variation	0.1735298	0.1774929	0.1194799	0.1745984	0.176024	0.1195737
51	Mean	0.2170716	0.2174072	0.2172394	0.9768221	0.9783324	0.9775773
	Variance	0.002106203	0.002294098	0.001003823	0.04265062	0.04645549	0.02032742
	Coefficient of Variation	0.2114205	0.2203089	0.1458445	0.2114205	0.2203089	0.1458445
35	Mean	0.2318945	0.2192878	0.2255912	1.015561	1.003053	1.009307
	Variance	0.003579336	0.003640749	0.001655616	0.06788312	0.07363536	0.0321547
	Coefficient of Variation	0.2579947	0.2751572	0.1803673	0.2565516	0.2705324	0.1776638
19	Mean	0.2586671	0.2239617	0.2413144	1.133721	1.083927	1.108824
	Variance	0.01014877	0.007669703	0.003966218	0.1852681	0.164015	0.07705648
	Coefficient of Variation	0.3894624	0.391035	0.2609787	0.3796594	0.3736301	0.2503468
11	Mean	0.3033415	0.3035425	0.303442	1.365037	1.365941	1.365489
	Variance	0.03397046	0.03037847	0.01466895	0.6879018	0.615164	0.2970462
	Coefficient of Variation	0.6076015	0.5742004	0.3991386	0.6076015	0.5742004	0.3991386

Table 5.2

Table of Means, Variances and Coefficients of Variation of (i) the Estimator obtained through the First Decile and the Median, (ii) the Estimator obtained through the Median and the Ninth Decile, and (iii) the SIA-Estimator of the Shape Parameter of the Log-Logistic Distribution for β=1.4 & β=2.0

		Estimator based on D_1 and Median	Estimator based on Median and D_9	SIA-Estimator	Estimator based on D_1 and Median	Estimator based on Median and D_9	SIA-Estimator
n ↓	$\beta \rightarrow$	β =1.4	β =1.4	β =1.4	β =2.0	β =2.0	β =2.0
99	Mean	1.465019	1.456347	1.460683	2.092505	2.080948	2.086727
	Variance	0.05148477	0.04760691	0.02299407	0.1050131	0.097207	0.04691946
	Coefficient of Variation	0.1548803	0.1498202	0.1038131	0.1548657	0.1498261	0.1038033
75	Mean	1.481135	1.476755	1.478945	2.114945	2.110568	2.112756
	Variance	0.06704314	0.06741457	0.03129011	0.1368916	0.1375257	0.06387609
	Coefficient of Variation	0.1748166	0.1758201	0.1196056	0.1749401	0.1757085	0.1196244
51	Mean	1.519501	1.52185	1.520676	2.170716	2.174072	2.172394
	Variance	0.103204	0.1124108	0.04918734	0.2106203	0.2294098	0.1003823
	Coefficient of Variation	0.2114205	0.2203089	0.1458445	0.2114205	0.2203089	0.1458445
35	Mean	1.574414	1.564555	1.569485	2.245093	2.238622	2.241858
	Variance	0.1640017	0.1777883	0.07755836	0.3346193	0.3623255	0.1580371
	Coefficient of Variation	0.2572202	0.2695012	0.1774423	0.2576566	0.2688863	0.1773257
19	Mean	1.752607	1.705764	1.729185	2.49386	2.452304	2.473082
	Variance	0.4397454	0.4009018	0.1853309	0.8878265	0.8241853	0.3773606
	Coefficient of Variation	0.3783696	0.3711932	0.2489616	0.3778262	0.3702015	0.2483933
11	Mean	2.123391	2.124797	2.124094	3.033415	3.035425	3.03442
	Variance	1.664553	1.488545	0.7187783	3.397046	3.037847	1.466895
	Coefficient of Variation	0.6076015	0.5742004	0.3991386	0.6076015	0.5742004	0.3991386

The means, variances and coefficients of variation of each of the three estimators of β depicted in Tables 5.1 and 5.2 have been obtained by drawing 3000 samples of different sizes from the Log-Logistic distribution with $\alpha = 1$ and four different values of β using R. From the tables, it is easy to see that:

- for each of the three estimators, the means of the sampling distribution $\hat{\beta}$ is approximately the same as the values of β for large values of n implying that $\hat{\beta}$ is asymptotically an unbiased estimator of β ;
- the variances of the SIA-based estimator are smaller than the variances of the estimator based on the first or the ninth deciles;
- the coefficients of variation of the SIA-based estimator are substantially smaller than the coefficients of variation of the estimator based on the first or the ninth deciles.

6. APPLICATION

In this section, we consider the data-set analyzed by Bjerkedal (1960), Gupta et al. (1997), and Kundu et al. (2008) that represents the life-lengths of guinea pigs that have been injected with different doses of tubercle bacilli. The 72 observations are as follows:

12, 15, 22, 24, 24, 32, 32, 33, 34,38, 38, 43, 44, 48, 52, 53, 54, 54,55, 56, 57, 58, 58, 59, 60, 60, 60,60, 61, 62, 63, 65, 65, 67, 68, 70,70, 72, 73, 75, 76, 76, 81, 83, 84,85, 87, 91, 95, 96, 98, 99, 109, 110,121, 127, 129, 131, 143, 146, 146, 175, 175,211, 233, 258, 258, 263, 297, 341, 341, 376.

The histogram of this data-set given in Fig.6.1 has been reproduced from Xavier and Habibullah (2015). The shapes of the PDF (3.1) with median identical to the median of the data and with three different values of the shape parameter are given in Fig. 6.2. The positively skewed shape of the histogram being similar to the shape of the distribution for some values of β , we decide to fit the log-logistic distribution to the data-set.

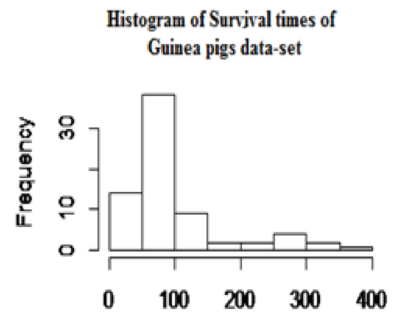


Fig. 6.1: Histogram of the Survival times of guinea pigs data-set Drawn using R language

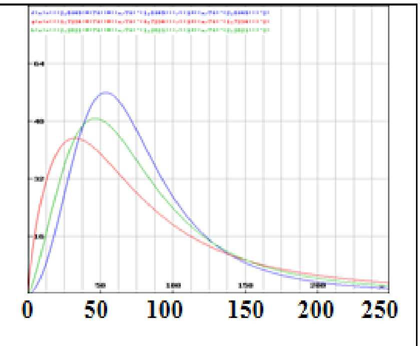


Fig. 6.2: PDF of Log-Logistic distribution when $A = 70$ *and* $\beta = 2.8049, 1.7234, 2.2821$ Drawn using R language

Initially, we fit the Log-Logistic model using $\hat{\beta}_1$ and $\hat{\beta}_2$ given in (4.1) and (4.2) respectively by using the Kolmogorov Smirnov Test. Using $\hat{\beta}_1$, the value of the KS-statistic D comes out to be 0.025 whereas, using $\hat{\beta}_2$, the value turns out to be 0.077. Each of these is less than the tabulated value of $D_{\text{critical\_value}\,(\alpha=0.05)} = 0.160$ indicating that each of the two estimators provides a good fit to the data. Next, we use the SIA-based estimator given in (4.3) which yields D=0.037. Thus we see that the value of the KS-statistic obtained through the SIA-estimator is smaller than the value of D obtained through $\hat{\beta}_2$ but, interestingly, not smaller than the value of D obtained through $\hat{\beta}_1$.

7. CONCLUDING REMARKS

In this paper, an SIA-estimator of the shape parameter of the Log-Logistic distribution has been proposed using the first and ninth deciles and, by conducting a simulation study based on 3000 samples of various sizes drawn from the Log-Logistic distribution for $A = 1$ and $\beta = 0.2,\ 0.9, 1.4, 2.0$ it has been shown that coefficient of variation of the sampling distribution of the SIA-estimator is *smaller* than the coefficient of variation of (i) the Non-SIA estimator based on the first decile and the median and (ii) the Non-SIA estimator based on the median and the ninth decile. Using each of the three estimators, we fitted the Log-Logistic distribution to a data-set picked up from the literature. For this particular data-set, we found that the SIA-estimator is providing a better fit than the Non-SIA estimator based on the median and the ninth decile whereas the Non-SIA estimator based on the first decile and the median seems to be a good competitor of the SIA-estimator.

REFERENCES

1. Habibullah, S.N. and Fatima, S.S. (2015). On a Newly Developed Estimator for More Accurate Modeling with an Application to Civil Engineering, *Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering* (ICASP12) organized by CERRA, Vancouver, BC, Canada.
2. Habibullah, S.N. and Saunders, S.C. (2011). A Role for Self-Inversion, *Proc. of International Conference on Advanced Modeling and Simulation (ICAMS)* published by Department of Mechanical Engineering, College of Electrical and Mechanical Engineering, National University of Science and Technology (NUST), Islamabad, Pakistan, Copyright 2011, ISBN 978-869-8535-11-7.
3. Habibullah, S.N., Memon, A.Z. and Ahmad, M. (2010). *On a Class of Distributions Closed Under Inversion*, Lambert Academic Publishing (LAP), (ISBN 978-3-8383-4868-1).
4. Saunders, S.C. (1974). A family of random variables closed under reciprocla. *J. Amer. Statist. Assoc.*, 69(346), 533-539.
5. Seshadri, V. (1965). On Random Variables which have the Same Dostribution as their Reciprocals. *Can. Math. Bull*., 8(6), 819-824.
6. Xavier, K. and Habibullah, S.N. (2015). On an SIA-Based Estimator of the Shape Parameter of the Log-Logistic Distribution, Proceedings of the *1st International Multi-Disciplinary Conference 2015* (IMDC 2015) held at The University of Lahore, Gujrat Campus, Gujrat, Pakistan.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 427-434

ON A DECILE-BASED SIA-ESTIMATOR OF THE SCALE PARAMETER OF THE LOG-CAUCHY DISTRIBUTION

Aiman Ali and **Saleha Naghmi Habibullah**
Kinnaird College for Women, Lahore, Pakistan
salehahabibullah@gmail.com

ABSTRACT

Self-Inversion at A (SIA) is a property possessed by some of the well-known distributions of non-negative continuous random variables. If Y is a non-negative continuous random variable with median A, self-inversion at A implies that the distribution of Y/A is indistinguishable from the distribution of A/Y. The class of SIA distributions possesses an interesting property that has recently been utilized for developing 'SIA-estimators' of distribution parameters i.e. estimators developed on the basis of the SIA property that are *more efficient* than their well-known counterparts and the adoption of which yields *a better fit* to real-life data. The Log-Cauchy distribution is a member of the class of SIA distributions and finds applications in situations where there is a possibility of encountering significant outliers or extreme results. It has also been proposed as a model for species abundance patterns in addition to having found an application in Bayesian statistics. In this paper, we obtain an SIA-estimator of the scale parameter of the Log-Cauchy distribution based on the first and ninth deciles. By conducting a simulation study based on 10,000 samples of various sizes drawn from the Log-Cauchy distribution with different values of the scale parameter, we show that the coefficient of variation of the newly developed SIA-estimator is *smaller* than the coefficient of variation of the non-SIA estimator based on the first decile and the median or the non-SIA estimator based on the median and the ninth decile. We fit the log-Cauchy distribution to a data-set picked up from the literature in order to demonstrate the *usefulness* of the newly developed SIA-estimator in achieving a better fit.

1. INTRODUCTION

Habibullah and Fatima (2015) have introduced the term "Self-inverse at A (SIA)" which implies that, for any non-negative random variable Y with median A, the probability distribution of Y/A is identical to that of A/Y. The class of SIA distributions contains probability models that are extensively used in engineering and reliability studies such as the Lognormal distribution and the Birnbaum Saunders distribution. Like many others, the Log-Cauchy distribution is also a member of the class of SIA distributions and has been used to model situations where there is a possibility of encountering significant outliers or extreme results. An example of a process where the Log-Cauchy distribution may be an appropriate model is the time-duration between a human becoming a patient of a particular type of cancer and exhibiting symptoms of the cancer which may be very long for some people. It has also been suggested as a model for species abundance patterns in addition to having found an application in Bayesian Statistics.

The SIA property can be used to develop a large number of efficient estimators of distribution parameters and, in this paper; we present an SIA-estimator of the scale parameter of the Log-Cauchy distribution based on the first and ninth deciles of the distribution. By conducting a simulation study based on 10,000 samples of various sizes drawn from the Log-Cauchy distribution with various values of σ, we show that the coefficient of variation of the newly developed SIA-estimator is *smaller* than the coefficient of variation of the non-SIA estimator based on the first decile and the median or the non-SIA estimator based on the median and the ninth decile.

2. LOG-CAUCHY DISTRIBUTION

A Log-Cauchy distribution is a probability distribution of a random variable the logarithm of which is distributed in accordance with a Cauchy distribution. In other words, if X is a random variable with a Cauchy distribution, then $Y = e^X$ has a Log-Cauchy distribution.

The probability density function (PDF) and the cumulative distribution function (CDF) of the Log-Cauchy distribution are, respectively,

$$f(x) = \frac{\sigma}{x\pi\left[(\ln x - \mu)^2 + \sigma^2\right]}, \quad 0 < x < \infty$$

and

$$F(x) = \frac{1}{\pi}\tan^{-1}\left(\frac{\ln x - \mu}{\sigma}\right) + \frac{1}{2}, \quad 0 < x < \infty$$

where μ is the location parameter and $\sigma > 0$ is the scale parameter of the distribution.

The Log-Cauchy distribution is also known as the generalized form of the Beta distribution of the second kind. This is known to be a heavy-tailed distribution.

Figures 3.1 to 3.4 have been reproduced from Ali and Habibullah (2015). Figures 3.1 and 3.2 present the shapes of the PDF and the CDF, respectively, for the various values of the scale parameter σ. Figures 3.3 and 3.4 present the shapes of the survival function and the hazard function, respectively.

3. ESTIMATION OF SCALE PARAMETER

For the Log-Cauchy distribution, median $= \mathrm{A} = e^{\mu}$ and, as such, we have

$$D_1 = e^{\mu} - e^{\tan\frac{2\pi}{5}\sigma} \Rightarrow = \mathrm{A} - e^{\tan\frac{2\pi}{5}\sigma}$$

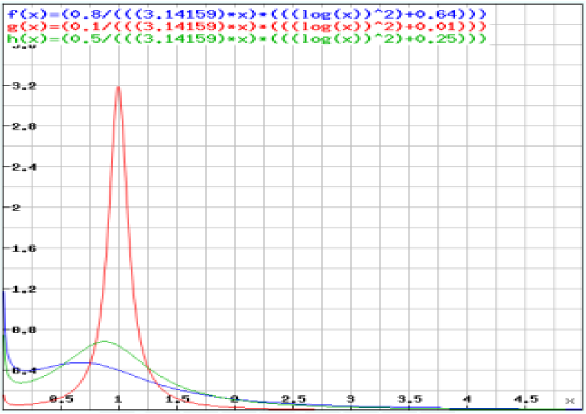


Fig. 3.1: PDF of the Log-Cauchy Distribution

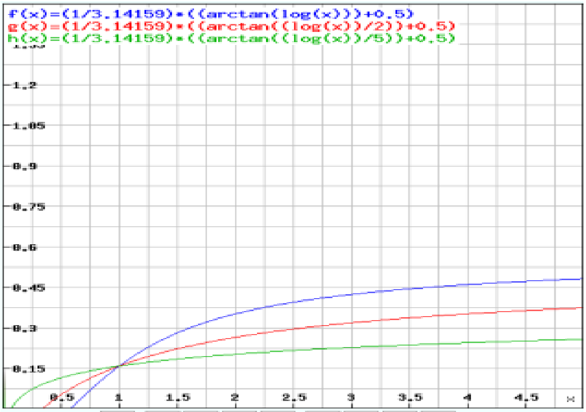


Fig. 3.2: CDF of the Log-Cauchy Distribution

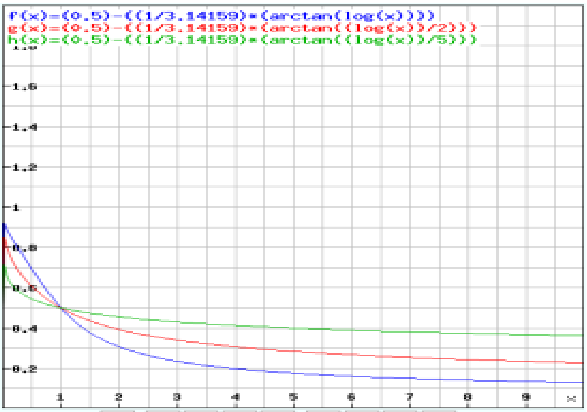


Fig. 3.3: Survival Function of the Log-Cauchy Distribution

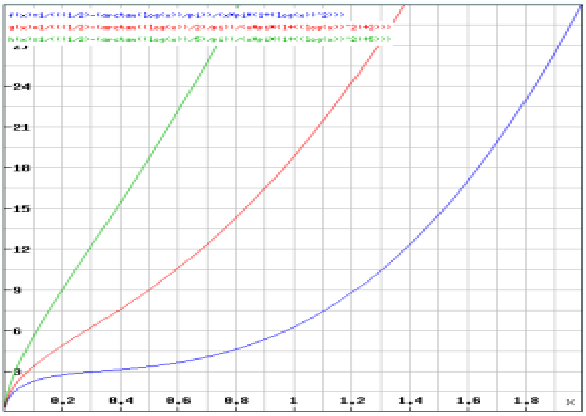


Fig. 3.4: Hazard Rate of the Log-Cauchy Distribution

implying that

$$\ln D_1 = \ln \mathrm{A} - \tan\frac{2\pi}{5}\sigma$$

so that the estimator of σ based on the first decile and the median is given by

$$\hat{\sigma}_1 = \frac{\ln \hat{\mathrm{A}} - \ln \hat{D}_1}{\tan\frac{2\pi}{5}} \tag{3.1}$$

also

$$D_9 = e^{\mu+\tan\frac{2\pi}{5}\sigma} = e^{\mu} + e^{\tan\frac{2\pi}{5}\sigma}$$

so that the estimator of σ based on the median and the ninth decile is given by

$$\hat{\sigma}_2 = \frac{\ln \hat{D}_9 - \ln \hat{\mathrm{A}}}{\tan\frac{2\pi}{5}} \tag{3.2}$$

On the other hand by using the SIA property of the Log-Cauchy distribution, we propose the following estimator of σ:

$$\hat{\sigma}_{\mathrm{SIA}} = \frac{\ln \hat{D}_9 - \ln \hat{D}_1}{2\tan\frac{2\pi}{5}} \tag{3.3}$$

4. SIMULATION STUDY

In this section, we present the results of a simulation study based on 10,000 samples of various sizes drawn from the Log-Cauchy distribution with $\mu = 0$ and various values of σ. Tables 4.1, 4.2 and 4.3 present the means, variances and coefficients of variation of

(i) the Non-SIA estimator of σ based on the first decile and the median, (ii) the Non-SIA estimator based on the median and the ninth decile, and (iii) the SIA-estimator based on the first and ninth deciles for various values of σ and six different sample sizes. From the tables, it is easy to see that, for all combinations of values of n and σ, the coefficient of variation of the SIA-estimator is *smaller* than the coefficient of variation of the estimator based on the first decile and the median as well as the coefficient of variation of the estimator based on the median and the ninth decile.

5. APPLICATION

In this section, we fit the Log-Cauchy distribution to a data-set considered by Bebbington, et al. (2006) on the electrical appliance failure times (1000's of cycles) that were originally presented by Lawless (2003). The data-values are given in Table 5.1. We fit the distribution three times using the median of the data to estimate the location parameter and using (3.1), (3.2) and (3.3) respectively to estimate the scale parameter.

Table 4.1
Table of Means, Variances & Coefficients of Variation of (i) the Estimator based on D_1 and the Median, (ii) the Estimator based on the Median and D_9, and (iii) the SIA-Estimator based on D_1 and D_9 of the Scale Parameter of the Log-Cauchy Distribution for some selected values of σ and for n=20, n=30

		n =20			n =30		
		Estimator based on D_1 and Median	Estimator based on Median and D_9	SIA-Estimator	Estimator based on D_1 and Median	Estimator based on Median and D_9	SIA-Estimator
σ=0.1	Mean	0.3103	0.1551	0.2840	0.2626	0.1259	0.2571
	Var.	0.4510	0.1127	0.1249	0.0430	0.0069	0.0174
	C.V.	**2.1645**	**2.1645**	**1.2447**	**0.7894**	**0.6622**	**0.5123**
σ=0.2	Mean	0.6692	0.3346	0.5891	0.5396	0.2502	0.5200
	Var.	2.1861	0.5465	0.5931	0.2255	0.0273	0.0821
	C.V.	**2.2095**	**2.2095**	**1.3073**	**0.8799**	**0.6600**	**0.5511**
σ=0.5	Mean	1.8749	0.9374	1.5637	1.4323	0.6196	1.3358
	Var.	16.0981	4.0243	4.3029	1.9780	0.1669	0.6497
	C.V.	**2.1400**	**2.1400**	**1.3265**	**0.9819**	**0.6594**	**0.6034**
σ=0.8	Mean	3.1759	1.5879	2.5832	2.3837	0.9866	2.1785
	Var.	43.3515	10.8373	11.5426	5.7693	0.4241	1.8341
	C.V.	**2.0732**	**2.0732**	**1.3152**	**1.0076**	**0.6601**	**0.6217**
σ=1.0	Mean	4.0718	2.0358	3.2766	3.0392	1.2307	2.7503
	Var.	68.9478	17.2360	18.3343	9.4765	0.6609	2.9783
	C.V.	**2.0393**	**2.0393**	**1.3068**	**1.0129**	**0.6606**	**0.6275**

Table 4.2
Table of Means, Variances & Coefficients of Variation of (i) the Estimator based on D_1 and the Median, (ii) the Estimator based on the Median and D_9, and (iii) the SIA-Estimator based on D_1 and D_9 of the Scale Parameter of the Log-Cauchy Distribution for some selected values of σ and for n=40, n=50

		n =40			n =50		
		Estimator based on D_1 and Median	Estimator based on Median and D_9	SIA-Estimator	Estimator based on D_1 and Median	Estimator based on Median and D_9	SIA-Estimator
σ=0.1	Mean	0.2543	0.1246	0.2518	0.2519	0.1243	0.2503
	Var.	0.0213	0.0047	0.0098	0.0153	0.0036	0.0073
	C.V.	**0.5738**	**0.5522**	**0.3940**	**0.4912**	**0.4845**	**0.3419**
σ=0.2	Mean	0.5142	0.2484	0.5055	0.5061	0.2482	0.5012
	Var.	0.0958	0.0187	0.0418	0.0635	0.0144	0.0297
	C.V.	**0.6018**	**0.5511**	**0.4043**	**0.4979**	**0.4836**	**0.3441**
σ=0.5	Mean	1.3270	0.6174	1.2809	1.2844	0.6181	1.2603
	Var.	0.7769	0.1154	0.3034	0.4481	0.0890	0.1976
	C.V.	**0.6642**	**0.5502**	**0.4300**	**0.5212**	**0.4826**	**0.3527**
σ=0.8	Mean	2.1780	0.9844	2.0733	2.0854	0.9844	2.0291
	Var.	2.2963	0.2935	0.8509	1.2626	0.2935	0.5327
	C.V.	**0.6958**	**0.5503**	**0.4449**	**0.5388**	**0.5503**	**0.3597**
σ=1.0	Mean	2.7605	1.2284	2.6087	2.6304	1.2315	2.5466
	Var.	3.8165	0.4574	1.3848	2.0741	0.3529	0.8564
	C.V.	**0.7077**	**0.5505**	**0.4511**	**0.5475**	**0.4824**	**0.3634**

Table 4.3
Table of Means, Variances & Coefficients of Variation of (i) the Estimator based on D_1 and the Median, (ii) the Estimator based on the Median and D_9, and (iii) the SIA-Estimator based on D_1 and D_9 of the Scale Parameter of the Log-Cauchy Distribution for some selected values of σ and for n=70, n=100

		n =70			n =100		
		Estimator based on D_1 and Median	Estimator based on Median and D_9	SIA-Estimator	Estimator based on D_1 and Median	Estimator based on Median and D_9	SIA-Estimator
σ=0.1	Mean	0.2480	0.1235	0.2475	0.2478	0.1236	0.2475
	Var.	0.0092	0.0025	0.0047	0.0063	0.0016	0.0031
	C.V.	**0.3869**	**0.4058**	**0.2774**	**0.3209**	**0.3266**	**0.2245**
σ=0.2	Mean	0.4968	0.2467	0.4951	0.4958	0.2471	0.4950
	Var.	0.0373	0.0100	0.0189	0.0254	0.0065	0.0124
	C.V.	**0.3886**	**0.4054**	**0.2778**	**0.3212**	**0.3264**	**0.2246**
σ=0.5	Mean	1.2491	0.6155	1.2400	1.2419	0.6170	1.2380
	Var.	0.2455	0.0621	0.1208	0.1607	0.0405	0.0776
	C.V.	**0.3966**	**0.4049**	**0.2803**	**0.3228**	**0.3262**	**0.2250**
σ=0.8	Mean	2.0114	0.9833	1.9890	1.9917	0.9864	1.9822
	Var.	0.6644	0.1583	0.3175	0.4192	0.1034	0.2002
	C.V.	**0.4053**	**0.4047**	**0.2833**	**0.3251**	**0.3260**	**0.2257**
σ=1.0	Mean	2.5250	1.2281	2.4906	2.4937	1.2324	2.4792
	Var.	1.0736	0.2469	0.5043	0.6639	0.1613	0.3148
	C.V.	**0.4104**	**0.4046**	**0.2851**	**0.3267**	**0.3259**	**0.2263**

Table 5.1
Electrical Appliance Failure Times (1000's of Cycles)

0.014	0.479	1.275	2.326	3.912
0.034	0.556	1.355	2.337	4.100
0.059	0.574	1.397	2.628	4.106
0.061	0.839	1.477	2.785	4.116
0.069	0.917	1.578	2.811	4.315
0.080	0.969	1.649	2.886	4.510
0.123	0.991	1.702	2.993	4.584
0.142	1.064	1.893	3.122	5.267
0.165	1.088	1.932	3.248	5.299
0.210	1.091	2.001	3.715	5.583

Applying the Kolmogorov Smirnov Test to test for the goodness of fit, we find that, using (3.1), the value of the KS-statistic D is 0.1709, using (3.2), D comes out to be 0.1616 and, using (3.3), D is 0.1205. The critical value of D is 0.1575 indicating that neither the D achieved through (3.1) nor the D obtained through (3.2) lies in the acceptance region. The value of D achieved through (3.3) is smaller than both of the other two values of D and is less than the critical value of D also. As such, it is obvious that, out of the three estimators under consideration, the SIA-estimator provides the best fit to this data-set.

