Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 33-46

MONTE CARLO SIMULATION OF ENVIRONMENTAL AND BUSINESS DATA WITH *MATLAB* AND *R*

Muhammad Arif Hussain[1] and **Adnan Alam Khan**[2]
Institute of Business and Technology, Karachi, Pakistan
Email: [1]arif.biztek@gmail.com, [2]wmarif2002@yahoo.com

ABSTRACT

Voluminous amount of data are available nowadays in every discipline of life. Statistical analysis of data is required in order to explore the physical systems generating these data sets. Modern computer applications like, *SAS*, *R*, *Matlab*, *SPSS*, and *Minitab*, etc., help to do analysis on such data and are widely used by scientists, economist, and engineers. Among these computer applications *R* and *Matlab* are widely used for data analysis, which cover several contemporary advanced research areas and provide enough functionality. Specifically, Monte Carlo simulation is easily implemented in *R*, *Matlab* and *SAS*. This communication emphasizes use of *R* and *Matlab* in estimating environmental and business data model parameters through simulation. Bootstrap simulation of regression parameters shows *Matlab* performance is slightly better than *R*. Similarly, time series data analysis using *Matlab* and *R* has revealed that Pakistan federal excise data follows *ARIMA*(1, 3, 1) forecast model. These packages can easily be used to analyze and understand the results of modern data analysis as these computer applications support programming languages, which have simple syntaxes. In terms of modern applied statistical tools, *R* libraries are somewhat richer than those of *Matlab*. Whereas, *Matlab* has much better graphics, which researcher can put in papers or in presentations. We also discuss the carefulness needed while using these packages in education and research.

KEYWORDS

Statistical analysis, *SAS*, *R*, *Matlab*, *SPSS*, *Minitab*, Monte Carlo simulation.

1. INTRODUCTION

Voluminous amount of data are available nowadays in every discipline of life. Statistical analysis of data is required in order to explore the physical systems generating these data sets. Modern computer applications like, *SAS*, *R*, *Matlab*, *SPSS*, and *Minitab*, etc., help to do analysis on such data and are widely used by scientists, economist, and engineers. For about last three decades educators, academicians and scientists around the globe have been using these tools both for teaching and for pursuing research in diverse disciplines of education. Monte Carlo simulations are in fact computer generated dynamic models of various real world physical phenomena or processes. Monte Carlo methods can also be used to simulate large classes of stochastic processes, and the prices of financial instruments [1]. Using above mentioned computer applications, one can do

simulations to investigate physical phenomena under study that cannot generally be explored completely by traditional means.

Personal computers are owned and used by the universities and their students for statistical analyses using Excel as it is automatically preinstalled in most of the computers nowadays. Researchers have recommended that Excel should be used for typical spreadsheet applications and for producing graphics such as bar charts, pie charts, and scatter plots but not for research level statistical analyses. Moreover, statisticians have discovered numerical instabilities in many of the algorithms present in Excel. Furthermore, the pseudorandom number generators that are used in Excel are also known to be faulty [2]. *Matlab* uses Marsaglia's generator, which is powerful and has very long cycle length [3]. Similarly, *R* implements pseudorandom number generators proposed by Reeds, Hubert, Abrahams [4]. The bootstrap simulation approach can provide a simple way to obtain standard errors when the formulas from asymptotic theory are complex, as is the case for some two-step estimators [5]. Bootstrapping helps in doing inference when the parametric inference requires very complicated formulas. Simulation approach also helps in reducing bias of estimated parameters of environmental data of small size [6] and imputation of missing data, which is a critical problem in South East Asian countries including Pakistan especially in case of data related to environment [7]. One should use at least two parallel computer applications in order to counter check critical results [8]. So, this communication emphasizes use of *R* and *Matlab* in simulation of environmental and business data. It also reviews from the perspectives of its contributions in different disciplines.

Section 2 explains useful *Matlab* and *R* commands related to data analysis. Data and methodology are discussed in Section 3. In Section 4, we demonstrate use of *Matlab* and *R* by taking a few environmental and business data sets, and Section 5 concludes this study.

2. BASIC *MATLAB* AND *R* COMMANDS

For serious computer analysis one needs *R* or *Matlab* computer applications as these application systems contain many modern analysis techniques in their libraries, which is growing each day. *R* libraries are somewhat richer than those of *Matlab*. In addition, *R* is free. One can also implement new algorithms as *R* and *Matlab* support programming languages with relatively simple syntaxes. On the other hand, *Matlab* has much better graphics [9]. *Matlab* is also used for teaching of mathematical and computational subjects [10, 11]. In this section, we explain some useful commands of these computer applications.

2.1 Basic Operations in *MATLAB*

Addition (or subtraction) of two numbers in *Matlab*:

```
>> 10 + 1/5
ans =
10.2000
```

This shows use of *Matlab* as a calculator. Now, we look at the vector representation of data array. For example, if we write x = [1, 2, 3, 5, 7] at the *Matlab* prompt >> and press enter, the variable ***x*** will represent a vector containing the five numbers [12].

```
>> x = [1, 2, 3, 5, 7];
>> x
x =
1   2   3   5   7
```

Now, consider a 2 by 3 matrix *p*:

```
>>p = [1, 9, 0; 1, 4, 1];
```

The variable *p* is now a matrix with 2 rows and 3 columns.

```
p =
1   9   0
1   4   1
```

It is very easy to create matrices with *Matlab* commands, for example the following command

```
>>q = zeros(4, 6)
```

creates a matrix *q*, with 4 rows and 6 columns filled with zeros.

Similarly, commands *unifrnd*() and *normrnd*() can be used to create vector (or matrices) containing random numbers from uniform and normal (or Gaussian) probability distributions respectively. Following command creates a variable *simrand* containing 10000 random numbers from normal probability distribution with mean 5 and variance 1.

```
>>simrand = normrnd(5, 3, 10000, 1);
```

Following is the histogram of above random numbers created with the command:

```
>>hist (simrand);
```

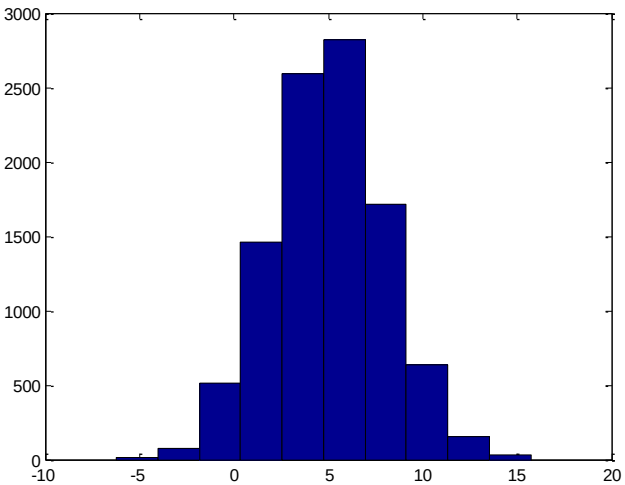


Fig. 1: Histogram of 10000 Simulated Random Numbers Using *Matlab*

Loading data into *Matlab*:

load() command loads data into *Matlab*. For example, to load a text file "sampledata.txt" into *Matlab*, following command is entered at the prompt.

>> *x* = load('sampledata.txt');

2.2 Basic Operations in *R*

Consider addition of two numbers in *R*:

```
>10 + 1/5

[1] 10.2000
```

Now, we look at the vector representation of data array. For example, if we write x = c(1, 2, 3, 5, 7) at the *R* prompt > and press enter, the variable ***x*** will represent a vector containing the five numbers.

```
> x = c(1, 2, 3, 5, 7);
> x
[1] 1 2 3 5 7
```

There are several ways of making a matrix. One can create one directly as follows [13].

```
> y <- matrix(c(1,1,9,4,0,1), nrow=2)
> y
[,1] [,2] [,3]
[1,] 1 9 0
[2,] 1 4 1
```

Command *rnorm*( ) can be used to create vector (or matrices) containing random numbers from normal (or Gaussian) probability distribution. Following command creates a variable *values* containing 10000 random numbers from normal probability distribution with mean 5 and variance 1 [14].

> *values* <- *rnorm*(10000,5,1)

Following is the histogram of above random numbers created with the command:

> hist(*values*, main="", xlab="Random numbers from Normal distribution with mean 5 and var 1")

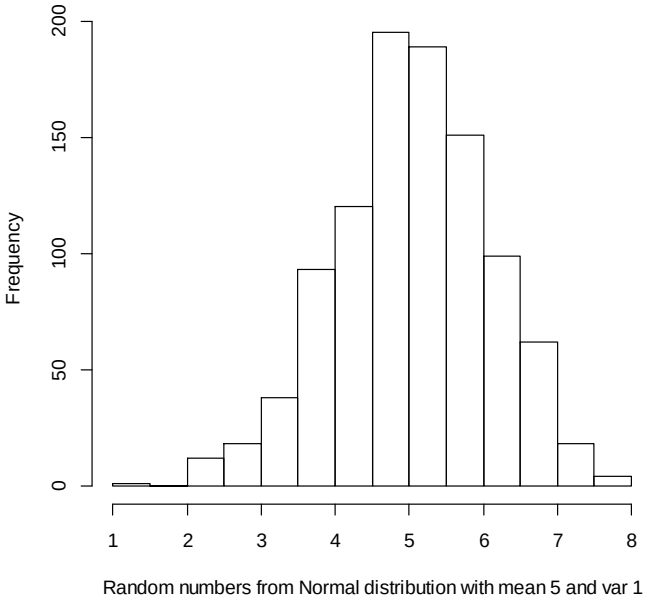


Fig. 2: Histogram of 10000 Simulated Random Numbers Using *R*

A comparison of Figures 1 and 2 shows that the histogram produced by *Matlab* is, to some extent, more symmetric as compared to the histogram produced by *R*. Now, *we describe* the *data sources* and the research *approach adopted* in this study.

3. DATA AND METHODOLOGY

A statistical analysis is a guided series of transformations of the data from its raw form to a simple summary regarding a question of interest [15]. To apply bootstrap resampling technique, mentioned in the next paragraph, we will use Duncan's regression of occupational prestige on the income and educational levels of 45 U. S. occupations [16]. To further emphasize its usefulness in optimized parameter estimation we will next apply the same approach to daily air quality measurements in New York which is daily readings of the following air quality values for May 1, 1973 to September 30, 1973 [17].

1) *Ozone*: Mean ozone in parts per billion from 1300 to 1500 hours at Roosevelt Island
2) *Solar.R*: Solar radiation in Langleys in the frequency band 4000–7700 Angstroms from 0800 to 1200 hours at Central Park
3) *Wind*: Average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia Airport
4) *Temp*: Maximum daily temperature in degrees Fahrenheit at La Guardia Airport.

Finally, in case 3, we will use Pakistan federal excise collection (in millions of Rs.) data during the1973 to 1993, to develop time series ARIMA model and simulate the model by using ***R*** functions.

Bootstrap is the nonparametric statistical simulation sampling technique in which sample is generated by sampling from a sample, and inferences are made for the population parameters. This technique is used to reduce the bias of the estimated parameters [18] as also mentioned in the introduction section. We employ bootstrap resampling to compute confidence intervals of estimated parameters. The principle is that if an estimator $\hat{\Theta}$ is normally distributed with mean 0 and standard deviation σ, then there is a probability of $1 - \alpha$ that the statement [19]

$$\hat{\Theta} - z_{\alpha/2}\,\sigma < \Theta < \hat{\Theta} + z_{\alpha/2}\,\sigma \tag{1}$$

holds for any random value of $\hat{\Theta}$, where $z_{\alpha/2}$ is the value that is exceeded with probability $\alpha/2$ for the standard normal distribution. The confidence interval is given by the rule

$$\text{estimate} \pm z_{\alpha/2} \times \text{bootstrap standard deviation.} \tag{2}$$

Plugging $z_{.025} = 1.96$ in Eq. (2), we get the standard 95% bootstrap interval of the estimated parameter.

In the next section we implement above method to regression parameters.

4. CASE STUDIES

This section considers three case studies. In case 1, bootstrap for regression coefficients is applied to Duncan's occupational prestige data.

CASE 1 DUNCAN data

```
> library(MASS)
>library(car)
> data(Duncan)
> library(boot)
>boot.nr <- function(data, indices, maxit=20){data <- data[indices,]
>mod1 <- lm(prestige ~ income + education, data=data, maxit=maxit)
>coefficients(mod1)}
>nr.boot <- boot(Duncan, boot.nr, 1999, maxit=100)
> nr.boot
```

Table 1 gives the output of above commands.

Table 1: Output from R

Bootstrap for Regression Coefficients Applied to Duncan's Occupational Prestige Data Bootstrap results are based on 1000 bootstrap samples

Model	***Bootstrap Estimates*** $\boldsymbol{\beta}$	***Bias***	***Standard Error***
Constant	–6.060	–0.030	2.952
Income	0.599	0.026	0.169
Education	0.546	–0.023	0.137

Matlab output

The bootstrap function of *Matlab* can be used directly with one of *Matlab*'s statistical functions, followed by its arguments [20]. Students and teachers should now useful commands of both computer applications.

```
>> data=load(' Duncan.dat');
>> y=data(:,3);
>> x1=data(:,1);
>> x2=data(:,2);
>> pred = [x1 x2];
>> x = [ones(size(data)),pred];
>> b = regress(y,x);
>> yfit = x*b;
>> resid = y - yfit;
>> se = std(bootstrp(...
 1000,@(bootr)regress(yfit+bootr,x),resid));
```

Table 2: Output from Matlab

Bootstrap for Regression Coefficients Applied to Duncan's Occupational Prestige Data Bootstrap results are based on 1000 bootstrap samples.

Model	***Bootstrap Estimates*** β	***Bias***	***Standard Error***
Constant	–6.065	–0.022	3.003
Income	0.599	0.021	0.163
Education	0.546	–0.018	0.135

Comparing Tables 1 and 2, we see that Bias is less in case of *Matlab* bootstrapping.

CASE 2 Air Quality data

A number of statistical computer application systems including R provide multi-panel conditioning plots (or *coplots*) under the name Trellis plots, which is a scatter plot of two variables (solar.R and ozone) conditional on the value of a third variable (wind and temperature) known as the or conditioning variable as shown in Fig. 3.

We use following ***R*** commands to produce conditioning plots.

```
> data(airquality)
> coplot(Ozone ~ Solar. R | Temp * Wind, number = c(4, 4), data = airquality, pch
= 21, col = "goldenrod", bg = "goldenrod")
```

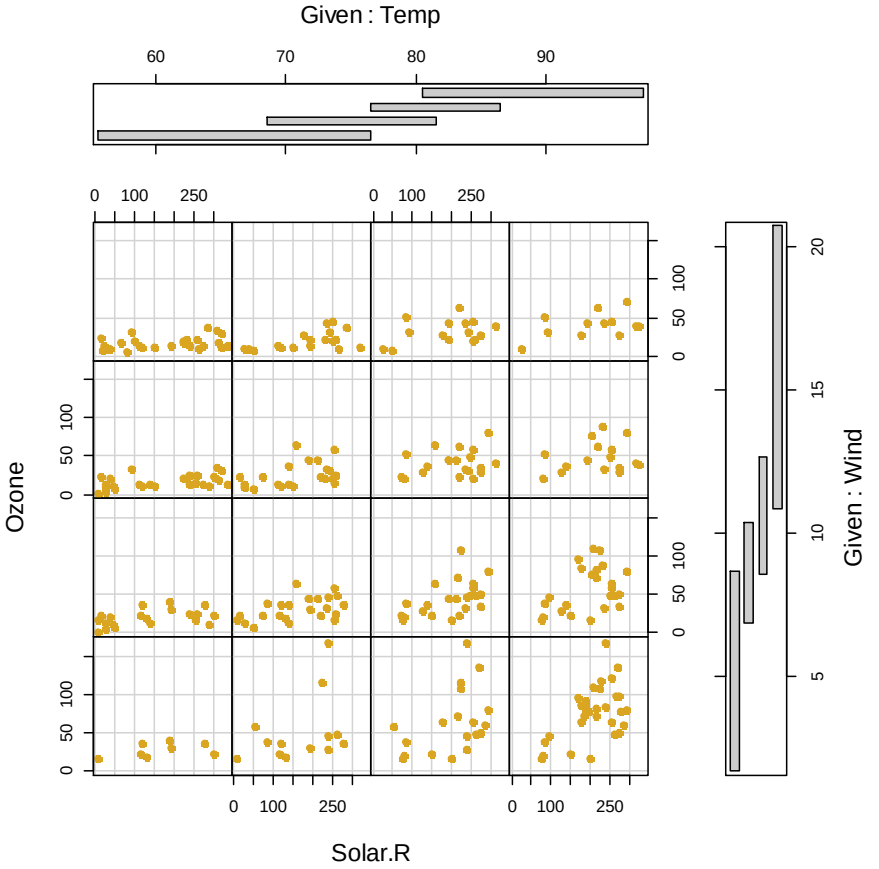


Fig. 3: Conditioning plot of solar radiation and ozone level conditional on wind and temperature values.

Model fitting and Simulation of "airquality" data with R.

```
> data(airquality)
> pairs(airquality,panel=panel.smooth)
> g <- lm(Ozone ~Solar.R + Wind + Temp,airquality)
> summary(g)
```

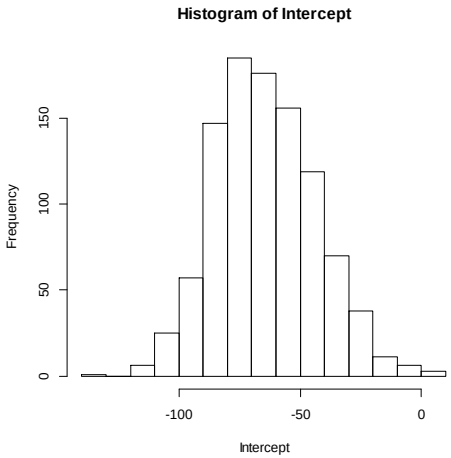
Coefficients: (p-values show that fitted regression model is adequate):

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-64.34208	23.05472	-2.791	0.00623
Solar.R	0.05982	0.02319	2.580	0.01124
Wind	-3.33359	0.65441	-5.094	1.52e-06
Temp	1.65209	0.25353	6.516	2.42e-09

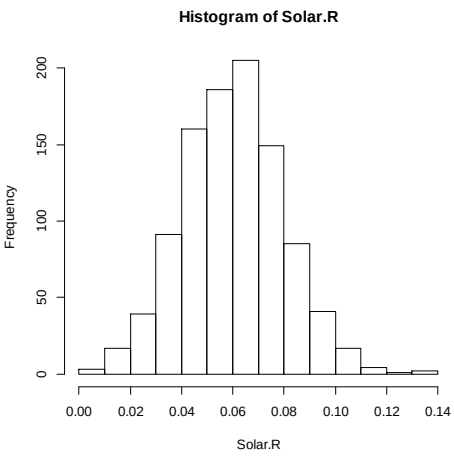
Multiple R-squared: 0.6059, Adjusted R-squared: 0.5948
F-statistic: 54.83 on 3 and 107 df, p-value: < 2.2e-16.

Bootstrapping a Regression Model:

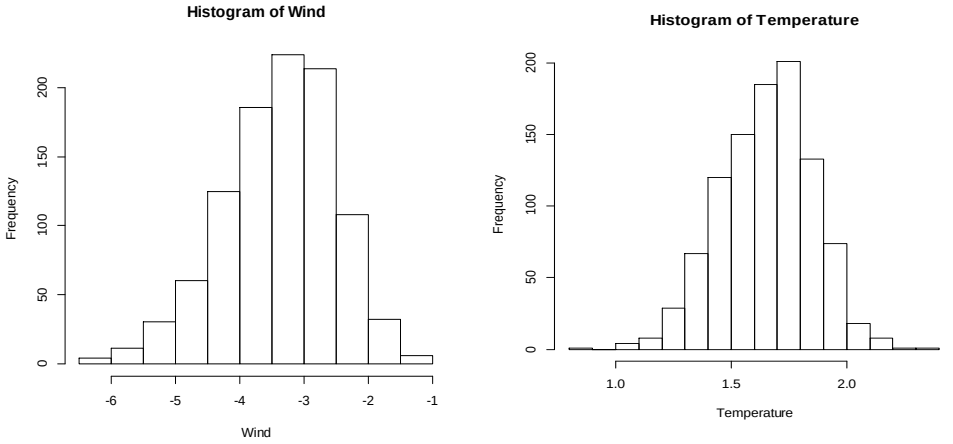
```
> require(simpleboot)
>data(airquality)
>attach(airquality)
>lmodel <- lm(Ozone ~ Solar.R + Wind + Temp, data = airquality)
>lboot <- lm.boot(lmodel, R = 1000)
## Get sampling distributions for coefficients
>s <- samples(lboot, "coef")
>Intercept <-s[1,]
>Solar.R <-s[2,]
>Wind <-s[3,]
>Temperature <-s[4,]
## Histogram for the intercept,b1,b2,b3
>hist(Intercept)
>hist(Solar.R)
>hist(Wind)
>hist(Temperature)
```



(a)



(b)



(c) (d)

Fig. 4: (a, b, c, d) Simulation of coefficients of air quality model shows that the mean values of these distributions are very close to the estimated coefficient values.

<u>Case Study 3</u>: Simulation of Business data:

Here, time series data of Pakistan Federal Excise collection during the year 1973 to 1993 is analyzed for appropriate model fitting. Time series plot of the data is shown in Fig. 5.

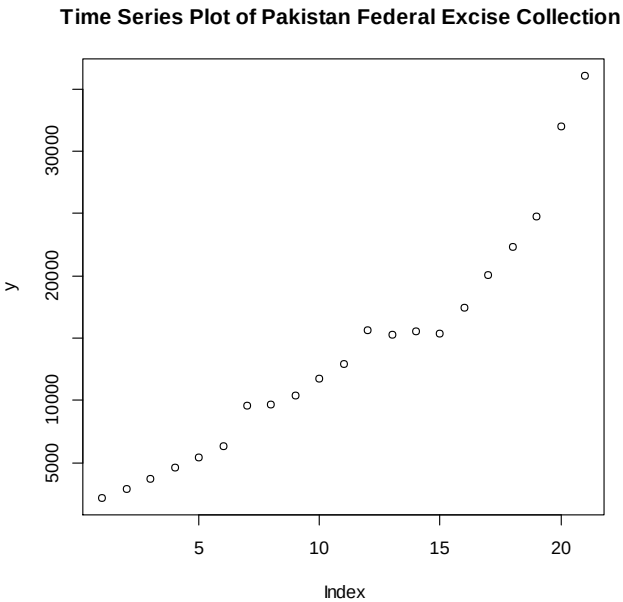


Fig. 5: Time series plot of Pakistan Federal Excise collection revealing presence of non-stationarity in the series.

Now, we use *Matlab* and *R* for fitting of *ARIMA* (*p*, *d*, *q*) models to Pakistan Federal Excise Data*.

* Several issues of CBR books and Budget Mamorandum

R-commands:

```
> require(astsa)
> xx <- c(2211, 2895, 3670, 4585, 5429, 6299, 9616, 9701, 10413, 11743, 12911,
    15652, 15313, 15515, 15361, 17399, 20038, 22341, 24740, 32015, 36021)
> sarima(xx, p=1, d=3, q=1, P =0, D = 0, Q = 0, S = 0, details = TRUE,tol =
    sqrt(.Machine$double.eps), no.constant = TRUE)
```

Following models were fitted to above data.

ARIMA (3,2,1), *ARIMA* (3,1,1), *ARIMA* (2,1,1), *ARIMA* (2,2,1), *ARIMA* (2,2,0), *ARIMA* (2,2,0),

ARIMA (1,2,0), *ARIMA* (1,2,1), *ARIMA* (1,3,1), *ARIMA* (1,1,0), *ARIMA* (0,1,1). *Based on Akaike information criterion (AIC)* one can identify appropriate *p*, *d*, and *q* values of the fitted *ARIMA* model [21]. The smallest value for *AIC* results for *ARIMA* (1, 3, 1).

Matlab used ARIMAAnalyzer,m program to sift out the optimum *ARIMA* model on the basis of *AIC* [22].

Summary of *ARIMA* (1, 3, 1) Model for Pakistan Federal Excise data: Time Series XX

R Output:
Coefficients:

	ar1	**ma1**
	-0.5182	-1.0000
s.e.	0.2130	0.4252

Figure 6 shows standard residuals, Autocorrelation Function (ACF), Normal QQ plot, and p-values of estimated parameters of Pakistan Federal Excise collection data model. By doing analysis of these graphs one can say that the fitted model is appropriate.

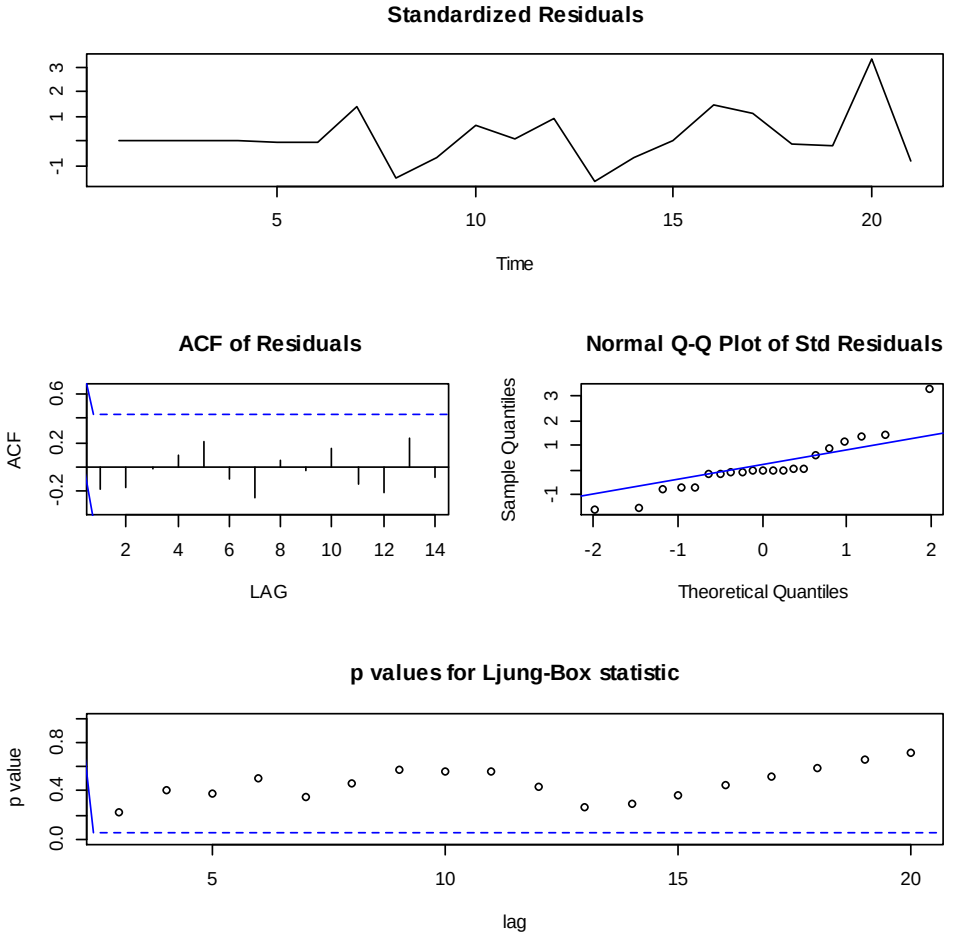


Fig. 6: Standard residuals, Autocorrelation function (ACF), Normal QQ plot, and p-values of estimated parameters of Pakistan Federal Excise collection data model reveal that the fitted model is adequate.

Now, we explain use of *R* in simulating *ARIMA*(1,3,1) model.

ARIMA Simulation using R:

```
> x = arima.sim(list(order=c(1,3,1), ar=-0.5182, ma=-1.00), n=30)
> plot(x, main=(expression(ARIMA(1,3,1))))
```

Figure 7 shows the above simulation result.

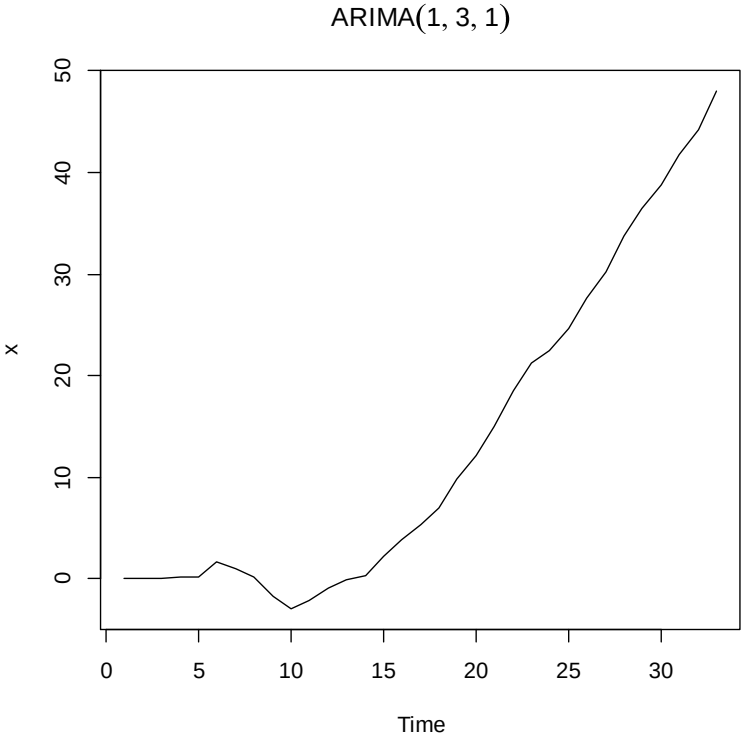


Fig. 7: Simulation of 30 Realizations of Pakistan Federal Excise Collection

Now, we conclude this study.

5. CONCLUSION

Matlab and *R* computer applications were used to guide students, researchers and teachers for doing data analysis. After brief introduction of basic *Matlab* and *R* commands, bootstrap Monte Carlo simulation approach was explained to reduce bias in standard errors of estimated population parameters, which actually helped in empirical assessment of environmental and business data model parameters through simulation. It has been emphasized that students and teachers should know use of appropriate commands in *Matlab* and *R* to obtain better results while performing analysis on data at hand. Bootstrapping of regression parameters using *Matlab* and *R* showed that simulation performed by *Matlab* was slightly better than *R*, while time series data analysis using both computer applications revealed that Pakistan federal excise data followed *ARIMA*(1, 3, 1) model. ***R*** simulation command was also used to simulate 30 realizations of above model. So, we can say that technology can be used to help users in getting more insights of the processes generating data. Further discussion on use of these computer applications to solve problems related to financial risks will be the topic of future paper.

REFERENCES

1. Cameron, A.C. and Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*, Cambridge University Press, New York, pp. 12.
2. Chernick, M.R. (2003). *Introductory Biostatistics for the Health Sciences: Modern Applications Including Bootstrap*, John Wiley, New Jersey, pp. 360.
3. Crawley, M.J. (2013). *The R Book*, Wiley, UK, pp. 224.
4. Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, London.
5. Enders, W. (2004). *Applied Econometric Time Series 2/e*, John Wiley & Sons, Asia, pp. 69.
6. Hussain M.A. (2008). *Mathematical Aspects of Impact of Urban Greenhouse Gas Emissions on Global Warming*, Ph.D. Thesis, Federal Urdu University, Karachi, Pakistan.
7. Hussain M.A., Khan, M.A. and Abbas, S. (2014). The Impact of Technology in Mathematics Education: A. Case Study of Pakistan. *J. of Basic and Applied Sciences* (*Accepted*).
8. Hussain, M.A. (2013). On the Solution of Economic Models Using Computer Application *Matlab*. *J. of Business Strategies*, 7(2), 29-38.
9. Hussain, M.A. and Rahim, I. (2012). Teaching Optimization Techniques using Matlab and Mathematica: A Comparative Study. *Journal of Information & Communication Technology*, 6(1), 18-26.
10. Klemens, B. (2009). Modeling with Data: *Tools and Techniques for Scientific Computing*, Princeton University Press, Princeton and Oxford, pp. 27.
11. Landau, D.P. and Binder, K. (2009). *A Guide to: Monte Carlo Simulations in Statistical Physics 3/e*, Cambridge University Press, New York, pp. 415.
12. Marques, de Sá J.P. (2007). *Applied Statistics using SPSS, STATISTICA, MATLAB and R 2/e*, Springer Berlin Heidelberg, New York, pp. 415.
13. Marsaglia, G. and Zaman, A. (1991). A new class of random number generators, *Annals of Applied Probability*, 3, 462-480.
14. Shumway, R.H., Azari, R.S., and Kayhanian, M. (2002). Statistical Approaches to Estimating Mean Water Quality Concentrations with Detection Limits. *Environ. Sci. Technol.*, 36(15), 3345-3353.
15. stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html (airquality data Accessed Mar 03, 2014).
16. Trauth, M.H. (2006). *MALAB Recipes for Earth Sciences*, Springer-Verlag Berlin Heidelberg, pp. 74.
17. Watson, R.T., Zinyowera, M.C. and Moss, R.H. (1999). *The Regional Impacts of Climate Change: An Assessment Vulnerability*, Cambridge University Press, Cambridge.
18. http://www.chrisrackauckas.com/publications.html (accessed March 02, 2014).
19. www.mathworks.com/help/pdf_doc/matlab/matlab_prog.pdf (Accessed Feb 28, 2014)
20. *www.sagepub.com/upm-data/21122_Chapter_21.pdf* (Duncan Refression, Accessed Mar 03, 2014).

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 47-52

EMPIRICAL ASSESSMENT OF THE EFFECTS OF ATMOSPHERIC HUMIDITY ON CLOUD COVER PAKISTAN UNDER LOWER ATMOSPHERE

Faisal Ahmed Khan Afridi[1] and **M. Ayub Khan Yousufzai[2]**
Department of Applied Physics and Solar-Terrestrial & Atmospheric Research Wing,
Institute of Space & Planetary Astrophysics, University of Karachi, Karachi, Pakistan
Email: [1]fakafridi@hotmail.com, [2]ayubzai@yahoo.com

ABSTRACT

In this communication, we have presented the empirical assessment of the effects of atmospheric relative humidity on cloud cover at the Quetta region, Pakistan using dataset on a monthly basis from 1980 to 2009. Relative humidity is an important variable in the study of the atmospheric dynamics and ecosystem. Clouds are also one of the most significant constituents of our atmosphere. Clouds deliver water from the atmosphere to the Earth's surface as rain or snow and are thus a key step in the hydrologic cycle.

To study the behavior and association of relative humidity on cloud cover at the Quetta region, we have adopted approaches such as Exploratory Data Analysis (EDA). It has been observed that the relative humidity has a significant and positive correlation on cloud cover at the Quetta region. In order to strengthen our claim, we have compared the data sets of relative humidity with cloud cover at the Quetta region and analyzed both the data sets that exhibit positive correlation along with the interpretations.

KEYWORDS

Clouds, cloud cover, humidity, relative humidity, climatology, correlation, regression.

1. INTRODUCTION

Being a source of precipitation, clouds have two climatic roles to play that they are themselves a highly visible part of climate and they significantly modify the local radiation streams. Therefore, influence through the energy balance, the whole climate [1].

Water vapor is the single most important trace species in the atmosphere. Water vapor is confined almost exclusively to the troposphere [2]. The amount of water vapor present in the air is called the humidity of the air. Warm air can hold more humidity than cold air. The humidity of the air is proportional to the temperature [3]. The proportions of dry air, water vapor and liquid or solid water vary within rather wide limits in the atmosphere as results of evaporation, condensation and precipit
ation. The amount of water vapor in the air is not a constant, but undergoes large fluctuations. There is an upper limit the moisture content normally will not exceed. Like other gases, water vapor exerts a pressure, called vapor pressure. Air is considered saturated when its vapor pressure equals that directly over a level water surface. The saturation vapor pressure is a quantity that depends on temperature. The saturation vapor pressure changes very slowly at low temperatures and more and more swiftly at high

temperatures. Water vapor occupies an especially important role in atmospheric science [4].

The quantity commonly stated in a weather report is relative humidity at or very near the ground, but usually vary with altitude because both water vapor pressure and temperature change with distance in the vertical. This fact is important for the formation of clouds and precipitation. The relative humidity of the ground can be less than 100%, but a couple of kilometers higher the air can be saturated. As a consequence, clouds are usually located above the ground. When the relative humidity of the ground reaches 100%, a fog may develop [6]. A necessary condition for the cloud formation is the increase in the relative humidity of an air parcel to a value exceeding 100%. This relative humidity increase is usually the result of cooling of a moist air parcel. Even if the water mass inside the air parcel does not change, its saturation water vapor concentration decreases as its temperature decreases and therefore its relative humidity increases [5]. The relative humidity (R.H) is the ratio of the actual water vapor pressure (w) to the saturation vapor pressure (w_s). It is generally expressed as a percentage, such that

$$R.H = (w/w_s) \times 100\% \text{ [1,7,8]}$$

Relative humidity is an important parameter in the study of the environment of plants and animals because it is a rough indicator of the rate of evaporation from animal or vegetable tissue [7]. It may seem strange that human comfort depends on relative humidity rather than the actual moisture content of the air. When the relative humidity is above 80%, even air with low moisture content feels sweaty. On the other hand, low relative humidity leads to increased evaporation and this is very important in warnings of forest-fire danger [4].

In the present work, a systematic study of the empirical assessment of the effects of atmospheric relative humidity on cloud cover at the Quetta region has been carried out for a longer period, 1980 to 2009. The significance of this study is enormous because relative humidity is an important variable in the study of the atmospheric dynamics and ecosystem and where clouds are also one of the most significant constituents of the atmosphere.

2. METHODOLOGY

The relative humidity and cloud cover datasets were attained by in-situ observations under the World Meteorological Organization (WMO) standards, by Pakistan Meteorological Department, Govt. of Pakistan. The co-relational study of our data analysis on relative humidity and cloud cover for duration 1980 to 2009 consists of Exploratory Data Analysis (EDA) and inferential Statistics using standard Statistical softwares such as Minitab and DataFit. The statistical approaches, we have adopted are well within the feasibility and approach of available computer softwares. By using various parameters such as regression analysis, best model equation, coefficient of multiple determination, residual analysis, variance analysis, model plot and residual normal probability plot, we have analyzed the assessment of the effects of atmospheric humidity on cloud cover at Quetta region, Pakistan.

3. RESULTS AND DISCUSSION

Scatter plots are particularly useful tools in an exploratory analysis, conveying information about the association between X and Y, the clustering of the points, the presence of outliers, etc [9]. Regression shows a relationship between the average values of two variables. The regression is very helpful in estimating and predicting the average value of one variable for a given value of the other variable [10]. The polynomial can be used to describe many data distributions. The order of the polynomial is given by the highest power of x and the best-fitting curve is described by the coefficients a_o, a_1, etc [11]. The Fig. 1 shows the best model is tenth order polynomial out of 249 non-linear continuous distributions. This illustrates that there is a positive correlation between relative humidity and cloud cover at Quetta region during 1980-2009. In Fig. 1, it also exemplified few outliers in the data.

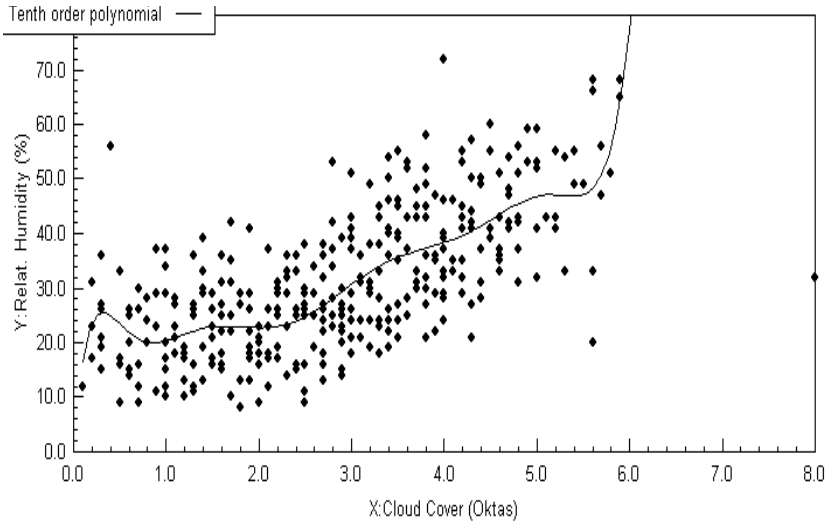


Fig. 1: Model Plot of Tenth Order Polynomial

Coefficient of Multiple Determination (R^2):

The coefficient of determination, R^2, expresses the proportion of y-variation explainable by the regression line and it is generally expressed as a percentage. We may test also the significance of the coefficient of determination, R^2, by the use of F-statistic. R^2 may be expressed as [11]:

$$R^2 = \frac{Variability\,\exp lainedbyregressionline}{Totalsample\,\mathrm{var}iability} = \frac{SS_{yy} - SSE}{SS_{yy}}$$

(SS_{yy}) = Sums of the squares of deviations from mean - y

(SSE) = Sums of the squares of deviations from predicted value $\hat{y}$ (errors)

The range value of R^2 may be from zero to unity. In Table I, the coefficient of determination, R^2 = 50%, which indicates that there is significant correlation between

relative humidity and cloud cover at Quetta region. The value of the Durbin-Watson Statistic is within the limit. The Table II shows all the coefficient values of the model equation. The computed value of Y at X is 2.032 while X is 1.0. The best model equation between relative humidity (Y) and cloud cover (X) of Quetta region is:

Best Model Equation:

$$Y = -0.008x^{10} + 0.273x^{9} - 3.947x^{8} + 31.733x^{7} - 155.790x^{6} + 481.933x^{5} - 935.297x^{4} + 1098.182x^{3} - 719.163x^{2} + 221.939x + 0.466$$

Table I
Statistical parameters of monthly mean total cloud cover with relative humidity at the Quetta region during 1980-2009

Parameters	Values
Number of Observations	360
Coefficient of Multiple Determination (R^2)	50%
Minimum Residual	-28.250
Maximum Residual	33.615
Residual Sum of Squares (Absolute)	29956.804
Stand. Error of the Estimate	9.264
Durbin-Watson Statistic	1.00

Table II
Regression Variable Results

Variable	Value	Standard Error	t-value	Prob. (t)
a	-0.008	0.003	-2.501	0.012
b	0.273	0.109	2.493	0.013
c	-3.947	1.590	-2.481	0.013
d	31.733	12.891	2.461	0.014
e	-155.790	64.095	-2.430	0.015
f	481.933	202.223	2.383	0.017
g	-935.297	404.480	-2.312	0.021
h	1098.182	497.107	2.209	0.027
i	-719.163	350.060	-2.054	0.040
j	221.939	122.919	1.805	0.071
k	0.466	15.572	0.029	0.976

Residual Analysis:

Residuals from any model are helpful in evaluating the adequacy of the model itself relative to the data and any assumptions might make in the analysis. A residual is simply the difference between the observed value of the response variable (y) and the value of "y" predicted by the model. There is one residual for each observation. Least squares residuals can be used to investigate the assumption of normality of the error terms [9]. The residuals also provide a useful means of assessing the validity of attempting regression [11].

The normal probability plot of the residuals is constructed to check on the normal distribution assumption of the residuals. In the normal probability plot, the residuals are ordered by size and plotted versus the normal probability scale [12-13]. In the Fig. 2, the plot approximates a straight line but some departure at the bottom tail, so it illustrates that the residuals follow a normal distribution, which is reasonably satisfied. The normality plot of the residuals is pretty much linear. In the Fig. 2, there are few outliers that suggest the observation is due to measurement error or keying mistake.

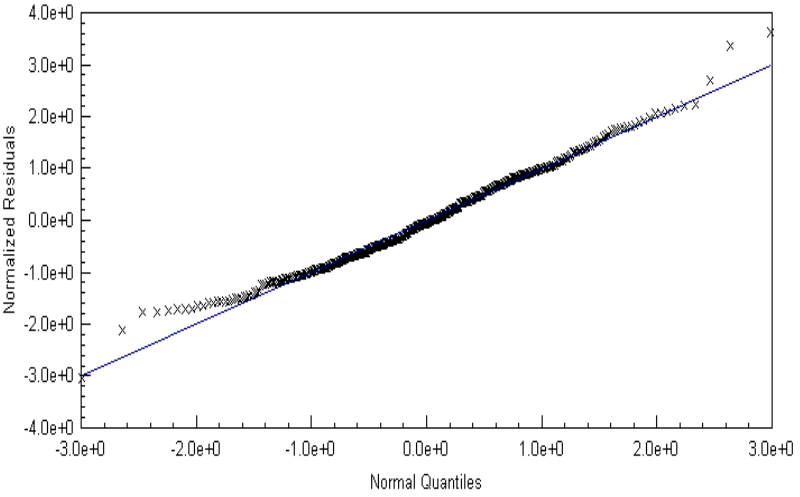


Fig. 2: Residual Normal Probability Plot

6. COMMENTS AND CONCLUSION

In this study, different parameters for relative humidity and cloud cover of Quetta region during 1980 to 2009 have been computed. The best model equation between relative humidity (Y) and cloud cover (X) of Quetta region is tenth order polynomial out of 249 non-linear continuous distributions. The relative humidity has a significant and positive correlation on cloud cover at the Quetta region. It has also illustrated few outliers in the data set. The coefficient of determination, R^2 is 50%, which indicates that there is significant correlation between relative humidity and cloud cover. The residual normal probability plot approximates a straight line but some departure at the bottom tail, which means that the residuals follow a normal distribution. The normality plot of the residuals is much linear in behavior.

ACKNOWLEDGEMENTS

The authors are grateful to the organizers of the 12th International Conference on Statistical Sciences, jointly organized by the ISOSS, Dow University of Health Sciences and DHA Suffa University, Karachi, Pakistan, held on March 24-26, 2014. The authors are also thankful to the officials of Pakistan Meteorological Department, Govt. of Pakistan, Karachi, Pakistan, who provides us the climatological data.

REFERENCES

1. Robinson, P.J. and Henderson-Sellers, A. (1999). *Contemporary Climatology*. 2nd ed. Essex, England, Longman Group UK Ltd., 98-120.
2. Salby, M.L. (1996). *Fundamentals of Atmospheric Physics*. California, USA, Academic Press, 25-26.
3. Welch, R. (2013). *A Change in the Weather*, Ice Cap Publishing, San Francisco, USA, 152-170.
4. Riehl, H. (1972). *Introduction to the Atmosphere*. 2nd ed. New York, USA, McGraw-Hill Inc., 63-90.
5. Seinfeld, J.H. and Pandis, S.N. (2006). *Atmospheric Chemistry & Physics: From Air Pollution to Climate Change*. 2nd ed. New Jersey, USA, John Wiley & Sons Inc., 13-777.
6. Hartmann, D.L. (1994). *Global Physical Climatology*. California, USA, Academic Press Inc., 232-233.
7. Fleagle, R.G. and Businger, J. A. (1963). *An Introduction to Atmospheric Physics*. New York, USA, Academic Press Inc., 64-65.
8. Frederick, J.E. (2008). *Principles of Atmospheric Science*. Sudbury, USA, Jones & Bartlett Publishers Inc., 83-84.
9. Du Toit, S.H.C., Steyn, A.G.W. and Stumpf, R.H. (1986). *Graphical Exploratory Data Analysis*. New York, USA, Springer-Verlag, 177-199.
10. Sumeet, A. (2010). *Comprehensive Statistical Methods*. New Delhi, India, S. Chand & Co. Ltd., 7.4-8.2.
11. Borradaile G. (2003). *Statistics of Earth Science Data: Their Distribution in Time, Space and Orientation*. Berlin, Germany, Springer-Verlag, 133-163.
12. Makridakis, S.G., Wheelwright, S.C. and McGee, V.E. (1997). *Forecasting: Methods and Applications*. Singapore, John Wiley & Sons Inc., 203-222.
13. Mickey, R.M., Dunn, O.J. and Clark, V.A. (2004). *Applied Statistics: Analysis of Variance and Regression*. 3rd ed. New Jersey, USA, John Wiley & Sons Inc., 43-47.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 53-60

SPECTRAL ANALYTICAL PERSPECTIVE OF SOLAR FLARE

Saifuddin Ahmed Jilani[1] and **M. Ayub Khan Yousuf Zai**[2]
[1] Department of Physics, University of Karachi,
Karachi, Pakistan. Email: sajilani@uok.edu.pk
[2] Institute of Space and Planetary Astrophysics (ISPA),
University of Karachi, Karachi, Pakistan.
Email: ayubzai@yahoo.com

ABSTRACT

The most energetic natural phenomenon on the solar disc is solar flare that is responsible for ozone layer depletion. Solar flares are closely related with sunspot cycle but it is important to find the hidden periodicities in both the data series separately because sometimes major flares are produced when sunspots are decaying. These flares adversely affect space weather including ozone layer depletion, malfunctioning in orbiting satellite and radio wave blackout. The tumultuous behavior of solar flare generates coronal mass ejection (CME) that is the stream of solar proton which often associated with gradual flares. By using other statistical techniques we have little insight into different types of signals that are blended together to make the recorded data. The technique of spectral analysis examines data series in terms of their frequency and time contents and explores the hidden periodicities. In addition to the mixed series of solar flares this communication evaluates their impulsive and gradual phases as well. This study further examines the activity of Solar Flares in terms of its duration corresponding to the series of ozone and Sun-spot number (SSN).

KEY WORDS

Fast Fourier transformation (FFT), Spectral analysis, Ozone Layer Depletion (OLD), Solar Flare Duration (SFD), Periodogram, Spectral density, Fourier coefficients, Sun-spot number (SSN).

INTRODUCTION

The general mechanism of solar flare involves the reconnection of magnetic field lines. The sun's tumultuous magnetic fields provide the fuel of flares. The sudden release of energy in a flare result from a process called reconnection, whereby oppositely directed magnetic field lines come together and partially annihilate each other. Scientists generally agree that the energy released in a flare must be stored in the sun's magnetic fields. Flares erupt from the part of the sun called active regions, where solar magnetic fields are much stronger than average. These areas are most easily identified by the presence of sunspots. These dark-looking patches host the most intense magnetic field on the sun. The Sun's dark spot are due to magnetic fields generated in the interior emerging on the surface. When these fields get tangled and reconnect like shorted electrical wires,

sudden explosion or solar flare results. Unlike regular sunlight, flares emit strongly in the radio, X-ray and γ -ray bands [1-6].

The most serious effects on our atmosphere occur during major geomagnetic storms. It is now understood that the major geomagnetic storms are induced by coronal mass ejection (CMEs). Coronal mass ejections are usually associated with flares, but sometimes no flare is observed when they occur. Like flares CME's are most frequent during the active phase of the sun's approximately 11 year cycle [7].

These flares occur when a great number of twisted tubes of magnetic field lines release their energy at once through the photosphere. Extremely energetic particles are blasted out of the Sun, reaching the Earth in about three days. They are responsible not only for spectacular auroras but also for disruption of earthly radio transmission. Normally, the Earth's magnetic field prevents particles of the solar wind from reaching the ionosphere by deflecting and trapping them, but the high-energy particles in the solar wind are able to penetrate to the ionosphere. When the ionosphere is disrupted by the solar wind, we may experience static in radio reception or even complete loss of signal [8-9].

METHODOLOGY

The methodology is the time series statistical techniques in analyzing flare behavior, to samples of flares. The first analytical techniques developed was Periodogram analysis, which is based on "finite Fourier series fit" to the data points. Although the periodogram itself has certain defects, it continues to occupy an important role in the analysis of stationary time series. Determination of the Fourier components of a time series can be used to determine a Periodogram that can be used to define Spectral Density (Spectrum) of the time series.

Through the use of Fourier transform, it enables us to get a better understanding of periodic signals through their frequency spectra but as aperiodic signal as well. In general, a periodic function contains the fundamental sinusoid, and a large number of harmonics. If we refer to the exponential form of the Fourier series then C_n, which is complex number, contains the information about the magnitude and phase of each component of the wave form. If $|C_n|$, the magnitude of Cn, is plotted against nw_o, we obtain the ***frequency spectrum*** of the magnitude of the periodic function. The spectrum of a magnitude of a periodic function is called its frequency spectrum. The value of 'C_n' have been shown only for positive 'n' and hence only for positive value of w. This is sufficient because for negative 'n' we get conjugate of C_n, with the same magnitude. It should also be noted that the frequency components are spaced w_o apart, although some of the components (even harmonics) have zero amplitude. Since the spectrum consists of discrete lines, it is called a discrete spectrum. This is characteristic of all periodic function. Since $C_n = (a_n - jb_n) / 2$, so that for an odd waveform containing only sine terms, a_n must be zero .Similarly, for even waveform, b_n, must be zero. Drichlet condition indicates that a periodic function need not be continuous for its Fourier series to exist [10].

A fundamental objective of spectral analysis is to identify the dominant frequencies in a series and to find an explanation of the system from which the measurements were derived. It is the tendency of observed data to show periodic kinds of fluctuations that justifies the use of frequency domain methods. One of the main goals of spectral analysis

is the estimation of periodogram and its counterpart called spectral analysis. The time domain approach may be thought of as regression of the present on the past, whereas the frequency domain approach may be considered as regression of the present on periodic sines and cosines [11].

Since each pair of coefficients (A_p, B_p) is associated with a frequency ω_p (or f_p), the amplitudes of the coefficients provide a measure of the relative importance of each frequency component to the overall signal variability. For example, if $(A_6^2 + B_6^2)^{1/2} >> (A_2^2 + B_2^2)^{1/2}$ we expect there is much more "spectral energy" at frequency ω_6 then at frequency ω_2. Here, spectral energy refers to the amplitudes squared of the Fourier coefficients which represent the variance, and therefore the energy, for that portion of the time series. To determine the energy distribution within a time series y (t), we need to find its Fourier transform. That is, we need to determine the coefficients A_p, B_p in the Fourier series [12].

The confidence interval is used to find the significance of spectral density function for the highest peak of each of the data sets. This will assure us that whether the value of the spectral density for white noise falls outside the interval. The 100(1–α) % confidence interval for f (ω) is given by,

$$\frac{\nu\hat{f}(\omega)}{\chi^2_{\nu,\alpha/2}} to \frac{\nu\hat{f}(\omega)}{\chi^2_{\nu,1-\alpha/2}}$$

The degree of freedom for the Hamming window is 2.50 (N/M) respectively. A useful rough guide is to choose M to be about $2\sqrt{N}$ [13].

DATA ANALYSIS

Often, non-stationary trends are present that should be eliminated before computing the periodogram. Trends introduce extremely low frequency components in the periodogram that tend to obscure the appearance at higher frequencies.

Figure 1 depicts a stationary series otherwise the partial autocorrelation decays exponentially. To be a series to stationary the spectral density cannot change with time. For the case of SFD the first half of a time series had component oscillations of f = 0.0123 with respect to the highest peak and for the second half this frequency found to be uniform. Hence this indicates stationary series. Figure 2 and 3 also shows stationary behaviors since both half of the series are uniform.

Figures 4-8 depicts periodicities of each of the series in years per cycle.

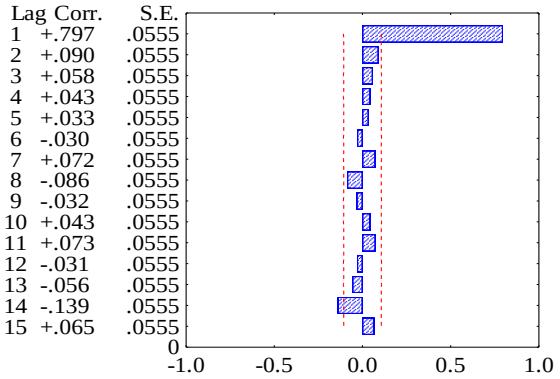


Fig. 1: Partial Autocorrelation Function for the series of SFD of mixed category of flares exhibit stationary behavior.

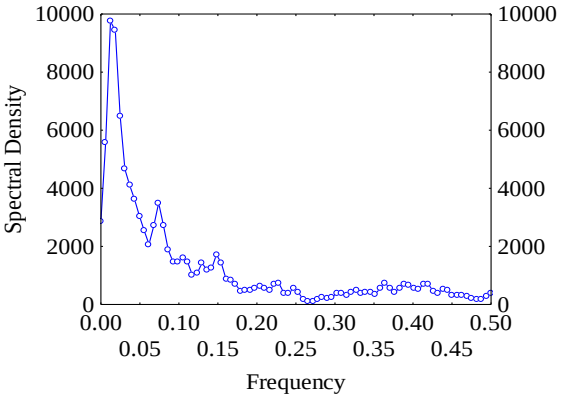


Fig. 2: Spectral analysis of the first half of the series of SFD consisting of first 162 cases.

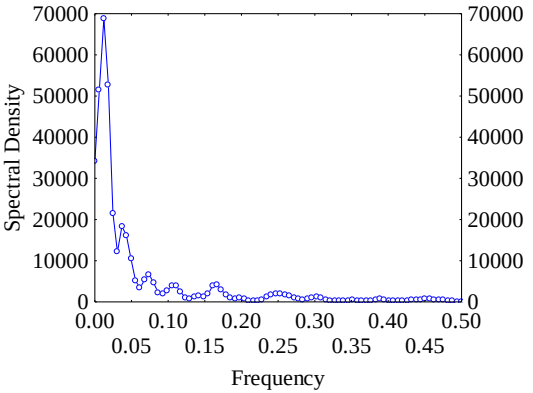


Fig. 3: Spectral analysis of the second half of the series of SFD consisting of next 162 cases.

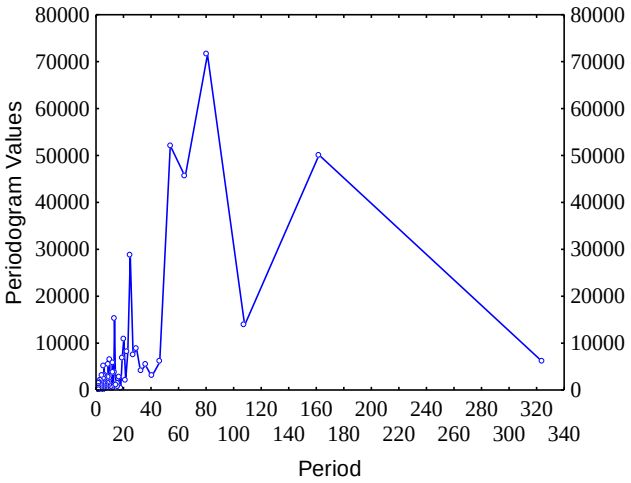


Fig. 4: Periodogram of SFD for mixed series of flares in time domain

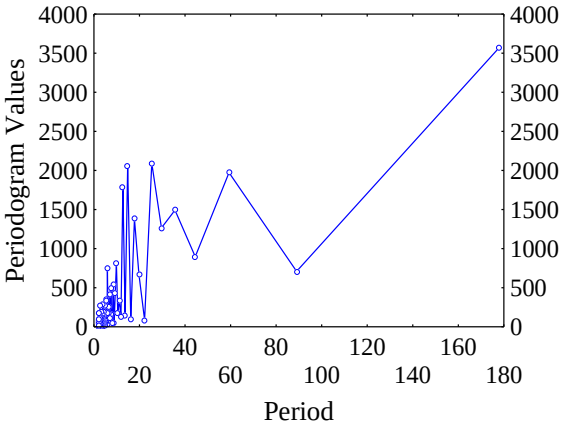


Fig. 5: Periodogram for the series of impulsive flares in time domain

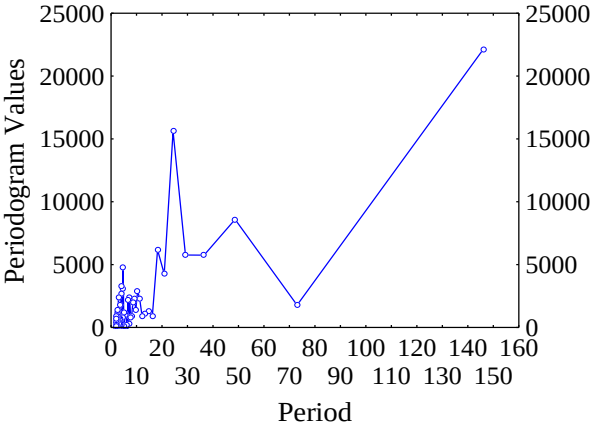


Fig. 6: Periodogram for the series of gradual flares in time domain

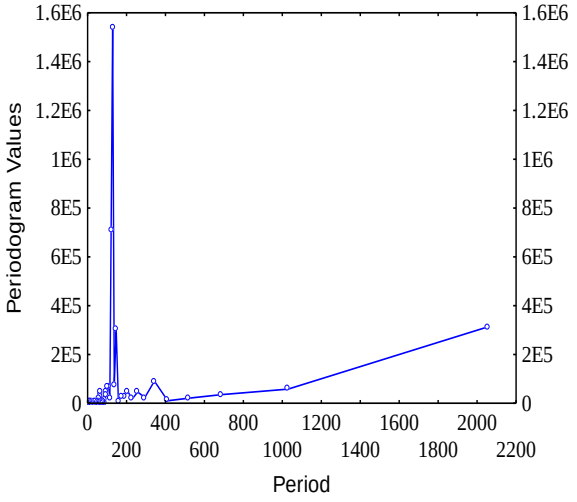


Fig. 7: Periodogram for the series of SSN in time domain

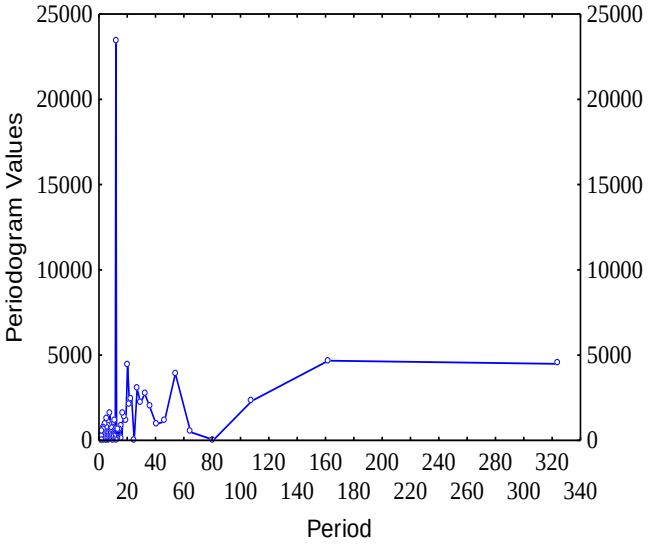


Fig. 8: Periodogram for the series of ozone in time domain

Table 1
Periodogram analysis of different data sets

Data Series	Period	Periodogram Value	Sine Coeff.	Cosine Coeff.
SFD-Mixed	6.75	71512.99	12.81	- 16.64
SFD-Imp	14.83	3572.21	1.244	6.211
SFD-Grad	12.16	22115.28	2.867	17.16
OLD	1.0	23384.45	7.076	9.709
SSN	11.91	653542.8	- 9.168	61.53

Table 2
Confidence Intervals depicted with Hamming window

Data Series	Frequency	C. I
SFD-Mixed	0.0123	0.0074-0.0242
SFD-Imp	0.0056	0.0031-0.0125
SFD-Grad	0.0068	0.0037-0.0162
OLD	0.0833	0.0503-0.1639
SSN	0.0069	0.0040-0.0140

CONCLUSION

It concludes that in all of the five data series analyzed, the series of gradual flares have periodicity very much close to that of the periodicity of sunspot number (SSN). This implies that gradual flares are closely associated with sunspot dynamics. Their frequencies and their periodicities are very much in accordance with that of sunspot numbers. Hence gradual flares in terms of their duration may be one of the alternate of sunspot numbers in the study of solar activity.

REFERENCES

1. Subba T.R., Priestley, M.B. and Lessi, O. (1997). *Application of Time Series Analysis in Astronomy and Meteorology*, Chapman & Hall.
2. Hong Don, Wang Jianzhong, Robert Gardner (2005). *Real Analysis with an Introduction to Wavelet and Applications*. Academic Press, Elsevier Inc.
3. May-Britt Kallenrode (2003). *An introduction to plasmas and particles in the heliosphere and magnetosphere*. Third, Enlarged Edition, and Springer Verlag Berlin Heidelberg New York.
4. Herman John R. and Richard A. Goldberg, (1997). *Sun, Weather and Climate*, Dover Publications, Inc. New York.
5. Karl F. Kuhn (1998). *In Quest of the Universe*, 2nd Edition, Jones and Bartlett Publishers, Sudbury.
6. Benestad, R.E. (2003). *Solar Activity and Earth's climate*, Springer-Parxis, London.
7. Kenneth JH Phillips (1992). *Guide to the Sun*. Cambridge University Press.
8. Robert C.W. and Sheo, S.P. (1985). *Ozone in the free Atmosphere*, Van Nostrand Reinhold Company Inc. New York.
9. Fynman, J. and Hundhausen, A.J. (1994). Coronal mass ejection and major solar flares: the great active center of March 1989, *J. Geophys*.
10. Nareshk Sinha (1991). *Linear Systems*, John Wiley & Sons. Inc.
11. Robert H.S. and David S.S. (2006). *Time Series Analysis and Its Applications*, Second Edition, Springer Science.
12. William J.E. and Richard, E.T. (2004). *Data Analysis Methods in Physical Oceanography*. Second & Revised Ed. Elsevier B.V Sara Burgerharstraat 25 Amsterdam, The Netherlands.
13. Chatfield, C. (1989). *Analysis of Time Series*, Chapman & Hall, London.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 61-68

PATTERNS OF AGE-SPECIFIC FERTILITY RATES OF PAKISTAN REGION-WISE

Hira Fatima Waseem[1] and **Farah Yasmeen**[2]
Department of Statistics, University of Karachi, Karachi, Pakistan
Email: [1]hira.waseem@live.com and [2]riazfarah@yahool.com

ABSTRACT

The Pakistan population is broadly distributed into two main regions, urban-rural areas. The purpose is to study the pattern of age-specific fertility rates (ASFRs) of Pakistan for these two regions: urban and rural separately. Modeling age-specific fertility rates instead of total fertility rates gives more accurate forecasts and also provides clear differences between age groups in contribution to overall variability. We fit functional time series (FTS) model to the secondary data of age-specific fertility rates region wise from 1984 to 2007, obtained from Pakistan Demographic Surveys (PDS). Firstly, we smoothed data by using non-parametric smoothing methods, before applying FTS models. Secondly, we employ weighted regression B-splines with a concavity constraint with fixed number of knots and a concavity constraint. Finally we fit a model with four basis functions, which explain the major modes of variations and obtain the fertility forecasts. We also obtain the 80% prediction intervals for one-year, ten-year and twenty years fertility forecasts.