6. CONCLUDING REMARKS

The class of distributions self-inverse at A possesses an interesting property by which it is possible to propose estimators of distribution parameters that are *more efficient* than well-known estimators. Such estimators may be called SIA-estimators. The Log-Cauchy distribution being a member of this class of distributions, in this paper, we have proposed an SIA-estimator of the scale parameter involving both the first and the ninth deciles and have compared it with two other estimators one of which involves the first decile and the median and the other the median and the ninth decile. Monte Carlo simulation based on 10,000 samples of various sizes drawn from the Log-Cauchy distribution shows that the sampling distribution of the SIA-estimator is *narrower* than the sampling distribution of either of the other two estimators. We fit the log-Cauchy distribution to a data-set taken from the literature in order to demonstrate the fact that a more efficient estimator is likely to yield more accurate modeling. Given that the Log-Cauchy distribution finds applications in important areas such as health and biology, the usefulness of efficient estimation of the distribution's scale parameter is obvious.

REFERENCES

1. Ali, A. and Habibullah, S.N. (2015). On a New Estimator of the Scale Parameter of the Log-Cauchy Distribution to Achieve Better Fit, Proceedings of the *1st International Multi-Disciplinary Conference 2015* (IMDC 2015) held at The University of Lahore, Gujrat Campus, Gujrat, Pakistan.
2. Habibullah, S.N. and Fatima, S.S. (2015). On a Newly Developed Estimator for More Accurate Modeling with an Application to Civil Engineering, *Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering* (ICASP12) organized by CERRA (Vancouver, BC, Canada, July 12-15, 2015).
3. Habibullah, S.N. and Saunders, S.C. (2011). A Role for Self-Inversion, *Proc. of International Conference on Advanced Modeling and Simulation (ICAMS)* published by Department of Mechanical Engineering, College of Electrical and Mechanical Engineering, National University of Science and Technology (NUST), Islamabad, Pakistan, Copyright 2011, ISBN 978-869-8535-11-7.
4. Bebbington, M., Lai, C.D. and Zitikis, R. (2006). Useful periods for lifetime distributions with bathtub shaped hazard rate functions. *IEEE Transactions on Reliability*, 55(2), 245-251.
5. Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd Edition, John Wiley and Sons, New York.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 435-442

EFFECTS OF COURSE REPETITION ON STUDENTS' SUBSEQUENT PROGRESS: NEED FOR SUPPLEMENTARY EXAM

Muhammad Hassan, Noor us Saba and **Faqiha Sultan**
FAST-National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: k142126@nu.edu.pk,
k142139@nu.edu.pk,
faqiha.sultan@nu.edu.pk

ABSTRACT

The aim of this study is to explore the impact of course repetition on students' subsequent academic progress at undergraduate level. This issue has attracted considerable attention among the university students. This research refers to the way chosen for clearance of a student's course, if one fails in the first attempt, which can be either by appearing in a supplementary exam or by repeating the complete course, where each method has its own implications and influences the subsequent academic success of the affected students. This study claims that offering a supplementary exam is much better than course repetition. To demonstrate this, students from different universities of Karachi are divided into two categories according to the way chosen for clearance of a failed course and both the groups are asked about their academic growth. From each category, a sample of 100 students is picked out and the survey is conducted through questionnaires. A comparative study between two categories is conducted and the obtained results support the hypothesis.

KEYWORDS

Academic progress; Course repetition; Supplementary Exam

1. INTRODUCTION

Currently, universities are divided into two categories based on the examination policy for a student failing a course. The first one requires a student to register the course again in the upcoming semester or year, fulfill all the norms accordingly and repeat the whole course. Whereas, the second category demands to attempt a supplementary exam only and there is no restriction on the number of times one can attempt a supplementary exam. Every university practices one particular method and the students have to experience the same examination system without the feasibility of a choice.

In defining the examination methods for failure students, there is also a need to understand the basic assessment methods and their impact on students' learning and success. During last few decades, there have been many changes in assessment methods realizing the fact that the examinations play an imperative part in the scholarly career of students and it is an idea that is rehashed throughout the academic year and one that decides future roles and placement of students. Miller and Parlett (1974) presented the collective responses of both students and instructors on different examination methods and their effects on teaching and learning. This research has given significant experimental confirmation to propose that student behavior and learning are very much influenced by evaluation methods. Erwin (1991) investigated issues caused by several assessment methods and proposed effective methods for better assessment. Later, Brown and Knight (1994) presented a comprehensive overview of the principles, purposes, practices and uses of assessment, with particular emphasis on student learning and development. McDowell (1995)investigated the impact of assessments, such as, non-conventional exams, oral presentations, group projects and peer assessment, on student learning via a series of case studies at the University of Northumbria. The examples of innovative assessment studied illustrate some very positive features which can do much to encourage student learning, but also indicate number of outstanding problems. Rust (2002) also investigated the impact of assessment on student learning and explored that how the research literatures practically help to inform the development of departmental assessment strategies and learner-centered assessment practices.

Many researches present in literature highlight the factors of failure in any exam. Hashmat et al. (2008)assessed examination related anxiety among final professional medical students and determined its factors. By considering a sample of 120 students, this study concluded that there is a moderate level of exam anxiety based on a Visual Analogue Scale contributed by factors such as extensive course load, lack of exercise and long duration of exams. Bishop (1998) claimed that curriculum-based external exit exam systems (CBEEESs) based on explicit content standards could improve the teaching and learning of core subjects. Descriptive analysis was used to conclude that the countries and Canadian provinces with such systems outperform other countries at a comparable level of development. Rehman et al. (2009) investigated the quality of faculty in public sector universities of Pakistan. By constituting sample of 300 teachers of public sector universities of Pakistan and using some statistical techniques, this research concluded that the teachers were competent enough to provide conductive learning environment. However, training in teaching methodology, curriculum planning and education required improvement. Majority of the teachers were dissatisfied with pay structure and withholding of housing facility which resulted in brain drain of teachers. Later, Gujjar et al. (2011)evaluated the prospective teachers’ teaching practice by selecting a sample of size 650 from 26 teacher training institutions consisting of both students and teachers. This study concluded that majority of the students and teachers are not given demonstration lessons in all subjects, and there choice is not considered in assigning lesson and duration of teaching practice. The evaluation method in practical component and proportion of theory and practical is inappropriate.

All the above literature focuses on different assessment methods, effects on students learning and success, and causes of failure. According to the best knowledge of authors, no research has been done to highlight the problems of the students who fail a course during any degree program and this is a fact that this issue is being currently discussed at large among university students. This research aims to fill this gap and claims that the supplementary exam is best for students’ academic progress and placement after graduation, whereas, course repetition affects students’ subsequent progress that declines due to various factors of the exam system. This investigation of the examination system can allow academic staff in universities to pay heed to this issue and adopt changes in examination system accordingly. This paper is divided into three major sections: methodology, results and discussion, and conclusion.

2. METHODOLOGY

The population used for the research is undergraduate students of all the universities and a sample of 100 students is selected from universities of each category of the exam system. The data is collected through questionnaires in which the respondent's CGPA and views about the current exam system are enquired using the two set of questions, one for each of the two categories of students and lastly, a set of questions enquired student's preference and its reasons.

The collected responses are tabulated with a value being assigned to each entry of the five points rating scale, that is, strongly agree = 5, agree = 4, neutral = 3, disagree = 2 and strongly disagree = 1. Results are extracted by using independent mean testing and chi square test of association.

3. RESULTS AND DISCUSSION

In this study, certain statistical techniques are employed to extract results from the collected data such as frequency plots, independent mean test (T-test), and chi square test of association. Firstly, the ratio of the respondent's preference presented in the frequency bar chart in Figure 1 clearly highlights that 80% of the 200 respondents prefer supplementary exam system over course repetition exam system.

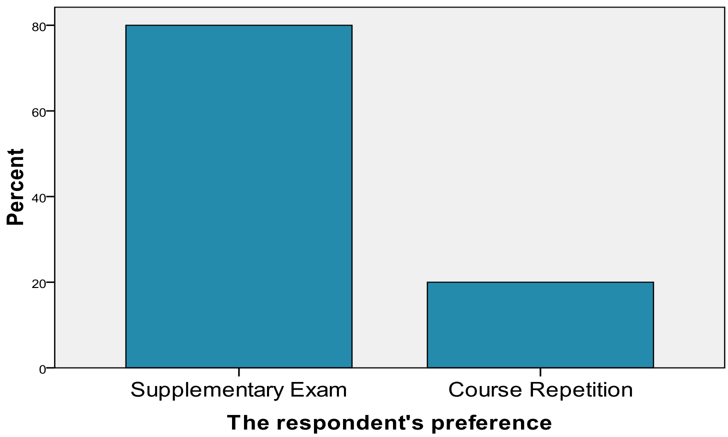


Fig. 1: Percentage Bar Chart Showing the Preference of Exam System

After frequency plot, independent mean test is applied to find out the mean CGPA of respondents of each category of the exam system. Normality test is done to ensure that the CGPA data is normal and Table 1 shows that the data is normal as significance value is more than the p value (0.05). This can also be seen from the histogram shown in Figure 2.

Table 1
Shapiro-Wilk Test of Normality

Exam System		Sig. value
CGPA of respondent	Supplementary Exam	.101
	Course Repetition	.657

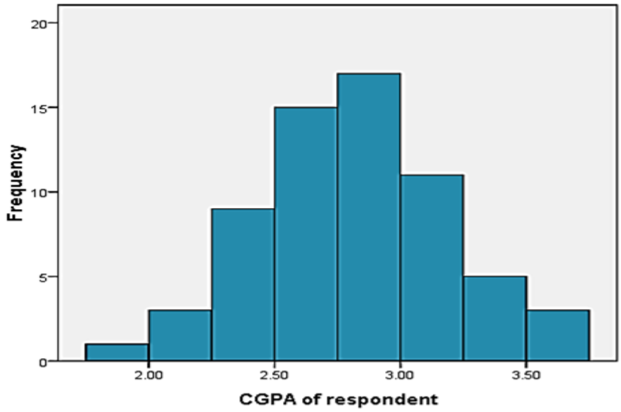


Fig. 2: Histogram of CGPA of Respondent

Upon calculating T-test for GPA of respondent in the second attempt of the course, no significant difference is found. This is concluded from Table 2 as it shows that the GPA of both the exam systems are not significantly different. However, the significance value in the T-test of CGPA of the respondent is less than the p value, indicating that there is significant difference in the mean of the two categories of exam system.

Table 2
Independent Samples Mean Test

Respondent's course GPA in second attempt.	Equal variances assumed	.160
	Equal variances not assumed	.162
CGPA of respondent	Equal variances assumed	.000
	Equal variances not assumed	.000

From Table 3, the mean of the CGPA of the respondents facing supplementary exam is greater than that of the respondents of course repetition exam system. Hence, it can be concluded from that the subsequent progress translated in terms of CGPA of the respondent is better of the students who face supplementary exam system.

Table 3
Mean CGPA of Respondent

	Exam System	N	Mean
CGPA of respondent	Supplementary Exam	100	3.1021
	Course Repetition	100	2.8378

Lastly, chi square test of association is applied on some pair of variables listed in Table 4 with their corresponding significance values to find out the association between them. It should be noted that the significant results are found in only those pair of variables whose significance value is less than the p value, that is, 0.05.

Table 4
Pair of Variables and their Pearson Chi-Square Sig. Value

Variable 1	Variable 2	Pearson Chi-Square Sig. Value
The exam system	The exam system helped the respondent	.005
The exam system	The exam system took much time	.000
The exam system	The exam system was costly	.000
The exam system	The exam system had much workload	.000
The exam system	Graduation was possible within the actual time	.000

The chi square test on the first set of variables explores the association between the exam system and the exam system helping the respondent. The histogram in Figure 3 shows that more of the students' facing supplementary exam strongly agree that the exam system helped the respondent whereas a few of course repetition exam system students agree with this which allows us to conclude that course repetition is comparatively less helpful than supplementary exam system.

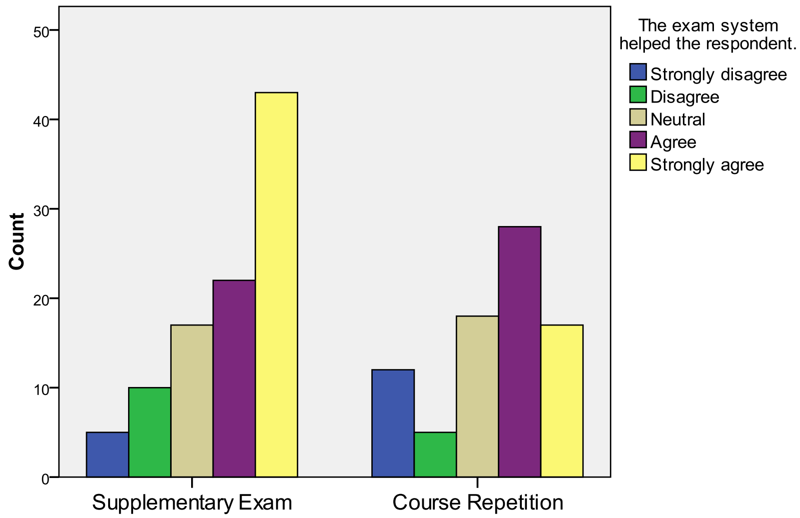


Fig. 3: Histogram of Association between the Exam System and the Exam System Helping the Respondent

The chi square test on the second set of variables explores the association between the exam system and which exam system takes more time. The histogram in Figure 4 shows that the respondents facing course repetition strongly agree that the exam system was too much time taking whereas opposite is the case with respondent's experiencing supplementary exam system. They highly disagree with this allowing us to conclude that supplementary exam system is not too much time taking. This is true as unlike spending a whole semester repeating a course, supplementary exam only takes time for appearing in the exam where preparation time varies according to student's will.

The chi square test on the third set of variables explores the association between the exam system and the exam system being costly. The histogram in Figure 5 shows that respondents facing course repetition strongly agrees that the exam system was costly whereas supplementary exam system respondents strongly disagree with this. This shows that according to respondents supplementary exam system is not costly. This can be backed up with the fact that there is a small fee for supplementary exam compared with the complete course fee paid in course repetition exam system.

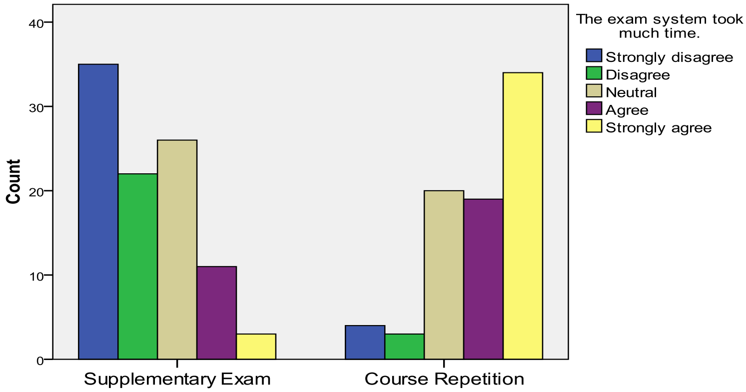


Fig. 4: Histogram of Association between the Exam System and the Exam System Taking Much Time

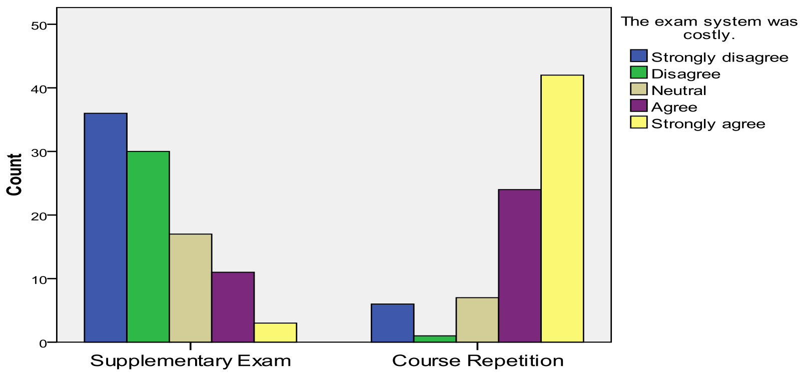


Fig. 5: Histogram of Association between the Exam System and the Exam System Being Costly

The chi square test on the fourth set of variables explores the association between the exam system and the exam system having much workload. The histogram in Figure 6 shows that respondents facing course repetition agree that the exam system has too much workload whereas most of the respondents of supplementary exam system strongly disagree with this fact. Hence, it can be concluded that supplementary exam system has less workload comparatively. This is true as in supplementary exam, only the time for appearing in exam is rigid. Other than this, every student can study and prepare for the exam as per one’s own will so there is not much workload.

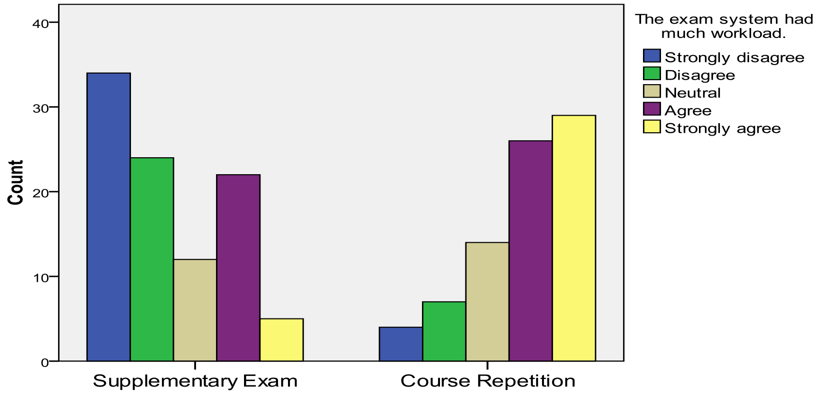


Fig. 6: Histogram of Association between the Exam System and the Exam System Having Much Workload

The chi square test on the last set of variables explores the association between the exam system and graduation being possible within the actual time. The histogram in Figure 7 shows that respondents facing course repetition strongly disagree that graduation was possible within actual time whereas most of the students of supplementary exam system agree with this which allows us to conclude that course repetition delays the student’s graduation. This is true as repeating a course in a semester does not allow the student to take fresh courses.

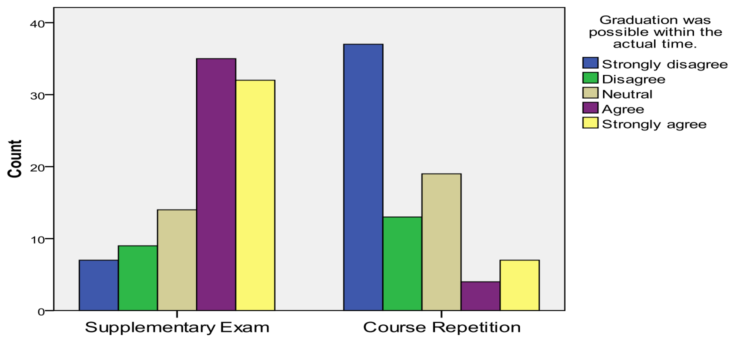


Fig. 7: Histogram of Association between the Exam System and Graduation Being Possible Within the Actual Time

4. COMMENTS AND CONCLUSION

The purpose of the research was to explore the impact of course repetition on students' subsequent academic progress at undergraduate level and to prove that offering supplementary exam is better than course repetition. The methodology used for data collection was questionnaires. The survey was conducted on a sample of 200 respondents, 100 from each category of the exam system.

The first main significant variable was the respondent's preference which indicated 80% respondents in favor of supplementary exam system. The second main significant variable was the CGPA of the respondent which is translated as students' subsequent progress, which showed that respondents facing supplementary exam system scored a better CGPA comparatively. The main factors of this are less workload and the exam system helping the students which results in better subsequent performance by the students.

5. ACKNOWLEDGEMENT

I would like to admit the sincere cooperation and help of many other individuals, without whom this project would not have been possible. I extend gratitude towards all those who provided us with valuable information in our survey. Furthermore, we would like to acknowledge financial support by FAST-NUCES to present this research in ISSOS. Lastly, I extend my appreciation to our parents and teachers who throughout the project helped and motivated us for completing this research work.

REFERENCES

1. Bishop, J.H. (1998). The effect of curriculum-based external exit exam systems on student achievement. *The Journal of Economic Education*, 29, 171-182.
2. Brown, S. and Knight, P. (1994). *Assessing learners in higher education*: Psychology Press.
3. Erwin, T.D. (1991). Assessing student learning and development: A guide to the principles, goals, and methods of determining college outcomes. Jossey-Bass, San Francisco.
4. Gujjar, A.A., Ramzan, M. and Bajwa, M.J. (2011). An evaluation of teaching practice: Practicum, *Pakistan Journal of Commerce and Social Sciences*, 5, 302-318.
5. Hashmat, S., Hashmat, M., Amanullah, F. and Aziz, S. (2008). Factors causing exam anxiety in medical students. *Journal-Pakistan Medical Association*, 58, 167.
6. McDowell, L. (1995). The impact of innovative assessment on student learning. *Programmed Learning*, 32, 302-313.
7. Miller, C.M. and Parlett, M. (1974). *Up to the mark: A study of the examination game*, Society for Research into Higher Education, London.
8. Rehman, S., Gujjar, A.A., Khan, S.A. and Iqbal, J. (2009). Quality of teaching faculty in public sector universities of pakistan as viewed by teachers themselves, *International Online Journal of Educational Sciences*, 1, 48-63.
9. Rust, C. (2002). The impact of assessment on student learning: How can the research literature practically help to inform the development of departmental assessment strategies and learner-centred assessment practices? *Active Learning in Higher Education*, 3, 145-158.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 443-452

PRIVATIZATION OF PAKISTAN INTERNATIONAL AIRLINES: EFFECTS ON GENERAL PUBLIC AND EMPLOYEES

Sauman Balkhi, Hammad Khan, Akif Jawed, Hasan Khan
and **Faqiha Sultan**
FAST-National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: k142107@nu.edu.pk
k142145@nu.edu.pk
k142093@nu.edu.pk
k142086@nu.edu.pk
faqiha.sultan@nu.edu.pk

ABSTRACT

The main purpose of this research is to recognize the effects the privatization of Pakistan international airlines (PIA) will throw on the public and PIA employees and their perspective about it. The changing worldwide air transport environment creates challenges for the administration of the airline. PIA is in decline and has had financial losses since last ten years that may push it into accepting fundamental changes very rapidly. This paper underlines latest changes in government arrangement and privatization of PIA, which expected to create a more successful, and cognizant air industry and influence the public and PIA employees. For this purpose, the data is collected from two groups affected by this change in PIA administration. 50 PIA professionals and 200 public individuals are surveyed using different questionnaire for each group. Statistical analysis for both the groups is carried out by using SPSS and the financial impact of PIA privatization on these groups is compared.

KEYWORDS

Privatization; Airline Industry; PIA Performance; General public.

1. INTRODUCTION

Privatization of any organization has its own advantages; it becomes more efficient, faces less political interference, less pressure from shareholders and emerge as a competitive company. However, it also have some liabilities like problems of debts, pension, insurance, and high rates for public and customers. In past, many organizations in all around the world were privatized and many researches were conducted to investigate the effects of privatization. Caves (1990) explored that the privatization of most of Britain's major public enterprises resulted in interesting conclusions such as liberalization to achieve productive efficiency, predicting slack absorption, and future productivity, emphasizing that the privatized sectors are more consistent than government sectors. Sader (1995) proposed that privatization in developing countries has also helped understanding the foreign investment as well as investment from direct investors. This helps in good decision-making and promoting sales and making country a good infrastructure for further investment. Another British survey calculated revenue, income, assets, debts, and load and concluded that they are more profitable after privatization. And that stock price of competitors fell by 7% after privatization (Eckel et al., 1997). Boubakri and Cosset (1998) examined the change in performance of 79 companies from 21 developing countries that were privatized fully or partially. The survey was conducted with the help of questionnaire and a constructive comparison was made analyzing the profits, capital, investments, productivity and employment criteria. Al-Jazzaf (1999) conducted an empirical study on privatization of firms by examining 10 privatized airlines and investigated its effects on the performance, efficiency, employment, and profit margins of the organization. The effects of privatization on inner data of the airlines such as sales, income, revenue, assets and expenses were also calculated in this study. Backx et al. (2002) presented a comparative study of private and state owned companies observing the profit margin between the two sectors as well as their rates. The study was conducted by dividing the companies into private, public owned and mixed ownership and evaluating their performance and grading them according to their success.

Privatization helps to lead improvement within the airline. Kenya Airways is another example for this where the revenue generated and the burden on employees is catered (Oyieke, 2002). Fischer et al. (2003) examined 37 Chilean state-owned enterprises (SOEs) that were privatized from 1981-2000. Financial status of this airline after privatization and data related efficiency were gathered using paper surveys. Debrah and Toroitich (2005)claimed that many more terms regarding privatization of any institution could be learnt from privatization in Kenya, which focuses more on organization management, achieving profits and dealing with problems. Starring on independent variables like political interference, development of ports and previous arrears.

The most important part in any company or organization is customer satisfaction, which is directly related to desired services within reasonable fares. This trend in turn generates repeat clientele, ensures a preferred supplier status from other enterprises, and enhances prospective market shares (Al-Jabaly & Khraim, 2014). However, privatization is based on the capability of different countries and it promote private sector in that particular country. Apart from efficiency, profit and job opportunities it also has some lacking on promoting private sector (Mugume, 2001). A survey on public enterprise tells that private sectors are about 12-100% more efficient than public ownership. Privatization for this study means deregulation, liberalization, market and sales (Domberger & Piggott, 1986).

The purpose of this research is to examine the effect of privatization on public and employees of PIA and how it will change the organization and its performance as well as efficiency. To investigate this, a survey is conducted to compare the pre and post privatization effects. The contrast between these two cases gives an estimate of the effects on public and employees due to privatization. In particular, the study tries to determine whether, following privatization, the firm will increase its profitability, operating efficiency, capital expenditures, and output. The technique used for this study is questionnaires survey from public and employees of PIA to explore their views on privatization and how it will change the services of PIA for customers and employees.

2. METHODOLOGY

In order to know the opinion of public and PIA employees the technique used was questionnaire. The population for this research was divided into two categories: PIA employees and public. A sample of size 150 was taken from public and 50 from employees. The questions selected for this research were about general views from public about their experience of PIA services and how privatization will make changes in fares, services and whether or not it will benefit public sector. On the other hand, questions from employees are about their job, salary, workload and post privatization effects.

The data collected from these source was tabulated and questions asked were graded as follow: a) strongly agree b) agree c) neutral d) disagree e) strongly disagree.

In this investigation, a statistical software SPSS-17 was used to perform Chi-square test for discovering the relationship between different variables and to find out the significant association between them. The Chi-square test is suitable for the situation here to find relation between four variables i.e. between fares and preference after privatization, benefits to the public sector and privatization is the only solution.

3. RESULTS

In this study, opinions of public and PIA employees collected through questionnaire survey are analyzed by using a statistical software SPSS-17. Data is summarized through charts and tables, and Chi-square test is used to find the association between some important variable.