KEY WORDS

Urban and rural, age-specific fertility rates, fertility pattern, functional time series, prediction intervals.

1. INTRODUCTION

Fertility is the second principal vital event and most important indicator for national and international demographers. Fertility rates give researcher a glimpse into the future of society and its population. Pakistan population has approximately 179.2 million (2012 est.) (http://data.worldbank.org/country/pakistan) and broadly distributed into two main regions: urban-rural areas. Modeling of fertility curves is main concerned for studying the reproductivity. Therefore, demographers always insist to model fertility curves.

In this paper, we study and model the pattern of the most important indicator of fertility which is age specific fertility rate of urban-rural areas from 1984-2007 and also studied the change of fertility rates from 1984-2007 in these areas. We apply functional time series independently for modeling age specific fertility rates of urban-rural areas of Pakistan and then forecast fertility curves. Basically, functional data analysis technique (FDS) encompasses data in the form of curves that are observed functional at regular intervals of age. The applications of functional time series model have been employed by Hyndman and Ullah (2007) and Yasmeen and Mahmood (2012) for demographic forecasting.

A variety of empirical literature has been proposed for studying the fertility pattern and techniques to model and forecast fertility sees: Hussain and Bittles (2004), Nasir, et al. (2009), Sathar and Casterline (1998), Zakria (2012), Kielman and Pham (2000), Nikotovic (2011). From existing literature, there is lacking the modeling approach for projecting fertility by FTS model region wise. Fertility forecast are of interest to governments in planning children's services. Finally, we also obtain the 80% prediction intervals for one-year, ten years and twenty years fertility forecasts.

Thus, the main objectives of this paper are:

- To study the pattern of the age specific fertility rates of urban-rural areas of Pakistan for available years
- To estimate the ASFR of urban-rural areas for missing years using interpolation spline
- To model the trends of ASFR of urban-rural by FTS
- To obtain fertility forecasts curve with 80% prediction intervals for one-year, ten-years and twenty years.

This paper is divided into six sections. Section 1 starts with introduction of the fertility and FTS model approach for demographic forecasting. In section 2, we are reviewed the fertility pattern of Pakistan. In section 3, some basic concepts of fertility of urban and rural areas by functional time series (FTS) model are discussed. In section 4, we have described source of data of fertility rates region wise: urban-rural and estimation of fertility rates for missing years, the application of FTS models to the age-specific fertility rates of urban-rural areas are discussed. We discuss the statistical results in section 5 and then final conclusion to overall analysis in section 6.

2. THE FERTILITY PATTERN IN PAKISTAN

Pakistan is the world's sixth most populous country. Pakistan population is broadly distributed into two main regions: urban and rural areas. During 1950–2011, Pakistan's urban population expanded over sevenfold, while the total population increased by over fourfold. The demographic researches have been observed decline in fertility after 1990s. TFR in urban area of Pakistan came down from 6.24 to 3.2 during study period (1984-2007); the estimates imply a decline of 3.04 births and TFR in rural area of Pakistan came down from 7.27 to 4 during study period (1984-2007); the estimates imply a decline of 3.27 births (Source: Pakistan Demographic Survey). However, the past trend especially in rural areas where the fertility rate is relatively high as compared to urban due to low contraceptive awareness in rural areas of Pakistan.

3. FORECASTING THROUGH FUNCTIONAL TIME SERIES ((FTS) MODELS

Hyndman and Ullah (2007) who introduced functional time series (FTS) model for demographic forecasting. They used several principal components obtained by applying singular value decomposition (SVD) to the matrix of log fertility rates. Here, we will apply functional time series model independently for forecasting age specific fertility rates of Pakistan separately for urban and rural population. The main FTS model is given by

$$y_t(x_i) = g_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i} \quad (1)$$

where $y_t(x_i)$ denotes the log fertility rate at age x_i in year 't', $y_t(x_i) = \log\left[f_t(x_i)\right]$ and assume that there is an underlying smooth function $g_t(x_i)$ that observing with error. Where x_i is the centre of age-group $i, (i = 1, ..., p)$, $\varepsilon_{t,i}$ are independent and identically distributed standard normal random variable and $\sigma_t(x_i)$ is the amount of noise to vary with age x.

Hyndman & Ullah 2007 also proposed the nonparametric smoothing to decompose the fitted curves via basis function using the model:

$$g_t(x) = \mu(x) + \sum_{k=1}^{K} \alpha_{t,k}\phi_k(x) + e_t(x) \quad (2)$$

where $\mu(x)$ is the mean log fertility rate across years, $\phi_k(x)$ is the set of orthogonal basis functions, $\alpha_{t,k}$ is univariate time series, and $e_t(x)$ is the model error which is assumed to be serially uncorrelated i.e.

$$e_t(x) \sim N\left(0, V(x)\right).$$

The *h*-step ahead forecast of $\hat{y}_{,n+h,}(x)$ can be obtained as

$$\hat{y}_{n+h}(x) = \hat{\mu}(x) + \sum_{k=1}^{k} \hat{\alpha}_{k,n+h}\hat{\phi}_k(x) \quad (3)$$

where $\hat{\mu}(x)$ and $\hat{\varphi}_j(x)$ are the estimates of the mean function and basis functions, respectively and denotes the *h*-step ahead forecast of $\hat{\alpha}_{k,n+h}$

4. MATERIAL AND METHODS

4.1. Available Data

A Secondary data of age specific fertility rates (ASFR) of urban-rural regions of Pakistan during 1984-2007 has been taken from the Pakistan Demographic Surveys (1984-1986, 1988-1992, 1995-1997, 1999-2001, 2003, 2005, 2006 and 2007). Age-specific fertility rates are missing for rest of the years, so we estimate those using interpolating splines.

4.2 Fitting FTS Models Indepenently to Age Specific Fertility Rates of Pakistan Separately for Urban-Rural Areas

Firstly, we plot ASFR separately for urban and rural areas of Pakistan for selected years 1984, 1990, 2000, 2005 and observed them as separate functional observations in figure 1a and 1b. After that, the ASFR for all available years are plotted in figure 2a and 2b. Secondly, we use non-parametric smoothing method to smooth these rates. For obtaining smooth curves, we use weighted regression B-spline for all available years, as plotted in figure 3a and 3b. Then, we apply FTS model introduced by Hyndman and

Ullah 2007 to analyse the available curves. Functional regression models with four basis functions provide an adequate fit to available curves. Thirdly, the time series coefficients are then forecast using univariate time series models. These forecasts are then multiplied with the estimated basis functions resulting in forecasts of the entire fertility curve. This technique also provides computation of prediction intervals (for details, see Yasmeen et al., 2010). We also estimate one-year, ten-year and twenty year predictions interval for rates of Pakistan separately for urban-rural areas. All statistical analyses were performed in R version 2.15.3.

5. RESULTS

Figure 1a and 1b displays the pattern of ASFR separately for urban and rural areas of Pakistan for the selected years. We have seen that the fertility rates have been decreased for all age groups in both regions

Figure 1a: Fertility Rates of Pakistan Urban Region

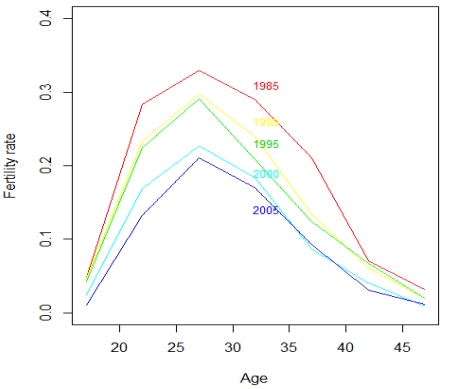


Figure 1b: Fertility Rates of Pakistan Rural Region

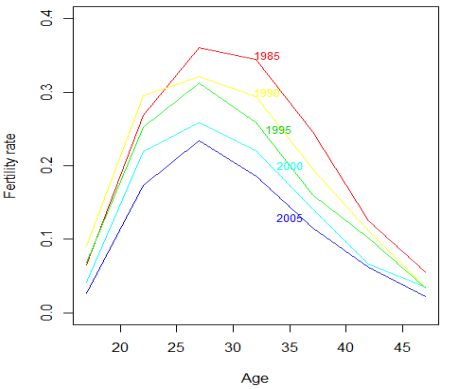
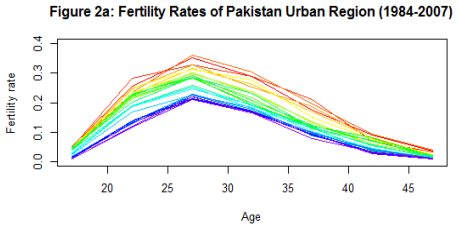
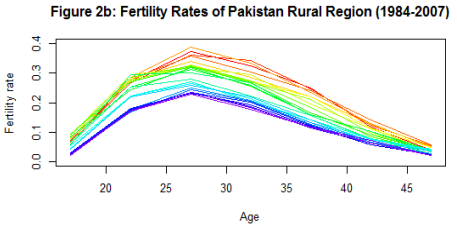
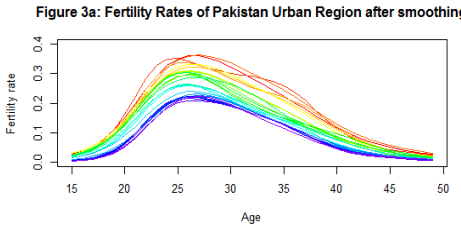
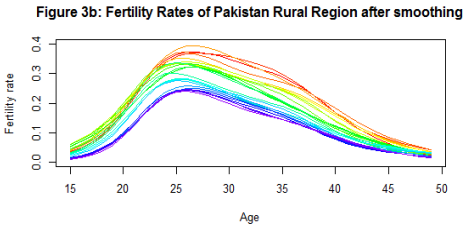


Figure 2a and 2b displays the pattern of ASFR separately for urban and rural areas of Pakistan for the study period which shows the rural curves are much wider than urban curves. At first, for application of FTS model, we smoothed log fertility rates which are given in Figure 3a and 3b separately for urban and rural regions of Pakistan. We use regression B spline for smoothing. Secondly, a model with four basis function is selected. The percentage variation due to basis function for urban region is: 90.8% 5.1% 2.4% 1.1%. Hence only four basis functions are sufficient to explain about 99.4% of the total variability in rates of urban women. Similarly, rural region is: 89.3% 7.0% 2.0% 0.8%. Hence only four basis functions are sufficient to explain about 99.1% of the total variability in rates of rural women.









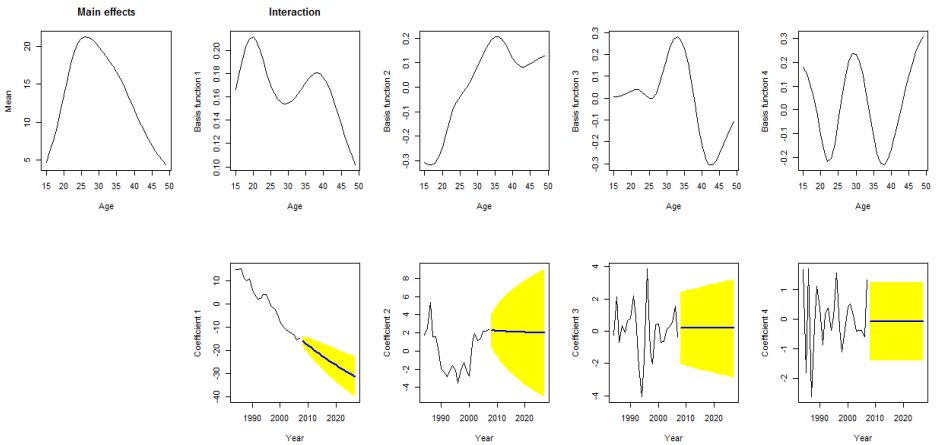


Figure 4a: Forecast for First Four Coefficients of FTS model for Urban Region

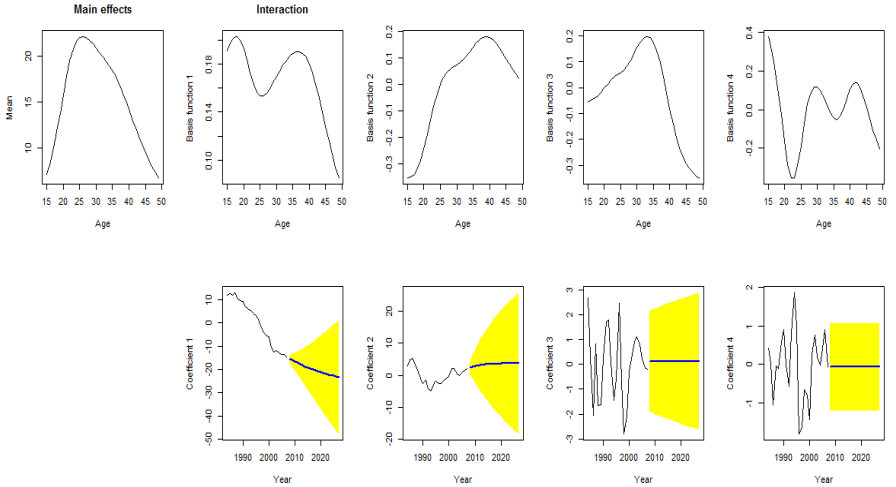


Figure 4b: Forecast for First Four Coefficients of FTS model for Rural Region

Figure 5a: Twenty-year forecasts Pakistan Urban (2008-2027)

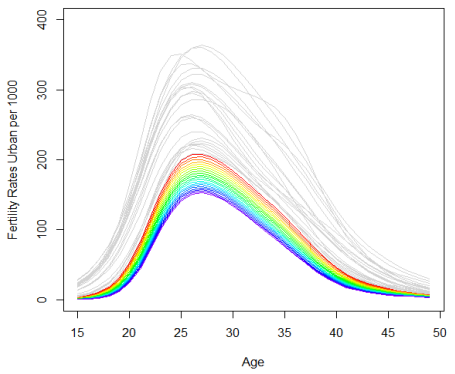


Figure 5b: Twenty-year forecasts Pakistan Rural(2008-2027)

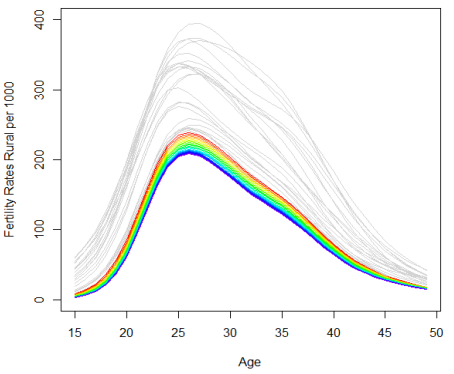
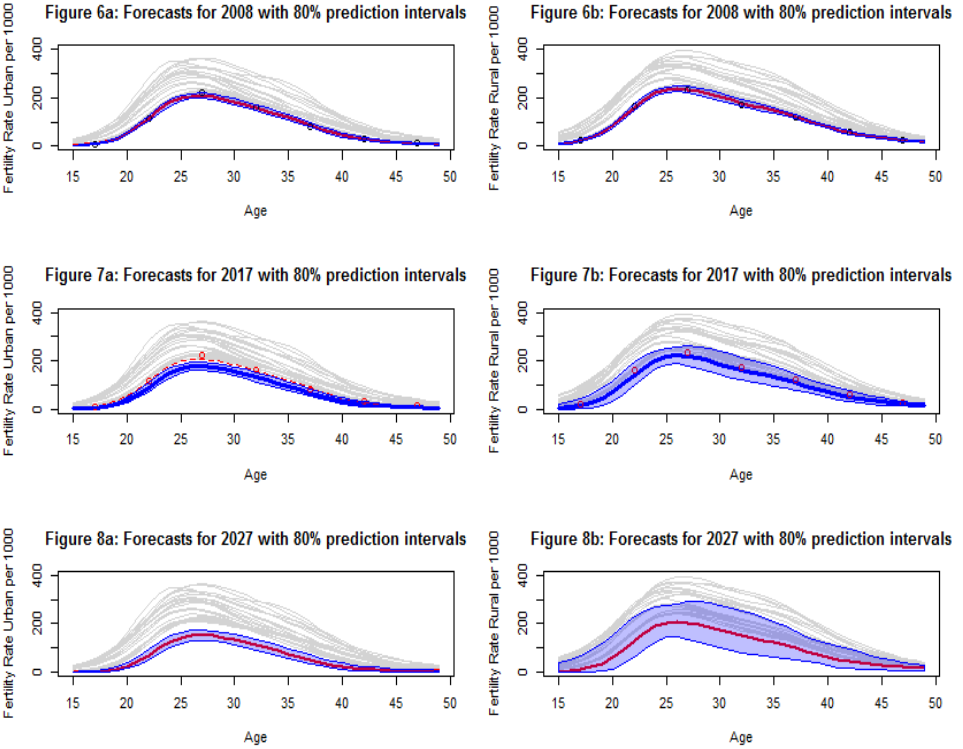


Figure 4a displays the mean function plots of first four basis functions and their corresponding coefficients for the FTS model for urban area. The major changes in fertility rates amongst Pakistani urban women occur at the ages 22 and 40 years, while their respective coefficients show a decline in future years. We do not attempt to interpret the other basis functions as they involve second-and higher-order effects. Figure 4b displays the mean function plots of first four basis functions and their corresponding coefficients for the FTS model for urban area The major changes in fertility rates amongst Pakistani rural women occur at the ages 18 and 37 years, while their respective coefficients show a decline in future years. We do not attempt to interpret the other basis functions as they involve second-and higher-order effects.

After obtaining basis function, we multiplying the forecasts of the coefficients with the basis functions of their respective regions and summing there results, we obtain forecasts of the entire fertility curves. Figure 5a and 5b displays the forecast urban and rural fertility for the next twenty years (2008-2027), the forecast displays the greater decline observed in urban fertility as compared to rural fertility.



Figures 6a and 6b, 7a and 7b, 8a and 8b represent the forecast with 80% prediction interval for 2008, 2017 and 2027 for both urban-rural regions. These plots exhibit the future fertility rates decline for all ages, but that the decline will be greatest for the middle-aged women and relatively slower for the youngest and oldest women while the prediction intervals of rural region are much broader than urban region.

6. COMMENTS AND CONCLUSION

In this paper, we study ASFR pattern of urban and rural areas of Pakistan. We also applied FTS model independently to ASFR for both regions of Pakistan. Fertility is the most important indicator for national and international demographers. We model age specific fertility rates instead of total fertility rate provide more accurate forecasts given clear differences between age groups in contribution to overall variability and also any improvements in fertility forecasting have an immediate impact of guiding policy decisions regarding the allocation of current and future resources.

The pattern of ASFR of both regions has a traditional V-shape. However, from existing literature FTS model is applied on ASFR of Pakistan and forecast for ASFR of Pakistan and also some polynomial models are also fitted to fertility rates of Pakistan and its urban-rural region but neither of the study forecast the future age specific fertility rates for two broadly distributed regions: urban and rural. To our knowledge, this is the first

study which shows fertility-age projections for urban and rural regions. From multiple sources, we imply that the fertility is declined in Pakistan after 1990s. Our results also confirm the previous research results. The least reduction has been seen in the rural areas as compared to urban region because of lack of educational facilities, exposure of mass media, early marriages and poor health facilities. We also obtain the twenty year prediction interval for ASFR of both regions. The forecast and their predication plots will decline in all age groups.

REFERENCES

1. http://data.worldbank.org/country/pakistan
2. CIA: The World Fact book: 2013, Population Growth Rate
3. Errbas, 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.
4. Government of Pakistan (2007). *Population Demographic Survey-2007*. Federal Bureau of Statistics, Statistics Division, Islamabad, Pakistan.
5. Hussain, R. and Bittles A.H. (2004). Assessment of Association between Consanguinity and Fertility in Asian Populations. *Journal Health Population Nutrition*, 22(1), 1-12.
6. 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.
7. Islam, M.R. and Ali, M.K. (2004). Mathematical modeling of age specific fertility rates and study the reproductivity in the rural area of Bangladesh during 1980-1998. *Pak. J. Statist.*, 20(3), 379-392.
8. 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.
9. Kielman, N. and Pham, D.Q. (2000). Predictive Intervals for Age-Specific Fertility. *Eur. J. Popu.*, 16, 41-66.
10. Sathar, Z.A. and Casterline, J.B. (1998). The onset of fertility transition in Pakistan. *Popu. and Dev. Rev.*, 24, 773-796.
11. Yasmeen, F. and Mahmood, Z. (2012). Forecasting age-specific fertility rates of Pakistan using functional time series model. *Proceedings of 12th Islamic Countries Conference on Statistical Sciences*, 23, 669-674, Doha, Qatar.
12. Yasmeen, F., Hyndman, R.J. and Erbas, B. (2010). Forecasting age-related changes in breast cancer mortality among white and black US women: A functional data approach. *Cancer Epidemiology*, 34(5), 542-549.
13. Zakria, M., Yasmeen, M., Muhammad, F. and Mahmood, S. (2012). Modeling the age specific fertility rates of rural and urban areas of Pakistan during 1984-2007. *Proceedings of 12th Islamic Countries Conference on Statistical Sciences*, 23, 571-582, Doha, Qatar.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 69-76

MODELING AND FORECASTING ELECTRICITY CONSUMPTION IN PAKISTAN

Farah Yasmeen[1] and **Muhammad Sharif**[2]
Department of Statistics, University of Karachi, Karachi, Pakistan
Email: [1]riazfarah@yahoo.com
[2]sharif_stats@yahoo.com

ABSTRACT

Now-a-days, different sectors of the economy are being significantly affected by the electricity variable. In this research, we analyzed the monthly electricity consumption for the period of January 1990 through December 2012 in Pakistan using linear and non linear modeling techniques. They include ARIMA, Seasonal ARIMA (SARIMA) and ARCH/GARCH models. Electricity consumption model reveals a significant trend due to socio-economic factors. The monthly behavior of our forecast values depicts that the electricity consumption is more for summer season and this demand will be increased in future. Forecast model and the forecast values reveal that the electricity consumption is increasing with the passage of time. We have checked the accuracy of the models by diagnostic tests and then compared the forecast values to select the more appropriate model. In this study, the most least out of sample forecast performance i.e. Mean Absolute Percentage Error (MAPE) value and most minimum forecast standard deviation value of ARIMA(3,1,2) show that among the four competing time series models, ARIMA(3,1,2) model is the most appropriate model to forecast electricity consumption of Pakistan.

KEY WORDS

ARCH/GARCH, ARIMA, Electricity Consumption, forecast accuracy, SARIMA.

1. INTRODUCTION

Electricity is a typical form of energy and plays an essential role in modern life, bringing benefits and progress in various sectors, including transportation, manufacturing, mining and communication sectors. Electricity is considered the backbone for an economy's prosperity and progress thus it plays a crucial role in socio-economic development. It is a tool that has capability of making useful contributions to planning and future policy formulation of energy sector.

Pakistan is a developing and agricultural country of South Asia. It is an electricity-deficient country and the electricity sector operates at bare capacity margin. Pakistan fulfills its energy requirement through different sources, which includes oil, gas, atomic, hydral, etc. The electric power sector in Pakistan is operated by the Water and Power Development Authority (WAPDA), and the Karachi Electricity Supply Corporation (KESC), with additional generation contribution from Independent (Private) Power Producers (IPPs).

Recently, [1] reported that the current electricity production of Pakistan is around 11,500 MW per day. According to [2], the economy's electricity jumped to around 20,000MW per day by 2010. This gap between the supply and demand in the energy sector in Pakistan suggests the magnitude of the energy crisis the economy faces. To fulfill increased electricity demand, planning and investment in infrastructure development is necessary. [3] justified that Pakistan has been facing severe imbalance between energy demand and energy supply. During the past 25 years, energy supply in Pakistan has been increased by around 40 times, but still the demand outstrips supply. According to research of [4], current electricity crisis has influenced the whole economy.

2. MODELING ELECTRICITY CONSUMPTION

Various studies of modeling and forecasting electricity consumption have been undertaken in the past. [5] applied ARIMA and cointegration analysis modeling for electricity demand estimation and forecast in Turkey. [6] Used modeling based on econometrics method for forecasting electricity demand in Bangladesh. In this research the researcher forecasted the electricity demand with respect to the GDP data. The results show that as far as we can keep pace with the electricity demand and energy generation, we can successfully avoid the burden of load-shedding.

For developed countries, [7] indicate that there is a strongly positive correlation between wealth and energy (or electricity) consumption, and the correlation between electricity and wealth is stronger than the correlation between total energy and wealth. However, the process of electricity production and consumption may emit air pollution and greenhouse gas which results global warming. Previous studies mainly utilize time-series or cross-sectional datasets to investigate the relevant issues of energy ([8], [9]). [10] studied Autoregressive Distributed Lag (ARDL) bounds testing procedure to identify the long run equilibrium relationship between electricity consumption and economic growth in Pakistan for the period of 1971 to 2008. [11] examined the relationship between electricity demand and temperature in Europe by means of a panel threshold regression model on 15 European countries over the last two decades. The results show that temperature is a major determinant of electricity consumption in Europe and that the sensitivity of electricity consumption to temperature has increased in the recent period. Given the concern about global warming, these findings support the renewed interest in energy related questions by the European policymakers.

Using the electricity consumption and real GDP per capita data of Pakistan over the period 1971 to 2008, [12] worked on electricity consumption and economic growth. The researchers used co-integration technique to model the relationship between electricity consumption and real GDP per capita. The empirical results indicate the existence of a long-run equilibrium relationship between the two variables.

3. ELECTRICITY CONSUMPTION IN PAKISTAN

In this research, we will use linear and non linear modeling techniques for modeling and forecasting electricity consumption in Pakistan. They include Auto Regressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), Auto Regressive

Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model.

3.1 Data description

As Pakistan is a developing country, there is no appropriate arrangement for saving of energy. The demand of electricity in the country almost remained greater than the supply. Therefore it is not possible to get directly the monthly electricity consumption data of Pakistan. In other words we can say that at least consumption of electricity is equal to its production. Therefore we used electricity production data as a *proxy* variable to electricity consumption data.

A series of monthly electricity consumption (EC) of Pakistan in million kilo watt hour (kwh), across the period from January 1990 through December 2011, has been used in this study. The data were recorded by department of Federal Bureau of Statistics Pakistan and plotted in Figure 1. It consists of electricity consumption in all economic sectors (industrial, residential, and commercial) of Pakistan, because regional or sectoral disaggregated data were not available.

The data are divided into three subsets: a training set January 1990– April 2012, validation set January 2000– December 2011 and test set May 2012- December 2012. The training subset is used to estimate the parameters. The second subset, validation set is used to monitor the estimation process. Third data set was not used for estimation, but was used for comparing different models.

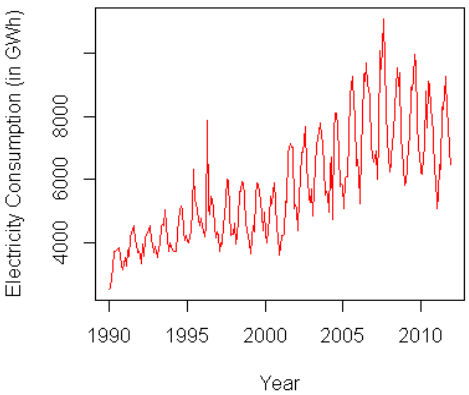


Fig. 1: Elec. Consumption in Pakistan

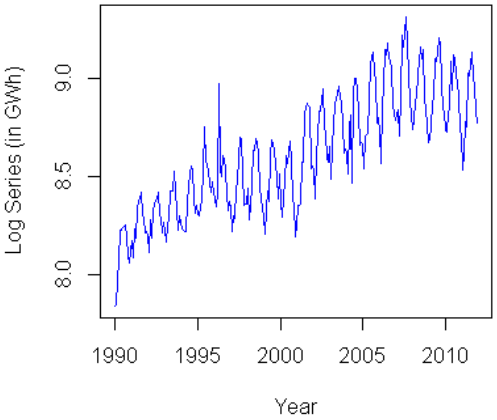


Fig. 2: Log Electricity Consumption

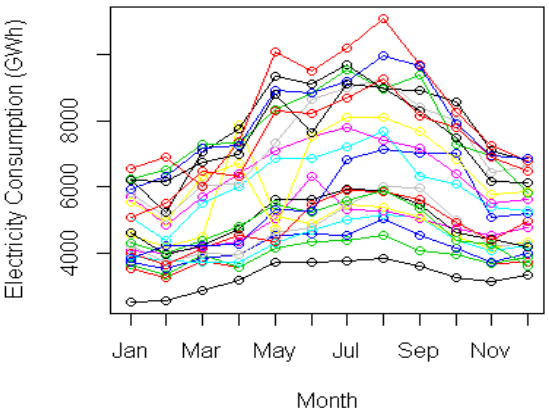


Fig. 3: Season Plot

3.2 Linear Modeling Techniques

Box-Jenkins Methodology

Box and Jenkins were the first to approach the task of estimating an ARMA model in a systematic manner. There are three steps to their approaches which include identification, estimation and the model diagnostic checking. Model identification involves determining the order of the model. The model can be identified using Information criteria like Akaike Information Criteria (AIC). The model can be estimated using least square or maximum likelihood, while model can be checked by deliberating over fitting and residual diagnosis

Forecast Accuracy Measurement

In literature, different criteria are used for assessing the accuracy of time series forecasting. Let us define $f_{t,s}$ as the forecast made at time 't' for 's' steps ahead and y_{t+s} as the realised value of 'y' at time $t+s$. Here, we use Mean Absolute Percentage Error

(MAPE) to check the forecasting performance of the models. Mean Absolute Percentage Error (MAPE) is defined as:

$$MAPE = 100 \times \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_{t+s} - f_{t,s}}{y_{t+s}} \right| \quad (1)$$

where 'N' shows total number of observations in the data.

3.2.1 Autoregressive Integrated Moving Average (ARIMA) Modeling

An ARMA model of order (p, q) is defined

$$y_t = \mu + \varphi_1 \, y_{t-1} + \varphi_2 \, y_{t-2} + \ldots + \varphi_p \, y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_q u_{t-q} + u_t \quad (2)$$

e.g. ARMA(2,2) model can be written as

$$y_t = \mu + \varphi_1 \, y_{t-1} + \varphi_2 \, y_{t-2} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + u_t \quad (3)$$

Considering $\mu = 0$ and putting the estimated parameter values, we have

$$\hat{y}_t = 1.400759 \, y_{t-1} - 0.787728 \, y_{t-2} + 1.360282 \text{ ut-1} - 0.897939 \text{ut-2} + u_t \quad (4)$$

3.2.2 Seasonal Autoregressive Moving Average (SARIMA) Modeling

In the case where seasonal components are included in ARIMA model, it is known as the SARIMA model. By seasonality, we mean periodic fluctuations. In addition to trend, stationary series quite commonly display seasonal behavior where a certain basic pattern tends to be repeated at regular seasonal intervals. The seasonal pattern may additionally frequently display constant change over time as well. Just as regular differencing was applied to the overall trending series, seasonal differencing (SD) is applied to seasonal non-stationary series as well.

The monthly seasonality graph of electricity consumption (Figure 3) shows that the electricity consumption begins to increase from March through August by using air conditioning systems and is maximum in July & August while minimum in January & February. The plot also shows that data have seasonal and trend patterns with increasingly variability in variance. It means that the data not yet satisfy the stationary condition, both in mean and variance at various time periods seasonality exist. Unfortunately, the daily data for electricity consumption was not available due to which we are unable to say anything about the daily profile of electricity consumption e.g. in working days and holidays.

3.3 Non- Linear Modeling Techniques

3.3.1 AutoRegressive Conditional Heteroskedasticity (ARCH) Modeling:

The linear structural and time series models assume constant error variance. This assumption is not realistic in most of the time series data especially at high frequency. Therefore we often move to some non-linear models (volatile models) like ARCH and GARCH models.

An ARCH model of order (q) is defined as

$$y_t = \beta_1 + \beta_2 x_{2t} + ... + \beta_k x_{kt} + u_t, \; u_t \sim N(0, h_t)$$

where, $h_t = \alpha_\circ + \alpha_1 u^2_{t-1} + \alpha_2 u^2_{t-2} + ... + \alpha_q u^2_{t-q}$ (5)

Using AIC and BIC information criteria for electricity consumption data, ARCH(2) is the most appropriate model because of most least values of AIC and BIC among others.

The estimated ARCH(2) model can be written as

$$\sigma^2_t = \alpha_\circ + \alpha_1 u^2_{t-1} + \alpha_2 u^2_{t-2} \quad (6)$$

$$\hat{\sigma}^2_t = 0.005462 + 0.235823 u^2_{t-1} + 0.245966 u^2_{t-2} \quad (7)$$

3.3.2 Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) Modeling

GARCH (*p*, *q*) can be written as

$$\sigma^2_t = \alpha_0 + \sum_{i=0}^{p} (\alpha_i u^2_{t-i}) + \sum_{j=0}^{q} (\beta_j \sigma^2_{t-j}) \quad (8)$$

Using AIC and BIC information criteria, GARCH(1,1) is the most appropriate model because of most least values of AIC and BIC among others. The estimated GARCH(1,1) model can be written as

$$\hat{\sigma}^2_t = 0.005802 + 0.197685 u^2_{t-1} + 0.181422 \sigma^2_{t-1} \quad (9)$$

The forecast evaluation of each model summarized in Table 1 shows that among the four time series models ARIMA (3, 1, 2) model has most least out of sample forecast performance i.e. least MAPE value, showing that ARIMA(3,1,2) model is the best model to forecast electricity consumption of Pakistan.

Table 1
Summary Table of AIC and MAPE using Different Time Series Models

S.No	Model	AIC	MAPE
1	ARIMA(3,1,2)	-1.496960	5.994240
2	SARIMA(2,1,2) with SMA(12)	-2.074949	10.51
3	ARCH(2)	-1.849054	11.71829
4	GARCH(1,1)	-1.827264	9.998946

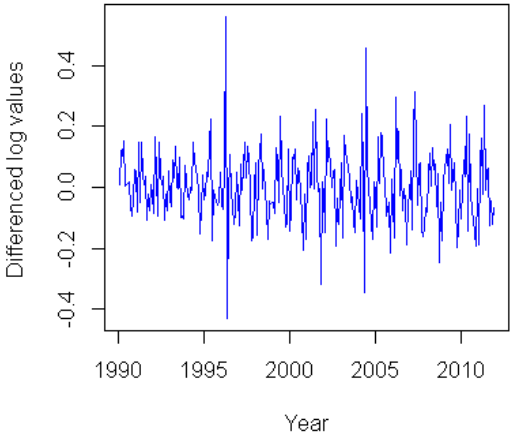


Fig. 4: Differenced Log Series

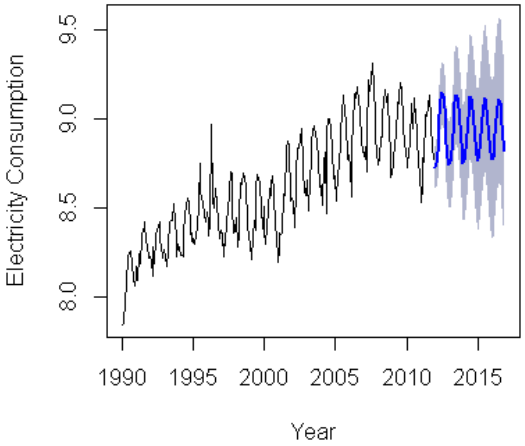


Fig. 5: Forecast from ARIMA(3,1,2)

4. CONCLUSION

In this research, we have analyzed and obtained the forecast of the electricity consumption in Pakistan using linear and non linear modeling techniques. They include ARIMA, Seasonal ARIMA (SARIMA), ARCH/GARCH model. The results of the study conclude that ARIMA (3,1,2) model is the most appropriate model for forecasting electricity consumption of Pakistan. The monthly behavior of our forecast values depicts that the electricity consumption is more for summer season and this demand will be increase with the passage of time. This indirectly indicates that the government of Pakistan must take effective steps to increase the electricity production through different energy sources to restore the economical status of the country by meeting the demand of electricity in the country.

REFERENCES

1. Ali, M., Iqbal, J.M. and Sharif, M. (2013). Relationship between extreme temperature and electricity demand in Pakistan. *Intl. J. Energy and Environ. Engg*. (IEEE) to appear.
2. Shahbaz, M. and Feridun, M. (2011). Electricity consumption and economic growth empirical evidence from Pakistan. *Springer Science Business Media B.V., Qual Quant* DOI 10.1007/s11135-011-9468-3.
3. Khan, M.A. and Qayyum, A. (2009). The Demand for Electricity in Pakistan. *OPEC Energy Review*, 33(1), 70-96.
4. Khan, A.M. and Ahmed, U. (2009). *Energy Demand in Pakistan: A Disaggregate Analysis*, Pakistan Institute of Development Economics, Islamabad. Available at http://mpra.ub.uni-muenchen.de/15056/ MPRA Paper No. 15056.
5. Erdogdu, E. (2007). Electricity demand analysis using cointegration and ARIMA modeling: A case study of Turkey. *Energy Policy*, 35(2), 1129-1146.
6. Shuvra, M.A., Rahman, M.M., Ali, A. and Khan, I.S. (2011). Modeling and Forecasting Demand for Electricity in Bangladesh: Econometrics Model. *International Conference on Economics, Trade and Development*, IPEDR 7 IACSIT Press, Singapore.
7. Ferguson, R., Wilkinson, W. and Hill, R. (2000). Electricity use and economic development. *Energy Policy*, 28, 923-934.
8. Wolde-Rufael, Y. (2006). Electricity consumption and economic growth: A time series experience for 17 African countries. *Energy Policy*, 34(10), 106-1114.
9. Wolde-Rufael, Y. (2009). Energy consumption and economic growth: The experience of African countries revisited. *Energy Economics*, 31(2), 217-224.
10. Bessec, M. and Fouquau, J. (2008). The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach. *Energy Economics*, 30, 2705-2721.
11. Shiu, A. and Lam, P.L. (2004). Electricity consumption and economic growth in China. *Energy Policy*, 32, 47-54.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 77-86

FUNCTIONAL TIME SERIES MODELS TO PREDICT AGE-SPECIFIC BREAST CANCER INCIDENCE RATES FOR WOMEN IN KARACHI

Sidra Zaheer[1] and **Farah Yasmeen**[2]

[1] School of Public Health, Dow University of Health Sciences, Karachi, Pakistan. Email: sidraz.ku@gmail.com

[2] Department of Statistics, University of Karachi, Karachi, Pakistan. Email: riazfarah@yahoo.com

ABSTRACT

Background: At present, breast cancer is the most common cancer among women in the Pakistan, although incidence prediction may contain important evidence to understand for allocating public health resources and control of the disease.

Objective: In this study we examined the recent trends of breast cancer incidence rates among women then we apply functional time series model to forecasting future time trends in age-specific breast cancer incidence rates for women of Karachi.

Methodology: This is a retrospective study; we obtained the secondary data of breast cancer from various locations in Karachi. They include Jinnah Hospital, Karachi Institute of Radiotherapy and Nuclear Medicine (KIRAN) and CIVIL hospital, Karachi where the data are available for the years 2004-2011. During this period, a total of 5331 new cases of female breast cancer were registered. We analyzed the data in 5-year age groups (15-75+). Age-specific incidence curves were obtained using nonparametric smoothing methods then the curves are decomposed using principal components analysis to fit functional time series model with four basis functions to forecast future time trends.

Results: From 2004 to 2011, overall age-specific breast cancer incidence rates (expressed as number of breast cancer per 100,000 women-years) were increased with age. From early ages (>15 years) incidence rates were showing slightly increasing trends through age of 50 and the rates showed variability for the age > 50. At early years from 2004 to 2005 the incidence rates at high ages (+70) started decline sharply and in recent years, the incidence rates were showing increasing patterns at high ages 70 and above. Overall ten year forecasts suggest continued stable trends for women aged 15 to 50 years, and slightly increased among women older than 55 with time.

Conclusion: Prediction of age related changes in breast cancer incidence rates will be provide useful information for controlling the overall burden of cancer in Pakistan and also utilized for health planning in future research. Moreover, these models will be most useful for modeling and projecting future trends of other cancers and chronic diseases.

KEYWORDS

Incidence rates; breast cancer; nonparametric smoothing; functional time series.

1. INTRODUCTION

The Functional Data Analysis (FDA) explores samples of data where each observation arises from a curve or function. With functional data methods, data can be smooth curves or functions rather than the discrete observations (Ramsay & Silverman 2005, Kneip & Utikal 2001). Such data arise in many different contexts including demography, epidemiology, finance, meteorology and elsewhere. Data of annual mortality and fertility rates with age as functional covariate refer to as functional time series (Hyndman & Ullah, 2007).

Previous methods of modeling do not utilize the advantages of functional data approach. Recently, functional data analysis (FDA) approach is used for breast cancer mortality forecasting by Erbas et al. (2007) and Yasmeen et al. (2010).

Breast cancer disease is the most common malignancy causing deaths in women worldwide, and affects the developed and the developing nations. In Pakistan, breast cancer is one of the most dreadful diseases affecting women, and the incidence of breast cancer in Pakistani women is higher than many other Asian countries (Naeem et al., 2008; Afsar et al., 2010).

In this study, we decided to work on breast cancer incidence data of women in Karachi. As described above, the functional time series (FTS) models have been applied to breast cancer incidence rates to find how the incidence curves changes with time and to forecast age-specific incidence of breast cancer in women. To the best of our knowledge, this will be the first study to report age-specific predictions/forecasts of breast cancer incidence rates in Pakistan.

2. STATISTICAL METHODS

Data:

In this study we obtained the secondary data of breast cancer from various locations in Karachi. They include Jinnah Hospital, Karachi Institute of Radiotherapy and Nuclear Medicine (KIRAN) and Civil hospital, where the data was available for the years 2004-2011. During this period, a total of 5331 new cases of female breast cancer were registered. Breast cancer incidence rates were calculated as the number of new cases of breast cancer divided by the corresponding population at risk during a year, expressed as the rates per 100,000 of the population. We analyzed the rates in 5-year age groups: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+. The annual populations at risk for females in these age groups were estimated, based on the 1998 census (obtained from the Sindh Bureau of Statistics).

Functional Data Analysis:

At first, we define a functional time series model. Let $r_t(x)$ denote the incidence rate for age x and year t, $t = 1, 2, ..., n$. In FDA, model the log rates, $y_t(x) = \log[r_t(x)]$ and assume that there is an underlying smooth function of x, $s_t(x)$ that we are observing with error. Thus,

$$y_t(x_i) = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i} \qquad (1)$$

where x_i is the center of age-group, $i, (i = 1, ..., p)$, $\varepsilon_{t,i}$ is an independent and identically distributed standard normal random variable and $\sigma_t(x_i)$ allows the amount of noise to vary with age 'x'.

We estimated these smooth functions by using nonparametric technique locally weighted regression (Loess) (Cleveland & Devlin, 1988), then we decomposed the smoothed curves using the basis function model proposed by Erbas et al. (2007) and Hyndman & Ullah (2007).

$$s_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\varphi_k(x) + e_t(x) \qquad (2)$$

where $\mu(x)$ is the mean log incidence rate across years, $\varphi_k(x)$ is a set of orthogonal basis functions, $\beta_{t,k}$ is a univariate time series, and the model error $e_t(x)$ assumed to be serially uncorrelated i.e. $e_t(x) \sim N(0, v(x))$.

The basis functions are computed by applying the functional principal components (PC) decomposition (Ramsay &Dalzell, 1991) to the smooth curves $\{s_t(x)\}$, as this approach gives a small number of basis function , and enables useful interpretation and gives coefficient which are uncorrelated each other. The coefficients $\{\beta_{t,k}\}$ are each forecast, and then multiplied by the basis functions $\{\varphi_k(x)\}$to compute forecasts of age curves through time by using exponential smoothing state space models.

Forecasting:

To prediction of h-step ahead forecast $\hat{s}_{n+h}(x)$ by forecasting the entire function $s_t(x)$ *for* $t = n+1, ..., n+h$

$$\hat{s}_{n+h}(x) = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{k,n+h}\hat{\varphi}_k(x) \qquad (3)$$

Where $\hat{\mu}(x)$ and $\hat{\varphi}_k(x)$ are the estimates of mean function and basis functions respectively, $\hat{\beta}_{k,n+h}$ denotes the h-step ahead forecast of $\beta_{k,t}$ coefficients.

Hyndman and Ullah (2007) showed that the forecast variance can be obtained by adding all the variances from each of the terms in equation (1) and (2).The estimated variances of the error terms are used to compute prediction intervals for the forecasts. Finally we estimated 10-year forecast of breast cancer incidence rates with 80% prediction intervals. All statistical analyses were performed in R version 3.0.1 using the demography package, available at CRAN.

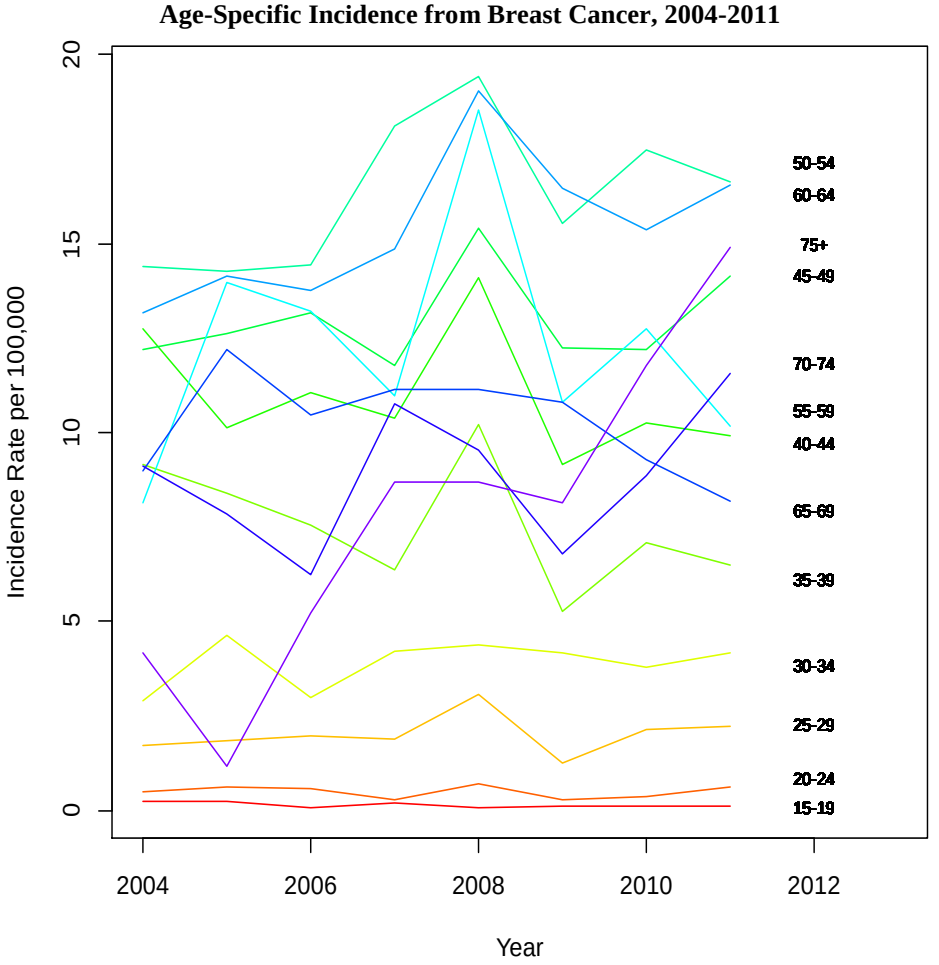


Fig. 1: Breast cancer incidence rates (per 100,000 women) by age-group, Karachi (2004-2011)

3. RESULTS

Breast cancer incidence patterns were observed annually by plotting incidence rates as a function of age. For each year, we plotted incidence-age curves to see the association between age and breast cancer incidence rates. Figure 1 displays the observed age-specific breast cancer incidence rates for the women (per 100,000 women) using rainbow plot (Hyndman & Shang, 2010).

Overall, the breast cancer incidence was relatively high in year 2008 for all age groups, whereas it was relatively stable for women aged 15-25 years (fewer than 25 year age). Except for women aged 60-64 years, there was a sharp increase in incidence between 2009 and 2010 for all age groups; for women aged 60-64 years, the incidence rates started to decline after year 2009. Incidence trends for women aged 70-74 years and

+75 years fluctuated during the study period; after year 2009 the rates sharply increased for that age group then there was a clear and sharp increasing trend between 2005 and 2011. Therefore, the incidence increased with age, and in each year, highest incidence occurred in women aged 50-54 years. The rates were stable at starting years, and then started increasing in 2006 to reach its maximum in 2008.

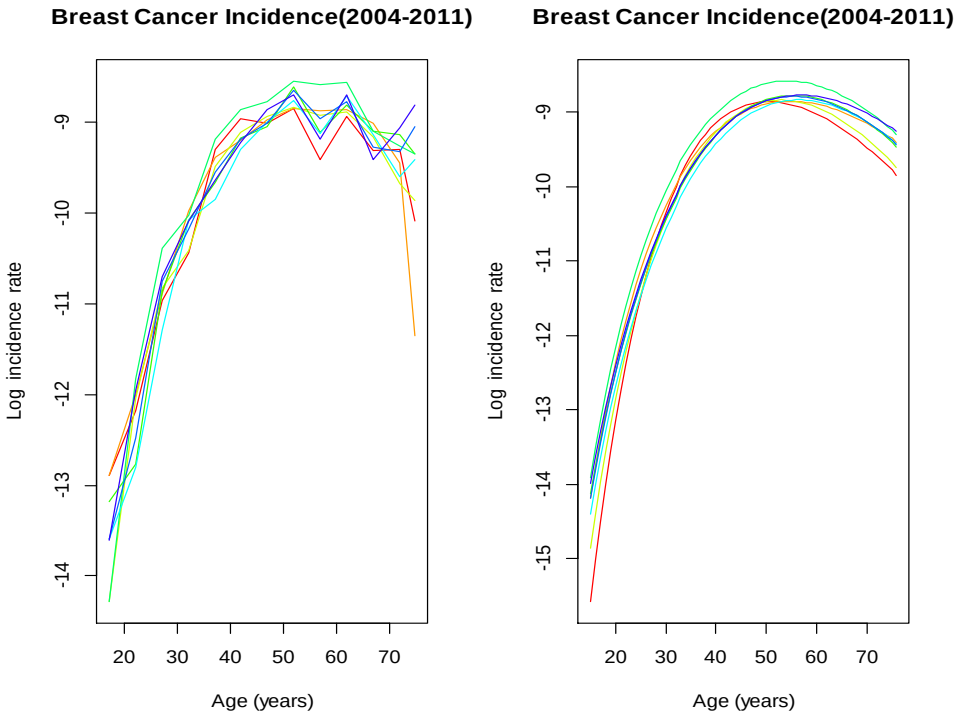


Fig. 2: Log incidence rates and Loess (local regression) smoothed curves by age (2004-2011)

Figure 2 shows the observed log incidence rates and smooth curves as functional data time series. From the graph we can see that the incidence rates increased with age. From early ages (>15) overall incidence rates were showing slightly increasing trends through age of 50 and the rates show variability for the age > 50. At early years from 2004 to 2005 the incidence rates at older ages (+70) started to decline sharply and in recent years, the incidence rates were showing increasing patterns at older ages (70 and above).

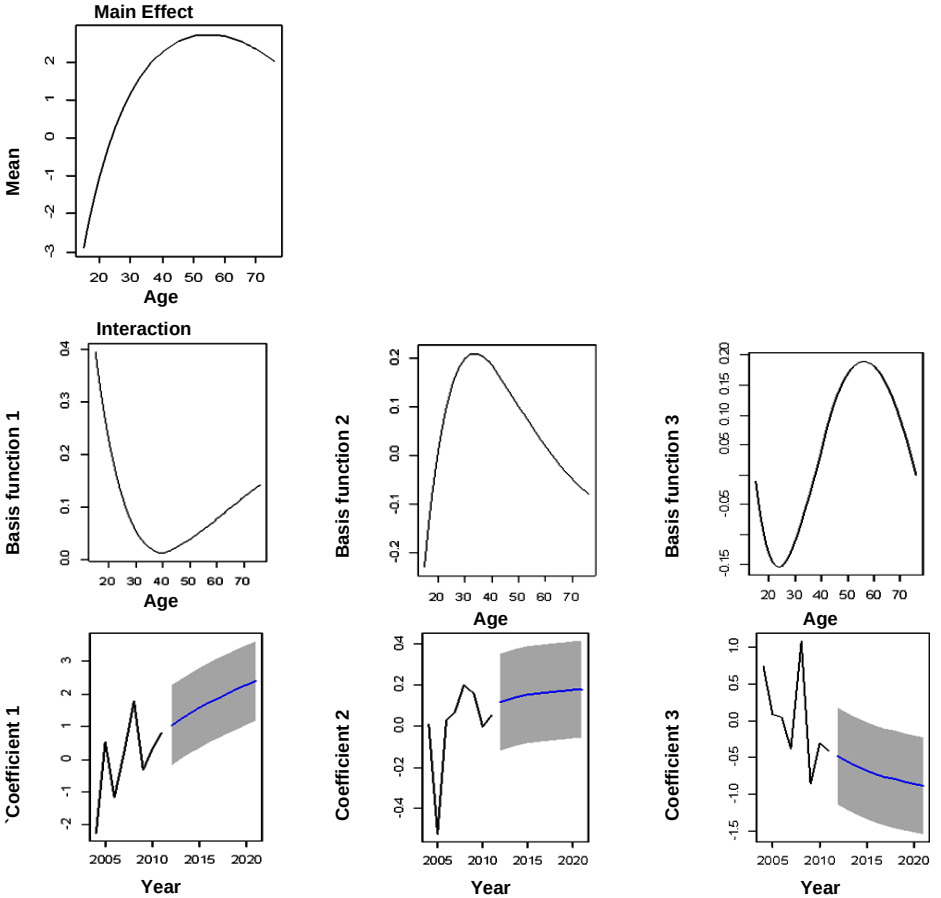


Fig. 3: The components of FTS model and ten-year forecasts of the time coefficients (2012-2021). The shaded region gives 80% prediction intervals.

We used principal components decomposition to estimate the basis functions, we selected a model with k=4 basis functions. The percentage variation explained by the first four basis functions are 77.7%, 19.3%, 2.3% and 0.7% respectively. The estimated basis functions and corresponding coefficients are shown in Figure 3. The first basis function showed increases in incidence with age, specifically high increases arising at the younger ages (age < 20). The plot of first time series coefficient showed fluctuation but overall increases with time. The second basis function explained variation in rates among women aged 40, compared to those who are much younger and much older. The third and fourth basis functions are complex and contrasts those between 20 and 30 or older women (over 70). Forecast for time series coefficient of first basis function β_1, which controls the overall change in trend of breast cancer incidence, is clearly showing increasing trend in future.

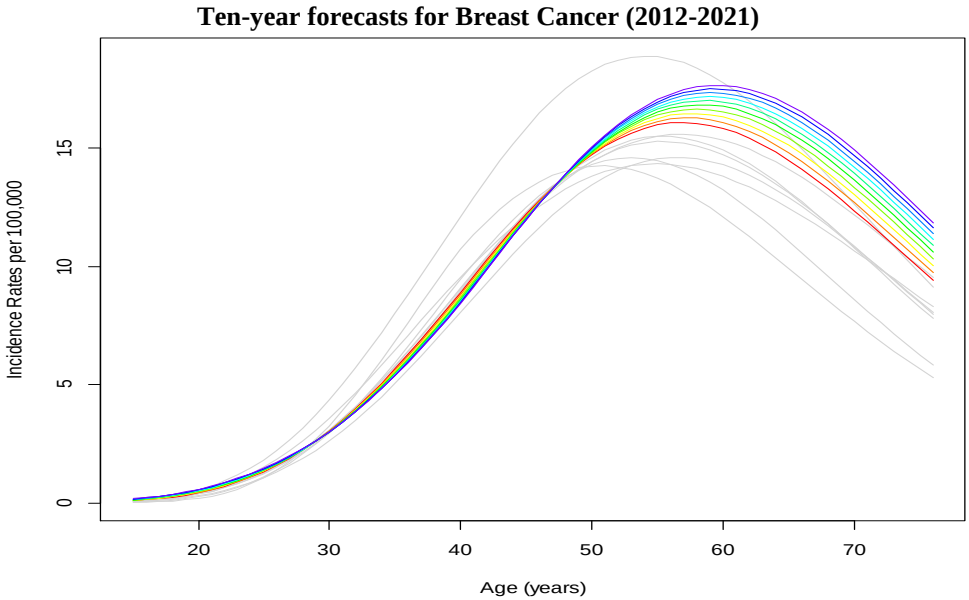


Fig. 4: Ten-year forecasts for breast cancer incidence rates (2012-2021). Gray curves show the actual smoothed incidence rates (2004-2011).

Forecasts for the entire incidence curves were obtained by summing the results after multiplying the forecasts of each coefficient with the basis functions; 10-year breast cancer incidence forecasts are shown in Figure 4. Forecast plots display that the future incidence rates will rise for those aged 50 and above, but the rates will slightly increase or remain stable for those younger than 50 (women aged 15 to 50).

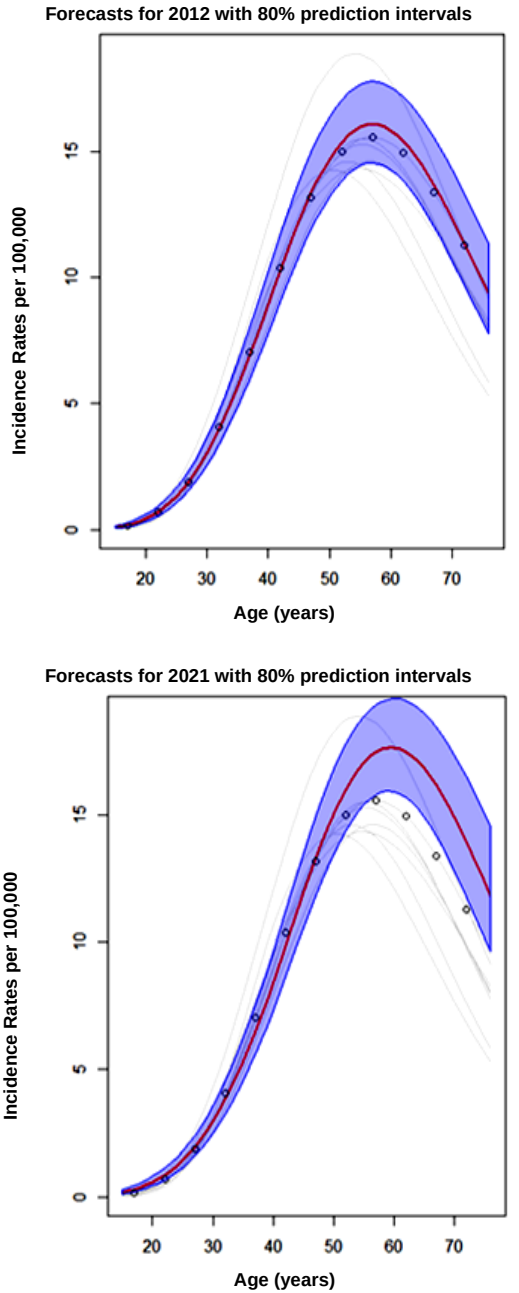


Fig. 5: Forecast for 2012 and 2021 along with 80% prediction intervals. Actual values for 2011 are shown as circles.

In Figure 5, we also plot the incidence forecasts for forecasts of 2012 and 2021, along with 80% prediction intervals to check the forecast uncertainty. There is a clear increasing trend in both forecasts. The forecast graph suggested that incidence rates will rise with time, especially for higher ages (age > 50).

4. DISCUSSION

In this study, we have estimated the future age-specific breast cancer incidence trends by using functional forecasting methods. We have shown that incidence rates increased with age, and in each year, highest incidence occurred in women aged 50-54 years. Ten-year forecasts for the entire incidence curves suggest that for younger ages (women aged 15 to 50) incidence rates are expected to remain stable through time, but are expected to rise for women aged 50 and above for the next ten years.

Our estimates are consistent with results of other studies on Pakistani population, showing that the breast cancer is becoming a serious health concern in many Asian countries due to its increasing incidence (Park et al., 2008).Like other developing countries, Pakistan is also facing a double burden of diseases, with significant increase in breast cancer incidence among women (Gilani et al., 2006). Pakistan shows an unusually high incidence of breast cancer that was otherwise observed among the highly affluent populations of the West. Risk of breast cancer among Pakistani women is even higher than several neighboring countries including India (One in nine as compared to one in twenty two in India) (Park et al., 2008).

The strength of our study is the fact that FTS approach requires few assumptions and has a visual character of the projections. Secondly, using FTS approach, we have calculated prediction intervals for future breast cancer incidence rates, which are required to adjust for uncertainty in the mathematical models, and variation in incidence rates. The prediction intervals of this approach are narrow for age specific incidence of breast cancer. The limitations of our study include the fact that we used hospital based data. The birth cohort effects which may alter the incidence of breast cancer, was not included in the model. However in the absence of population based registries, where standardized incidence figures are not available on national level, this will be provide useful information for controlling the overall burden of cancer in Pakistan; and also serve to be utilized for health planning in future research. In conclusion, the breast cancer patterns and future incidence rates reported in our study provide valuable leads for cancer control in Pakistan.

5. ACKNOWLEDGEMENT

We are deeply thankful to the oncologists, Dr. Amir Maqbool, Clinical Oncologist, KIRAN; and Dr. Noor Muhammad Soomro, Clinical Oncologist, Civil Hospital, Karachi, for their help and cooperation in provide data sources.

We are also grateful for the support and guidance provided by the Department of Statistics, University of Karachi.

REFERENCES

1. Afsar, N.A., Kulsoom, B., Mateen, A., Ahmed, S., Tehseen, M. and Ahmed, A. (2010). Breast cancer pattern and chemotherapy response: An institutional study in Pakistan. *Asian Pacific Journal of Cancer Prevention* 11(3), 825-830.
2. Cleveland, W.S. and Devlin, S.J. (1988). Locally weighted regression; an approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.*, 83, 596-610.
3. Erbas, B., Hyndman, R.J. and Gertig, D.M. (2007). Forecasting age-specific breast cancer mortality using functional data analysis. *Stat. Med.*, 26, 458-70.
4. Gilani, G.M., Kamal, S. and Gilani, S.A.M. (2006). Risk factors for breast cancer for women in Punjab, Pakistan. *Pak. J. Statist. & Oper. Res.*, 2(2), 17-26.
5. Hyndman, R.J. and Ullah, M.S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Comput. Stat. Data Anal.*, 51, 4942-56.
6. Hyndman, R.J. (2013). Demography: forecasting mortality and fertility data. R package version 1.15". [http://robjhyndman.com/software/demography]
7. Hyndman, R.J. and Shang, H.L. (2010). "Rainbow plots, bag plots and box plots for functional data. *J. Comput. Graph Stat.*, 19(1), 29-45.
8. Kneip, A. and Utikal, K.J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.*, 96(454), 519-532.
9. Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting US mortality. *J. Amer. Statist. Assoc.*, 87(419), 659-671.
10. Naeem, M., Khan, N., Aman, Z., Nasir, A., Samad, A. and Khattak, A. (2008). Pattern of breast cancer: experience at lady reading hospital, Peshawar. *Journal of Ayub Medical College Abbottabad*, 20(4), 22-25.
11. Park, S., Bae, J., Nam, H.B. and Yoo, K.Y. (2008). Cancer etiology in Asia. *Asian Pac. J. Cancer. Prev.*, 9, 371-80.
12. Ramsay, J.O. and Dalzell, C.J. (1991). Some tools for functional data analysis with discussions. *J. Roy. Statist. Soc. Series B*, 53,539-572.
13. Ramsay, J.O. and Silverman, B.W. (2005). *Functional data analysis*. Springer.
14. Yasmeen, F., Hyndman, R.J. and Erbas, B. (2010). Forecasting age-related changes in breast cancer mortality among white and black US women: A functional data approach. *Cancer Epidemiology*, 34(5), 542-549.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 87-92

THE USE OF ANALYSIS OF VARIANCE IN DATA MONTHLY SUMMARY REPORT: A CASE STUDY OF EAST NILE MODEL HOSPITAL

Siraj O. Omer
Department of Medical Statistics, East Nile Model Hospital, Khartoum, Sudan
Email: sirajstat@yahoo.com

ABSTRACT

Monthly reports are providing management with a summary of the activities of the health care. The application of analysis of variance development and availability of sophisticated statistical software has facilitated the evaluation of health care and their services. The purpose of this study is to use the approach of analysis of variance in monthly data reports for evaluating the differences between the three periods (June, July and August), which as were represents duplicates into three treatment groups. Frequency of patients such as internists, general surgery, urology, bones, kids Obstetrics and Gynecology Clinic has been measured. The result show that there were statistically significant between treatment group ($p>0.0001$) with $R2=96\%$. This paper provides guidelines for organizing and analyzing the case study applied in health data so that they produce interpreting results, significantly outperform and decision making.