Table 1 shows the significance of relation between public opinion about change in fares and their preference of PIA after privatization and secondly, relation between benefits to public sector due to privatization and privatization being the only solution. The main reason to know the significant relation between fares and preference of PIA after privatization was to know how much effect change of fares are going to hamper the preference of PIA after privatization. After applying chi-square test on two variables, the Pearson Chi-square value was 0.002, which shows its significance. 50% of the people who strongly agree that privatization will have an effect on fare will use PIA after privatization. 87.5% agree that privatization will have an effect on fare and they will use PIA after privatization. 28.6% remained neutral on the point that privatization will have an effect on fare and they will use PIA after privatization. 71.4% of the people who disagreed that privatization will have an effect on fare and they will use PIA after privatization. 100 of the people who strongly disagreed that privatization will have an effect on fare and they will use PIA after privatization.

Table 1
Cross Tabulation of Public Opinion about Change in Fares and PIA Preference after Privatization

			Will you prefer PIA even after privatization?		
			Yes	No	Total
Privatization will have effect on the fares	Strongly agree	Count	14	14	28
		% within Privatization will have effect on the fares	50.0%	50.0%	100.0%
		% within Will you prefer PIA even after privatization?	22.2%	50.0%	30.8%
	Agree	Count	35	5	40
		% within Privatization will have effect on the fares	87.5%	12.5%	100.0%
		% within Will you prefer PIA even after privatization?	55.6%	17.9%	44.0%
	Neutral	Count	2	5	7
		% within Privatization will have effect on the fares	28.6%	71.4%	100.0%
		% within Will you prefer PIA even after privatization?	3.2%	17.9%	7.7%
	Disagree	Count	10	4	14
		% within Privatization will have effect on the fares	71.4%	28.6%	100.0%
		% within Will you prefer PIA even after privatization?	15.9%	14.3%	15.4%
	Strongly disagree	Count	2	0	2
		% within Privatization will have effect on the fares	100.0%	.0%	100.0%
		% within Will you prefer PIA even after privatization?	3.2%	.0%	2.2%

Figure 1 shows the response of public when asked that will privatization effect the fares. The result shows that mostly agree or strongly agree that fares will be effected due to privatization.

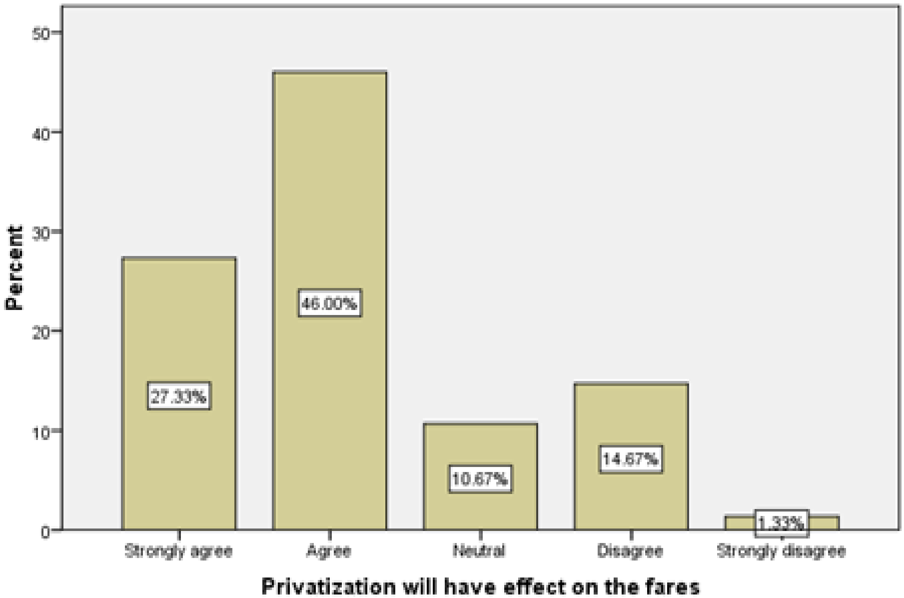


Fig. 1: Public Response on Expected Change in Fares after Privatization of PIA

The second most significant relation is benefits to public sector due to privatization and privatization being the only solution. The significance of this relation is to know that will privatizing PIA only benefit PIA or also the public sector. After applying chi-square test on the following relation results were produced which are shown in Table 2. After applying chi-square test on 2 variables the Pearson Chi-square value was 0.005. 16.7% strongly agree that privatizing PIA will benifet public sector and they also strongly agree that privatization is the only solution, 41.7% agree, 8.3% remained nuetral, 25% disagreed and 8.3% strongly disagreed that privatizing PIA is the only solution and also strongly agreed that privatizing will benifet the public sector. 13.2% agreed that privatizing PIA will benifet public sector and they also strongly agree that privatization is the only solution, 26.3% agree, 21.1% remained nuetral, 34.2% disagreed and 5.3% strongly disagreed that privatizing PIA is the only solution and also agreed that privatizing will benifet the public sector.0% remained nuetral on the point that privatizing PIA will benifet public sector and they also strongly agree that privatization is the only solution, 11.8% agree, 35.3% remained nuetral, 29.3% disagreed and 23.5% strongly disagreed that privatizing PIA is the only solution and remained nuetral that privatizing will benifet the public sector. 0% disagreed that privatizing PIA will benifet public sector and they also strongly agree that privatization is the only solution, 0% agree, 20% remained nuetral, 50% disagreed and 30% strongly disagreed that privatizing PIA is the only solution and also disagreed that privatizing will benifet the public sector.

Table 2
Cross Tabulation of Public Opinion about the Benefits PIA Privatization will give to Public and Privatization be the only Solution

A		B	Privatization is the only Solution: Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Total
Privatizing PIA will benefit public sector	Strongly agree	Count	2	5	1	3	1	12
		% within A	16.7%	41.7%	8.3%	25.0%	8.3%	100.0%
		% within B	28.6%	29.4%	5.3%	9.4%	6.3%	13.2%
	Agree	Count	5	10	8	13	2	38
		% within A	13.2%	26.3%	21.1%	34.2%	5.3%	100.0%
		% within B	71.4%	58.8%	42.1%	40.6%	12.5%	41.8%
	Neutral	Count	0	2	6	5	4	17
		% within A	.0%	11.8%	35.3%	29.4%	23.5%	100.0%
		% within B	.0%	11.8%	31.6%	15.6%	25.0%	18.7%
	Disagree	Count	0	0	4	10	6	20
		% within A	.0%	.0%	20.0%	50.0%	30.0%	100.0%
		% within B	.0%	.0%	21.1%	31.3%	37.5%	22.0%
	Strongly disagree	Count	0	0	0	1	3	4
		% within A	.0%	.0%	.0%	25.0%	75.0%	100.0%
		% within B	.0%	.0%	.0%	3.1%	18.8%	4.4%
Total		Count	7	17	19	32	16	91
		% within A	7.7%	18.7%	20.9%	35.2%	17.6%	100.0%
		% within B	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

7.7% strongly disagreed that privatizing PIA will benifet public sector and they also strongly agree that privatization is the only solution, 18.7% agree, 20.9% remained nuetral, 35.2% disagreed and 17.6% strongly disagreed that privatizing PIA is the only solution and also strongly disagreed that privatizing will benifet the public sector.

Figure 2 shows the response of public when asked that will PIA face losses after privatization. The result shows that majority remained neutral or disagreed.

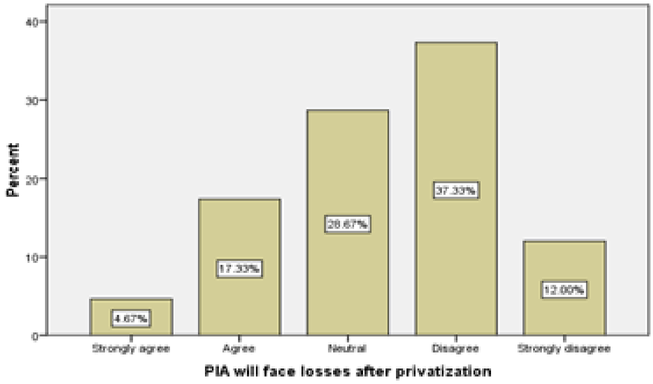


Fig. 2: Public Perspective about Loss to PIA after Privatization

Figure 3 shows the percentage of employees fearing for their job after privatization. The chart shows that majority of the employees fear losing their job. Figure 4 shows the response of employees when asked that is privatization the best solution. The result shows that most of them disagreed to this statement.

A separate questionnaire was prepared for the PIA employees and Chi-Square test was applied on the results, which produced one significant relation that was between salaries and working environment. The result produced after applying the Chi-Square test is shown in table 3. The Pearson Chi-Square value after applying the test was 0.004, which is less than 0.005. Through this test, we get to know that 94.1% believe that their salaries will increase and working conditions will improve, 5.9% think that conditions will remain the same and their salary will increase while none of those who thought that salaries would increase also believed that conditions would deteriorate. 50% employees believe that salaries will increase while working conditions will improve and the remaining 50% believe that their salaries will decrease and working condition would deteriorate. 100% employees believed that salaries will remain the same and working condition would improve.

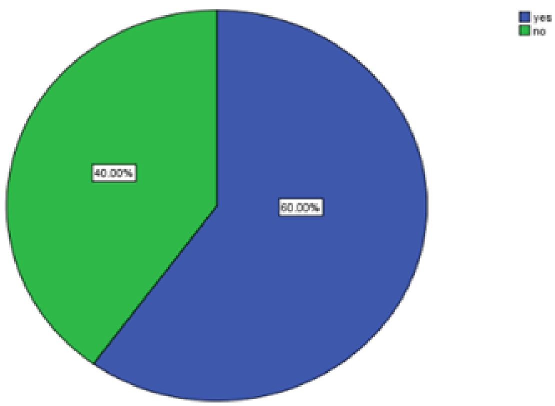


Fig. 3: Do you Fear Losing Job after Privatization

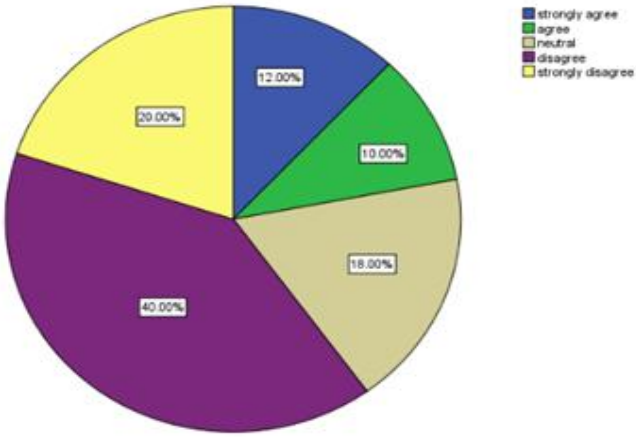


Fig. 4: Privatization is the Best Solution for the Employees

Table 3
PIA Employees Response to Salary Increment and Working Environment after Privatization
Salaries of Employees will * Working Environment will Cross Tabulation

Working Environment Will						
			Improve	Deteriorate	Remain the Same	Total
salaries of employees will	increase	Count	16	0	1	17
		% within salaries of employees will	94.1%	.0%	5.9%	100.0%
	decrease	Count	1	1	0	2
		% within salaries of employees will	50.0%	50.0%	.0%	100.0%
	remain the same	Count	11	0	0	11
		% within salaries of employees will	100.0%	.0%	.0%	100.0%
Total		Count	28	1	1	30
		% within salaries of employees will	93.3%	3.3%	3.3%	100.0%

Table 4
Pearson Chi-Square Coefficients Showing Association in Important Variables

Variable 1	Variable 2	Pearson's Chi Square Value
1) Privatizing PIA will benefit public sector	Privatization is the only solution	.005
2) Privatization will have effect on the fares	Will you prefer PIA even after privatization?	.002
3) Salaries of employees	Working environment	.004

4. COMMENTS AND CONCLUSION

The analysis based on public and employees of PIA suggests that privatization of PIA has conflicting effects on public and employees of PIA. Some surprising relations between important variables were found, majority of the employees were not in favor of privatization and feared that they would lose their job after privatization. Similarly, the public believe that fares will increase after privatization and most of them prefer to travel on PIA after privatization. However, the actual effect of privatization on public and employees could not investigated and needs to be researched in few years when PIA will be privatized.

5. ACKNOWLEDGEMENT

We would like to express our gratitude towards Manager Finance PIA for helping us in collection of the valuable information regarding PIA employees. Furthermore, we would also like to acknowledge FAST-NUCES for bearing all the financial responsibilities of our research.

REFERENCES

1. Al-Jabaly, S.M. and Khraim, A.S. (2014). The effect of perceived value and customer satisfaction on perceived price fairness of airline travelers in Jordan. *Universal Journal of Management*, 2, 186-196.
2. Al-Jazzaf, M.I (1999). Impact of privatization on airlines performance: An empirical analysis. *Journal of Air Transport Management*, 5, 45-52.
3. Backx, M., Carney, M. and Gedajlovic, E. (2002). Public, private and mixed ownership and the performance of international airlines. *Journal of Air Transport Management*, 8, 213-220.
4. Boubakri, N. and Cosset, J.C. (1998). The financial and operating performance of newly privatized firms: Evidence from developing countries. *The Journal of Finance*, 53, 1081-1110.
5. Caves, R.E. (1990). Lessons from privatization in Britain: State enterprise behavior, public choice, and corporate governance. *Journal of Economic Behavior & Organization*, 13, 145-169.
6. Debrah, Y.A. and Toroitich, O.K. (2005). The making of an African success story: The privatization of Kenya Airways. *Thunderbird International Business Review*, 47, 205-230.

7. Domberger, S. and Piggott, J. (1986). Privatization policies and public enterprise: A survey. *Economic Record*, 62, 145-162.
8. Eckel, C., Eckel, D. and Singal, V. (1997). Privatization and efficiency: Industry effects of the sale of British Airways. *Journal of Financial Economics*, 43, 275-298.
9. Fischer, R.D., Gutierrez, R. and Serra, P. (2003). *The effects of privatization on firms and on social welfare: The Chilean case*. Research Department Publications: Inter-American Development Bank.
10. Mugume, A. (2001). *The privatisation process and its impact on society*, A report submitted to Uganda National NGO Forum, Structural Adjustment Participatory Review Initiatives.
11. Oyieke, S. (2002). *Kenya Airways: A case study of privatization* (Vol. 119): African Economic Research Consortium.
12. Sader, F. (1995). *Privatizing public enterprises and foreign investment in developing countries, 1988-93* (Vol. 5): World Bank Publications.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 453-462

BAYESIAN ANALYSIS OF VAN BAAREN MODEL (V) FOR PAIRED COMPARISON ALLOWING TIES

Sadia Qamar[1§], Amna Nazeer[2] and **SaminaSatti[3]**

[1] Department of Statistics, University of Sargodha, Pakistan
[2] COMSATS Institute of Information Technology, Islamabad, Pakistan
[3] University of Wah, Wah Cantt, Pakistan
[§] Corresponding author Email: diyaqau@gmail.com

ABSTRACT

Sometimes it may be tricky for a panelist to compare two or more treatments (objects, items or options) when there is tie effect. In this study Bayesian technique is used for analysis of the van Baaren model (V) for paired comparison. Analysis of the van Baaren model (V) taking tie effect into account is done in unique way. It is hoped that this study will enhance the importance of tied observations in paired comparisons and will serve to increase interest to analyze the paired comparison through Bayesian paradigm. The two non-informative priors (Uniform and Jeffreys prior) are used for Bayesian analysis of the model by using Quadrature method. The posterior means, preference, predictive and posterior probabilities for hypothesis testing are computed. Graphical presentations of marginal densities are also provided. The appropriateness of the model has been checked by using Chi-square distribution.

KEYWORDS

The van Barren Model; Non-informative prior; Bayesian Inference; Predictive Probability; Quadrature method.

1. INTRODUCTION

This study elaborates the Bayesian analysis of paired comparison (PC) data through a PC model proposed by van Barren in 1978. This model is an extended form of the basic Bradley-Terry PC model. Since Glenn and David (1960) modified the Thrustone-Mosteller PC model by accommodating ties in a PC experiment. Rao and Kupper (1967) and Davidson (1970) made similar extension for the Bradley-Terry PC model. A wide-ranging bibliography on PCs is found in Davidson and Farquhar (1976). Baaren (1978) had further experimented with the Bradley-Terry PC model in a different way and modified the Bradley-Terry PC model to produce six different extensions or in other words proposed six new PC models. These new suggested models accommodated the parameter of no preference in different ways. The model (V) is one of the interesting extensions of those six PC models used for the comparison of several objects pairwise, simultaneously. One modification is made in the model (V) here that the order of presentation of the objects is not given consideration and hence mainly focuses on the tie parameter.

The Bayesian logic offers a way to measure things that were previously immeasurable, allowing us to test hypothesis and predictions and thereby refine our conclusions and decisions. Bayesian analysis of PC models has been given by Davidson and Solomon (1973). From the past two centuries Bayesian approach has been widely refined, accepted and used in almost all research areas. Bayesian approach is proved to be an excellent approach and is widely used now-a-days for the analysis of PC models. Some comprehensive work has been presented on PC data under Bayesian frame work in different aspects by Kim (2005), Glickman (2008), Abbas and Aslam (2009), Altaf and Aslam (2012) and Hussain and Aslam (2013). The main objective of this article is to present a Bayesian analysis for PC model recommended by van Baaren (1978) using two non-informative prior distributions namely Uniform prior and Jeffreys prior. It is hoped that this study will increase the importance of tied observations in PC experiments and will serve to boost the interesting analysis of PC allowing ties through the Bayesian approach.

2. VAN BAAREN MODEL (V)

Let there are$m$ treatments to be compared pairwise in a PC experiment and θ_i be the worth parameter of the i^{th} treatment, the van Baaren model (V) with tie effect 'v' is then defined as follows:

$$\begin{aligned}
\theta_{i.ij} &= \frac{\theta_i}{\left(\theta_i + \theta_j + v\sqrt{\dfrac{\theta_j}{\theta_i}}\right)} \\
\theta_{j.ij} &= \frac{\theta_j}{\left(\theta_i + \theta_j + v\sqrt{\dfrac{\theta_i}{\theta_j}}\right)} \qquad (1) \\
\theta_{o.ij} &= \frac{v^2 + 2v\sqrt{\theta_i\theta_j}}{\left(\theta_i + \theta_j + v\sqrt{\dfrac{\theta_j}{\theta_i}}\right)\left(\theta_i + \theta_j + v\sqrt{\dfrac{\theta_i}{\theta_j}}\right)}
\end{aligned}$$

$\theta_{i.ij}$ represents the probability of preference of i^{th} treatment over j^{th} treatment when the pair-wise comparison is made among i^{th} and j^{th} treatment and $\theta_{0.ij}$ is the probability of stating tie. By letting v=0; the model reduced to the basic Bradley-Terry PC model.

3. BAYESIAN ANALYSIS OF THE MODEL

Likelihood function and the prior distribution are the key ingredients of any Bayesian analysis. The likelihood function of outcome observed in **x** of trial is:

$$l(\mathbf{x};\theta_1,\theta_2...\theta_m,v) = \prod_{i<j=1}^{m}\prod_{k=1}^{r_{ij}} P_{ijk}$$

$$l(\mathbf{x};\theta_1,\theta_2...\theta_m,v)=\prod_{i<j=1}^{m}\prod_{k=1}^{r_{ij}}\left[\frac{\theta_i}{A_{ij}}\right]^{n_{i.ijk}}\left[\frac{\theta_j}{B_{ij}}\right]^{n_{j.ijk}}\left[\frac{v^2+2v\sqrt{\theta_i\theta_j}}{A_{ij}B_{ij}}\right]^{n_{o.ijk}}$$

Simplest form of likelihood function is as,

$$l(\mathbf{x};\theta_1,\theta_2...\theta_m,v)=\prod_{i<j=1}^{m}\frac{K_{ij}\theta_i^{n_{i.ij}}\theta_j^{n_{j.ij}}\left(v^2+2v\sqrt{\theta_i\theta_j}\right)^{t_{ij}}}{\left(A_{ij}\right)^{w_{i.ij}}\left(B_{ij}\right)^{w_{j.ij}}} \qquad (2)$$

where P_{ijk} is the probability of the observed results in the k^{th} repetition of the pair (T_i,T_j), $n_{i.ij}$ is the number of times i^{th} treatment is preferred, t_{ij} is the number of times tie is stated and $w_{i.ij}=n_{i.ij}+t_{ij}$, $w_{j.ij}=n_{j.ij}+t_{ij}$.

$$A_{ij}=\left(\theta_i+\theta_j+v\sqrt{\frac{\theta_j}{\theta_i}}\right),\qquad B_{ij}=\left(\theta_i+\theta_j+v\sqrt{\frac{\theta_i}{\theta_j}}\right)$$

and

$$K_{ij}=r_{ij}\Big/\left(n_{0.ij}!\,n_{i.ij}!\,n_{j.ij}!\right),\ 0\le\theta_i\le 1,\ (i=1,2,....,m),\ \sum_{i=1}^{m}\theta_i=1 \text{ and } v>0\ .$$

3.1. Prior Distribution Choice

Prior distribution choice is a matter of fact in any Bayesian analysis. Elicitation of readily available prior information and further formulation of it into distribution is not an easy task in the multi parameter situations. In situations where such prior elicitation is hard to find and less prior information are given, one can do analyses with conventionally chosen priors that may imitate less prior information. These kind of priors are known as non-informative or indifference, ignorance and vague priors. Several approaches for the choice of the non-informative prior has been argued but here we used the Uniform prior and Jeffreys prior for analysis.

3.2. Bayesian Analysis of Using Uniform Prior for Three Treatments

Let there are three parameters θ_1,θ_2 and θ_3 which are measuring worth of three treatments evaluated pair wise and v is threshold parameter. The likelihood function for m=3 using expression (2) is as follows:

$$l(\mathbf{x};\theta_1,\theta_2...\theta_m,v)=C\,\frac{\theta_1^{n_1}\theta_2^{n_2}(1-\theta_1-\theta_2)^{n_3}\prod_{i(<j)=1}^{3}\left(v^2+2v\sqrt{\theta_i\theta_j}\right)^{t_{ij}}}{\prod_{i(<j)=1}^{3}\left(A_{ij}\right)^{w_{i.ij}}\left(B_{ij}\right)^{w_{j.ij}}} \qquad (3)$$

In order to perform the Bayesian analysis for this case the data is taken from Aslam (2005).

Table 1
Data Allowing Ties for m=3

Pairs	$n_{i.ij}$	$n_{j.ij}$	$n_{o.ij}$	r_{ij}
(1,2)	5	24	1	30
(1,3)	7	19	4	30
(2,3)	20	7	3	30

3.3.3. Posterior Distribution of Parameters

Uniform Prior for analysis is defined as:

$$p(\theta_1,\theta_2,\theta_3,v) \propto 1, \qquad 0 \le \theta_i \le 1, \quad i = 1,2,3 \text{ and } v > 0 . \tag{4}$$

While joint posterior distribution for parameters by utilizing (3) and (4) is given as:

$$p(\theta_1,\theta_2;v|x) = K^{-1} \frac{\theta_1^{n_1}\theta_2^{n_2}(1-\theta_1-\theta_2)^{n_3} \prod\limits_{i(<j)=1}^{3} \left(v^2 + 2v\sqrt{\theta_i\theta_j}\right)^{t_{ij}}}{\prod\limits_{i(<j)=1}^{3} \left(A_{ij}\right)^{w_{i.ij}} \left(B_{ij}\right)^{w_{j.ij}}} \tag{5}$$

3.3.4. Posterior Marginal Distribution of Parameters

For parameter θ_1 the marginal posterior distribution is given as under:

$$p(\theta_1|x) = \int\limits_{\theta_2}^{1-\theta_1} \int\limits_{v=0}^{\infty} \frac{K^{-1}\theta_1^{n_1}\theta_2^{n_2}(1-\theta_1-\theta_2)^{n_3} \prod\limits_{i(<j)=1}^{3} \left(v^2 + 2v\sqrt{\theta_i\theta_j}\right)^{t_{ij}}}{\prod\limits_{i(<j)=1}^{3} \left(A_{ij}\right)^{w_{i.ij}} \left(B_{ij}\right)^{w_{j.ij}}} dv\,d\theta_2, \; 0 \le \theta_i \le 1$$

Similarly other marginal posterior distributions can be obtained. Graphical representation of marginal posterior distribution for above data is presented as follows:

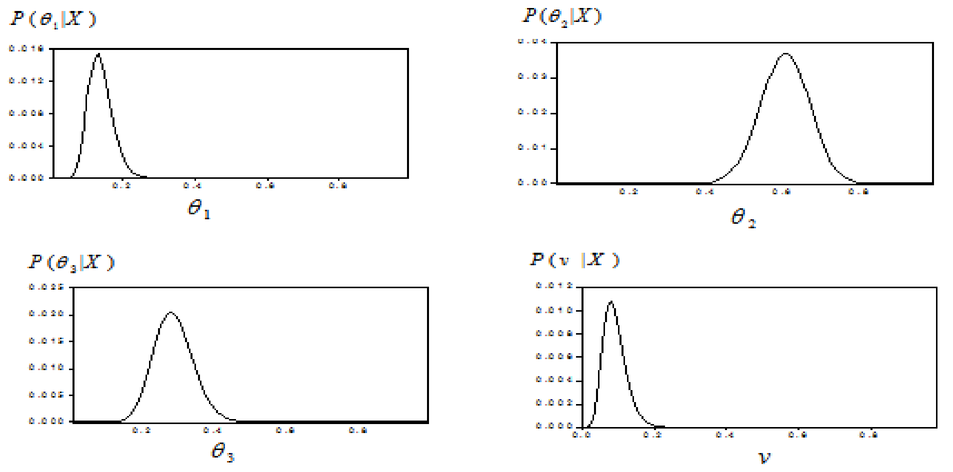


Figure 1: Graph of the Marginal Densities.

The figure indicates that posterior marginal distributions of parameters are approximately symmetrical about modal values.

3.3.5. Posterior Estimates and Preference Probabilities

We have calculated posterior estimates by using Quadrature method in SAS package to get required results and are provided in Table 2. Estimated chances of preferring one treatment instead of any other treatment in a single comparison are identified by preference probabilities. By putting posterior estimates in (1) preference probabilities can be obtained and given in Table 2.

3.3.6. Bayesian Hypotheses Testing

Different techniques are used for testing of hypotheses in Bayesian analysis, such as posterior probabilities, posterior odds and Bayes factor. Posterior probabilities are used to test the following two hypotheses H_{ij} and $\bar{H}_{ij}$ *(i<j=1,2,3)*:

$$\bar{H}_{ij} : \theta_j \geq \theta_i$$

For $H_{ij} = p_{ij} = P(\theta_i > \theta_j)$ the probability is taken as:

$$p_{ij} = P(\phi > 0|X) = \int_{\phi=0}^{1} \int_{\xi=\phi}^{(1+\phi)/2} \int_{v=0}^{\infty} P(\phi, \xi, v|X)\, dv\, d\xi\, d\phi\,,$$

While posterior probability for $\bar{H}_{ij} = q_{ij} = 1 - p_{ij}$. Decision rule utilized for acceptance or rejection of above hypothesis is according to Aslam (2002). The results thus obtained are discussed in Table 2.

3.3.7. Predictive Probabilities

The predictive probabilities of treatments for future single comparison is computed using following expression:

$$P_{i.ij} = P(T_i > T_j) = \int_{\theta_i=0}^{1} \int_{\theta_j=0}^{1-\theta_i} \int_{v=0}^{\infty} \theta_{i.ij}\, p(\theta_i, \theta_j, v|x)\, dv\, d\theta_j\, d\theta_i\,,$$

where $\theta_{i.ij}$ is given in (1) and $p(\theta_i, \theta_j, v|x)$ is posterior probability of θ_i, θ_j and v .No preference would be given when treatments T_i and T_j are compared in a single future comparison the predictive probability is:

$$P_{o.ij} = P(T_i = T_j) = \int_{\theta_i=0}^{1} \int_{\theta_j=0}^{1-\theta_i} \int_{v=0}^{\infty} \theta_{\mathrm{o.ij}}\, p(\theta_i, \theta_j, v|x)\, dv\, d\theta_j\, d\theta_i\,,$$

Results are illustrated in Table 2.

3.3.8. Bayesian Analysis of Model Using Jeffreys Prior for M=3

Another prior which is non-informative named as Jeffreys prior for parameters is chosen for the Bayesian analysis. Jeffreys prior is square root of Fisher's Information matrix. Derivation of Fisher's information matrix determinant is done by taking logarithm of likelihood function. The simplification of the determinant is complicated so it can be used numerically. By designing the program in SAS package, we calculate the numerical form of Jeffreys prior and use it for the calculation of joint and marginal posterior distributions and for further analysis.

3.3.9 Posterior Distribution and Posterior Marginal Distribution of Parameters

For the case of three treatment parameters posterior distribution by using Jeffreys prior, which is defined as the determinant of the Fisher information matrix, is given as:

$$p(\theta_1,\theta_2;v|x)=K^{-1}p_j(\theta_1,\theta_2,\theta_3,v)\frac{\theta_1^{n_1}\theta_2^{n_2}(1-\theta_1-\theta_2)^{n_3}\prod_{i(<j)=1}^{3}\left(v^2+2v\sqrt{\theta_i\theta_j}\right)^{t_{ij}}}{\prod_{i(<j)=1}^{3}\left(A_{ij}\right)^{w_{i.ij}}\left(B_{ij}\right)^{w_{j.ij}}}$$

$$0\le\theta_i\le 1,\ v\ge 0,\ \theta_1+\theta_2+\theta_3=1. \qquad (6)$$

where $p_j(\theta_1,\theta_2,\theta_3,v)$ is the Jeffreys prior distribution and K^{-1} is the normalizing constant and $\theta_3=1-\theta_1-\theta_2$ is the constraint used. The simplification of the determinant is complicated so it can be approximated numerically. By designing the program in SAS package, we calculate the numerical form of Jeffreys prior and use it for the calculation of joint and marginal posterior distributions and for further analysis.

The marginal posterior distribution of θ_1 is as follows:

$$p(\theta_1|x)=\int_{\theta_2}^{1-\theta_1}\int_{v=0}^{\infty}K^{-1}p_j(\theta_1,\theta_2,\theta_3,v)\frac{\theta_1^{n_1}\theta_2^{n_2}(1-\theta_1-\theta_2)^{n_3}\prod_{i(<j)=1}^{3}\left(v^2+2v\sqrt{\theta_i\theta_j}\right)^{t_{ij}}dv\,d\theta_2}{\prod_{i(<j)=1}^{3}\left(A_{ij}\right)^{w_{i.ij}}\left(B_{ij}\right)^{w_{j.ij}}} \qquad (7)$$

were $0\le\theta_i\le 1, v\ge 0$. Similarly, the marginal posterior distribution for θ_2,θ_3 and v can be derived. The graphical presentation of marginal posterior distributions is given as under in Figure 2.

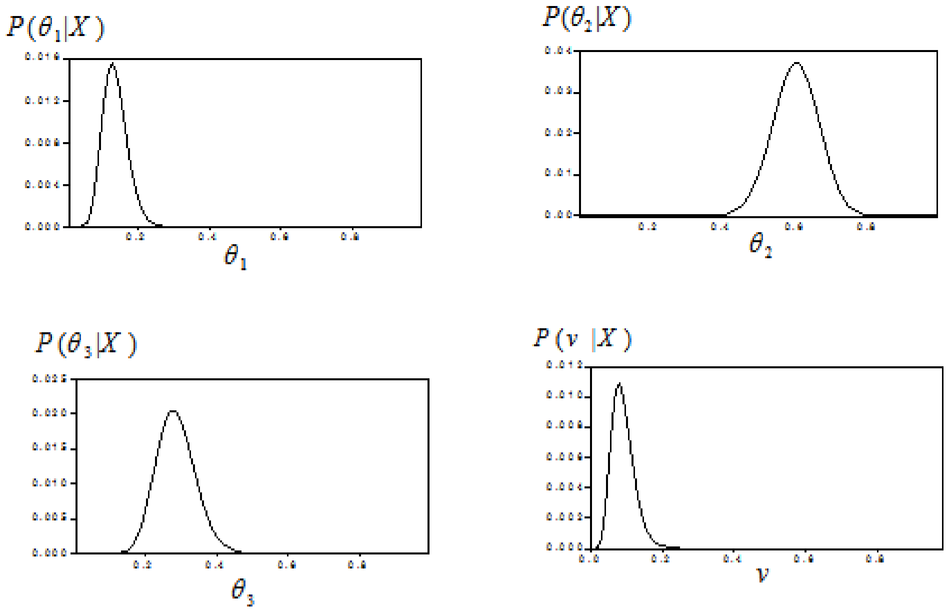


Figure 2: The Marginal Densities Graphical Representation.

The above figure indicates that posterior marginal distributions of parameters are approximately symmetrical about modal values. The required analysis on the model is repeated for Jeffreys Prior using posterior distribution given in (7) as already done by using Uniform Prior. The posterior estimates, preference probabilities, predictive probabilities and hypothesis testing are calculated through Quadrature method and given in Table 2.

4. RESULTS AND DISCUSSION

The results obtained from both non-informative priors are given below.

Table 2
Summary of Results

Results	Parameters	Via Uniform Prior	Via Jeffreys Prior
Posterior Means	θ_1	0.1289	0.12296
	θ_2	0.5961	0.61109
	θ_3	0.2751	0.26596
	v	0.0792	0.07640
Posterior Modes	θ_1	0.1216	0.1264
	θ_2	0.6164	0.6024
	θ_3	0.2619	0.2704
	v	0.0705	0.0757
Preference Probabilities	$\theta_{1.12}$	0.14393	0.13596
	$\theta_{2.12}$	0.78251	0.79536
	$\theta_{0.12}$	0.07356	0.06868
	$\theta_{1.13}$	0.24796	0.24529
	$\theta_{3.13}$	0.60043	0.60326
	$\theta_{0.13}$	0.15161	0.15145
	$\theta_{2.23}$	0.64444	0.65889
	$\theta_{3.23}$	0.27849	0.27647
	$\theta_{0.23}$	0.07707	0.07354
Predictive Probabilities	$P_{1.12}$	0.14573	0.14317
	$P_{2.12}$	0.78160	0.78631
	$P_{0.12}$	0.07265	0.07051
	$P_{1.13}$	0.25029	0.26659
	$P_{3.13}$	0.60061	0.58185
	$P_{0.13}$	0.14908	0.15156
	$P_{2.23}$	0.64441	0.65799
	$P_{3.23}$	0.27975	0.26923
	$P_{0.23}$	0.07582	0.07278
Posterior Probabilities	p_{12}	8.5×10^{-8}	8.5×10^{-8}
	q_{12}	0.9999	0.9999
	p_{13}	0.0028	0.00280
	q_{13}	0.9971	0.99719
	p_{23}	0.99579	0.9958
	q_{23}	4.21×10^{-3}	4.2×10^{-3}

We have compared the results obtained under Uniform prior and Jeffrey's prior using Quadrature method. Both the priors give approximately same results. Moreover, both the method assigns the same rank to the parameters as indicated by preference probabilities. It was noted from the results that treatment 2 is preferred among treatment 1 and 3. The treatment 3 is preferred among treatment 1 means that $T_2 \rightarrow T_3 \rightarrow T_1$. By using both uniform and Jeffrey's priors posterior and predictive probabilities are also calculated and they are similar by using both priors. Most of calculated values agree up to two significant digits. Since, both the priors give the similar result so it is concluded that one can use either of two non-informative prior. By calculating posterior probabilities it is cleared that $\bar{H}_{12}$, $\bar{H}_{13}$ and H_{23} is accepted with very high probability.

5. APPROPRIATENESS OF THE MODEL FOR M=3

Observed number and expected number of preferences are compared for appropriateness of model. If discrepancies are small the solution is considered to be internally consistent. For testing purpose we use χ^2 statistic.
The two hypotheses are

H_o : The model is true for some values of $\theta = \theta_o$
H_1: The model is not true for any value of θ
χ^2 statistic has the fallowing form:

$$\chi^2 = \sum_{i<j}^{m} \left\{ \frac{(n_{i.ij} - \hat{n}_{i.ij})^2}{\hat{n}_{i.ij}} + \frac{(n_{j.ij} - \hat{n}_{j.ij})^2}{\hat{n}_{j.ij}} + \frac{(n_{o.ij} - \hat{n}_{o.ij})^2}{\hat{n}_{o.ij}} \right\} \tag{8}$$

With degrees of freedom{$m(m$-2$)$}.

Table 3
Results of Observed and Expected Number of Preferences

Pairs(i , j)		**Observed Values**	**Expected Values**	
			Via Uniform Prior	**Via Jeffreys Prior**
(1,2)	$n_{1.12}$	5	4.32	4.08
	$n_{2.12}$	24	23.48	23.86
	$n_{0.12}$	1	2.21	2.06
(1,3)	$n_{1.13}$	7	7.44	7.36
	$n_{3.13}$	19	18.01	18.1
	$n_{0.13}$	4	4.55	4.54
(2,3)	$n_{2.23}$	20	19.33	19.77
	$n_{3.23}$	7	8.35	8.04
	$n_{0.23}$	**3**	**2.31**	**2.2**

The value of χ^2 statistic computed through equation (8)is found to be 1.37 and p-value is 0.71, so no evidence is found that model does not fit.

6. SUMMARY

We have conducted Bayesian analysis of van Baaren model (V) for paired comparison considering two non-informative priors. Estimated posterior means and posterior modes for the model parameters helped to rank the treatments under study. Bayes ranking for the worth parameter is calculated under both priors and shown to coincide for two estimates. Preference probabilities also favor the same ranking most of the calculated results are similar up to two significant digits. As both the priors give the similar result it is suggested that any of two priors can be utilized as non-informative prior for future studies.

REFERENCES

1. Abbas, N., and Aslam, M. (2009). Prioritizing the items through paired comparison models, a Bayesian approach. *Pakistan Journal of Statistics*, 25(1), 59-69.
2. Altaf, S. and Aslam, M. (2012). Paired comparison analysis of the van Baaren model using Bayesian approach with non-informative prior. *Pakistan Journal of Statistics & Operation Research*, 8, 259-270.
3. Aslam, M. (2002). Bayesian analysis for paired comparison models allowing ties and not allowing ties. *Pakistan Journal of Statistics*, 18(1), 53-69.
4. Aslam, M. (2005). Bayesian comparison of the paired comparison models allowing ties. *Journal of Statistical Theory and Application,* 4, 161-171.
5. Barren, V.A. (1978). On a class of extension to the Bradley-Terry model in paired Comparisons. *Statistica Neerlandica*, 32, 57-67.
6. Bradley, R.A. and Terry, M.E. (1952). Rank analysis of Incomplete Block Designs, I. The method of Paired Comparisons. *Biometrika*, 39, 324-345.
7. Davidson, R.R. (1970). On Extending the Bradley-Terry Model to Accommodate Ties in the Paired comparison Experiments. *Journal of the American Statistical Association*, 65, 317-328.
8. Davidson, R.R. and Solomon, D.L. (1973). A Bayesian Approach to Paired Comparison Experiments. *Biometrika*, 60, 477-487.
9. Davidson, R.R. and Farquhar, P.H. (1976). A Bibliography on the Method of Paired Comparisons. *Biometrics,* 32, 241-252.
10. Glenn, W.A. and David, H.A. (1960). Ties in paired comparison experiments using a modified Thrustone-Mosteller model. *Biometrics*, 16, 86-109.
11. Glickman, M.E. (2008). Bayesian locally optimal design of knockout tournaments. *Journal of Statistical Planning and Inference*, 138, 2117-2127.
12. Hussain, S. and Aslam, M. (2013). Bayesian Inference of Log-Linear Version of the Bradley-Terry Model for Paired Comparisons Using Uninformative Priors. *International Journal of Probability and Statistics*, 2, 43-49.
13. Kim, D.H. and Kim, H.J. (2014). A Bayesian approach to paired comparison of several products of Poisson rates. *Proceedings of the Autumn Conference*, *Korean Statistical Society*.
14. Rao, P.V. and Kupper, L.L. (1967). Ties in paired comparison experiment. A Generalization of Bradley-Terry Model. *Journal of the American Statistical Association*, 62, 194-204.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 463-468

STATISTICAL STUDY OF MALAYSIAN HEALTH SECTOR

Imran Anwar Ujan, M. Malook Rind and **Mohd Adam Suhimi**
Kulliyyah of Information & Communication Technology,
International Islamic University, Malaysia

ABSTRACT

Malaysia is an energetic and active country increase economic growth and political constancy day by day. Peoples are healthier, live long life, more creative as compared by other developing countries. The healthcare attained is the key measures of the achievement of nation. Great health empowers Malaysians to lead profitable and satisfying lives. Abnormal state of health pays to expanded thriving and social quality.

1. POPULATION STRUCTURE OF MALAYSIA

According to the Population and Housing Census 2010, the country population in 2012 was 29.34 million, 1.3 was the annual population growth rate in 2011-2012. The population increased 0.38 million in 2012, in 2011 it was 28.96 million. The population of Selangor had highest 5.65 million, though Federal Territory of Putrajaya the lowest 0.08 million population (Table 1). Though, Federal Territory of Putrajaya had highest annual population growth rate 3.9, while Perak and Perlis was the lowest annual growth rate of 0.7.

Table 1
Population and Annual Population Growth Rate by State, Malaysia 2011-2012

State	Population ('000) 2011	Population ('000) 2012	Annual Population Growth Rate 2011/2012
Perlis	237.5	239.4	0.8
Kedah	1,973.1	1,996.8	1.2
Pulau Pinang	1,593.6	1,611.1	1.1
Perak	2,397.6	2,416.7	0.8
Selangor	5,577.4	5,650.8	1.3
FT Kuala Lumpur	1,694.5	1,713.4	1.1
FT Putrajaya	76.4	79.4	3.9
Negeri Sembilan	1,042.9	1,056.3	1.3
Melaka	833.0	842.5	1.1
Johor	3,401.8	3,439.6	1.1
Pahang	1,524.8	1,548.4	1.5
Terengganu	1,074.0	1,092.9	1.8
Kelantan	1,615.2	1,640.4	1.5
Sabah	3,316.4	3,371.7	1.7
FT Labuan	89.8	91.6	1.9
Sarawak	2,516.2	2,545.8	1.2
MALAYSIA	28,964.3	29,336.8	1.3

The population in Malaysia is comparatively young, where below 20 years was 35.9% of the total population, and 8.3% of 60 years and more shows in table 2. In 2012, the population 15 to 64 years of age was 20.0 million which was 68.3% of the full population, although the less than 15 years and above 65 years of age was 9.3 million which are 31.7% of the full population.

Table 2
Statistics Related to Population, 2012

Population	Number ('000)	% of Total Population
Male	15,093.7	51.4
Female	14,243.1	48.6
Youths (below 20 years)	10,544.4	35.9
Elderly (60 years and above)	2,438.6	8.3
Economically-productive (age 15-64 years)	20,034.5	68.3
Economically-dependent (age below 15 & above 64 year)	9,302.3	31.7

2. HEALTH STATUS

Health position be gaged by utilization of health position pointers. Pointers, for example, future during childbirth, mortality and dismalness position of nation were among the markers that can be measured, and assist as a sign of condition for health of people, and hence the strength of general populace.

Life Expectancy at Birth

The measure of quantity years, on a normal, that a man can expectation to live. With the change in the nutritious and financial position of the populace, Malaysian people hope to live more than previously. The evaluated future during childbirth in light of the assessed 2012 information has expanded to 72.3 years for male and 77.2 years for female individually, when contrasted with 70.8 years for male and 75.3 years for female recorded in 2002.

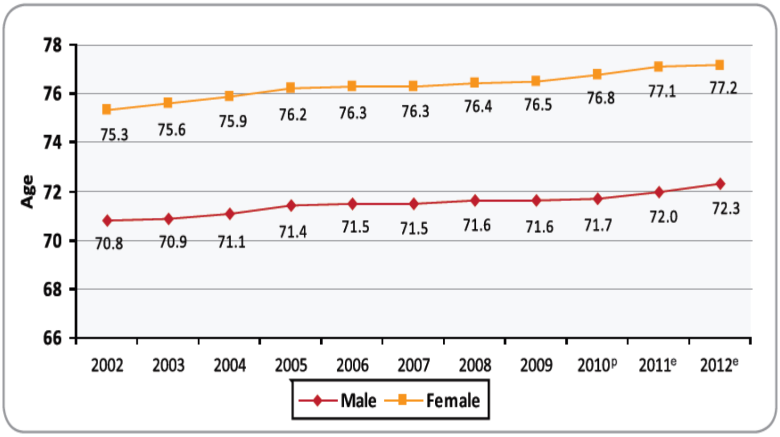
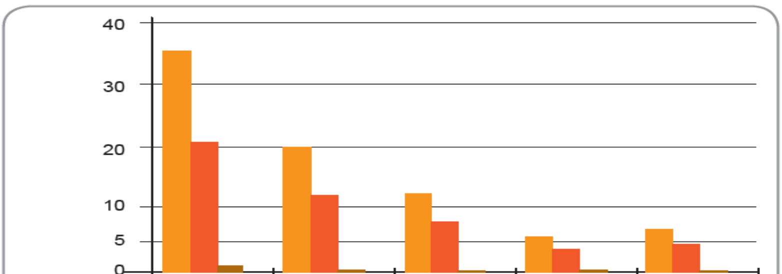


Fig. 1: Life Expectancy at Birth (in years) by Sex, Malaysia, 2002-2012

3. MORTALITY

Mortality information gives a helpful endpoint to measuring health. These information give a far reaching photo of the soundness of the group, since it covers each person. A wide range of sorts of measures are utilized to give perspectives of health from varying points of view.

For as far back as 40 years (1972-2012), the death rates in Malaysia had been diminishing. The pattern of maternal mortality proportion (MMR), newborn child death rate (IMR) and neonatal death rate (NMR) in Malaysia.



	1971	1982	1992	2002	2011p
IMR	36.1	19.5	12.1	6.5	6.6
NMR	21.0	12.1	7.9	3.8	4.2
MMR	1.0	0.5	0.2	0.3	0.3

Fig. 2: IMR, NMR and MMR, Malaysia, 1971-2011

The pattern for the other death rates remains generally the same from 2007 to 2011 (Table 3). Escalated inoculation endeavors and other related customized were done by both general society and private parts to enhance this rates. These information can likewise be credited to the nutritious status change of the youngsters, change of safety, and enhancing ecological conditions

Table 3
Mortality Rates in Malaysia, 2007-2011

Indicator	2007	2008	2009	2010	2011
Crude Death Rate (per 1,000 population)	4.5	4.7	4.8	4.6	4.6
Maternal Mortality Ratio (per 100,000 live births)	29.0	27.3	27.0	26.1	25.5
Infant Mortality Rate (per 1,000 live births)	6.2	6.2	6.9	6.7	6.6
Neonatal Mortality Rate (per 1,000 live births)	3.8	3.9	4.3	4.3	4.2
Under Five Mortality Rate (per 1,000 live births)	7.9	8.0	8.5	8.4	8.1
Toddler Mortality Rate (per 1,000 population aged 1-4 years)	0.4	0.4	0.4	0.4	0.4
Stillbirth Rate (per 1,000 births)	4.4	4.3	4.4	4.5	4.5
Perinatal Mortality Rate (per 1,000 births)	7.2	7.3	7.6	7.7	7.6

4. MORBIDITY

The health status of a group is generally measured regarding grimness, which concentrates on the frequency or commonness of sickness, and mortality, which portrays the extent of death in a populace.

Hospitalization shows the seriousness of sickness that needs facilitate treatment, adjustment of patients or the need of seclusion with a specific end goal to keep the spreading of the maladies to others. For the time of 2000-2012, the quantity of confirmations in MoH Hospitals had expanded 45.6% to 2,264,019 in 2012 from that of 1,555,133 in 2000. The 10 important reasons for hospitalization in the MoH Hospitals for 2012 are appeared in Table 4. The maladies were regrouped to groupings in light of the International Statistical Classification of Disease tenth Revision (ICD10). In 2012, "Pregnancy, labor and the puerperium" (27.32%) was the top reason for affirmations in MoH clinics took after by "Illnesses of the respiratory system"(11.02%).

Table 4
10 Principal Causes of Hospitalisation in MoH Hospitals, 2012

Principal Causes	Percentage to Total Admissions
1. Pregnancy, childbirth and the puerperium	27.32
2. Diseases of the respiratory system	11.02
3. Injury, poisoning and certain other consequences of external causes	8.22
4. Diseases of the circulatory system	7.55
5. Certain conditions originating in the perinatal period	7.55
6. Certain infectious and parasitic diseases	6.82
7. Diseases of the digestive system	4.88
8. Diseases of the genitourinary system	4.48
9. Factors influencing health status and contact with health services	3.64
10. Neoplasms	3.34

So also, the quantity of deaths (for all causes) in MoH Hospitals for the time of 2000-2012 expanded 67.7% from 30,319 in 2000 to 50,849 in 2012. “Diseases of the circulatory system" was the top reason for death in MoH healing centers recorded in 2012 (24.69%), trailed by “Diseases of the respiratory system" (18.80%) and “Certain infectious and parasitic diseases" (17.17%). The 10 foremost reasons for deaths in the MoH Hospitals for 2012 are as appeared in Table 5.

Table 5
10 Principal Causes of Death in MoH Hospitals, 2011

Principal Causes	Percentage to total deaths
1. Diseases of the circulatory system	24.69
2. Diseases of the respiratory system	18.80
3. Certain infectious and parasitic diseases	17.17
4. Neoplasms	11.64
5. Injury, poisoning and certain other consequences of external causes	5.32
6. Diseases of the digestive system	5.07
7. Diseases of the genitourinary system	4.18
8. Certain conditions originating in the perinatal period	3.43
9. Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	1.87
10. Diseases of the nervous system	1.75

5. HEALTH FACILITIES AND FACILITY UTILIZATION

In 2012, there were 919 Health Clinics, 1,831 Community Clinics and 106 Maternal and Child Health Clinics. In 2010, 1 Malaysia Clinic was propelled in chose urban territories. These offices give fundamental medicinal administrations to diseases and wounds, for example, fever, hack, colds, wounds and cuts, diabetes, and hypertension. To date, there are 178 1 Malaysia Clinics that give prompt medicinal services to populace.

Concerning doctor's facilities, there were 132 government MoH doctor's facilities and 8 Special Medical Institutions with bed correlative of 34,078 and 4,900 beds separately. General Bed Occupancy Rate (BOR) for MoH doctor's facilities and Institutions in 2012 was 72.13% (Table 6).

TABLE 6
Health Facilities by Type, Total Bed Complements and BOR, 2008-2012

Facility	2008	2009	2010	2011	2012
Number of MoH Hospital	130	130	131	132	132
Number of Special Medical Institution	6	6	6	6	8
Total Beds Complement 1	38,004	38,057	37,793	36,148	38,978
Bed Occupancy Rate (%)	65.46	65.45	66.26	68.63	72.13
Number of Health Clinics	802	808	813	879	919
Number of Community Clinics	1,927	1,920	1,916	1,864	1,831
Number of Maternal and Child Health Clinics	95	90	104	106	106
Number of 1 Malaysia Clinics	---	---	53	109	178

6. CONCLUSION

It is concluded that the Ministry of Health, Government of Malaysia pay attention to health sector more and increase the facilities day by day. Their mission is that nation working together for better health.

REFERENCES

1. Annual Report Ministry of Health 2012 (**Source of Data**)
2. Bush, M., Lederer, A.L., Li, X., Palmisano, J. and Rao, S. (2009). The alignment of information systems with organizational objectives and strategies in health care. *International Journal of Medical Informatics*, 78(7), 446-456.
3. Heller, P.S. (1982). A model of the demand for medical and health services in Peninsular Malaysia. *Social Science & Medicine*, 16(3), 267-284.
4. Panis, C.W. and Lillard, L.A. (1994). Health inputs and child mortality: Malaysia. *Journal of Health Economics*, 13(4), 455-489.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 469-478

FACTORS EFFECTING FIXED-LINE TELECOM SERVICES' CUSTOMER RETENTION: A STUDY OF PAKISTAN

M. Malook Rind, Tami Alzabi, Imran Anwar Ujan
and **Asadullah Shah**
Kulliyyah of Information & Communication Technology,
International Islamic University Malaysia
Email: engineermalook@gmail.com
alzabi.tami@gmail.com
iujan@yahoo.com
asadullah@iium.edu.my

ABSTRACT

The use of fixed-line telecom services in Pakistan is declining gradually since past few years. The purpose of this research is to propose an integrative model that helps to evaluate the main determinants of fixed-line telecom services' post purchase intention. This research proposes a comprehensive integrated framework that may help to measure the relationship among stability, content quality, responsiveness, staff attitude, functional value, perceived customer value, monetary value, perceived service quality, customer satisfaction and their impact on post-purchase intention of telecom services. Based on the proposed research framework built through a thorough literature review, authors intend to adopt quantitative, clustered sampling approach for data collection. A Structured Equation Modeling (SEM) approach will be adopted to test and validate the acquired data. Findings of this research will help operators to make proactive, feasible decisions towards best the utilization of their resources in maximizing profit, retaining customer base and achieving organization goals.

KEYWORDS

Fixed-line telecom services, customer satisfaction, post-purchase intention, Perceived quality, Perceived value.