KEYWORDS

Monthly reports; Analysis of Variance; Hospital.

1. INTRODUCTION

Data analysis is very important in health care research (Corbin et al, 2005). In order to achieve some targeted progress in health quality and strengthening evaluation for development services, the use of statistical methodology have been applied by simulation in all area (Wang et al. 2013). Monthly reports are recorded by each health centre as a means of providing management with a summary of the activities of the health care (Liu, 2009). Hospital has collected almost data records of outpatient appointments by department of Statistics for daily, monthly and annually writing report for example frequency of patients or outpatients from internists, general surgery, urology, bones, kids Obstetrics and Gynecology Clinic etc. (Williams, 1987). Outpatient data originally released on an 'information basis and outpatient services are defined as services for patients who may attend a consultant or other medical clinic, or who have an arranged meeting with a consultant or a senior member of their team out with a clinic session. Outpatients are categorized as new outpatients or follow-up (return) outpatient (Zivot, 2005).

2. ANALYSIS OF VARIANCE (ANOVA)

Analysis of variance (ANOVA) is an approach used to identify differentially expressed genes in complex experimental designs (Julien and Gibson, 2007). A method which partitions the total variation in the response into the components (sources of variation) in

the above model is called the analysis of variance (ANOVA). It requires the following four assumptions to be met (i) addictively of factors effects, (ii) constancy of error variance, (iii) normality of experimental errors, (iv) independence of experimental errors (Singh et al. 2012). The primary purpose of analysis of variance (ANOVA) is to estimate experimental error variance. In case of frequency of patients /outpatients in different clinic its need a test for significant of the treatment factor, in terms of equality of means of all the treatments and a global comparison (Geoffrey et al. 2004). Usually after finding that that treatment is statistically significant using the ANOVA, one would want to know which of the means contributed to the significance of the treatment (i.e., which treatment levels are particularly different from each other) (Siraj et al. 2010). ANOVA is a method of great complexity and subtlety with many different variations, each of which applies in a particular experimental context (Armstrong et al. 2002). A two-way ANOVA considers a set of treatments caused by varying two factors simultaneously

There are three main advantages of two-way ANOVA:

1. It is more efficient to study two factors simultaneously rather than separately.
2. We can reduce the residual variation in a model by including a second factor thought to influence the response.
3. We can investigate interactions between factors.

In practice, the statistical descriptive of outpatient's frequency observes variation in an illness property from to place and in sample from the same bag filled at any one place. Various analytical processes advocated in medical statistics would lead to meaningful comparison (Zhuang, 2011).

3. MONTHLY REPORT

Monthly reports are one of a series of reports that evaluate the methods of the medical care utilization (PHS, 198). Monthly report provides information on monthly reporting to the medical statistics to be adopted for all major capital investment projects from inception to the completion of construction; commissioning and hand over and also assisting with their governance responsibility to:

- Simplify maintaining of health data with monitoring of the scope and status;
- help in making decisions and providing endorsements to other roles associated with good project management;
- Provide key issues for action and attendant information knowledge of health care and recommendations arising from the analysis undertaken (Thomas, 2012)

4. STATEMENTS OF THE PROBLEMS

One of executive monthly reports summary is to provide the variations, recommendations and endorsements; therefore the analysis of variance is popular method use statistical inference of health data in this study. On writing report need an evaluation so, the use of analysis of variance will sharing evaluation results and on communication strategies. And share the draft report with stakeholders and revise as needed to be followed by follow up as well as disseminate evaluation report by scientific method rather than traditional manner.

5. OBJECTIVE OF STUDY

The purpose of this study is to use the analysis of variance in monthly data reports for evaluating the differences between the three periods (June, July and August), which as were represents duplicates into three treatment groups.

6. METHODOLOGY

We define frequency data variables as those observed at the fundamental of clinic people consisted of 9 variables/ treatment as x1, x2, …, X9 and define frequency data monthly as fixed variables as those observed at sampling considered of 3 treatment /samples either as temporal aggregates or as skip samples of their high frequency values. A variable is skip sampled when, for example, it is generated every month but is sampled every third month, say, in the last month of every quarter.

6.1 Model of Frequency Data

In this section, we define frequency data as two way analysis of variance and frequency of outpatients as the factors and month representative as sampling. The study is about a n experiment involving frequency of patients and their monthly. The model for the Two Way ANOVA.

Let x_{ijk} represent the kth observation from population having factor A (Months) at level I and factor B (Clinics) at level j. The statistical model is

$$x_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

where $i = 1,\ldots,I$ and $j = 1,\ldots,J$ and $k = 1, n,\ldots,n_{ij}$, which the deviations e are from normal distribution $N(0,\sigma)$.

6.2 Data Collection and Processing

Data management and analysis were performed using SPSS software, Statistical descriptive analysis of variance the primary outcome of this data monthly report. The fieldwork began on monthly data report at East Nile Model Hospital a frequency of outpatient clinic were selected as two way analysis of variance. Data collection was carried out during June July and August, 2010.

7. RESULTS AND DISCUSSIONS

A number of summary tables from analyses of the two set of data keeping in view the results. This was obtained using two -way analysis using SAS software.

Table 1:
Analysis of Variance of Frequencies of Outpatient Over Three Months

Source	DF	SS	MS	F	P
Clinics	7	1938118	276874	86.39	0.0001
Month	2	43188	21594	6.74	0.009
Error	14	44867	3205		
Total	23	2026173			

Table 1 show the analysis of variance of frequencies of outpatients. The overall F-test is significant for these data by reference to this distribution of Fcator A and B respectively are (*FA* =86.36) and (FB= 6.74), indicating that the means are statistically significant factor A (Clinics) (p>0.0001) and Factor B (Month) (p>0.009). The analysis of variance summary show the coefficient of variation (CV) is 17.17 and coefficient of determination is 96%.

Table 2
Total Frequency by the Patient in Outpatient

Clinic	June	July	August	Means
Internists,	647	765	499	637.00 b
kids Obstetrics	238	308	242	262.67 cd
General surgery,	380	401	227	336.00 c
Gynecology	981	938	839	919.33 a
Urology,	99	219	81	133.00 def
Kidney Surgery	84	109	73	88.67 ef
Dental	314	212	189	238.33 cde
Heart	23	22	23	22.67 e
Total	2782	2974	1274	2343.33

Frequency of patients such as internists, general surgery, urology, bones, kids obstetrics and gynecology clinic has been measured. These have the same interpretation as univariate *t*-tests, but they test the pair wise comparison of which clinics show high patients difference between the months. In this example a Teukey test for pair-wise comparison were used. For the analysis of the experimental removal of grazers, there was a significant difference among all pairs of treatments: assemblages colonizing surfaces in cages differed from those in the open or in cage controls thus there are strong differences among the clinics in different months. ANOVA enables not only the effect of individual factors to be estimated but also their interactions; information which cannot be obtained readily when factors are investigated separately.

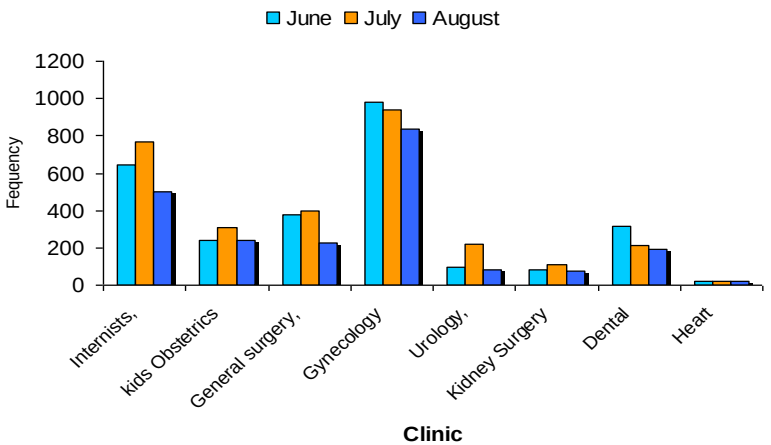


Fig. 1: Shows the Monthly Frequency of Patients from June, July & August 2012

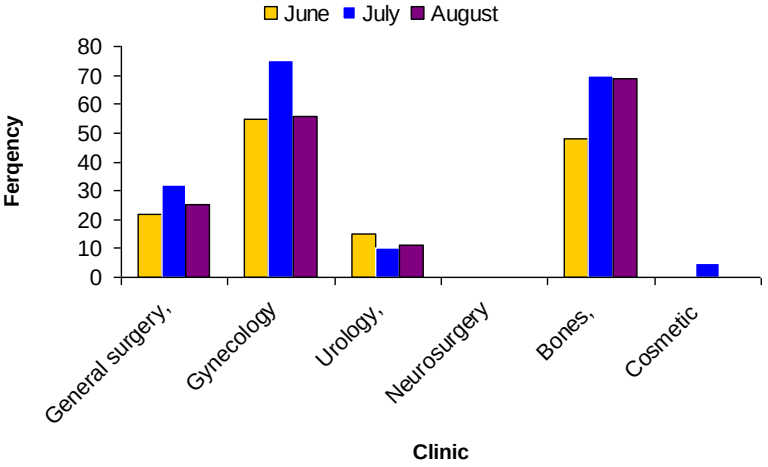


Fig. 2: Shows the Monthly Surgical Operations of Patients from June, July & August 2012

These figures above show the monthly report of frequencies patients by type of clinic and take on these results the decisions and recommendations on the movement of frequencies patients and size of patients and its relationship to the number of medical staff. And determine optimum service to patients in the hospital and also depends on these forms in comparisons between the months with the size frequency by type of kidney clinic. The analysis of variance is applicable for examining how the difference in data monthly Reports of hospital every three months will stays followed a similar pattern in annual report. This method can be used on monthly data reports, for example, data for three months first and then add the next and so on until the end of the year we have 12 months. Performance information on outpatient services is limited. It has highlighted three key areas that give cause for concern. Despite the fact that more people encounter secondary care in this setting than in any other, there is a lack of robust data and information upon which to plan and manage outpatient services, from the information that is available there are some worrying patterns. For example, there are doubts about whether the available costing data re sufficiently robust to support benchmarking and more work needs to be done to ensure variations do not simply reflect different accounting treatments.

8. CONCLUSION

Monthly data report is useful for evolution the health needs The use of analysis of variance will improving most measures and providers efficient Medicare frequency in hospital, it be a key question is whether current Medicare are adequate to cover the view patients. Therefore, the knowledge the size of the doctors most wanted to meet the needs of the hospital medical staff. This paper provides guidelines for organizing and analyzing the case study applied in health data so that they produce interpreting results, significantly outperform and decision making. Statistical information such as analysis of variance on monthly data report of health access to and use of medical services,

associated charges and sources of payment, and health insurance coverage. Further development is that the use ANOVA nonparametric, the approach is summation that all patients are equivalent, in fact one may expect variation in frequency of patient or outpatient are not data normally distribution.

REFERENCES

1. Armstrong, R.A., Slade, S.V. and Eperjesi, F. (2000). An introduction to analysis of variance (ANOVA) with special reference to data from clinical experiments in optometry. *Ophthal. Physiol. Opt*., 20, 235-241.
2. Ayroles, J.F. and Gibson, G. (2007). Analysis of Variance of Microarray Data. *DNA Microarrays, Methods in Enzymology*, 411, 214-233.
3. Keyes, G.R., Singer, R., Iverson, R.E., McGuire, M., Yates, J., Gold, A. and Thompson, D. (2004). Analysis of Outpatient Surgery Center Safety Using an Internet-Based Quality Improvement and Peer Review Program, *Plast. Reconstr. Surg.*, 113(6), 1760-70.
4. Liu, Q. (2009). On Portfolio Optimization: How Do We Benefit from High-Frequency Data. *Journal of Applied Econometrics*, 24(4), 560-582.
5. PHS (1987). Evaluation of Data Collection Frequency and the Use of a Summary in the National Medical Care Utilization and Expenditure Survey, Public Health Service, National Center for Health Statistics Series. *A Methodological Report* No. 17.
6. Singh, A.S. and Masuku, M.B. (2012). An Insight in Statistical Techniques and Design in Agricultural and Applied Research. *World Journal of Agricultural Sciences*, (6), 568-584.
7. Siraj, O.M., Eltahir, S.A. and Murari, S. (2010). Comparing Statistical Software for Outputs on Analysis of Variance in Agricultural Experiments, Proceedings of the *15th European GenStat and ASRem Applied Statistics Conference*). http://www.vsni.co.uk/ resources/events/european-genstat-and-asreml-applie statistics- conference-2010.
8. Thomas, K.T. (2012). *Monthly Statistical Report*, Research and Planning Division www.doc.alabama.gov Compiled and Published.
9. Wang, O., Li, M., Zu, H., Gao, M., Cao, C. and Xia, L.C. (2013). A quantitative evaluation of health care system in US, China and Sweden. *Journal of Society for Development in New Net Environment in B&H HealthMED*, 7(4).
10. Williams, B.K. (1987). The use of analysis of variance procedures in biological studies. *Applied Stochastic Models and Data Analysis*, 3(4), 207-226.
11. Zhuang, X. (2011). Portfolio Selection with High-Frequency Data. Economics 201FS at the Duke Community Standard was upheld in the completion of this project.
12. Zivot, E. (2005). *Analysis of High Frequency Financial Data with S-PLUS*. Unpublished paper by Bingchen Yan and Eric Zivot. Data and S-PLUS scripts are available at http:\\faculty.washington.edu\ezivot\ezresearch.htm.

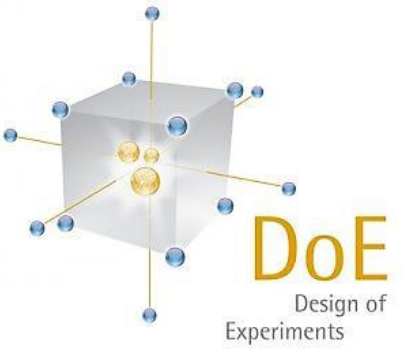
Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 93-106

OPTIMIZATION THROUGH FRACTIONAL FACTORIAL DESIGN: EVALUATING THE FACTORS THAT INFLUENCE THE CRISPINESS AND SOFTNESS OF THE PIZZA DOUGH

Zara Omar[1] and **Hafiz Mohammad Aamir**[2]
[1] Institute of Business Management, Karachi, Pakistan
Email: zara.omar@iobm.edu.pk
[2] Hamdard University, Karachi, Pakistan

ABSTRACT

The objective of the study to evaluates the factors that influence the crispiness and softness of the pizza dough. The scope of the study was limited to fractional factorial experiment of Dawlance and Panasonic products. A most important advantage ensuing from the application of the fractional factorial design is that the number of experiments required to estimate multifactor interactions is restricted. In present study, the required number of experiments was condensed to 32 from the 64 essential with conservative factorial designs. Six experimentally proscribed factors were investigated at two levels. The six factors were (A) Flour type (B) Yeast (C) Time of fermentation (D) Temperature of Baking (E) Time taken in baking (F) Two Brand of Microwaves. Significant main effects were found for factors B, C, and D; whereas significant interactive effects were found only for factor A with factor D. In light of the findings concrete recommendation were made.



KEYWORDS

Fractional design; multifactor interaction.

INTRODUCTION

Pizza is actually modern Italian for pie. The origin of the term is murky but is said to be derived from an Old Italian word meaning "a point," which later evolved to the Italian word pizzicare, meaning "to pinch or pluck." A Neapolitan dialect of the word first shows up in print in 1000 A.D. in the form of picea or piza, assumably referring to the way the hot pie is plucked from the oven.

Early forms of pizza were most likely what we call focaccia today. The tomato did not reach Italy until the 16th century, hence no tomato sauce for the pie we know as pizza. The addition of mozzarella cheese (initially made from buffalo milk) did not come about until the 19th century.

The standard tomato, basil, and mozzarella pizza is first credited to Raffaele Esposito of the Pizzeria di Pietro, who dedicated the creation to Queen Margherita in 1889. The colors nicely coincided with the colors of the Italian flag.

Italian immigrants brought pizza to the United States, but it was mostly confined to small areas of the Northeast. Once in America, the pizza stretched out to 18 inches in diameter or more. The first pizzeria opened in New York City in 1896. Yet pizza did not really gain popularity outside the Northeast until returning World War II servicemen returned from their tours in Italy with a hankering for the pizza they enjoyed in Naples. Pizza is described as a flat pie made from a yeast dough topped with various cheeses, vegetables, meats, seasonings, and other ingredients. Nowadays, you will find many variations, including dessert pizzas. The topping of choice by the majority is pepperoni, followed by Italian sausage and mushrooms.

The basic formulation of the dough includes flour, water, salt, sugar, and bakers' yeast. The pizza dough is a significant fraction of the product and its appearance, texture, and taste are important attributes for its acceptance by consumers and identification.

Wheat flour is a structural component and an essential ingredient for obtaining the dough thus presenting a key role in the quality of baked foods due to its ability of forming extensive and elastic dough that expands. The process steps responsible for increased dough volume are: mixing and forming the dough, dividing and molding, fermenting, rolling, and baking. These steps are designed to develop the dough, forming a sponge-like structure, and stabilize the porous structure during baking.

The dough is usually prepared by the direct process using baker's yeast to ferment at temperatures between 23 and 26 ºC for a period of time required for leavening or the time necessary to raise the initial volume 2-5 fold before shaping the pizza discs.

Some authors indicate the fermentation temperature of 35 ºC for 60 minutes (CLARKE; FARRELL, 2000), 25 ºC for 6-8 hours (FORMATO; PEPE, 2005), and Wang et al. (2005), who produced pizza dough from mixtures of soybean and wheat flour, pre-cooked by extrusion, considered a shorter time of fermentation of 25 ºC for 40 minutes . It is important to note that there is not a standardization of the pizza dough production method, and that it also reflects in the lack of identity for parameters as final product appearance, height, texture, and other important quality aspects. As written by Coppola, Pepe and Maureiello (1998), this kind of product is widely characterized by highly variable and often unsatisfactory quality.

In this paper, the focus was to produce characteristic soft pizza dough with bubbles and crispy edges and soft in the center. The dough was prepared with the usual ingredients, fermented at a temperature range from room temperature for 1 to 1.5 hour enlarged and baked. The influence of the variables flour, yeast, time and temperature on the dough was confirmed with positive and significant effect, observed for the development or maximum height of the dough. The results showed that the significance of the factors on the pizza dough. The Experiment was conducted and the analysis yielded the following factors as the most important factors. An interaction effect was also noticed between the factors. A residual analysis was done to test for defects such as non-normality, non-independent and non-constant variance. Therefore, the objective is to produce characteristic soft, airy, crispy pizza dough evaluating these attributes.

LITERATURE REVIEW

According to Antony and Capon (1998) “A number of successful applications of Experimental Design to improve their process performance have, process variability reducing improve process efficiency, etc. by many manufacturers in the last fifteen years reported. Research has shown that the application of the technique of erectile dysfunction by the engineering fraternity in industry and services is limited and often performed incorrectly in the application. In other words, there is no knowledge gap in the understanding statistics of engineers with Experimental Design as tools to solve problems”. To finding the best combination of the factors producing Pizza Doe the statistical background and application must be strong.

Sudarsanam and Frey (May 2005) stated in research article “Experimentation is a key driver of the development and marketing costs. The pressure to reduce the resource requirements of experimentation especially in economically competitive industries. The common objectives of erectile dysfunction are: improving the efficiency of scientific research and reliable, efficient product and process optimization and improving the robustness of the system to changing or uncertain environmental conditions, internal breakdown, manufacture, use or customer profiles”.

According to Fannin and Marcus ed. (1981) “Fractional factorial designs allow testing of a number of factors and their interactions simultaneously with a substantial reduction in the number of individual experiments (experimental runs) required, but without loss of important information”. Changing the factor levels is often expensive, so it is not advisable to leave the number of necessary changes to chance. For this reason, have run jobs that offer the minimum number of factor level changes and simultaneously sought to minimize the possible influence of adverse factors on the experiments.. In this paper, we present the best sequences for designs with 32 experiments.

METHODOLOGY

Experimental Design for Pizza Dough Production

A 2^{6-2} fractional factorial design with six factors (flour type, yeast type, time of fermentation and temperature, brand of microwave and time of baking), two levels coded as -1 and +1 and two repetitions were used for producing the pizza dough samples. 16 treatments were performed. The developing times and temperatures of the dough were 30, 36, and 42 minutes and 27, 30 and 33 ºC respectively.

Preparing the pizza dough

Pizza Dough Recipe
Ingredients:
1 1/2 cups warm water
2 tsp yeast
Pinch of sugar
600g (4 cups) plain flour
1 tsp salt
60ml (1/4 cup) oil

1. **Step 1**

Combine the water, yeast and sugar in a small bowl. Set aside for 5 minutes or until foamy. Combine the flour and salt in a large bowl and make a well in the centre. Add the yeast mixture and oil. Use a round-bladed knife in a cutting motion to mix until the mixture is combined. Use your hands to bring the dough together in the bowl.

2. **Step 2**

Brush a bowl lightly with oil. Turn the dough onto a lightly floured surface and knead for 10 minutes or until smooth and elastic. Place in the prepared bowl and turn to coat in oil. Cover with plastic wrap and set aside in a warm, draught-free place to rise or until dough doubles in size.

The ingredients of the pizza dough were based on the flour. The base formulation was wheat flour (BRASIL, 2005), crystal sugar , refined salt , soybean oil , baker's yeast and drinking water between 4-5 ºC (GIANNOU; KESSOGLOU; TZIA, 2003) with changes in the percentage of sugar, yeast, and water. After mixing the dry ingredients, water was incorporated and the dough beat for 6-7 minutes until obtaining a homogeneous, smooth dough with full development of gluten performing the “windowpane test" by flattening and stretching the dough with the fingers until it is translucent.

After that the dough was divided into several portions, rounded and fermented according to the experimental design.

STATISTICAL ANALYSIS

Statement of the Problem:

The crispness and softness of the pizza dough is measured by the different factors. Our objective was to find out what factors affect these factors. We wanted to find the levels of these significant factors at which the dough is soft and crisp. We stated our problem, choose the factors, their levels and ranges and we selected the response variable.

Our objective can be summarized as follows:

Objective: To find the factor which influence the crispiness and softness of pizza dough the responses of the experimental design for the statistical analysis considering the results obtained by the dough.

Choice of Factors and Levels

The following factors and their level were considered for the experiment

S.No	*Factors Levels*	*Factors*	*High Level(+)*	*Low Level(-)*
1	*Flour Type*	*A*	*Whole-wheat*	*Grain wheat*
2	*Yeast*	*B*	*Instant*	*Active*
3	*Time of fermentation*	*C*	*Max (90mint)*	*Min(60mint)*
4	*Temperature of baking*	*D*	*33C*	*30C*
5	*Time taken in baking*	*E*	*Max (36mint)*	*Min(30mint)*
6	*Brand of Microwave*	*F*	*Dawlance*	*Panasonic*

Response variable

The response variable is the height of the dough measure with the vernier caliper. In conducting the experiment object varying the factors as per the design matrix. The height of the dough was done on a scale of 1 – 6 with 6 being the highest ranked and 1 being the lowest.

Choice of design

The different number of factors that were decided in conducting the experiment was 6. So initially the team decided to conduct a 2^6 design. On further research found out that of the six factors only four main factors were really important. Therefore going by the Sparsity of effects principle these higher order interactions could be neglected. But it was not sure about the two-order interactions. This called for resolution IV design. We chose two levels of each factor, in an attempt to screen factors not significantly affecting the response variables.

Furthermore we know that a fractional factorial such as one-quarter fractional factorial is helpful to estimate the same with less number of runs. So the final design selected as a screening experiment was 2^{6-2}.

Design Test Matrix

The design test matrix for the 2^{6-2} is as shown below. Design Expert was used to arrive at the above design matrix to test combination and run orders. For this study, six major factors were selected from a broader range of ideas. The six factors were flour

type, yeast, time of fermentation, temperature of baking, time taken in baking and two brand of microwaves brand. Our study was designed around a two level factorial model. All variables had only two values. To limit the continuous variables, range-finding trials were conducted to set low and high levels for each of the experiments.

A statistically desirable array of combinations of the low and high levels was built, for a total of 16 runs, half the total number (32) of combinations possible. Such a fractional factorial design is sufficient. In fact, making more runs would not add to our knowledge. It is not necessary to run all 32 combinations to study the interactions between factors. The runs were randomized to protect the study against lurking variables such as changes in the protocol that could otherwise confound the study. To simplify the administration of such a study, we used Minitab software. It handled randomizing the samples and the statistical analysis.

Table shows the standard array for six factors and 16 experiments. It also shows the run order and observed responses. To estimate pure error, two repeat runs were planned. These extra experiments were meant to be run at factors fixed at low (-) or high (+).

I + ABCE + ADEF + BCDF
A + BCE + DEF + ABCDF
B + ACE + CDF + ABDEF
C + ABE + BDF + ACDEF
D + AEF + BCF + ABCDE
E + ABC + ADF + BCDEF
F + ADE + BCD + ABCEF
AB + CE + ACDF + BDEF
AC + BE + ABDF + CDEF
AD + EF + ABCF + BCDE
AE + BC + DF + ABCDEF
AF + DE + ABCD + BCEF
BD + CF + ABEF + ACDE
BF + CD + ABDE + ACEF
ABD + ACF + BEF + CDE
ABF + ACD + BDE + CEF

Analysis of Normal Probability Plot

The analysis of the collected data was done using “Minitab” analysis packages.

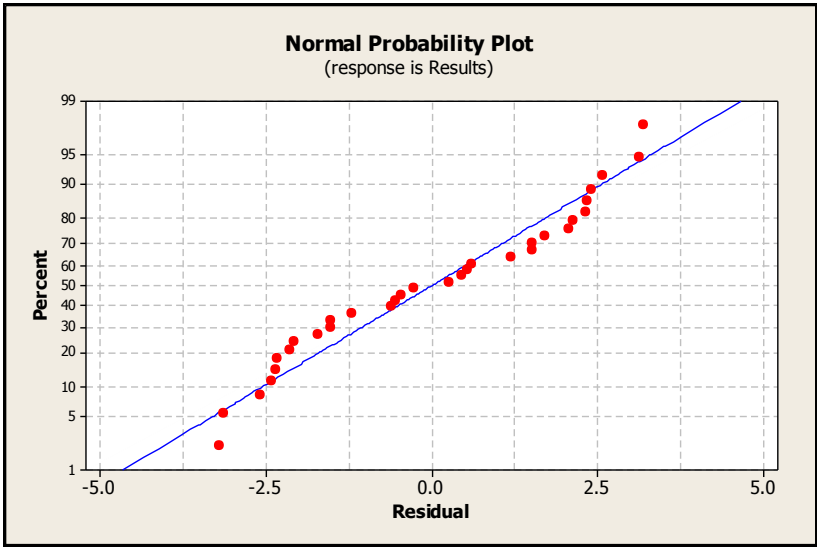


Fig. 1: Normal Probability Plot

The data was tested for normality using “Minitab” package. The Normal plot of the residuals (Fig 1) showed that almost the residuals would pass the fat pencil test, which proved that the residuals were normally distributed.

Analysis of the Half Normal Plot and Normal Plot:

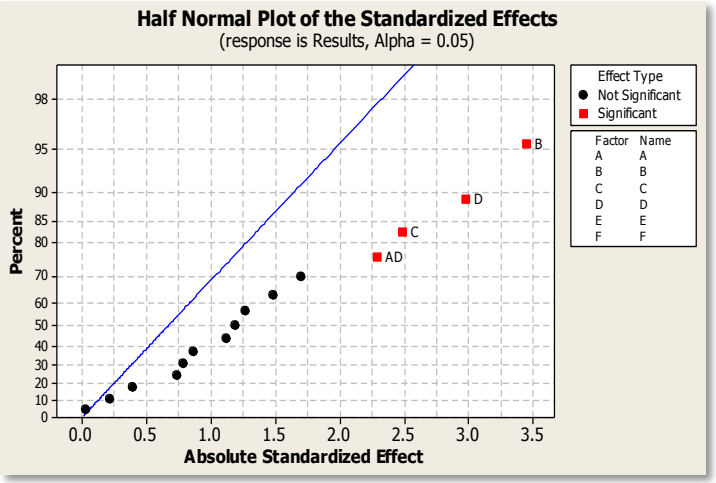


Fig. 2: Half Normal Plot

The half plot was plotted which shows that the factors B, C and D are significantly away from the straight line. We can conclude that these factors are not normally distributed with a mean of zero and constant variance and these factors are significant. The effect of the other main factors and the interaction effects are fairly near to the line and can thus we can conclude that they are normally distributed with a mean of zero and a constant variance and do not have significant effects. The Half-Normal plot is shown in figure 2.

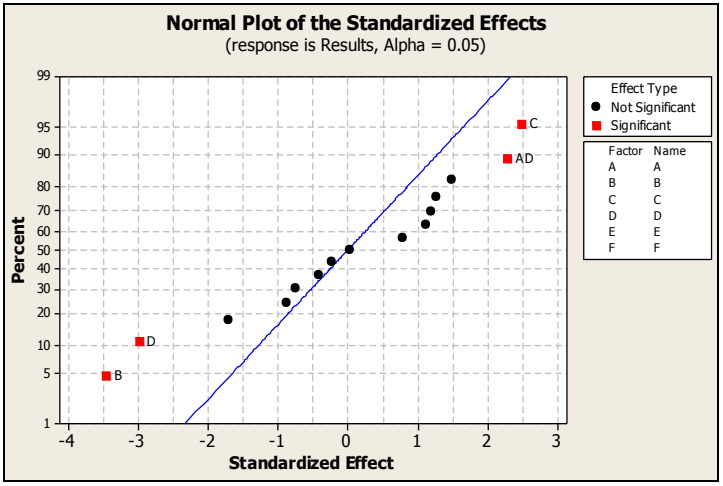


Fig. 3: Normal Plot

After analyzing the normal graph (Fig 3), we found that the effects B, C and D were the only effects that were significantly away from the straight line. So, we concluded that these factors were not distributed with a mean zero and a constant variance. So, these three effects were the only effects that were significant. Since the factors are significant, we had a doubt that their interaction was also significant.

Analysis of Variance (ANOVA):

After studying the "Half Normal plot" and the "Normal graph", we proceeded with the analysis if variance. The model had 1 degree of freedom, which comprised of one degree of freedom for each of the significant effects. The block had one degree of freedom and the residuals had 15 degrees of freedom. The total correlative degrees of freedom were 16. The value of the Type I error was assumed to be 0.05. The ANOVA table is shown in Table 1.

Table 1: ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	6	225.73	225.73	37.6216	4.85	0.005
A	1	9.79	9.79	9.7903	1.26	0.278
B	1	92.48	92.48	92.48	11.92	0.003
C	1	48.265	48.265	48.2653	6.22	0.024
D	1	69.031	69.031	69.0312	8.9	0.009
E	1	5.78	5.78	5.78	0.74	0.401
F	1	0.383	0.383	0.3828	0.05	0.827
2-Way Interactions	7		86.907	12.4154	1.6	0.206
A*B	1	0.008	0.008	0.0078	0	0.975
A*C	1	12.5	12.5	12.5	1.61	0.222
A*D	1	40.725	40.725	40.7253	5.25	0.036
A*E	1	1.24	1.24	1.2403	0.16	0.695
A*F	1	17.111	17.111	17.1113	2.21	0.157
B*D	1	11.045	11.045	11.045	1.42	0.25
B*F	1	4.278	4.278	4.2778	0.55	0.469
3-Way Interactions	2	27.418	27.418	13.7089	1.77	0.203
A*B*D	1	22.613	22.613	22.6128	2.91	0.107
A*B*F	1	4.805	4.805	4.805	0.62	0.443
Residual Error	16	124.135	124.135	7.7584		
Pure Error	16	124.135	124.135	7.7584		
Total		31	464.19			

In the analysis, we have six main effects, overall three effects B, C and D are almost significant whereas, effects A, E, F are not significant effect, and the overall two-way interactions are not significant except "AD". In three – way interaction all the effects are not significant but we only have one degree of freedom of error - so this makes this a very low-power test. However, this is the value that you would benefit with a fractional factorial.

Let's look at another plot that we haven't used that much yet - the Pareto plot. This type of plot looks at the effects and orders them from largest to smallest showing you the relative sizes of the main effects.

Main Effects Plot

We can find that all the factors except F have significant affects to the response variable. And the high level of all the three factors will result in soft and crispy dough. To get more information of the interaction effects, we had the interaction plot for result.

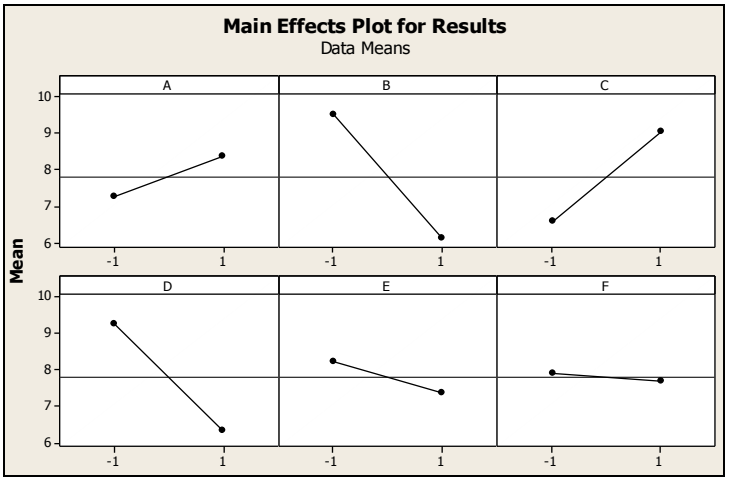


Fig. 4: Main Effects Plot

The data analysis identifies that the factor A and C have a significant linear effect on lower level and the remaining factors B, D and E has effect on higher level a very significant effect on the softness and crispiness of pizza dough. Main effect is the effect of an independent variable on a dependent variable averaging across the levels of any other independent variables.

Interaction Plot

There are no obvious interactions from the plot even though A, B ,C , D , E and F have very weak interactions. When two factors interact, it means that changes in the dependent variable cannot be explained by independent effects of the two factors. Rather, the explanation must be more complicated. The effect of one factor depends on what has happened to the other factor.

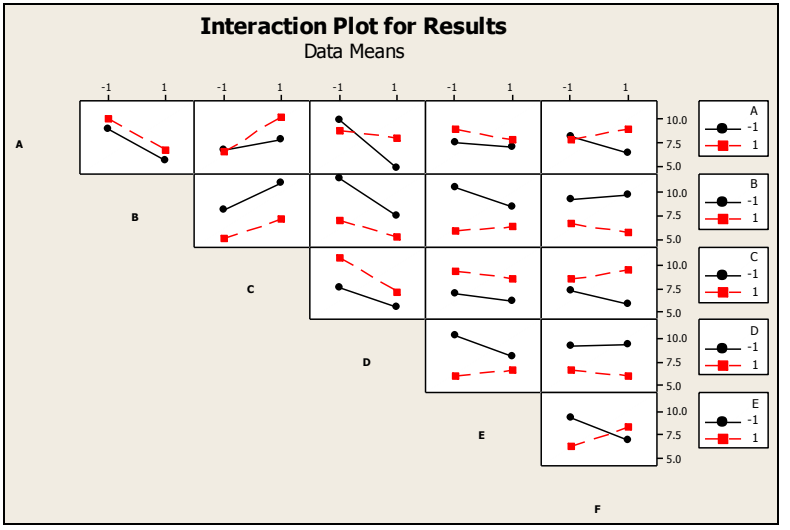


Fig. 5: Interaction Effects Plot

CONCLUSIONS

The statistical analysis of the data clearly indicates the following factors significantly affect the dough

a. The flour type, yeast, temperature and time of rise of flour affect the softness and crispiness of dough.
b. One of the best way to produce high rise of the dough is possible at high level of time of yeast.
c. Factor A – This should be maintained at the low level.
d. Factor B – This should be maintained at the high level.
e. Factor C – This should be maintained at the high level.
f. Factor D – This should be maintained at the high level.
g. Factor AD Interaction – The best results are obtained at low level of factor A and low level of factor D.
h. The model satisfies the normality, independency and constant variance assumptions.
i. Also we had an estimate and we were fairly certain that A, B, C, D would be important and therefore they were kept in the basic design. The relatively less significant factors E and F were then generated using the generators.

Fractional factorial designs allow a large number of variables to be analyzed with a small number of experiments. Many effects and interactions are confounded. The resolution of a design is the sum of the order of confounded effects. A design with higher resolution is considered better.

RECOMMENDATIONS

In this experiment, we did not use different types of ready dough to compare the performance of the home made pizza dough. So in the future using different type of ready-made dough can also be used to perform the experiment. The selection of the instant yeast and different brands should also be randomized. This design would give more true results as compared to the present one. This could introduce certain nuisance variables, which may have affected our results.

Fractional factorials are useful when hope or expect that not all of the factors are going to be significant. We are screening for factors to drop out of the study. In this example, we started with a 2^{6-2} design but when we dropped A, E and F we ended up with a 2^3 design with 1 observation per cell.

This is a typical scenario; you begin by screening a large number of factors and end up with a smaller set. We still don't know much about the factors and this is still a pretty thin or weak design but it gives the information that you need to take the next step. You can now do a more complete experiment on fewer factors.

REFERENCES

1. Angioloni, A.; Berti, C.; Rosa, M.D. (2004). Messa a punto di um preparato per ottenere impasti tipo pizza in brevissimi tempi di miscelazione. *Industrie Alimentari*, 42(442), 1281-1285.
2. Associação brasileira de normas técnicas - abnt. Nbr 14140: alimentos e bebidas - análise sensorial - teste de análise descritiva quantitativa (adq). Abnt, 1998.
3. Autio, K. and Laurikainen, T. (1997). Relationships between flour/dough microstructure and dough handling and baking properties. *Trends in Food Science & Technology*, 8, 181-185.
4. Brasil. Ministério de agricultura pecuária e abastecimento. Associação brasileira da indústria e massas alimentícias. Instrução normativa nº 8, de 2 de junho de 2005. Regulamento técnico de identidade qualidade de farinha de trigo. Diário oficial da república federativa do brasil, brasília, df, 3 jun. 2005. Seção 1, n. 105, p. 91. Disponível em:
5. Barros Neto, B. and Scarmini, I.S. (202). Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria. Campinas: unicamp, p. 223-227, 261-262.
6. Battochio, J.R. et al. (2006). Perfil sensorial de pão de forma integral. *Ciência tecnologia de alimentos*, 26(2), 428-433.
7. Coppola, S.; Pepe, O.; Maureiello, G. (1998). Effect of leavening microflora on pizza dough properties. *Journal of Applied Microbiology*, 85, 891-897.
8. Clarke, C.H. and Farrell, G. (2000). The effects of recipe formulation on the textural characteristics of microwave-reheated pizza bases. *Journal of the science of food and agriculture*, 80, 1237-1244.
9. Czuchajowska, Z. and Pomeranz, Y. (1993). Gas formation and gas retention.i. The system and methodology. *Cereal Foods World*, 38, 499-503.
10. Dobraszczyk, B.J.; (2003). Morgenstern, M.P. Rheology and bread making process. *Journal of Cereal Science*, 38, 229-245.
11. Dutcosky, S.A. (2007). Análise sensorial de alimentos. *Editora Champagnat*, 123-180.
12. Ferreira, V.L.P. (2000). Análise sensorial testes descritivos e afetivos. *Profiqua; SBCTA*, 92-94, 97-98. (manual - série qualidade).
13. Formato, A. and Pepe, O. (2005). Pizza dough differentiation by principal component analysis of alveographic, microbiological, and chemical parameters. *Cereal chemistry*, 82(4), 356-360.
14. Gan, Z.; Ellis, P.R. and Schofield, J.D. (1995). Mini review gas cell stabilization and gas retention in wheat bread dough. *Journal of Cereal Science*, 21, 215-230.
15. Garruti, D.S., Brito, E.S., Brandão, T.M., Uchôa, P., Silva, M.A.A.P. (2003). Desenvolvimento do perfil sensorial e aceitação de requeijão cremoso. *Ciência e tecnologia de alimentos*, 23(3), 434-440.
16. Giannou, V.K.; Kessoglou, V. and Tzia, C. (2003). Quality and safety characteristics of bread made from frozen dough. *Trends in Food & Technology*, 14, 199-108.
17. Larsen, D.M.; Setter C.S.; Faubion, J.M. (1993). Effects of flour type and dough retardation time on the sensory characteristic of pizza crust. *Cereal chemistry*, 70(6), 647-650.

18. Pinho, B.H.S.; Machado, M.I. and Furlong, E.B. (2001). Propriedades físico-químicas das massas de pizza semiprontas e sua relação com o desenvolvimento de bolores e leveduras. *Revista instituto adolfo lutz*, 60(1), 35-41.
19. Rosell, C.M. and Collar, C. (2009). Effect of temperature and consistency on wheat dough performance. *International Journal of Food Science and Technology*, 44, 493-502.
20. Santana, L.R.R. et al. (2006). Perfil sensorial de iogurte light, sabor pêssego. *Ciências e tecnologia de alimentos*, 26(3), 619-625.
21. Scanlon, M.G. and Zghal, M.C. (2001). Bread properties and crumb structure. *Food Research International*, 34, 841-864.
22. Stone, H. and Sidel, J.L. (1992). Sensory evaluation practices. 2. Ed. San diego: academic press. 308 p.
23. Stone, H. et al. (1974). Sensory evaluation by quantitative descriptive analysis. *Food technology*, 28(11), 24-34.
24. Teófilo, R.F. and Ferreira, M.M. (2006). Quiometria II: planilhas eletrônicas para cálculos de planejamentos experimentais, um tutorial. *Química nova*, 29(2), 338-350.
25. Wang, S.H., Oliveira, M.F., Costa, P.d.S., Rosa, A.G. and Ascheri, J.L.R. (2005). Farinhas de trigo e soja pré-cozidas por extrusão para massas de pizza. *Pesq. Agropec. Bras., Brasília*, 40(4), 389-395.
26. Correaa, A.A. Grimab, P. and Tort-Martorellb, X. (2012). Experimentation order with good properties for 2^k factorial designs. *Journal of Applied Statistics*, 36(7), 743-754.
27. Fannin, T.E., Marcus, M.D., Anderson, D.A. and Bergman, H.L. (1981). Use of a fractional factorial design to evaluate interactions of environmental factors affecting biodegradation rates. *Appl. Environ. Microbiol.*, 42(6), 936-43.
28. Li, X., Sudarsanam, N. and Frey, D.D. (2005). Regularities in Data from Factorial Experiments, *Complexity*, 11(5), 32-45.
29. Vanhatalo, E., Bergquist, B. and Vännman, K. (2013). Towards improved analysis methods for two-level factorial experiments with time series responses. *Quality and Reliability Engineering International*, 28(5), 725-741.
30. Taam, W. and Hamada, M. (1990). Detecting spatial effects from fractional experiments: an application in integrated-circuit manufacturing, IIQP Research Report RR-90-02.
31. Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978). *Statistics for experimenters.* New York, NY: John Wiley & Sons Inc.
32. Montgomery, D.C. (2005). *Design and Analysis of Experiments.* 6th ed. New York, NY: John Wiley & Sons Inc.

ANNEXURE A
FRACTIONAL FACTORIAL DESIGN

Factors:	6	Base Design:	6, 16	Resolution:	IV
Runs:	32	Replicates:	2	Fraction:	1/4
Blocks:	1	Center pts (total):	0		

Design Generators: E = ABC, F = BCD
Alias Structure

A	B	C	D	E	F	Results
-1	-1	-1	-1	-1	-1	11.00
1	-1	-1	-1	1	-1	6.50
-1	1	-1	-1	1	1	2.75
1	1	-1	-1	-1	1	8.75
-1	-1	1	-1	1	1	11.75
1	-1	1	-1	-1	1	12.00
-1	1	1	-1	-1	-1	12.25
1	1	1	-1	1	-1	5.50
-1	-1	-1	1	-1	1	4.75
1	-1	-1	1	1	1	5.75
-1	1	-1	1	1	-1	7.25
1	1	-1	1	-1	-1	1.00
-1	-1	1	1	1	-1	3.00
1	-1	1	1	-1	-1	9.25
-1	1	1	1	-1	1	3.00
1	1	1	1	1	1	7.50
-1	-1	-1	-1	-1	-1	14.40
1	-1	-1	-1	1	-1	7.60
-1	1	-1	-1	1	1	5.80
1	1	-1	-1	-1	1	4.60
-1	-1	1	-1	1	1	14.80
1	-1	1	-1	-1	1	14.40
-1	1	1	-1	-1	-1	6.00
1	1	1	-1	1	-1	10.20
-1	-1	-1	1	-1	1	4.20
1	-1	-1	1	1	1	10.40
-1	1	-1	1	1	-1	3.00
1	1	-1	1	-1	-1	7.40
-1	-1	1	1	1	-1	7.80
1	-1	1	1	-1	-1	14.40
-1	1	1	1	-1	1	4.20
1	1	1	1	1	1	8.40

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 107-120

ASTRONOMICAL IMPROVE MODEL OF PRAYER TIMING WITH ERROR ANALYSIS

Maqsood Alam[1], Rabia Tabasum[2] and **Yousra Waseem[3]**
[1] Department of Basic Sciences, DHA Suffa University, Karachi, Pakistan. Email: maqsood.dsu@gmail.com
[2] Department of Humanities & Sciences, National University (FAST), Karachi, Pakistan
[3] Department of Mathematical Sciences, University of Karachi Karachi, Pakistan

ABSTRACT

Islam is a religion that teaches us the several ways to praise the uncountable kindness of Allah Almighty offering Salaat (Prayer) on time is one of them. The knowledge of starting and ending times have the fundamentals importance to the Muslims, As it is *farz* (obligatory) for Muslims to offer five times Salaat a day which are *Fajr*, *Zuhar*, *Aser*, *Maghrib* and *Esha*.

In this research, the astronomical improve mathematical models of Salaat timing are developed in line with the major school of thought of Islamic jurisprudence. The tracking of a true Sun, for calculating the *hour angles* on *elliptic* at five different places and developing the models of *Zuhar* and *Aser* Salaat timing according to *Shafi* and *Hanfi* school of thought are the key objectives of this research. There are various astronomical measures like Astronomical Islamic twilight (AIT), Depression angle, Declination of sun and Equation of time are used in this study to construct the models. Finally, validation of model with available data and error analysis of model outcomes using SPSS 19 will be presented.

Further these modeling and algorithms can be used to make a Java programming application for smart phones without get connected with internet, otherwise all existing applications regarding Salaat timing depends on high speed internet for collecting the initial parameters.

KEYWORDS

Spherical trigonometry; Twilight; Declination of sun; Equation of time; True sun

1. INTRODUCTION

Salaat is the most beautiful way to praise the Allah Almighty for his uncountable kindness. It is compulsory for every Muslim to perform these prayers at correct timing. Prayer time in the time of prophet (S.A.W) was based on human observation of physical phenomena like sunrise, afternoon, sunset, dawn and twilight. Nowadays, the modelling of prayer time based on astronomical calculation and different methods have been developed over the years. The Namaz timing is dependent on the geographical location of

country or city and has place to place and day to day variations. A model for Salaat timing for any location over the globe can be constructed knowing some basic parameters like latitude of any particular location along with declination of sun and equation of time for specific day which is mathematically more effective and efficient way as compare to explain it theoretically. Though Islam gives some relaxation of about +5 or -5 minutes in actual prayer timings, a mathematical modelling is a tool which helps in getting accuracy.

2. LOCATION OF SUN ON CELESTIAL SPHERE

It is compulsory for every Muslim to perform their five time prayer on time. The permissible period of time for performing each prayer will start from one prayer time to another.

i) Fajr Prayer

This prayer starts with the beginning of new day at dawn. There are two different descriptions for dawn: when first light appears at the horizon, second when the light of the early day spreads horizontally across the horizon. Fajr is to be performed at the second dawn known as *Subh sadiq* and end just before sunrises.

ii) Zuhar Prayer

The noontime prayer performed when the sun declines after touched the highest position (Zawal) in the sky (Zenith). According to different school of thought zuhr is performed after 5 min past to the zenith. Zuhr ends at the start of aser time.

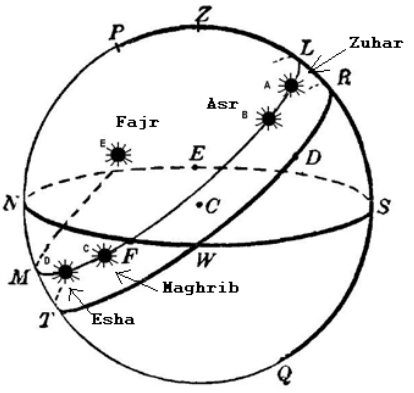


Figure 1

iii) Aser Prayer

Aser Prayer timing depends on the length of the shadow of a stick fixed in the ground. According to Hanfi school of thought, asr timing starts when length of the shadow become twice the original length of the object, whereas Shafi school of jurisprudence asr begin when the length exceed the length of the object.

iv) Maghrib Prayer

The prayer perform just after sunset, however physical condition like refraction of light or height of the building in a city leads us to fix the time of this prayer 3 minutes after the theoretical time of sunset as it appears in newspapers and ends at start of isha.

v) Isha Prayer

The timing of Isha prayer starts at dusk. There are two type of dusk, both called Shafaq. Just after sunset when the sky is in red color known as “Shafaq al Ahmar” and later when red colour disappears, known as Shafaq al Abyad”.The major school of thought fix the isha prayer either at the disappearance of Shafaq al Ahmar or at the disappearance of Shafaq al Abyad.

3. ASTRONOMICAL COORDINATES SYSTEM

There are different coordinates systems to locate any celestial object like true sun on the celestial sphere. The first one mention below is observer dependent whereas others are observer independent.

i. (Altitude and Azimuth) = $(a\ , A)$

ii. (Hour Angle and Declination) = $(H\ , \delta)$

iii. (Right Ascension and Declination) = $(\alpha\ , \delta)$

iv. (Celestial Latitude and Celestial Longitude) = $(\varphi\ , \lambda)$

The coordinate systems $(\alpha\ , \delta)$ and $(H\ , \delta)$ are used directly or indirectly to determine the position of sun on celestial sphere as shown in the figure 1, 2 and 3 (See Appendix).

4. MODEL CONSTRUCTION

4.1) Zuhar Prayer Model

There are several methods and technique that have been used to derive the formula for prayer timing but this is a unique way to model prayer timing in which mathematics of longitude play a main role. In figure 2 shows the position of "True Sun" just after ZAWAL time (12:00^h Noon). According to equation of time

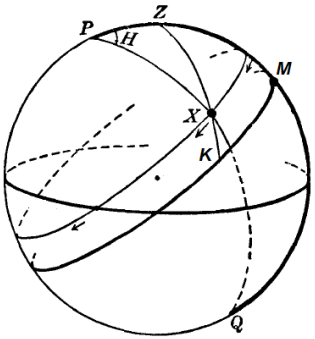


Figure 2

$$E = R.A.M.S - R.A\ \Theta$$
$$E = \gamma M - \gamma K \qquad (1)$$

or can say that

$$H.A\Theta = RM + MK \qquad (2)$$

Form fig 4

$$RM = H.A.M.S$$

and $\gamma M = K\gamma + MK$

$$MK = M\gamma - K\gamma$$

from equation 1 $MK = E$ Put in (2)

$$H.A\Theta = H.A.M.S + E \qquad (3)$$

Convert the above formula in to local mean time (LMT)

$$L.M.T = H.A.M.S \pm 12$$

$H.A.M.S = L.M.T \pm 12$ put in (3)

$$H.A\Theta = L.M.T \pm 12^h + E \qquad (4)$$

$H.A\Theta = L.M.T \pm longtitude\ of\ l$ (Position)

$L.M.T = U.T \pm longtitude\ of\ l$ Put in (4)

$$H.A\Theta = (U.T \pm long\ of\ l) + 12^h + E \quad (5)$$

$U.T = Z.T \pm long\ of\ l$ put in (5)

$$H.A\Theta = L.M.T \pm 12^h + E \quad (6)$$

To make above formula more accurate we can multiply 1/15 with difference of lambda because ,there are 24 meridians line on the globe which has 360 degree, so each longitudinal difference has 360/25 = 15 degree.

$$H.A\Theta = 12^h + \left(\frac{\lambda_{ZoneTime} - \lambda_{Local}}{15}\right) + \frac{E}{60}$$

where $\lambda = \lambda_{ZoneTime} - \lambda_{Local}$

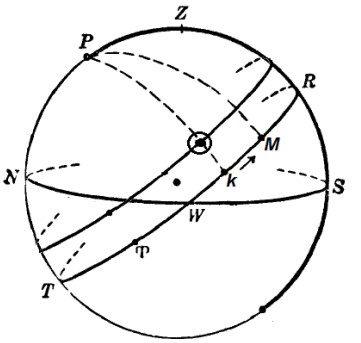


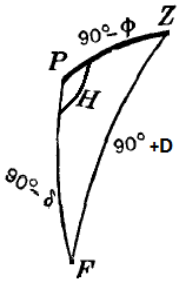
Figure 3

$$Zuhar = 12^h + \frac{\lambda}{15} + \frac{E}{60} + 1^{\min}(\text{Sun's semi diameter}) \quad (7)$$

The above equation (7) represents a mathematical model of *Zuhar* prayer.

4.2) Maghrib Prayer Model

The position of sun soon after sunset tells about the starting of Maghrib prayer as shown point *G* in the following fig 4. With the help of Spherical trigonometry we can make hour angle formula for Maghrib prayer.



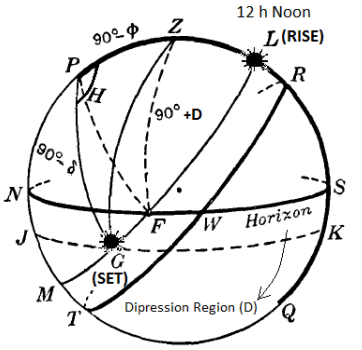


Figure 4

Applying the Cosine Law on Spherical triangle PZF

$$Cos(90+D) = Cos(90-\varphi)Cos(90-\delta) + Sin(90-\varphi)Sin(90-\delta)CosH$$
$$-SinD = Sin\varphi Sin\delta + Cos\varphi Cos\delta CosH$$

$$Magrib = Zuhar + \frac{1}{15}Cos^{-1}\left[\frac{-SinD - Sin\varphi Sin\delta}{Cos\delta Cos\varphi}\right] \quad (8)$$

Here *D* id known as the depression angle, which is the horizontal refraction angle from Almanac and its value is 34 minutes plus angular semi diameter of the sun which is 16 minutes. To change it into degree measure $^{50}/_{60} = 0.833^{o}$ so, D = 0.833 degree.

The above equation (8) represents a mathematical model of *Maghrib* prayer.

4.3) Isha Prayer Model

The time of the *Isha* prayer starts at dusk. The point *G* in the following figure 5 is showing the position of sun at isha.

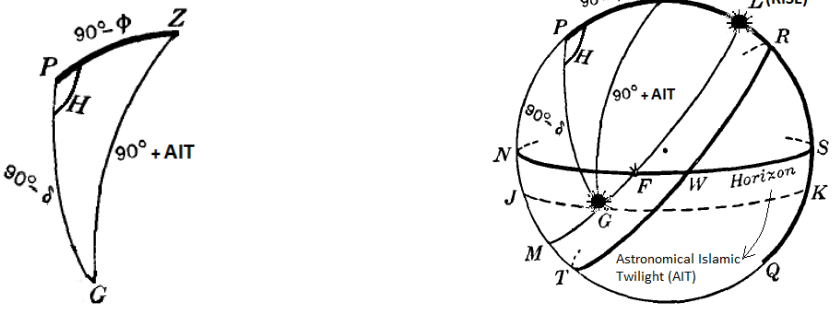


Figure 5

Applying the Cosine Law on spherical triangle PZG

$$Cos(90+AIT) = Cos\ (90-\varphi)Cos\ (90-\delta) + Sin\ (90-\varphi)Sin(90-\delta)CosH$$
$$-Sin(AIT) = Sin\varphi Sin\delta + Cos\varphi Cos\delta CosH$$

$$H = \frac{1}{15}Cos^{-1}\left[\frac{-Sin(AIT) - Sin\varphi Sin\delta}{Cos\delta Cos\varphi}\right]$$

$$Esha = Zuhar + \frac{1}{15}Cos^{-1}\left[\frac{-Sin(AIT) - Sin\varphi Sin\delta}{Cos\delta Cos\varphi}\right] \quad (9)$$

The above equation (9) represents a mathematical model of *Isha* prayer.

4.4) Fajr Paryer Model

The formula for *Fajr* prayer is same as the formula for *Isha* but the only difference in AIT at Dusk. The value of AIT at Dusk and at Dawn is same for many countries. (See Appendix 2)

$$Fajr = Zuhar - \frac{1}{15} Cos^{-1}\left[\frac{-Sin(AIT) - Sin\varphi\, Sin\delta}{Cos\delta\, Cos\varphi}\right] \tag{10}$$

The above equation (10) represents a mathematical model of *Isha* prayer.

4.5) Length of Minimum shadow

In The following figure 6, length AB represent a pole of length L, AC is its shadow at Zawal (Minimum shadow) and $\angle CAB = z$ (Zenith distance of sun)

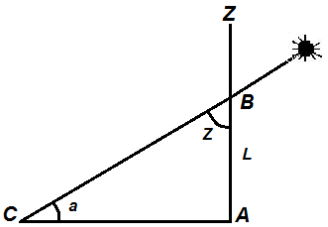


Figure 6

At upper transit $z = |\varphi - \delta|$

Min shadow $= L\tan|\varphi - \delta|$

AB is pole of length L. AC is its shadow length.

At beginning of *Aser Shafi*: $AC = L + \min shadow = L + L\tan|\varphi - \delta|$

At the beginning of *Aser Hanafi*: $AC = 2L + L\tan|\varphi - \delta|$

and angle $\angle BCA = a$ (Altitude of the sun)

for *Aere Shafi* consider ΔABC,

$$\cot a = \frac{AC}{AB} = \frac{L + L\tan|\varphi - \delta|}{L};\ a = \cot^{-1}\left(1 + \tan|\varphi - \delta|\right) \text{ similarly}$$

for *Aare Hanafi*

$$\cot a = \frac{AC}{AB} = \frac{2L + L\tan|\varphi - \delta|}{L};\ a = \cot^{-1}\left(2 + \tan|\varphi - \delta|\right)$$

4.6) Aser Prayer Model

When the sun is at upper transit (Local Noon time) the shadow of vertical pole is minimum.

i) According to “Hazrat Imam Shafai R.A ”, the Aser time begins when the length of the shadow become equal to the length of the minimum shadow plus length of the object.

ii) According to “Hazarat Imam Hanfai R.A”, the time of the Aser begins when the length of the shadow becomes equal to the length of the minimum shadow plus twice the length of the body.

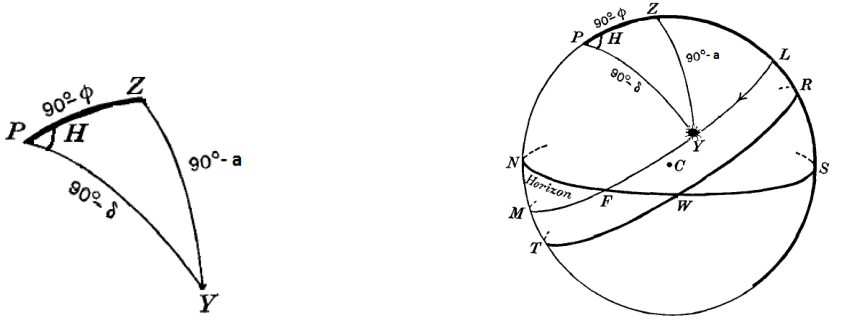


Figure 7

First we derive the formula of length of the minimum shadow.

From Spherical triangle ΔPZY, $PZ = 90^0 - \varphi$ $\quad PY = 90^0 - \delta$

ZY = Zenith distance of sun = $90^0 - a$

By applying Cosine law $\sin a = \sin\varphi \sin\delta + \cos\varphi \cos\delta \cos H$

$$\cos H = \frac{\sin a - \sin\varphi\sin\delta}{\cos\varphi\cos\delta};\ H = \cos^{-1}\left(\frac{\sin a - \sin\varphi\sin\delta}{\cos\varphi\cos\delta}\right)$$

$$Aser = = Z + \frac{1}{15}\cos^{-1}\left(\frac{\sin a - \sin\varphi\sin\delta}{\cos\varphi\cos\delta}\right) \qquad (11)$$

where $a = \cot^{-1}\left(1 + \tan|\varphi - \delta|\right)$ is used for *Aser Shafi* and $a = \cot^{-1}\left(2 + \tan|\varphi - \delta|\right)$ for *Aser Hanfai.*

The above equation (11) represents a mathematical model of *Aser prayer*.

5. COMPARISON OF RESULTS BY OLD & NEW IMPROVED FORMULA

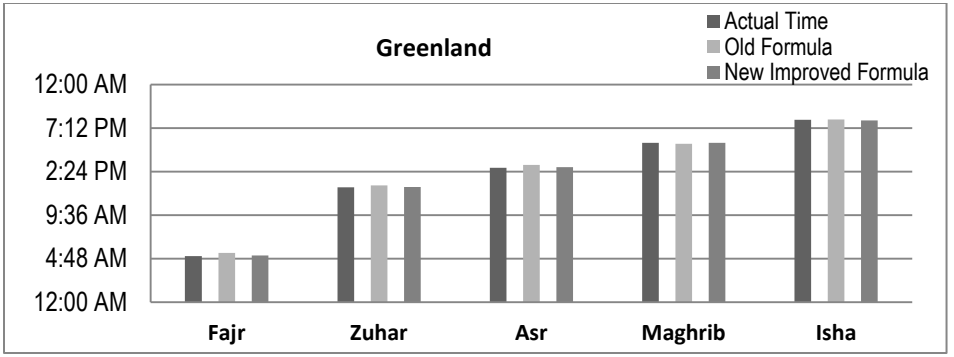
The fundamental formula for finding the Hour angle of sun's setting or rising on celestial sphere is introduced in the book of Spherical astronomy by Smart (1977). The $Cos\, z = \sin\varphi\, Sin\delta + Cos\varphi Cos\delta\, CosH$ formula can be used for finding the hour angle of sun with error as the result shown in Table 1. The improved formulas which derived above are used to find Prayer time with less error. The results of old and new formulas on dated 26th February 2014 with T.E = 8.8' (Equation of time) given in the following table.

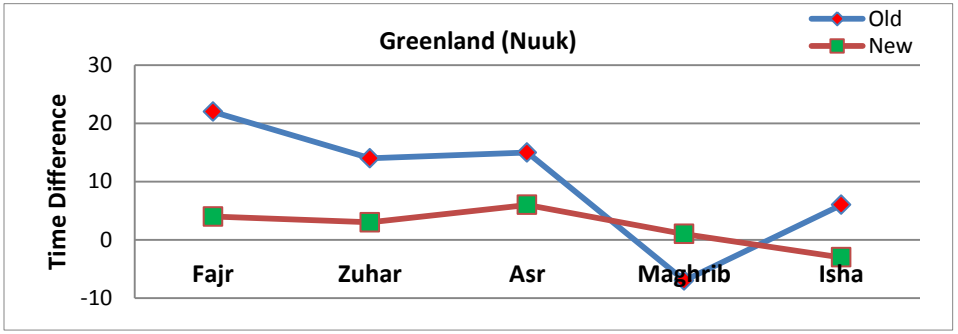
$z = 12 + T_Z - \left(\frac{L}{15} + T_E\right)$ Where L = Longitude of the place and T_z = Time zone.