1. INTRODUCTION

Keeping in view the increasing tough competitive market environment, it has been realized that the customer satisfaction and customer value with respect to service quality are the most important factors for the success and survival of any service oriented and manufacturing organization. Analysis of the determinant of fixed-line telecom services' customer retention can help service providers to optimize their services for the post-purchase intention behavior. Bandwidth hungry application like IP based premium HDTV (High Definition Television), high speed internet, video calling, video conferring, CCTV (Closed-Circuit Television) video surveillance and others, require high bandwidth core and access network with robust systems at backend. Traditionally, fixed line telecom service providers competed fiercely with wireless service provider for new customers (Lai et al., 2009). Fixed-line telecom services' user ratio is declining compared to mobile-cellular subscription (ITU-T, 2014). Gaining new customers may cost four times more than retaining existing customers (Grönroos, 1996). Fornell (1992) argued that high customer satisfaction leads to a stronger competitive position and a significant determinant of customer loyalty and post-purchase intention. Satisfied customers not only become loyal and return to buy again but also tell other people about their experiences (Fornell, et al., 1996). Overall customer satisfaction has a central role and is a strong predictor of Post-purchase intention (McQuitty et al., 2000; Oliver, 1993).

Pakistan is a huge telecom market with 3.172 million fixed-line telecom services' users (PTA, 2014). In 2003, Ministry of Information Technology, government of Pakistan introduced telecom deregulation policy (MOITT, 2003), with the aim to break the centralized monopolistic telecom market and create an open competitive market for telecom service providers. As a result, at present 13 telecom operators are competing and offering telecom services out of which 6 operators including Pakistan Telecommunication Company Limited (PTCL), National Telecommunication Company (NTC), Brain Telecom Limited, World Call, Union Telecom, and Naya Tel are providing fixed-line telecom services in 14 telecom regions of Pakistan (MOITT, 2014). Huge investment has been done to improve overall network quality, attain excellence in quality of service to attract and retain fixed-line customers. With the new entrants in telecom industry of Pakistan, the potential of this growing industry have faced a new hyper competition (Ali et al. 2010). This study aims to propose a conceptual framework that checks the relationship between perceived service quality and customer perceived value towards the overall customer satisfaction and the post-purchase intention of fixed-line telecom services.

2. LITERATURE REVIEW, RESEARCH MODEL AND HYPOTHESES

There is strong relationship between service quality and customer satisfaction, which results an increase in post-purchase intention. Understanding the effect of the antecedents of customers' satisfaction has been under academic debates since past few decades. Customer always demand good quality products and services in a reasonable and affordable prices (Peng Leong, 2006). Various empirical studies for the analysis of significant factors related to customer satisfaction, service quality, post-purchase intention, customer loyalty, customer retention has been done in different contextual settings (Kuo et al. 2009; Wang et al. 2004; Deng et al. 2010; Mittal et al., 1998; Basheer et al., 2012; Ghobadian et al., 1994; Wang & Lo, 2002; Lee et al., 2006). User satisfaction is quite important key indicator of success for organizations (Alzabi, T., & Shah, A. 2016). Researchers projected many ways of measuring the quality of the service. One famous model for measuring the service quality proposed and named by Parasuraman et al., (1988) is the SERVQUAL model. This model measures the variation among customers expectation perception on the basis of five factors i.e reliability, tangibles, assurance, empathy and responsiveness. Later on Grönroos (2004), presented seven factors to perceive the service quality including: behavior and attitude of employees, skills and professionalism, flexibility and convenience, recovery of services, trustworthiness and reliability, credibility and reputation, and scope of service. In another research study conducted by Wang & Lo (2002) on the service quality, customer satisfaction and behavior intention of customer in china's telecommunication industry prospectus. Researchers found that service provider's responsiveness, quality of service and behavior towards customer are important factors involved the overall customer satisfaction in telecommunication industry. As per Deng et al., (2010), customer's trust and satisfaction has direct influence on the enhancement of customer loyalty in services sector. Furthermore, others (Anderson & Fornell 1994; Anderson 1993; Jones & Sue 2000) suggest that customer satisfaction can be helpful in maintaining customer base which would result in long-term customer loyalty. Measuring customer satisfaction for services is difficult as compared to products. Malik et al., (2012) argued that "it is easy to evaluate the quality of any product as compared to the service quality; because the services are composed of strange characteristics".

After a thorough investigation of existing literature it can be concluded that until now there has been no study conducted on the simultaneous investigation on the interrelationships between post-purchase intention, customer satisfaction, customer perceived value and perceived service quality. On the basis of literature review, this research posits a comprehensive theoretical framework which integrates service quality factors (stability, content quality), customer perceived values (functional value, monetary value), responsiveness, staff attitude and customer satisfaction to identify the antecedents of post-purchase intention in Pakistan's fixed-line telecom services sector. The conceptual model in Figure 1 shows the relationship of various factors involved in this research and their hypothesized relationships.

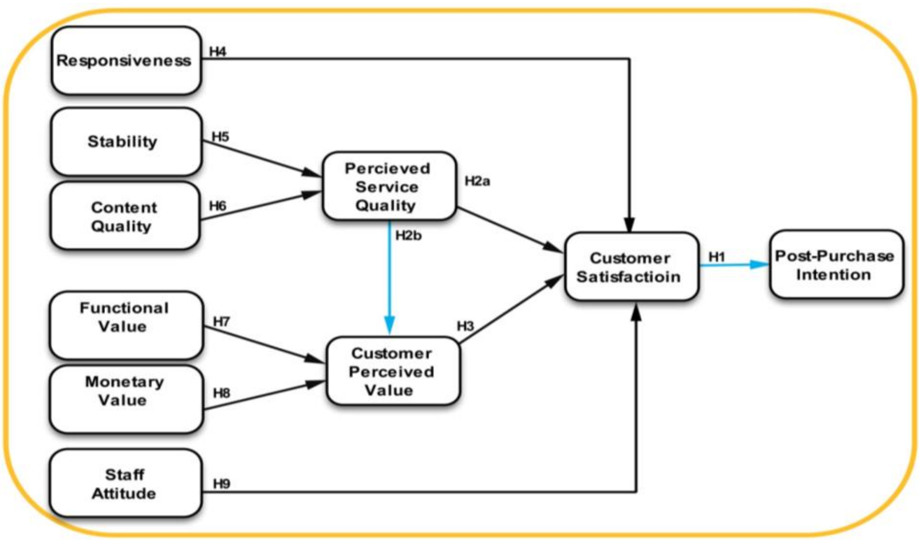


Figure 1: Research Model

2.1 Customer Satisfaction and Post-Purchase Intention

Customer satisfaction is considered as the most important determinant of customer's post-purchase intention (Deng et al., 2010). Satisfaction is defined as the buyer's state of being sufficiently satisfied in buying a particular service or product against his or her sacrifice made (Basheer et al., 2012). Since customer satisfaction is a measurement of the degree to which a customer feels positive in context of service provider, it is vital that service providers must understand what makes customer to be loyal (Mittal et al. 1998). Post-purchase intention is the intention of customer that he or she has tendency to repurchase goods of services from the same firm (Kuo et al., 2009). Post-purchase intention play major role both in customer retention and in winning new customers through the WOM (Word Of Mouth) (Bansal & Voyer 2000). It is obvious that better service quality can convince consumer to have a positive post-purchase intention of a service or product. Thus, on the basis of literature evidences it is hypothesized that:

H1 *Customer satisfaction has significant positive effect on post-purchase intention.*

2.2 Perceived Service Quality

In recent decades, many service firms have considered service quality as a differentiation tool. According to Seth et al. (2005), there is a significant impact service quality on customer satisfaction, business performance, business costs, customer profitability and loyalty. Perceived service quality is an outcome of the comparison of customers' expectations with perceived performance of services (Parasuraman, et al., 1988). Some researchers argue that service quality is a comparison of performance perceptions against the service expectations. The finding of the study of Parasuraman and Grewal (2000) indicate that perceived service quality enhances perceived customer value, which, in turn, enhances customer loyalty and the post-purchase intention. Similarly, Zeithaml and Bitner (2003) offered deeper definition of service quality and proposed that service quality can be treated as an agent that has direct effect on customer satisfaction, which causes the repetition of purchase behavior which concludes in the organization's long-term profitability. Thus we hypothesize that:

H2a *Perceived service quality has significant positive effect on customer satisfaction.*

H2b *Perceived service quality has significant positive effect on customer perceived value.*

2.3 Customer Perceived Value

Customers always expect more benefits in terms of value against what they pay. It is however a pertinent issue that, how and to what extent telecom service provider charge their customers in providing value. From customers' perceptive, value is of prime importance while calculating what is sacrificed and what is received. As per Babin and Darden (1994), the perceived value is a customer's view of the subjective worth of some action or item considering every single net advantage and expenses of utilization. Customer perceived value is not only related to what customers get, but also what he or she sacrifices (Wang et al., 2004). Perceived value can be measured with the help of two questions: overall quality given price and overall price given quality. McDougall and T (2000); Anderson et al., (1994); Ravald and C (1996); On the basis of provided literary evidence it can be hypothesized that:

H3 *Customer perceived value has significant positive effect on customer satisfaction.*

2.4 Responsiveness

Consumers are more likely to adhere to an organization that provide quick response to their issue as well as interested to take their feedback into consideration. Organization need to optimize their service and goods based on feedback and understand customer requirements in order to enhance customer satisfaction and post-purchase intention. Similarly, Rahaman & Rahman (2011) defines responsiveness as the willingness to help customers and provide prompt service. Wang & Lo, 2002; Lee et al., (2006); Kuo et al., (2009); Lai et al., (2009) found a significant impact of responsiveness on customer satisfaction. In context of our research, responsiveness reflects the extent to which customers' feedback is taken into consideration and the promptness fixed-line telecom service provider shows in responding the queries. Thus we hypothesize that:

H4 *Responsiveness has significant positive effect on customer satisfaction.*

2.5 Stability

Stability is one of important antecedents of perceived service quality. Garvin, D. (1988) has identified eight main attributes of service quality. One of them is the stability of system or a service. It is defined as the probability that the system, service or a product will operate as per its prescribed value over a specified time under stated conditions of use. System and network stability has direct impact on system quality. Kuo et al., (2009) and Martin (2004), highlighted system reliability by means of its reliability, error freeness, information richness. These researchers believe that there is significant impact of service stability of telecom services on service quality. Thus it is hypothesized that:

H5 *Stability has significant positive effect on perceived service quality.*

2.6 Content Quality

In telecom services content quality is an appealing feature and important antecedent of overall service quality. As per Garvin (1988), content quality can be defined as the customer's perception of service or a product quality and performance based on the provider's reputation. Kuo et al., (2009) defined content quality as a combination of various values related to product or service, like it is appropriate, complete, rich, regularly updated and easily understood and accessed. They also found that content quality is the predicator of service quality. Thus we hypothesize that:

H6 *Content quality has significant positive effect on perceived service quality.*

2.7 Functional Value

Functional value is a customer perception of any product or service. Deng et al., (2010) found that functional value is a multidimensional construct from the customer perspective, i.e customer realize that the service or product is reliable, it poses good functions, fulfils need and requirements, and very well managed as per demands. Deng et al. (2010) also concluded that the functional value of a service or a product have significant effect on customer value. If users' functional value is satisfied, they will experience more satisfaction towards the product or services. It is therefore hypothesized that:

H7 *Functional value has significant positive effect on customer perceived value.*

2.8 Monetary Value

Monetary value is the customer's perception about the value of any product or a service regarding its overall value and the financial cost is paid for it. It is customers feeling that the price is reasonable, economical and comparatively good (Deng et al. 2010). According to Kuo et al. (2009), customer's monetary or economic perspective for the product or service is the amount he is willing to pay for a service or product and the amount practically he paid. Thus it is hypothesized that:

H8 *Monetary value has significant positive effect on customer perceived value.*

2.9 Staff Attitude

Organizational behavior is usually reflected by the attitude of the front staff which has direct interaction with customers. Staff attitude is an important pillar of marketing and service sector organizations. Chase and Bowen (1991), argue that when service delivery occurs between service provider employees and the customers, attitudes and behaviors of seller can influence customers' perceptions. Staff attitude play an important role, as customers' perceptions of service or product quality and its value has influence on customer satisfaction, which further effects customer's purchase intentions (Lee et al. 2006). Johnston (2005) see staff attitude on greater scale including courtesy showing the politeness, respect and propriety of staff, flexibility to listen customers, friendliness with customers. Thus we hypothesize that:

H9 *Staff attitude has significant positive effect on customer satisfaction.*

3. RESEARCH METHODOLOGY

3.1 Sample and Data Collection

Researchers intend to employ quantitative method using cross-sectional questionnaire survey technique to collect data for the proposed model for post-purchase usage intention of fixed-line telecom services. This methodology is in confirmation with the prior studies conducted in telecommunication, mobile value-added services and information systems acceptance domain (Lai et al. 2009; Kuo et al. 2009; (Chandio 2011); Fida Hussain Chandio et al. 2013). According to (PTA, 2014) recent report, there are approximately

3.172 million fixed-line telecom services' users in Pakistan. Users of all five Fixed-line telecom service providers (PTCL, NTC, Brain Telecom Limited, World Call, Union Telecom, and Naya Tel) will be selected used as target population. The inclusion criterion for the participants would be those respondents who has been using the fixed-line telecom service or have used these services in past.

3.2 Data Analysis

Structural Equation Modeling (SEM) technique will be used to assess the proposed model. Hair et al. (2006), suggest that SEM as very helpful tool for researchers in measuring the relationships among multiple variables simultaneously. As per Hair et al. (2006); and Tabachnick & Fidell, (2001) SEM is a powerful statistical technique with rigorous procedures that helps to deal with complex research models. Different statistical procedures will be applied to deal with missing values, outliers and normality issues. Data will be tested for validity and reliability of every item using Confirmatory Factor Analysis (FA). Two–step approach will be adopted to test the proposed model and the hypothesized relationships. In first step CFA will be run to diagnose the proposed model with SEM based Analysis of Moment Structures (AMOS) software version 23.0. In second step the specification of structural model based on dependent and independent variables will be conducted to test the hypotheses.

4. COMMENTS AND CONCLUSION

The foremost aim of this research study was to propose a framework that can help us examine the post-purchase intention behavior of fixed-line telecom users and to understand the relative strength of antecedents that play role in customer satisfaction. This research paper provides a comprehensive integrated framework to understand the dynamic relationship among various dimensions of fixed-line telecom services in Pakistan. Proposed model is still in theoretical phase and will be evaluated in future research. Synopsis of crucial factors identified by this research would offer theoretical contribution and understanding of customers post purchase intention towards their existing service providers. Additionally, findings of this study may be very helpful for fixed-line telecom service providers in developing economies to find practical insights and optimize customer satisfaction and retention practices to survive in the competitive market environment.

REFERENCES

1. Ali, J.F., Ali, I., Rehman, K., Yilmaz, A.K., Safwan, N. and Afzal, H. (2010). Determinants of consumer retention in cellular industry of Pakistan. *African Journal of Business Management*, 4(12), 2402-2408.
2. Anderson E.M. (1993). The antecedents and consequences of customer satisfaction for firms. *Mark. Sci.*, 12(2), 125-143.
3. Anderson, E.W., Fornell, C. and Lehmann, D.R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *The Journal of Marketing*, 58, 53-66.

4. Babin, B.J., Darden, W.R. and Griffin, M. (1994). Work and/or fun: measuring hedonic and utilitarian shopping value. *Journal of Consumer Research*, 20(4), 644-656.
5. Bansal, H.S. and Voyer, P.A. (2000). Word-of-mouth processes within a services purchase decision context. *Journal of Service Research*, (3(2), 166-177.
6. Awad, B.A.A.A.G. (2012). The relationships between service quality, satisfaction, and behavioral intentions of Malaysian spa center customers. *International Journal of Business and Social Science*, 3(1), 198-205.
7. Chandio, F.H. (2011). *Studying acceptance of online banking information system: A structural equation model*. (Doctoral dissertation, Brunel University Brunel Business School PhD Theses).
8. Chase, R.B. and Bowen, D.E. (1991). Service quality and the service delivery system: A diagnostic framework. In *Service Quality: Multidisciplinary and Multinational perspectives*, SW Brown, E. Gummesson, B. Edvardsson and B. Gustavsson (editors), Lexington Books, New York, 157-78.
9. Deng, Z., Lu, Y., Wei, K.K. and Zhang, J. (2010). Understanding customer satisfaction and loyalty: An empirical study of mobile instant messages in China. *International journal of information management*, 30(4), 289-300.
10. Chandio, F.H., Abbasi, M.S., Nizamani, H.A. and Nizamani, Q.U.A. (2013). Online banking information systems acceptance: a structural equation modelling analysis. *International Journal of Business Information Systems*, 12(2), 177-193.
11. Fornell, C. (1992). A national satisfaction barometer: the Swedish experience. *Journal of Marketing*, 56, 6-21.
12. Fornell, C., Johnson, M.D., Anderson, E.W., Cha, J. and Bryant, B.E. (1996). The American customer satisfaction index: nature, purpose, and findings. *The Journal of Marketing*, 60(4), 7-18.
13. Garvin, D.A. (1988). *Managing quality: The strategic and competitive edge*. The free press, New York.
14. Ghobadian, R., Speller, S. and Jones, W. (1994). Service quality concepts and models. *International Journal of Quality Management*, 11, 43-66.
15. Gronroos, C. (1996). Relationship Marketing: Strategic and Tactical Implications. *Management Decision*, 34(3), 5-14.
16. Hair, J.F., Jr., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. (2006). *Multivariate Data Analysis*, 6th ed., New Jersey. Pearson Prentice Hall, Pearson Education, Inc., Upper Saddle River.
17. Jones, M.A. and Sue, J.J. (2000). Transaction-specific satisfaction and overall satisfaction: an empirical analysis. *Serv. Mark.*, 14, 147-159.
18. Kuo, Y., Wu, C. and Deng, W. (2009). The relationships among service quality, perceived value, customer satisfaction, and post-purchase intention in mobile value-added services. *Computers in Human Behavior*, 25(4), 887-896.
19. Lai, F., Grif, M. and Babin, B.J. (2009). How quality, value, image, and satisfaction create loyalty at a Chinese telecom. *Journal of Business Research*, 62, 980-986.
20. Lee, H., Lee, Y. and Yoo, D. (2000). The determinants of perceived service quality and its relationship with satisfaction. *Journal of services marketing*, 14(3), 217-231.

21. Malik, M.E., Ghafoor, M.M. and Iqbal, H.K. (2012). Impact of brand Image, service quality and price on customer satisfaction in Pakistan telecommunication sector. *International Journal of Business and Social Science*, 3(23), 123-130.
22. O'Neill, M. and Palmer, A. (2004). Cognitive dissonance and the stability of service quality perceptions. *Journal of Services Marketing*, 18(6), 433-449.
23. McDougall, G. and Levesque, T. (2000). Customer satisfaction with services: putting perceived value into the equation. *Journal of Services Marketing*, 14(5), 392-410.
24. McQuitty, S., Finn, A. and Wiley, J.B. (2000). Systematically varying consumer satisfaction and its implications for product choice. *Academy of Marketing Science Review*, 2000, 1.
25. Mittal, V., Ross, W. and Baldasare, P. (1998). The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intention. *Journal of Marketing*, 61(1), 33-47.
26. MOITT (2003). *De-Regulation Policy for the telecom industry*, Available at: http://www.moitt.gov.pk. [Accessed: 25-Aug-2015].
27. MOITT (2014). *MoIT Broadband Report*. Available at: http://www.moitt.gov.pk. [Accessed: 25-Aug-2015].
28. Oliver, R.L. (1993). Cognitive, affective and attribute bases of the satisfaction response. *Journal of Consumer Research*, 20, 418-430.
29. Parasuraman, A., Zeithamal, A.V. and Berry, L.L. (1988). Servqual: Amultiple-item Scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 12-40.
30. Parasuraman, A. and Grewal, D. (2000). The impact of technology on the quality-value-loyalty chain: A research agenda. *Journal of Academic of Marketing Science*, 28, 168-174.
31. Peng, L.Y. and Wang, Q. (2006). Impact of relationship marketing tactics (RMTs) on switchers and stayers in a competitive service industry. *Journal of Marketing Management*, 22(1-2), 25-59.
32. PTA (2014). Annual cellular Mobile Teledensity in Pakistan, May, 2014. Annual cellular Mobile Teledensity. Available at: http://www.pta.gov.pk/index.php?option =com_content&task=view&id=269&Itemid=658. [Accessed: 25-Aug-2015].
33. Rahaman, M.M., Abdullah, M. and Rahman, A. (2011). Measuring service quality using SERVQUAL model: A study on PCBs (Private Commercial Banks) in Bangladesh. Business Management Dynamics, 1(1), 1-11.
34. Ravald, A. and Gronroos, C. (1996). The value concept and relationship marketing. *European Journal of Marketing*, 30(2), 19-30.
35. Seth, N., Deshmukh, S.G. and Vrat, P. (2005). Service quality models: a review. *International Journal of Quality & Reliability Management*, 22(9), 913-49.
36. Tabachnick, B. and Fidell, L. (2001). *Using Multivariate Statistics*, Allyn and Bacon, Boston.
37. Alzabi, T.H. and Shah, A. (2006). The different perceptions of factors that influence user satisfaction among mobile communication subscribers: Study in Kuwait. *ARPN Journal of Engineering and Applied Sciences*, 10(23), 17698-17704.

38. Wang, Y. and Lo, H.P. (2002). Service quality, customer satisfaction and behavior intentions: Evidence from China's Telecommunication Industry. *Info*, 4(6), 50-60.
39. Wang, Y., Lo, H.P. and Yang, Y. (2004). An Integrated Framework for Service Quality, Customer Value, Satisfaction: Evidence from China's Telecommunication Industry. *Information Systems Frontiers*, 6(4), 325-340.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 479-486

CONCEPT OF DRUG DISPOSAL AMONG KARACHIITES

Rabia Shahid, Alishba Ahmed, Sana Islam, Faksheena Anjum, Omair Anwar and **Sana Ghayas**
Dow College of Pharmacy, Dow University of Health Sciences

ABSTRACT:

Objective: To study drug disposal practices among the Karachiites.

Methodology: This survey was conducted from August to December, 2015 using anespecially designed Performa which were filled by the respondents in the presence of the researchers. These were then collected by the researchers for analysis using descriptive statistics.

Result: Out of 350 Performa, 300 were filled completely by the participants (Response Rate=86%). It was found that 58% persons had no unused or expired medicines at home and approx. 51% kept medicines at home until expired. It was mentioned by 56% respondents that they kept unused medicines in case they needed them later. Most of the respondents (45%) disposed their medicines with household trash. It was found that 73% respondents had never been educated about medicines disposal and according to about 24% respondents, Pakistan Pharmacy Council is responsible to create awareness about proper medicines disposal. More than 40% persons believed that not taking or finishing medication course as directed by the physician is a strong reason for the increase in expired medicines. About 66% respondents thought that medicines are hazardous to environment if not disposed properly. Around 88% respondents agreed to provide information regarding availability of a secured disposal location to others.

Conclusion: Awareness about drug disposal to general public is necessary since drugs are potentially hazardous substances. Proper and secure disposal location should be made available in order to ensure pollution free environment.

KEY WORDS

Drug disposal, Pollution free environment, Unused or expired medicines, Disposal location.

1. INTRODUCTION

Medication disposal has always been an ignored topic to create awareness among the general public. Karachi is one of the biggest industrial city of Pakistan and has population of about 16 million. Accordingly large amount of waste is produced in the city which also comprises of unused and expired medicines which is mostly dumped into landfills. Waste disposing into sewerage goes into sea and out of the disposed wasted only 20% is treated while rest is untreated (Ahsaan A, Nousheen M, Tariq M, 2013). Pharmaceuticals enter the environment as they are disposed of as solid waste into landfills or through sewerage into water system. Throughout the world the use of pharmaceuticals is increasing with time and because of these pharmaceuticals, relatively newly recognized pollutants can become a threat to the environment in future (Daughton C.G. 2008). Pharmaceuticals can find their way in to the environment through the excretion of urine and feces containing drugs (un-metabolized as well as bioactive metabolite), topical applications (release from the skin during washing or bathing) drug manufacturing process, animal excretion and disposal of expired, unused or unwanted drugs into the sewerage or trash (Daughton CG 2007; Gielen et al., 2009; Sarmah et al., 2008; Chee-Sanford, J.C., et al. 2001; Boxall, A 2008.) With globally increased use of pharmaceuticals has lead to enhanced international awareness of potential detrimental effects on the environment from disposal of these compounds either to landfills and aquatic environment (Md Abul Kalam, Md Reza-ul-Haque, Md Akhtar-uz-zaman, 2012). Safe disposal programs help prevent diversion in an environmentally friendly manner, however, due to how our bodies metabolize the medicine we take, some trace amounts of pharmaceuticals may still be found in surface waters. When a drug passes through the body without being fully metabolized, trace amounts of the medication enter the wastewater treatment system and find their way into surface waters. Researches and surveys have increased on practice of medication disposal among public. There have been recent reports of pharmaceutical compounds and metabolites in the environment (Alfred Y.C, Barrie M, Rhiannon B, 2011). Many studies have shown presence of trace amount of drug in water system and long term existence may produce toxic effects (Ahsaan A, Nousheen M, Tariq M, 2013). Presence of antibiotic in water can produce antibiotic resistance as presence of antibiotics have effect on the bacteria present in the water ways and even trace levels of ethinyl estradiol (the active compound of oral contraceptive) found in waterways has been shown to impair sexual development and feminization of fish (Md Abul Kalam, Md Reza-ul-Haque, Md Akhtar-uz-Zaman, 2012) (Ahsaan A, Nousheen M, Tariq M, 2013). Improperly dispose of medications potentially pose a significant environmental risk and storage of unwanted or unused medication in the household provides an increased risk of accidental childhood poisonings.

This survey was conducted to check the awareness of people on disposal of medications and their environmental implications because the basic step to protect the environment from pharmaceuticals is by creating awareness proper disposal methods.

2. METHODOLOGY

Study Design: This was a descriptive cross-sectional survey based on a structured questionnaire format.

Study Population: The study populations include adults from different areas of Karachi. This study period was conducted from August to December 2015. The study sample comprised of 300 conveniently selected individuals aged between 15-80 years.

Survey Instrument: The questionnaire was placed by the interviewer to the individual respondents and collected question-sheet with their answers. The questionnaire was divided into three sections. The first section asked for demographic data including age, gender, ethnic group and educational level. In the second section, questions about various type of medications waste were listed namely solid, liquid and semi-solid medications. In the third section, the individuals were asked about the awareness regarding the effect of pharmaceuticals on environment.

Statistical Analysis: Data were entered into the Statistical Package for Social Sciences (SPSS, version 16.0) and descriptive analysis conducted. Demographic characteristics and knowledge and attitude scores were summarized using descriptive statistics. Correlation between respondents' education level and their knowledge was performed using Chi-square test. The results were reported as percentage response. In all statistical analyses, a p-value of < 0.05 was considered to be statistically significant.

3. RESULT AND DISCUSSION

Out of two hundred Performa, 158 were filled completely by the participants (response rate=79%). Most of the respondents were females (66.7) aged between 21-25 (38.3%) and undergraduate (34.3%) (Table 2). However, education level contributed significantly to the statement regarding drug disposal practice. Age and gender did not relate significantly with the respondents knowledge on drug disposal. Education level and drug disposal were correlated but result didn't show any positive result however irregularity were seen among the result and it was concluded that the most wrong practice was seen among post grad students and least was seen in primary students when asked about to rinse the drug down the toilet, the most wrong practice was seen among primary students and least was seen in intermediate students when asked about to flush the drug down the toilet, the most wrong practice was seen among secondary students and least was seen in primary students when asked about throw the drugs in trash and most of the primary students said that they will return drugs to pharmacy and least result was seen among under grad students (Figure 1). When result was observed between people ho were previously educated about proper drug disposal to people who weren't, positive result was obtained however not a very vast difference but people who were educated followed the right way to dispose the drug as compare who weren't educated. Most of the respondents claimed that the reason of excessive drugs stored is not taking or finishing the course as directed by physician (Figure 3). Most of the respondents said that Pakistan Pharmacy Council is responsible for spreading awareness for proper drug disposal (Figure 4).

Irregular result was observed when people were asked about medication disposal but most of them agreed that disposing drugs through trash is acceptable as compare to disposing them down the sink, flush or returning back to pharmacy (Figure 5). Most of the people said that they keep unused medicines in case if they are needed later (Figure 6).

4. DISCUSSION

Karachi being a metropolitan city faces many problems related to sanitation and health. Large population demands organized solid waste management this is because the management in the past few decades had been shifted to different organizations and still number of institutes and organizations are involved (government and non-government) making the process complicated and slow. The city has two landfill areas which can accommodate the waste produced on daily basis but hardly 40 % waste can reach to those sites while the rest remained dumped on their primary waste sites in the city and may undergo degradation and leachate to the ground surface water. Drugs reportedly, consumed completely going into the environment through excretion are also threatening the environmental safety. Only 30% of the sewerage water is treated and remaining goes into the sea through *nallahs and naddi. Modern treatment plants are incapable to remove the pharmaceuticals and their metabolites showing their presence in water thereby affecting the aquatic life. (Daughton. C.G. and Ternes; 1999).* In Pakistan 80% vegetables and fruits are irrigated through sewerage water (John et al, 2005 and Khuda B. and Sarfaraz H. 2006) which may contaminates them and harm the population eating those vegetables or fruits. Shepherds also bring their animals for grazing grass, small herbs or vegetation growing around sewerage water and in some portions of the dumping sites. These may lead to harm due to potential entry into the food chain. With time, population will increase with more utilization of pharmaceuticals which will become the part of environment through the waste.

Table 1
Guidelines for Proper Drug Disposal by FDA

Follow any specific disposal instructions on the prescription drug labeling or patient information that accompanies the medicine. Do not flush medicines down the sink or toilet unless this information specifically instructs you to do so.	• Take advantage of programs that allow the public to take unused drugs to a central location for proper disposal. Call your local law enforcement agencies to see if they sponsor medicine take-back programs in your community. Contact your city's or county government's household trash and recycling service to learn about medication disposal options and guidelines for your area. • Transfer unused medicines to collectors registered with the Drug Enforcement Administration (DEA). Authorized sites may be retail, hospital or clinic pharmacies, and law enforcement locations.
If no disposal instructions are given on the prescription drug labeling and no take-back program is available in your area, throw the drugs in the household trash following these steps:	1. Remove them from their original containers and mix them with an undesirable substance, such as used coffee grounds, dirt or kitty litter (this makes the drug less appealing to children and pets, and unrecognizable to people who may intentionally go through the trash seeking drugs). 2. Place the mixture in a sealable bag, empty can or other container to prevent the drug from leaking or breaking out of a garbage bag.

Table 2
Demographic Data

GENDER	Male	30.7%
	Female	66.7%
AGE	15-20	25.3%
	21-25	38.3%
	26-30	8%
	31-35	3%
	36-40	3.3%
	41-45	3%
	46-50	2.7%
	50 onwards	3%
EDUCATION LEVEL	Primary	1.3%
	Secondary	9.7%
	Intermediate	12.7%
	Under Grad	34.3%
	Grad	15.3%
	Post Grad	12 %

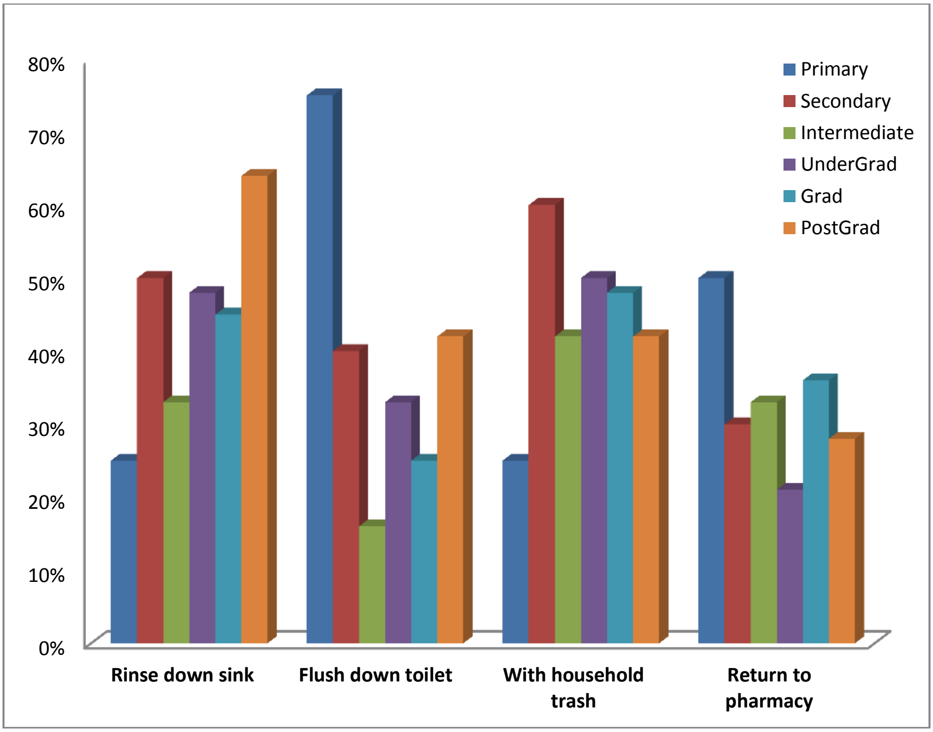


Figure 1: Correlation of Education Level to Drug Disposal Practices

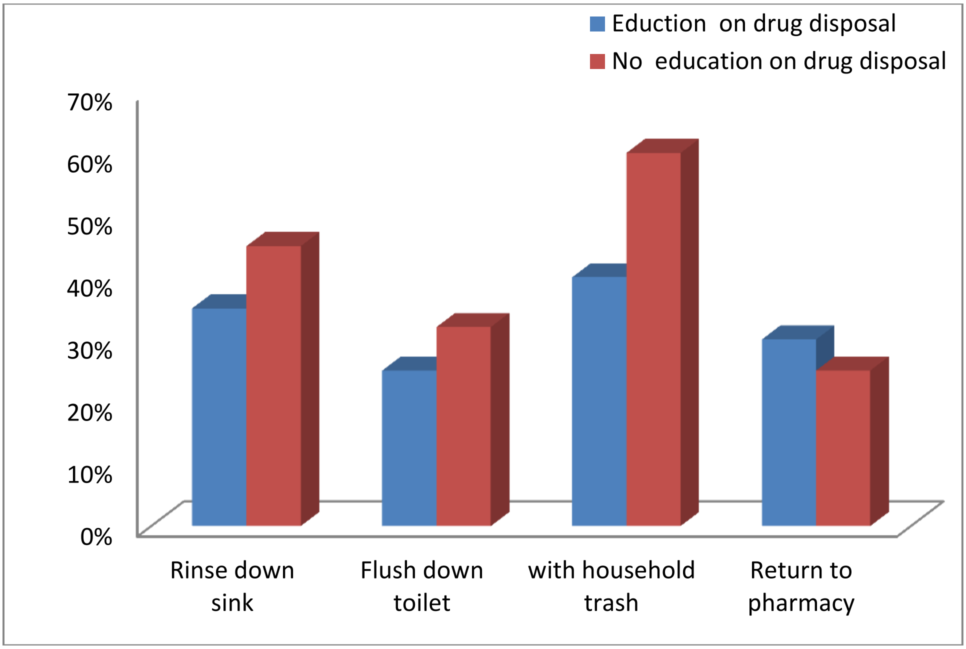


Figure 2: Correlation of Education on Disposal to Drug Disposal Practices

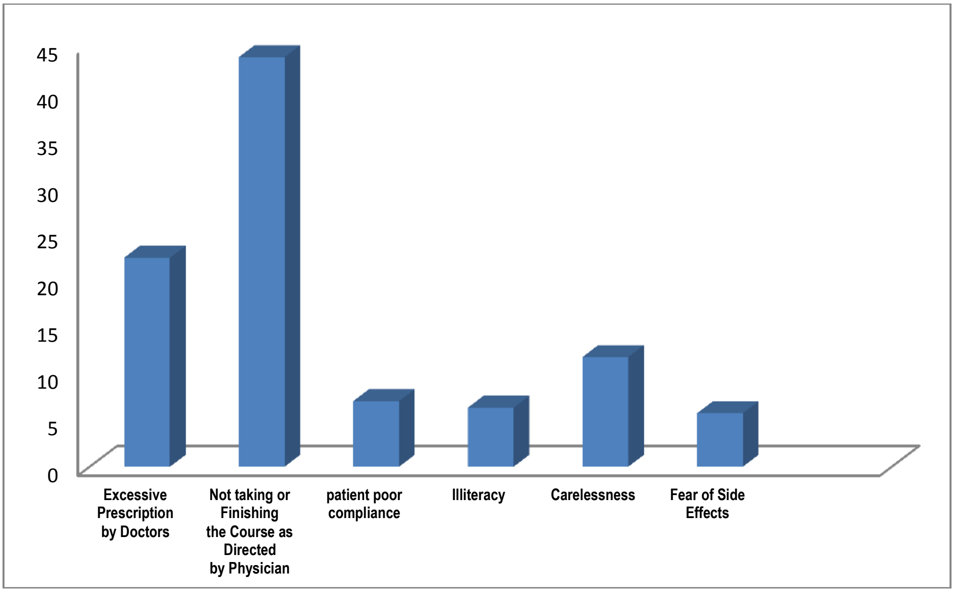


Figure 3: Reason behind increase in expired medicines

Table 3

Questions	Yes	No	Don't Know
There must be official guidelines for drug disposal	85.2%	5.3%	2%
Pharmacies take back unwanted/expired medicines	37.9%	31%	30.8%
Medications are hazardous to environment if not disposed properly	72.2%	8.3%	18.3%
Provision of secure disposal location	88.8%	1.2%	10.1%
educated about proper drug disposal	14.7%	72.7%	11%
Unused/expired medicines at home	58%	38.7%	1%

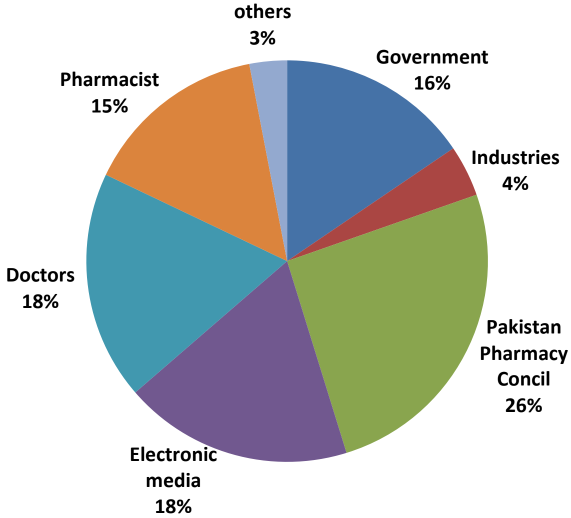


Figure 4: Awareness for Proper Drug Disposal

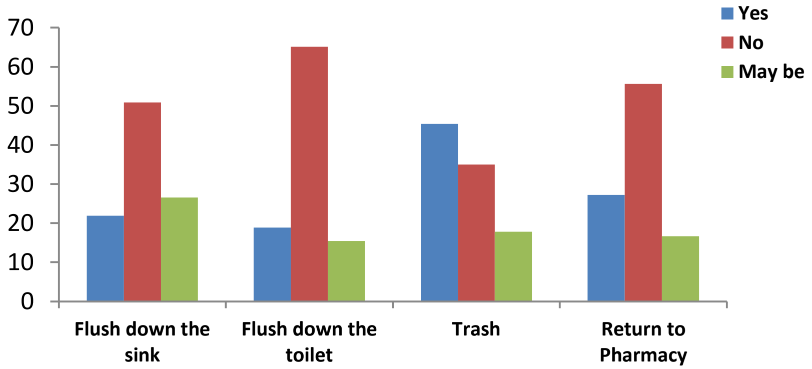


Figure 5: Medication Disposal Practice

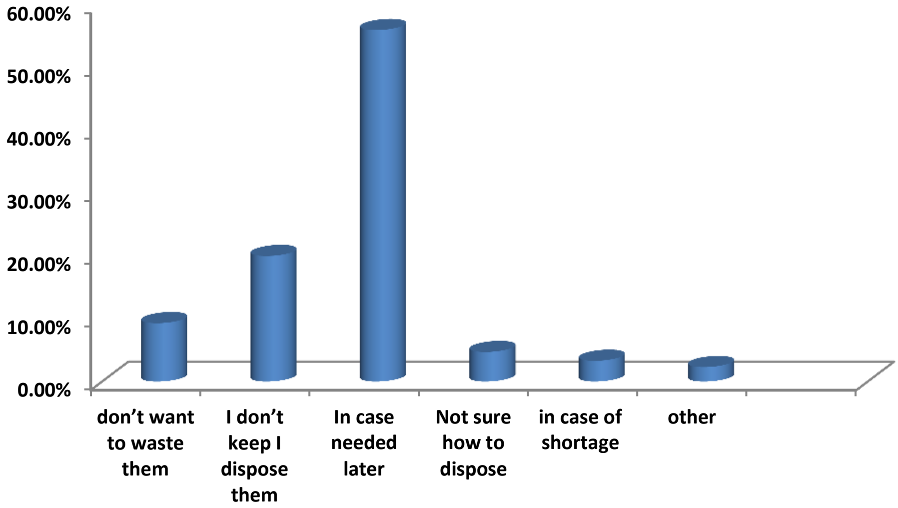


Figure 6: Keeping Unused Medicines

REFERENCES

1. Ahmed, A., Mushtaq, N., Tariq, M., Durrani, M., Arif, M. and Yasmeen, G. (2007). Disposal practices of unused and expired pharmaceuticals in Karachi and their impact on health and environment. *Journal of University Medical and Dental College*, 4(2), 42-48.
2. Azad, M.A.K., Ansary, M.R.H., Akhter, M.A., Al-Mamun, S.M. and Uddin, M. (2012). Disposal Practice for Unused Medications amongthe Students of the International Islamic University Malaysia. *Journal of Applied Pharmaceutical Science*, 2(7), 11.
3. Glassmeyer, S.T., Hinchey, E.K., Boehme, S.E., Daughton, C.G., Ruhoy, I.S., Conerly, O., Daniels, R.L., Lauer, L., McCarthy, M., Nettesheim, T.G. and Sykes, K. (2009). Disposal practices for unwanted residential medications in the United States. *Environment International*, 35(3), 566-572.
4. Seehusen, D.A. and Edwards, J. (2006). Patient practices and beliefs concerning disposal of medications. *The Journal of the American Board of Family Medicine*, 19(6), 542-547.
5. Tong, A.Y., Peake, B.M. and Braund, R. (2011). Disposal practices for unused medications around the world. *Environment International*, 37(1), 292-298.
6. Wilson, T.N., Weiss, L.B., Malone, J.O. and Garnier, K. (2011). Physician knowledge and perception of the need for drug disposal guidelines. *Osteopathic Family Physician*, 3(2), 48-52.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 487-491

CLASSIFICATION OF POTATO (SOLANUM TUBEROSUM L.) GENOTYPES BY USING CLUSTER ANALYSIS

Irum Raza[1§], Sobia Naheed[1], Muhammad Zubair Anwar[1], M. Asif Masood[1] and **Shazia Erum[2]**

[1] Social Sciences Research Institute, National Agricultural Research Centre, Islamabad, Pakistan,

[2] Plant Genetics Research Institute, National Agricultural Research Centre, Islamabad, Pakistan,

[§] Corresponding author Email: irumraza83@gmail.com

ABSTRACT

In the present study cluster analysis by average linkage method was used for classification of potato genotypes on the basis of phenotypic traits. Cluster analysis is a multivariate technique used to group the data into clusters in a way that the observations within in a cluster are similar whereas the clusters are distinct from one another. Data on thirty five potato genotypes were evaluated in the field experiment for 3 qualitative traits namely number of eyes/tuber, tuber shape and tuber skin type. The amalgamation steps in the cluster analysis showed that the similarity level decreased by increases of about 6 or less until it decreased by 13 at the step from four clusters to three. This shows that four clusters are logically sufficient for final partition. The resulting dendrogram displayed the information in the form of a tree diagram. In the present data, potato genotypes such as CIP01, CIP25, CIP2, CIP4, CIP10, CIP24, CIP32, CIP 09, CIP20, CIP27, CIP29, CIP33, ASTRIX, DESIRRE, ZINA Red, HZD2 1499, CIP 07, CIP16 and CIP30 made up the first cluster, CIP12, CIP14, CIP28, CIP31and CIP17 made up the second cluster, CIP03, CIP06, CIP08, CIP11, CIP15, CIP05, CIP13, CIP18, CIP19, CIP34 made up the third cluster and CIP22 made up the fourth cluster. These genotypes are grouped into clusters based on the similar phenotypic characteristics that is number of eyes/tuber, tuber shape and skin type are associated in each of the distinct cluster. This classification of genotypes with the help of cluster analysis could assist breeders to select and identify genotypes with desirable characteristics for inclusion in variety breeding programs.

KEYWORDS

Potato, Genotypes, Phenotypic traits, Cluster Analysis, Dendrogram.

1. INTRODUCTION

Potato (*Solamum tuberosum* L.) is an annual crop from family Solanaceae and ranked fourth in the world after wheat, rice and maize. Potato, is an annual cool season crop, which need most favorable temperature ranging from 16ºC to 20ºC for the excellent growth of tubers (Mathur, 2003). Pakistan is the seventh largest potato producing country in the world. The production of potatoes during the year 2014-15 increased by 6.3 percent (GoP, 2014).

Classification of crop genotypes with improved characteristics is important for the breeding program. This results in better attributes for food quality and marketability (Chirimi et al., 1999). Multivariate statistical techniques such as cluster analysis have been widely used in agriculture for the classification of improved varieties. Categorization or classification of genotypes on the basis of phenotypic traits have been made by my studies such as (Bos et al., 2000; Kaplan, 2001; Lacroix C, et al., 2005; Li et al., 2009; K'Opondo, 2011).

Cluster is defined as a collection of data objects similar to one another within the same cluster and dissimilar to the objects in other clusters. Cluster analysis is a multivariate technique used to find similarities between data according to the characteristics found in the data and grouping similar data objects into clusters (Hair,J F., et al., 2006)

Cluster analysis has been applied in many fields of interest such as in agriculture for the classification of varieties on the basis of similar traits or characteristics, marketing to identify persons with similar buying habits, biology for identification of diseases and their stages and in land use for identification of areas of similar land use in an earth observation database et (Arabie, P et al. 1996).

Mainly there are two types of cluster analysis: hierarchical Clustering Technique that merge those two units that are most similar and repeat the process till all are merged to form one entity and another is non-hierarchical Clustering Technique in which data are divided into K partitions or groups with each partition representing a cluster. Methods for computing distance between clusters comprise of single linkage: that works on the principal of smallest distance or nearest neighbor, complete linkage**:** It works on the principle of distant neighbor or dissimilarities, average linkage**:** This works on the principle of average distance, centroid: This method assigns each item to the cluster having nearest centroid (means) Ward's: It forms clusters by maximizing within cluster homogeneity (Kaufman, L., & Rousseeuw, P. J. 2009).

Keeping in view the significance of cluster analysis technique the present study was designed on the demand of potato program of NARC for classifying the high yielding potato genotypes based on an appropriate statistical technique. Therefore cluster analysis was used for the analysis and classification of the genotypes. Therefore the prime goal of the study was to classify potato genotypes on the basis of their qualitative characteristics such as number of eyes/tuber, tuber shape and tuber skin type.

2. MATERIAL AND METHODS

Data on thirty five potato genotypes for 3 qualitative traits namely number of eyes/tuber, tuber shape and tuber skin type were collected from the field experiment of Potato program NARC. Cluster analysis by average linkage method was used for classification of potato genotypes on the basis of qualitative traits (Hardiman, R. T., R. Lacey, and Yang Mu Yi, 1992). MINITAB software version 16 was used for analysis of data. In average linkage, the distance between two clusters is the average distance between an observation in one cluster and an observation in the other cluster. In terms of the distance matrix,

$$d_{mj} = N_k d_{kj} + N_l d_{lj} \,/\, N_m$$

where, N_k, N_l, and N_m are the number of observations in clusters k, l, and m.

3. RESULTS AND DISCUSSION

The results of the cluster analysis showed that the similarity level decreased by increases of about 6 or less until it decreased by 13 at the step from four clusters to three. This shows that four clusters are logically sufficient for final partition. The resulting dendrogram (Figure 1) displayed the information in the form of a tree diagram. In the present data, potato genotypes such as CIP01, CIP25, CIP2, CIP4, CIP10, CIP24, CIP32, CIP 09, CIP20, CIP27, CIP29, CIP33, ASTRIX, DESIRRE, ZINA Red, HZD2 1499, CIP 07, CIP16 and CIP30 made up the first cluster. CIP12, CIP14, CIP28, CIP31and CIP17 made up the second cluster. CIP03, CIP06, CIP08, CIP11, CIP15, CIP05, CIP13, CIP18, CIP19, CIP34 made up the third cluster. CIP22 made up the fourth cluster. These genotypes are grouped into clusters based on the similar qualitative characteristics that is number of eyes/tuber, tuber shape and skin type are associated in each of the distinct cluster (Table 1). This classification of genotypes with the help of cluster analysis assisted breeders to select and identify genotypes with desirable characteristics for inclusion in variety breeding programs. These results differ from another study by (Tairo, F et al., 2008) in which morphological and agronomical characterization of sweet potato germplasm has been made on the basis of cluster analysis. Cluster analysis classified accessions into two major groups with dissimilarity between them and association within the observations of the groups.

Table 1
Distribution of 35 Potato Genotypes into 4 Clusters

Clusters	Number of Observations	Within Cluster Sum of Squares	Average Distance from Centroid	Maximum Distance from Centroid
1	19	12.842	0.747	1.672
2	10	6.000	0.725	1.414
3	5	4.000	0.826	1.341
4	1	0.000	0.000	0.000

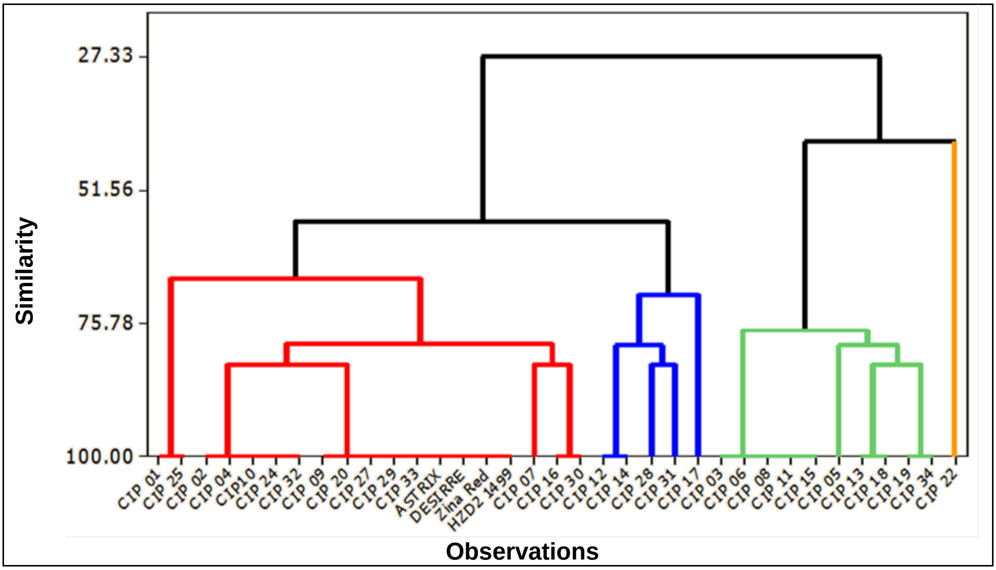


Figure 1: Dendrogram showing Classification of Potato Genotypes by Average Linkage Method

4. CONCLUSION AND RECOMMENDATION

Cluster analysis technique has proven to be functional in the classification of 35 potato genotypes into 4 clusters. This technique helped the breeders to select those genotypes with genetic similarity within the clusters and dissimilarity between the clusters. This technique will also be helpful for the selection of genotypes of different crops.

REFERENCES

1. Arabie, P., Hubert, L.J. and De Soete, G. (Eds.). (1996). *Clustering and classification*. World Scientific.
2. Bos, H.J., Vos, J. and Strulk, P.C. (2000). Morphological analysis of plant density effects on early leaf area growth in maize. *Netherlands J. Agric. Sci*., 48, 199-212.
3. Chirimi, B., Kapinga, R., Kanju, E., Rugutu, C. and Caley, T. (1999). On-farm evaluation of Sweet potato varieties in the Lake Zone of Tanzania. In Potatoes for Poverty Alleviation, *African Potato Association Conference Proceedings*. (Edited by Adipala, E; Nampala, P. and Osiru, M). 29 May-2 June 2000, Kampala, Uganda. Vol. 5, 125-130.
4. Government of Pakistan (2014). Economic Survey of Pakistan, Ministry of Finance, Government of Pakistan, Islamabad.
5. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. (2006). *Multivariate data analysis* (Vol. 6). Upper Saddle River, NJ: Pearson Prentice Hall.

6. Hardiman, R.T., Lacey, R. and Yi, Y.M. (1990). Use of cluster analysis for identification and classification of farming systems in Qingyang County, Central North China. *Agricultural Systems*, 33(2), 115-125.
7. K'opondo, F.B. (2011). Morphological characterization of selected spider plant (Cleome gynandra L.) types from western Kenya. *Annals of Biological Research*, 2(2), 54-64.
8. Kaplan, D.R. (2001). The science of plant morphology: definition, history, and role in modern biology. *Am. J. Bot.*, 88(10), 1711-1741.
9. Kaufman, L. and Rousseeuw, P.J. (2009). *Finding groups in data: An introduction to cluster analysis* (Vol. 344). John Wiley & Sons.
10. Lacroix, C., Jeune, B. and Barabe, D. (2005). Encasement in plant morphology: An integrative approach from genes to organisms. *Can. J. Bot*., 83, 1207-1221.
11. Li, P., Wang, Y., Sun, X. and Han, J. (2009). Using microsatellite (SSR) and morphological markers to assess the genetic diversity of 12 falcata (Medicagosativaspp. falcata) populations from Eurasia. *Afr. J. Biotechnol*., 8(10), 2102-2108.
12. Mathur, A. (2003). *Studies on phosphorylation status of starch in potato tubers (Solanum tuberosum L.)*. MSc. Thesis, Department of Biotechnology and Environmental Sciences, Thapar Institute of Engineering and Technology, Patiala. pp.10-14
13. Minitab version 15 (2007). Statistical Data Analysis Software.
14. Tairo, F., Mneney, E. and Kullaya, A. (2008). Morphological and agronomical characterization of sweet potato [Ipomoea batatas (L.) Lam.] germplasm collection from Tanzania. *African Journal of Plant Science*, *2*(8), 77-85.