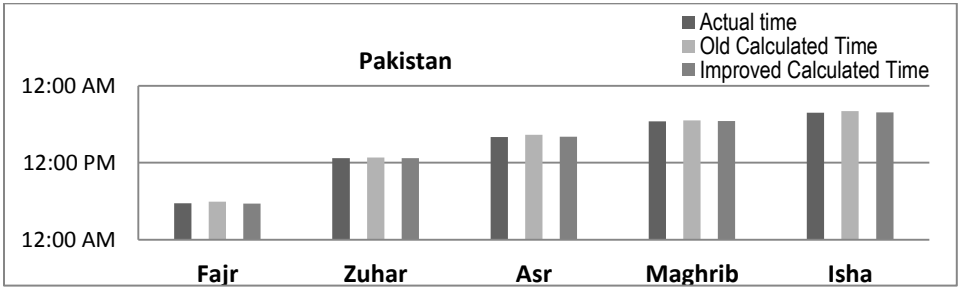
6. DATA ANALYSIS AND RESULTS

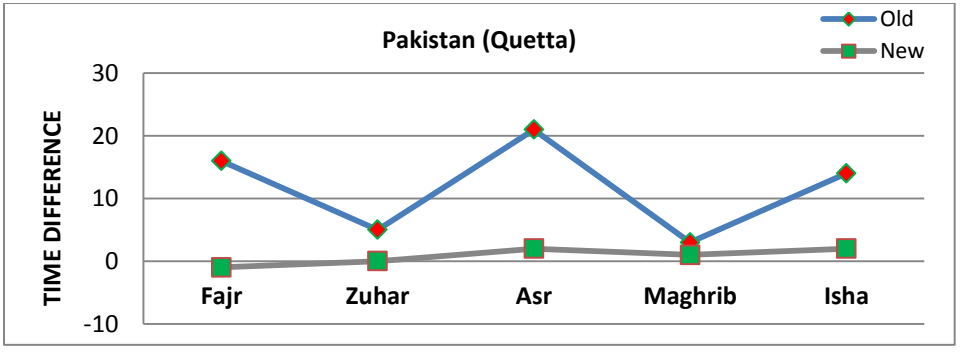
Table 1

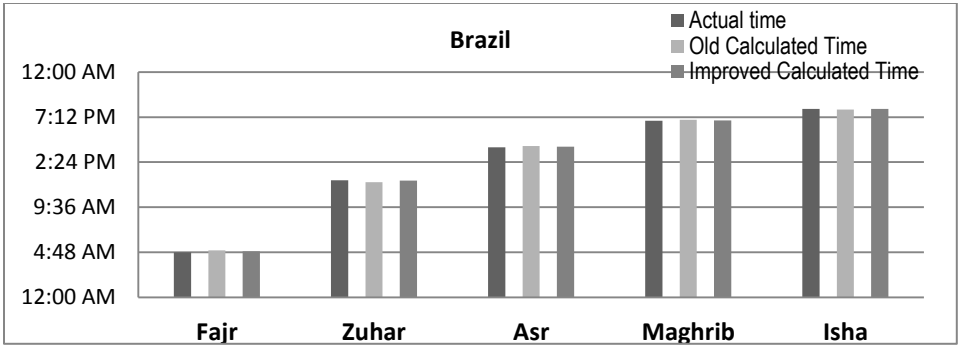
Prayer Name	Actual Time*	Result with Old Formula	Result with New Improved Formula	Time Difference (OLD)	Time Difference (NEW)	Error (OLD)%	Error (NEW) %
Greenland (Nuuk, Sermersooq)							
Fajr	5:05 AM	5:27 AM	5:09 AM	+22 min	+4 min	6.35	1.15
Zuhar	12:39 PM	12:53 PM	12:42 PM	+14 min	+3 min	4.04	0.87
Asr	2:48 PM	3:09 PM	2:54 PM	+15 min	+6 min	4.33	1.73
Maghrib	5:35 PM	5:28 PM	5:34 PM	-7 min	+1 min	2.02	0.29
Isha	8:06 PM	8:09 PM	8:03 PM	+6 min	-3 min	1.73	0.87
Pakistan (Quetta, Balochistan)							
Fajr	5:41 AM	5:57 AM	5:40 AM	+16 min	-1 min	4.61	0.29
Zuhar	12:44 PM	12:49 PM	12:44 PM	+5 min	0 min	1.44	0.00
Asr	4:00 PM	4:21 PM	4:02 PM	+21 min	+2 min	6.06	0.58
Maghrib	6:28 PM	6:35 PM	6:29 PM	+3 min	+1 min	0.87	0.29
Isha	7:48 PM	8:02 PM	7:50 PM	+14 min	+2 min	4.04	0.58
Brazil (Gaspar, Santa Catarina)							
Fajr	4:47 AM	5:01 AM	4:54 AM	+14 min	+7 min	4.04	2.02
Zuhar	12:28 PM	12:16 PM	12:27 PM	+12 min	-1 min	3.46	0.29
Asr	3:59 PM	4:08 PM	4:02 PM	+9 min	+3 min	2.60	0.87
Maghrib	6:49 PM	6:54 PM	6:50 PM	+5 min	+1 min	1.44	0.29
Isha	8:04 PM	8:00 PM	8:04 PM	-4 min	0 min	1.15	0.00
Australia (New Castle , New South Wales)							
Fajr	5:12 AM	5:23 AM	5:13 AM	-11 min	+1 min	3.17	0.29
Zuhar	1:05 PM	1:19 PM	1:10 PM	+14 min	+5 min	4.04	1.44
Asr	4:42 PM	4:31 PM	4:40 PM	-11 min	-2 min	3.17	0.58
Maghrib	7:32 PM	7:25 PM	7:28 PM	+7 min	+4 min	2.02	1.15
Isha	8:53 PM	8:52 PM	8:50 PM	-1 min	-3 min	0.29	0.87
Kuwait (Salwa)							
Fajr	4:55 AM	5:09 AM	4:55 AM	+14 min	+0 min	4.04	0.00
Zuhar	12:00 PM	12:07 PM	12:01 PM	+7 min	+1 min	2.02	0.29
Asr	3:17 PM	3:20 PM	3:18 PM	+3 min	+1 min	0.87	0.29
Maghrib	5:44 PM	5:41 PM	5:42 PM	-2 min	-2 min	0.58	0.58
Isha	7:14 PM	7:15 PM	7:13 PM	+1 min	-1 min	0.29	0.29

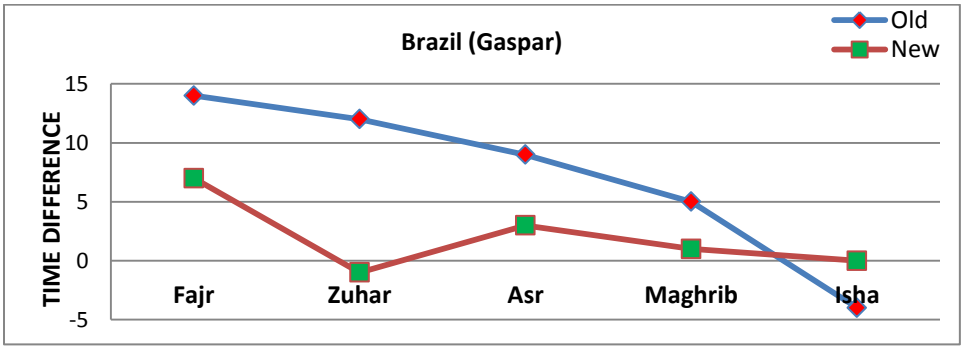

Greenland
Actual Time
Old Formula
New Improved Formula
12:00 AM
7:12 PM
2:24 PM
9:36 AM
4:48 AM
12:00 AM
Fajr
Zuhar
Asr
Maghrib
Isha

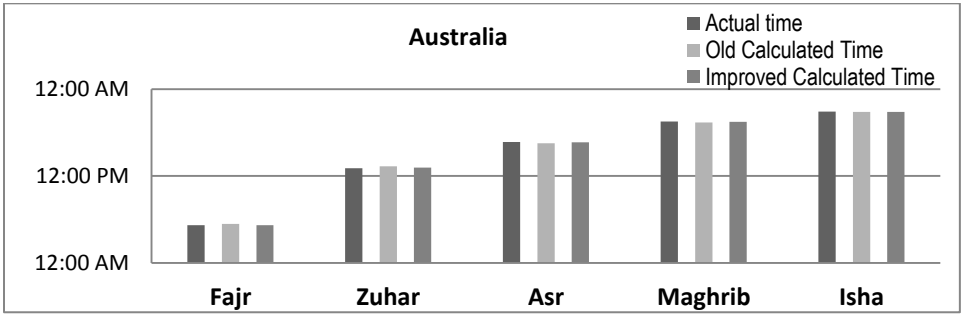

Greenland (Nuuk)
Old
New
Time Difference
30
20
10
0
-10
Fajr
Zuhar
Asr
Maghrib
Isha

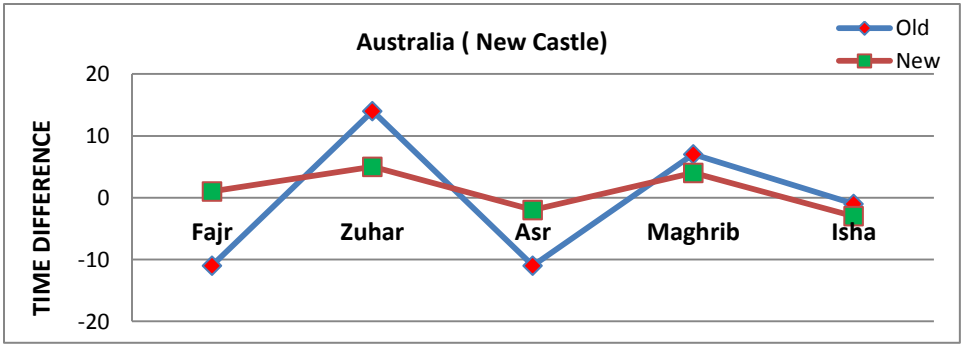

Pakistan
Actual time
Old Calculated Time
Improved Calculated Time
12:00 AM
12:00 PM
12:00 AM
Fajr
Zuhar
Asr
Maghrib
Isha

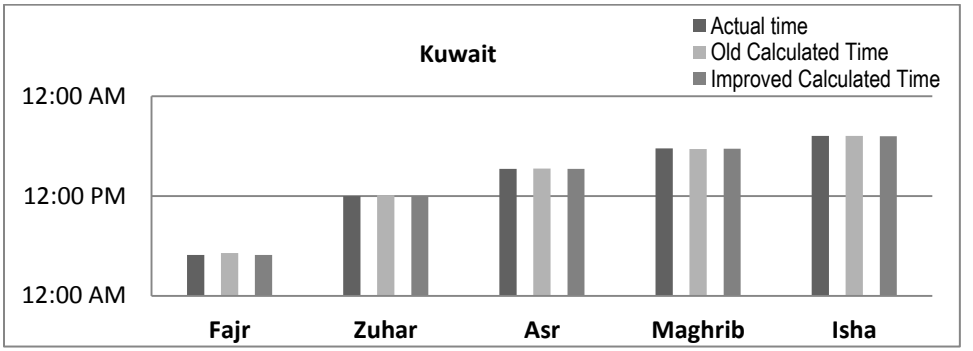

Pakistan (Quetta)
Old
New
TIME DIFFERENCE
30
20
10
0
-10
Fajr
Zuhar
Asr
Maghrib
Isha

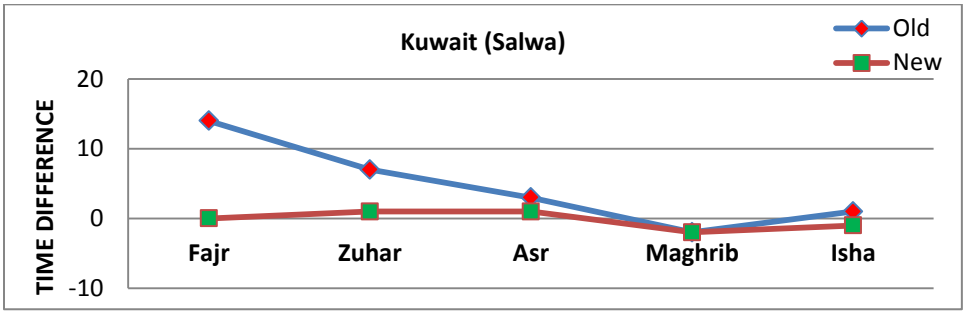

Brazil
Actual time
Old Calculated Time
Improved Calculated Time
12:00 AM
7:12 PM
2:24 PM
9:36 AM
4:48 AM
12:00 AM
Fajr
Zuhar
Asr
Maghrib
Isha

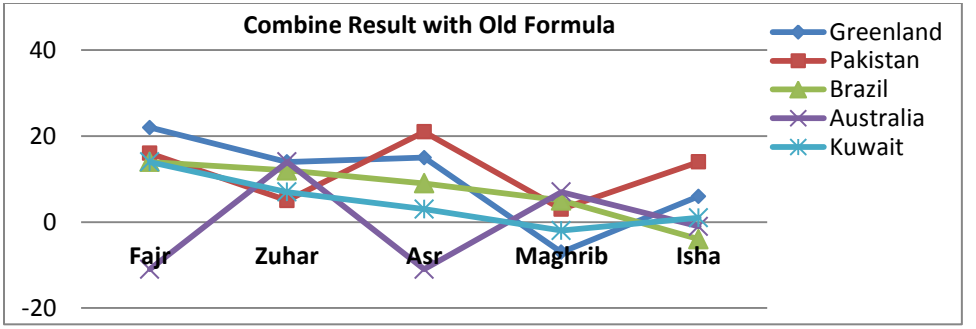

Brazil (Gaspar)
Old
New
TIME DIFFERENCE
15
10
5
0
-5
Fajr
Zuhar
Asr
Maghrib
Isha

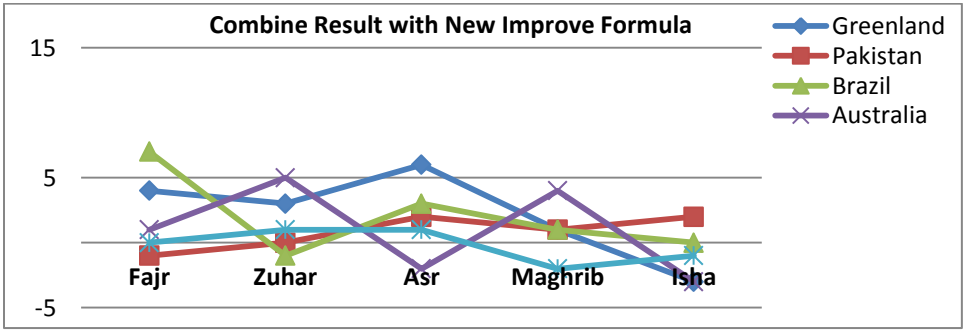

Australia
Actual time
Old Calculated Time
Improved Calculated Time
12:00 AM
12:00 PM
12:00 AM
Fajr
Zuhar
Asr
Maghrib
Isha


Australia (New Castle)
Old
New
TIME DIFFERENCE
20
10
0
-10
-20
Fajr
Zuhar
Asr
Maghrib
Isha


Kuwait
Actual time
Old Calculated Time
Improved Calculated Time
12:00 AM
12:00 PM
12:00 AM
Fajr
Zuhar
Asr
Maghrib
Isha


Kuwait (Salwa)
Old
New
TIME DIFFERENCE
20
10
0
-10
Fajr
Zuhar
Asr
Maghrib
Isha


Combine Result with Old Formula
Greenland
Pakistan
Brazil
Australia
Kuwait
40
20
0
-20
Fajr
Zuhar
Asr
Maghrib
Isha


Combine Result with New Improve Formula
Greenland
Pakistan
Brazil
Australia
15
5
-5
Fajr
Zuhar
Asr
Maghrib
Isha

7. CONCLUSION

The accuracy regarding prayer timing is important though Islam gives some relaxation of about +5 or -5 minutes in actual timing without any issue. The above results with new improved formula shows that almost all timing of selected countries fall within the time interval of -5 to +5, whereas the result with old formula do not lie in this timeframe. So finally, it is concluded that the new improved formula gives best result with high precision. Furthermore many other parameters like altitude of some place can also be included in the improved modelling.

REFERENCES

1. Awang, A., Zakaria, Z., Johar, M.F. and Othman, A.F. (2006). In-Flight Prayer Times and Qiblat Direction - A Preliminary Study, *International RF and Microwave Conference*, Putrajaya, Malaysia, pp 1-5.
2. Roy, A.E. and Clarke, D. (2003). *Astronomy Principal and Practice*, 4th Edition.
3. Bagvi, M.B.A. (1970). Determination of the Direction of Qibla and the Islamic Timings, *Ashraful- Madaris*, Karachi, Pakistan.
4. Cameron, D. (1962). *Elements of Mathematical Astronomy*, London.
5. Goodwin, H.B. (1907). *Plane and Spherical Trigonometry*, Longmans, Green and Co.
6. Ilyas, M. (1984). *A Modern Guide to Astronomical Calculation of Islamic Calendar*, Times & Qibla, Berita Publishing Sdn Bhd., Kuala Lumpur, Malaysia.
7. Gummre, J. (1837). *Elementary Treatise on Astronomy*: England.
8. Lateef, A.B. (1986). *Perpetual Prayer Time Table for the Whole World*, Published by Abdul Majeed Qureshi, Karachi, Pakistan.
9. Moche, D.L. (2009). *Astronomy – A Self-Teaching Guide*, 7th Edition, Wiley California Association for Research in Astronomy.
10. Meeus, J. (1991). *Astronomical Algorithms*, 2nd edition, Willmann -Bell Inc; Richmond.
11. Niweateh, H. (2002). *Al-Falak*, Islamic Academy, Bangkok (Thai).
12. Nassau, J.J. (1984). *Practical Astronomy*, Mc-Graw-Hill Book Company, New York.
13. Qasmi, M.Y. and Muneer, T. (1989). *Prayer Times for United Kingdom & Ireland*. Islamic Research Institute of Britain, Dewsbury, UK.
14. Smart, W.M. (1977). *Textbook on Spherical Astronomy*, 6th edition, Cambridge University Press: London.
15. Simon, N. (1906). *A Compendium of Spherical Astronomy*, New York.
16. Willis, E.J. (1907). *Mathematical Geography*, A Merical Book Company.

APPENDIX: A

Zenith (Z) = Point directly above to observer, on the celestial sphere.

Nadir (N) = Point opposite to Zenith.

Horizon (SWN) = Plane passes through the O and perpendicular to OZ.

Vertical Circle (ZSN) = Great circle (G.C) passing through the Zenith' and Nadir and perpendicular to the horizon.

Altitude (AX) = Angular distance ***a*** of a star above the horizon.

Zenith Distance (ZX) = Distance from the zenith (Z) to the Star, Z.D = 90^{o} - a

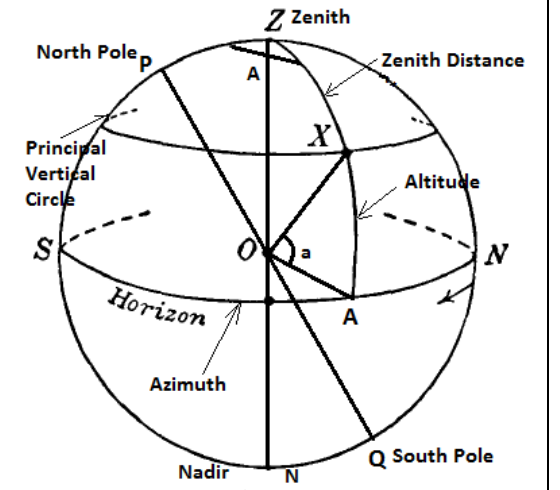


Figure 8

Pole of the Celestial sphere (OP & OQ): Line parallel to the axis about which the earth.

Principal Vertical Circle (SPZ) = Vertical Circle passing through pole *P* or *Q* cut the horizon N or S.

Azimuth (SA) = Angle between vertical plane of star (X) and principal vertical circle.

Colatitude of Observer (PZ) = Angle between radius of earth which passes through the observer's position and earth axis of rotation (PZ = 90 – φ).

Celestial Equator (RWT) = Projection of terrestrial equator in to space.

Transit or Culmination = When the star (X) is at L on the observer meridian. F = Set L = Rise (Transit) Altitude at F = 0^{0} & Z.D = $90^{0.}$ Altitude at L = 90 – φ + sigma.

Observer Meridian (PZRSQ) = Semi great circle to fix completely the position of any star on celestial sphere.

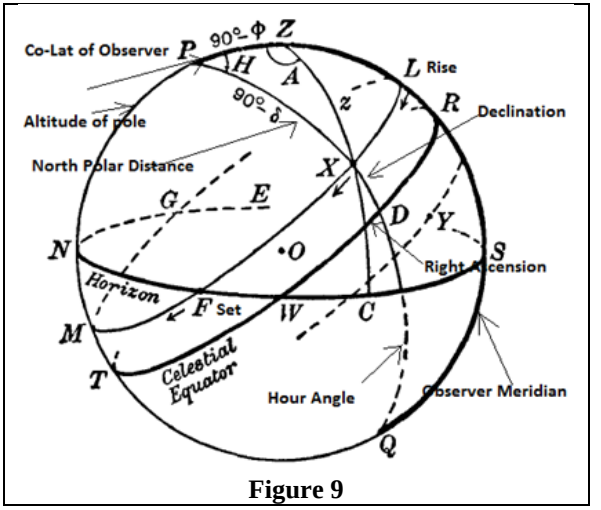


Figure 9

Declination (DX) = A star (X), on small circle arc LXM parallel to Celestial equator then arc DX is called Declination δ . The arc PX = 90 - δ is known as North Polar distance (N.P.D).

Hour angle (XL) = Star's position on the parallel of Declination is specified by the angle at *P* between observer meridian and the Great Circle.

Right Ascension (γD) = Independent of observer's location Coordinate system in which we can determine the position of a star. $\gamma R = \gamma D + RD$

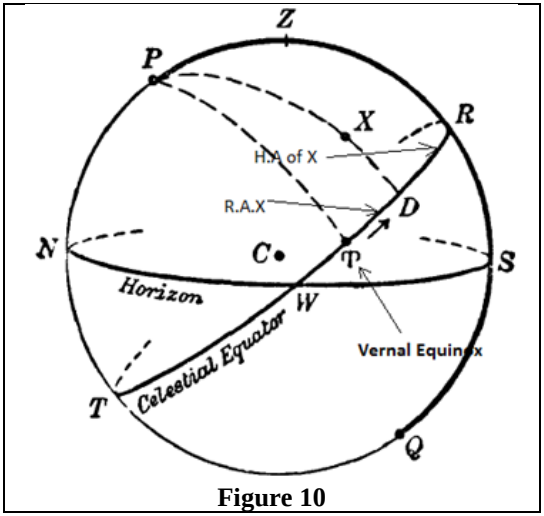


Figure 10

Sidereal Time (S.T) = Hour angle of γ called sidereal time. γ is on observer meridian then S.T is 0^h S.T = H.A.X + R.A.X therefor S.T = H + α

Sidereal Day = interval between two consecutive transit of γ across observer meridian. One sidereal day = $23^h\ 56^m\ 23^S$

Equation of Time = Difference at any moment between R.A of Mean sun and of the true sun. E = R.A.M.S – R.A.O

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 121-128

ON THE IMPACT OF PAIR WORK ON THE SPEAKING SKILLS OF L2 LEARNERS

Nasreen Z. Pasha[§] and **Isabel William**
Department of English, Kinnaird College for Women, Lahore, Pakistan
Email: nasreen.pasha@kinnaird.edu.pk
[§]Corresponding author

ABSTRACT

Language learners and language teachers both need to explore new strategies to enhance language skills. This paper contains an account of an experiment that was conducted in a leading women's college of Lahore to explore the impact of pair work on oral skills as exhibited by an improvement in vocabulary, increase in confidence level, reduction in anxiety, etc. of L 2 learners comprising a mixed ability group at the undergraduate level. The study was conducted on 16 students enrolled in the Compulsory English course in the second semester of the B.Sc. Honors Program. A number of communicative activities were administered on the participants over a span of six sessions, and every activity session was video-recorded. The researcher engaged in focused observation of the participant behavior during the activities as well as during the course of an interactive discussion with the participants immediately after each session. Data was collected from the participants by administering a set of Pre and Post-Activities Questionnaires. Analysis of the collected data is indicative of an increase in the level of confidence of the students regarding the speaking of English with their peers and teachers. A very interesting finding is that, although one would have expected that the students with Urdu-Medium background would benefit from the pair work activities to a much higher extent than those with an English-Medium background, the data did not substantiate this assumption. It appears that pair-work is strategically important in improving not only oral but also written language skills for learners coming from either of the two language backgrounds.

1. INTRODUCTION

The general aim of the set English courses is to develop all the four language skills, that is, listening, speaking, reading, and writing in the students. However, the focus is mainly on writing and reading skills; emphasis on listening and speaking skills is minimal or none at all.

Speech is basically spontaneous i.e., in most situations; people do not plan ahead of time what they are going to say. Both teachers and students think it neither realistic nor necessary to teach or learn speaking. Pattern drills are often arranged in each unit of textbooks, most of which are not authentic but mechanical. In addition, many teachers themselves are not fluent in speaking English, nor do they know how to teach it in a big class.

Furthermore, the tightly structured nature of the activities limits the natural creativity and imagination of the learners, thus, restricting their spontaneity and incentive to take part

in any verbal interaction. The majority of the students tend to remain passive, shy and silent and only a few competent students dominate the class. Some students feel challenged when asked by teachers to answer questions or make comments in English and are pressurized by the judgmental eyes of their peers. In this scenario, pair work and group work activities provide a more conducive atmosphere enabling students to learn and interact freely.

2. LITERATURE REVIEW

In recent years, cooperative learning has emerged as a significant concept and instructional practice in the field of second language education. Attention has been paid to the pedagogical and psychological rationale of its use in second language classrooms (Long & Porter, 1985) and to possible benefits it might have in bilingual programs and second or foreign language settings (Coelho, 1994).

Instructional processes and planning of second language education are also compared with those of cooperative learning methods to see if there are any parallels (McGroarty, 1993). In addition, studies have been conducted to examine, among other things, the impact of cooperative learning on second language acquisition, maintenance of first language, the integration of language and content learning, and second language learners' perceptions.

Hess (2001) opines that collaboration means working together and cooperating, and asserts that students who work together will participate, learn how to compromise, negotiate meaning, become better risk takers and self-evaluators. Working together supports learner autonomy, because students will learn not to count on teacher all the time but rather figure out the unknown expression on their own.

The advantages to pairwork and small-group work over teacher-dominated classes, including a higher possibility of engaging an individual's preferred learning style, a proportional increase in each student's chance to be actively involved in language use, and a greater variety of types of language that students can use, have been extensively discussed in the literature on language teaching methodology (McDonough & Shaw, 1993; Nunan, 1991; Long et al., 1976). Communicative or fluency activities, where the focus of the learners is primarily on the communicative message rather than on displaying control over a specific target structure (Ellis, 1988), and the role output as well as as meaningful input is recognised, are given considerable support from the field of second language acquisition theory (Ellis, 1988; Nunan, 1991; Pica, 1994). Accepting, then, the usefulness of pair work fluency activities in the language classroom, it stands to reason that if students are to reap the benefits of these activities fully, they must take place in the L2. McDonough and Shaw (1993) point out that one disadvantage of pair work is that "monolingual classes readily use their mother tongue instead of the target language, particularly where discussion is animated, (and) interacting in English ... may initially be perceived as artificial."

3. OBJECTIVES OF THE STUDY

The aim of this research is to explore the impact of pair work as an effective teaching strategy on the speaking skills of L 2 learners at the undergraduate level. The objective is to gauge the pedagogical implication of pair work dynamics as a determinant factor of an L 2 teaching and learning environment consisting of a mixed ability group.

4. RESEARCH METHODOLOGY

The research was based on an experiment that was carried out in a women's college of Lahore, Pakistan to investigate the effect of pair and group work on the speaking skills of 16 students enrolled in the compulsory English language course in the second semester of the B.Sc. Honors Program. The students were asked to participate in a series of communicative activities over a period of six sessions of 2 hours duration. Each session was video recorded. The researcher observed the participants behavior during the activities and held interactive discussions following each activity. A pair of semi-structured questionnaires was devised and administered on the *participants,* the *pre-activities questionnaire* prior to the activities and the *post-activities questionnaire* after the completion of the activities for purposes of data collection.

4.1 Research Site and Participant Sample

This study took place at Kinnaird College for Women in Lahore, Pakistan. Participants were enrolled in the second semester of the B.Sc. Honors Program. The class meets twice a week for 60 minutes each class session. The Program offers an English Core course of 3 credit hours during the Fall semester. A sample of sixteen female students was randomly chosen from this class for purposes of data collection. Henceforth, we will refer to them as the participants.

4.2 Summary of Activities

Activities were designed to motivate students to interact in the target language i.e. English. These include *role play activities, picture talk, storytelling, debates, interviews* and *open discussions.* An outline of the various activities is as follows:

- **Role Play:** The class was divided into pairs and was asked to enact the role of parents and how they behave, while the teacher moved around in class listening to their dialogues
- **Debate:** The attention of the students was focused on the recently held debates and focused on argumentative writing. Students were given the topic "uniforms in college" and asked to discuss and note down their ideas in pairs, and a student from each group was asked to present arguments against and in favor of the topic.
- **Open Discussion:** The participants were asked to take part in an open discussion, about how they felt in class that day in comparison to their first day in class.
- **Peer Interview:** The students were put into randomly selected pairs. A set of questions written on slips of paper were given to each pair for their reading, in preparation for the interview activity. Participants were instructed to come to the front of the class and ask each other the questions provided for about two to three minutes.
- **Telephone Conversations:** The students were put into pairs, each student was given a situation to be used as telephone conversation. The list situations comprises (i) Patient seeking appointment from Doctor's assistant, (ii) Customer complaining to Pizza Hut manager, and (iii) Daughter calling up mother about staying late in college.
- **Picture talk:** The students were put into pairs and given a picture each. They were given time to discuss how they would interpret the image in each picture.

- **Retelling Stories:** Students were divided into two groups. A different story was told to each group. As some students began to write down the story, they were advised to only listen. All students were asked to pair up with a member of the other group. They were asked to exchange their stories and retell them in their own words.
- **Short Speeches:** Students were asked to select one of the topics given below and asked to discuss with the student next to her, subsequent to which, each one was asked to speak aloud three to four sentences about their respective topics in front of the class. Some of the topics are as follows: Bangles; Hijab; Eid-ul-Azha; Suicide bombing.

5. DATA ANALYSIS & RESULTS

As indicated above, two semi-structured questionnaires were formulated and administered on the participants, the pre-activities questionnaire prior to the activities undertaken, and the post-activities questionnaire subsequent to the completion of the activities. The salient results of each of the two questionnaires are presented in this section.

5.1 Pre-Activities Questionnaire:

Main findings with reference to the pre-activities questionnaire are summarized below:

- "How do you rate your speaking skills?" In reply to this question, 37.5% of the students have satisfactory speaking skills, 43.75% had average and 18.75% were weak.
- "Do you feel confident while speaking English?" 44% of the students were confident while speaking and 56% were not confident.
- "Do you feel upset or frightened if asked to speak in English in class?" 31.25% of students indicated that they were upset or frightened while speaking in English and 68.75% were not.
- "Do you feel hesitant in speaking before your teachers?" 56.25% of students do not hesitate before teachers while speaking and 43.75% do not hesitate.
- "Do you feel hesitant in speaking before your classmates?" 37.5% students said they do hesitate before classmates while speaking English and 62.5% said they do not.
- With reference to problems faced while speaking:
  i. Shyness: 37.50% students shy while speaking English, 56.25% said they do not shy and 6.25% did not respond.
  ii. Nervousness: 56.2% said Yes.
  iii. Anxiety: 37.50% students said they were anxious while speaking English and 62.50% said they were not.
  iv. Grammar/Vocabulary: 56.25% students were having grammar problem while speaking English, 37.50% said they do not have grammar problem and 6.2% do not respond.
  v. Any other embarrassment? Interestingly, not a single participant identified any other embarrassment.

- "Where you taught how to speak English in school?" 62.50% students said they were taught English in schools and 37.50% said that they were not.
- "Do teachers in Kinnaird encourage you to speak in English?" An overwhelming majority of 93.75% of the students said teachers encourage to speak in English at KC whereas a small minority of 6.25% said no.
- "Do you get ample chance to speak in class?" 68.75% of the students said they do get a chance to speak in class and 31.25% students said they do not.
- "Do language teachers use pair work?" In response to this question, 75% of the respondents said that their teachers do use pair work in class whereas 25 % said "no".
- "Do you feel comfortable working in pairs?" Every single one of the 16 participants replied in the affirmative.

5.2 Evaluation of the Activities:

Presented below is an account of the evaluation of each activity referred to above vis-a-vis its effectiveness in enhancing the oral fluency and the level of confidence in the participants.

Role Play:

This activity generated a lot of spontaneous questions and comments, making the dialogue more interesting. It was noted that although there were a lot of grammatical errors, the expressions were natural and imaginative. The students kept on asking questions that were not written down and were additional to the original ones. At the end of the activity, the teacher asks for students views from children about different parental attitudes towards their children. The students offered their opinions on the strictness of parents and advocated more freedom for children. It proved that subjects close to the learners' real life experiences were able to trigger voluntary discourse in the target language.

Debate:

It was observed that the students were very excited and everyone was anxious to speak up on this particular topic, since the topic was something they could all relate to and felt strongly about. Topics of learners' interests lead them to think creatively and voice out their opinions or views, thus motivating them to speak more. Their note taking skill served as preparatory skill for their oral communication.

Open Discussion:

Upon completion of this activity, the participants admitted that they were more confident now, earlier they felt shy and now they could say anything without hesitation. Asked to give the reason for the change, the participants responded that because they were interacting more with each other, they were actively learning more too. One of the participants went on to say that if the teacher has some knowledge to share with the students, the students should seize that opportunity and gain from the teacher.

Peer Interview:

It was observed that all the participants engaged in this activity with interest and enthusiasm. Some participants read out the questions whereas most of the participants managed to speak extempore. In the dialogue ensuing, many participants put forth questions and comments of their own. It appeared that this activity was instrumental in enhancing the confidence and interest level of the participants. Some participants still had

to refer back to their books to read out the questions. But asking direct questions helped in overcoming their hesitation of speaking in English and speaking in front of the camera for the first time.

Telephone Conversations:

It was observed that the variety of situations was highly motivating for students. Some pairs began to enact the situations for example, the patient asking for appointment started to cough, so this activity carried role-play elements. This showed that peer interaction can lead to a practice of other related communicative skills like drama. Another pair created sound effects of the telephone buzzing, creating an atmosphere of excitement and laughter amongst the others present.

Picture Talk:

It was observed that the picture input served as a source of interest and information for the students to engage in meaningful discussion. Not only did the students benefit in improving their communication skills but also in the development of their socio-cultural awareness and real-life issues. Lot of language was generated. Students presented their opinions on various serious aspects related to their pictures, for example, child-labor and Government's responsibility towards child-labour, women rights etc.

Retelling Stories:

This activity involved the students listening skills. The students' involvement and excitement could be seen from their facial expressions and spontaneous laughter. During the narration, vocabulary items were explained as well. The tonal variations of the narrator were much enjoyed by the students. It seemed that the psychological barriers like anxiety and inhibition had been overcome. The class was highly interactive and noisy as a student searched for words others helped in providing the missing links in the story. The students said that they had immensely enjoyed the humor and surprise endings of these stories.

Short Speeches:

This activity served as a kind of test of individual spoken skills. Setting a time limit and word limit proved to be useful for students and encouraged them to think fast and speak out without worrying about saying too much or too less. Talking about the topic "stars", the student said that life was governed by stars. "I'm a Gemini" The teacher was happy to see all of the students spoke with very little hesitancy, and were able to say something, for even a topic like "Honda" .The bottom line given to the students by the teacher was that the more students talk the better. If the learners wanted to learn the new language they must talk a little more than usual and must also read. Fluency would ultimately lead to accuracy.

5.3 Post-Activities Questionnaire:

- "Did you enjoy the activities?" It is very encouraging to find that all sixteen participants enjoyed the activities.
- "Did you use Urdu?" In response to this question, only 6 of the 16 participants said that they did not use Urdu whereas the remaining 10 confessed to the use of the Urdu language, at least to some extent.
- "How confident do you feel in speaking English now?" 62.5 % of the respondents feel confident to a large extent whereas 37.5 % feel confident to some extent.

- "Has your vocabulary improved through these activities?" Although 3 of the 16 participants said that their vocabulary has not improved, 13 indicated that their vocabulary is better now, at least to some extent.
- "Do you think such activities are needed?" Every single one of the participants replied in the affirmative, 88% indicating that such activities are needed to improve the students' confidence level and speaking skills.
- "Did all activities involve fun?" An overwhelming majority of 94% of the participants responded favorably. See Figure 4.17.
- "Can you speak in English before fellow students now?" In response to this question, 15 of the 16 participants (i.e. a huge majority) indicated that they were able to do so.
- "Can you speak in English before teachers now?" All 16 participants replied in the affirmative.
- "Which skills did you use while taking part in these activities?" In response to this question, 12 participants said "thinking skills" whereas 4 said "speaking skills".

5.4 Some Striking Results:

The study has lead to a number of interesting and important findings, some of which are as under:

- Interesting input drawn from real-world content, such as was incorporated in the pair-work activities, will produce ***spontaneous l***anguage output making even the ***weakest*** student speak in the target language i.e. English.
- Some of the results could be regarded as ***eye-openers.*** For example, we find that, among those participants of the study who had been taught to speak English in school, the proportion of those who felt confident, not nervous or hesitant was not as high as one would have expected.
- It is gratifying to note that the pair-work activities acted as a ***catalyst*** for the utilization of ***thinking skills*** by the learners.
- Contrary to expectations, 67% of the participants who had not been taught English in school rated their English speaking skill as satisfactory whereas ***only 20%*** of the participants who had been taught English in school did so.
- Among the participants who had not been taught to speak English at school, 40% had to face problems pertaining to grammar / vocabulary whereas among the participants who had been taught to speak English in school, ***as many as 70%*** said that they had to face these problems.

6. CONCLUDING REMARKS

The use of pair work activities is justified both on pedagogical and theoretical arguments. Interactive activities can be brought into a classroom at any language level and at any point in the lesson. Opportunities for interaction can make classroom learning more meaningful. Seeing the benefits that learners gain from interactive activities especially in pair work activities, teachers should re-examine their present classroom practices and incorporate more opportunities for interaction. Careful implementation of communicative strategies for teaching speaking skills can positively influence students' learning, making them fluent first and ultimately leading to accuracy.

The study was conducted on college students who had already some proficiency in the English language but their speaking skills showed room for improvement. This situation is representative of students in general. The activities selected were communicative in nature and allowed the students to interact with each other, and gave ample opportunity for language production in the target language i.e. English. Learners normally are used to a teacher-dominated classroom in which they are supposed to do exactly what the teacher directs. In the activities conducted, however, the learner participants were able to use their imagination and creativity in carrying them out. The general procedure was to let the students be the focus in the classroom rather than the teacher.

Evaluation of participant behavior during the pair-work activities as well as analysis of the pre and post-activities questionnaires leads to the conclusion that pair work possesses the potential to act as an effective tool for developing and enhancing the speaking skills of learners. The study supports the research hypothesis, that is, pair work is indeed an effective learning strategy for teaching speaking skills to second language learners.

ACKNOWLEDGMENT

The authors would like to acknowledge the motivation provided by Dr. Nadia Anjum and the assistance provided by Dr. Kalsoom Akhtar in analyzing the collected data. The paper could not have taken the shape that it did without the committed efforts of Dr. Saleha Naghmi Habibullah.

REFERENCES

1. Coelho, E. (1994). *Learning together in the multicultural classroom*. Markham, Ontario: Pippin Publishing Limited.
2. Ellis, R. (1988). *The Role of Practice in Classroom Learning*, in Kasper, G. (ed.), 20-39.
3. Hess, N. (2001). *Teaching Large Multilevel Classes*. Cambridge: Cambridge University Press.
4. Long, M. and Porter, P. (1985). Group work, inter-language talk, and second language acquisition. *Tesol Quarterly*, 19, 207-228.
5. Long, M.H., Adams, L., McLean, M. and Castanos, F. (1976). Doing this with words: Verbal interaction in lockstep and small group classroom situations. In J. Fanselow & R. Crymes (Eds.), On TESOL ’76 (pp. 137-153). Washington, DC: TESOL.
6. McDonough, J. and Shaw, C. (1993). *Materials and Methods in ELT*. Blackwell Publishers.
7. McGroarty, M. (1993). Cooperative learning and second language acquisition. In D.D. Holt (Ed.), *Cooperative learning: A response to linguistic and cultural diversity* (pp. 19-46). *McHenry*, IL: Center for Applied Linguistics and Delta systems.
8. Nunan, D. (1991). *Language Teaching Methodology*. Prentice Hall.
9. Pica, T. (1994). Research on Negotiation: What Does it Reveal about Second Language Learning Conditions, Processes, and Outcomes? *Language Learning*, 44(3), 493-527.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 129-146

USING PCA, POISSON AND NEGATIVE BINOMIAL MODEL TO STUDY THE CLIMATIC FACTOR AND DENGUE FEVER OUTBREAK IN PAKISTAN

Syed Afrozuddin Ahmed[1] and **Junaid Saghir Siddiqi**[2]
Department of Statistics, University of Karachi, Karachi, Pakistan.
Email: safrozuddin@yahoo.com; jsdr123@yahoo.com

ABSTRACT

Various studies have reported that global warming causes unstable climate and many serious impact to physical environment and public health. The increasing incidence of dengue incidence is now a priority health issue and become a health burden of Pakistan. The study aims to understand, explore and compare the climatic factors of Karachi and Lahore that causes the emergence or increasing rate of dengue fever incidence that effects the population and its health. Principal component analysis (PCA) is performed for the purpose of finding if there is/are any general environmental factor/structure which could be considered as Pakistani climate. The correlation coefficient of Karachi monthly and weekly data shows no significant correlation among dengue incidence and climatic variables while Lahore daily and monthly results show significant correlation with maximum and minimum temperatures. We developed an early warning model for the prediction of dengue outbreak in Karachi and Lahore. This has been done by using Poisson regression and Negative binomial regression model. For this purpose we use monthly and weekly data of Karachi and daily and weekly data of Lahore. The negative binomial model with lag (4) for Karachi weekly data for climatic variable is best model and similar to Lahore for lag (28) for daily data. For Karachi weekly relative humidity, maximum temperature and minimum temperature effecting the chance of occurrence positively by 1.71% times, 19.48% times and 11.51% times respectively, while wind speed effect negatively on the occurrence by 7.41% times. For Lahore daily and weekly maximum temperature effect negatively and for the past 28 days it is estimated to negatively influence on the dengue occurrence by 26.1% times. Daily wind speed is effecting negatively by 14.7% times and minimum temperature effect positively for the past 28 days by 86.7% times.

KEY WORDS

Dengue Fever; Global warming; Environmental condition; Principal component analysis.

1. INTRODUCTION

Climate is a composite term and this refers to weather conditions over a period of time. Climate is an average or general conditions of temperature, humidity, atmospheric pressure, wind speed and rainfall of a place. These elements act as climatic factors. Climate is an important subject of scientific enquiry, particularly as it has such an impact on vegetation, soil and health etc. Its importance increases manifold because of its

influence on human life. Natural hazards like storms, floods and desertification are result s of climatic changes. Extreme temperatures, weather hot or cold affect human beings, crops, animals and mosquitoes lifecycles especially Aedes agepty mosquitoes which causes the Dengue Fever [1]. In short, there are many ways in which climate affects our daily lives. The viral disease which is called dengue transmitted to human body through the mosquito named Aedes agepty [2]. In the medical terminology it's another name is "break bone fever" and the headache, pain in joints and muscle, headache, rashes in skin close to the measles are its symptoms. Sometime this disease is threatening to the life as results in blood disorder causes the lack of platelets and shortage of plasma and due to these problems dengue hemorrhagic shock syndrome may be observed also patients may go in to problem of low blood pressure. No vaccines are still commercially available for this disease; however prevention is the way through which we may reduce the mosquito habitat and exposure to bites. This viral disease has spread in most part of the world [3].About two million people are affected by dengue fever in the world [3]. "Den-1, Den-2, Den-3 and Den-4" are used the names of serotypes which causes severe dengue [4]. The dengue infectious disease spreads in warm climate [5]. Aedes Agepty mosquitoes are seriously sensitive by environmental conditions. The climatic variables i.e., precipitation, humidity, temperature and wind speed are helping for the survival and reproduction of dengue mosquitoes. High temperature helps to decrease the time lag for the replicate of the dengue mosquito virus. This process is called "Extrinsic Incubation Period", in this period virus is transferred from mosquito to human body through the mosquito's salivary glands. In the high temperature mosquito become infectious faster and has higher probability too dangerous for human being before completing its life cycle. So we may say that the critical factor is climate by which the situation may be controlled to the mosquito habitat. We can see vast amount of information from the researchers Afrozuddin et al. (2013)[6] Hii YL et al. (2009), Johansson et al. (2009) Su GL (2008), [7][8][9] have studied and reported that there is a positive correlation between the amount of rainfall and relative humidity with dengue occurrence. Similar results have also reported by other studied groups Barbazan et al. (2010), G.Chowell et al. (2003) [10],[11].

In Karachi, Pakistan first Dengue fever outbreak reported in 1994 [12] 1 patient out of 145 died. Many researchers have studied and reported epidemics of dengue and its causes from different parts of Pakistan [13][14][15]. Pakistan has experience a number of dengue fever outbreaks since 1994. In October 2005, Dengue again hit Karachi after 10 years and 21 deaths out of total 103 confirmed cases were recorded. Since then, the disease has become widely accepted as one of the major public health problems. Over 21,204 people were reportedly infected in the country in 2010. The reasons why and how the dengue epidemics become endemic diseases in Pakistan, different reasons are for this problems like poor hygiene and inefficient sewerage system as results these are the ideal habitat for the dengue vector to lay eggs and flourish. From the last few years transmission of dengue virus situation is high in the country, Karachi and Lahore are the cities which are heavily threatened by dengue epidemics in pre and post monsoon periods. In the periods of flood this adverse situation may also be observed. The internally displaced peoples who migrated from the arm conflicted areas and they were living without shelter and proper healthy environment suffered with the dengue and other viral diseases.

2. STATISTICAL MODEL

A. Generalized Linear Model

The generalized linear model is used in the situations where an independent sequence of random variables is taken and a function of their means is linked to a linear predictor based on known covariates[16][17]. This allows regression techniques to be used for non-normal (but exponential family) distributions on a non-linear scale. These techniques may conveniently be used more general and some function of μ_i is assumed to be a linear combination of β's [18][19]. The generalized linear model formulation is the usual method of fitting factors and variables to data which involve an exponential family error structure [20]. Thus, the GLM allows the extension of regression techniques to non-normal sampling models with effects on a suitable non-linear scale.

Use of Poisson and negative binomial regression models depends on the nature of the distribution of the dependents variables [21]. Researchers prefer to use Negative binomial regression because of the assumption of Poisson models which are hard to observe with the data, some researchers also prefer to use both Poisson and negative binomial model within the same study [22]. The model parameters give the same information in both Poisson and Negative binomial regression model. Poisson regression model are commonly used for count data [23]. In our study the "events" nature of severe and non-severe dengue cases. We have used the both Poisson and negative binomial regression models in our study where we have use the count data of dengue occurrence. Poisson and negative binomial regression models are different in regards to their assumptions of the mean and variance of the dependents variables, in the case of Poison model the assumption is the mean and variance of the distributions are equal i.e. equi dispersion whereas the negative binomial regression models we do not use the assumptions of equal mean and variance and is particularly correct for using this model when the over dispersion is in the data. Which shows that mean is less than variance.

Some researchers have noted that count data for occurrence of diseases often have over dispersion so they have preferred to use negative binomial regression in those types of studies [24]. We have also preferred to use negative binomial regression models on our data because it also shows the over dispersion.

B. Principal Component Analysis

In this study our focused is to exploring the climatic variables and/or the climatic factors (structure) which may influence in the increase or decrease in the number of dengue fever cases in Karachi and Lahore. To find the climatic structure we used Principle component analysis method. PCA technique is recognized as reduction and extraction for dimensionality of the data and rating as much of the variation present in the original data set. This is one of the way of identifying patterns in the data [25]. It is difficult to find patterns in the data for this purposed we use PCA which highlight the similarities and difference in the data. The pattern in the data can be found by squeeze the data in other word by suppressing the dimensionality of the data by avoiding the losses of information. This procedure is also use in the image compression or image reorganization. The purpose to apply the PCA is to reduce the manifest variables, in this way the set of components will be reduced [26]. The new components are called PC1,

PC2, PC3 and so on, (for the first, second and third principal components) are independent and decrease the amount of variance from the original data set. PC1 (the first component) captures most of the variance, PC2 captures the second most of the variance and so on until all the variance is accounted for, in this way very few will be retain for the further study[27]. The components have been treated as climatic factors or climatic structures. The Principle component method transforms correlate observed variables in to uncorrelated variables which are linear combination of observed data. The required condition to apply this technique is correlation/covariance can be defined. So we don't find any problem in the application of principle component analysis.

3. DATA AND METHODOLOGY

This study was conducted in Karachi and Lahore capital of Sindh and Punjab province, these two cities are the largest and most populous metropolitan cities of Pakistan. The climate of Karachi features an arid climate, albeit a moderate version of this climate. Karachi is located on the coast and as a result has a relatively mild climate. Karachi has two main seasons; Summer & Winter, while spring and autumn are very short [28]. The features of Lahore are clearly based on five season semi-arid climate with five seasons. The climate data of Karachi and Lahore were collected from Pakistan Meteorologist Department, on monthly basis from 1980 -2012 and daily basis from 2010-2012. Names of five climatic variables which we have studied are precipitation (P), Maximum temperature (Mx), Minimum temperature (Mn), Humidity (H) and Wind speed (W). The dengue cases in Karachi from 2010 to 2012 are reported on weekly basis while dengue cases in Lahore are reported in daily basis from 2011 to 2012. The data on the number of deaths due to Dengue fever are not available with us; therefore our study will only focus on the dengue occurrence cases only. In 2010 Govt. of Sindh and in 2011 Govt. of Punjab, Pakistan has established the Dengue surveillance cell for not only keeping the records of Dengue Fever reported cases, but also providing the information about the DF and the health facilities to the people in this regard. The dependent and explanatory variables along with the corresponding codes which are used to express these variables in equations, tables etc. and unit of measurements are listed in the Table 1. In Table 2 we presented number of dengue fever cases by months, for the years 2010, 2011 and 2012.

Table 1:
Climate variables, Dengue Fever their labels and their unit of measurement

S.No.	Variables	Code	Unit of Measurement
1	Precipitation	P	mm
2	Relative humidity	H	%
3	Maximum temperature	Mx	°C
4	Minimum temperatrure	Mn	°C
5	Wind speed	W	Knots
6	Dengue Fever	DF	

Table 2
No. of Dengue cases reported in Karachi (2010-2012) and Lahore (2011-2012)

Year		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Karachi	2010	11	3	14	7	7	12	25	224	128	2510	1902	406	5249
	2011	18	5	8	4	17	8	32	40	393	415	104	35	1079
	2012	11	14	8	12	16	28	39	40	6	141	262	102	734
Lahore	2011	0	0	4	5	15	3	6	1124	8254	5805	1531	55	16802
	2012	0	0	0	0	5	1	0	6	174	58	15	1	260

Source: Dengue surveillance cell Lahore, Govt. of Punjab, Pakistan

Graphs shown in Fig. 1 to Fig. 4 also reveal this fact, in these graphs high peak is during September October 2011, while this is negligible for the year 2012. The graph for daily, weekly and monthly data are similar, spikes can be seen in the daily graph but, it is completely smooth in the monthly data graph. The continuous efforts of the government have reduced the dengue outbreak which can be observed from the year 2012 in the given Table 2.

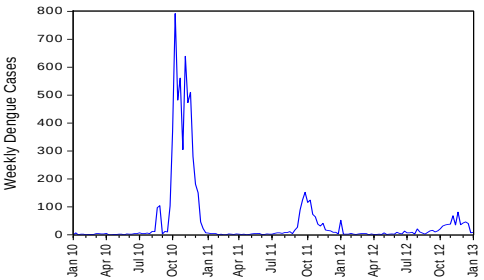


Fig. 1: Karachi Weekly Dengue Cases 2010-2012

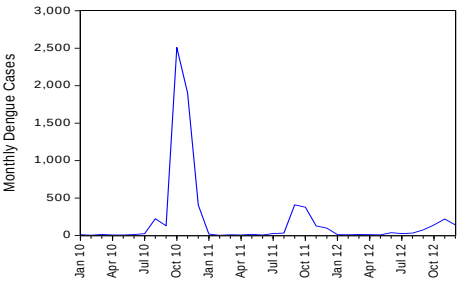


Fig. 2: Karachi Monthly Dengue Cases 2010-2012

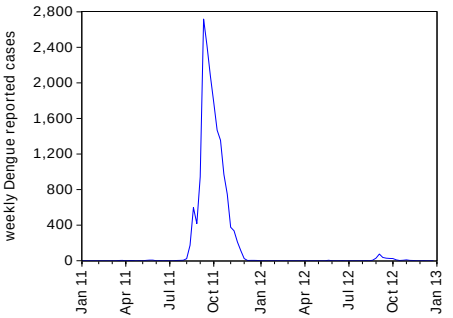


Fig. 3: Lahore Weekly Dengue Cases (2011-2012)

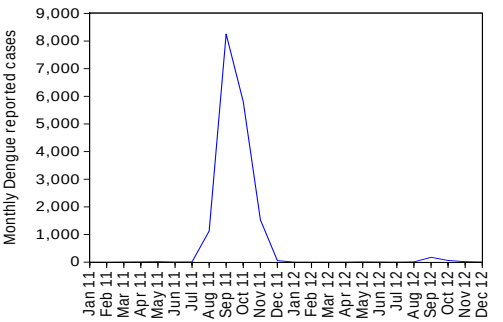


Fig. 4: Lahore Monthly Dengue Cases (2011-2012)

4. RESULTS AND DISCUSSION

A. PCA to explore climatic structure

For the purpose of statistical studies, we have applied PCA and other statistical techniques also fitting different statistical models to find the relations between DF and climatic variables. Hii et al. (2009), Jury (2008) and Su (2008) have also discussed the relation between climatic variable and dengue incidence.

For the purpose of statistical analysis we have taken the Karachi weekly and monthly data in dengue period 2010-2012 and Lahore daily, weekly and monthly data from 2011-2012. The reason why we have considered the data of these periods is that dengue data is available for these periods only.

We have applied Principal Component Analysis technique for daily weekly and monthly data of 2010-2012 and 2011-2012 to explore if there is any difference in the structure of climate in these time period. Sadiq et al. (2010) Jacquelynne et al. (1999), Benzl et al. (1997), Ehrendorfer (1987), Gibson et al. (1984), Molteni et al. (1983) and Jolliffe et al. (1972) have also applied Principal Component Analysis technique for the characterization of climatic variables.

For explanation and understanding purposes we present all four PCs for considered period in Table 3. These tables contains, in columns; the linear combination (PCs) in approximate and simplified form, percentage of variation explained by respective Principle components and abbreviation of labels of the PC.

According to the elbow rule, all plots show that two Factor model (PC) is sufficient to explain the variation in the data. But for the comments and interpretation we consider four PCs: as the percentage explained by 3rd and 4th PC in some case are quite large. The small variation in weather may affect the general atmospheric condition can affect life. Table 3 is prepared for the monthly and weekly basis for the period 2010-2012.

The first four PCs for the period 2010-2012 of Karachi weekly and monthly data are almost the same. For weekly dengue period data the percentage of variance of PC1 and PC2 are highly different to each other, they are around 44% and 22% which shows that there are only 66% of total variation factor of Karachi climate is namely PC1 and PC2 variation in temperature with humidity and wind speed and contradiction of wetness and temperature.

The third component in the variation of Karachi climate is “reverse of variation of wetness and wind speed “which constitutes 18% of variation. Whereas the fourth component is the contradiction of “temperature” and “wind speed “which is 11.2% of variation. Explanation and interpretation of PCs one by one, for all data sets, shown in Table 3.

PC1s of Karachi is being constructed as the linear combination of Humidity, Maximum temperature, Minimum temperature and Wind speed, in other words it is the weighted average of these four climatic variables and all the four variables are related to ***General Atmospheric Condition (GAC)***. For the reader to understand that how we have interpreted and labeled the PCs, let us take the example of first PC of the monthly data of year 2010-2012. It is (PC1) weighted average; [PC1= 0.9(H+Mx)+Mn+W] of the

Humidity (H), Maximum temperature (Mx), Minimum temperature (Mn) and Wind speed (W). Thus it may interpret that major source of variation in the data occur is due to "General Atmospheric Condition". Then we label it or name it as ***General Atmospheric Condition*** GAC**.** This PC (GAC) contributes/explain (44%) of total variation varies from 44 to 65.6%. The PC1s for Lahore daily data of 2011-2012 are constructed as linear combination of same climatic variables, that are **Mx,Mn** and **W**. Therefore this component can be interpreted or labeled as "*Windy and hot* **"(HtW)** ". While in PC1 of weekly period of 2011-2012, one more climatic variable **P** has been added and these can be interpreted as " *Windy, hot* and rainy" (**WHtR)**. PC1 of monthly period of 2011-2012 is made of temperatures and **P** this can be interpreted as "*hot* and rainy" (**HtR)**. Temperatures, Mx and **Mn** Contributes highly in PC1 of all data. The percentage contribution of all PC1s varies from 43% to 45% which are not as high as in case of Karachi data. One can generally conclude that major climatic factor is variation in temperature together with high wind.

PC2 for the Karachi Dengue period is the ***contrast between precipitation and wind speed*** and the contribution of PC2s varies from 21 to 22 percentages. As second PC, this contribution may consider as large contribution. Thus we may conclude that this factor play an important role in the formation of weather of city Karachi.

PC2 of Lahore daily period of 2011-2012 are linear combination of **H** and **P** thus, this factor can be named as "*Wetness*" **(Wt)**. But monthly data of 2011-2012, which is for Dengue period, has one more variable **W** with negative sign so we may label it as "*Wetness with low wind*"**(WtLW)**. The percentage contribution of all PC2s varies from 27% to 41%. Therefore we can say that the second major component of Lahore is the rainy with high humid weather.

Table 3:
PCA for Karachi weekly and monthly Climatic variables: (2010–2012)

PCs	2010 -2012 *(Weekly Dengue Period)*			2010 -2012 *(Monthly Dengue Period)*		
	Linear combination	***Label***	***%***	***Linear combination***	***Label***	***%***
PC1	0.6H+0.7Mx+0.8Mn+0.8W	**GAC**	44	0.9H+0.9Mx+Mn+0.7W	**GAC**	65.6
PC2	0.5P+0.7H-0.4Mx-0.4Mn	**C[PH][T]**	22	0.9P-0.5W	**CPW**	21.8
PC3	0.7P-0.4H+W	**C[P][WH]**	18	0.5W-0.5Mx	**CWT**	10
PC4	0.5Mx-0.4W	**CTW**	11	0.2P+0.1Mx-0.3H	**C[PT][H]**	2.4

Table 4:
PCA for Lahore Daily, weekly and Monthly Climatic variables: (2011–2012)

PCs	2011-2012 (Daily Dengue period)			2011-2012 (Weekly Dengue period)			2011 -2012 (Monthly Dengue period)		
	Linear combination	Label	%	Linear combination	Label	%	Linear combination	Label	%
PC1	0.9(Mx+Mn)+0.5W	**HtW**	43	.9(Mx+Mn)+.7W+.5P	**WHtR**	50	0.9(Mx+Mn)+.7P	**HtR**	45
PC2	0.75(P+H)	**Wt**	27	0.9H+0.8P	**Wt**	31	0.9H+0.5P-0.8W	**WtLW**	41
PC3	0.7W-0.5P	**WD**	15	0.6W-0.3(Mx+Mn)	**WLT**	12	0.6W+0.4P	**WWt**	11
PC4	0.5(H-P)	**HnR**	14	0.3H-0.4P	**HnR**	6	0.2(H-P)	**HnR**	2

where:

C[PT][H] is Contrast between the combination of Precipitation and Temperature with Humidity.
C[PH][T] is Contrast between the combination of Precipitation & Humidity with Temperature.
C[P][WH] is Contrast between of Precipitation and the combination with Wind speed & Humidity.
GAC is General Atmospheric condition.
CTP is Contrast between Temperature & Precipitation
CTW is Contrast between Temperature & Wind speed.
CWH is Contrast between Wind speed & Relative Humidity.
CPW is Contrast between Precipitation & Wind speed**.**
CWT is Contrast between Wind speed & Temperature.
WWt is windy and wet.
HnR is humid but no rain.
WtLW is Wetness with low wind.
WD is wind and dry.
HtR is hot and rainy.
WHtR is windy hot and rainy.
Wt is wetness.
HtW is windy and hot

PC3 for the Karachi weekly based data can be seen as difference between precipitation (P) and weighted mean of humidity and wind speed. This factor shows that during weeks of Dengue period the weather condition is different from non-Dengue period, and it contributed 18% in the variation of Karachi climate.

PC3 for Lahore weekly period 2011-2012 is the contrast between Wind and average temperatures, we can label these factors "windy with low temperature "(WLT). PC3 of monthly and daily period of 2011-2012 are found to be contrast between W and P that we may label as "windy and dry "(WD).

PC4 for Karachi Dengue (Monthly) period is constructed as ***contrast of humidity against weighted average of precipitation and maximum temperature,*** which contributes only 2.4 of the total variation. PC4 for the weekly data is the ***contrast between Mx and W*** and it explaining 11.2% of variation.

PC4 for Lahore of all three sets of data are contradiction between H and P that indicate in Lahore there is a climatic factor which is "humid but no rain" (HnR). The percentage of contribution varies from 2% to 14%. It is interesting to see that for daily period of 2011-2012 the contribution is quite high i.e. 14% and for monthly data of 2011-2012 is only 2%

B. Modeling DF with Climatic variables with different lags:

Karachi weekly data

We have applied the negative binomial and Poisson regression models on Karachi weekly and Lahore daily climatic variables and climatic factors (principal component score) for the Dengue period 2010-2012 and 2011-2012. In both negative binomial and Poisson model the dependent variable is DF cases.

Poisson model is found to be not suitable because of very large (high) over dispersion [23] and their AIC's values are very large varying from 45 to 48, so we will not consider it. We found negative binomial models much better than Poisson models. The AIC values in case of negative binomial models are very low as compare to AIC value of Poisson models, the AIC values from 7.44 to 7.51. The values of AIC decreases as lag period increases. We have shown the estimation of Poisson parameter for the purpose of comparison.

The following expressions are established from negative binomial models from the Tables 5 and 6. In these equation we have only consider only those variables whose coefficient are significant at least $\alpha = 0.05$.

Log (DF) = 0.018H + 0.118Mx -0.115W
Lag (-1) [7.513] KNBCV(-1)
Log (DF) = 0.022H + 0.139Mx – 0.136 W
Lag (-2) [7.464] KNBCV(-2)
Log (DF) = -4.183 + 0.021H + 0.163Mx + 0.107Mn -0.123W
Lag (-3) [7.441] KNBCV(-3)
Log (DF) =- 5.1882 + 0.017H + 0.178Mx + 0.109Mn – 0.077W
Lag (-4) [7.443] KNBCV(-4)

log(DF) = - 5.1882 + 0.017(H) + 0.178(Mx) + 0.109 (Mn) -0.077(W)
DF = (e-5.1882) (e0.017H) (e0.178Mx) (e0.109Mn) (e-0.077W)

We consider negative binomial climatic variables of Karachi [KNBCV (-3)] model as the best predictive model for number of DF cases because its AIC value is the smallest (AIC=7.441) among the above 4 models. But the AIC value at lag three and lag four are almost equal as four weeks lag is equivalent to one month and the variables involved in [KNBCV (-3)] & [KNBCV (-4)] are same. Therefore we have going to consider [KNBCV (-4)]. In all the four lag model relative humidity, maximum temperature and wind speed shows significant effect on the occurrence of 'DF' while wind speed 'W' has significant effect on prevention of 'DF'.

The main effecting explanatory variables are humidity, maximum temperature, minimum temperature and wind speed, where humidity, maximum temperature and minimum temperature effecting positively on the occurrence of dengue fever cases, while

the wind speed seems to be causing in the reduction of dengue fever cases. On the basis of the model expressed above following are some comments.

Weekly relative humidity is effecting the chance of occurrence positively and in the past 4 weeks (28 days) humidity value is estimated to positively influence on the chances of dengue occurrence by 1.71% times.

Weekly maximum temperature is effecting the chance of occurrence positively and in the past 4 weeks (28 days) maximum temperature value is estimated to positively influence on the chances dengue occurrence by 19.48% times.

Minimum temperature is effecting the chance of occurrence positively on the weekly DF occurrence. For the past 4 weeks the minimum temperature affects positively on the chances of dengue occurrence by 11.51% times.

Wind speed is effecting negatively on the weekly occurrence of dengue fever by 7.41%.

Lahore weekly data

The same models negative binomial and Poisson regression models are applied to Lahore weekly climatic data of the Dengue period 2011-2012. The following equations are established from negative binomial models from the Tables 6 and 7.

Log (DF) = 22.993- 0.164H -1.056Mx + 1.223Mn - 0.610W
Lag (-1) [6.346]
Log (DF) = 19.030 -0.146H-0.983Mx + 1.232Mn – 0.545 W
Lag (-2) [6.260]
Log (DF) =14.766 -0.948Mx + 1.252Mn – 0.521W
Lag (-3) [6.198]
Log (DF) = 14.136 -0.116H -0.849Mx+1.149Mn -0.442W
Lag (-4) [6.214]

The climatic variables for the Dengue period 2011-2012 are used in negative binomial and Poisson regression model with different lags. Again the Poisson model is not suitable because of large (high) over dispersion also their AIC's values are very large so we have avoided to consider the Poisson model.

$\log(DF)$ =14.76 -0.948(Mx) +1.252(Mn) -0.521(W)
$DF = (e^{14.1766})\,(e^{-0.948Mx})\,(e^{1.25Mn})\,(e^{-0.521W})$

We have shown the estimates of Poisson model parameter in the Table 6 and 7 for the purpose of comparison while the negative binomial models are suitable for all lags where the AIC values are very small (see table 6 and 7)

However the best model is negative binomial model with climatic variables with three weeks i.e. 21 days lag with AIC value 6.198. This model says that maximum temperature and Wind speed negatively effect on dengue occurrence whereas minimum temperature cause positively on occurrence of dengue (see table 6 & 7). This model and the model LNBCV (-28) (for daily data) give the same information, i.e. Mx and W make negative effect while Mn effects positively.

The models with lag 1, 2 and 4 contain variables H, Mx, Mn and W while the model with smallest value of AIC (6.198) missed H.

The main effecting explanatory variable are maximum temperature, minimum temperature and wind speed, where minimum temperature effecting positively on the occurrence of dengue fever cases, while the remaining two variables seems to be causing in the reduction of dengue fever cases,

1) Weekly maximum temperature is effecting negatively and in the past 4 weeks (28 days) maximum temperature value is estimated to negatively influence on the dengue occurrence by 61.2%.
2) Weekly wind speed is effecting negatively and in the past 4 weeks (28 days) wind speed value is estimated to negatively influence on the dengue occurrence by 40.6%.
3) Minimum temperature is effecting very highly on the daily DF occurrence. For the past 28 days the minimum temperature affects positively on the dengue occurrence by as high as 249.06%.

C. Model with Climatic Factors with different lags:

Karachi Weekly data

We have applied the Negative Binomial and Poisson regression models on Karachi weekly climatic factors for the Dengue period 2010-2012 complete results (output) are given in Tables (4 & 5). Below we write the models negative binomial considering those variables whose coefficients are significant with at least $\alpha = 0.05$. Also apart from lag differences, AIC values are given. These expressions are obtained from table 4 and 5. Principle components analyses are applied and components are extracted and available in Table 3 and the interpretations are made in earlier section. But in the expressions we have used the notation in term of PCs for the reason of simplicity.

Log (DF) = 3.421 + 0.81 PC4
Lag (-1) [8.388] KNBPC (-1)
Log (DF) = 3.369 + 0.9262 PC4
Lag (-2) [8.366] KNBPC (-2)
Log (DF) = 3.257 + 0.452PC1+1.003 PC4
Lag (-3) [8.310] KNBPC (-3)
Log (DF) = 3.228 + 0.792 PC1 + 0.927PC4
Lag (-4) [8.308] KNBPC (-4)

$$\log(DF) = -3.228 + 0.792(PC1) + 0.927(PC4)$$
$$DF = (e^{-3.228})\,(e^{0.792PC1})\,(e^{0.927PC4})$$

The main effecting explanatory climatic factors are PC1 and PC4. PC1 is interpreted *as general atmospheric condition* **(GAC)** and it effects positively on the occurrence of dengue fever cases. Whereas PC4 which we have labeled as *contrast between temperature and wind speed* **(CTW)** also seems to be causing in the increases of dengue fever cases. On the basis of expressions the following comments are made:

1) PC1 is effecting the chance of occurrence positively and in the past 4 weeks (28 days) PC1 value is estimated to positively influence on the chances of dengue occurrence by 120.78% times.
2) PC4 is effecting the chance of occurrence positively and in the past 4 weeks (28 days) PC4 value is estimated to positively influence on the chances dengue occurrence by 152.72% times.

Again the Poisson model is not suitable because of large (high) over dispersion [70] also their AIC's values are very large that is why have avoided to consider the Poisson model. For the comparison we have shown the estimation of Poisson parameter, while the negative binomial models are found to be suitable for all lags where the AIC values are very small.

We have used the first four PCs to model, DF data, using Poisson and negative binomials. The advantage of applying PC is that they are perfectly uncorrelated and thus there is no question of collinearity (among the explanatory variable) in the model [72]. The other advantage in this case is instead of original variable environmental factors variables are considered [73].

In this case of negative binomial models, all four models have shown quite good behavior, however for the model KNBPC (-4) is the best with smallest value of AIC. All four models PC4 is the best predictor. We interpret the model KNBPC (-4). This model says that PC1 and PC4 are having significant effect on DF. As PC1 represents ***general atmospheric condition***, it means that variation in general atmospheric condition effects significantly (P-value 0.000) the occurrence of DF. The PC4 is interpreted as ***contrast between maximum temperature and wind speed*** has also significant effect (P-value=0.000) on the occurrence of DF, that is, high temperature with low wind speed is suitable for the spread of DF.

Lahore Weekly data

After the application of negative binomial and Poisson regression models using weekly climatic factors from Lahore on the Dengue period 2011-2012. The following equations are established from negative binomial models from the Tables 6 and 7. The climatic factors for the Dengue period 2011-2012 are used in negative binomial and Poisson regression model with different lags. The following equations are performed form Negative binomial Model with different lags using PC'S as explanatory variables from Table 6 &7. Again we have only included in these equations those variables whose coefficients are significant at $\alpha = 0.05$.

Log (DF) = 3.680 + 1.460 PC1 – 1.574PC3
Lag (-1) [6.519]
Log (DF) = 3.259 + 2.019PC1 –1.626 PC3 +0.920 PC4
Lag (-2) [6.474]
Log (DF) = 2.835 + 2.110 PC1 -1.837 PC3 + 1.452 PC4
Lag (-3) [6.399]
Log (DF) = 2.5492+2.287PC1+1.133PC2-1.614PC3+0.898PC4
Lag (-4) [6.382]

Again the Poisson model is not found suitable because of large (high) over dispersion also their AIC values are very large so we have avoided considering the Poisson model. For the comparison we have shown the estimation of Poisson parameter, while the negative binomial model is suitable for all lags where the AIC values are very small.

As we have stated earlier that the advantage of applying PC is that they perfectly uncorrelated and thus there is no question of collinearity (among the explanatory variable) in the model. The other advantage in this case is instead original variable environmental factors variables are used.

The interpretation of PC can be found from the Table 4. In this case of negative binomial models, all four models are shown quite good behavior, but for the model LNBPC (-4) is the best with smaller value of AIC. We interpret the model LNBPC (-4). This model says that PC1 i.e. general atmospheric condition. PC2 is the combine effect of precipitation and humidity that explains 31% of variation whereas the PC4 which is the contradiction of precipitation and humidity that explains only 6% of variation having positive effect on the dengue outbreak however the PC3 which is the contrast of wind speed with temperatures having negatively affected on dengue occurrence i.e. prevention of dengue outbreak.

Note:

We feel that there is no need to NB/Poisson models for monthly data because when applied these model to weekly data, 4-week lag shows the best model. Therefore model for monthly data may not show better results.

Table 5:
Negative binomial and Poisson Regression Models for Karachi Weekly Climatic Variables and PCs score with Lags (1=1 Week, 2=2 Week)

		Lag 1=1 week				Lag 2=2 week			
C.V	*PCA*	*Poisson*		*Negative Binomial*		*Poisson*		*Negative binomial*	
		KPCV(-1)	KPPC(-1)	KNBCV(-1)	KNBPC(-1)	KPCV(-2)	KPPC(-2)	KNBCV(-2)	KNBPC(-2)
		C.V	PCA	C.V	PCA	C.V	PCA	C.V	PCA
Constant	Constant	-0.673822*	3.289924*	-1.095004	3.421568*	-1.283149*	3.307232*	-2.610924	3.369246*
P-value	P-value	(0.0004)	(0.0000)	(0.4957)	(0.0000)	(0.0000)	(0.0000)	(0.0944)	(0.0000)
Rainfall	PC1	-0.000960*	0.073784*	0.004614	-0.006991	-0.002222*	0.246229*	0.006360	0.195029
P-value	P-value	(0.2009)	(0.0000)	(0.6073)	(0.9655)	(0.0025)	(0.0000)	(0.4753)	(0.2218)
Humidity	PC2	0.025574*	0.208700*	0.018027*	0.121896	0.028595*	0.270955*	0.022406*	0.090866
P-value	P-value	(0.0000)	(0.0000)	(0.0048)	(0.4056)	(0.0000)	(0.0000)	(0.0004)	(0.5380)
Max-Temp	PC3	0.092952*	-0.093486*	0.118338*	-0.062715	0.083167*	-0.031850*	0.139797*	0.080677
P-value	P-value	(0.0000)	(0.0000)	(0.0470)	(0.6283)	(0.0000)	(0.0083)	(0.0166)	(0.5620)
Min-Temp	PC4	0.063665*	1.103704*	0.042956	0.809898*	0.104455*	1.068109*	0.079271	0.926157*
P-value	P-value	(0.0000)	(0.0000)	(0.3055)	(0.0000)	(0.0000)	(0.0000)	(0.0501)	(0.0000)
Windspeed		-0.144869*		-0.114828*		-0.36728*		-0.135944*	
P-value		(0.0000)		(0.0048)		(0.0000)		(0.0004)	
DF		150	151	151	154	148	150	151	153
AIC		48.59230	104.1927	7.512931	8.388649	46.94487	107.4916	7.463632	8.366574

Table 6:
Negative Binomial and Poisson Regression Models on Karachi Weekly Climatic Variables and PCs score with Lags (3=3 Week, 4=4 Week).

		Lag 3=3 week				Lag 4=4 week			
C.V	*PCA*	*Poisson*		*Negative Binomial*		*Poisson*		*Negative binomial*	
		KPCV(-3)	KPPC(-3)	KNBCV(-3)	KNBPC(-3)	KPCV(-4)	KPPC(-4)	KNBCV(-4)	KNBPC(-4)
		C.V	PCA	C.V	PCA	C.V	PCA	C.V	PCA
Constant	Constant	2.121113*	3.195835*	-4.183118*	3.256533*	-2.843588*	3.246479*	-5.188223*	3.227487*
P-value	P-value	(0.0000)	(0.0000)	(0.0050)	(0.0000)	(0.0000)	(0.0000)	(0.0003)	(0.0000)
Rainfall	PC1	-0.003240*	0.455423*	0.007874	0.451676*	-0.004586*	0.611525*	0.008183	0.791778*
P-value	P-value	(0.0000)	(0.0000)	(0.3774)	(0.0028)	(0.0000)	(0.0000)	(0.3378)	(0.0000)
Humidity	PC2	0.030179*	0.259762*	0.020994*	0.036624	0.029054*	0.333114*	0.017063*	0.074926
P-value	P-value	(0.0000)	(0.0000)	(0.0005)	(0.8076)	(0.0000)	(0.0000)	(0.0044)	(0.6332)
Max-Temp	PC3	0.077762*	-0.009479	0.163378*	0.212483	0.073605*	-0.064236*	0.177967*	0.289465
P-value	P-value	(0.0000)	(0.4267)	(0.0034)	(0.1682)	(0.0000)	(0.0000)	(0.0011)	(0.0967)
Min-Temp	PC4	0.144178*	1.36116*	0.107020*	1.003453*	0.169875*	1.022751*	0.109589*	0.926595*
P-value	P-value	(0.0000)	(0.0000)	(0.0089)	(0.0000)	(0.0000)	(0.0000)	(0.0064)	(0.0000)
Windspeed		-0.159716*		-0.123031*		-0.137838*		-0.077030*	
P-value		(0.0000)		(0.0012)		(0.0000)		(0.0432)	
DF		147	150	148	151	148	148	148	150
AIC		45.43362	101.0056	7.441136	8.310884	45.58021	107.5101	7.443659	8.308137

Table 7:
Dengue, Climate Variable and PCA with different lags of Lahore Weekly (2011-2012)

		Lag 1=1 week				Lag 2=2 week			
C.V	*PCA*	*Poisson*		*Negative Binomial*		*Poisson*		*Negative binomial*	
		LPCV(-1)	LPPC(-1)	LNBCV(-1)	LNBPC(-1)	LPCV(-2)	LPPC(-2)	LNBCV(-2)	LNBPC(-2)
		C.V	PCA	C.V	PCA	C.V	PCA	C.V	PCA
Constant	Constant	25.6303*	4.208996*	22.99307*	3.679522*	24.11909*	4.030225*	19.02968*	3.259218*
P-value	P-value	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)
Rainfall	PC1	0.151803*	0.971068*	0.075160	1.459905*	0.125212*	1.203691*	0.104581	2.019065*
P-value	P-value	(0.0000)	(0.0000)	(0.5516)	(0.0189)	(0.0000)	(0.0000)	(0.3552)	(0.0005)
Humidity	PC2	-0.197634*	0.218382*	-0.164177*	0.637811	-0.186571*	0.335563*	-0.145725*	0.649315
P-value	P-value	(0.0000)	(0.0000)	(0.0074)	(0.3168)	(0.0000)	(0.0000)	(0.0046)	(0.22850)
Maxtemp	PC3	-1.153103*	-1.073113*	-1.056105*	-1.573685*	-1.183044*	-1.029282*	-0.982974*	-1.625876*
P-value	P-value	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Mintemp	PC4	1.302340*	-0.023035*	1.223284*	0.980951	1.360018*	0.036117*	1.231869*	0.920636*
P-value	P-value	(0.0000)	(0.0016)	(0.0000)	(0.0435)	(0.0000)	(0.0000)	(0.0000)	(0.0337)
Windspeed		-0.607758*		-0.610050*		-0.485962*		-0.545110*	
P-value		(0.0000)		(0.0000)		(0.0000)		(0.0000)	
D.F		98	99	99	101		98	98	99
AIC		222.90	464.47	6.346	6.519		435.52	6.260	6.474

Table 8:
Dengue, Climate Variable and PCA with different lags of Lahore Weekly (2011-2012)

		Lag 3=3 week				Lag 4=4 week			
C.V	*PCA*	*Poisson*		*Negative Binomial*		*Poisson*		*Negative binomial*	
		LPCV(-3)	LPPC(-3)	LNBCV(-3)	LNBPC(-3)	LPCV(-4)	LPPC(-4)	LNBCV(-4)	LNBPC(-4)
		C.V	PCA	C.V	PCA	C.V	PCA	C.V	PCA
Constant	Constant	18.01813*	3.767791*	14.76573*	2.834607*	16.05917*	3.423748*	14.13624*	2.592102*
P-value	P-value	(0.0000)	(0.0000)	(0.0059)	(0.0000)	(0.0000)	(0.0000)	(0.0065)	(0.0000)
Rainfall	PC1	0.039401*	1.486695*	-0.030708	2.109989*	0.166028*	1.668099*	0.163236	2.286586*
P-value	P-value	(0.0000)	(0.0000)	(0.7771)	(0.0002)	(0.0000)	(0.0000)	(0.1065)	(0.0000)
Humidity	PC2	-0.130130*	0.483574*	-0.080481	1.277328*	-0.118620*	0.597242*	-0.116017*	1.132829*
P-value	P-value	(0.0000)	(0.0000)	(0.1502)	(0.0119)	(0.0000)	(0.0000)	(0.0273)	(0.0019)
Maxtemp	PC3	-1.007152*	-0.953186*	-0.947992*	-1.837171*	-0.861272*	-1.005102*	-0.848551*	-1.613902*
P-value	P-value	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)
Mintemp	PC4	1.218032*	0.205335*	1.251960*	1.452142*	1.043985*	0.059440*	1.148540*	0.898104*
P-value	P-value	(0.0000)	(0.0000)	(0.0000)	(0.0006)	(0.0000)	(0.0000)	(0.0000)	(0.0053)
Windspeed		-0.274910*		-0.521240*				-0.441874*	
P-value		(0.0000)		(0.0000)				(0.0001)	
D.F		96	97	98	97	97	96	96	96
AIC		235.87	403.45	6.198	6.399	208.42	325.14	6.214	6.382

5. CONCLUSIONS AND RECOMMENDATIONS

The environmental conditions of country are studied using PC and negative binomial models. Karachi and Lahore are the largest cities of Pakistan there environmental conditional is changing. Karachi is getting warmer and Lahore is getting cooler. Negative binomial model for Karachi and Lahore data with one month lag give a good fit to DF cases. This is also true for environmental manifest data and environmental factors (PCs). Therefore we may conclude that the cases of dengue fever are effected by the change of month. In this study relative humidity and minimum temperatures at 28 days or 4 weeks lag (approximate to one month) identified as significant predictor for the dengue incidence in Karachi and Lahore. Thus relative humidity and minimum temperature have some impact on the occurrence of dengue fever. If the minimum temperature recorded increases from previous day, the number of dengue cases is expected to increases in next 28 days. It is suggested that surveillance teams should keep eyes on the changes in relative humidity and minimum temperature and if the minimum temperature for several take serious remedial measures. Since in months of August and September large number of dengue cases are expected so remedial measures should be taken in the months of June and July. The data we have is related to the number of dengue cases reported with respect to time, but we suggest further studies the number of death due to dengue locality from

where the patient are coming, measure of awareness of the people are also needed to be studied.

We also suggest some Bio-statistical studies of the patient and the cause of spread of diseases.

Effect of increase of population of the country on the spread of several diseases is also requires attention by the researcher and the governments. We find that over population of the cities not only creating social problems but also creates health problems. It is suggested that government should not only control the population and control cities from having over crowded population by making new cities where people can get job and can have health facilities and daily necessities. We also suggest creating dengue surveillance cells not only in provincial's capitals also at district levels whose function should be awareness, medical and fumigation on regular basis.

6. REFERENCES

1. Afrozuddin S.A. (1993). *Random Effects in Generalized Linear Models*. Unpublished M.Phil. Thesis, Department of Mathematical Statistics and Operational Research, University of Exeter, Exeter, England.
2. Afrozuddin, S.A. (1995). Normaility Assumptions in Generalized Linear Models. *Second International Seminar on Islamic Economics*, Organized by Department of Statistics, University of Peshawar, Pakistan. 127-142.
3. Afrozuddin, S.A. and Junaid, S.S. (2014). *Principal Component Analysis to explore climatic variability and Dengue outbreak in Karachi*. Unpublished paper.
4. Barbazan, P., Guiserix, M., Boonyuan, W., Tuntaprasart, W., Pontier, D., Gonzalez, J.P. (2010). Modelling the effect of temperature on transmission of dengue. *Med. Vet. Entomol*., 24, 66-73.
5. Bock, R.D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179-197.
6. Bradley, E.L. (1973). The equivalence of maximum likelihood and weighted least squares in the exponential family. *J. Amer. Statist. Assoc*., 68, 199-200.
7. Chowell, G. and Stinchez, F. (2006). Climate-Based Descriptive Models of Dengue Fever: The 2002 Epidemic in Colima, Mexico. *Journal of Environmental Health*, 41-44.
8. David, L. (2000). *Linear algebra and its applications*. Addison-Wesley. New York.
9. Dobson, A.J. and Barnett, A.G. (2008). *An Introduction to Generalized Linear Models. London*: Chapman and Hall.
10. Farrar, J., Focks, D., Gubler, D., Barrera, R., Guzman, M.G., Simmons, C., et al. (2007). Towards a global dengue research agenda. *Trop. Med. Int. Health.*, 12, 695-9.
11. Gubler, D.J. and Kuno, G. (1997). *Dengue and dengue hemorrhagic fever New York*, NY: CAB International.
12. Hakim, S.T., Saleem, M. and Nadeem, S.G. (2011). An Experiencewith Dengue in Pakistan: An Expanding Problem. *Ibnosina J. Med.*, 3(1), 3-8.
13. Hardin, J.W. and Hilbe, J.M. (2012). *Generalized Linear Models and Extensions*. A Stata Press Publication. Texas, USA.

14. Hii, Y.L., Rocklo¨v, J., Ng, N., Tang, C.S., Pang, F.Y. et al. (2009). Climate variability and increase in incidence and magnitude of dengue incidence in Singapore. *Global Health Action* 2.
15. Humayoun, M.A., Waseem, T., Jawa, A.A., Hashmi, M.S. and Akram, J. (2010). Multiple dengue serotypes and high frequency of dengue hemorrhagic fever at two tertiary care hospitals in Lahore during the 2008 dengue virus outbreak in Punjab, Pakistan. *Int. J. Infect Dis.*, 14S3, e54-e59.
16. Jahan, F. (2011). Dengue Fever (DF) in Pakistan. *Asia Pac. Fam. Med.*, 10(1), 1.
17. Johansson, M.A., Dominici, F. and Glass, G.E. (2009). Local and Global Effects of Climate on Dengue Transmission in Puerto Rico, *PLoS Negl Trop Dis.*, 3(2), *World Academy of Science, Engineering and Technology*, 38, 2010 908.
18. Knudsen, D.C. (1992). Generalizing Poisson regression: Including a priori information using the methods of offsets. *The Professional Geographer*, 44, 202-208.
19. Lovett, A. and Flowerdew, R. (1989). Analysis of count data using Poisson regression. *The Professional Geographer*, 41, 190-198.
20. MacDonald, J. and Lattimore, P. (2010). Count models in criminology. In Piquero A. and Weisburd, D. (eds.) *Handbook of Quantitative Criminology*. Springer.
21. Riaz, M.M., Mumtaz, K., Khan, M.S., Patel, J., Tariq, M., Hilal, H.. Siddiqui, S.A. and Shezad, F. (2009). Out break of Dengue Fever in Karachi 2006: A Clinical Perspective. *J. Pak. Med. Assoc.*, 59(6), 9-4.
22. Shlens, J. (2005). *A tutorial on Principal Component Analysis*, 1-13.
23. Su, G.L. (2008). Correlation of climatic factors and dengue incidence in Metro Manila, Philippines. *Ambio*, 37, 292-4.
24. Will, Todd (1999). *Introduction to the singular value Decomposition*, Davidson College. *www. Davidson.edu/academic/math/will/svd/index.html*
25. World Health Organization. (2009). *Dengue: Guidelines for diagnosis, treatment, prevention and control*. New edition. France.
26. www.cdc.gov/dengue/entomologyEcology/climate.html.
27. www.karachi.com/v/geography
28. www.news-medical.net/health/Dengue-Fever.aspx

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 147-154

ON MODIFICATION OF ESTIMATOR OF THE PERCENTILE COEFFICIENT OF KURTOSIS IN THE CASE OF DISTRIBUTIONS SELF-INVERSE AT UNITY

Saleha Naghmi Habibullah[1§] and **Syeda Shan-e-Fatima**[2]
[1] Department of Statistics, Kinnaird College For Women
Lahore, Pakistan. Email: salehahabibullah@gmail.com
[2] Department of Statistics, Government College University
Lahore, Pakistan
[§] Corresponding Author

ABSTRACT

A random variable is said to be 'Self-Inverse at beta' when the probability distribution of X/beta is identical to that of beta/X. The case beta=1 yields a distribution for which the $(1\text{-}q)^{th}$ quantile is the reciprocal of the q^{th} quantile and, as a result, the median of which lies at unity. Habibullah and Saunders (2011) and Fatima et al. (2013) propose self-inversion-based modifications to well-known estimators of the ***cumulative distribution function*** and the ***cumulative hazard function*** respectively. Fatima and Habibullah (2013a) take up the estimation of measures of ***central tendency*** of reciprocal-invariant distributions, and propose self-inversion-based modifications of L-estimators including the ***mean, mid-hinge*** and ***mid-range***. By performing simulation studies based on well-known distributions, they demonstrate that gains in efficiency can be achieved by adopting the proposed modified formulae. Fatima and Habibullah (2013b) take up the estimation of measures of ***dispersion***, and present self-inversion-based modifications to L-estimators of dispersion including the ***range, inter-decile range*** and ***inter-quartile range***. Fatima and Habibullah (2013c) propose a modification to the formula of the ***Bowley's coefficient of skewness*** in the case of distributions self-inverse at unity, and show that the modified coefficient is a more efficient estimator of the population skewness as measured by the Bowley's coefficient. In this paper, we propose a self-inversion-based modification to the formula of the well-known ***Percentile Coefficient of Kurtosis*** and, by carrying out simulation studies based on 255 samples of size 10 each drawn from the Birnbaum Saunders and lognormal distributions self-inverse at unity, show that the proposed modification yields ***gains in efficiency*** which, obviously, has important implications for accurate modelling.

1. INTRODUCTION

Habibullah and Saunders (2011) have introduced the term "Self-inverse" by saying that a random variable X will be said to be 'Self-Inverse at β' when the probability distribution of β / X is identical to that of X / β. The case $\beta = 1$ yields "self-inversion at unity" i.e. the situation when the $(1\text{-}q)]^{th}$ quantile of the distribution of the random variable X is the reciprocal of the q^{th} quantile and, as a result of which, the median lies at unity. Such distributions can also be regarded as being invariant under the reciprocal

transformation. For details regarding such distributions, see Seshadri (1965), Saunders (1974) and Habibullah et al. (2010).

Habibullah and Saunders (2011) utilize the property of self-inversion to modify the formula of the well-known estimator of the ***cumulative distribution function*** and show that the modified formula is more efficient than the original formula when sampling from a distribution self-inverse at unity; Fatima et al. (2013) achieve something similar for the ***cumulative hazard function***.

Fatima and Habibullah (2013a) take up the estimation of measures of ***central tendency*** of reciprocal-invariant distributions, and propose self-inversion-based modifications of ***L-estimators*** and the ***cumulative hazard*** function respectively. Fatima and Habibullah (2013b) take up the estimation of measures of ***dispersion***, and present self-inversion-based modifications to L-estimators of dispersion including the ***range, inter-decile range*** and ***inter-quartile range***. Fatima and Habibullah (2013*c*) focus on the ***asymmetry*** of distributions defined on the positive half-line, and propose a self-inversion-based modification of the formula of the ***Bowley's coefficient of skewness.*** By performing simulation studies based on well-known distributions, Fatima and Habibullah (2013*a,b & c*) demonstrate that, by adopting the proposed modified formulae, ***gains in efficiency*** can be achieved in the estimation of the corresponding population parameters when sampling from a distribution self-inverse at unity.

In this paper, we focus on the property of peakedness of a distribution self-inverse at unity and propose a modification to the formula of the well-known ***percentile coefficient of kurtosis***. By carrying out simulation studies based on 1000 samples of size 10 each drawn from Birnbaum Saunders and lognormal distributions, we demonstrate the proposed modification yields ***gains in efficiency.***

2. PERCENTILE COEFFICIENT OF KURTOSIS AND ITS MODIFICATION IN THE CASE OF SELF-INVERSION AT UNITY

The well-known Percentile Coefficient of Kurtosis is given by

$$K = \frac{Q_3 - Q_1}{2\left(P_{90} - P_{10}\right)} \tag{2.1}$$

and it is well-known that $0<K<0.5$.

Utilizing the property of self-inversion at unity, we propose the following modification to formula (2.1):

$$K_{SIU} = \frac{\left[\left(Q_3 - \frac{1}{Q_3}\right) - \left(Q_1 - \frac{1}{Q_1}\right)\right]}{2\left[\left(P_{90} - \frac{1}{P_{90}}\right) - \left(P_{10} - \frac{1}{P_{10}}\right)\right]} \tag{2.2}$$

3. SIMULATION STUDY

In this section, we present the results of a simulation study that has been carried out in order to demonstrate that the modified estimator provide ***gains in efficiency***. Self-inverse at unity versions of the lognormal and Birnbaum Saunders distributions have been considered for this purpose.

3.1 Sampling from the Lognormal Distribution

In this section, we present a comparison of ranges, variances and coefficients of variation of the sampling distributions of the percentile coefficient of kurtosis and the self-inversion-based modified estimator. Subsections 3.1.1 to 3.1.3 present the results of a simulation study based on 1000 samples of size 10 each drawn from the lognormal distribution with $\mu = 0,\ \sigma^2 = 1$.

3.1.1 Comparison of Ranges

Table 3.1 presents the maximum and minimum values as well as the ***ranges*** of the sampling distributions of percentile coefficient of kurtosis and the self-inversion-based modified estimator based on 1000 samples of size 10 each drawn from a lognormal distribution with $\mu = 0, \sigma^2 = 1$. It is seen that the range of the sampling distribution of the newly proposed estimator is less than the range of the sampling distribution of the well-known percentile coefficient of kurtosis.

Table 3.1
Minimum and Maximum values and Ranges of the sampling distributions of the Percentile Coefficient of Kurtosis and Modified Percentile Coefficient of Kurtosis when sampling from the lognormal distribution $\mu = 0, \sigma^2 = 1$.

Well-Known Estimator			**Newly Proposed Estimator**		
Minimum	Maximum	**Range**	Minimum	Maximum	**Range**
0.0201	0.4429	0.4228	0.029038	0.44384	0.414802

3.1.2 Comparison of Variances

Table 3.2 presents a comparison of the variances of the sampling distribution of the well-known Percentile Coefficient of Kurtosis and Modified Percentile Coefficient of Kurtosis based on the 1000 samples of size 10 each from the lognormal distribution with $\mu = 0, \sigma^2 = 1$.

Table 3.2
Variances of the sampling distributions of the Percentile Coefficient of Kurtosis and Modified Percentile Coefficient of Kurtosis when sampling from the lognormal distribution with $\mu = 0, \sigma^2 = 1$.

Variance of Well-Known Estimator	**Variance of Newly Proposed Estimator**
0.0089	0.00599

We see that the variance of the sampling distribution of the modified estimator is smaller than the variance of the sampling distribution of the well known Percentile Coefficient of Kurtosis.

3.1.3 Comparison of Coefficients of Variation

Table 3.3 presents the means of the sampling distributions of the well-known Percentile Coefficient of Kurtosis and Modified Percentile Coefficient of Kurtosis obtained by drawing 1000 samples of size 10 each from the lognormal distribution with $\mu = 0, \sigma^2 = 1$.

Table 3.3
Means of the sampling distribution of the Percentile Coefficient of Kurtosis and Modified Percentile Coefficient of Kurtosis when sampling from the lognormal distribution $\mu = 0, \sigma^2 = 1$.

Mean of Well-Known Estimator	Mean of Newly Proposed Estimator
0.2115	0.200451

Utilizing the values given in Table 3.3, we obtain the coefficients of variation of the sampling distribution of the well-known Percentile Coefficient of Kurtosis and the corresponding modified estimator based on the 1000 samples of size 10 each from the lognormal distribution with $\mu = 0, \sigma\text{^}2 = 1$. These coefficients of variation are presented in Table 3.4.

Table 3.4
Coefficients of variation of the sampling distributions of the Percentile Coefficient of Kurtosis and Modified Percentile Coefficient of Kurtosis when sampling from the lognormal distribution with $\mu = 0, \sigma\text{^}2 = 1$.

Coefficient of Variation of Well-Known Estimator	Coefficient of Variation of Newly Proposed Estimator
0.4448	0.386118

Once again, we see that the coefficient of variation of the sampling distribution of the modified estimator is ***smaller*** than the coefficient of variation of the sampling distribution of the well known Percentile Coefficient of Kurtosis.

3.2 Sampling from the Birnbaum Saunders Distribution

The procedure adopted in the case of the lognormal distribution is replicated in the case of the Birnbaum Saunders distribution with $\alpha = \beta = 1$. The simulation study is based on 1000 sample of size 10 each drawn from this distribution.

3.2.1 Comparison of Ranges

Table 3.5 contains maximum, minimum and as well as values of ranges of the sampling distributions of the well-known Percentile Coefficient of Kurtosis and the modified estimator. We see that the range of the sampling distribution of the self-inverse-

based modified estimator is less than that of the well-known Percentile Coefficient of Kurtosis.

Table 3.5
Minimum and Maximum values and Ranges of the sampling distributions of the Percentile Coefficient of Kurtosis and Modified Percentile Coefficient of Kurtosis when sampling from the Birnbaum Saunders distribution with $\alpha=\beta=1$.

Well-Known Estimator			**Newly Proposed Estimator**		
Minimum	Maximum	**Range**	Minimum	Maximum	**Range**
0.04426	0.44849	0.40423	0.06292	0.448086	0.385166

3.2.2 Comparison of Variances

Table 3.6 contains variances of the sampling distributions of the well-known Percentile Coefficient of Kurtosis and the modified estimator based on the 1000 samples of size 10 each drawn from Birnbaum Saunders Distribution with $\alpha=\beta=1$. The variance of the modified estimator is smaller than that of the well known Percentile Coefficient of Kurtosis.

Table 3.6
Variances of the sampling distributions of the Percentile Coefficient of Kurtosis and Modified Percentile Coefficient of Kurtosis when sampling from the Birnbaum Saunders distribution with $\alpha=\beta=1$.

Variance of Well-Known Estimator	**Variance of Newly Proposed Estimator**
0.00797	0.005185

3.2.3 Comparison of Coefficients of Variation

Table 3.7 presents the means of the sampling distribution of the Percentile Coefficient of Kurtosis and the modified estimator obtained by drawing 1000 samples of size 10 each from the Birnbaum Saunders distribution with $\alpha=\beta=1$.

Table 3.7
Means of the sampling distributions of the Percentile Coefficient of Kurtosis and Modified Percentile Coefficient of Kurtosis when sampling from the Birnbaum Saunders distribution with $\alpha=\beta=1$.

Mean of Well-Known Estimator	**Mean of Newly Proposed Estimator**
0.23078	0.222934

Table 3.8 contains a comparison of the coefficients of variation of the sampling distributions of the Percentile Coefficient of Kurtosis and the corresponding modified estimator based on the 1000 samples of size 10 each drawn from BirnBaum Saunders

Distribution with $\alpha=\beta=1$. We see that the coefficient of variation of the modified estimator is ***smaller*** than that of the well known quantile-based measure of kurtosis.

Table 3.8
Coefficient of Variations of the sampling distributions of the Percentile Coefficient of Kurtosis and Modified Percentile Coefficient of Kurtosis when sampling from the Birnbaum Saunders distribution with $\alpha=\beta=1$.

Coefficient of Variation of Well-Known Estimator	Coefficient of Variation of Newly Proposed Estimator
0.38696	0.322994

4. CONCLUDING REMARKS

In this paper, we have utilized the property of self-inversion at unity to propose a modification to the formula of the Percentile Coefficient of Kurtosis and, through simulation studies, have demonstrated that the modified formula yields an estimator that is more efficient than the original estimator in the case when one is sampling from a distribution that is self-inverse at unity. It appears that adoption of the proposed modification may lead to ***more accurate modelling*** of life-data in situations where one has reasons to believe that a multi-parameter distribution defined on the positive half-line and invariant under the reciprocal transformation is a suitable probability model for the data at hand.

REFERENCES

1. Fatima, S.S. and Habibullah, S.N. (2013a). Self-Inversion-Based Modifications of L-Estimators of Central Tendency for Probability Distributions in the Field of Reliability and Safety. *International Conference on Safety, Construction Engineering and Project Management (ICSCEPM 2013), Issues, Challenges and Opportunities in Developing Countries* organized by NUST August 19-21, 2013, Islamabad, Pakistan.
2. Fatima, S.S. and Habibullah, S.N. (2013b). On Modifications of L-Estimators of Dispersion in the Case of Self-Inverse Distributions. Proceedings of *11th International Conference on Statistical Sciences: Social Accountability, Global Economics and Human Resource Development with Special Reference to Pakistan* (Indus International Institute (NCBA&E Sub-Campus), Dera Ghazi Khan, Pakistan, October 21-23, 2013).
3. Fatima, S.S. and Habibullah, S.N. (2013c). On a Modification of the Bowley's Coefficient of Skewness in the Case of Self-Inverse Distributions. Oral Presentation by S.N. Habibullah as *Invited Speaker* at the *Second National Conference on Space Science organized in order to celebrate the World Space Week* (Karachi, Pakistan, October 7-8, 2013).
4. Fatima, S.S., Habibullah, S.N. and Saunders, S.C. (2013). Some Results Pertaining to the Hazard Functions of Self-Inverse Life-Distributions. *Third International Conference on Aerospace Science and Engineering* (ICASE), Aug 21-23, 2013 organized by Institute of Space Technology (IST), Islamabad, Pakistan.

5. Habibullah, S.N. and Saunders, S.C. (2011). A Role for Self-Inversion. *Proceedings of International Conference on Advanced Modeling and Simulation (ICAMS, Nov 28-30, 2011)* 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.
6. 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.
7. Saunders, S.C. (1974). A Family of Random Variables Closed Under Reciprocation. *J. Amer. Statist. Assoc.*, 69(346), 533-539.
8. Seshadri, V. (1965). On Random Variables which have the Same Distribution as their Reciprocals. *Can. Math. Bull.*, 8(6), 819-824.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 155-166

DATA MINING TECHNIQUES FOR RESEARCHERS

Nadia Qasim[1], M. Qasim Rind[2] and **M. Saleem Sheikh[2]**
[1] Research Associate Kings College, London
Email: nadia.nadya@yahoo.com
[2] Preston University, Islamabad, Pakistan
Email: qasim.rind@yahoo.com;
drsmsaleem@yahoo.com

ABSTRACT

Data mining is a computational process of discovering patterns in large data sets involving data mining techniques. The main goal of the data mining process is to extract information from a data set and transform it into an understandable format for further use. İt also involves database and data management aspects, data pre-processing, model and inference considerations, post-processing of discovered structures, visualization, and online updating.

Thus, data mining uses information from past data to analyze the outcome of a particular problem or situation that may arise. Data mining analyzes data stored in data warehouses. The data may come from all parts of business, from the production to the management. Managers can use data mining to decide upon marketing strategies for their product. They can use data to compare and contrast among competitors. Data mining interprets its data into real time analysis that can be used to increase sales, promote new product, or delete product that is not value-added to the company.

In this paper we have described several data mining techniques, including association rules, sequence mining, classification, clustering, deviation detection, and k-nearest neighbors. We have also illustrated the use of some of these techniques. It is concluded that data mining can assist researchers by speeding up their data analyzing process; thus, allowing them more time to work on other projects.

KEY WORD

Data mining, Artificial intelligence, Visualization, Data warehouses.

1. INTRODUCTION

The manual extraction of patterns from data has occurred for centuries. Early methods of identifying patterns in data include Bayes' theorem (1700s) and regression analysis (1800s). The proliferation, ubiquity and increasing power of computer technology has dramatically increased data collection, storage, and manipulation ability. As data sets have grown in size and complexity, direct hands-on data analysis has increasingly been augmented with indirect, automated data processing, aided by other discoveries in computer science, such as neural networks, cluster analysis, genetic algorithms (1950s), decision trees (1960s), and support vector machines (1990s). Data mining is the process

of applying these methods with the intention of uncovering hidden patterns (Kantardzic, 2003) in large data sets. It bridges the gap from applied statistics and artificial intelligence to database management by exploiting the way data is stored and indexed in databases to execute the actual learning and discovery algorithms more efficiently.

2. DATA MINING GOAL

Data mining is defined as a sophisticated data search capability that uses statistical algorithms to discover patterns and correlations in data. The term is an analogy to gold or coal mining; data mining finds and extracts knowledge buried in corporate data warehouses, or information that visitors have dropped on a website, most of which can lead to improvements in the understanding and use of the data. The data mining approach is complementary to other data analysis techniques such as statistics, on-line analytical processing (OLAP), spreadsheets, and basic data access. In simple terms, data mining is another way to find meaning in data.

Data Mining is further defined as an art and science of discovering and exploiting new, useful, and profitable relationships in data, in any business; we will study what happens in order to improve it. We will study our potential and actual customers. The main objective is to find patterns of behavior from the data. In the past study of patterns was limited by the amount of human effort involved, and by the expense of gathering the necessary data. In the last few years, the strategy has changed. More data is readily available to most businesses than ever before. More computing power is available to process the data, and automated techniques are used to find patterns in the data with limited human intervention (Albion Research Ltd, 2002).

Data mining, also known as Knowledge-Discovery in Databases (KDD), is the process of automatically searching large volumes of data for patterns. Data mining is extraction of useful patterns from data sources, e.g., databases, texts, web, image. Thus, data mining applies many older computational techniques from statistics, machine learning and pattern recognition. Technically, data mining is the process of finding correlations or patterns in large relational databases.

Data mining helps business analysts to generate hypotheses, but it does not validate the hypotheses. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Hence, the main goal of data mining is prediction and predictive data mining is the most common type of data mining.

3. MAIN TOOLS FOR DATA MINING

There are several data mining techniques for the purpose of prediction and/or for description. Let us briefly describe the main tools used for data mining:

3.1 Artificial Neural Networks

Non-linear predictive models that learn through training and resemble biological neural networks in structure.

3.2 Decision Trees

Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.

3.3 Rule induction

The extraction of useful if-then rules from data based on statistical significance.

3.4 Genetic algorithms

Optimization techniques based on the concepts of genetic combination, mutation, and natural selection.

3.5 Nearest neighbor

A classification technique that classifies each record based on the records most similar to it in an historical database.

4. BREAKDOWN OF DATA MINING PROCESS

Data mining tools take data and construct a representation of reality in the form of a model. The resulting model describes patterns and relationships present in the data. From a process orientation, data mining activities fall into three general categories such as:

4.1 Discovery

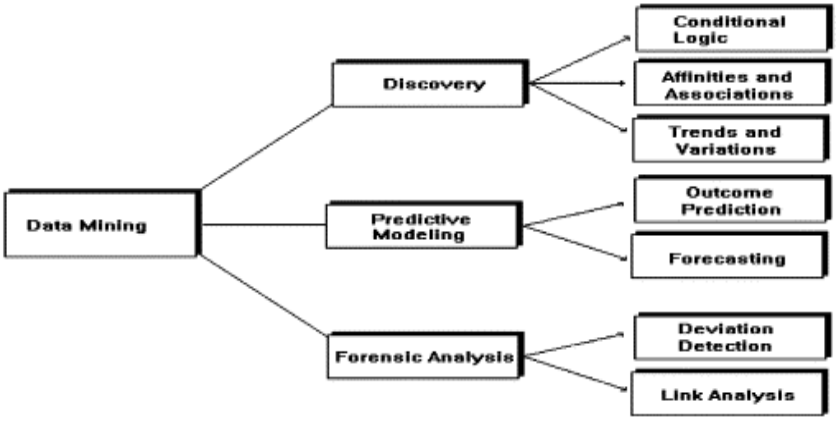
The process of looking in a database to find hidden patterns without a predetermined idea or hypothesis about what the patterns may be.

4.2 Predictive Modeling

The process of taking patterns discovered from the database and using them to predict the future.

4.3 Forensic Analysis

The process of applying the extracted patterns to find anomalous or unusual data elements.



Source: Information Discovery, Inc.

Fig. 1: Breakdown of Data Mining from a Process Orientation

5. PROCESSING STEPS OF DATA MINING

First of all understand the application domain, then collect and create the target dataset. Clean and transform the target dataset. Select features, reduce dimensions, apply data mining algorithms, interpret, evaluate, and visualize patterns. These steps are explained in Fig. 2

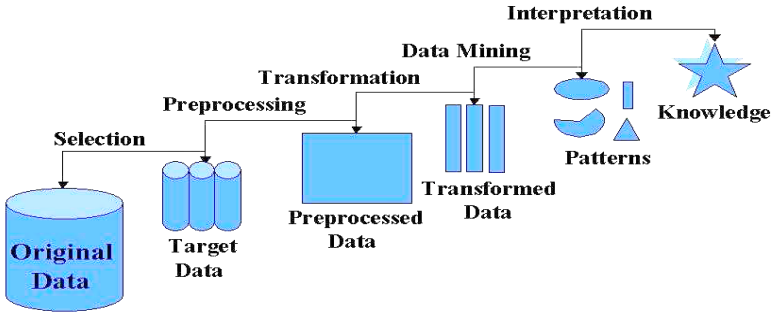


Fig. 2: The Data Mining Process

Data mining is about processing data and identifying patterns and trends in that information so that one can judge and decide the pattern in the data set. There are three stages of data mining processes which are explained as under:

Exploration - This stage usually starts with data preparation which may involve cleaning data, data transformations, selecting subsets of records

Model building and validation - This stage involves considering various models and choosing the best one based on their predictive performance

Deployment - That final stage involves using the model selected as best in the previous stage and applying it to new data in order to generate predictions or estimates of the expected outcome

Hence, data mining consists of five major elements such as: extract, transform, and load transaction data onto the data warehouse system, store and manage the data in a multidimensional database system, provide data access to business analysts and information technology professional, analyze the data by application software and present the data in a useful format, such as a graph or table.

5.1 Pre-processing

Before data mining algorithms can be used, a target data set must be collected. As data mining can only uncover patterns actually present in the data, the target data set must be large enough to contain these patterns. A common source for data is a data mart or data warehouse. Pre-processing is essential to analyze the multivariate data sets before data mining. The target set is then cleaned. Data cleaning removes the observations containing noise and those with missing data.

5.2 Data Preparation

Data mining itself relies upon building a suitable data model and structure that can be used to process, identify, and build the information that is required. While considering the

combination of the business requirements for the data mining, the identification of the existing variables (customer, values, country) and the requirement to create new variables that you might use to analyze the data in the preparation step. You might compose the analytical variables of data from many different sources to a single identifiable structure (for example, you might create a class of a particular grade and age of customer, or a particular error type). Several core techniques are used in data mining describe the type of mining and data recovery operation.

6. DATA MINING MODELS

Data mining is used to construct eight types of models aimed at solving business problems: classification, regression, time series, prediction, clustering, summarization association analysis, and sequence discovery. The first two, classification and regression, are used to make predictions, while association and sequence discovery are used to describe behavior. Clustering can be used for either forecasting or description. Data mining involves six common classes of tasks which are narrated below (Fayyad, 1996).

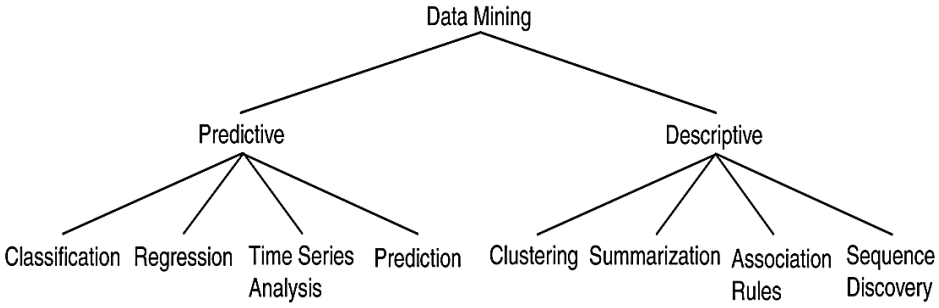


Fig. 3: Data Mining Models

6.1 Predictive Models

6.1.1 Classification – It can be used to build up an idea regarding type of customer, item, or object by describing multiple attributes to identify a particular class. For example, one can easily classify cars into different types (Toyota, Honda, Ford etc.) by identifying different attributes (number of seats, car shape, driven wheels). Given a new car, one might apply it into a particular class by comparing the attributes with known definition. One can apply the same principles to customers, for classifying them by age and social group.

6.1.2 Prediction - Prediction is a wide topic and runs from predicting the failure of components or machinery, to identifying fraud and even the prediction of company profits. Used in combination with the other data mining techniques, prediction involves analyzing trends, classification, pattern matching, and relation. By analyzing past events or instances, one can make a prediction about an event. Using the credit card authorization, for example, you might combine decision tree analysis of individual past transactions with classification and historical pattern matches to identify whether transaction is fraudulent.

6.2 Descriptive Data Mining Models

6.2.1 Association - Association is probably the better known and most familiar and straightforward data mining technique. Here, you make a simple correlation between two or more items, often of the same type to identify patterns. For example, when tracking people's buying habits, you might identify that a customer always buys cream when they buy strawberries, and therefore suggest that the next time that they buy strawberries they might also want to buy cream. Association rule learning (Dependency modeling) will searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.

6.2.2 Clustering - By examining one or more attributes or classes, you can group individual pieces of data together to form a structure opinion. At a simple level, clustering is using one or more attributes as your basis for identifying a cluster of correlating results. Clustering is useful to identify different information because it correlates with other examples so that you can see where the similarities and ranges agree. Thus, clustering is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.

6.2.3 Sequential patterns - Sequential patterns are a useful method for identifying trends, or regular occurrences of similar events. For example, with customer data you can identify that customers buy a particular collection of products together at different times of the year. In a shopping basket application, you can use this information to automatically suggest that certain items be added to a basket based on their frequency and past purchasing history. Sequential pattern mining finds sets of data items that occur together frequently in some sequences. Sequential pattern mining, which extracts frequent subsequences from a sequence database, has attracted a great deal of interest during the recent data mining research because it is the basis of many applications, such as: web user analysis, stock trend prediction, DNA sequence analysis, finding language or linguistic patterns from natural language texts, and using the history of symptoms to predict certain kind of disease.

6.2.4 Decision trees - Related to most of the other techniques (primarily classification and prediction), the decision tree can be used either as a part of the selection criteria, or to support the use and selection of specific data within the overall structure. Within the decision tree, you start with a simple question that has two (or sometimes more) answers. Each answer leads to a further question to help classify or identify the data so that it can be categorized.

7. RESULTS VALIDATION

The final step of knowledge discovery from data is to verify that the patterns produced by the data mining algorithms occur in the wider data set. Not all patterns found by the data mining algorithms are necessarily valid. It is common for the data mining algorithms to find patterns in the training set which are not present in the general data set. This is called over fitting. To overcome this, the evaluation uses a test set of data on

which the data mining algorithm was not trained. The learned patterns are applied to this test set and the resulting output is compared to the desired output. For example, a data mining algorithm trying to distinguish "spam" from "legitimate" emails would be trained on a training set of sample e-mails. Once trained, the learned patterns would be applied to the test set of e-mails on which it had not been trained. The accuracy of the patterns can then be measured from how many e-mails they correctly classify. If the learned patterns do not meet the desired standards, then it is necessary to re-evaluate and change the pre-processing and data mining steps. If the learned patterns do meet the desired standards, then the final step is to interpret the learned patterns and turn them into knowledge.

8. MAJOR APPLICATIONS OF DATA MINING

Data mining is applied in every field; some of its applications are narrated below.

8.1 Business Data Mining

Data mining is the analysis of historical business activities, stored as static data in data warehouse databases, to reveal hidden patterns and trends. Data mining software uses advanced pattern recognition algorithms to sift through large amounts of data to assist in discovering previously unknown strategic business information. Hence, business data mining is used for performing market analysis to identify new product bundles, finding the root cause of manufacturing problems, to prevent customer attrition and acquire new customers, cross-sell to existing customers, and profile customers with more accuracy (O'Brien, 2011).

In today's world raw data is being collected by companies at an exploding rate. For example, Wal-Mart processes over 20 million point-of-sale transactions every day. This information is stored in a centralized database, but would be useless without some type of data mining software to analyze it. If Wal-Mart analyzed their point-of-sale data with data mining techniques they would be able to determine sales trends, develop marketing campaigns, and more accurately predict customer loyalty. Every time we use our credit card, a store loyalty card, or fill out a warranty card data is being collected about our purchasing behavior.

In new business era, the importance of Customer Relationship Management (CRM) applications has been increased many fold. According to Swift, 2001 and Kracklauer, Mills, and Seifert (2004), the Customer Relationship Management (CRM) consists of four dimensions:

1. Customer Identification;
2. Customer Attraction;
3. Customer Retention;
4. Customer Development.

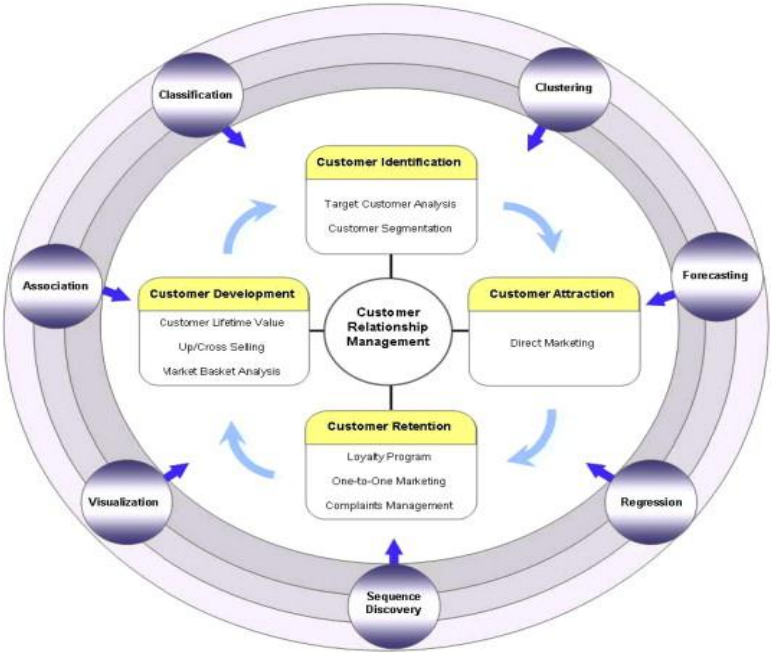


Fig. 4: Classification Framework for Data Mining Techniques in CRM

Many people search stored information about us from companies, such as Google, Facebook, and Amazon. Although there is the potential for our personal data to be used in harmful, ways it is also being used to make our lives better. For example, Ford and Audi may collect information about customer driving patterns so they can recommend safer routes and warn drivers about dangerous road conditions.

8.2 Medical Data Mining

In Medical field data mining can help to healthcare insurers detect fraud and abuse, to healthcare organizations make customer relationship management decisions, to physicians identify effective treatments and best practices, and to patients receive better and more affordable healthcare services. In the study of human genetics, sequence mining helps address the important goal of understanding the mapping relationship between the inter-individual variations in human DNA sequence and the variability in disease susceptibility. In 2011, the case of Sorrell v. IMS Health, Inc., decided by the Supreme Court of the United States, ruled that pharmacies may share information with outside companies. This practice was authorized under the 1st Amendment of the Constitution, protecting the "freedom of speech."

8.3 Spatial Data Mining

Spatial data mining is the application of data mining methods to spatial data. The end objective of spatial data mining is to find patterns in data with respect to geography. So far, data mining and Geographic Information Systems (GIS) have existed as two separate technologies, each with its own methods, traditions, and approaches to visualization and

data analysis. Particularly, most contemporary GIS have only very basic spatial analysis functionality. The immense explosion in geographically referenced data occasioned by developments in IT, digital mapping, remote sensing, and the global diffusion of GIS emphasize the importance of developing data-driven inductive approaches to geographical analysis and modeling (Goodchild & and Rhind, 1991) .

8.4 Sensor Data Mining

Wireless sensor networks can be used for facilitating the collection of data for spatial data mining for a variety of applications such as air pollution monitoring. A characteristic of such networks is that nearby sensor nodes monitoring an environmental feature typically registers similar values. This kind of data redundancy due to the spatial correlation between sensor observations inspires the techniques for in-network data aggregation and mining. By measuring the spatial correlation between data sampled by different sensors, a wide class of specialized algorithms can be developed (Ma, Richards, Ghanem, Guo, & Hassard, 2008).

8.5 Visual Data Mining

Visual data mining is a new idea that uses recent technology to apply some specific principles to how humans interpret data. It is a very general method, applied in detailed ways to get specific results, in areas like finance, medicine, public administration and government, transportation, and much more. In visual data mining, programmers build interfaces that allow for visual presentations to be a part of how human users interpret the data that they see. Computers can collect and display a huge amount of data. With data mining, humans can interpret this data in different ways. That is the power that data mining brings to the human community, and the potential that its practitioners are looking at for improving modern methodologies. Information visualization and visual data mining can help to deal with the flood of information. The advantage of visual data exploration is that the user is directly involved in the data mining process. Studies suggest visual data mining is faster and much more intuitive than is traditional data mining (Keim, 2002).

8.6 Music Data Mining

In the Internet age, a gigantic amount of music-related information is easily accessible. For example, music and artist information can be obtained from the artist and record company web sites, song lyrics can be downloaded from the lyrics databases, and music reviews are available from various discussion forums, blogs, and on-line magazines. Data mining techniques, and in particular co-occurrence analysis, has been used to discover relevant similarities among music corpora (radio lists, CD databases) for the purpose of classifying music into genres in a more objective manner (Pachet, 2001).

9. DATA MINING SOFTWARE

There are so many data mining software's available in the market depending upon the field of specialization. As well as, there are source wise data mining software's. They are classified in two categories, (Haughton & Deichmann, 2003).

9.1 Free open-source data mining software and applications

- Carrot2: Text and search results clustering framework.
- Chemicalize.org: A chemical structure miner and web search engine.

- ELKI: A university research project with advanced cluster analysis and outlier detection methods written in the Java language.
- GATE: a natural language processing and language engineering tool.
- SCaViS: Java cross-platform data analysis framework developed at Argonne National Laboratory.
- KNIME: The Konstanz Information Miner, a user friendly and comprehensive data analytics framework.
- ML-Flex: A software package that enables users to integrate with third-party machine-learning packages written in any programming language, executes classification analyses in parallel across multiple computing nodes, and produce HTML reports of classification results.
- NLTK (Natural Language Toolkit): A suite of libraries and programs for symbolic and statistical natural language processing (NLP) for the Python language.
- SenticNet API: A semantic and affective resource for opinion mining and sentiment analysis.
- Orange: A component-based data mining and machine learning software suite written in the Python language.
- R: A programming language and software environment for statistical computing, data mining, and graphics. It is part of the GNU Project.
- Rapid Miner: An environment for machine learning and data mining experiments.
- UIMA: The UIMA (Unstructured Information Management Architecture) is a component framework for analyzing unstructured content such as text, audio and video – originally developed by IBM.
- Weka: A suite of machine learning software applications written in the Java programming language.

9.2 Commercial data-mining software and applications

- Angoss Knowledge STUDIO: data mining tool provided by Angoss.
- BIRT Analytics: visual data mining and predictive analytics tool provided by Actuate Corporation.
- Clarabridge: enterprise class text analytics solution.
- IBM DB2 Intelligent Miner: in-database data mining platform provided by IBM, with modeling, scoring and visualization services based on the SQL/MM - PMML framework.
- IBM SPSS Modeler: data mining software provided by IBM.
- KXEN Modeler: data mining tool provided by KXEN.
- LION solver: an integrated software application for data mining, business intelligence, and modeling that implements the Learning and Intelligent Optimization (LION) approach.
- Microsoft Analysis Services: data mining software provided by Microsoft.
- Oracle Data Mining: data mining software by Oracle.
- Predixion Insight: data mining software by Predixion Software.
- SAS Enterprise Miner: data mining software provided by the SAS Institute.
- STATISTICA Data Miner: data mining software provided by StatSoft.

10. COMMENTS AND CONCLUSION

Data mining is a process of running complex queries on the data which are stored in the database. To identify the format of information that is needed for analysis. When you have information in the format you need, you can apply the different data mining techniques.

This paper describes some of the benefits to be gained from advanced data mining techniques, including how your organization can achieve a greater return on investment (ROI). It offers practical guidance for incorporating additional data types and sources, expanding the scope of data mining projects and deploying results more effectively throughout your organizations.

Data mining can aid direct marketers by providing them with useful and accurate trends about their customers' purchasing behavior. Data mining can assist financial institutions in areas such as credit reporting and loan information. Data mining can aid law enforcers in identifying criminal suspects as well as apprehending these criminals by examining trends in location, crime type, habit, and other patterns of behaviors.

It is also felt that there are some security issues in data mining process, they do not have sufficient security systems in place to protect the personal information. Some cases are reported regarding misuse of information. Some of the company will answer your phone based on your purchase history. If you have spent a lot of money or buying a lot of product from one company, your call will be answered really soon. So you should not think that your call is really being answer in the order in which it was received.

We have given an overview of data mining processes. We have described several data mining techniques, including association rules, sequence mining, classification, clustering, deviation detection, and k-nearest neighbors. We have also illustrated the use of some of these techniques. Sources of data mining software's are listed for testing and applications. It is concluded that data mining can assist researchers by speeding up their data analyzing process; thus, allowing those more time to work on other projects.

REFERENCES

1. Fayyad, U. (1996). From Data Mining to Knowledge Discovery in Databases. *Articles* (pp. 37-54). Al Magzine.
2. Goodchild, M.F. and Rhind, D.W. (1991). *Geographic Information Systems: Principles and Applications principles and applications.* London, UK Longman.
3. Haag, Stephen; Cummings, Maeve; Phillips, Amy (2006). *Management Information Systems for the information age*. Toronto: McGraw-Hill Ryerson. "http://en.wikipedia.org/wiki/International_Standard_Book_Number" \o "International Standard Book Number" ISBN, "http://en.wikipedia.org/wiki/Special:BookSources/0-07-095569-7" \o "Special: Boo k Sources /0-07-095569-7" 0-07-095569-7.
4. Haughton, D., and Deichmann. (2003). A Review of Software Packages for Data Mining. *The American Statistician*, 57, 290-309.
5. Kantardzic, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. New York: John Wiley & Sons.

6. Keim, D.A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1-8.
7. Ma, Y., Richards, M., Ghanem, M., Guo, Y. and Hassard, J. (2008). Air Pollution Monitoring and Mining Based on Sensor Grid in London. *IEEE Sensor Journal*, 11(3), 641-648.
8. O'Brien, J.A. and Marakas, G.M. (2011). *Management Information Systems.* New York: McGraw-Hill/Irwin.
9. Pachet, F.W. (2001). Musical Data Mining for Electronic Music Distribution Proceedings of the *1st WedelMusic Conference*, Firenze, Italy.
10. Zhu, Xingquan; Davidson, Ian (2007). *Knowledge Discovery and Data Mining: Challenges and Realities*. New York, NY: Hershey. p. 18.
11. Zernik, Joseph (2010). Data Mining as a Civic Duty – Online Public Prisoners' Registration Systems, *International Journal on Social Media: Monitoring, Measurement, Mining*, 1, 84-96. http://www.scribd.com/doc/38328591/

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 167-172

ANALYSIS FOR THE DEGREE OF PURCHASE INTENTION IN PRODUCT TEST

Hina Gul[1] and **Muhammad Amin**[2]
[1] Prontolinks, Karachi, Pakistan. Email: gul.hina08@gmail.com
[2] School of Mathematical Sciences, Dalian University of Technology
Dalian, Liaoning, China. Email: amin@dlut.edu.cn

ABSTRACT

In this paper we have used verbatim analysis technique to draw the main impact and the actual direction of purchase intention of product based on different attributes. The main objective of this research is to interpret the large number of attributes and find their effect on the degree of purchase intention. For this purpose we have used verbatim analysis technique with some modification.

Three beverages are tested which were coded as Alpha, Sigma and Delta. Central location test (CLT) technique was used for the survey. The study was conducted amongst 120 males of age 15-35 year old belonging to A, B and C Socio- economic Class. Each respondent tasted all three brands and the questionnaire was filled after each test. Same questionnaires were used for each product. Purchase Intention, overall Likeness, Carbonation Level, Color, Flavor and Aroma were rated on five point scale. Except these attributes around 40 verbatim were also rated on 5 point scale.

The main task was that how to analyze the 40 pre-defined attributes which explained the feeling of respondent after taste and calculate the effect on purchase intention. For this purpose verbatim technique was used after some modification. This analysis has been applied before on qualitative research but here it was used for quantitative data. This analysis gives the clear direction of purchase intention of all products. Delta was winner in all aspects and it was the accepted direction of purchase while sigma was negative in direction and alpha was approximately neutral. While basic statistics didn't display the clear result.

Furthermore this analysis can help to find the actual degree and direction of purchase on the basis of different attributes for different products.

1. INTRODUCTION

Willingness of purchase of any product is the prime factor of the product testing survey; however the accurate measurement is the most important. Mostly we use average rating of purchase intention to interpret it. But there are some confusion, may be any respondent give high rate to the purchase intention just to fulfill the requirement of questionnaire but he may not like the features of the product. However, this average rating is memory less or not cover whole questionnaire in the same manners. For example in this study we have asked about taste, color and some other characteristics of the product. Suppose a respondent likes taste of a product but not its color. He might get confused about that either he would purchase the product or not. He will rate neutral or

positive response for purchase of that product but will not accept that product. Thus we gather all feature and cross tab them with the purchase intention, and find a single solution for degree of purchase.

The data was collected through CLT survey method which is the most common and suitable method to collect the response of any product. In this method each respondent had tasted all the three products and after taste he gives his response on each product through pre-defined questionnaire. The Questionnaire contains basic demography of respondents, mostly used product likeness of different attributes. Purchase intention and some other questions, which explain the expectation and feelings of respondent after testing the related product. Mainly these questions are our target for this study. Actually respondents show their opinion which indirectly effects on purchase

2. THE BASIC ANALYTICAL STRUCTURE

It was defined above that we had collected the data on five point scale where

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

The purchase intention was also rated on 5 point scale, which follows as

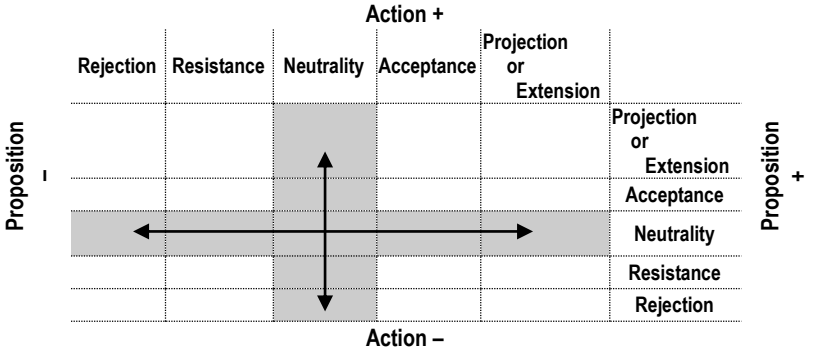
1. Definitely would not buy ~ Rejection of Product
2. Would not buy ~ Resistance of Product
3. Can't say either buy or not ~ Neutrality
4. Would buy ~ Acceptance
5. Definitely would buy ~ Projection

3. METHOD OF ANALYSIS

Walter G. Mitchell had used this method for the analysis of advertising research Verbatim he said that "verbatim responses fall into one or more of five categories and may be synthesized by a vector diagram revealing their general direction, so all the verbatim were categorized in to five categories of propositions and five levels of action, after the categorization of these verbatim he analyzed them and found a single vector solution, while our data of this research was already categorized on five point scale so that for the analyzing point of view we take the purchase intention on action scale and the other attributes on proposition scale. We cross tab these by using SPSS 12 and got a 5X5 table which is the main base of our analysis.