Proc. 14th International Conference on Statistical Sciences
Karachi, Pakistan – March 14-16, 2016, Vol. 29, pp. 493-517

INEQUALITY MEASURES AND CHARACTERIZATION OF "EXPONENTIATED MOMENT EXPONENTIAL DISTRIBUTION (EMED)"

Fiaz Ahmad Bhatti and **Munir Ahmad**
National College of Business Administration and Economics, Lahore, Pakistan.
Email: fiazahmad72@gmail.com; drmunir@nebae.edu.pk

ABSTRACT

In this paper, some new statistical properties of the "Exponentiated Moment Exponential distribution" are derived. Moments, negative moments, Central Moments and some other properties are computed. Different reliability models of proposed distribution are theoretically established. Uncertainty measures like information generating function, Renyi entropy, Q-entropy, Havrda and Chavrat entropy, and Tsallis entropy are calculated. Inequality Measures like Incomplete moments, mean deviation, Lorenz curve, Bonferroni curve, Zenga curve, Mean residual life function and Mean inactivity time are also studied. Exponentiated Moment Exponential distribution is characterized through conditional moments, hazard rate functions, reverse hazard rate functions, truncated first Mean moments.

KEYWORDS

Elasticity; Uncertainty; Inequality; Characterization.

1. INTRODUCTION

Moment distributions have many applications in Probability, Quality Control, Ecology, Medicine and Advance Sciences. Biased data are controlled by application of Weighted Distribution, when observations are recorded from a particular probabilistic model, the observations will not have same probabilistic model unless each observation has alike chance of being recorded. Rao (1965) established moment distributions from the intimation of Fisher (1934). Patil and Rao (1977) studied different weight functions for continuous and discrete models. Patil and Rao (1978) studied application of weighted distributions in sampling Problems. Sunoj and Maya (2006) assisted to repair a system with connections between weighted and original variables. If the random variable X has pdf $f(x)$ then the corresponding weighted distribution has pdf

$$f(x;\theta)=\frac{w(x)\,f_0(x;\theta)}{E\left[w(x)\right]} \tag{1.1}$$

where $w(x)$ is a non-negative weight function such that $E\left[w(x)\right]$ exists.

Dara and Ahmad (2012) suggested a distribution function of Moment Exponential distribution and developed some basic properties like Moments, Skewness, Kurtosis, Moment generating function and Hazard function. Hasnain and Ahmad (2013) proposed two Parameters "Exponentiated Moment Exponential distribution" and studied some properties. Hasnain et al. (2015), studied "on Exponentiated Moment Exponential (EME) distribution".

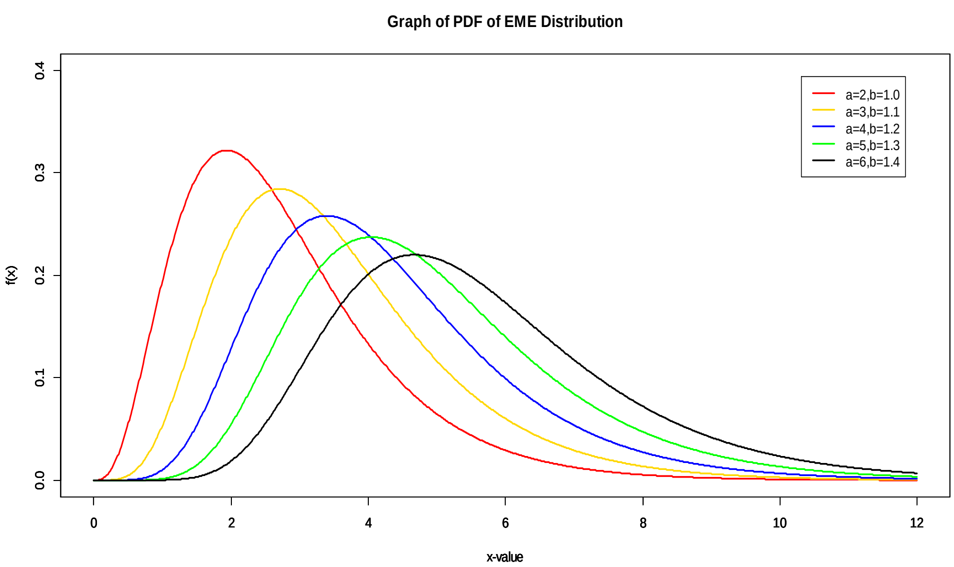
The article is written as follows, in section 2, some new properties of the Exponentiated Moment Exponential distribution are developed. Some plots for its cumulative distribution function (cdf), probability density function (pdf) and survival function are given. In Section 3, moments, negative moments, central moments and some other properties are computed. In section 4; Different reliability models of proposed Distribution are theoretically studied. In section 5, Uncertainty measures like information generating function, Renyi entropy, Q-entropy, Havrda and Chavrat entropy, and Tsallis entropy are derived. In section 6, Inequality measures like incomplete moments, mean deviation, Lorenz curve, Bonferroni curve, Zenga curve, mean residual life, and mean inactivity life are computed. In section 7, "Exponentiated Moment Exponential distribution" is characterized through conditional moments, hazard rate functions, reverse hazard rate functions, and truncated first mean moments. Finally, in Section 8, we provide some concluding remarks.

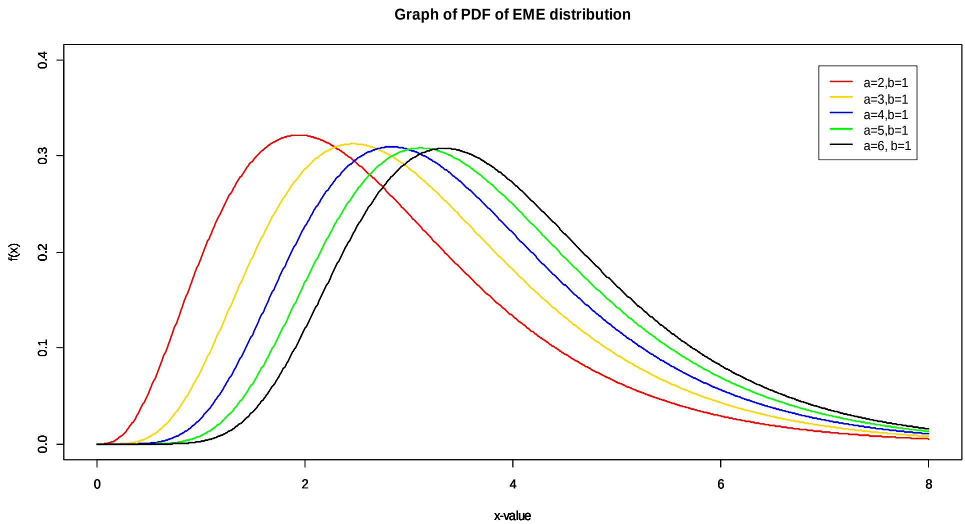
2. EXPONENTIATED MOMENT EXPONENTIAL DISTRIBUTION

2.1 Probability Density Function

If a random variable X has exponentiated moment exponential probability distribution with α shape and β scale parameters, then probability density function (pdf) of exponentiated moment exponential distribution is given as

$$\mathrm{f}(x)=\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}, \quad x>0,\ \alpha>1,\ \beta>0 \tag{2.1}$$

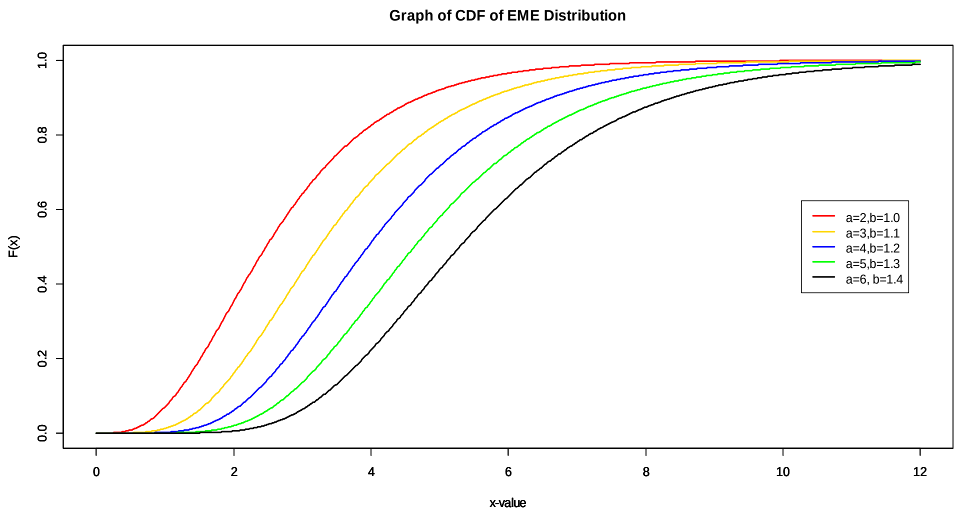


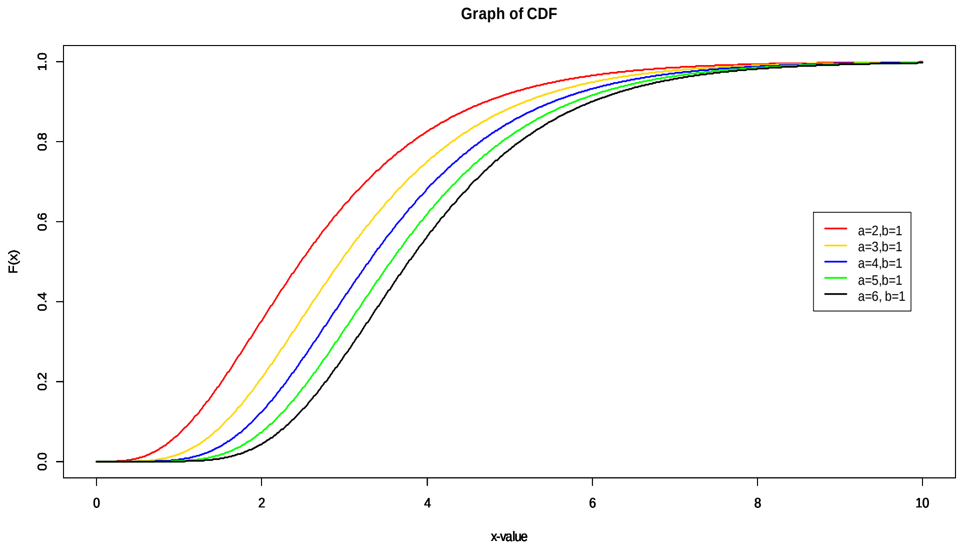


2.2 Cumulative distribution Function (cdf)

If a random variable X has EME probability distribution with α shape and β scale parameters, the cumulative distribution function of exponentiated moment exponential distribution is given as

$$F(x)=\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha},\ x>0,\ \alpha>1,\ \beta>0 \tag{2.2}$$

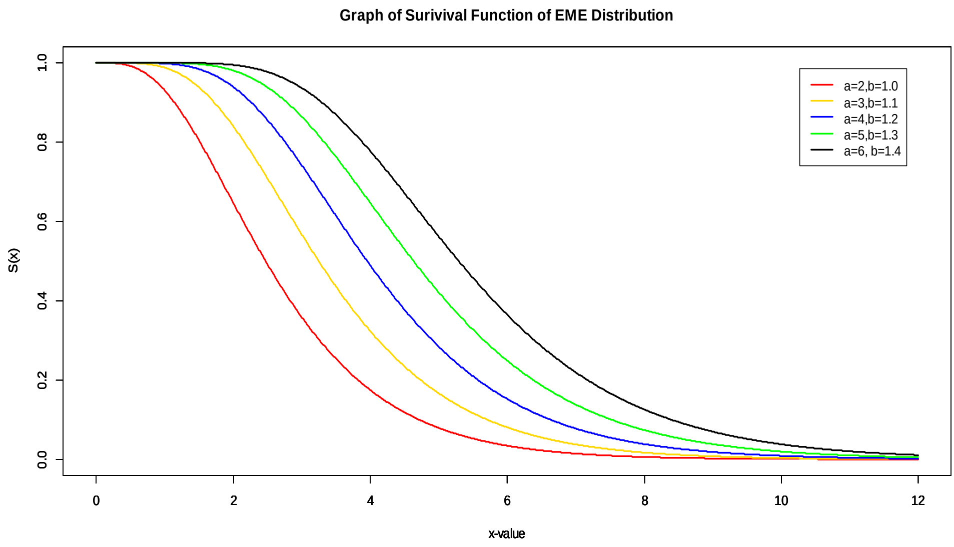


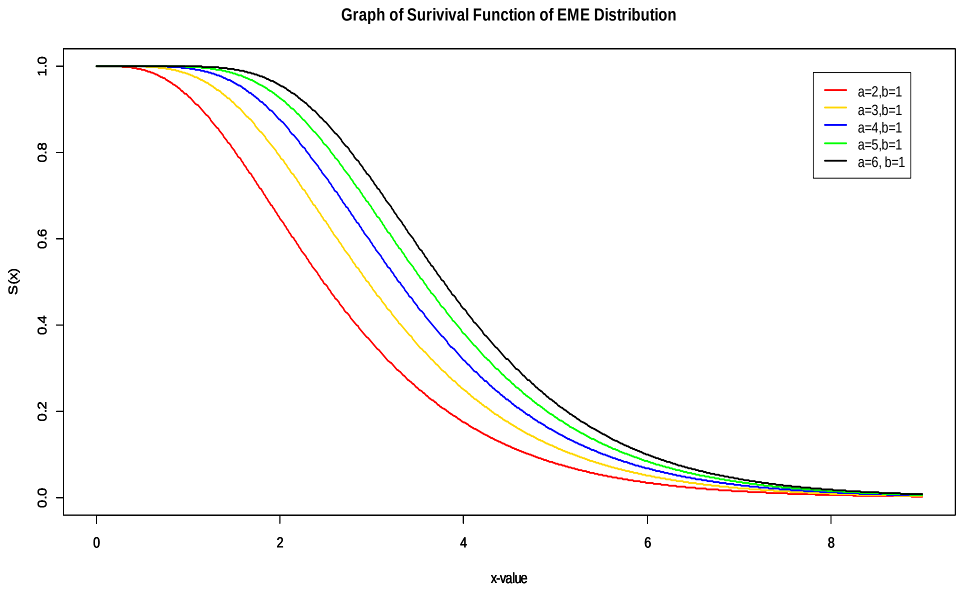


2.3 Survival Function

If a random variable X has $EMED(\alpha,\beta)$, then Survival function of exponentiated moment exponential distribution is given as

$$S(x)=1-\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha} \tag{2.3}$$





2.4 Hazard Rate Function

If a random variable X has $EMED(\alpha,\beta)$, then hazard rate function of EME distribution is given as

$$h(x)=\frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta}}{1-\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}} \tag{2.4}$$

2.5 Reverse Hazard Rate

If a random variable X has $EMED(\alpha,\beta)$ probability distribution, then reverse hazard rate function of EME distribution is given as

$$r(x)=\frac{\frac{\alpha}{\beta^2}xe^{-x/\beta}}{\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]} \tag{2.5}$$

2.6 Cumulative Hazard Rate

If a random variable X has exponentiated moment exponential probability distribution with α shape and β scale parameters, then Cumulative hazard rate function of EME distribution is given as

$$H(x) = -\ln\left(1-\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right) \tag{2.6}$$

2.7 Mills Ratio

The ratio between Survival function and Probability function is called Mills ratio (Inverse of hazard rate function). If a random variable X has exponentiated moment exponential probability distribution with α shape and β scale parameters, then Mills ratio is given as

$$m(x) = \frac{\beta^2 e^{x/\beta}}{\alpha x}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right] \tag{2.7}$$

2.8 Elasticity of EME Distribution

If a random variable X has exponentiated moment exponential probability distribution with α shape and β scale parameters, then Elasticity of EME distribution is given as

$$e(x) = xr(x) = \frac{\frac{\alpha}{\beta^2}x^2 e^{-x/\beta}}{\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]} \tag{2.8}$$

2.9 Sub Models of EMED Distribution

i) For $\beta = 1$, EME Distribution will become Exponentiated Exponential Distribution

ii) For $\alpha = 1$, EME Distribution will become Moment Exponential Distribution

iii) For $\alpha = 1$ *and* $\beta = 1$, EME Distribution will become Exponential Distribution.

3. PROPERTIES OF EME DISTRIBUTION

Some new properties EME Distribution are derived to understand scope and application of distribution. We derive Moments, Factorial Moments, Negative Moments, Moment generating function, Lower and Upper Incomplete raw moments,

3.1 Moments about Origin

Moments are necessary and important to study the features and characteristics of a distribution (e.g. tendency, dispersion, skewness and Kurtosis). If a random variable X has $EMED(\alpha,\beta)$, then the Moments are given as

$$\mu_r^{/} = E\left(X^r\right) = \int_0^\infty x^r f(x)\,dx$$

$$E\left(X^r\right) = \frac{\alpha}{\beta^2}\int_0^\infty \left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} x^{r+1}e^{-x/\beta}dx.$$

After substitution, we obtain

$$E\left(X^r\right)=\alpha\beta^r\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}(-1)^i\binom{\alpha-1}{\mathrm{i}}\binom{i}{j}\frac{\Gamma(j+r+2)}{(1+i)^{j+r+2}} \tag{3.1}$$

3.2 Factorial Moments

If a random variable X has $EME(\alpha,\beta)$ probability distribution, then the factorial moments of X are given as follows

$$E[X]_n=\sum_{i=1}^{n}\alpha_i E(X^i)=\sum_{i=1}^{n}\alpha_i\mu_i^{'} ,$$

where $[X]_i=X(X+1)(X+2)......(X+i-1)$.

3.3 Negative Moments of EMED

If a random variable X has $EMED(\alpha,\beta)$, then the factorial moments of X are given as

$$E\left(X^{-r}\right)=\alpha\beta^{-r}\sum_{i}^{\infty}\sum_{j=0}^{r-2}(-1)^i\binom{\alpha-1}{\mathrm{i}}\binom{i}{j}\frac{\Gamma(j-r+2)}{(1+i)^{j-r+2}}.$$

3.4 Moment Generating Function of EME Distribution

If a random variable X has $EME(\alpha,\beta)$ distribution, then the Moment generating function for X are given as

$$M_x(t)=E\left[e^{tx}\right]$$

$$M_x(t)=\sum_{r=1}^{\infty}\frac{t^r}{\mathrm{r}!}\mu_r^{/}$$

$$M_x(t)=\sum_{r=1}^{\infty}\frac{t^r}{\mathrm{r}!}\left[\alpha\beta^r\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}(-1)^i\binom{\alpha-1}{\mathrm{i}}\binom{i}{j}\frac{\Gamma(j+r+2)}{(1+i)^{j+r+2}}\right]$$

3.5 Lower Incomplete Moments

If a random variable X has $EME(\alpha,\beta)$ distribution, then the rth Lower Incomplete raw moments for X are given as

$$M_r(X)_{X\le x}=\frac{\alpha}{\beta^2}\int_0^x\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^{\alpha-1}t^{r+1}e^{-t/\beta}dt\,.$$

$$M_r(X)_{X\le x}=\alpha\beta^r\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\gamma\left(\frac{x}{\beta},j+r+2\right)}{(1+i)^{j+r+2}} \tag{3.5.1}$$

3.6 Upper Incomplete Moments

If a random variable X has $EME(\alpha,\beta)$ distribution, then the rth Upper Incomplete raw moments for X are given as

$$M_r(X)_{X\geq x}=\frac{\alpha}{\beta^2}\int_x^{\infty}\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^{\alpha-1}t^{r+1}e^{-t/\beta}dt\ .$$

$$M_r(X)_{X\geq x}=\alpha\beta^r\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\Gamma\left(\frac{x}{\beta},j+r+2\right)}{(1+i)^{j+r+2}} \qquad (3.6.1)$$

4. STRESS-STRENGTH RELIABILITY OF EME DISTRIBUTION

4.1 Stress-strength model

If $X_1\sim EME(\alpha_1,\beta), X_2\sim EME(\alpha_2,\beta)$, X_1 represents "strength" and X_2 represents "stress" and X_1 and X_2 follow a joint pdf $f(x_1,\mathrm{x}_2)$, then reliability of the component is:

$$\mathrm{R}=\Pr(\mathrm{X}_2<X_1)=\int_{-\infty}^{\infty}\int_{-\infty}^{x_1}f(x_1,x_2)dx_2dx_1$$

If the random variables are statistically independent, then

$$\mathrm{R}=\Pr(\mathrm{X}_2<X_1)=\int_{-\infty}^{\infty}\int_{-\infty}^{x_1}f(x_1)\,f(\mathrm{x}_2)\,dx_2dx_1$$

$$F_{x_2}(x_1)=\int_{-\infty}^{x_1}f(x_2)dx_2$$

$$R=\int_0^{\infty}f_{x_1}(x)F_{x_2}(x)\,dx$$

$$\mathrm{R}=\Pr(\mathrm{X}_2<X_1)=\int_0^{\infty}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha_2}\frac{\alpha_1}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha_1-1}xe^{-x/\beta}dx$$

$$\mathrm{R}=\Pr(\mathrm{X}_2<X_1)=\frac{\alpha_1}{\alpha_1+\alpha_2} \qquad (4.1)$$

Therefore

(i) R is independent of β

(ii) For $\alpha_1=\alpha_2$ $R=0.5$,

It means that X1 and X2 are i.i.d and there is equal chance that X1 is bigger than X2.

4.2 Reliability in Multicomponent Stress-Strength Based On EME Distribution

Let the random samples $y,x_1,x_2,....x_k$ be independent, G(y) be the continuous distribution function of Y and F(x) be the common continuous distribution function of

$x_1, x_2, ...x_k$. The reliability in a multicomponent stress strength model developed by Bhattacharyya and Johnson (1974) is

$$R_{s,k} = P[at\ least\ s\ of\ (x_1, x_2, ...x_k)\ exceed\ Y] \tag{4.2}$$

$$R_{s,k} = \sum_{i=s}^{k} \binom{k}{i} \int_0^\infty [1 - G(y)]^i [G(y)]^{k-i}\, dF(y)$$

where $x_1, x_2, ...x_k$ independently and identically distributed (iid) are with common distribution function F (x), this system is subjected to common random stress Y.

The probability in (4.2) is called reliability in a multicomponent stress-strength Model (Bhattacharyya & Johnson, 1974).

Suppose a system, with k identical components, functions if s $(1 < s < k)$ or more of the components simultaneously operate. In its operating environment, the system is subjected to a stress Y which is a random variable with distribution function G(.). The strengths of the components, that are the minimum stresses to cause failure, are independent and identically distributed (iid) random variables with distribution function F(.). Then the system reliability, which is the probability that the system does not fail, is the function $R_{s,k}$ given in (4.2). The estimation of reliability in a multicomponent stress-strength system when the stress, strength variates are following exponentiated moment exponential distribution is not studied.

Estimator of $R_{s,k}$

Let $X \sim EME(\alpha_1, \beta), Y \sim EME(\alpha_2, \beta)$ with unknown shape parameters $\alpha_{1,}$ α_2 and common scale parameter β, where X and Y are independently distributed. The reliability in multicomponent stress- strength for exponentiated moment exponential distribution using (3) results in:

$$R_{s,k} = \sum_{i=s}^{k} \binom{k}{i} \int_0^\infty \left\{ 1 - \left[1 - \frac{y+\beta}{\beta} e^{-y/\beta} \right]^{\alpha_1} \right\}^i$$

$$\left[\left[1 - \frac{y+\beta}{\beta} e^{-y/\beta} \right]^{\alpha_1} \right]^{k-i} \frac{\alpha_2}{\beta^2} \left[1 - \frac{y+\beta}{\beta} e^{-y/\beta} \right]^{\alpha_2 - 1} y e^{-y/\beta} \mathrm{dy}$$

$$where\ \ t = 1 - \frac{y+\beta}{\beta} e^{-y/\beta},\ \ dt = \frac{y e^{-y/\beta}}{\beta^2} \mathrm{dy}$$

$$R_{s,k} = \sum_{i=s}^{k} \binom{k}{i} \int_0^1 \{1 - t^{\alpha_1}\}^i [t^{\alpha_1}]^{k-i} \alpha_2 t^{\alpha 2 - 1} dt$$

$$where\ \ z = t^{\alpha_1}\ \ , t = z^{\frac{1}{\alpha_1}}\ \ \ dt = \frac{z^{\frac{1}{\alpha_1} - 1}}{\alpha_1} dz$$

$$R_{s,k} = \sum_{i=s}^{k}\binom{k}{i}\int_0^1\left\{(1-z)^i (z)^{k-i}\,\alpha_2 \,(z^{\frac{1}{\alpha_1}})^{\alpha 2-1}\,\frac{z^{\frac{1}{\alpha_1}-1}}{\alpha_1}\right\}dz$$

$$R_{s,k} = \sum_{i=s}^{k}\binom{k}{i}\int_0^1\left\{(1-z)^i (z)^{k-i}\,\frac{\alpha_2}{\alpha_1}\,z^{\frac{\alpha_2}{\alpha_1}-\frac{1}{\alpha_1}+\frac{1}{\alpha_1}-1}\right\}dz$$

$$R_{s,k} = v\sum_{i=s}^{k}\binom{k}{i}\int_0^1\left\{(1-z)^i (z)^{k-i+v-1}\right\}dz$$

$$where\ v = \frac{\alpha_2}{\alpha_1},$$

$$R_{s,k} = v\sum_{i=s}^{k}\binom{k}{i}\,\beta(i+1,k-i+v)$$

After simplification we get

$$R_{s,k} = v\sum_{i=s}^{k}\frac{k!}{(k-i)!}\left[\prod_{j=0}^{i}(k+v-j)\right]^{-1}$$

5. MEASURES OF UNCERTAINTY

In this section, we present information generating function, Renyi entropy, Q-entropy, Havrda and Chavrat entropy, and Tsallis entropy, for the EME distribution.