After getting this table we had followed all the steps of Walter G Mitchell analysis, as I had mentioned this analysis was used before for the advertising research, for qualitative data and here it is being used on quantitative data. We now discuss all procedure of analysis in following steps.

1. Cross tab the purchase intention with others attributes and got the basic structure the following basic structure follow in this study
2. Convert these frequencies into percentages such that the sum up makes 100% for the entire structure.



3. Convert percentages in any non-neutral (non perpendicular) square to a higher percentage by computing the straight- line distance from the center of the square to the origin (this is the hypotenuse of the right triangle with the X axis 1 or 2 units and the y axis 1 or 2 units).
4. Draw a vector diagram of all these forces radiating out in the proper directions from the neutral origin point with a force length equal to the correct percentages. Solve this diagram for the single resultant vector. This vector shows the degree and direction of the purchase intention.

4. CALCULATION AND RESULTS

Out of 120 respondents 50% males and 50% females, equal size of samples were taking from A, B and C Class. Age of respondent b/w 16-35 years, 59% respondent were employees while 41% were students, According to basic Statistics Mountain dew was most frequently used drink among all users, than Pepsi and 7-up. When we asked about last ten brand consumption, 71 % respondents says that Mountain Dew was the present in last 10 brands while Pepsi share just 49% with coca cola. Overall likeness of delta is 4.2 while 3.2 for the sigma and alpha so here tie between alpha and sigma. When level of sweetness of Sigma was 3.1 while Delta and alpha were 3.0, fuzziness level of Delta and alpha were 3.1 and sigma was 2.8. Average rating of main attributes is giving in table 1.

Table 1

Attributes	**Alpha**	**Delta**	**Sigma**
Overall Likeness	3.2	4.2	3.2
Fulfill ness of Expectation	3.2	3.9	3.1
Sweetness Level	3.0	3.0	3.1
Fuzziness Level	3.1	3.1	2.8
Color Likeness	3.7	3.9	3.7
Bitterness Felt	3.5	4.0	3.5
Taste Intensity	3.3	3.2	3.0
Aroma Likeness	3.5	3.9	3.3
Excitement Felt	2.6	3.0	2.5

According to these results delta leads among these tree products and the purchase intention shown in Figure 1, which are high for delta if we consider above results. The average rating lies around neutral that means this value of purchase intention not cover the whole impact of products

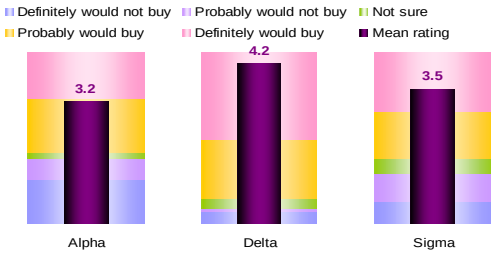


Figure 1

For the better assessment about product we include some other question in questionnaire which is not directly related to the product likeness, but most important for the consumption behavior and pattern. There is 5-point likert scale average rating show and fig 2, we are unable to conclude any perception about the product, because other than delta, sigma and alpha rating did not give clear result. Sigma rated high at some points but at other points it's less than Alpha

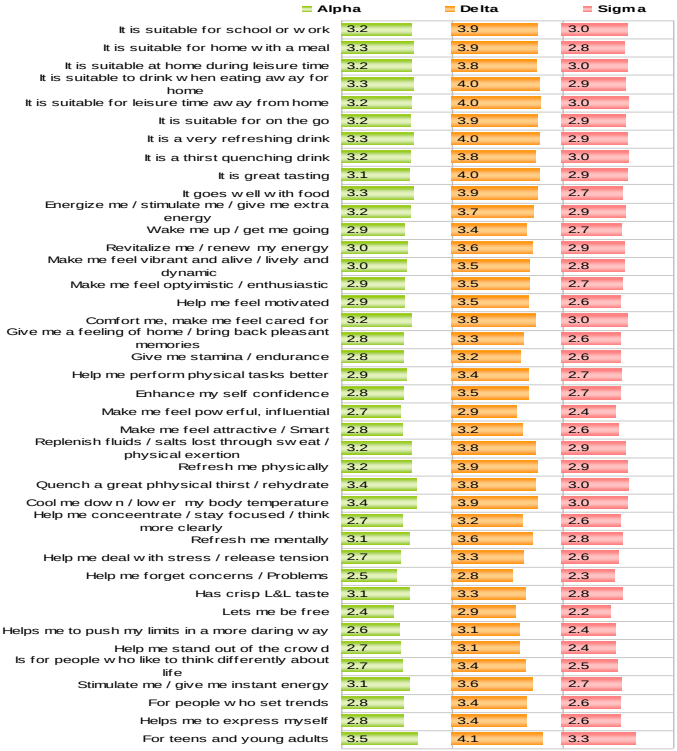


Figure 2

So that we used modified verbatim analysis for all the procedures already we have explained the detail Calculation for product alpha is by the cross tabulation of purchase intention with others attributes, we got the basic structure which shown in table 2 response percentages given in brackets. The sum of all categories is 100%.

Table 2

Proposition	Action +				
		5 (0.1)	4 (0.1)	92 (1.9)	624 (13)
	34 (0.7)	66 (1.4)	49 (1)	526 (11)	676 (14.1)
–	70 (1.5)	75 (1.6)	135 (2.8)	459 (9.6)	147 (3.1)
	90 (1.9)	417 (8.7)	120 (2.5)	127 (2.6)	82 (1.7)
	406 (8.5)	237 (4.9)	92 (1.9)	116 (2.4)	151 (3.1)
	Action –				Proposition +

Further analysis conducted on this table and finally we got the resultant vector for the alpha which was 54.27 and the angle was .428^{o} and resultant vector for delta and sigma were 129.7 and 32.17 respectively, While degree of purchase intention were 24.83^{o} and -54.27^{o}. The results are represented clearly in figure 3.

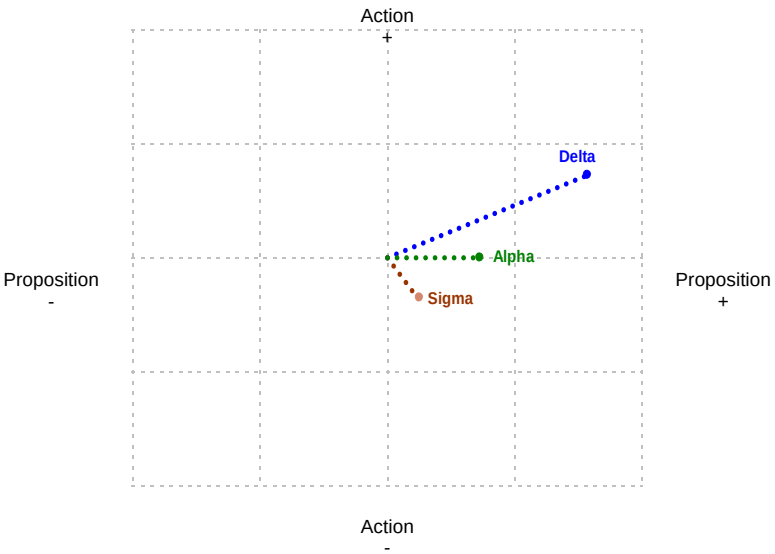


Figure 3

This figure shows that all product have positive proposition but delta fall in acceptance region of action, and its direction towards the extension, so this product lead all over the products. Alpha is neutral while Sigma in negative direction falls in rejection region of action. However, this analysis gives the clear picture of purchase intention.

6. REASONS FOR APPLYING THIS ANALYSIS

As it was defined above that mainly average rating or percentage against each rate used to find the amount of purchase of any product but did not help us to find out the conclusion for Alpha and sigma purchase. Because both perception change attribute to attribute while with the help of current analysis we found out the clear impact of these attribute on purchase.

We can also use this for find the impact of other attributes which are common and directly ask the respondent about the likeness of product for future studies.

REFERENCES

1. Mitchell, Walter G. (2000). Systematic Synthesis of Advertising Research Verbatims. *J. Adver. Res.*, 12, 1-8.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 173-178

INVESTIGATING THE RELATIONSHIP BETWEEN SELF-TALK AND SELF-DETERMINATION ON PSYCHOLOGICAL WELL-BEING OF PRE-LINGUISTIC DEAF STUDENTS

Sadia Batool, Rubab Afzal and **Muhammad Qasim Rind**
Preston University, Islamabad, Pakistan
Email: drsadiagardezi@gmail.com
rubabafzal@hotmail.com
dr_rind@yahoo.com

ABSTRACT

Psychological well-being is much prominent feature of our society because people can only be self-actualized when they are psychologically healthy; similarly, young adulthood (20-24) is a very crucial period of one's life; so psychological well-being of pre-linguistic deaf is of vital importance at this stage. Pre-linguistic deaf is a person whose hearing impairment is congenital; he renders the disability more severe because the sufferer is unable to communicate from the outset. Self-talk can have a really great impact on the deaf student self-esteem and confidence. There is positive and negative self-talk and they both have an impact on how one feels. Self-talk permeates throughout life and can help to improve attitude, performance and self-efficacy; it has impact on overall psychological well-being.

Self-determination is the belief that one can control his own destiny. It is a combination of attitudes and abilities that lead people to set goals for them, and to take the initiative to reach these goals. It was hypothesized that positive self-talk and self-determination is positively related to psychological well-being of the deaf students, negative self-talk has negative relationship with self-determination among the deaf students and negative self-talk is negatively related with psychological well-being. The self-talk inventory, self-determination scale, and psychological well-being scale was used for collecting the data. Data were personally collected from 60 deaf students (30 boys and 30 girls) from schools located at Rawalpindi, Islamabad and Lahore. Data were analyzed using the descriptive statistics (Mean, Standard Deviation) and Pearson-r. It was observed that high self-determination is positively related to psychological Well-Being of pre-linguistic deaf students, Negative self-talk is negatively related with psychological well-being, Positive self-talk and psychological well-being are positively related and negative self-talk is negatively related with self-determination.

KEYWORDS

Pre-linguistic deafness; Psychological well-being; Self-talk; Self-determination; Young adulthood.

1. INTRODUCTION

Pre-linguistic deafness is the hearing impairment in which the impairment is a hereditary condition that exists at birth. It is the time before the individual has acquired

speech and language capability thus rendering the disability more severe because the sufferer is unable to use the language from the very beginning (Dhawan, 2005). The problems of severely hearing impaired persons in terms of social, psychological and cognitive development throughout the life cycle are not merely due to the biological condition of deafness but rather to interference in the complex interactional process (Maher, 1989). Whereas, Psychological well-being is a subjective term that means different things to different people. It is a driving force behind all human behavior; and a universal human desire to fulfill one's potential. It is very similar to the concept of self-actualization postulated by Abraham Maslow. Self-actualization literally means to 'become what we are' (Boeree, 1997). Ryff (1995) operationally defined psychological well-being as: self-acceptance, autonomy, environmental mastery, purpose in life, positive relations with others, and personal growth.

Research on well-being has flourished in recent decades (Diener, Suh, Lucas, & Smith, 1999; Kahneman, Diener, & Schwarz, 1999). It refers to positive mental health. Research has shown that psychological well-being is a diverse multidimensional concept (Macleod & Moore, 2000; Ryff, 1989), which develops through a combination of emotional regulation. Ryan and Deci's (2001) integrative review organized the field of well-being into two broad traditions: one dealing with happiness (hedonic well-being), and the other dealing with human potential.

Ryff and Connell (1989) have brought about a shift in focus from a subjective to an objective conception of psychological well-being. Their research is theoretically and conceptually grounded on Maslow's (1968) conception of self-actualization, which has resulted in a new objective psychological well-being measurement, with the components such as autonomy, personal growth, environmental mastery, purpose in life, positive relation with others and self-acceptance (Conway & Macleod, 2002; Keyes, Shmotkin & Ryff, 2002; Ruini et al., 2002; Ryff & Keyes, 1995).

Self-talk is a healthy behavior that has potentially far-reaching effects, it is most likely used by those who have a high internal locus of control. Self-talk is categorized as being positive or negative. As its label implies, positive self-talk has good implications for people's mental and physical well-being. However, the negative is not all bad. The key to using self-talk is to strive for an appropriate balance (which is a tenet of holistic medicine itself) between the two. Communication and medical professionals have researched the physiological components of self-talk, to conclude that what people say to themselves does affect their ability to combat and ward off illnesses. Individuals can tap into the power of their own self-talk by recognizing it for what it is, reducing harmful negativity, and increasing the number of positive internal messages.

Self-Talk frequently results in positive or negative effect on one's performance (Sellars, 1997). It has been found, for instance, depressed individuals are less able to maintain interest in the positive Self-Talk (Ronan, 1994). Negative self-talk can serve as a predictor for anxious behavior (Kendall & Treadwell, 2007).

Self-determination is an approach to human motivation and personality that highlights the importance of human' evolved inner resources for personality development and behavioral self-regulation (Ryan, Kuhl and Deci, 1997). Theories have shown that when people conceive their behavior as internally motivated-determined by their own volition,

there are stronger consequences (Condry 1977; Deci 1992; Weiner 1979). It is the degree to which an individual's behavior is self-endorsed and self-determined (Ryan & Deci, 2000)

As young adulthood is very crucial period of life so, psychological well-being is of immense importance. In keeping with the theoretical grounding of objective well-being in models of adult development, the relationship between age and well-being has been of central interest to investigators and is associated with advantages and disadvantages in mental health (Ryff, 1989; Ryff, Keyes, & Hughes, 2004).Young adults with low psychological well-being may encounter lower levels of happiness, satisfaction, and self-esteem, while experiencing high levels of distress (Amato, 1994).

2. HYPOTHESES

In order to meet the objective of the research, following hypotheses were formulated:

1. Positive Self-Talk and self-determination is positively related to psychological Well-Being of pre-linguistic young adults.
2. Negative Self-Talk is negatively correlated with self-determination among pre-linguistic young adults.
3. Negative self-talk is negatively related with psychological well-being of pre-linguistic young adults.

3. DATA SAMPLE

The participants consisted of 60 young hearing impaired adults, 30 boys and 30 girls. The ages of the respondents ranged from 20 to 24. Fifty six percent children were orally educated; the remaining children used total communication. All children were judged as being able to understand simple, orally presented test instructions. All this training was given by using pre-linguistic vocal development process, which is the process by which children begin to produce increasingly complex speech and language with the help of therapists, audiologists, and teachers of deaf children. The researcher has also done specialization in hearing impairment and also having the 8 years of teaching experience with hearing impaired children.

4. RESEARCH INSTRUMENTS

The following research instruments were used to collect the data from pre-linguistic young adults, from schools located at Rawalpindi, Islamabad and Lahore.

4.1 Psychological well-being

Index of Personal Growth (IPG) is an e-factor model of self-actualization proposed by Jones and Crandall (1986). It consisted of 35 items, each item anchored on 5 point scale ranging from "not very probable incorrect" to "very probable".

4.2 Self-talk Inventory

(Calvete, E., Estivez A., Landin, C., Yolanda, Olga, M., Lourdes, C., Villardon, and Villa, A., 2005). It consisted of 52 items, 26 for measuring negative self-talk and 26 items for measuring positive self-talk. Each item anchored on 4 point scale ranging from "not very probable incorrect'' to "very probable".

4.3 The Self-Determination Scale (Ryan & Deci, 2002)

The SDS is a short, 10-item scale, with 5 anchored items

5. DATA ANALYSIS

The data were analyzed using statistical methods and processed through SPSS package. Following tables shows the summarized results of the analysis.

Table 1
Correlation between Self-talk and Psychological well-being

Psychological well-being			
	M	**SD**	**P**
Positive Self-talk	95.4	20.4	0.17
Negative self-talk	74.5	18.6	0.12

It is observed from the above Table 1 that positive self-talk and psychological well-being of pre-linguistic deaf children are positively related to one another The mean value is 95.4, standard deviation value is 20.4 and P value is 0.17; whereas, negative self-talk is negatively related with self-determination. The mean value calculated is 74.5, standard deviation is 18.6 and p value is 0.12.

Table 2
Correlation between Self-determination and Psychological well-being

Psychological well-being			
	M	**SD**	**p**
Self-determination	58	8.37	0.01

It is seen from the above table. 2 that high self-determination of pre-linguistic deaf children is positively related to their psychological well-being it is proved from statistical methods that the mean of above variables is 58, standard deviation is 8.37 and P value is 0.01.

Table 3
Correlation between Negative Self-Talk and Self determination

Negative Self-talk			
	M	**SD**	**p**
Self-determination	38.6	8.37	0.001

It is felt from above Table 3, that negative self-talk of children is negatively related with their psychological well-being. Result shows that mean value of variables 38.6, standard deviation is 8.37, and p value is 0.001.

6. DISCUSSION

Psychological well-being is a point of much emphasis in society today, because people can be self-actualized and use their full potentials, when they are psychologically

healthy. Young adulthood is very crucial period of life so, psychological well-being is very imperative. In keeping with the theoretical grounding of objective well-being in models of adult development, the relationship between age and well-being has been of central interest to investigators and is associated with advantages and disadvantages of mental health. The psychological well-being of hearing impaired children and their self-determination are much affected by their language handicap, as they cannot be understood by everyone and they even feel the shame to face the society, so research in this area is indispensible.

7. CONCLUSION

The present study was designed to explore the effects of self-talk and self-determination on psychological well-being of pre-linguistic hearing impaired children. As far as researcher's knowledge is concerned research on this topic has not yet been conducted. Results have shown that high self-determination of pre-linguistic deaf children is positively related to their psychological well-being. Negative self-talk of these children is negatively related with their psychological well-being, positive self-talk and psychological well-being of pre-linguistic deaf children is positively related to one another. Whereas, negative self-talk is negatively related with self-determination.

8. RECOMMENDATIONS

We make decisions all the times in our lives and sometimes we wonder why we did or say something that did not work out the way we want. In order to overpower the unfavorable ideas, we must talk to ourselves like ' I can do this or I will manage this alone' and keep repeating these words until subconscious quits letting the self-talk become action. This technique is very useful in all walks of life like while losing weight, reinforcing muscles, or alleviates stress, working on a project, withdrawing the stage fright etc.

Dr. Field and Dr. Hoffman (1994) developed self-determination instructional interventions. They believed instructional activities in the area such as increasing self-awareness improving decision making, goal setting, goal attainment skills enhancing relationship skills may enhance student self-determination of children with disability. Modeling is the best way to develop self-determination.

It is recommended that Pre-linguistic hearing impaired children may be given number of opportunities to make their own decisions, families can provide their children with disabilities the opportunities to explore employment, housing options, and community recreation programs in their community by utilizing their own network of relatives and friends, as well as formal service systems.

REFERENCES

1. Amato, P.R. (1994). Father-child relations, mother-child relations, and offspring psychological well-being in early adulthood. *Journal of Marriage and Family*, 56, 1031-1042.
2. Boeree, C.G. (1997). Alfred Adler. URL: http://www.ship.edu/~cgboeree/adler.html [Online, accessed 15 Oct. 2000].

3. Calvete, E., Estivez A., Landin, C., Yolanda, Olga, M., Lourdes, C., Villardon, and Villa, A. (2005). Self-Talk and Affective Problems in College Students: Valence of Thinking and Cognitive Content Specificity. *The Spanish Journal of Psychology*, 8(1), 56-67
4. Conway, C. and MacLeod, A.K. (2002). Well-being: Its importance in clinical research and practice. *Clinical Psychology*, 16, 26-29.
5. Deci, E. (1992). On the nature and functions of motivation theories. *Psychological Science*, 3, 167-171.
6. Deci, E.L. and Ryan, R.M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227-268.
7. Dhawan, M.L. (2005). *Education of Children with Special Needs*. Isha Books, India.
8. Diener, E., Suh, E.M., Lucas, R.E. and Smith, H.E. (1999). Subjective well-being: Three Decades of Progress. *Psychological Bulletin*, 125, 276-302.
9. Kahneman, D. Diener, E. and Schwarz, N. (1999). *Well-being: The foundations of Hedonic Psychology*, New York: Russell Sage Foundation.
10. Kendall, P.C. and Treadwell, K. (2007). The role of self-statements as a mediator in treatment for anxiety-disordered youth. *Journal of Consulting and Clinical Psychology*, 75, 380-389.
11. Keyes, C.L.M., Shmotkin, D., and Ryff, C.D. (2002). Optimizing well-being: The empirical encounter of two traditions. *Journal of Personality and Social Psychology*, 82, 1007-1022.
12. MacLeod, A.K. and Moore, R. (2000). Positive thinking revisited: Positive cognitions, well-being and mental health. *Clinical Psychology and Psychotherapy*, 7, 1-10.
13. Maher, T.F. (1989). The Psychological Development of Pre-linguistic Deaf Infants. *Clinical Social Work Journal*, 17, 209-222
14. Maslow, A. (1968). *Toward a psychology of being* (2nd ed.). New York: Van Nostrand.
15. Ruini, C., Rafanelli, C., Conti, S., Ottolini, F., Fabbri, S., Tossani, E., Grandi, S. and Fava, G.A. (2002). Psychological well-being and residual symptoms in patients with affective disorders. I. *Rivista di Psichiatria*, 37, 171-178.
16. Ryan, R.M., & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78.
17. Ryan, R.M. and Connell, J.P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57, 749-761.
18. Ryff, C. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57, 1069-1081.
19. Ryff, C.D. and Keyes, C.L.M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69, 719-727.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 179-184

RISK ESTIMATION OF HEPATITIS (B AND C) AMONG PROVINCES OF PAKISTAN USING ONLY DESCRIPTIVE STATS FROM PMRC HOUSES HOLD SURVEY REPORT

Syed Adnan Ali[1] and **Maria Rahim**[2]
[1] Dow University of Health Sciences, Karachi Pakistan
Email: ali.adnan032@gmail.com
[2] Pakistan Journal of Rehabilitation at Ziauddin University, Karachi, Pakistan. Email: maria_rhm@yahoo.com

ABSTRACT

As it is known viral hepatitis, including hepatitis A, hepatitis B, and hepatitis C, are different diseases that affect the liver and have different hepatitis symptoms and treatments. Hepatitis type is determined by costly laboratory tests like HBsAg and Anti HCV. It is the beauty of this article that we focused on application and techniques of statistics in biostatistics, Objective was to estimate the risk of Hepatitis among provinces of Pakistan by application of logistic regression model, checked the association of Hepatitis among provinces, and to associate the quantity of injection the provinces of Pakistan by using only descriptive statistics results were reported in the survey report of PMRC of 47043 surveyed participants available at their website. We worked on secondary data. Results revealed that, out of 47043 patients, 7.5% (3547) cases diagnosed as Hepatitis (B , C or Both) this prevalence was exactly match that was reported in survey report of PMRC. Logistic regression analysis showed that, Hepatitis 1.36 times more in Sind and 1.65 times more in Punjab as compare to Baluchistan, however KPK founded 0.412 times less affected with Hepatitis as compare to Baluchistan. It was found that, Number of injections was also associated with the provinces, with $p<0.01$, in NWFP 10% (764) patients found with more than ten injections. This study showed with the help of only descriptive results, we can apply logistic regression to estimate the odds of outcome with the help of single variable, can also estimate 95% confidence interval, and can see the significance of results using chi square test of association, without holding the whole data set.

KEY WORDS

Hepatitis, Risk Estimation, Logistic Regression, Survey, PMRC, Pakistan.

INTRODUCTION

Hepatitis B is a vaccine-preventable disease.[1,2] Worldwide, the rate of Hepatitis B infection ranges from 0.1%-20%.[2] Globally the prevalence of hepatitis B is divided into three categories i.e. high (>8%), intermediate (2-7%) and low (<2%).[3,4]

In countries having low prevalence of hepatitis unprotected sex and the use of intravenous drug are the leading causes of transmission 4-5 While the countries having

high and intermediate endemicity of hepatitis infection, prenatal infection is the predominant cause.

Considering Hepatitis B 'e' antigen (HBeAg) prevalence as a major indicator of mother-to-child-transmission (MTCT), the highest rates were detected among young females from East Asia, like China, where the prenatal transmission is the most common route of infection, and the lowest in Sub-Saharan Africa.[5]

Worldwide HCV infection is also an important cause of chronic liver disease and the third-leading cause of all death from cirrhosis and hepatocellular carcinoma (HCC). Approximately 3% of the world's populations (160 million people) are currently infected with HCV, which in most cases establishes a lifelong chronic infection.[8] However, 25%-30% of infected individuals spontaneously clear the virus during acute infection.[9]

The risk of developing the infection is inversely related to the age at which infection acquire and among the newly infected, approximately 90% of prenatal infections, 25-50% of infections amongst toddlers and 5% Virus become chronic when occurs at adult life infections. [5, 7, 10]

Literature shows that, most of the studies are on risk factors of Hepatitis infection focus mainly on the individual level. Few studies have used spatial regression techniques to analyze risk factors Hepatitis infection at the population level.[11,13] Hepatitis type is determined by costly laboratory tests like HBsAg and Anti HCV. It is the beauty of this article that we focused on application and techniques of statistics in biostatistics. Estimate the risk of Hepatitis among provinces of Pakistan by application of logistic regression model, checked the association of Hepatitis among provinces, and to associate the quantity of injection the provinces of Pakistan.

OBJECTIVE OF THE STUDY

Objective was to estimate the risk of Hepatitis among provinces of Pakistan by application of logistic regression model, to see the association of Hepatitis among provinces. To See the Association of Hepatitis to the quantity of injection in the provinces of Pakistan and to accomplish all above objectives using only secondary information reported in survey report of PMRC without the primary data on each subject.

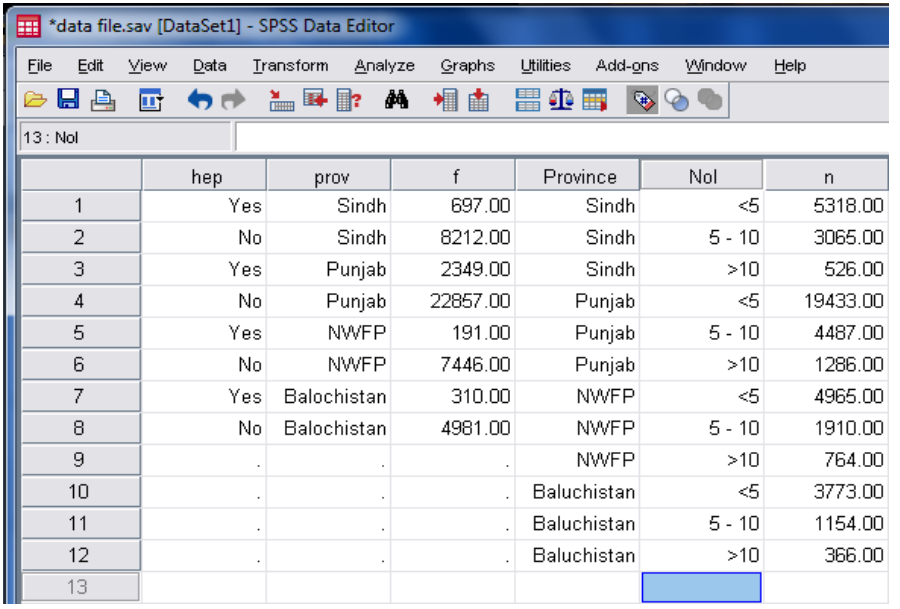
METHODOLOGY

This Survey was conducted by Pakistan Medical Research Council (PMRC) and we are using their secondary information was available on their website.[14] Cluster sampling was performed at primary stage to collect the Samples from all provinces of Pakistan. It was the beauty of this research that we focused on application and techniques of statistics and biostatistics. We extracted reported percentages for Hepatitis Prevalence, HBsAg and Anti HCV Test outcome, total used of injection in provinces reported in the survey report of PMRC on "survey on Prevalence of Hepatitis B and C in Pakistan". Using these percentages we find actual totals of each category with respect to the total cases of 47043,

$$\text{Actual Total} = \frac{\text{Reported Percentage}}{100} \times 47043$$

In statistical software SPSS, we did give the weights of these actual totals against the variable category by defining variables in our data file. We applied Pearson chi square test to see the association and logistic regression model to estimate the odds of Hepatitis among provinces.

The SPSS overview for the data setting can be seen in the following diagram.



	hep	prov	f	Province	Nol	n
1	Yes	Sindh	697.00	Sindh	<5	5318.00
2	No	Sindh	8212.00	Sindh	5 - 10	3065.00
3	Yes	Punjab	2349.00	Sindh	>10	526.00
4	No	Punjab	22857.00	Punjab	<5	19433.00
5	Yes	NWFP	191.00	Punjab	5 - 10	4487.00
6	No	NWFP	7446.00	Punjab	>10	1286.00
7	Yes	Balochistan	310.00	NWFP	<5	4965.00
8	No	Balochistan	4981.00	NWFP	5 - 10	1910.00
9	.	.	.	NWFP	>10	764.00
10	.	.	.	Baluchistan	<5	3773.00
11	.	.	.	Baluchistan	5 - 10	1154.00
12	.	.	.	Baluchistan	>10	366.00
13						

RESULTS AND DISCUSSION

This study revealed that, out of 47043 patients, 7.5% (3547) cases diagnosed as Hepatitis (B , C or Both) this prevalence was exactly matched that was reported in survey report of PMRC.[14] Punjab was the most affected province with 2349 (9.3%) of cases several studies are a sign of that, the rate of positivity for HCV is much higher in rural areas than the urban areas of Pakistan.[15] Chi square test gives the significant association between provinces and cases with hepatitis, with $p<0.001$, test of significance was not applied in the survey report.[14]

Logistic regression analysis showed that, Hepatitis risk 1.36 times more in Sind and 1.65 times more in Punjab as compare to Baluchistan, with 95% confidence intervals' (1.56, 1.18) and (1.86, 1.46) in sindh and Punjab respectively, KPK founded 0.412 times less affected with Hepatitis as compare to Baluchistan, 95% C.I was (0.49, 0.34) all odds were found significant with $p<0.001$. It was found that, Number of injections was also associated.[4-5] present study showed the association between injections and provinces, with $p<0.01$, in NWFP 10% (764) patients found with more than ten injections.

Table 1:

Province	Hepatitis B or C or Both		Total
	No	Yes	
Sind	8212	697	8909
	92.2%	7.8%	
Punjab	22857	2349	25206
	90.7%	9.3%	
KPK	7446	191	7637
	97.5%	2.5%	
Baluchistan	4981	310	5291
	94.1%	5.9%	
Total	43496	3547	47043
	92.5%	7.5%	
Chi Square test :p-value<0.01*			

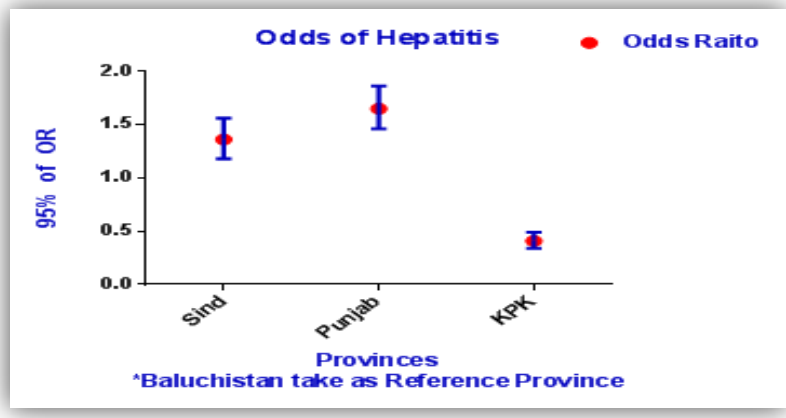


Table 2:

Province	Number of Injections			Total
	<5	5 - 10	>10	
Sind	5318	3065	526	8909
	59.7%	34.4%	5.9%	
Punjab	19433	4487	1286	25206
	77.1%	17.8%	5.1%	
KPK	4965	1910	764	7639
	65.0%	25.0%	10.0%	
Baluchistan	3773	1154	366	5293
	71.3%	21.8%	6.9%	
Total	33489	10616	2942	47047
	71.2%	22.6%	6.3%	
Chi Square test :p-value<0.01*				

CONCLUSION

Odds of Hepatitis found higher in Punjab and Sind as compare to Baluchistan, and there was significant association of provinces with Hepatitis and quantity of injections. This study showed with the help of only descriptive results, we can apply logistic regression to estimate the odds of outcome with the help of single variable, can estimate 95% confidence intervals, and can see the significance of results using chi square test of association, without holding the whole data set.

RECOMMENDATION

Government of Pakistan, conducted large scales surveys by investing million rupees funds, so, they should utilized these data by publishing to help the health sector in policy making, reporting only descriptive are not enough.

ACKNOWLEDGMENT

Thanks to PMRC for conducting survey and shared descriptive results on his website.

REFERENCES

1. Lavanchy, D. (2004). Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. *J Viralhepatol*., 11, 97-107.
2. Weinbaum, C.M., Mast, E.E. and Ward, J.W. (2009). Recommendations for identification and public health management of persons with chronic hepatitis B virus infection. *Hepatology*, 49, S35-S44.
3. Dooley, J.S., Lo, A., Burrough, A., Heathcot, J. (2011). *Sherlok's Diseases of the liver and biliary system*. In: Lok ASF. Hepatitis B.12th ed. UK: Wiley Blackwell, 370-1.
4. Maddrey, W.C. (2001). Hepatitis B--an important public health issue. *Clin. Lab.*, 47, 51-5.
5. Chang, M.H. (2007). Hepatitis B virus infection. *Semin. Fetal Neonatal. Med.* 12, 160-7.
6. Arfaoui, D., Fkih, M., Hafsa, A.E., Kaabia, N., Azzouz, M. (2010). Hepatitis B and pregnancy. *Tunis Med.*, 88, 383-9.
7. Edmunds, W.J., Medley, G.F., Nokes, D.J., O'callaghan, C.J., Whittle, H.C. and Hall, A.J. (1996). Epidemiological patterns of hepatitis B virus (HBV) in highly endemic areas. *Epidemiol. Infect.*, 117, 313-25.
8. Lavanchy, D. (2011). Evolving epidemiology of hepatitis C virus. *Clin. Microbiol. Infect.*, 17, 107-115.
9. Fierro, N., Castro-Garcia, F. and Panduro, A. (2013). Rethinking cytokine function during hepatitis A and hepatitis C infections. *Adv. Biosci. Biotechnol.*, 4, 13-18.
10. Rizzetto, M. (1999). Hepatitis B vaccination: current status. *Minerva Gastroenterol Dietol.*, 45, 199-205.
11. Zhuang, X., Wang, Y., Chow, E.P., Liang, Y. and Wilson, D.P. (2012). Risk factors associated with HIV/HCV infection among entrants in methadone maintenance treatment clinics in China: a systematic review and meta-analysis. *Drug Alcohol Depend*., 126, 286-295.
12. Yen, Y.F., Yen, M.Y., Su, L.W., Li, L.H. and Chuang, P. (2012). Prevalences and associated risk factors of HCV/HIV co-infection and HCV mono-infection among injecting drug users in a methadone maintenance treatment program in Taipei, Taiwan. *BMC Public Health*, 12, 1066. doi: 10.1186/1471-2458-12-1066.
13. Kretzschmar, M., Zhang, W., Mikolajczyk, R.T., Wang, L. and Sun, X. (2008). Regional differences in HIV prevalence among drug users in China: potential for future spread of HIV? *BMC Infect. Dis.*, 8, 108.
14. www.pmrc.org.pk/Hepatitis-Survey-Data.ppt en.wikipedia.org/wiki/Hepatitis_B

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 185-190

THE DETERMINANTS OF FDI'S IN PAKISTAN

Syed Monis Jawed
Department of Statistics, University of Karachi, Karachi, Pakistan
Email: monis_jawed@hotmail.com.

ABSTRACT

Since the emergence of 'Asian Tigers' on the global financial arena, Foreign Direct Investments (FDI's) has gained the attention of many econometricians and policy makers as they are treated as a driving force of economic growth and development, particularly for emerging economies. These studies were made under the umbrella of export led growth hypothesis (ELGH), because FDI's are included in the balance of trade of any country, hence they act as a channel of capital flows in any economy. The study is intended to be made in the context of Pakistan for time period 2004 to 2013. There are several determinants of FDI's in any economy. In this study we are going to describe FDI's via interest rate, stock market growth, exchange rate, openness and inflation. Thus, a causal model is intended to be developed here. The causal model here is developed using two approaches - Vector Autoregression (VAR) model and Structural Vector Autoregression (SVAR). VAR model, since its inception by Sims (1980) is very frequently used to explore causal relationships. VAR is basically a reduced form model, and hence it ignores all sorts of structural relationships. An alternative approach for studying relationships in macroeconomics is structural VAR (SVAR). This model is basically formed by imposing restrictions in VAR model. The results which will be obtained through VAR and SVAR approach will further be analyzed using Impulse Response Function and Variance Decomposition. The SVAR model is not yet use for explaining any causal relationship related to FDI's in the context of Pakistan, to the best of my knowledge. Another major contribution of this study is that causal relationships here are developed by two different competing models. Thus, the choice between different models is helpful in figuring out the model which would be more desirable for the case of Pakistan.

1. INTRODUCTION

Since the emergence of 'Asian Tigers' on the global financial arena, FDI has gained the attention of many econometricians and policy makers as they are treated as a driving force of economic growth and development, particularly for emerging economies. Soumare and Tchana (2011) considered FDI's as an important source for boosting developing economies and are significant for their economic growth. Prasad et al. (2003) indicated that there are a number of channels such as augmentation of domestic savings, reduction in cost of capital, transfer of technology etc. that increase the growth rate of the host economies. These studies were made under the umbrella of export led growth hypothesis (ELGH), because FDI's are included in the balance of trade of any country, hence they act as a channel of capital flows in any economy.

Foreign Direct Investments are made in the form of acquisitions, mergers and formation of multinational enterprises in the host country. The effects of such types of investments are supposed to be long lasting, as the firms which have made direct investments in the host country have stakes in it and hence they cannot drawback their investments quickly.

$$FDI_t = f(RGR_t^+, RER_t^+, OP_t^-, SMG_t^+, INF_t^+)$$

where,

SMG_t is stock market growth (embedded in stock market indices)
RGR is real growth rate
RER is real exchange rate
OP is the degree of openness
INF_t is the rate of inflation

GDP growth rate is supposed as the most appropriate macroeconomic factor for depicting the prosperity of any economy. It means that growth rate give signals to investors- local and foreign both. A rise in growth rate implies that capital stocks are growing, consequently that economy would be perceived as a suitable avenue for making investments. Prasad et al (2003) argue that capital inflows are generally attributed to higher growth rates. For measuring growth, here we are using large scale manufacturing index as a proxy for indicating the overall growth of the economy.

Exchange rate exhibits an interesting relationship with reference to FDI's. Exchange rate is taken as a relevant factor for explaining FDI's in the works of Adam and Tweneboah (2008), Ndikumana and Verick (2008), Ekeocha (2008) and Soumare and Tchana (2011). Louis and Eldomiaty (2010) have also included exchange rate as an important factor for explaining stock market developments. The rise in such rate cannot be taken as a positive development for economy. The increase in exchange rate in the host country used to increase the foreign investment because foreign investors would get better returns as verified by Karoui (2006). Thus, we are expecting a positive relationship between foreign investments and increase in exchange rate.

Openness is one of the most important determinants of foreign investment and is indicated in the works of Ndikumana and Verick (2008), Ekeocha (2008) and Soumare and Tchana (2011). Openness basically is the degree up to which extend foreign countries can participate in local economy. Several proxies are used in empirical work for modeling openness. Few works have used dummies to classify economies as open or closed. But, the methodology of Ndikumana and Verick (2008) and Ekeocha (2008) is applied here that is to use interest rate, which is the cost of usage of capital in the host country. If the interest rate increases, then it becomes difficult for firms to extend their operations. Thus rise in interest rate is expected to adversely affect the FDI flows.

Investments are an integral component of national income. Stock market activity is always conceived as one of the best proxy for assessing investments happening in any economy. De Santis and Ehling (2007) figure out stock market developments as the most important driving force for FDI's. Thus, there has to be a positive relationship existing between FDI's and capital markets.

2. DATA AND METHODOLOGY

In this study, we have considered the above mentioned factors for the period of approximately ten years that is from 2004 to 2013. For all the variables, monthly frequency has been used. All the variables are taken at log level, in order to assure that the data remains stationary.

For exploring causal relationships, econometricians frequently encountered the problems such as simultaneity. Sims (1980) developed vector autoregression (VAR) models to address the issue of simultaneity in time series analysis. The VAR model can be represented by the following equation

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \dots\dots\dots\dots + A_p y_{t-p} + u_t \quad (1)$$

In the above equation, A_i are the K x K dimension matrix, provided that y_t is a variable having k dimensions, that is, $y_t = (y_{1t}, y_{2t}, \dots, y_{kt})'$. Whereas u_t is a k dimensional white noise process, that is $u_t \sim N(0, \Sigma_u)$

Once we estimate VAR model for any system we may extend our causal analysis by applying impulse response function (IRF) and variance decomposition [Pfaff (2008)]. Various modifications of VAR had been developed over the period of time.

Another approach which has recently become popular amongst econometricians who are particularly interested in studying relationships in macroeconomics is structural VAR (SVAR). This model is basically formed by imposing restrictions in VAR model. As mentioned earlier, VAR is a reduced form model, that is, it ignores all sort of economic structures. But structures are there in various models of macroeconomics, where we encounter simultaneous relationships. The reason for having simultaneous relationships is in the theoretical frameworks involved in macroeconomics, where each factor is dependent upon several endogenous and exogenous factors. Thus, reduced form model (such as VAR) may not be that much useful in encountering such relationships. Therefore, in such conditions, its structural form can convincingly be applied. The structural form of an equivalent reduced form (that is VAR model) can be expressed by the following equation.

$$Ay_= A_1^* y_{t-1} + A_2^* y_{t-2} + \dots\dots\dots\dots + A_p^* y_{t-p} + B\varepsilon_t \quad (2)$$

where $A_i^* = A^{-1}A_i$ and $B\varepsilon_t = Au_t$. Here, $\Sigma_u = A^{-1}BB'A^{-1'}$.

Hegerty (2011) applied SVAR model to establish relationship between FDI's, FPI's and macroeconomic factors. Louis and Eldomiaty (2010) applied SVAR model to establish relationship between stock markets and macroeconomic factors. Here, in this study also, relationship between stock market and those macroeconomic factors which are related to foreign cash flows will be established.

3. EMPIRICAL RESULTS

Firstly, the stationary properties for all the variables are checked by Philips-Perron Test. All variables comes out to be stationary at level, which is an indication all the given series are cointegrated. The estimates for FDI that are been obtained through VAR are shown in Table 1 below

Table 1

	Estimates	Prob.
FDI	-0.06925961	0.4914053
SMG	-0.94836059	0.7900188
OP	-0.25649070	0.9424510
FX	3.61108592	0.8834817
RGR	-5.92323685	0.1693838
const	-0.14590301	0.2823707

The above results are indicating that the VAR estimates of our given system are not significant at any appropriate level of significance. Also the signs which we obtained for stock market growth and real growth rate are not in accordance with theory. These VAR results are further supported by IRF analysis that is presented in Fig. 1.

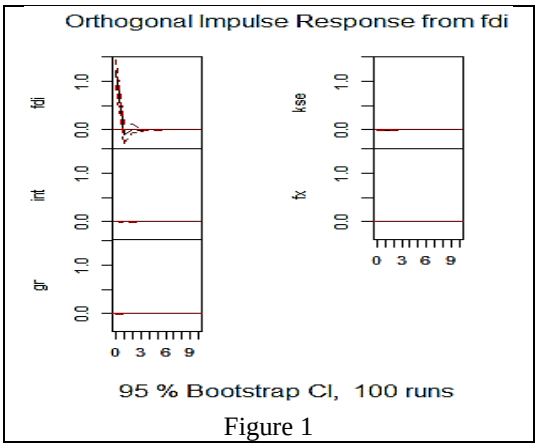


Figure 1

The IRF of the VAR estimates are indicating that no disturbance in any of the relevant describe theoretical macroeconomic variable has any impact on the FDI in case of Pakistan. However, a positive impact for FDI's on FDI's themselves are apparent in the IRP plot. The variance decomposition of error term that has been shown in Table 2

Table 2

Period	**FDI**	**SMG**	**OP**	**FX**	**RGR**
1	1.0000	0.0000	0.0000	0.0000	0.0000
2	0.9812	0.0005	0.0000	0.0001	0.0181
3	0.9803	0.0005	0.0000	0.0001	0.0186
4	0.9802	0.0006	0.0000	0.0004	0.0187
8	0.9802	0.0006	0.0000	0.0004	0.0187
12	0.9802	0.0006	0.0000	0.0004	0.0187

The above table re asserts the conclusions that are obtained from IRF, that is only FDI is explaining the variations in FDI's and no other macroeconomic factor is explaining it.

After attempting to establish a causal relationship by means of VAR, another approach for developing causal relationship on the same model is investigated through VAR. The estimates of SVAR model are mentioned below in Table 3.

Table 3

	Estimates	Prob.
FDI	1.0000000	0.01
SMG	0.0021605	0.01
OP	-0.0015027	0.01
FX	0.0004475	0.01
RGR	-0.0012506	0.01
const	0.0000	0.01

In the above table it is apparent that all the coefficients, that are been obtained are significant, and the signs are also in accordance with theory, except that for growth rate. One of the explanations of this unexpected sign was given by Prasad et al (2003), which argue that if the local infrastructure is not well developed then the benefits of FDI cannot be absorbed, and hence ELGH gets rejected. Another explanation, for this result can be due to the usage of large scale manufacturing index as a proxy of economic growth. Because in the given period, that is 2004-2013, more FDI's were diverted towards services sector, nor the manufacturing sector.

The IRF model for the SVAR estimates is given in Figure 2.

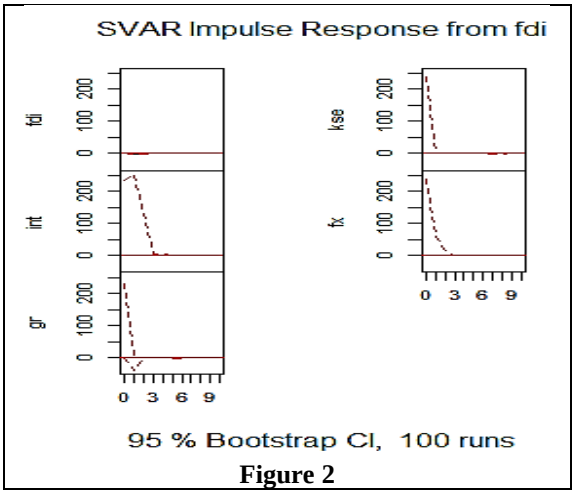


Figure 2

It can be seen in the above plot that shocks among all macroeconomic factors are positively affecting the FDI's. The largest impact can be observed for the openness which is been depicted here by means of interest rate. It implies that increase in interest rate would temporarily stimulate FDI's. The second largest impact is obvious for the shocks taking place in foreign exchange rate. Apart from it variation is obvious in case of growth rate, initially in the positive direction and later in negative direction. This behavior, seems to confirm the argument of Prasad et al. (2003). Table 4 below has variance decomposition results for the SVAR model.

Table 4

Period	**FDI**	**SMG**	**OP**	**FX**	**RGR**
1	1.0000	0.0000	0.0000	0.0000	0.0000
2	0.0200	0.0179	0.0013	0.2603	0.7006
3	0.0140	0.0144	0.0009	0.4694	0.5012
4	0.0137	0.0144	0.0010	0.4750	0.4957
8	0.0137	0.0144	0.0010	0.4764	0.4944
12	0.0137	0.0144	0.0010	0.4764	0.4944

The above table shows us the estimates forecast error variance decomposition for the SVAR model. It can be seen that largest source of variation is arising from real growth rate, followed by the variation in foreign exchange rate. Thus, the results are re asserting the conclusions about RGR and FX that are obtained through IRF, which have the largest trajectory.

4. COMMENTS AND CONCLUSIONS

This study explains the limitations of reduced form causal model for deducing conclusions in investigating macroeconomic relationships, where structures are inherent. In this case SVAR model has given results that are non-spurious. Several studies about the determinants of FDI's are been made in the context of Pakistan, but to the best of my knowledge, this study is the first attempt to establish a causal relationship on this issue via a structural model. This analysis can further be extended by adding infrastructure as a factor in the model, because it can explain why the real growth rate is not responsible for stimulating FDI's.

REFERENCES

1. Abdullah, A.M., Mansor, A.S. and Puah, H.C. (2010). Determinants of International Cash Flows: The Case of Malaysia. *Global Economy and Finance Journal*, 3, 31-43.
2. Adam, A.M. and Tweneboah, G. (2008). Foreign Direct Investment (FDI) and Stock Market Development: Ghana Evidence. *MRPA Paper*, 11261.
3. De Santis, R.A. and Ehling, P. (2007). Do International Portfolio Investors Follows Firms' Foreign Investment Decisions, *ECB Working Paper*, 815.
4. Desai, A.M., Foley, F.C. and Hines Jr., J.R. (2004). Capital Controls, Liberalization and Foreign Direct Investments. *NBER Working Paper*, 10337.
5. Ekeocha, P.C. (2008). Modelling the Long Run Determinants of Foreign Portfolio Investment in an Emerging Market: Evidence from Nigeria, *International Conference on Applied Economics-ICOAE*, 289-296.
6. Hegerty, W.S. (2011). Capital Flows to Russia, Ukraine and Belarus: Does 'Hot' Money Respond Differently to Macroeconomic Shocks. *New York Economic Review*, 42, 47-62.
7. Karoui, A. (2006). The correlation between FX rate volatility and Stock Exchange returns volatility: An emerging market overview *SSRN Working Paper*.
8. Louis, J.R. and Eldomiaty, T. (2010). How do Stock Markets Respond to Fundamental Shocks in the Case of United States? Evidence from NASDAQ and DJIA. *The Quarterly Review of Economics and Finance*, 50, 310-322.
9. Ndikumana, L. and Verick, S. (2008). The Linkages between FDI and Domestic Investment: Unraveling the Developmental Impact of Foreign Investment in Sub-Saharan Africa, *IZA Discussion Paper*, 3296.
10. Pfaff, B. (2008). VAR, SVAR and SVEC Model: Implementation within R Package vars, *Journal of Statistical Software*, 27, 1-32.
11. Prasad, S.E., Rogoff, K., Wei, S-J. and Kose, A.M. (2003). Effects of Financial Globalization on Developing Countries: Some Empirical Evidence. *IMF Occasional Paper*, 220.
12. Soumare, I. and Tchana, F.T. (2011). Causality between FDI and Financial Market Development: Evidence from Emerging Markets, *MRPA Paper*, 31328.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 191-200

A COMPARISON BETWEEN FTS MODELS AND APC MODELS

Saba Mughal and **Farah Yasmeen**
Department of Statistics, University of Karachi, Karachi, Pakistan
Email: saba-mughal@live.com, riazfarah@yahoo.com

ABSTRACT

Analysis, precise interpretation and forecasting of incidence and mortality rates are very important aspects in epidemiology and demography. The purpose of this study is to apply two different methodologies, namely the functional time series (FTS) models and age-period-cohort (APC) models on a single data set. We intend to compare the results obtained and the performances of these two approaches.

First we use FTS models on age-specific incidence rates taking age as a continuous function that is varying by time. Here we examine the age variations by using Functional Principal Component (FPC) analysis. We also forecast the entire incidence curves. Then we apply APC models on the same data where we check the age, period and cohort effects separately. We illustrate these approaches by using lung cancer incidence rates for male in Denmark, obtained from R-package "Epi" available on CRAN.

There was a rapid increase in lung cancer incidence rates since 1960, and the highest rates were found in the year 1985. After that they started level-off. A continuous increment also found in rates since 1958 birth cohort. These rates stabilized in 1905 and started decreasing since 1925. The first four basis functions of the FTS model explain 98.5%, 1.2%, 0.2%, and 0.1% of the total variation respectively. We also obtained 20-year predictions and suggested that future trends for the lung cancer incidence in male will be decrease in all ages.

KEYWORDS

Functional time series models, forecast, age-period-cohort models, age-specific incidence rates.

1. INTRODUCTION

Functional data analysis (FDA) is a branch of Statistics that analyzes data that provide information about curves, or surfaces instead of single observations that vary over time. These types of data sets are referred as functional data. In functional time series (FTS) these functional data provide information about curves which considered as unit of observations that are observed in regular intervals. The functional time series (FTS) models are first introduced by Hyndman & Ullah (2007), where these models are used in forecasting a series of functional data observed over time. The functions are observed, with error at times $t = 1, \dots, n$, and are forecasted for times $t = n + 1, \dots, n + h$. For example in demography mortality or fertility rates can be forecasted taking age as a continuous function in time t. The model is a generalization of the Lee–Carter (LC)

model commonly used in mortality and fertility forecasting. Lee & Carter (1992) proposed a methodology for modeling and extrapolating trends in mortality rates having long term periods. There are many applications involving functional time series including those studied in Erbas, Hyndman & Gertig (2007). Also these models are applied in forecasting the mortality rates of chronic diseases such as breast cancer by Yasmeen et al. (2010).

The age-period-cohort (APC) models are one of the most commonly used statistical methods that are used to analyze trends for the rates in demography and epidemiology, (e.g., Jatoi et al. 2005). These rates are observed over a long time period and for a broad age range, that have categorized by age (at death/at follow up), period (year of death /date of follow up) and cohort (date of birth). The purpose of designing age-period-cohort (APC)models is to estimate the effects of three variables age, period and cohort, (e.g., Holford, 1992).The age effect can be determined by the differing risks correspond to different age groups, the period effect indicates a change in rate which is associated with all age groups like a change in treatment that would reduce mortality rate in all ages, and the cohort effects are associated with long-term exposures or long-term habits whereby different generations are exposed to different risks. Period and cohort effects explain changes in the rates associated with time.

The major problem with age-period-cohort (APC) models is an exact linear dependency among the three factors, discussed in Holford (1992), because if we have the information of any two variables then we can easily calculate the third variable as $c = a - p$. In this model effect of these variables get non-linear on the log rate scale hence the modelling of the rates and parameterization has to be done carefully. This identifiability problem is well known, and a number of solutions have been proposed. The first paper that showed the effects by using different arbitrary constraints on the period and cohort parameters so that one could ensure the problem of identifiability proposed by Holford (1983).To deal with the identification problem, standard Clayton and Schifflers (1987a,b) procedures are used, where the mortality rate is decomposed into a common linear trend (drift), a non-linear period effect and a non-linear cohort effect, this is the classical references which showed the effect of using different arbitrary constraints on the parameters to ensure identifiability. Recently a method is proposed by Carstensen (2007) where he suggested putting no constraint on age parameters, and to choose a reference point for either period or cohort, and constrain the other to be 0 on average. This method is used by Su et al. (2013), to check the impact of a new screening program on the mortality rates in cervical cancer in Taiwan, for the two different time periods.

Above discussed methodologies are widely used in the field of demography and epidemiology, and both of the methods have their own importance in these fields. In this paper we intend to compare the results obtained and check the performances of these two approaches. To the best of our knowledge, this is the first study to show the performance of above mentioned methodologies using a single data set, so that we can check the accuracy of the fitted models and discuss their limitations.

This paper is divided into five sections. Section 1 is an introductory section, methodologies are introduced in section 2, whereas the applications of the methods are

given in section 3. In later two sections strength and limitations of the models are discussed and conclusion is drawn.

2. METHODOLOGIES

2.1 Fitting of functional time series (FTS) models:

Hyndman & Ullah (2007) proposed this methodology to fit model on log rates and obtain future curves, where each curve is considered as a unit of analysis rather than discrete observations. The main model is

$$y_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, \quad (1)$$

where $y_{t(x_i)}$ is the log incidence rates at age x in year t, where $i = 1, \dots, p.$, $f_{t(x_i)}$ is the smooth function of x_i, $\sigma_t(x_i)$ allows heteroskedasticity in x_i, and $\varepsilon_{t,i}$ is independent and identically distributed standard normal random variable.

First step is to estimate smooth functions by using non parametric smoothing in order to reduce the randomness of the observed data. Then the smooth curves are decomposed into basis functions and their corresponding coefficients by using functional principle component analysis. Following model is used for decomposition

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x)\ , \quad (2)$$

where $\mu(x)$ is a measure of location of log incidence rates over t, $\phi_k(x)$ is a set of orthogonal basis functions, $\beta_{t,k}$ is the univariate time series for each $k = 1, \dots, K$ and $e_t(x)$ is independent and identically distributed random functions with zero mean and they are assumed to be serially uncorrelated.

These time series cofficients are forecasted by univariate time series models, the h-step foreacst can be obtained by multiplying these foracasts with the estimated basis functions and summing the results.

$$\hat{y}_{n+h}(x_i) = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{n+h,k}\widehat{\phi}_k(x), \quad (3)$$

where $\hat{\mu}(x)$ and $\widehat{\phi}_k(x)$ denote the estimates of mean and basis functions, respectively and $\hat{\beta}_{n+h,k}$ is the h-step ahead forecast of $\beta_{t,k}$. Forecast variance and the prediction intervals for the h-step forecaste curves can also be determined by the proposed method of Hyndman andUllah (2007).

All options for model fitting used are available in the 'demography' package for R.

2.2 Fitting of age-period-cohort (APC) models:

Age-period-cohort (APC) models are used to investigate the age, period and cohort effects independently; however these effects cannot be analyzed simultaneously, as the model suffers a problem of identifiability. The classical models to estimate the effects are :

$$\rho_{ac} = \alpha'_a\gamma'_c, \quad (4)$$

which is the age-cohort model, here ρ_{ac} are the observed rates at age a for the cohort c, parameter α'_a show age-specific rates and parameter γ'_c show their dependence upon

cohort. If we fix one of the γ_c equal to zero then antilog of this (γ_c') is the relative risk of every single cohort that is relative to the reference cohort and α_a' show fitted age-specific rates for this reference cohort.

$$\rho_{ap} = \alpha_a' \beta_p', \tag{5}$$

which is the age-period model, here ρ_{ap} are the observed rates at age a in period p, parameter α_a' show age-specific rates and parameter β_p' show their dependence upon period.

The age-period-cohort model can be expressed as:

$$Y_{ap} = \alpha_a + \delta_p(p - p_o) + \delta_c(c - c_o), \tag{6}$$

where Y_{ap} log rates at age a in period p for the cohort $c = p - a$, α_a represent relationship between age and incidence, δ_p & δ_c represent (constant) change in log-rates between adjacent period and cohort, respectively and p_o& c_o are the reference period and cohort, respectively. By taking antilog above expression can be re-write

$$\rho_{ap} = \alpha_a' \left(\delta'_p\right)^{(p-p_o)} (\delta'_c)^{(c-c_o)}, \tag{7}$$

where δ'_p and δ'_c are the relative risks between adjacent period and cohort, respectively.

Recently Carstensen (2007) proposed a methodology consist of two mathematical steps. First step is to model the age effect and the cohort effect by choosing a suitable reference for the cohort, in this way the age-effects would be the log rates for the reference cohort and the cohort effects would be the log RRs relative to this. $\hat{f}(a)$ and $\hat{h}(c)$ are used as age and cohort effects. Then the results are used as offset variable in a model with period effects,

$$log[\lambda(a,p)] = \left[\hat{f}(a) + \hat{h}(c)\right] + g(p), \tag{8}$$

In this way the period effects would be the residuals log RRs by period. This small trick is used for model fitting (parameterization) where natural spline functions are used to smooth rates for modeling. All options for model fitting used are available in the 'Epi' package of R available on CRAN.

3. APPLICATIONS

To illustrate the calculations we consider data on lung cancer incidence for men in Denmark from 1943-1995, in the ages 40-89, where cases and person-years are tabulated by age and date of diagnosis (period) as well as date of birth (cohort) in 5-year classes, obtained from R-package "Epi". The source of the data for this package is the Danish cancer registry and statistics Denmark.

3.1 Results of functional time series (FTS) models:

Here is the plot of functional time series taking age as a continuous function. Figure 1 represents the curves in rainbow colors where earlier years are in red and recent years are in violet. Here we can see that incidence rates were low in oldest years but high in recent years for almost all ages. It is clear that from 1988 rates were relatively low as compare

to the previous years. There was a rapid increase in incidence from the age of 55 followed by a decline from 78. There was a high risk found in the ages of 70's for incidence.

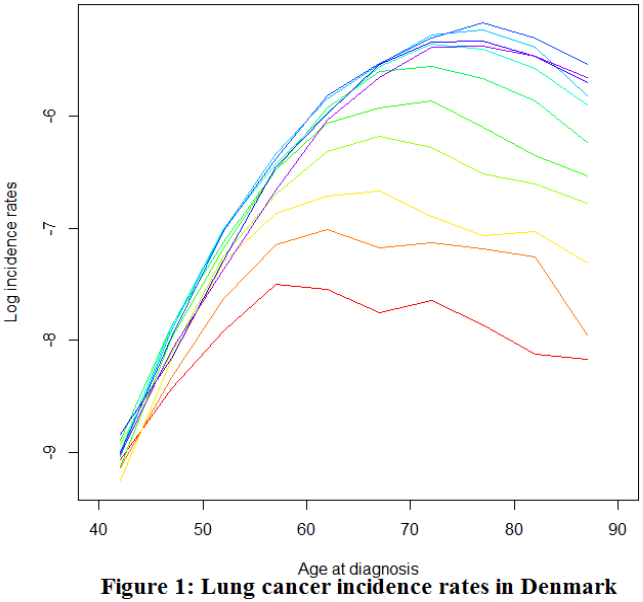


Figure 1: Lung cancer incidence rates in Denmark

Smooth rates are estimated by using weighted regression B-spline with a concavity constraint, and use functional regression model in order to obtain an adequate fit to the data. We obtain four basis functions with their coefficients which explain 98.5%, 1.2%, 0.2%, and 0.1% of the total variation.

Figure 2 depicts the overall mean function, first four estimated basis functions and their forecasted coefficients with 80% prediction intervals. Basis function models incidence rates in different age ranges as we can see $\emptyset_1(x)$ shows an increase in incidence for ages and maximum variation in men around 70, where $\emptyset_2(x)$ shows variation in a contrast those between 40 and 60 While their respective coefficients show a decline in future years. $\emptyset_3(x)$ and $\emptyset_4(x)$ are complex as they involve second and higher order effects so we don't attempt to interpret them.

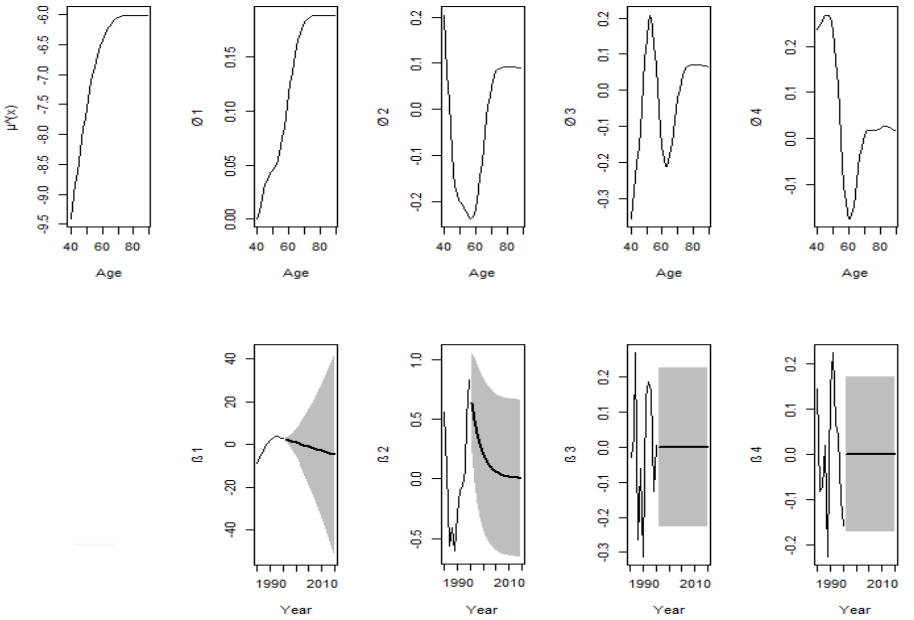


Figure 2: Basis functions and forecasts of the coefficients

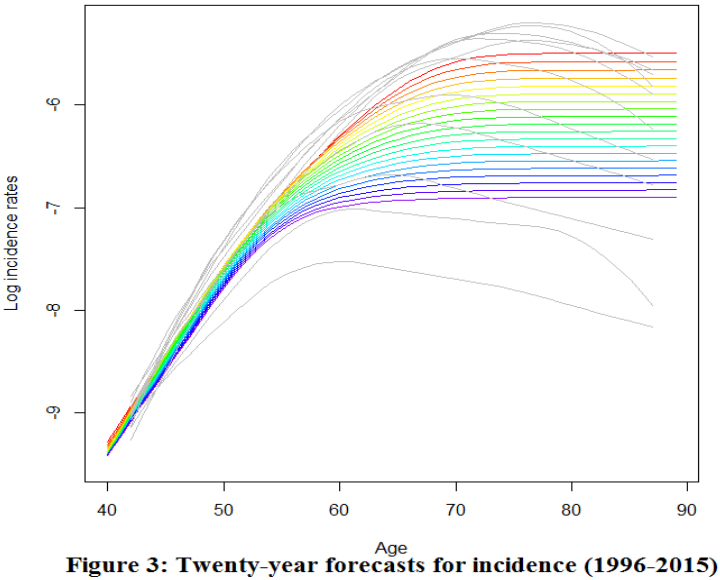


Figure 3: Twenty-year forecasts for incidence (1996-2015)

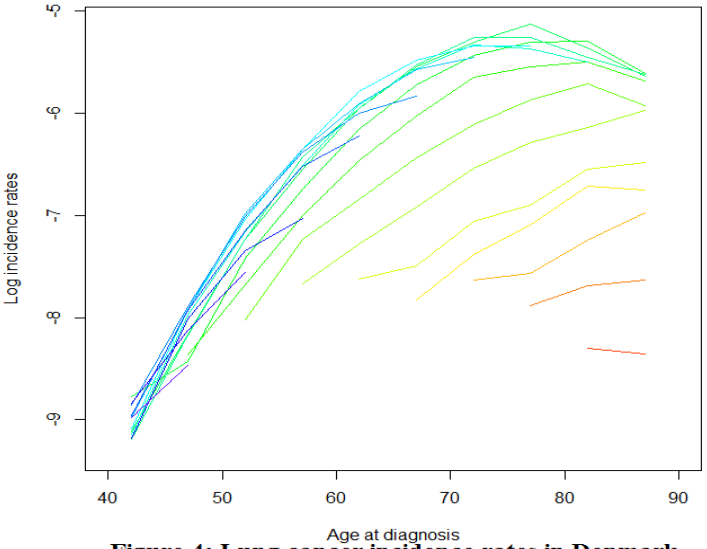


Figure 4: Lung cancer incidence rates in Denmark

Figure 3 shows the twenty-year forecasts for the lung cancer incidence for men from 1996 to 2015 by multiplying the forecasted coefficients with the estimated basis functions and summing the results. The gray curves show the smoothed data from 1943-1995. It shows that there will be a decline in the incidence rates for almost all ages. A slower decline will be present in the ages of 40 – 50 and a relatively large decline for the ages of 50 – 55. But incidence rates will decrease clearly from the age of 65 and above.

3.2 Results of age-period-cohort (APC) models:

Figure 4 shows incidence rates of lung cancer in Denmark for males, where each curve in this graph represents a longitudinal series of rates. It depicts the evolution of incidence within a birth cohort against age. Here we can see that the rates for earlier birth cohorts were relatively low but birth cohorts for the middle years show a sharp increment in the incidence rates especially for the ages of 60 to 75.

Age, period and cohort effects are estimated by using regression model based on Poisson distribution in order to attain an adequate fit for the smooth curves using natural spline.

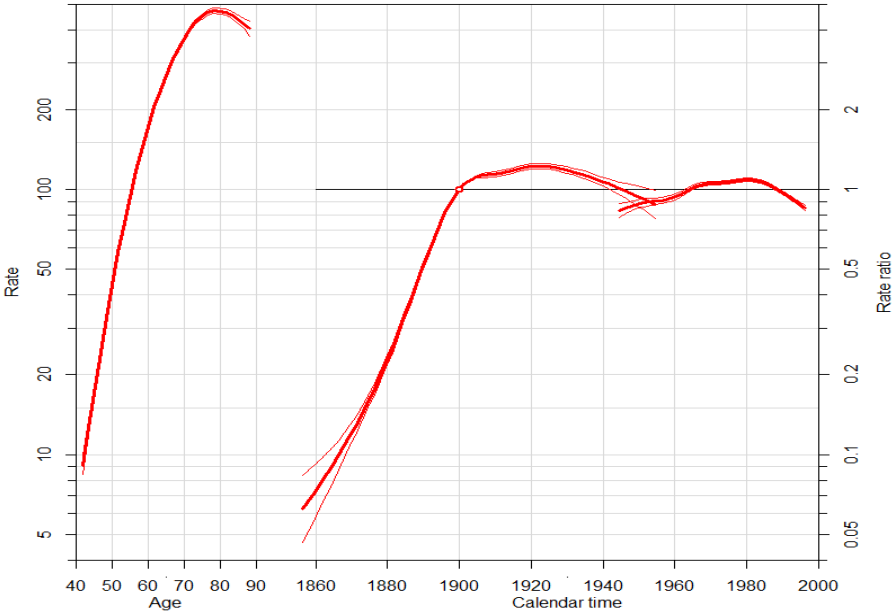


Figure 5: The age, period and cohort effects from APC model

Figure 5 shows the estimated effects of age at diagnosis on the age scale, whereas the right most curve which is the shortest one shows the period effects and the left most and longest of the two curve shows the cohort effects on the calendar time scale with 95% CIs for the Denmark age-specific incidence rates of lung cancer from 1943 – 1995. The vertical scales show rate-scale for the age-effects and the relative risk-scale for the period and cohort effects. Here we use the sequential approach by first fitting the age-cohort model and then the period model to the residuals. The reference cohort can be found at $(c_o, 1)$ where c_o= 1900, the corresponding value at the rate ratio scale is 1. The age-effect refers to the reference cohort. This curve shows an increasing trend throughout the ages. For the 70 – 80 high rates were found followed by a small decline after 80. The cohort effect can be read as the RR relative to the reference cohort. There is a clear tendency that cohorts born earlier show lower rates and higher for those who born later, relative to the reference cohort. From 1900 – 1920 the rates were stable but in the cohort of 1930 the highest rates were found after that it got level off. The curve of period effect shows increasing rates for the incidence followed by a decline after 1980.

4. STRENGTH AND LIMITATIONS:

Every model has its own advantages and limitations. Functional time series (FTS) model is the most widely used method to forecast rates; it deals with the curves where we can analyze subtle variations among years. Age function can be analyzed in detail with their associated time series coefficients as model decomposes curves via basis function extension. It also gives forecast results better than any other approach. But this model

does not incorporate the cohort term in the model. Cohort effect has a strong impact on the causes of the chronic diseases where we require prolong exposures.

On the other hand age-period-cohort (APC) models are the commonly used statistical method which incorporate all three effects and give better understanding by providing their estimates. By using the model one can also extrapolate the future rates. But the parameter identification problem is considered as a unique and unfortunate limitation of the model. For this reason the predictions also depend on the parameterization is chosen. Long term predictions are notoriously unstable because at any future point changes in the parameterization will cancel out the predictions, as the model is not uniquely parameterized.

5. CONCLUSION

We have discussed the functional time series (FTS) models that forecast a functional time series data, for better illustration we also used one of this method's applications, where we forecasted twenty-year incidence curves for the lung cancer in Denmark. We also discussed the age-period-cohort (APC) models using the same application where we fitted model sequentially in order to get estimates for the age, period and cohort curves, and found significant results for the period and cohort effects in incidence trends.

In Pakistan, these methodologies will be useful especially in the field of demography and epidemiology. By using functional time series (FTS) models fertility or mortality/incidence due to any chronic disease can be forecasted which will be very helpful in building public health policies, implementing screening programs and treatments for the disease prevention.

Whereas age-period-cohort (APC) models can be used if one aims to capture the all three effects that play important part in the medium of time trends in disease rates. As in mortality rates period effect counts a lot if sudden changes occur such as implementation of new screening program, similarly some times in chronic diseases certain generations or cohorts have greater exposure where long-term habits effects the disease rates.

REFERENCES

1. Carstensen, B. (2007). Age–period–cohort models for the Lexis diagram. *Statistics in medicine*, 26(15), 3018-3045.
2. Clayton, D. and Schifflers, E. (1987). Models for temporal variation in cancer rates. I: age–period and age–cohort models. *Statistics in medicine*, 6(4), 449-467.
3. Clayton, D. and Schifflers, E. (1987). Models for temporal variation in cancer rates. II: age–period–cohort models. *Statistics in medicine*, 6(4), 469-481.
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(2), 458-470.
5. Holford, T.R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, 311-324.
6. Holford, T.R. (1992). Analyzing the temporal effects of age, period and cohort. *Statistical Methods in Medical Research*, 1(3), 317-337.

7. 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, 4942-4956.
8. Jatoi, I., Anderson, W.F., Rao, S.R. and Devesa, S.S. (2005). Breast cancer trends among black and white women in the United States. *Journal of Clinical Oncology*, 23(31), 7836-7841.
9. Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659-671.
10. Su, S. Y., Huang, J.Y., Ho, C.C. and Liaw, Y.P. (2013). Evidence for cervical cancer mortality with screening program in Taiwan, 1981–2010: age-period-cohort model. *BMC Public Health*, 13(1), 13.
11. Yasmeen, F., Hyndman, R.J. and Erbas, B. (2010). Forecasting age-related changes in breast cancer mortality among white and black US women: A functional data approach. *Cancer Epidemiology*, 34(5), 542-549.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 201-208

ASSESSMENT OF GENETIC VARIATION ACROSS A HETEROGENEOUS POPULATION USING ANTHROPOMETRIC DATA

Nuzhat A. Akram[§], Bilal Abdullah Sheikh, Mansoor Ahmed, Aisha Abbas, Warda Ramesa Malik and **Salman Azmi**
Hamdard College of Medicine, Hamdard University, Karachi, Pakistan
[§]Corresponding author Email: smr.akram@gmail.com

ABSTRACT

Objectives: Genetic variation in a heterogeneous population can be explored through anthropometric indices. Karachi the largest cosmopolitan city of Pakistan was chosen as the model population to assess genetic variation across its inhabitant ethno-linguistic groups.

Methods: Two anthropometric indices were used to assess of variation across the ethno-linguistic groups. Each individual was investigated for his/her self-declared ethnicity. Convenient sampling method was used for this purpose. Total number of individuals were five hundred and fourteen (n= 514). ANOVA was performed across anthropometric indices to show statistically significant differences across the population strata.

Results: Analyses of variances showed significant non homogeneity of the population for anthropometric measures.

Conclusions: Significant differences across the five population strata were found which indicates that these population strata should be considered as separate entities for sampling purposes in various fields of research like clinical, forensics and genetics.

KEYWORDS

Anthropometric indices, convenient sampling, population strata, significant variation.

1. INTRODUCTION

Pakistan is considered an ethnically and linguistically complex state of South Asia (Cohen, 2004). It is comprised of five administrative units or provinces. These are Balochistan, Gilgit Baltistan, Khyber PakhtunKhwa (KPK), Punjab, and Sindh. Each of its provinces is associated with a single ethno-linguistic group namely Balochistan with Balochis, KPK with Pashtuns (or Pathans), Punjab with Punjabis, and Sindh with Sindhis. However, Karachi the capital city of Sind hosts all of these ethno-linguistic groups (Blood, 1994). The census population of Karachi is composed of five major ethno-linguistic groups each sharing more than three percent in census population (Pakistan Census Organization, 1998). These groups are designated as Punjabis, Pathans, Sindhis, Balochis and Muhajirs (Fig. 1). People from different areas of Pakistan migrated to Karachi as it is Pakistan's most industrialized city and the commercial and financial

capital. Hence, Karachi's inhabitants constitute a heterogeneous population making the city an urban agglomeration of different ethno-linguistic groups.

Molecular markers data have been used to classify human populations according to their similarities and differences (Nei and Roychoudhury, 1993; Cavalli Sforza et al., 1994; Brown, 2002; Adams, 2008). Variation across the distribution of these molecular markers have been used to reconstruct evolutionary histories using a wide variety of statistical tests including principal component analysis, cluster analysis, discriminant analysis, phylogenetic analysis etc. Very often these analyses of the molecular markers data coincided closely with groups defined by self- identified race or ethnicity (Bamshad et al. 2004; Mountain and Risch, 2004). It implies that genetic factors contribute to phenotypic variations between the races and ethnic groups.

Anthropometric measurements particularly facial anthropometry has become an important tool in anthropology and forensic investigation such as in determining age, sex and racial and ethnic groups (Williams et al., 1995; Oladipo, 2007; Kasai, 2008; Olotu, 2009). Among facial anthropometric dimensions two indices are very important. These include cephalic index and nasal index. These indices are calculated by measuring length (height) and breadth of skull and nose respectively. On the basis of cephalic index the head shapes are classified into four to six categories, that include dolichocephaly, brachycephaly, mesocephaly (or mesaticephaly) and hyperbrachycephaly (Table 1). Nasal index is also a very useful tool in anthropology in delineating racial and ethnic variations (Franciscus and Long, 2001; Porter and Olson 2003; Aung, 2000).

2. MATERIALS AND METHODS

Samples

Five hundred and fourteen individuals were selected across the five population groups through convenient sampling for anthropometric study (Fig. 2). Each population group was identified on the basis of self-declared ethnicity and taken as a stratum.

Inclusion Criteria

Only those healthy individuals were taken who were more than 20 years of age. All individuals with history of nutritional disorders or craniofacial injury or defects were excluded.

Study Design

This is a cohort study design in which each population group based on self-declared ethnicity is taken as cohort. Within each cohort individuals were selected through stratified convenient sampling.

Anthropometric Measurements

Individuals were asked to sit in a relaxed state, straight and looking forward. Head measurements were determined by spreading calliper or measuring tape. Head length was calculated from summit of glabella to inion (occipital point). Head breadth is taken across maximum horizontal diameter (across parietal protuberances).

Nasal length was measured as the distance from the nasion to the tip of the nose and nasal height was measured across the nasal bridge.

Cephalic and nasal indices

Cephalic index was calculated using the formulae *cephalic index= (head length/head breadth) × 100.* Nasal index was calculated using the formulae *nasal index= (nasal length/nasal breadth) × 100.*

Assessing variation across population strata

Mean value for the all the anthropometric measures along with the two indices across the five population groups were calculated with 95% confidence intervals (Table 2). Analysis of Variance (ANOVA) was performed for anthropometric measurements across the population groups (Table 3).

3. RESULTS

Results shows that all five population groups of the present study fall under the categories of hyper-dolichocephalic (cephalic index = 65.0-69.9) or ultra-dolichocephalic (cephalic index < 64.9). Analysis of variance (ANOVA) showed statistically significant (p value < <0.05) difference across the five population groups for all the four craniofacial measurements (skull length, skull breadth, nasal length, nasal height) and cephalic indices.

4. DISCUSSION

Population heterogeneity has been attributed partly to genetic and partly to environmental factors. Phenotypic characters like craniofacial anthropometry are likely to vary as a result of complex interaction of both these factors (Boas and Boas, 1913; Sparks and Jantz, 2002; Clerence et al., 2003). In the present study craniofacial measures of five population groups of Pakistan were reported. These anthropometric traits are considered to be specific to major human races or ethnic groups. Cephalic indices were in the range of 59.4 to 67.6. Therefore all the population groups of the present study fall under the categories of hyper- dolichocephalic (cephalic index = 65.0-69.9) or ultra-dolichocephalic (cephalic index < 64.9). Results are contradictory to those that reported Pakistani Punjabi population (Saqib et al., 2012) to be dolichocephalic (cephalic index 70.0-74.9) and neighbouring Indian populations (Salve et al., 2011; Yagain et al., 2012) to be masaticephalic (cephalic index 75.0-79.9) and brachycephalic (80.0-84.9). This contradiction can be attributed to the environmental factors that along with the genetic component also play an important role in expression of such phenotypic traits. It is worth mentioning that Boas has also reported significant difference in cephalic indices of offspring from their immigrant parents (Boas and Boas, 1913). It shows that plasticity of cranium is greatly affected by the environmental factors.

Genetic components of anthropometric trait also make it possible to infer genetic variation across population groups on the basis of phenotypic variation. It is interesting that clustering of the population based on the molecular marker data is almost the same as shown by the phenotypic characters (Mountain and Risch, 2004). In the present study all four craniofacial measures namely skull length, skull breadth, nasal length and nasal breadth showed significant differences across the five population groups. This finding was in accordance with the studies that reported variation across the same population groups on the basis of blood groups ABO and Rh and genetic markers (Akram and

Farooqi, 2014). Although it is not justified to infer population history or structure on the basis of variability of one or a few markers (Tarazona-Santos et al., 2001), however, the study will help to plan future research especially in the fields of forensics, genetics and anthropology.

ACKNOWLEDGEMENT

We would like to thank 3rd year MBBS students Sana Jameel, Komal Soomro. Saba Younus, Rasheed Achakzai, Shah Zain and many others who helped us in data collection and compilation.

REFERENCES

1. Adams, J. (2008). Human evolutionary tree. *Nature Education,* 1(1).
2. Akram, N. and Farooqi, S. (2014). A model-based 'varimax' sampling strategy for a heterogeneous population. *Ann. Hum. Biol.,* 41(2), 116-124
3. Aung, S.C., Foo, C.L. and Lee, S.T. (2000). Three dimensional lasers can assessment of the Oriental nose with new classification of oriental nasal types. *Br. J. Plast. Surg.,* 53(2), 109-116.
4. Bamshad, M., Wooding, S., Salisbury, B.A. and Stephens, J.C. (2004). Deconstructing the relationship between genetics and race. *Nat. Rev. Genet.,* 5, 598-609.
5. Blood P.R. (1994). *Pakistan: A country study.* Library of Congress. Federal Research Division. Washington DC.
6. Boas, F. and Boas, H.M. (1913). The head – forms of the Italians as influenced by heredity and environment. *Am. Anthropol.,* 15, 163-188.
7. Brown, T.A. (2002). *Genomes*, 2nd edition. Wiley-Liss, Oxford, England.
8. Cavalli Sforza, L.L., Menozzi P. and Piazza, A. (1994). *The history and geography of human genes*. Princeton University Press. New Jersey.
9. Cohen, S.P. (2004). The idea of Pakistan. The Brookings Institution Press. Washington DC.
10. Franciscus, R.G. and Long, J.C. (1991). Variation in human nasal height and breadth. *Am. J. Phys. Anthropol.,* 85(40), 419-427.
11. Gravlee, C.C., Bernard, H.R. and Leonard, W.R. (2003). Heredity, environment, and cranial form: A reanalysis of Boas's immigrant data. *Am. Anthrop.,* 105(1), 125-138.
12. Kasai, K.L.C., Richard, T. and Brown, T. (1993). Comparative study of craniofacial morphology in Japanese and Australian aboriginal population. *Hum. Biol.,* 65, 821-834.
13. Mountain, J.L. and Risch, N. (2004). Assessing genetic contributions to phenotypic differences among racial and ethnic groups. *Nature Genetics Supplement,* 36(11), S48-S53.
14. Nei, M. and Roychoudhury, A.K. (1993). Evolutionary relationship of human populations on global scale. *Mol. Biol. Evol.,* 10, 927-943.
15. Oladipo, G.S., Olabiyi, A.O., Oremosu, A.A. and Noronha, C.C. (2007). Nasal indices among major ethnic groups in Southern Nigeria. *Scientific Res. Ess.,* 2(1), 20-22.

16. Olotu, J.E., Eroje, A., Oladipo, G.S. and Ezon-Ebidor, E. (2009). Anthropometric study of the facial and nasal length of adult Igbo ethnic group in Nigeria. *The Int. J. Biol. Anthropol.*, 2(2).
17. Population Census Organization. (1998). http://www.census.gov.pk/.
18. Porter, J.P. and Olson, K.L. (2003). Analysis of the African American female nose. *Plast. Reconst. Surg.*, 111(2), 620-626.
19. Salve, V.M., Thotal, N.R. and Patibandla, A. (2011). The study of cephalic index of Andhra region. *Asian J. of Med. Sci.*, 2, 53-55.
20. Saqib, A.S., Quddus-ur-Rehman and Mujeebullah (2012). Significance of cephalic index in race determination. *Medical Forum*, 23(7), 38-40.
21. Sparks, C.S. and Jantz, R.L. (2002). A reassessment of human cranial plasticity: Boas revisited. *Proc. Natl. Acad. Sci. USA*, 99(23), 14636-14639.
22. Tarazona-Santos, E., Carvalho-Silva, D.R., Pettener, D., Luiselli, D., De Stefano, G.F., Labarga, C.M., Rickards, O., Tyler-Smith, C., Pena, S.D.J. and Fabrício, R.S. (2001). Genetic differentiation in South Amerindians is related to environmental and cultural diversity: Evidence from the Y chromosome. *Am. J. Hum. Genet.*, 68(6), 1485-1496.
23. Williams, P., Dyson, M., Dussak, J.E., Bannister, L.H., Berry, M.M., Collins, P., Ferguson, M.W.J. (1995). Skeletal System. *Gray's Anatomy,* 38th ed; Churchll Livingstone, Edinburgh, 609-612.
24. Yagain, V.K., Pai, S.R., Kalthur, S.G., Chethan, P. and Hemalatha, I. (2012). Study of cephalic index in Indian students. *Int. J. Morphol.*, 30(1), 125-129.

Table 1
Different categories of cephalic index

Types of Cephalic Index	Cephalic Index Value (Maximum skull breadth*100/Maximum skull length)
Ultradolichocephalic	< 64.9
Hyperdolichocephalic	65.0-69.9
Dolichocephalic	70.0-74.9
Mesaticephalic	75.0-79.9
Brachycephalic	80.0-84.9
Hyperbrachycephalic	85.0-89.9
Ultrabrachycephalic	> 90.0

Table 2
Descriptive statistics for all the anthropometric measures. These measures included skull length, skull breadth, nasal length, nasal height, cephalic index and nasal index

Anthropometric Measures	Population Groups	Total No. of Individuals	Mean	Standard Deviation	Standard Error	95% Confidence Interval for Mean	
						Lower Bound	Upper Bound
Skull Length	Balochi	62	33.43	2.22	0.48	32.42	34.44
	Muhajir	122	33.01	2.46	0.22	32.57	33.45
	Pathan	154	34.09	2.13	0.17	33.75	34.43
	Punjabi	112	33.38	2.25	0.21	32.96	33.80
	Sindhi	64	33.18	2.31	0.29	32.60	33.76
Skull Breadth	Balochi	62	22.55	3.49	0.76	20.96	24.14
	Muhajir	122	18.71	3.09	0.28	18.15	19.26
	Pathan	154	22.11	4.79	0.39	21.35	22.87
	Punjabi	112	21.73	5.10	0.48	20.78	22.69
	Sindhi	64	21.62	4.53	0.57	20.49	22.75
Nasal Length	Balochi	62	5.95	0.77	0.17	5.60	6.30
	Muhajir	122	5.62	0.67	0.06	5.50	5.74
	Pathan	154	6.03	0.65	0.05	5.92	6.13
	Punjabi	112	5.80	0.64	0.06	5.68	5.92
	Sindhi	64	5.66	0.57	0.07	5.51	5.80
Nasal Height	Balochi	62	5.83	0.70	0.15	5.52	6.15
	Muhajir	122	5.24	0.77	0.07	5.10	5.38
	Pathan	154	5.93	1.36	0.11	5.72	6.15
	Punjabi	112	5.58	0.84	0.08	5.42	5.74
	Sindhi	64	5.62	1.00	0.12	5.37	5.87
Cephalic Index	Balochi	62	67.62	10.56	2.30	62.81	72.43
	Muhajir	122	56.87	9.26	0.84	55.21	58.53
	Pathan	154	65.00	14.08	1.13	62.76	67.24
	Punjabi	112	65.10	14.46	1.37	62.39	67.80
	Sindhi	64	65.44	14.46	1.81	61.83	69.06
Nasal Index	Balochi	62	49.46	6.66	1.45	46.43	52.50
	Muhajir	122	47.06	7.45	0.67	45.72	48.39
	Pathan	154	49.68	12.78	1.03	47.64	51.71
	Punjabi	112	48.41	7.37	0.70	47.03	49.79
	Sindhi	64	49.89	8.52	1.06	47.76	52.01

Table 3
Analysis of variance (ANOVA) for the distribution of anthropometric measures across the five population groups. These groups are Balochi, Muhajir, Pathan, Punjabi and Sindhi. Significance level was predetermined at 0.05.

Anthropometric Measures		Sum of Squares	Degree of Freedom	Mean Square	F	P value
Skull Length	Between Groups	91.64	4	22.91	4.42	.002
	Within Groups	2424.15	510	5.18		
	Total	2515.79	514			
Skull Breadth	Between Groups	960.21	4	240.05	12.36	.000
	Within Groups	9091.33	510	19.43		
	Total	10051.54	514			
Nasal Length	Between Groups	13.80	4	3.45	8.19	.000
	Within Groups	197.07	510	0.42		
	Total	210.88	514			
Nasal Height	Between Groups	33.97	4	8.49	7.86	.000
	Within Groups	505.74	510	1.08		
	Total	539.71	514			
Cephalic Index	Between Groups	6515.97	4	1628.99	9.61	.000
	Within Groups	79316.80	510	169.48		
	Total	85832.77	514			
Nasal Index	Between Groups	587.43	4	146.86	1.59	.176
	Within Groups	43200.98	510	92.31		
	Total	43788.41	514			

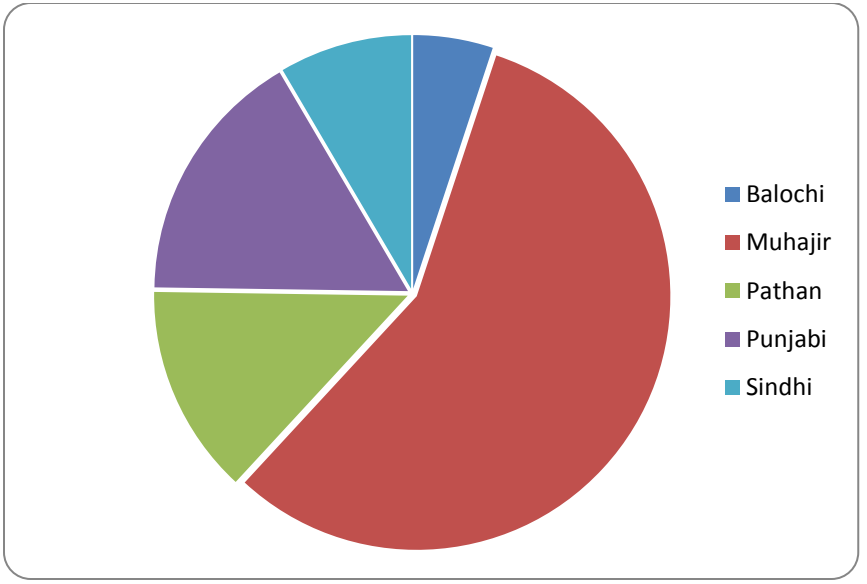


Fig. 1: Proportion of the five population groups of the present study in census population of Karachi according to census 1998. http://www.census.gov.pk/.

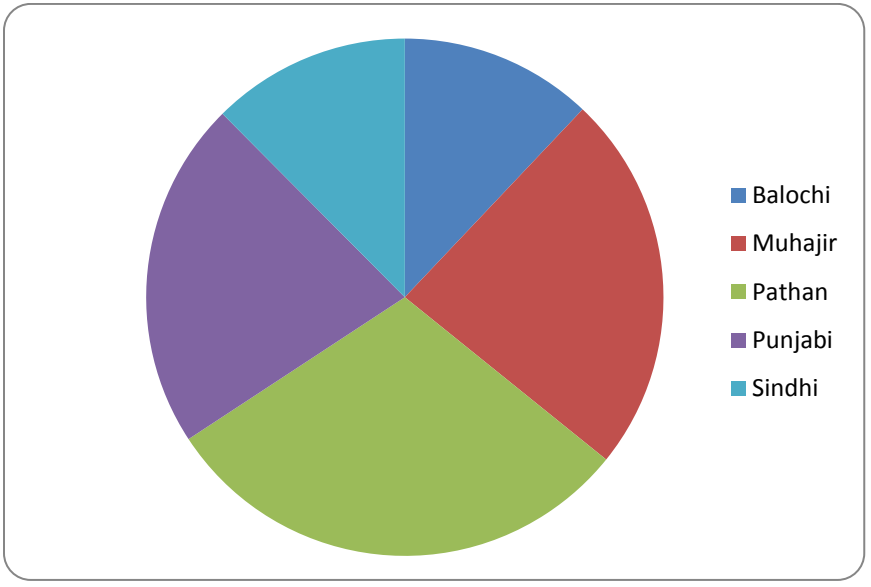


Fig. 2: Number of samples (individual) across the five population groups of the present study. Each individual was selected through stratified convenient sampling for anthropometric measures

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 209-218

A STATISTICAL ANALYSIS OF NEED AND POSSIBILITY OF ADMINISTRATION OF BACTERIOPHAGE THERAPY

Ahmad Rahman
Department of Computer Science, National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: k122026@nu.edu.pk

ABSTRACT

Most of the antimicrobial agents are predicted to become less effective, if not completely useless, in near future due to the growing development of pathogenic multi drug resistant bacteria. The therapeutic use of Bacteriophages has been employed for almost a century for treating bacterial infections in humans as well as other species, a process known as phage therapy. In this review we explore the statistical backing for the use of Bacteriophages instead of, or in combination with antibiotics to treat bacterial infection such as Tuberculosis. Three main groups were selected by convenient sampling technique - patients, doctors and researchers. A total of 210 individuals were surveyed using questionnaires and structured interviews. The results showed that the anti-microbial agents used today have a variety of side effects which could be eliminated by the use of phage therapy. The results also indicated the public's negative view of viruses. There is a need of some awareness campaign to cross the psychological barrier that people have so that they may become willing to use Bacteriophages as therapy for treating bacterial infections. Furthermore the result indicated that the use of most common antibiotics pose a great threat of being a cause of evolution of a resistance against multiple broad spectrum antibiotics in pathogenic bacteria. Some widely accepted Medline citations from Poland, USA, France and Soviet Union, related to the therapeutic use of Bacteriophages, have also been reviewed. The administration of Bacteriophages as bactericidal agents, in combination with antibiotics, may turn out to be a valuable approach.

1. INTRODUCTION

With fears of a post-antibiotic era, as old and new antibiotics alike are failing rapidly to resistant bacteria, a more dynamic solution is needed. A solution that can keep up with bacterial resistance mutation for mutation with its own tenacity for life. This solution may well have been found with bacteriophage therapy.

Prior to the discovery and widespread use of antibiotics, it was suggested that bacterial infections could be prevented and/or treated by the administration of Bacteriophages. Although the early clinical studies with Bacteriophages were not vigorously pursued in the United States and Western Europe, phages continued to be utilized in the former Soviet Union and Eastern Europe. The results of these studies were extensively published in non-English (primarily Russian, Georgian, and Polish) journals and, therefore, were not readily available to the western scientific community.

Problem with broad spectrum antibiotics are that they contribute to cancer, they can cause allergic reactions, they destroy beneficial bowel flora, they are the cause of development of resistant species of micro-organisms, they cause human immune suppression, they cause overgrowth of dangerous intestinal infections, they cause chronic fatigue syndrome, they treat effects rather than causes, they cost too much, etc. There for phage therapy has to be thought of seriously.

In this review we explore the statistical backing for the use of Bacteriophages instead of, or in combination with antibiotics to treat bacterial infection such as Tuberculosis.

2. MATERIALS AND METHODS

The research design included surveys using the structured interviews and questionnaires were done in Karachi city. Three main groups were selected by convenient sampling technique - patients, doctors and researchers. Three different questionnaires were made, one for each group, to get the answers of the questions posed by the problem and its proposed solution.

Within the group of patients, a cluster was selected including only patients that were infected by bacteria; among them 100 individuals were selected randomly for survey by structured interviewing. Patients interviewed were of different sex, belonged to different age groups, had different lifestyles and had different type of bacterial infections. Since bacterial infections are very common, there was no difficulty in finding the patients with bacterial infections. Patients were interviewed from Jinnah medical hospital, PAF Hospital Faisal Base, Trimax Hospital, T.O. Clinic, CMH Malir Cantonment and Rab Medical Center.

Patients were asked a total of 10 questions in the structured interview by which the following variables were studied:

A) Symptoms that appear after broad spectrum anti-biotic therapy
B) Percentage of incomplete course of broad spectrum anti-biotic therapies
C) Success Rate of broad spectrum anti-biotic therapies
D) Perception of antibiotics in the minds of general public

Within the group of doctors, 100 individuals were selected by convenient sampling technique for survey by structured interviewing. Doctors interviewed were of different fields, had different work experience and had different type of bacterial infections. Doctors from Liaquat National Hospital, Jinnah medical hospital, PAF Hospital Faisal Base, Trimax Hospital and Abbasi Shaheed Hospital were interviewed.

Doctors were asked a total of 10 questions in the structured interview by which the following variables were studied:

A) Causes of evolution of infectious bacteria into drug resistant bacteria
B) Bacteriophage therapy as an alternative of broad spectrum anti-biotic therapy
C) Percentage of patients complaining about the side effects of antibiotics

Within the group of researchers, a cluster was selected including only the researchers that were doing or did a research related to bacterial infections, behavior of Bacteriophages or alternative medicine; among them 10 individuals were selected randomly for survey by questionnaires. Researchers were from Aga Khan University Hospital.

Researchers were also asked a total of 10 questions in the questionnaire by which the following variables were studied:

A) Alternatives of broad spectrum anti-biotic therapies
B) Study of patients treated with phage therapy
C) Problems with phage therapy

The responses of patients, doctors and researchers were recorded by IBM SPSS Data Collection in a spread sheet which was then feed into the IBM SPSS Statistics software for analysis.

Results

Fig. 1.1 shows that most of the patients interviewed have neutral opinion about the effectiveness of antibiotics in treatment of bacterial infections among which 45% patient experience earlier mentioned side effects of antibiotics.

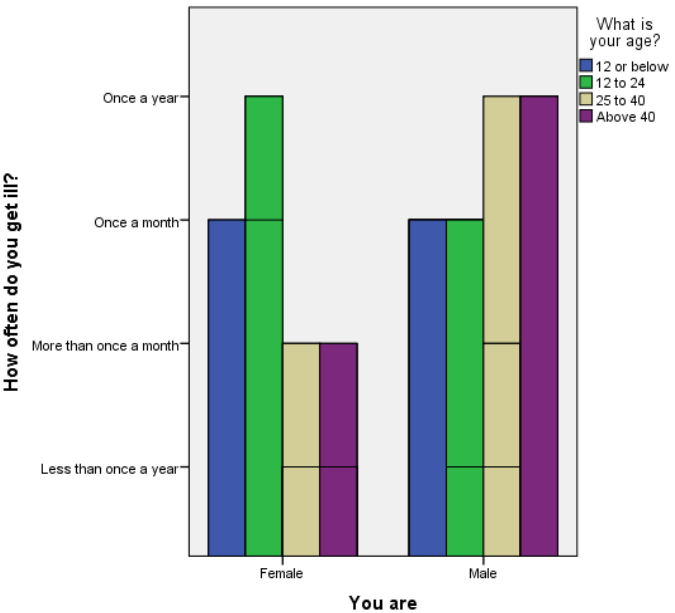


Fig. 1.1

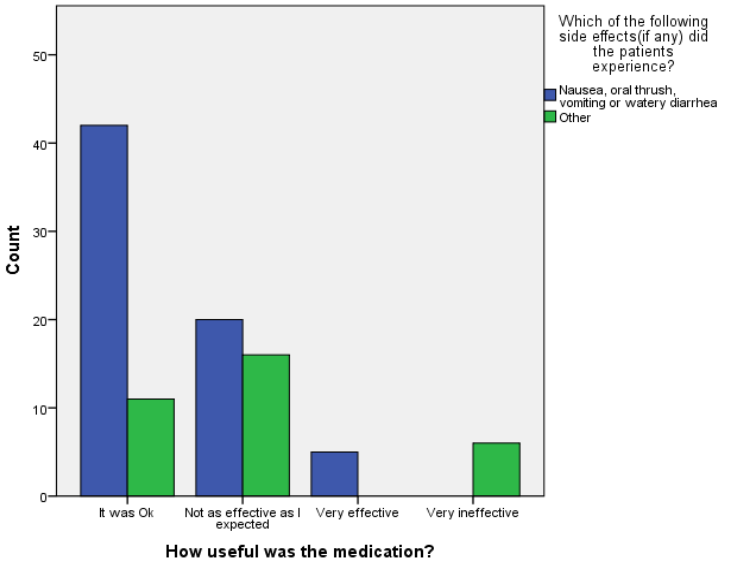


Fig. 1.2

Fig. 1.2 shows that among the patients interviewed number of males and females were even and it also shows that females of more than 25 years of age get ill very often whereas males of less than 25 years of age get ill relatively more often.

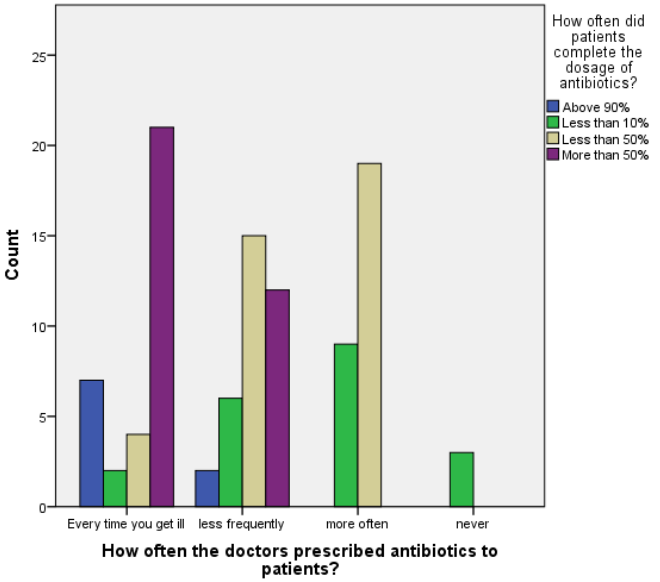


Fig. 1.3

Fig. 1.3 shows that patients that often suffer from bacterial infections among which most of them complete the antibiotic course more than 50% of times.

Fig. 2.1 shows that most of the doctors interviewed prescribe antibiotics to more than 90% of patients with bacterial infections among which in 30% of cases bacteria become resistant to antibiotics.

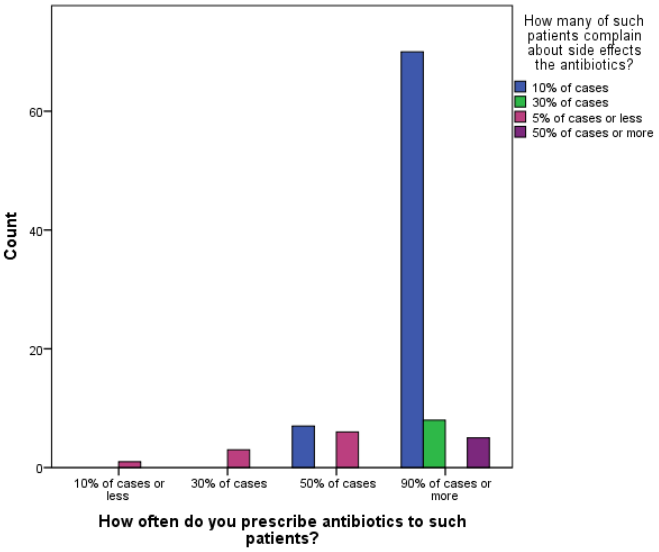


Fig. 2.1

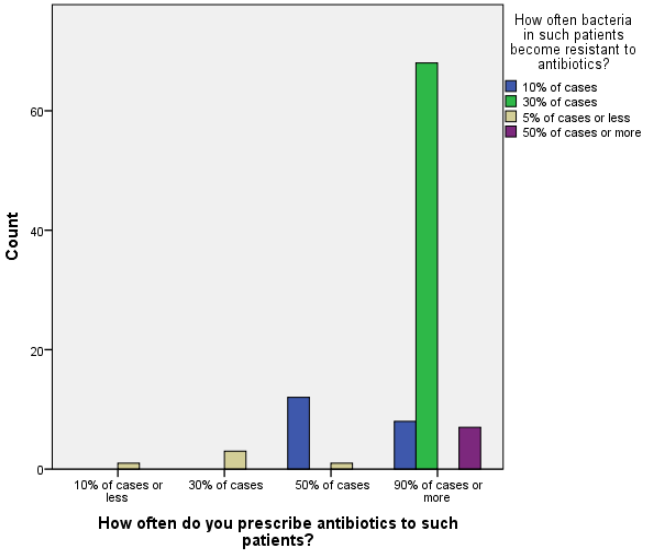


Fig. 2.2

Fig. 2.2 shows that most of the doctors interviewed prescribe antibiotics to more than 90% of patients with bacterial infections among which 10% of subjects complained about the side effects antibiotics.

3. CONCLUSION

Most of the patients with bacterial infections have faith in antibiotics. Antibiotics in most cases have proved helpful but the increasing occurrences of evolution of multi drug resistant bacteria pose a great threat. Results showed that most of the patients in Karachi don't complain to their doctor about the side effects however this doesn't mean that patients don't experience side effects. Most of the antibiotics available today have unwanted side effects such as diarrhea, vomiting, etc. The results also showed that most of the doctors prescribe antibiotics to the patients with bacterial infections among which in 30% of cases bacteria developed resistance against the antibiotic. These results indicate the need of an alternative therapy to treat bacterial infections among which phage therapy is a strong candidate. The results also indicated the public's negative view of viruses. There is a need of some awareness campaign to cross the psychological barrier that people have so that they may become willing to use Bacteriophages as therapy for treating bacterial infections.

Most of the researchers think that phage therapy has still much further to go in experimentation and research but the need of research in the field is very dire.

4. DISCUSSION

Phage therapy involves the targeted application of Bacteriophages that, upon encounter with specific pathogenic bacteria, can infect and kill them. As typically practiced phages then lyse those bacteria, releasing virion progeny that can continue the cycle, including migrating to other sites of infection anywhere in the body. The actual phage-mediated bacterial killing, however, occurs well prior to the lysis step—e.g., such as in the first minutes of infection for a phage such as phage T4 1—as the phage converts the cell into a factory for making new phages. Phages are unique among antibacterial agents in their ability to increase their numbers when in the presence of bacterial targets. Of similar importance, phages only minimally impact non-target bacteria or body tissues.

Bacterial Host Specificity

The bacterial host range of phage is generally narrower than that found in the antibiotics that have been selected for clinical applications. Most phage are specific for one species of bacteria and many are only able to lyse specific strains within a species. This limited host range can be advantageous, in principle, as phage therapy results in less harm to the normal body flora and ecology than commonly used antibiotics, which often disrupt the normal gastrointestinal flora and result in opportunistic secondary infections by organisms such as Clostridium difficile. The potential clinical disadvantages associated with the narrow host range of most phage strains is addressed through the development of a large collection of well-characterized phage for a broad range of pathogens, and methods to rapidly determine which of the phage strains in the collection will be effective for any given infection.

Advantages over Antibiotics

Phage therapy can be very effective in certain conditions and has some unique advantages over antibiotics. Bacteria also develop resistance to phages, but it is incomparably easier to develop new phage than new antibiotic. A few weeks versus years are needed to obtain new phage for new strain of resistant bacteria. As bacteria evolve resistance, the relevant phages naturally evolve alongside. When super bacterium appears, the super phage already attacks it. We just need to derive it from the same environment. Phages have special advantage for localized use, because they penetrate deeper as long as the infection is present, rather than decrease rapidly in concentration below the surface like antibiotics. The phages stop reproducing once as the specific bacteria they target are destroyed. Phages do not develop secondary resistance, which is quite often in antibiotics. With the increasing incidence of antibiotic resistant bacteria and a deficit in the development of new classes of antibiotics to counteract them, there is a need to apply phages in a range of infections.

5. LITERATURE REVIEW

United States

Interest in phages and the use of phage therapy spread quickly to the United States during the above-noted "enthusiastic period", that is, the 1920s and 1930s, which has been described in some detail in reference **29, 16** and **28**. One of the first studies of subcutaneous phage administration was carried out at the Michigan Department of Health, where Larkum **17** reported treating 208 patients with chronic furunculosis; 78% of them had no recurring infections for at least 6 months after treatment and only 3% showed no improvement. Schultz **19** and Schless **6** reported remarkable success with staphylococcal septicemia and meningitis, respectively.

Several well-known pharmaceutical companies became actively involved in efforts to produce therapeutic phage preparations in the 1930's. Eli Lilly, for example, produced sterile-filtered phage-lysates (Staphylo-lysate, Colo-lysate, Ento-lysate, Neiso-lysate) and the same preparations in a water-soluble jelly base (e.g., Staphylo-jel) for treating abscesses, purulent wounds, vaginitis, mastoiditis and respiratory infections. As reported by Straub and Applebaum**,7** E.R. Squibb and Sons and a division of Abbott Labs were also involved in commercial therapeutic phage production. Unfortunately, they found that all of them had problems with quality control, stability and establishment of efficacy.

Georgia

The Georgian experience with phage therapy centers around what is now the Eliava Institute in Tbilisi, which was founded in the 1930s by George Eliava, in association with Félix d'Hérelle. The pattern followed was one of receiving strains of pathogenic bacteria from throughout the Soviet Union, against which phages were isolated, tested and adapted. Testing included checking for virulence against target bacteria, along with host range, using appropriate panels of what currently were the most problematic strains of the bacteria in question. The result ultimately was a large operation employing 1,200 people, most of whom were involved in phage production, with a production capacity of approximately 2 tons per week. The bulk of their output was shipped to the Soviet military—for treatment of diarrhea and wounds, predominantly—while the rest was available in various forms to the general public. Georgia declared independence from the

states of the Soviet Union just before the latter's collapse in 1991. The Georgian military employed phage formulations, with soldiers in 1991 and 1992 carrying canisters of phages during battles in the disputed territory of Abkhazia, and developed special formulations for the key battlefield strains. Phage treatment was again used extensively during the battles following the 2008 dispute in the region between Georgia and Russia (Gvasalia G, personal communication).

Russia

The story of Russia's role in the development of phage therapy is long, inaccessible to the non-Russian speaker, and has not been subject to the same kind of scrutiny as has the Georgian experience. Sulakvelidze **32** and Kutter **19** and Häusler **18** provide substantially more detail of the older Soviet-era work. A review by Letarov et al. on rational phage therapy **21** gives insight into some of the more recent thought and work.

France

Human phage therapy has been practiced in France since 1919, when d'Hérelle first successfully treated several children at the Hospital des Enfants Malades in Paris who were suffering from severe dysentery, using the phage he had first isolated from the stools of soldiers he had observed at the Pasteur Institute.**20,2,3** He delayed both publication and further clinical work, though, until he had carried out very extensive studies of the properties of phages, particularly those relevant to clinical applications, with work especially in fowl typhoid and in cholera. His study of the role of phages in combating infections are laid out in a series of books, including five that have been translated from French into English **13,14,15,6,10.** His methods for the preparation of therapeutic phages were particularly well laid out in an appendix accompanying one of his later works, the first English translation of which is found elsewhere in this issue.**11** However, although d'Hérelle carried out the first human therapeutic phage trial, the first article documenting phage therapy was on research conducted in Belgium by Bruynoghe and Maisin in 1921.**4** They reported that injecting phages targeting Staphylococcus near the base of cutaneous boils (furuncles and carbuncles), in six patients, led to improvement within 48 hours: reduction in pain and swelling and some reduction in fever.

Poland

Thousands of patients have been treated with phages in Poland, particularly in association with the Hirszfeld Institute of Immunology and Experimental Therapy in Wrocńaw, which was founded in 1954. This work has been more thoroughly documented than any other in the English-language literature, mainly in the Institute's own journal in the earlier years and much of the work is available at their web site, www.aite.wroclaw.pl, and/or at www.evergreen.edu/phage. Treatment was performed by physicians from throughout the region, using phages specifically selected and prepared for each patient from the large Institute collection, and with detailed records kept. Every one of their 550 patients from 1981–1986 were included in a series of overview articles and specific discussions of particular conditions. **25,26** Reported cure rates for specific infection types ranged from 75 to 100%.**27**

7. REFERENCES

1. Bruynoghe, R. and Maisin, J. (1921). Essais de thérapeutique au moyen du bactériophage du Staphylocoque. *Compt. Rend. Soc. Biol*., 85, 1120-1121. (Fre).
2. d'Hérelle, F. (1921). *Le Bactériophage: Son Rôle dans l'Immunité*. Paris: Masson. Et. cie; (Fre).
3. d'Hérelle, F. (1922). *Der Bakteriophage und seine Bedeutung für die Immunität; nach einem erweiterten und verbesserten. Braunschweig*: F. Vieweg & Sohn. (Ger).
4. d'Hérelle, F. (1923). *Les Défenses de l'Organisme*. Paris: Flammarion. (Fre).
5. d'Hérelle, F. (1924). *Drie Voordrachten over het Verschijnsel der Bacteriophagie*. Groningen: J.B. Wolters. (Ger).
6. d'Hérelle, F. (1926). *Le Bactériophage et son Comportment*. Paris: Masson. et. Cie. (Fre).
7. d'Hérelle, F. (1929). Études sur le Choléra. *Alexandrie*: Impr. A. Serafini. (Fre).
8. d'Hérelle, F. and Smith, G.H. (1930). *The Bacteriophage and its Clinical Application*. Springfield, IL: Charles. C. Thomas. Publisher.
9. d'Hérelle, F. (1933). *Le Bactériophage et ses Applications Thérapeutiques*. Paris: Doin. (Fre).
10. d'Hérelle, F. (1938). *Le Phénomène de la Guérison dans les Maladies Infectieuses*. Paris: Masson. et. cie; (Fre).
11. d'Hérelle, F. and Smith, G.H. (1922). *The Bacteriophage: Its Role in Immunity*. Baltimore, MD: Williams. and. Wilkins. Co.,/Waverly Press.
12. d'Hérelle, F. (1924). *Immunity in Natural Infectious Disease*. Baltimore: Williams & Wilkins Co.
13. d'Hérelle, F. and Smith, G.H. (1926). *The Bacteriophage and its Behavior*. Baltimore, MD: Williams & Wilkins Co.
14. d'Hérelle, F., Malone, R.H. and Lahiri, M.N. (1930). *Studies on Asiatic Cholera*. Calcutta: Thacker. Spink & Co.
15. d'Hérelle, F. (1938). (Translated by Kuhl S, Mazure H, 2011). Preparation of therapeutic bacteriophages. Appendix 1 of *Le. Phénomè. De. La. Guérison. Dans. Les. Maladies. Infectieuses. Bacteriophage*. 2011, 1, 55-65.
16. Eaton, M.D. and Bayne-Jones, S. (1934). Bacteriophage therapy: Review of the principles and results of the use of bacteriophage in the treatment of infections (III) *J. Am. Med. Assoc*., 103, 1934-1939.
17. Fortuna, W., Miedzybrodzki, R., Weber-Dabrowska, B. and Górski, A. (2008). Bacteriophage therapy in children: facts and prospects. *Med. Sci. Monit*., 14, RA126-RA132. [PubMed]
18. Häusler, T. (2006). *Viruses vs. Superbugs: A Solution to the Antibiotic Crisis*. Macmillan.
19. Kutter, E. (2008). Phage Therapy: Bacteriophages as naturally occurring antimicrobials. In: Goldman E, Green LH, editors. *Practical Handbook of Microbiology*. Boca Raton, FL: CRC Press; 713-730.
20. Larkum, N.W. (1929). Bacteriophage treatment of Staphylococcus infections. *J. Infect. Dis*., 45, 34-41. doi: 10.1093/infdis/45.1.34. [Cross Ref]
21. Letarov, A.V., Golomidova, A.K. and Tarasyan, K.K. (2010). Ecological basis of rational phage therapy. *Acta Naturae.*, 2, 60-71. [PMC free article] [PubMed]

22. Pirnay, J.P., De, V.D., Verbeken, G., Merabishvili, M., Chanishvili, N., Vaneechoutte, M. (2011). The Phage Therapy Paradigm: Prêt-à-Porter or Sur-mesure? *Pharm Res*., 28, 934-937. doi: 10.1007/s11095-010-0313-5. [PubMed] [Cross Ref]
23. Rhoads, D.D., Wolcott, R.D., Kuskowski, M.A., Wolcott, B.M., Ward, L.S. and Sulakvelidze, A. (2009). Bacteriophage therapy of venous leg ulcers in humans: results of a phase I safety trial. *J. Wound. Care*., 18, 237-238. [PubMed]
24. Slopek, S., Weber-Dabrowska, B., Dabrowski, M. and Kucharewicz-Krukowska, A. (1987). Results of bacteriophage treatment of suppurative bacterial infections in the years 1981-1986. *Arch. Immunol. Ther. Exp.* (Warsz), 35, 569-583. [PubMed]
25. Slopek, S., Durlakova, I., Weber-Dabrowska, B., Kucharewicz-Krukowska, A., Dabrowski, M. and Bisikiewicz, R. (1983). Results of bacteriophage treatment of suppurative bacterial infections. I. General evaluation of the results. *Arch. Immunol. Ther. Exp.* (Warsz), 31, 267-291. [PubMed]
26. Slopek, S., Durlakova, I., Weber-Dabrowska, B., Kucharewicz-Krukowska, A., Dabrowski, M. and Bisikiewicz, R. (1983). Results of bacteriophage treatment of suppurative bacterial infections. II. Detailed evalulation of the results. *Arch. Immunol. Ther. Exp*. (Warsz), 31, 293-327. [PubMed]
27. Slopek, S., Durlakowa, I., Weber-Dabrowska, B., Dabrowski, M. and Kucharewicz-Krukowska, A. (1984). Results of bacteriophage treatment of suppurative bacterial infections. III. Detailed evaluation of the results obtained in further 150 cases. *Arch. Immunol. Ther. Exp.* (Warsz), 32, 317-335. [PubMed]
28. Slopek, S., Kucharewicz-Krukowska, A., Weber-Dabrowska, B. and Dabrowski M. (1985). Results of bacteriophage treatment of suppurative bacterial infections. IV. Evaluation of results obtained in 370 cases. *Arch. Immunol. Ther. Exp*. (Warsz), 33, 219-240. [PubMed]
29. Slopek, S., Kucharewicz-Krukowska, A., Weber-Dabrowska, B., Dabrowski, M. (1985). Results of bacteriophage treatment of suppurative bacterial infections. V. Evaluation of the results obtained in children. *Arch. Immunol. Ther. Exp*. (Warsz), 33, 241-259. [PubMed]
30. Slopek, S., Kucharewicz-Krukowska, A., Weber-Dabrowska, B., Dabrowski, M. (1985). Results of bacteriophage treatment of suppurative bacterial infections. VI. Analysis of treatment of suppurative staphylococcal infections. *Arch. Immunol. Ther. Exp*. (Warsz), 33, 261-273. [PubMed]
31. Smith, H.W. and Huggins, M.B. (1983). Effectiveness of phages in treating experimental Escherichia coli diarrhoea in calves, piglets and lambs. *J. Gen. Microbiol*., 129, 2659-2675. [PubMed]
32. Sulakvelidze, A. and Kutter, E. (2005). Bacteriophage therapy in humans. In: Kutter E, Sulakvelidze A, editors. *Bacteriophages: Biology and Application*. Boca. Raton, FL: CRC Press, 381-436.
33. Schultz, E.W. (1929). The bacteriophage as a therapeutic agent. *Cal. West. Med.*, 31, 5-10. [PMC free article][PubMed]
34. Schless, R.A. (1932). Staphylococcus aureus meningitis: 5 treatment with specific bacteriophage. *Am. J. Dis. Child*., 44, 813-822.
35. Krylov V. (2002). *Phagotherapy: myths and realities*. Rus. Acad. Sci. Pres., 4, 40-46.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 219-228

THE IMPACT OF AFTER SALE SERVICES ON CONSUMER BUYING BEHAVIOR IN BROADBAND INDUSTRY OF KARACHI

Muhammad Ibrahim Shamsi[1], Muhammad Haris Abdullah[2]
and **Farhal Naseem Ansari[3]**
National University of Computer & Emerging Science–FAST, Karachi, Pakistan
Email: [1]ibrahim.shamsi@nu.edu.pk
[2]k122079@nu.edu.pk
[3]k122025@nu.edu.pk

ABSTRACT

In country like Pakistan after sale services play dominant role in consumer purchase decision. After sale service sustains competitive edge, it also helps in profit enhancement and creating brand image. Broadband companies ensure high flexibility in after sale services because this flexibility is reflected in consumer decision. Customers carefully take decision while purchasing any broadband facility because they usually have to invest some amount to get the service. Purchasing decision in this industry highly depends on the services that customer receives after purchase.

The assumption of our study is that there is a positive association of after sales services on consumer buying behavior in Broadband industry of Karachi. Nine factors '1-service quality', '2-service charges', '3-consistency in promised speed', '4-frequency of internet breakdown', '5-response time', '6-consistency in promised download volume', '7-brand image', '8-convenience of usage' and '9-portability', were proposed as the influential attributes that affect the buying decision. The dependent variable 'buying behavior' is measured through the 'opinion' of those 9 attributes from consumers' point of view. The independent variable 'after sales service' is measured through the 'importance' perceived by the customer regarding those 9 attributes.

Data was collected from 203 broadband users through closed ended questionnaire. Chi square test of independence were used to test the association between the opinion and importance variables. Results show that except service quality and promised speed all the attributes were significant and good predictors of consumer buying behavior in broad band connection. It is recommended that broad band companies can enhance the buying behavior by lowering service charges, reducing frequency of internet breakdown, providing the promised download volume, improving brand Image and improving the portability and convenience of using of the internet device.

1. INTRODUCTION

The broadband internet industry had a boom after 2007 when a majority of the population started buying the internet service. There are numerous broad band companies providing their services in Karachi city. It is a challenge for any new born broad band company to differentiate their services from its rivals. Consumer buying behavior is an important dimension to be explored before entering into a new industry. The buying

decision in the broad band industry greatly depends on the facilities and services the buyer or customer receives after the purchase of the broad band connection. After sale services play central role in consumer buying assessment. These services contribute in additional buyer satisfaction since the customer and buyers relationship do not end with the purchase of goods. After sale services generate trademark image. This study is based on determining the impact of after sale services on consumer buying behavior in broad band industry of Karachi. Our belief is that there is an optimistic link between after sales services and consumer buying behavior. There are Nine key factors that may influence the buying decision of broad band customers, namely '1-service quality', '2-service charges', '3-consistency in promised speed', '4-frequency of internet breakdown', '5-response time', '6-consistency in promised download volume', '7-brand image', '8-convenience of usage' and '9-portability'. The independent variable 'after sale service' is measured by considering the importance of these 9 attributes from consumer point of view. While the dependent variable 'buying behavior' is measured by considering the opinion regarding those 9 attributes according to the broad band service the consumer is experiencing. The result of this study will help any new company to differentiate themselves in better services, and the existing internet service providers to identify the areas of their services need to improve in order to get a better market share.

2. LITERATURE REVIEW

There is a cycle in which the consumer uses a product. Many factors affect the decision of the consumer to buy a certain product. One of such factors is the after sales services, which includes facilities such as providing the promised services and making them consistent, these can easily change the customer's decision in buying a product. After sales services can greatly improve customer satisfaction as told by Clark and Armistead (1990). A research by Ehinlanwo and Zairi (1996) produces almost the same result that after sales services is very important for a firm. Long term use of a product by a certain user can only be assured by after sales services (Rosen and Surprenant 1998). They also concluded that competitive firms try to stay ahead of each other by means of improving their after sales services. People tend to buy the same product again if it can provide them with a good amount of satisfaction and also if they are really satisfied then they will recommend the product to a second user and so on and thus sales of a certain product improves. Therefore a firm's achievements become apparent only when a customer makes the decision to buy a certain product again. Levitt T (1983) emphasized that in the modern day consumer market the seller must change his focus from just making an exchange to making sure the buyer remains happy till the day he discontinues using the product. He also describes the end result of this by saying that the companies that take after sale services seriously will get a better market share. The buyer expects a lot of things from the seller, they believes that the exchange that took place should guarantee a high quality product and quick response from the company in case of a problem. Levitt T (1983) further discussed that in the world of full of complex products, people not only buy a product but buy a whole bunch of services that are usually not mentioned in the contract. He argued that for products that are difficult to manage for the user, the buyer buys expectations along with the product, and if the expectations are not met the buyer's anxieties build up after the purchase.

Zeithami, et al., (1996) discussed the behavioral consequences of service quality and offer a conceptual model that superior quality leads to superior buying behavior and leads to ongoing revenue. They explained further that customers facing no problems will have the best buying behavior, while customers facing problems will have an intermediate buying behavior if their problem was quickly resolved. Zeithami (1996) emphasized that companies should keep a routine check on customers and measure the profits that resulted due to good after sale services and focused on changing buying behavior and find out if it is due to unfavorable services. Rosen D., Surprenant C. (1998) pointed out that taking care of the customers satisfaction is the job of all the departments, if all the work force doesn't get the job done of keeping the consumer happy, the bond between the customer and the company will take a heavy blow. They further highlighted that the buyer sees the exchange as a favor bestowed upon the company, since he/she believes that the price of the product covers everything from implementation of the service or part to its maintenance.

The major role of after sales services should be to provide maximum satisfaction and comfort and facilities as customers find it attractive and it also increases sales (Rigoupoulo & et al., 2008). In a good relation, the buyer and the seller will enjoy long term benefits, specially the seller due to long term customer loyalty. Carrete L.D., Lopez S.I., Cervantes T. G. (2008) explored the factors that influence the brand loyalty and dealer loyalty of the industry. They explained loyalty or more specifically brand loyalty influences a person's dilemma of buying a product, and it is totally based on after sale services. People may even buy a bad product released by the company that they adore because of the whole concept of brand loyalty. They said that the term fan was created to signify the adoration one has for a specific brand but another word "fan boy" today signifies the loyalty of a person towards a brand. It signifies patriotism, respect and even the need to defend the brand when it's in need. After sale service is a second name of brand promise which ensures the customers that they are paying real value for the product. A reliable after sale service will increase the acceptance rate of the new product from same brand because customers feel secure in adopting new products from them (Carrete L.D. et al., 2008).

Saeed, et al., (2013) studied the factors which affect the buying decision of customers in automobile industry in Sahiwal, Pakistan. Three factors were discussed 1) Economic spare parts 2) Easy availability of spare parts and 3) Easy availability of mechanics. The factor economic spare parts have a very strong impact on the dependent variable Buying Behavior. He emphasized that customers need surety while buying a vehicle whether it is easily repairable or not and the spare parts are present in the local market at an economical price, plus the mechanics in their region have the expertise to fix the problem in car.

Looking at present times, the whole outlook towards buying a product is changing. Customers now not only purchase something but a whole bunch of services. Companies today don't have the monopolistic advantage that companies in the past may have had. Because of this customers today don't have just one company to go to when they desire something. Companies can never provide a product that will never malfunction but they can provide appropriate services to fix those malfunctions. This is exactly what after sales

service does. It acts a bridge between the company and the customer and provides assistance when the customer is in need (Ahmed, 2013).

3. MODEL FRAMEWORK

After sales service has become one of the most important techniques companies use in their path to maintain their position as the best out there. Customers need to know that they are putting their money in the right place and that is only assured by the trust that is generated by a good relationship between the company and the customer. The good relationship can be built up in broad band industry if we provide them good service quality, cheap service charges, maintaining the promised claim in internet speed, download volume, lowering the frequency of breakdown, appealing brand image and introducing portable device. The figure-1 shown below describes a theoretical model framework of our study.

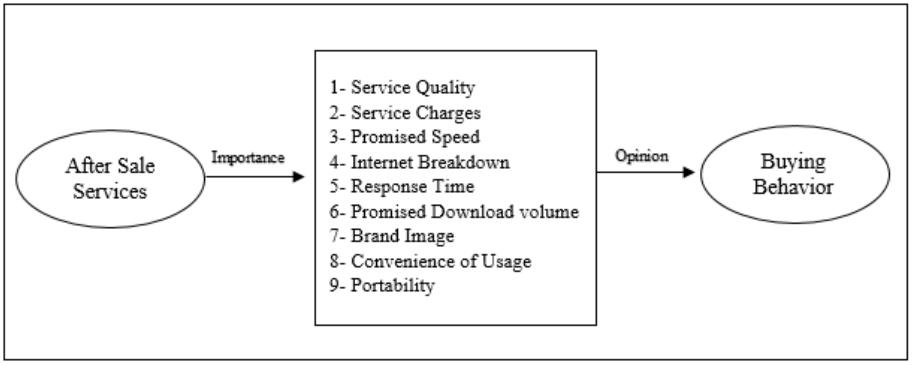


Fig. 1: Theoretical Model Framework

4. RESEARCH HYPOTHESES

Based on the previous research studies and the theoretical justification we stated the hypothesis that the after sales services in broadband industry are positively associated with consumer buying behavior. Or the behavior of broad band consumers greatly depends on the services they received after purchasing the connection. The attributes of broad band customers measured through 9 key factors (1)Services Quality, (2)Services Charges, (3)Consistency in Promised Speed, (4)Frequency of Internet Breakdown, (5)Response Time, (6)Promised Download Volume, (7)Brand Image, (8)Convenience of Usage, and (9)Portability. Each respondents were asked to rate those 9 factors according to their importance and opinion. For a single factor there were two questions. One was to rate the opinion about that factor after experiencing the broad band connection and second was to rate how important is the factor to be considered while choosing a broad band connection. The criteria for deciding a factor to be a predictor of buying behavior was made on the fact that "if opinion regarding any of the factor is independent of the importance of the same factor than that factor has no impact on buying behavior otherwise it has an impact on buying behavior".