5.1 Information generating Function

The differentiation of information generating function at point 0 or 1, help to derive the measures of information which are otherwise difficult to characterize and compute. The information generating function of a probability distribution is defined as

$$H(f) = E\left[\left(f(X)\right)^{v-1}\right] = \int_0^\infty \left(f(x)\right)^v dx$$

$$E\left[\left(f(x)\right)^v\right] = \frac{\alpha^v}{\beta^{v-1}}\sum_{i=0}^{v(\alpha-1)}\sum_{j=0}^{i}(-1)^i\binom{v(\alpha-1)}{i}\binom{i}{j}\frac{\Gamma(j+v+1)}{(v+1)^{(j+v+1)}}$$

If a random variable X has EME distribution, then information generating function is given as

$$H(f) = \frac{\alpha^v}{\beta^{v-1}}\sum_{i=0}^{v(\alpha-1)}\sum_{j=0}^{i}(-1)^i\binom{v(\alpha-1)}{i}\binom{i}{j}\frac{\Gamma(j+v+1)}{(v+1)^{(j+v+1)}} \tag{5.1}$$

The Shannon Entropy can be found by $\left.\frac{d}{dv}H(f)\right|_{v=1}$

5.2 Entropy

Entropy is brilliant device to measure the amount of information (randomness or uncertainty) contained in random observation about its parent distribution. Entropies are used to measure expected information content or uncertainty or randomness of random variable x of probability distribution. A large value of entropy is indication of the greater uncertainty in the data.

5.3 Renyi Entropy

Renyi entropy for a continuous probability distribution with f(x) pdf is defined as

$$I_R(v) = \frac{1}{1-v}\log\left(\int_{-\infty}^{\infty}\left(f(x)\right)^v dx\right) \quad v \neq 1, v > 0 \tag{5.2}$$

If a random variable X has EME distribution, then Renyi entropy is given as

$$I_R = \frac{1}{1-\nu}\log\left\{\frac{\alpha^v}{\beta^{v-1}}\sum_{i=0}^{v(\alpha-1)}\sum_{j=0}^{i}(-1)^i\binom{v(\alpha-1)}{i}\binom{i}{j}\frac{\Gamma(j+v+1)}{(v+1)^{(j+v+1)}}\right\} \tag{5.3}$$

5.4 Q-Entropy

Q-Entropy for a continuous probability distribution with f(x) pdf is defined as

$$H_q(f) = \frac{1}{1-q}\log\left[1-I(q)\right]$$

If a random variable X has EME distribution, then Q-Entropy is given as

$$H_q(f) = \frac{1}{1-q}\log\left[1-\frac{\alpha^q}{\beta^{q-1}}\sum_{i=0}^{q(\alpha-1)}\sum_{j=0}^{i}(-1)^i\binom{q(\alpha-1)}{i}\binom{i}{j}\frac{\Gamma(j+q+1)}{(q+1)^{(j+q+1)}}\right] \tag{5.4}$$

5.5 Havrda and Chavrat Entropy (1967)

Havrda and Chavrat Entropy (1967) for a continuous probability distribution with f(x) pdf is defined as

$$I_R(v) = \frac{1}{v-1}\log\left(\int_{-\infty}^{\infty}(f(x))^v dx\right) \quad v \neq 1, v > 0$$

If a random variable X has EME distribution, then Havrda and Chavrat Entropy (1967) is given by

$$I_R = \frac{1}{\nu-1}\log\left\{\frac{\alpha^v}{\beta^{v-1}}\sum_{i=0}^{v(\alpha-1)}\sum_{j=0}^{i}(-1)^i\binom{v(\alpha-1)}{i}\binom{i}{j}\frac{\Gamma(j+v+1)}{(v+1)^{(j+v+1)}}\right\}, \quad v \neq 1 \tag{5.5}$$

5.6 Tsallis Entropy

Tsallis Entropy was introduced by Constnatino Tsallis (1988). Tsallis Entropy for a continuous probability distribution with f(x) pdf is defined as

$$S_q\left(f(x)\right)=\frac{1}{q-1}\left(1-\int\left(f(x)\right)^q dx\right)$$

If a random variable X has EME distribution, then Tsallis Entropy Is given by

$$S_q\left(f(x)\right)=\frac{1}{q-1}\left(1-\frac{\alpha^q}{\beta^{q-1}}\sum_{i=0}^{q(\alpha-1)}\sum_{j=0}^{i}(-1)^i\binom{q(\alpha-1)}{i}\binom{i}{j}\frac{\Gamma(j+q+1)}{(q+1)^{(j+q+1)}}\right) \tag{5.6}$$

6. MEAN DEVIATIONS AND INEQUALITY MEASURES

In this section, we present the mean deviation about the mean, the mean deviation about the median, Bonferroni and Lorenz curves. Bonferroni, Lorenz curves and Zenga curve are income inequality measures that are also useful and applicable in other areas including reliability, demography, medicine and insurance.

6.1 Mean Deviations

The amount of dispersion in population may be measured with totality of absolute values of deviations from mean or median.

The mean deviation about the mean and mean deviation about the median are defined by

$$D(\mu)=\int_0^\infty |x-\mu|\,\mathrm{f}(x)dx,\quad and\ \ D(M)=\int_0^\infty |x-M|\,\mathrm{f}(x)dx,\ \text{respectively,}$$

where μ= E(X) and M = Median (X) = $F^{-1}(1/2)$ is median of F. These measures $D(\mu)$ and D(M) can be calculated using the relationship

$$D(\mu)=2\mu G(\mu)-2\mu+2\int_\mu^\infty x\,\mathrm{f}(x)dx=2\mu G(\mu)-2\int_0^\mu x\,\mathrm{f}(x)dx$$

$$D(\mu)=2\mu G(\mu)-2\mu+2\int_\mu^\infty x\,\mathrm{f}(x)dx=2\mu G(\mu)$$

$$-2\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\gamma\left(\frac{\mu}{\beta},j+3\right)}{(1+i)^{j+3}} \tag{6.1.1}$$

$$D(M)=-\mu+2\int_M^\infty xg(x)dx=\mu-2\int_0^M xg(x)dx$$

$$D(M)=-\mu+2\int_M^\infty xg(x)dx=\mu-2\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\gamma\left(\frac{M}{\beta},j+3\right)}{(1+i)^{j+3}} \tag{6.1.2}$$

6.2 Lorenz Curve

If a random variable X has $EME(\alpha,\beta)$ distribution, then Lorenz Curve is given by

$$L(\mathrm{F}(x)) = \frac{\int_0^x t\,\mathrm{f}(t)dt}{E(x)} \ ,$$

$$L(\mathrm{F}(x)) = \frac{1}{\mu}\int_0^Q t\,\mathrm{f}(t)dt$$

$$L(\mathrm{F}(x)) = \frac{1}{\mu}\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\gamma\left(\frac{Q}{\beta}, j+3\right)}{(1+i)^{j+3}} \tag{6.2.1}$$

6.3 Bonferroni curve

If a random variable X has $EME(\alpha,\beta)$ distribution, then Bonferroni curve is given by

$$B(\mathrm{F}(x)) = \frac{L(\mathrm{F}(x))}{\mathrm{F}(x)}$$

$$B(\mathrm{F}(x)) = \frac{1}{p\mu}\int_0^Q t\,\mathrm{f}(t)dt$$

$$L(\mathrm{F}(x)) = \frac{1}{p\mu}\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\gamma\left(\frac{Q}{\beta}, j+3\right)}{(1+i)^{j+3}} \tag{6.3.1}$$

where $Q(p) = F^{-1}(p)$, $x = F^{-1}(p)$ *and* $T(x) = \int_0^x t\,\mathrm{f}(t)dt$

6.4 Zenga Curve

If a random variable X has $EME(\alpha,\beta)$ distribution, then Zenga (2007) Purposed a curve based on ratios of Lower and Upper means

$$\mu^-(x) = \frac{\int_0^x tf(t)dt}{F(x)} = \frac{\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\gamma\left(\frac{x}{\beta}, j+3\right)}{(1+i)^{j+3}}}{F(x)}$$

$$\mu^+(x) = \frac{\int_x^\infty tf(t)dt}{1-F(x)} = \frac{\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\Gamma\left(\frac{x}{\beta}, j+3\right)}{(1+i)^{j+3}}}{1-F(x)}$$

where $Q(p)=F^{-1}(p)$, $x=F^{-1}(p)$, $\mu^{-}(x)$ and $\mu^{+}(x)$ are lower and upper means.

$$A(x)=1-\frac{\mu^{-}(x)}{\mu^{+}(x)} \tag{6.4.1}$$

6.5 Mean Residual Life (MRL) Function

The mean residual life (MRL) function or the life expectancy at a given time t, say m(t), measures the expected remaining lifetime of an individual of age t and it given by

$$m(t)=\frac{1}{S(t)}\left[E(t)-\int_0^t xf(x)\,dx\right]-t$$

$$m(t)=\frac{1}{S(t)}\left[\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\Gamma(j+3)}{(1+i)^{j+3}}-\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\gamma\left(\frac{t}{\beta},j+3\right)}{(1+i)^{j+3}}\right]-t$$

$$m(t)=\frac{1}{S(t)}\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\left[\frac{\Gamma(j+3)}{(1+i)^{j+3}}-\frac{\gamma\left(\frac{t}{\beta},j+3\right)}{(1+i)^{j+3}}\right]-t \tag{6.5.1}$$

6.6 Mean Waiting Time (MWT) or Mean Inactivity Time

The mean waiting time (MWT) or mean inactivity time represents the waiting time elapsed since the failure of an item on condition that this failure had occurred in the interval [0, t]. The MWT of X, say $\bar{m}(t)$, is defined by

$$\bar{m}(t)=t-\frac{1}{F(t)}\left[\int_0^t xf(x)\,dx\right]$$

$$\bar{m}(t)=t-\frac{1}{F(t)}\left(\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\gamma\left(\frac{t}{\beta},j+3\right)}{(1+i)^{j+3}}\right) \tag{6.6.1}$$

7. CHARACTERIZATION

7.1 Characterization with Conditional Moments

Theorem 7.1.1:

Let X be a non-negative random variable having an absolute continuous distribution function $F(x)$ with $F(0)=0$ and $0<F(x)<1$ for all $x>0$, and then its distribution function is $F(x)=\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}$, $x>0,\ \alpha>1,\ \beta>0$

If and only if

$$E\left[\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m \middle| X>t\right]=\frac{1}{(m+\alpha)\bar{F}(t)}\left\{1-\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^{m+\alpha}\right\}$$

Proof:

$$f(x)=\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}$$

$$E\left[\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m \middle| X>t\right]$$

$$=\frac{1}{\bar{F}(t)}\int_t^\infty\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{m+\alpha-1} xe^{-x/\beta}dx$$

$$E\left[\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m \middle| X>t\right]=\frac{1}{(m+\alpha)\bar{F}(t)}\left\{1-\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^{m+\alpha}\right\} \quad (7.1)$$

Conversely

$$\frac{1}{\bar{F}(t)}\int_t^\infty\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m f(x)dx=\frac{1}{(m+\alpha)\bar{F}(t)}\left\{1-\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^{m+\alpha}\right\}$$

Differentiating both sides

$$\int_t^\infty\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m f(x)dx=\frac{1}{(m+\alpha)}\left\{1-\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^{m+\alpha}\right\}$$

$$-\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m f(x)=-\frac{m+\alpha}{(m+\alpha)}\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha+m-1}$$

$$f(x)=\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}$$

Theorem 7.1.2:

Let X be a non-negative random variable having an absolute continuous distribution function $F(x)$ with $F(0)=0$ and $0<F(x)<1$ for all $x>0$, and then its distribution function is $F(x)=\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^\alpha$, $x>0,\ \alpha>1,\ \beta>0$

If and only if $E\left[\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m \middle| X \le t\right] = \frac{1}{(m+\alpha)}\left[1-\left(1+\frac{t}{\beta}\right)e^{-x/\beta}\right]^m$

Proof:

$$f(x) = \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}$$

$$E\left[\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m \middle| X \le t\right] = \frac{1}{F(t)}\int_0^t \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{m+\alpha-1} xe^{-x/\beta}dx$$

$$E\left[\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m \middle| X \le t\right] = \frac{1}{(m+\alpha)F(t)}\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^{m+\alpha}$$

$$E\left[\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m \middle| X \le t\right] = \frac{1}{(m+\alpha)}\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^m \tag{7.2}$$

Conversely

$$E\left[\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m \middle| X \le t\right] = \frac{1}{(m+\alpha)}\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^m$$

Differentiating both sides

$$\int_0^t \left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m f(x)dx = \frac{1}{(m+\alpha)}\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^{m+\alpha}$$

$$\int_0^t \left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m f(x)dx = \frac{1}{(m+\alpha)}\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^{m+\alpha}$$

$$\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^m f(t) = \frac{(m+\alpha)}{(m+\alpha)}\left[1-\left(1+\frac{t}{\beta}\right)e^{-t/\beta}\right]^{m+\alpha-1}$$

$$\mathrm{f}(x) = \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}, \quad x>0,\ \alpha>1,\ \beta>0$$

7.2 Characterization through Hazard Rate Function

Proposition 7.2.1

The probability density function of a continuous random variable X is

$$\mathrm{f}(x) = \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}, \quad x>0,\ \alpha>1,\ \beta>0$$

if and only if the hazard function satisfy the equation

$$f(x) = ce^{\int\left(\frac{\eta_F^{/}(x)}{\eta_F(x)} - \eta_F(x)\right)dx}$$

Proof:

$$\eta_F(x) = \frac{f(x)}{1-F(x)} = \frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}{1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}}$$

$$\frac{\eta_F^{/}(x)}{\eta_F(x)} - \eta_F(x) = (\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta} - e^{-x/\beta}\frac{1}{\beta}\right\} + \frac{1}{x} + \left(-\frac{1}{\beta}\right)$$

$$f(x) = k\exp\left(\int\left(\frac{\eta_F^{/}(x)}{\eta_F(x)} - \eta_F(x)\right)dx\right)$$

$$f(x) = k\exp\left[\int\left((\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta} - e^{-x/\beta}\frac{1}{\beta}\right\} + \frac{1}{x} + \left(-\frac{1}{\beta}\right)\right)du\right]$$

$$f(x) = k\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{(\alpha-1)} xe^{-\frac{x}{\beta}}$$

For $\int_0^\infty f(x)dx = 1$ so $k = \frac{\alpha}{\beta^2}$

$$\text{f}(x) = \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}, \quad x>0,\ \alpha>1,\ \beta>0$$

Proposition 7.2.2

The probability density function of a continuous random variable X is

$$\text{f}(x) = \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}, \quad x>0,\ \alpha>1,\ \beta>0$$

if and only if the hazard function satisfy the equation

(i)
$$\frac{\beta^2}{\alpha}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-\alpha+1} x^{-1}e^{x/\beta}\eta'(x)-\eta(x)\frac{\beta^2}{\alpha}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-\alpha+1} x^{-1}e^{x/\beta}$$
$$\left\{(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta}-e^{-x/\beta}\frac{1}{\beta}\right\}+\frac{1}{x}-\frac{1}{\beta}\right\}$$
$$=\frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}{\left(1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)^2}$$

(ii)
$$\eta'(x)\left(1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)-\eta^2(x)\left(1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)=\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}$$
$$\left\{(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta}-e^{-x/\beta}\frac{1}{\beta}\right\}+\frac{1}{x}-\frac{1}{\beta}\right\}$$

Proof:

(i)
$$\eta(x)=\frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}{1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}}$$

$$\frac{\beta^2}{\alpha}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-\alpha+1} x^{-1}e^{x/\beta}\eta'(x)-\eta(x)\frac{\beta^2}{\alpha}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-\alpha+1} x^{-1}e^{x/\beta}$$
$$\left\{(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta}-e^{-x/\beta}\frac{1}{\beta}\right\}+\frac{1}{x}-\frac{1}{\beta}\right\}$$
$$=\frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}{\left(1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)^2}$$

$$\eta'(x)-\eta(x)xe^{-x/\beta}\left\{\begin{matrix}(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\\ \left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta}-e^{-x/\beta}\frac{1}{\beta}\right\}+\frac{1}{x}-\frac{1}{\beta}\end{matrix}\right\}-\eta^2(x)=0$$

$$\frac{d}{dx}\left[\frac{\eta(x)}{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}\right]=\frac{d}{dx}\left(\frac{1}{1-\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}}\right) \qquad A$$

(ii)

$$\eta'(x)\left(1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)-\eta^2(x)\left(1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)=\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}$$

$$\left\{(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta}-e^{-x/\beta}\frac{1}{\beta}\right\}+\frac{1}{x}-\frac{1}{\beta}\right\}$$

$$\eta'(x)\left(1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)-\eta^2(x)\left(1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)=\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}$$

$$\left\{(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta}-e^{-x/\beta}\frac{1}{\beta}\right\}+\frac{1}{x}-\frac{1}{\beta}\right\}$$

$$\frac{d}{dx}\left[\eta(x)\left(1-\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)\right]=\frac{d}{dx}\left(\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}\right) \qquad B$$

Solution of Eq. A and B

$$\eta(x)=\frac{f(x)}{1-F(x)}=\frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}{1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}}$$

$$\ln\left(1-F(x)\right)=\ln\left(1\text{-}\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)$$

$$F(x)=\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha},\ x>0,\ \alpha>1,\ \beta>0$$

7.3 Characterization through Hazard Rate Function

Proposition 7.3.1

The probability density function of a continuous random variable X is

$$\mathrm{f}(x)=\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta},\quad x>0,\ \alpha>1,\ \beta>0$$

if and only if the Reverse hazard function satisfy the equation

$$f(x) = ce^{\int \left(\frac{r_F^{/}(x)}{r_F(x)} + r_F(x) \right) dx}$$

Proof:

$$r(x) = \frac{f(x)}{F(x)} = \frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}{\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}}$$

$$\frac{r_F^{/}(x)}{r_F(x)} + r_F(x) = (\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta} - e^{-x/\beta}\frac{1}{\beta}\right\} + \frac{1}{x} + \left(-\frac{1}{\beta}\right)$$

$$f(x) = k\exp\left[\int\left(\frac{r_F^{/}(x)}{r_F(x)} + r_F(x)\right)dx\right]$$

$$f(x) = k\exp\left[\int\left((\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta} - e^{-x/\beta}\frac{1}{\beta}\right\} + \frac{1}{x} + \left(-\frac{1}{\beta}\right)\right)du\right]$$

$$f(x) = k\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{(\alpha-1)} xe^{-\frac{x}{\beta}}$$

For $\int_0^\infty f(x)dx = 1$ so $k = \frac{\alpha}{\beta^2}$

$$\text{f}(x) = \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}, \quad x>0,\ \alpha>1,\ \beta>0$$

Proposition 7.3.2

The probability density function of a continuous random variable X is

$$\text{f}(x) = \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}, \quad x>0,\ \alpha>1,\ \beta>0$$

if and only if the Reverse hazard function satisfy the equation

$$r_F'(x) - \left\{(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left(\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta} - e^{-x/\beta}\frac{1}{\beta}\right) + \frac{1}{x} - \frac{1}{\beta}\right\} r_F(x) + r_F{}^2(x) = 0$$

(i)
$$\left[r'(x) - \left\{ \begin{aligned} &(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-\alpha}\frac{1}{\alpha} \\ &+\left(\frac{1}{x}-\frac{1}{\beta}\right)\frac{\beta^2}{\alpha}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-\alpha+1} x^{-1}e^{x/\beta} \end{aligned} \right\} r(x) \right]$$

$$= -\frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}{\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{2\alpha}}$$

(ii)
$$\left\{ r'(x)\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha} + r(x)\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta} \right\}$$

$$= \left[\left\{ (\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\frac{x}{\beta^2}e^{-x/\beta} + \frac{1}{x}-\frac{1}{\beta} \right\} \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta} \right]$$

Proof:

$$r_F(x) = \frac{f(x)}{F(x)} = \frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}{1-\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}}$$

(i)
$$\left[r'(x) - \left\{ \begin{aligned} &(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-\alpha}\frac{1}{\alpha} \\ &+\left(\frac{1}{x}-\frac{1}{\beta}\right)\frac{\beta^2}{\alpha}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-\alpha+1} x^{-1}e^{x/\beta} \end{aligned} \right\} r(x) \right]$$

$$= -\frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}{\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{2\alpha}}$$

$$\frac{d}{dx}\left[\frac{r(x)}{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} xe^{-x/\beta}}\right] = \frac{d}{dx}\left(\frac{1}{\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}}\right)$$

(ii) $$\left\{r'(x)\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}+r(x)\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta}\right\}$$

$$=\left[\begin{array}{c}\left\{(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\frac{x}{\beta^2}e^{-x/\beta}+\frac{1}{x}-\frac{1}{\beta}\right\}\\ \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta}\end{array}\right]$$

$$\frac{d}{dx}\left[r(x)\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right]=\frac{d}{dx}\left(\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta}\right)$$

$$\ln\left(F(x)\right)=\ln\left(\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}\right)$$

$$F(x)=\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha},\ x>0,\ \alpha>1,\ \beta>0$$

7.4 A New Characterization of the EME Distribution through First Truncated Moment

Suppose that EME random variable X has cdf $F(x)=\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}$, with pdf $f(x)=\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta}$ and $E(X/X\le x)$ exists for all real x $(0,\infty)$. Then

$E(X/X\le x)=g(x)\eta(x)$ and $\eta(x)=\dfrac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta}}{\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}}$

where $g(x)=\dfrac{\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\dfrac{\gamma\left(\frac{x}{\beta},j+3\right)}{(1+i)^{j+3}}}{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta}}$ is differentiable function w r t x for all real $x\ \varepsilon(0,\infty)$, if

$$f(x) = c e^{\int_{-\infty}^{x} \frac{u - g'(u)}{g(u)} du}$$

where c is determined such that $\int_{-\infty}^{\infty} f(x)dx = 1$

Proof:

Necessary Part: Suppose that

$$f(x) = \frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} x e^{-x/\beta}$$

$$\eta(x) = \frac{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} x e^{-x/\beta}}{\left[1-\frac{x+\beta}{\beta}e^{-x/\beta}\right]^{\alpha}}$$

$$E(X / X \le x) = g(x)\eta(x)$$

We have

$$\frac{\int_{-\infty}^{x} u f(u)du}{F(x)} = \frac{g(x)f(x)}{F(x)}$$

$$\int_{-\infty}^{x} u f(u)du = \alpha\beta \sum_{i=0}^{\alpha-1}\sum_{j=0}^{i} \binom{\alpha-1}{i}\binom{i}{j}(-1)^i \frac{\gamma\left(\frac{x}{\beta}, j+3\right)}{(1+i)^{j+3}}$$

Thus,

$$g(x) = \frac{\alpha\beta \sum_{i=0}^{\alpha-1}\sum_{j=0}^{i} \binom{\alpha-1}{i}\binom{i}{j}(-1)^i \frac{\gamma\left(\frac{x}{\beta}, j+3\right)}{(1+i)^{j+3}}}{\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1} x e^{-x/\beta}}$$

Differentiating both sides of the equation with respect to x, we obtain

$$g'(x)=\frac{\left(\begin{array}{l}\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta}\,\frac{d}{dx}\left[\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\gamma\left(\frac{x}{\beta},j+3\right)}{(1+i)^{j+3}}\right]\\ \qquad -\alpha\beta\sum_{i=0}^{\alpha-1}\sum_{j=0}^{i}\binom{\alpha-1}{i}\binom{i}{j}(-1)^i\frac{\gamma\left(\frac{x}{\beta},j+3\right)}{(1+i)^{j+3}}\,\frac{d}{dx}\left[\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta}\right]\end{array}\right)}{\left(\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta}\right)^2}$$

$$g'(x)=x-g(x)(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta}-e^{-x/\beta}\frac{1}{\beta}\right\}+\frac{1}{x}+\left(-\frac{1}{\beta}\right)$$

After simplification we get

$$\frac{f'(x)}{f(x)}=\frac{x-g'(x)}{g(x)}=(\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta}-e^{-x/\beta}\frac{1}{\beta}\right\}+\frac{1}{x}+\left(-\frac{1}{\beta}\right)$$

Integrating the above equation we

$$f(x)=k\exp\left[\int\left(\frac{x-g'(\mathrm{x})}{g(\mathrm{x})}\right)dx\right]$$

$$f(x)=k\exp\left[\int\left((\alpha-1)\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{-1}\left\{\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\frac{1}{\beta}-e^{-x/\beta}\frac{1}{\beta}\right\}+\frac{1}{x}+\left(-\frac{1}{\beta}\right)\right)du\right]$$

$$f(x)=k\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{(\alpha-1)}xe^{-\frac{x}{\beta}}$$

For $\int_0^\infty f(x)\,dx=1$ so $k=\frac{\alpha}{\beta^2}$

$$\mathrm{f}(x)=\frac{\alpha}{\beta^2}\left[1-\left(1+\frac{x}{\beta}\right)e^{-x/\beta}\right]^{\alpha-1}xe^{-x/\beta},\quad x>0,\ \alpha>1,\ \beta>0$$

8. CONCLUDING REMARKS

In this paper we have studied different statistical properties of Exponentiated Moment Exponential Distribution. Reliability Models, Uncertainty Measures, Inequality Measures, Mean Residual Life, and Mean Inactivity Life are studied. Exponentiated Moment Exponential distribution is characterized through conditional Moments, hazard rate functions, reverse hazard rate functions, truncated first Mean moments.

REFERENCES

1. Dara, S.T. (2012). *Recent Advances in Moment Distributions and their Hazard Rate,* (Unpublished Ph.D. Thesis). National College of Business Administration and Economics Pakistan, Lahore.
2. Bhattacharyya, G.K. and Johnson, R.A. (1974). Estimation of reliability in multicomponent stress-strength model. *JASA*, 69, 966-970.
3. Fisher, R.A. (1934). The Effects of Methods of Ascertainment upon the Estimation of Frequencies. *Annals of Eugenics*, 6(1), 13-25.
4. Hasnain, A.S. and Ahmad, M. (2012). *Exponentiated Moment Exponential Distribution*. (Unpublished material). Submitted to National College of Business Administration and Economics Pakistan, Lahore.
5. Havrda, J. and Charvat, F. (1967). Quantification Method of Classification Processes. Concept of Structural Entropy. *Kybernetika*, 3, 30-35.
6. Patil, G. and Rao, C. (1977). Weighted Distributions a survey of their applications in: Krishnaiah. P.R. (ed.) *Applications of Statistics*, Amsterdam, North- Holland 383-405.
7. Patil, G. and Rao, C. (1978). Weighted Distributions and Size-biased Sampling with Applications to Wildlife Populations and Human Families. *Biometrics*, 34, 179-189.
8. Rao, C.R. (1965). On Discrete Distributions Arising out of Methods of Ascertainment Classical and Contagious Discrete Distributions. *Statistical Publishing Society*, 27(4), 320-332.
9. Rényi, A. (1961). On Measures of Entropy and Information, Proc. *4th Berk. Symp. Math. Statist. and Probl.*, University of California Press, Vol. 1, 547-461.
10. Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52*:* 479-487. Bibcode*:*1988JSP....52..479T.doi*:*10.1007/BF01016429*.*
11. Zenga, M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni*, 5, 3-27.
12. Sunoj, S.M. and Maya, S.S. (2006). Some Properties of Weighted Distributions in the Context of Repairable Systems. *Commun. in Statist. - Theo. and Meth.*, 35, 223-228.