We proposed the following 9 research hypotheses on the basis of those nine attributes.

H1: The opinion regarding broad band internet *services quality* depends on its importance while choosing a broad band connection

H2: The opinion regarding broad band internet *services charges* depends on its importance while choosing a broad band connection

H3: The opinion regarding *consistency in promised internet speed* depends on its importance while choosing a broad band connection

H4: The opinion regarding *frequent internet breakdown* depends on its importance while choosing a broad band connection

H5: The opinion regarding *response time to resolve problems* depends on its importance while choosing a broad band connection

H6: The opinion regarding *consistency in promised download volume* depends on its importance while choosing a broad band connection

H7: The opinion regarding *brand image* depends on its importance while choosing a broad band connection

H8: The opinion regarding *convenience of usage* depends on its importance while choosing a broad band connection

H9: The opinion regarding *portability* depends on its importance while choosing a broad band connection

5. RESEARCH METHODOLOGY

A questionnaire was constructed to survey the broad band consumer using 5 point likert scale questions. The question has three sections. The first section was about demographics variables of the respondent; those were 1-gender, 2-age group and 3-occupation. The second section was about the experiographics (experience of the services used) variables; those were 1-broad band service provider name, 2-service usage time, 3-speed of the internet connection and 4-download volume. The third section consist of two main questions; one was targeting the ‘opinion’ (consumer view point after experiencing any broad band service) regarding the nine factors discussed earlier, and second was targeting the ‘importance’ (how they feel to be considered while choosing a broad band connection) regarding the same factors.

The survey was conducted through convenient sampling methodology due to unknown size of population or broad band consumers. The sampling frame was Karachities. We have targeted 250 broad band customers working in different industries, student and teachers. Only 203 were responded positively, thus the response rate was 81.2% which is very much higher because the survey was done through physical meetings. Data were analyzed through the statistical package SPSS There were three demographic variables Age, Gender, Occupation and four

experiographics variables Broadband Services, Services Usage Time, Speed of Internet Connection, Download Volume. The 9 factors were asked in two ways opinion and Importance, therefore there were 9 variables for opinion and 9 variables for importance. The descriptive statistics of all demographics and experiographics variables are shown in table 1 and chart 1. The research variables (9 opinions and 9 importances) were analyzed through Chi-Square Test of Independence. Every factor was checked whether there is a dependency between the opinion and importance of that factor or not. The descriptive statistics of all nine factors and result of Chi-square test is shown in the table 2.

Table 1:
Description of Demographics Varaibles

		Count	Percent
Gender	Male	193	95.1
	Female	10	4.9
Age Group	20-29	92	45.3
	30-39	62	30.5
	40-49	29	14.3
	Above 49	20	9.9
Occupation	Student	44	21.7
	Employee	99	48.8
	Self Employed	41	20.2
	Other	19	9.4

6. RESULTS AND DISCUSSION

To test the reliability of the questionnaire all 18 likert scaled variables were analyzed using SPSS. The Cronbach's Alpha Value is 0.674 indicated that the data tool was reliable. The survey was conducted from 203 respondents, 193 out of 203 (95.1%) were male. Out of 10 female 7 were of age group 20-29 and 3 were more than 29. Nineteen male were above 49 which is only 9.8% of total sample. Out of 203, one hundred and fifty four were below 40 years, which is 75.8% of the sample size.

Forty four out of 203 were students, 99 were employees, 41 were self-employed and the remaining 19 were others. Four bar charts are shown in the chart 1, the first one shows the frequency distribution of the respondents with respect to their broadband internet connection. The users of PTCL broad band service (47.78%) is significantly higher than all the other service providers, the second highest is WITRIBE (19.21%) and the third is cable net users (14.29%). Majority of the respondents were using the broadband service between 1 and 3 years (43.84%), only 25.12% were between 3 and 5

years, 11.82% had experience of using the connection for more than 5 years and only 19.21% had been using the internet for less than 1 year.

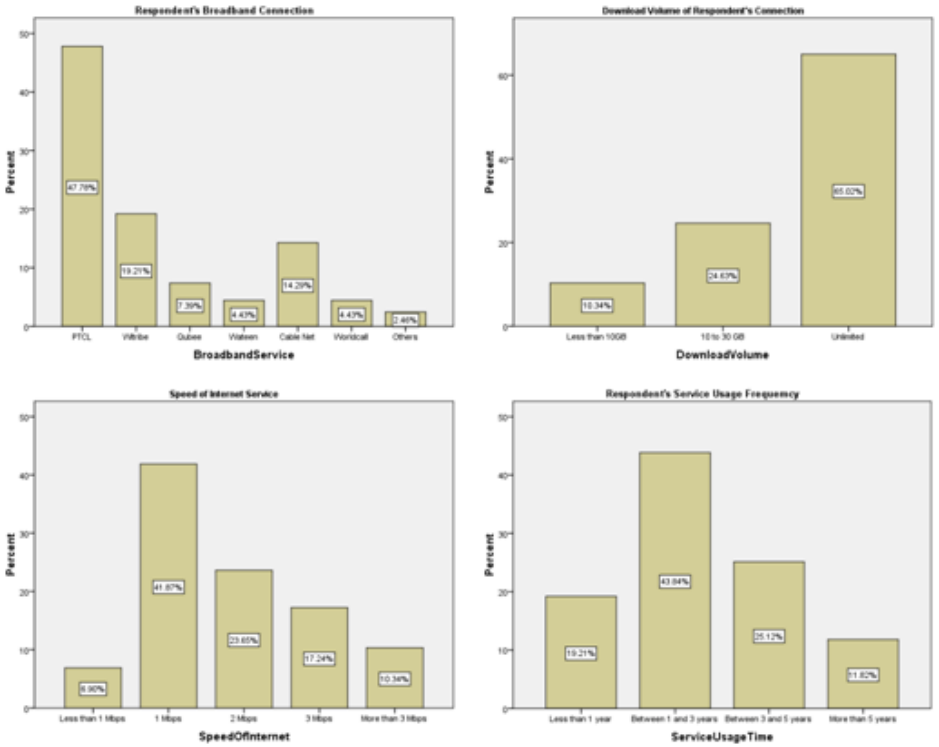


Chart 1: Bar Charts of Experiographics Variables

Speed of internet is very important while choosing the broad band connection. Majority of the respondents (41.87%) were using 1 Mbps speed connection, 23.65% were 2 Mbps , 17.24% were 3 Mbps , and the remaining had speed greater than 3 Mbps or less than 1 Mbps. Only 10.34% users had broadband connection with less than 10 GB downloads limit, 24.63% had between 10 and 30 GB, and the majority of the respondents (65.02%) were having unlimited download volume. Within unlimited download volume, 53% were PTCL customers, 18.2% were Cable net and the rest were others.

The 18 research variables regarding the 9 factors consisting of 9 opinion attributes and 9 importance attributes were analyzed using SPSS. The descriptive statistics (mean and standard deviations) of each factor were calculated and shown in table 2. The average opinion about the 9 factors of the respondent's broad band connection is in between 3 and 4. The attribute service charges (3.44335), internet breakdown (3.0936), response time (3.202) and portability (3.26108) showed the average opinion less than 3.5, it indicate that the consumer experience is below good regarding these four attributes. Respondents considered service quality (4.3399) as highly important attribute, that should be

considered while choosing the broadband connection, but chi-square sig value (0.323) indicates that there is no impact of importance of service quality on the opinion of service quality.

This leads us to the conclusion that the importance of service quality cannot change the opinion about the service quality of any broadband service provider. Although the users have a strong belief on the service quality of any service provider but opinion depends after experiencing their service. The second most important attribute considered by the respondents was promised speed (4.26601), but chi square test again showed insignificant (0.576) in this case also. This concludes that the importance of promised speed cannot predict the opinion regarding the promised speed. The third attribute response time that was considered most significant (4.22167) by the respondent, also not significant (x^2 sig value = 0.066).

Table 2:
Description of 9 Factors and Pearson Chi Square test results

	Opinion		Importance		Pearson Chi Square Value	df	Asymp. Sig. Value	Comments
	Mean	Std. Deviation	Mean	Std. Deviation				
H1: Service Quality	3.82759	.935984	4.33990	.806983	18.011	16	0.323	Not Significant
H2: Service Charges	3.44335	1.057970	4.02956	.843797	31.57	16	0.011	Significant
H3: Promised Speed	3.54680	.995797	4.26601	.788643	14.299	16	0.576	Not Significant
H4: Internet Breakdown	3.09360	1.106286	3.91133	1.165519	30.299	16	0.017	Significant
H5: Response Time	3.32020	1.121763	4.22167	.925504	25.197	16	0.066	Significant at 7%
H6: Promised Download	3.74384	.950976	4.04433	.846077	53.293	16	0.000	Significant
H7: Brand Image	3.65025	.939158	3.60099	1.016400	43.939	16	0.000	Significant
H8: Usage Convenience	3.76355	.891795	3.78325	.986199	44.988	16	0.000	Significant
H9: Portability	3.26108	1.260918	3.73892	1.036852	34.539	16	0.005	Significant

These findings lead us to read the minds of the broadband customers that they consider service quality, promised speed of internet and complain resolution time as important, but it will not influence their buying decision. Therefore the internet service providers should not focus on the above mentioned three attributes. The attribute service charges (0.011), internet breakdown (0.017), promised download volume (0.000), brand image (0.000), usage convenience (0.000) and portability (0.005) were significant and had a strong influence on the buying decision.

7. COMMENTS AND CONCLUSION

Based on the findings we can conclude that broadband service providers should focus on lowering service charges, reducing the frequency of internet breakdown, providing the promised download volume, improving brand Image and introducing portability and convenience of using of the device.

The companies should also provide the said attributes as after sale service to enhance the level of satisfaction of their customers. It was also noted that the service quality, promised speed and complain response time were very serious attributes and are not predictable. Customers consider these very important before selecting the internet connection, but their opinion is always volatile and not predictable.

REFERENCES

1. Ahmad, S. (2013). Exploring the Effectiveness of Performance-based Logistics: A Quantitative Examination. *International Journal of Operations and Logistics Management*, 2(3), 71-81
2. Armistead C. and Clark G. (1990). After-Sales Support Strategy. *OMA-UK Conference on Manufacturing Strategy, Warwick*, 5(1), 1-66.
3. Carrete L.D., Lopez S.I. and Cervantes T.G. (2008). Factors that Influence the Brand Loyalty and Dealer Loyalty of the Industry. *The Latin American Advances in Consumer Research*, 2, 235-237.
4. Ehinlanwo O.O. and Zairi, M. (1996). Best Practice in The Car After-Sales Service: An Empirical .Study of Ford, Toyota, Nissan and Fiat in Germany. *Business Process Re-engineering & Management Journal*, 2(2), 39-56.
5. Levitt T. (1983). After the Sale is Over. *Harvard Business Review*, 61(5), 87-93.
6. Rosen D. and Surprenant C. (1998). Evaluating Relationships: Are Satisfaction and Quality Enough? *International Journal of Service Industry Management*, 9(2), 103-125.
7. Rigoupoulo, I.D., Chaniotakis, I.E., Lymperopoulos, C. and Siomkos, G.I. (2008). After Service Quality as an Antecedent of Customer Satisfaction. *Managing Service Quality*, 18(5), 512-527.
8. Saeed R., Lodhi R.N., Majid, M.B., Rana, M.I., Mahmood, Z. and Ahmad, M. (2013). Impact of After Sale Service on Consumer Buying Behavior in Sahiwal Pakistan. *Management and Administrative Sciences Review*, 2(5), 555-562.
9. Zeithami, V., Berry, L. and Parasurman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, 60, 31-46.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 229-236

ASSESSMENT OF STATISTICAL APPROACHES TO MODEL LOW COUNT DATA: AN EMPIRICAL APPLICATION TO YOUTH DELINQUENCY

Taimoor Malik[1] and **Syed Arif Ali**[2]
Dow University of Health Sciences, Karachi, Pakistan
Email: [1]mtaimoor23@gmail.com
[2]arif.ali@duhs.edu.pk

ABSTRACT

Objectives: The aim of this study was to identify the risk factors associated with number of crime committed by youth (Youth Delinquency) between ages 10-17, using Ordinary Least Square (OLS), Poisson, Negative Binomial (NB) & Zero Inflated Negative Binomial (ZINB) with the aim to choose the most appropriate model for the count observed data.

Methodology: The data in the study was collected from youth whose mothers enrolled in Philadelphia Collaborative Perinatal Project (CPP). School and delinquency record (between ages 10-17) was obtained by the Centre for studies in Criminology and Criminal Law. Literature search suggest that factors associated with child delinquency can be divided into four main factors as Individual, Family, School and Peer. Therefore we included variables in the analysis accordingly.

Result: For OLS scatter plot of residuals versus estimated counts showed definite pattern of heterogeneity (non-constant variance).The likelihood-ratio (LR) test of over dispersion yields the significant p-value, which implied that the outcome variable is overdispersed. The plot of the difference between the actual probabilities and the mean predicted probabilities for each model showed that PRM has poor predictions for low counts (0-2). NBRM and ZINB both performed well, however fit statistics revealed that NBRM has provided more closed predication as compare ZINB.

Conclusion: NB modeling techniques provides much more compelling and accurate results instead of basic PRM or those available through simple linear or log-linear modeling techniques.

KEY WORDS

Count Data, Overdispersed models, Negative Binomial Regression, Zero Inflated models.

INTRODUCTION

Count Variables indicate how many times specific event has occurred. A count variable can take only zero or positive integer values as an event cannot occur a negative number of times. Three are several multidisciplinary examples including number of doctor visits (Cameron & Trivedi, 1998), number of alcohol drink consumed (Armeli, 2005), number of road traffic accidents (Chin & Quddus, 2003), number of publications

(Long, 1997) ,number of children ever born to a women (Poston, 2003).Studies that model theses counts and their association with other variables provide information leading to better understanding of the problem under study.

When the mean of the count variable is relatively high, OLS regression techniques provide reasonable results. However, when the mean of the count is low, OLS regression yields Inefficient, inconsistent and biased estimates (Cameron & Trivedi, 1998; Long, 1997). In this regard, Poison Regression Model (PRM) has been served as the basic model as it lends itself well with the nature of count data and relatively easy to understand. However Poisson Regression Model (PRM) assumes that the mean and variance of count variable are equal, a property known as equidispersion. Practically, in most of applications actual data may be overdispersed (that is variance exceeds the mean), which has similar consequences as the failure of homoscedasticity (constant variance) assumption in OLS regression. (Cameron & Trivedi, 1998; Cameron & Trivedi, 1986)

Due to this restriction, Negative Binomial regression model (NB) has been developed which has the capability to account for overdispersion. But sometimes data in hand contains far more zeros (excessive number of zeros) than are allowed (Cameron & Trivedi, 1998; Long, 1997; Hilbe, 2011). Although overdispersed count model that is NBRM can be used to model the data having zero counts. The problem of excessive zero counts can be handled by using Zero Inflated Negative Binomial Regression Model (ZINB) introduced by lambert (Lambert, 1992) and Greene

(Greene, 1994) and there may be little advantage in fitting ZINB. In ZINB data are assumed to come from a mixture of two distributions. The structural zeros from a binary distribution are mixed with the non-negative integer outcomes (including zeros) from a count distribution. Logistic regression is usually used to model the structural zeros, and negative binomial regression is used for the count outcomes. (Hilbe, 2011)

The aim of this study is to identify the risk factors associated with number of crime committed by youth between ages 10-17 (Youth Delinquency) using, PRM, NB & ZINB with the aim to choose the most appropriate model for the observed data.

METHODOLOGY

The data we used in the study was collected from youth whose mothers enrolled in Philadelphia Collaborative Perinatal Project (CPP).Upon registration, each mother was administered a series of interviews. Data recorded for each child from birth through age seven were recorded on several variables. School and delinquency record (between ages 10-17) were obtained by the Centre for studies in Criminology and Criminal Law (Denno, 1985).

Recent literature search suggest that factors associated with child delinquency can be divided into four main factors as Individual, Family, School and Peer. (Keenan, 2003; Shader, 2001)Therefore we included variables in the analysis accordingly. **Table-1**

STATISTICAL ANALYSIS

The adequacy of OLS model was examined by scatter plot of residuals versus estimated counts. To compare count models, the difference between the observed probabilities for each count and the average predicted probabilities for each model were plotted on the same graph. All regression analyses were performed using Stata version 11.1 (Stata Corp LP, 2010).

RESULTS

The results of the estimated Regression Models presented in **Table-2**. For OLS scatter plot of residuals versus estimated counts showed definite pattern of heterogeneity (non-constant variance)**,** indicating OLS regression model was not appropriate choice.

To choose between PRM and NBRM, the likelihood-ratio (LR) test of over dispersion yields the significant p-value, **chibar 2 (01) =183.72, p =<.01**, which implied that the outcome variable is overdispersed and hence NBRM is the better choice.

Additionally, the results of Voung test, a test to determine whether ZINB is statistically preferred over NBRM, yields the insignificant p-value of 0.105, supporting NRBM as compare to ZINB.

Figure-1 is the plot of the difference between the actual probabilities and the mean predicted probabilities for each model. It is apparent that PRM has poor predictions for low counts (0-2).Among NBRM and ZINB, both performed well, however investigation based on fit statistics revealed that NBRM has provided more closed predication as compare ZINB.

Gender, Income, evidence of disciplinary problems at school and two subscales of CAT were found to be the statistically significant risk factors associated with youth delinquency in all models. Additionally, one subscale of CAT was significant only in PRM and one subscale was significant both in OLS and PRM. Since, the results obtained from OLS and PRM could lead to incorrect inference and also as Figure-1 suggest that the NBRM is the most appropriate model, therefore, we focused only on NB Model.

Being a male youth, increases the expectation of committing a crime. Income at registration in CPP was negatively associated with delinquency. Similarly youth with low achievement score are more opt to have an offense record. Whereas, evidence of disciplinary problems at school, presented in the model as the number of times an individual enrolled in a disciplinary program was positively associated with youth delinquency. **(Table-2)**

Estimated Coefficients of NB and ZINB can be presented can be interpreted in the same way as in PRM. For example, results of NB regression revealed that being a female youth decreases the expected number of crimes by a factor of 0.38. Results can also be expressed in term of percentage as being a female youth decreases the expected number of crimes by 62%. **(Table-2)**

DISCUSSION

In this study we have compared modern approaches of analyzing count outcome by using empirical data of Youth delinquency. The study showed that gender, lower family income, lower achievements & disciplinary problems at school were the strongest predictors of youth delinquency.

While PRM is the good starting point due to its simplicity, it rarely explains the data in hand. ((Cameron & Trivedi, 1998; Chin, 2003; Hara, 2010) The NBRM that induces overdispersion may be sometimes more suitable. When there is high frequency of zero count in dataset, ZINB may be more appropriate. (Lambert, 1992; Greene, 1994)

In this study NBRM completely outperformed the basic PRM. On the other hand although the proportion of zero observations in dataset was approximately 25% but still NBRM was the best fitted model which suggests that when fitting a series of models without any theoretical justification, it is easy to over fit the data. The fit of count models was assessed by comparing the predicted average proportion of each count outcome to observed proportion. (Freese & Long, 2006; Long& Freese, 2001)

The results of our study suggest that serious over dispersion can results into inflated test statistics which could eventually lead to incorrect statistical inference .For example two variables two CAT variables were not significant in both PRM and NBRM attained statistical significance in PRM. Due to its structure, the Negative Binomial modeling techniques provides much more compelling and accurate results instead of basic PRM or those available through simple linear or log-linear modeling techniques.

REFERENCES

1. Armeli, S., Mohr, C., Todd, M., Maltby, N., Tennen, H., Carney, M.A. and Affleck, G. (2005). Daily evaluation of anticipated outcomes from alcohol use among college students. *Journal of Social and Clinical Psychology*, 24(6), 767-792.
2. Cameron, A.C. and Trivedi, P.K. (1998). *Regression analysis of count data* (No. 53). Cambridge University Press.
3. Cameron, A.C. and Trivedi, P.K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, 1(1), 29-53.
4. Chin, H.C. and Quddus, M.A. (2003). Modeling Count Data with Excess Zeroes An Empirical Application to Traffic Accidents. *Sociological Methods & Research*, 32(1), 90-116.
5. Denno, D.W. (1985). Sociological and human developmental explanations of crime: conflict or consensus? *Criminology*, 23(4), 711-741.
6. Freese, J. and Long, J.S. (2006). *Regression models for categorical dependent variables using Stata*. College Station.
7. Greene, W.H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. *NYU Working Paper* No. EC-94-10. Available at SSRN: http://ssrn.com/abstract=1293115.
8. Hilbe, J.M. (2011). *Negative binomial regression*. Cambridge University Press.
9. Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1-14.

10. Keenan, K., Tremblay, R.E., Coie, J.D., Herrenkohl, T.I., Loeber, R. and Petechuk, D. (2003). *Risk and protective factors of child delinquency*. US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.
11. Long, J.S. and Freese, J. (2001). Predicted probabilities for count models. *The Stata Journal*, 1(1), 51-57.
12. O'Hara, R.B. and Kotze, D.J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, 1(2), 118-122.
13. Poston Jr, D.L. and McKibben, S.L. (2003). Using zero-inflated count regression models to estimate the fertility of US women. *Journal of Modern Applied Statistical Methods*, 2(2), 10.
14. Scott Long, J. (1997). *Regression models for categorical and limited dependent variables*. Advanced quantitative techniques in the social sciences, 7. Sage Publications: Thousand Oaks, CA.
15. Shader, M. (2001). *Risk factors for delinquency: An overview*. US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Table-1
Description of Variables used in the analyses

Variable	Description
Gender	0= Male 1= Female
Income	Family income at the time of Registration in CPP
Remedial Disciplinary Codes in School	Number of times an individual enrolled in a disciplinary program during High School
Subscales of CAT[1]	
CAT_1	Vocabulary Score
CAT_2	Comprehension Score
CAT_3	Reading Score
CAT_4	Computation Score
CAT_5	Concept& Problem Score
CAT_6	Total Maths Score
CAT_7	Mechanics Score
Subscales of WISC[2]	
WISC-1	Information Score
WISC-2	Comprehension Score
WISC-3	Vocabulary Score
WISC-4	Digit Span Score
WISC-5	Block Design Score
WISC-6	Coding Performance Score
WISC-7	Verbal I.Q SCORE
WISC-8	Performance I.Q Score

Table 2
Estimated Models to analyze Youth Delinquency

Regression Model	Poisson				Negative Binomial				Zero Inflated Negative Binomial			
Variable	Coeff	SE	e^b	P-Value	Coeff	SE	e^b	P-Value	Coeff	SE	e^b	P-Value
Gender												
Female	-0.90944	0.1056669	0.4027	<0.01	-0.9658702	0.1822669	0.3807	<0.01	-0.939789	0.1799274	0.3907	<0.01
Family Income	-0.0001202	0.0000259	0.9999	<0.01	-0.0001161	0.0000475	0.9999	0.014	-0.0001213	0.0000461	0.9999	0.008
Remedial Disciplinary Codes in School	0.2475586	0.0235893	1.2809	<0.01	0.3250835	0.1318478	1.3841	0.014	0.2728055	0.1117078	1.3136	0.015
Retarded	-0.0866352	0.0504951	0.917	0.086	-0.0650381	0.0855598	0.937	0.447	-0.0520493	0.0938061	0.9493	0.579
Subscales of CAT[1]												
Vocabulary Score	-0.0017094	0.0029647	0.998	0.564	-0.0048391	0.0057505	0.9952	0.4	-0.0042933	0.005579	0.9957	0.442
Comprehension Score	-0.0094624	0.0038414	0.991	0.014	-0.0084006	0.0072804	0.9916	0.249	-0.0050991	0.0070953	0.9949	0.472
Reading Score	-0.0222458	0.0040609	0.978	<0.01	-0.0215281	0.0068799	0.9787	0.0020	-0.024581	0.0066807	0.9757	<0.01
Computation Score	0.0103111	0.0028505	1.010	<0.01	0.012072	0.0058244	1.0121	0.038	0.0117496	0.0056746	1.0118	0.038
Concept& Problem Score	-0.0054292	0.0031216	0.9946	0.082	-0.005009	0.0055582	0.995	0.367	-0.0040867	0.0053547	0.9959	0.445
Total Maths Score	0.0174498	0.0035963	1.0176	<0.01	0.0106158	0.0069551	1.0107	0.127	0.0105066	0.0068229	1.0106	0.124
Mechanics Score	-0.0047984	0.0044883	0.9952	0.285	-0.0003341	0.0079657	0.9997	0.967	0.0008428	0.0077938	1.0008	0.914
Subscales of WISC[2]												
Information Score	-0.1550568	0.1535085	0.8564	0.312	0.0500873	0.3296517	1.0514	0.879	-0.0383882	0.3266869	0.99	0.906
Comprehension Score	-0.2002367	0.1540842	0.8185	0.194	-0.0177752	0.3293874	0.9824	0.957	-0.0818141	0.3264187	0.99	0.802
Vocabulary Score	-0.1788257	0.1551708	0.8363	0.249	0.0073434	0.3310648	1.0074	0.982	-0.2109335	0.3293144	0.99	0.522
Digit Span Score	-0.2133848	0.1547352	0.8078	0.168	-0.0187186	0.3312071	0.9815	0.955	-0.0853319	0.3279963	0.99	0.795
Block Design Score	-0.0054404	0.029305	0.9946	0.853	-0.0299404	0.0614853	0.9705	0.626	-0.0184781	0.0597649	0.99	0.757
Coding Performance Score	0.0292024	0.0256732	1.0296	0.255	0.0263197	0.0509468	1.0267	0.605	0.0341352	0.0492585	0.99	0.488
Verbal I.Q SCORE	0.11941	0.0976037	1.1268	0.221	-0.0006731	0.2083913	0.9993	0.997	0.0426289	0.2062347	0.99	0.836
Performance I.Q Score	-0.0142703	0.0083109	0.9858	0.086	-0.0067547	0.0166896	0.9933	0.686	-0.0100017	0.016097	0.99	0.534
1. California Achievement Test												
2. Wechsler Intelligence Scale for Children												

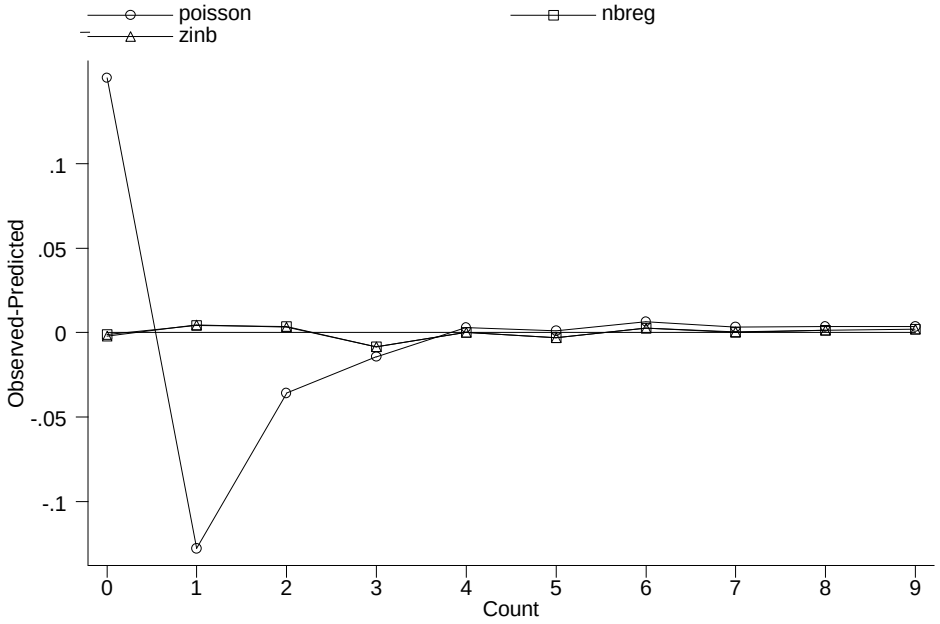


Fig. 1: Difference Between the Observed Probabilities and the Average Predicted Probabilities

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 237-248

THE IMPACT OF INSTITUTIONAL DEVELOPMENT ON SHORT-TERM RETURN PREDICTABILITY IN AN EMERGING MARKET OF PAKISTAN

Abida Hafeez[1] and **Mohammad Nishat**[2]
[1] Applied Economics Research Centre, University of Karachi
Karachi, Pakistan. Email: abida_aerc@yahoo.com
[2] Institute of Business Administration, Karachi, Pakistan.
Email: mnishat@iba.edu.pk

ABSTRACT

This research paper investigates short-term stock return predictability during 2002 to 2012 in the Karachi Stock Market, an emerging stock of Pakistan. This paper employs non-parametric stock market efficiency tests consisting of run test and autocorrelation function. The results from run test and autocorrelation function on daily and weekly return data of KSE 100 index reject the null hypothesis of no autocorrelation. So the presence of autocorrelation leads us to draw a conclusion that this emerging stock market is inefficient and implies institutional imperfections.The results are similar to those of many emerging markets and indicate the presence of autocorrelation on the Karachi stock exchange Institutional development and financial liberalisation in the emerging stock market of Pakistan have not been succeeded to overcome anomalous behaviour of the market. Moreover, this leads to predict returns and to reject random walk hypothesis (RWH). However, financial liberalization plays an important role to distort the return pattern thereby causing market rationality.

KEYWORD

Return Predictability, Emerging Market, Autocorrelation, Institutional Development, Efficiency, Random Walk Hypothesis (RWH).

INTRODUCTION

One of the researchable tasks in finance is return predictability through the channel of efficient market Hypothesis (EMH). Others include whether the financial market is efficient in allocating or using economic resources and information or not along with volatility, speculation, and anomalies which are also related to market efficiency and all are interdependent (Islam and Oh, 2003; Mills, 1999; Cuthbertson, 1996). In the 1960s and early 1970s, the focus was to measure the extent to which successive changes in stock prices were independent of each other or whether stock prices followed a random walk. The early tests to answer this question were conducted by Fama (1965) and Samuelson (1965) in which they concluded that most of the evidence seems to have been consistent with the efficient market hypothesis (EMH). Stock prices followed a random walk model and the predictable variations in equity returns, if any, were found to be statistically insignificant. Other studies in the US with similar findings included those of Sharpe (1966), Friend *et al*. (1970) and Williamson (1972). Throughout the 1980s, EMH has provided the theoretical basis for much of the research, and most empirical studies during these years focused on

predicting prices from historical data, while also attempting to produce forecasts based on variables such as P/E ratios (Campbell and Shiller 1987), dividend yield (Fama and French 1989), term structure variables (Harvey 1991) and announcement of various events, i.e. earnings, stock splits, capital expenditure, divestitures, and takeovers (Jensen and Ruback 1983; McConnell and Muscarella 1985; Kettel 2001).

The issue of EMH in relation to stock prices is fundamental for an investigation of the characteristics of the Karachi stock market. The results from testing the EMH can assist in the identification of these factors, which could be seen as the influence of anomalies (Nassir and Mohammad 1987; Ho 1990; Berument and Kayimaz 2001), insider trading and asymmetric information (Jaffe 1974; Jegadeesh and Titman 1993), stock splits (Ikenberry *et al*., 1996), dividend initiations and omissions (Michaely *et al*., 1995), etc. Efficiency in developed financial markets was in the main focus for empirical analysis (Kendall, 1953; Fama, 1970) but the focus of researchers has dramatically repositioned towards emerging markets (Classens *et al*., 1995; Chang and Ting, 2000; Abeysekera, 2001; and Kim and Shamsuddin, 2008). It is long-familiar that equities from emerging markets are characterized differently as compared to equities from developed markets.

The basic motive for this study is apparent from the fact that there is lack of research work on asset return predictability in the Karachi stock exchange (KSE). Investigation of asset return predictability[1] is crucial in detecting anomalies and in probing the extent of regulatory policy changes and institutional development to overcome the anomalous behavior of the stock market and in achieving market efficiency as well as changes in the magnitude of portfolio investment. The null hypothesis for short term predictability is that whether the successive returns are independent during the reform periods.

The present study used non-parametric stock market efficiency tests such as run test and autocorrelation function (ACF) to determine a conclusion about efficient market hypothesis (EMH) in an emerging stock market of Pakistan. Non-parametric testing of market efficiency is based on the premise of no arbitrage opportunities, i.e., the opportunities for earning unusual returns do not exist (Fama 1970; Jensen 1978). The non-parametric methods like run test and autocorrelation function being pursued in this paper are aiming consistent statistical characteristics of stock prices and returns profile, using few interlinkages with a specific model of asset pricing. If the Karachi stock exchange (KSE) is efficient, the stock prices would correctly and fully reflect all available information. As a result no arbitrage opportunities would exist. However, the rejection of the null hypothesis of efficiency would reject market efficiency for the Karachi stock market. The implication of efficiency, broadly speaking, is that stock returns always reflect their intrinsic worth and can be taken at their face value.

This research paper is organized as follows; Section 2 reviews the previous studies about the market efficiency hypothesis (EMH) and short term return predictability using non-parametric tests. Section 3 sheds light on conceptual framework. Section 4 demonstrates data and methodology. The empirical findings and interpretation are shown in section 5. Section 6 provides concluding remarks.

[1] Investors 'expectations about the present value of future cash flows

2. LITERATURE REVIEW

The various research studies employed non-parametric tests; run test and autocorrelation function to investigate market efficiency. Sharma and Kennedy (1977) probe the behaviour of stock indices of three markets such as the Bombay stock exchange, London and New York stock markets over the period of 1963-73 employing run test and spectral techniques. The results support the random movement of stock indices for all the three markets. The conclusion shows the random walk behaviour at the BSE and is efficient in weak-form.

Ramachandran (1986) examines the weak-form of EMH using weekly stock prices from 1976 to 1981. He applies filter rule tests along with correlation and run tests and finds the weak-form of efficient market hypothesis (EMH).

Yalawar (1988) investigates the Bombay stock exchange from the efficiency point of view. He analyses the monthly stock prices listed at the BSE from 1963 to 1982. This study uses the non-parametric tests along with Spearman's rank correlation test and finds the random behaviour of security prices.

Seiler and Walter (1997) investigate random behaviour of stock returns listed at the New York stock exchange (NYSE) from 1885 to 1962. The results show that fluctuations in stock returns are totally random.

Keasey and Mobarek (2000) investigate the weak-form efficiency of Dhaka stock exchange (DSE) of Bangladesh from 1988 to 1997 by using both parametric and non-parametric tests. The data of daily price indices has been used in this study to analyse the efficiency of the Dhaka stock exchange. The observation shows that DSE is not efficient in weak-form.

Pandey (2003) argues that market efficiency has an impact on the investment strategy of a stock investor. This research study tests the efficiency and existence of the random walk hypothesis in the context of Indian stock market from January 1996 to June 2002, using the run test and the autocorrelation analysis. The study has indicated that the Indian stock market is inefficient. Autocorrelation analysis and runs test enable to conclude that the series of stock indices do not show random behaviour. The autocorrelation function shows that the pattern of stock returns does not support the random walk hypothesis in the context of the Indian stock exchange.

Islam *et al.* (2005) propose two non-parametric tests to find out weak-form efficiency for an emerging market of Thai from 1992-2001. The study employs run test and autocorrelation analysis to make a critical decision about the strength of EMH in the context of emerging stock markets. The finding of this study indicates that there exists an autocorrelation in stock returns of the emerging market of Thai. The prevalent inefficiency in this market rejects an efficient market hypothesis and entails financial and institutional imperfections.

Sharma and Mahendru (2009) analyse the necessary condition of efficient market hypothesis for the BSE using weekly stock prices between July 2007 to October 2007 by employing run test and autocorrelation (ACF) analysis. The observation reveals that the BSE is efficient in weak-form and the security prices have very meagre impact on future stock returns which entails that it is not possible to make excess return.

A study by Dong Loc *et al*. (2010) focuses on the question whether the market is weak-form efficient. An important element of the investigation using run test and autocorrelation analysis refers the possible bias of the results caused by the thin trading that characterizes the STC. Stock market returns are adjusted for this. The main conclusion is that the STC is inefficient in the weak form.

Khan *et al.* (2011) conclude that the efficiency of the market provides an important insight for the stock investors, financial analyst, financial regulators, government, etc. The present research is crucial in the era of globalisation. This study examined the necessary condition of EMH for the Indian stock market. The efficiency of the Indian stock market is examined through the daily data of the indices of NSE and BSE from April 2000 to March 2010 by using run test. Alternate hypothesis is rejected based on the result of runs test and demonstrates that Indian stock market is inefficient and does not show random walk behaviour.

Beer and Bhuyana (2011) use several non-parametric and parametric econometric methods. The research study tests for the weak form efficiency of the Indonesian stock market, one of the predicting emerging markets of the world. After analyzing daily data from 1983 to 2007, the finding is that Indonesian stock market is not weak-form efficient. This is yet another financial market where stock prices can be anticipated from information in the past stock prices.

Gimba (2012) examines the weak-form of efficient market hypothesis of the NSE by supposing normal distribution and random walk of the return series using run and autocorrelation techniques. Daily and weekly All Share indices and five bank stocks of the NSE are tested for 3-year period for the daily data and for the weekly data from June 2005 to December 2009. The results indicate that the autocorrelation function rejects the null hypothesis of the existence of a random walk. Generally, it is revealed that the NSE is not efficient in the weak form of market efficiency. Given the empirical evidence that the stock market is inefficient in weak-form and it is thought that return predictability exists in the NSE.

Ogege and Mojekwu (2013) investigate the weak-form of the Nigerian stock market from the perspective of the efficient market hypothesis. The modeling of stock prices is to test for weak-form of efficiency to find out serial correlation, which might indicate whether current stock prices depend on previous prices or not. Using monthly data of the Nigerian stock price index from 1985-2010, a regression analysis, two non-parametric test including the runs test and correlogram, correlation matrix along with the least squares estimation techniques are employed. The findings reveal that investors can use past data to predict the future stock prices which represent inefficiency.

3. CONCEPTUAL FRAMEWORK

An efficient market from the efficient informational execution point of view is known as the Efficient Market Hypothesis (EMH). The EMH can be defined as where "stock prices 'fully reflect' all available information" (Fama 1970) and efficiency can be categorized in the following ways: weak, semi-strong and strong-form of market efficiency.

3.1 Weak Form Market Efficiency Models

The definition of efficient market hypothesis is where stock prices contain all the available information. This definition can be expressed in model form based on the assumption of expected returns i.e. an expression for the conditions of market equilibrium. The expected returns can be written as:

$$E(r_{j,t+1}|\Phi_t) = \frac{E(P_{j,t+1}|\Phi_t) - P_{j,t}}{P_{j,t}} \quad (1)$$

i.e.

$$E(P_{j,t+1}|\Phi_t) = [1 + E(r_{j,t+1}|\Phi_t)]P_{j,t} \quad (2)$$

where, E is the expected value operator, r indicates one period percentage return, $P\,j$ is the price of stock j, $\Phi\,t$ represents information set and $P\,j,\,t+i$ is the cumulative dividends.

The value of $\tilde{r}_{j,t+1}|\Phi_t$ is defined by the expected return theory. The expected return is determined by information set (Φt). It is also noteworthy that the assumption of market equilibrium in terms of expected returns is a mathematical concept (Campbell *et al.*, 1997; Keane, 1983).

The ***random walk hypothesis*** defines the successive price movements are independent. If the co-variance equation (1) holds, then returns are independent of each other, representing *random walk 1* and *random walk 2* models. In *random walk 3*, equation (1) entails that returns are not serially correlated. *Random walk 1* and *2* are a stronger sense of fair game than the *martingale* model because the first two cases of the random walk hypothesis implicate that even the non-linear functions of price increments are uncorrelated (Campbell *et al.*, 1997; Keane, 1983).

Random Walk 1(IID Increments) means that price increments are independently and identically distributed (IID), the process of $P\,t$ is expressed as:

$$P_t = \mu + P_{t-1} + e_t, \qquad e_t \sim IID(0, \sigma^2) \quad (3)$$

where, μ denotes drift, and IID $(0, \sigma^2)$ implies that the increments e_t are independently and identically distributed with a zero mean and variance of σ^2. The assumption of identically distributed increments is obviously not viable for stock prices for long time periods. It is unlikely to anticipate identical price changes due to unrelated and non-identical pieces of information. Likewise, it cannot be expected that stock prices will remain same after making changes in regulatory policy, institutional development and financial liberalisation that have a significant impact on the stock market (Campbell *et al.*, 1997; Damodaran, 1996). Variation in volatility and deviation from normality are peculiar characteristics of financial time series (Lo, 1997; Hall and Urga, 2002). However, the main focus should not be normality against the non-predictability and the absence of correlation. Thus, statistical tests should be employed with cautious because robustness might be questionable due to the many forms of heteroskedasticity and non-normality. Statistical tests for the IID assumption include sequences and reversals, and runs.

Random walk 2 (Independent Increments) takes the assumption of independent increments into account because of heteroskedasticity in the e_t's and to capture for the time variation in the volatility of financial time series of stock returns. Considering only

the independent increments assumption does not affect independently distributed e_t's, i.e., "… any arbitrary transformation of future price increments [cannot be estimated] using any arbitrary transformation of past price increments" (Campbell *et al.*, 1997). The unconditional and conditional probability of P_{t-1} do not change, i.e.

$$\Pr(P_{t+1} = P_t | P_{t-1}, P_{t-2} \ldots) = \Pr(P_{t+1} = P_t). \quad (4)$$

There are two lines of empirical research that can be viewed as a kind of "economic" test of RW2: filter rules, and technical analysis. (Campbell *et al.*, 1997)

Random walk 3 (Uncorrelated Increments) allows the identical distribution assumption to hold processes. *Random walk 3* to hold rather than *random walk 1* and *2* is any process where $Cov\ (e_t , e_{t-k}) = 0$ for all $k \neq 0$, but where $Cov\ (e^2_t , e^2_{t-k}) \neq 0$ for some $k \neq 0$. The uncorrelated increments do not mean that they are clearly independent as the squared increments are correlated (Merton, 1990; Damodaran, 1996; Campbell *et al.*, 1997). Tests of random walk 3 include autocorrelation coefficients, portmanteau statistics, and variance ratios.

4. ECONOMETRIC METHODOLOGY

The returns for KSE 100 index from 2002 to 2012 are calculated by using the following equation:

$$R_{At} = \{In\ (P_{At}/P_{A,\,t-1})\}*100 \quad (5)$$

where R_{At} is the stock return at a particular day, P_{At} indicates today's stock price and $P_{A,\,t-1}$ represents previous day stock price.

4.1 Non-parametric Stock Market Efficiency Tests

The methodology contains non-parametric tests such as run test and autocorrelation function (ACF) to analyze market efficiency which can be used as evidence for or against the EMH. Keane (1983) explains market inefficiency through an idea "Gambler's Fallacy describing the belief that what "goes up must come down". This phenomenon indicates time series behavior of stock prices exhibiting movements in opposite directions such as "rise in stock prices for a period of time and so is deemed to be "due for a fall". Generally speaking, after observing the relationship of the previous price to recent price movements, thus investor can better estimate the likely direction of future stock price movements, i.e. historical data can be used to predict future stock prices. This provides credibility to the argument that the market is predictable and inefficient. Therefore, the objective is to see whether the stock market is predictable or not by detecting serial dependence of stock returns. In this paper, two well-known tests of stock market efficiency consisting of the run test and autocorrelation function (ACF) test which can analyze serial dependence of stock returns.

4.1.1 Run Test

The run test is a non-parametric approach, based on the number of sequence of successive returns whether positive or negative analyzed under the random walk hypothesis (Campbell *et al.*, 1997; Gujarati, 2003). A run is the reoccurrence of the same value of a variable. It is based on two parameters like the type and the length of the run. There are three types of run such as negative, positive, or no change. The length can be defined as how frequently a run type occurs in succession (Islam *et al*., 2005). This non-parametric test is used to capture the autocorrelation, thereby, helpful in return

predictability. The following steps are involved, Firstly, find out the number of runs (consider the signs of prices whether positive, negative or no change. Secondly, the calculation of the mean and variance based on the following formulas respectively:

a. Expected return

$$\mu = \frac{N(N+1) - \sum_{i=1}^{3} n_i^2}{N} \tag{6}$$

where, *N* is the total number of price changes and *ni* the numbers of price changes of each sign.

b. Variance

$$\sigma_{\mu} = \left[\frac{\sum_{i=1}^{3}[\sum_{i=1}^{3} n_i^2 + N(N+1)] - 2N(\sum_{i=1}^{3} n_i^3 - N^3)}{N^2(N-1)} \right]^{\frac{1}{2}} \tag{7}$$

4.1.2 Autocorrelation Function Test and Correlogram

Another non-parametric technique is employed to capture the random behavior of the stock returns. The autocorrelation function (ACF) analysis is used to find out the degree of autocorrelation in a financial time series. It evaluates the correlation between the lagged and the current values of stock returns of the time series. It can be expressed as:

$$p_k = \frac{\sum_{t=1}^{n-k} (R_t - \bar{R})(R_{t+k} - \bar{R})}{\sum_{t=1}^{n} (R_t - \bar{R})^2} \tag{8}$$

where, *k* is the number of lags and R_t is the stock return. There are two ways of calculating autocorrelation such as the standard error test for individual lags and the Box Pierce Q (BPQ) test is used to identify the significance of coefficients at the group level (Islam *et al*., 2005). A test of stationary is based on autocorrelation function. The ACF at lag *k*, is defined as:

$$\rho_k = \gamma_k / \gamma_{0 = \text{Covariance lag k/Variance}} \tag{9}$$

Note that $k = 0$, $\rho_0 = 1$ since both covariance and variance are measured in the same units of measurement, ρ_k *is* a pure number. It lies between $-1 \leq \rho \leq 1$. If we plot ρ_k against *k*, the graph is known as the Correlogram. A correlogram makes it possible to find out a particular series stationary or non-stationary.

5. EMPIRICAL RESULT

The objective of this section is to find out the reliability of the notion from the perspective of Pakistan under the umbrella of institutional development that "future movements in financial markets are predictable, is inconsistent with the Efficient Markets Hypothesis (EMH), which states that security prices reflect all available information and as a result there are no opportunities for acquiring excess returns". For this purpose, the following Hypothesis has been tested:

H_0: No Autocorrelation *(Successive Returns are independent)*
H_1: Autocorrelation *(Successive Returns are not independent)*

Table 1
Run Test for Daily Data Set of KSE 100 Index

Period	Observed No. of Runs	Expected No. Runs	Negative	Positive	No Change	Test Values
Reform Period-1	405	456	371	538	0	0.022
Non-Reform Period	345	364	315	423	1	0.325
Reform Period-II	494	539	783	554	35	0.002
Overall	1244	1359	1169	1515	36	0.000

Source: Researcher's Estimation

A noteworthy aspect of runs of all sub-periods consisting of reform period-I, non-reform period and reform period-II, is that expected numbers of runs are greater than the observed numbers of runs. This indicates that the sign of residuals change frequently; therefore, there exists a positive correlation. Table 1 indicates the findings for the daily data.

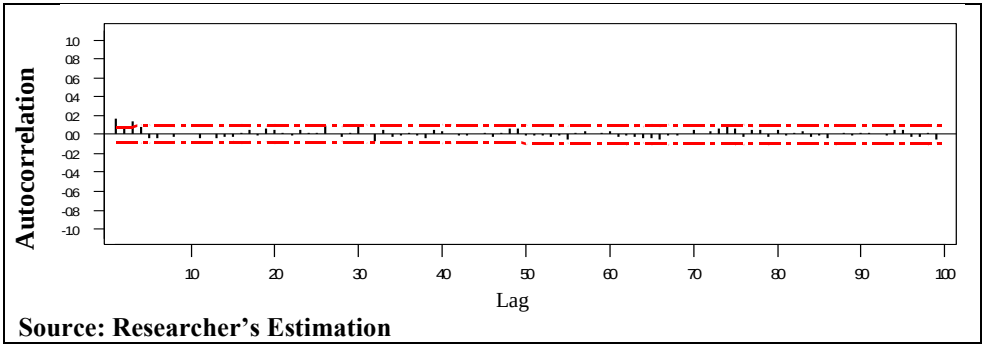
Table 2
Run Test for Weekly Data Set of KSE 100 Index

Period	Observed No. of Runs	Expected No. Runs	Negative	Positive	No Change	Test Values
Reform Period-1	62	94	57	133	0	0.000
Non-Reform Period	61	72	58	91	0	0.031
Reform Period-II	105	114	80	146	3	0.478
Overall	228	280	195	370	3	0.000

Source: Researcher's Estimation

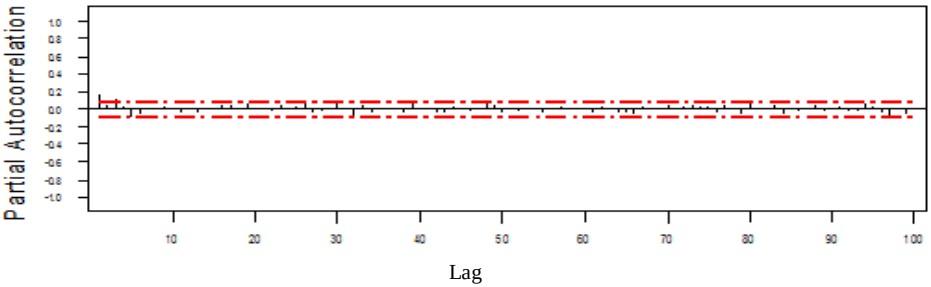
The findings for weekly data in table 2 show almost the same degree of autocorrelation as that of daily results. In addition, the overall test value is significant at 5% level of significance and it can be concluded that, for both type of data either weekly or daily, the rejection of the null hypothesis reveals that there exists autocorrelation.

According to Gujarati (2003), ACF is computed of about one-quarter to one-third of the length of a particular time series. In this study, weekly data has been used to calculate ACF. Both figures below contain the correlograms of both the autocorrelation function and partial autocorrelation function.



Source: Researcher's Estimation

Fig. 1: Autocorrelation Function for Weekly Data



Source: Researcher's Estimation

Fig. 2: Partial Autocorrelation Function for Weekly Data

Graphical representation indicates movements of autocorrelation at different lags. The results show the presence of positive autocorrelation. This explains the non-stationarity time series. However, returns are not distributed randomly. Moreover, lack of randomness leads investors to predict returns. Therefore, the test of weak form market efficiency from the perspective of Pakistan indicates inefficiency prevails in Karachi Stock Exchange and implies institutional imperfection. However, financial liberalization plays an important role to distort the return pattern but to lesser extent. The results exhibit a small level of positive autocorrelation of the daily and weekly returns on the stock during 2002-2012. The ACF and run test of weekly return are similar in that both tests produce a positive autocorrelation.

ACF Results for Daily Returns: Overall, at lag 100 the Q statistic value is 184.78 which is greater than χ^2 values such as 135.81, 124.34 and 118.49 at 1%, 5% and 10% respectively; shows the presence of autocorrelation.

a) ***Reform Period-I***

At lag 100, LBQ value is 124.19**,** the tabulated value is 118.50 at 10% at lag 100 respectively at 100 DF indicates presence of autocorrelation.

b) ***Non- Reform Period***

At lag 100, calculated value is 134.31 and χ^2 values (100 DF) are 129.56, 124.34 and 118.50 at all level of significance, shows the presence of autocorrelation.

c) ***Reform Period-II***

In the second reform period, calculated value is 267.56 at lag 100 which is greater than 129.56, 124.34 and 118.498 at 1%, 5% and 10% respectively at 100 DF indicates presence of autocorrelation.

ACF Results for Weekly Returns: Estimation provides evidence that there is autocorrelation in the KSE and there prevails inefficiency in Karachi stock market. The calculated value is 126.04 is greater than118.50 indicates the presence of autocorrelation in Karachi stock market.

a) ***Reform Period-I***

The ACF test provides evidence regarding the presence of autocorrelation. The calculated *Q* value at lag 45 yields 53.04 is greater than tabulated value at the 5% level of significance. Conclusion leads us to reject the null hypothesis and secondly, the autocorrelation is stronger in the daily data.

b) Non- Reform Period

At lag 37 the calculated value is 41.16 and tabulated chi-square is 40.26. So, at 10% level of significance there has been found a positive autocorrelation.

c) Reform Period-II

The result pertaining to the reform period-II also indicates that there exists positive autocorrelation. The *Q* value at lag 55 yields 77.85 is significant at 5% and 10% as compared to these values 63.17 and 67.81 respectively.

6. CONCLUSION

The empirical investigation of short-term return predictability through the efficient market hypothesis (EMH) contributes a lot to the understanding of an emerging market of Pakistan. The results from the run test on daily and weekly return data of KSE 100 index reject the null hypothesis of no autocorrelation. So the presence of autocorrelation leads us to draw a conclusion that this emerging stock market is inefficient and implies institutional imperfections. The empirical results of this research study are also supported by the previous studies of many emerging markets like Indian, Thai and Nigerian (Pandey, 2003; Islam *et al*., 2005; Ogege and Mojekwu, 2013), etc. Institutional development and financial liberalisation in the emerging stock market of Pakistan have not been succeeded to overcome anomalous behaviour of the market. Moreover, this leads to predict returns and to reject random walk hypothesis (RWH) means there is a strong chance that investors or security analysts could use historical data to earn abnormal profit while trading in the market. However, financial liberalization plays an important role to distort the return patterns but to some extent and to accept the hypothesis of market rationality means the absence of persistent anomaly.

The theoretical and empirical studies of the EMH make an important contribution to the understanding of an emerging market of Pakistan. The findings exhibit inconsistency and instability as result of regulatory policy changes. The implication is that the benefits of a well-functioning stock market are not being realized in the economy. Indeed, the weak form inefficiency of the stock market found in this study is most likely caused by a combination of the lack of its development and the implication of policy choices. The present study provides more insights into the operation and characteristics of the stock market of Pakistan in terms of its efficiency to make an assessment of the empirical characteristics of the Karachi stock market.

REFERENCES

1. Beer, F.M. and Bhuyan, R. (2011). Additional evidence against the Weak-form of the Efficient Markets Hypothesis, the Indonesian Stock Exchange. *Pan-Pacific Journal of Business Research*, 2(1), 75-103.
2. Benninga, S. (2008). *Financial Modeling*. Massachusetts London, The MIT Press Cambridge.
3. Berument, H. and Kayimaz, H. (2001). The Day of the Week Effect on Stock Market Volatility. *Journal of Economics and Finance*, 25,181-193.
4. Campbell, J.Y. and Shiller, R.J. (1987). Cointegration and Tests of Present Value Models. *Journal of Political Economy*, 95, 1062-1088.
5. Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997). *The Econometrics of Financial Markets*. New Jersey, Princeton University Press, Princeton.

6. Cuthbertson, K. (1996). *Quantitative Financial Economics, Stocks, Bonds, Foreign Exchange*. John Wiley and Sons, London.
7. Damodaran, A. (1996). *Investment Valuation*. New York: John Wiley and Sons, Inc.
8. Dong Loc, T., Lanjouw, G. and Lensink, R. (2010). Stock-Market Efficiency in Thin-Trading Markets, the Case of the Vietnamese Stock Market. *Applied Economics*, 42(27), 3519-3532.
9. Fama, E.F. (1965). The Behaviour of Stock Market Prices. *Journal of Business*, 38, 34-105.
10. Fama, E.F. (1970). Efficient Capital Markets, a Review of Theory and Empirical Work. *Journal of Finance*, 25, 383-417.
11. Fama, E.F. (1976). *Foundations of finance*. Basic Books, New York.
12. Fama, E.F. (1991). Efficient Capital Markets, II. *Journal of Finance,* 96, 1575-1617.
13. Fama, E.F. and French K.R. (1989). Business Conditions and Expected Returns to Stocks and Bonds. *Journal of Financial Economics*, 25, 23-50.
14. Fama, E.F., French, K, Booth, D. and Sinquefield, R. (1993). Differences in the Risks and Returns of NYSE and NASD Stocks. *Financial Analysts Journal*, 49, 37-41.
15. Friend. I., Blume, M. and Crockett, J. (1970). *Mutual Funds and Other Institutional Investors, a New Perspective.* McGraw-Hill, New York.
16. Fry, M.J. (1995). *Money, Interest and Banking in Economic Development*. 2nd ed. The Johns Hopkins University Press, Baltimore.
17. Gimba, Victor K. (2012). Testing the Weak-form Efficiency Market Hypothesis, Evidence from Nigerian Stock Market. *Journal of Applied Statistics*, 3 (1), 117-136.
18. Gujarati, D.N. (2003). *Basic Econometrics*. 4th ed Delhi, McGraw Hill Inc.
19. Hall, S. and Urga, G. (2002). Testing for Ongoing Efficiency in the Russian Stock Market. *http://www.nes.ru/~agoriaev/Papers/Hall-Urga%20ME%20timevarying%20Russia%2002.pdf.*
20. Harvey, C.R. (1991). The world price of covariance risk. *Journal of Finance*, 46, 111-157.
21. Ho, Y.K. (1990). Stock Return Seasonalities in Asia Pacific Markets. *Journal of International Financial Management and Accounting*, 2, 44-77.
22. Ikenberry D, Rankine G and Stice EK (1996). What Do Stock Splits Really Signal? *Journal of Finance*, 48, 65-91.
23. Islam, S., M.N. and Oh, K.B. (2003). *Applied Financial Econometrics in E-Commerce. Contributions to Economic Analysis*. North Holland Publishing, Amsterdam.
24. Islam, S., M.N., Watanapalachaikul, Sethapong and Colin Clark (2005). Are Emerging Financial Markets Efficient? Some Evidence from the Models of the Thai Stock Market, Unpublished paper.
25. Jaffe, J.R. (1974). Special Information and Insider Trading. *Journal of Business*, 47, 410-428.
26. Jegadeesh, N. and Titman, S. (1993). Returns to Buying Winners and Selling Loosers, Implications for Stock Market Efficiency. *Journal of Finance*, 48, 65-91.
27. Jensen, M. (1978). Some Anomalous Evidence Regarding Market Efficiency. *Journal of Financial Economics*, 12, 33-56.
28. Jensen, M. and Ruback, R.S. (1983). The Market for Corporate Control, the Scientific Evidence. *Journal of Financial Economics*, 11, 5-50.
29. Keane, S. (1983). *Stock Market Efficiency*. Phillip Allan Publishers, Oxford.

30. Keasey, K. and Mobarek, A. (2000). Weak-Form Market Efficiency of an Emerging Market, Evidence from Dhaka Stock Market of Bangladesh. The paper presented at the ENBS Conference held on Oslo, May 2000.
31. Kettel, B. (2001). *Financial Economics: Making Sense of Market Information*. Financial Times Prentice Hall, London
32. Khan, A.Q., Ikram, S. and Mariyam, M. (2011). Testing Weak Form Market Efficiency of Indian Capital Market, A Case of National Stock Exchange (NSE) and Bombay Stock Exchange (BSE). *African Journal of Marketing Management*, 3(6), 115-127.
33. Lo, A.W. (1997). *Market efficiency: Stock Behaviour in Theory and in Practice*. Cheltenham: Edward Elgar Publishing Limited.
34. McConnell, J.J. and Muscarella, C.J. (1985). Corporate Capital Expenditure Decisions and the Market Value of the Firm. *Journal of Financial Economics*, 14, 399-422.
35. Merton, R.C. (1990). *Continuous-Time Finance*. Cambridge, M.A: Blackwell Publishers.
36. Michaely, R., Thaler, R. and Womack, K. (1995). Price Reactions to Dividend Initiations and Omissions, Overreaction or Drift? *Journal of Finance*, 50, 573-608.
37. Mills, T.C. (1999). *The Econometric Modeling of Financial Time Series*. Cambridge University Press, Cambridge
38. Nassir, A. and Mohammad, S. (1987). The January Effect of Stocks Traded at the Kuala Lumpur Stock Exchange, an Empirical Analysis. *Hong Kong Journal of Business Management*, 5, 35-50.
39. Ogege, Samson and Mojekwu, J.N. (2013). Econometric Investigation of the Random Walk Hypothesis in the Nigerian Stock Market. *Journal of Emerging Issues in Economics*, 1(5), 381-400.
40. Pandey, A. (2003). *Efficiency of Indian Stock Market*. Time Series Course.
41. Ramachandran, J. (1986). Behavior of Stock Market Prices, Information Assimilation and Market Efficiency. Unpublished FPM Thesis, IIM Ahmedabad.
42. Samuelson, P.A. (1965). Proof That Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, 6, 41-50.
43. Seiler, M.J. and Walter, Rom (1997). A Historical Analysis of Market Efficiency, Do Historical Returns Follow A Random? *Journal of Financial Strategic Decisions*, 10(2), 49-57.
44. Sharma G.D. and Mahendru, M. (2009). Efficiency Hypothesis of the Stock Markets, A Case of Indian Securities. *International Journal of Business and Manage*ment 4(3), 136-144.
45. Sharma, J.L. and Kennedy, R.E. (1977). A Comparative Analysis of Stock Price Behaviour on the Bombay, London and New York stock Exchanges. *Journal of Financial and Quantitative Analysis*, 12, 391-413.
46. Sharpe, W.F. (1966). Mutual Fund Performance, *Journal of Business*, January
47. Williamson, J. (1972). Measuring Mutual Fund Performance. *Financial Analysts Journal*, November/ December
48. Yalawar, Y.B. (1988). Bombay Stock Exchange, Rates of Return and Efficiency. *Indian Economic Journal*. 35, 68-121.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 249-264

SHORT-TERM RETURN PREDICTABILITY USING A PARAMETRIC STOCK MARKET EFFICIENCY TEST: EVIDENCE FROM THE KARACHI STOCK EXCHANGE

Abida Hafeez
Applied Economics Research Centre (AERC), University of Karachi.
Email: abida_aerc@yahoo.com

ABSTRACT

This study examines the efficiency of the Karachi stock exchange using a parametric stock market efficiency test. The objective is to test the short-term return predictability through univariate autoregressive integrated moving average (ARIMA) model as one of the forecasting techniques using weekly data of the KSE 100 index from 2002 to 2012. The parametric test strongly recommends that return series does not follow the random walk model and rejects the hypothesis of weak form efficiency. The results from the empirical analysis powerfully propose that the KSE is inefficient in weak form. The observed findings reveal that the prevailed inefficiency in KSE could affect the decision making regarding the portfolios investment and asset allocation, trading in the financial market, risk management conditions and information available to the investors. Also, the investors could use the level of efficiency to predict and to quantify return volatility before making their trading decisions.

KEYWORDS

Efficiency, Forecasting, Volatility, Return Predictability, Autoregressive Integrated Moving Average (ARIMA).

1. INTRODUCTION

Forecasting return volatility is a crucial and thought-provoking financial task. It is agreed to large extents that although daily and monthly stock returns are not suitable to make a good forecast, whereas volatility is highly predictable. Forecasting has significance for financial economics and risk management (Torben *et al.*, 2009). Autoregressive Integrated Moving Average (ARIMA) has been used in different fields other than business and financial economics due to its popularity and suitability for short-term forecasting. For instance, it has been used in agriculture economics to predict poultry retail prices (Fahimifard *et al.*, 2009); Cocoa Bean Price (Assis *et al.*, 2010) and Wheat Production (Saeed *et al*., 2000); in macroeconomics this model has been employed for policy making and to predict inflation in Ghana (Alnaa and Ahiakpor, 2011) and again, to predict inflation in Ghana (Samnel and Ferdinand, 2011).

Some researchers have been blended ARIMA, to forecast the long-term memory of OPEC oil prices through Autoregressive Fractionally Integrated Moving Average

(ARFIMA) (Mostafaei and Sakhabakhsh, 2011); seasonal time series ARIMA (SARIMA) to predict international tourists' footfalls in India (Purna, 2011) and SARIMABP which is a combination of seasonal time series ARIMA and the neural network back propagation (BP) considered as a hybrid forecasting model (Tseng *et al.*, 2002). According to Elena and Storis (2009), movements in financial markets are somewhat observable and stock returns may be happened volatile. Such trends in the markets have been under analysis for a long time and are still being studied strongly.

Forecasts have traditionally been made using econometric models such as the univariate time series models known as autoregressive integrated moving average (ARIMA) models based on Box and Jenkins (1970). These models have been extensively used to make forecast for any economic time series, inventory and sales modelling (Brown, 1959; Holt *et al.*, 1960). Several methods for identifying special cases of autoregressive integrated moving average (ARIMA) models have been proposed by Univariate Box-Jenkins (UBJ) models are in particular suitable for short-term forecasting. Makridakis *et al.* (1982) and Meese and Geweke (1982) discuss the methods of recognizing as being univariate models. Amongst other researchers Jenkins and Watts (1968), Yule (1926, 1927), Bartlett (1964), Quenouille (1949), Ljunge and Box (1978) and Pindyck and Rubinfeld (1981) have also proposed to employ ARIMA methodology for forecasting.

Autoregressive Integrated Moving Average (ARIMA) or Univariate Box-Jenkins (UBJ) models are in particular suitable for short-run forecasting. Pankratz (1983) studies short-run forecasting, because most ARIMA models emphasized on the recent past rather than the distant past. This model emphasizes on the recent past forecast because short-run forecasts are more authentic than those of long-run from ARIMA models.

This research paper is organized as follows; Section 2 reviews literature on short term return predictability using a parametric test (ARIMA). Section 3 discusses conceptual framework. Section 4 contains data and methodology. The empirical findings are shown in section 5. Section 6 provides concluding remarks.

2. LITERATURE REVIEW

Gujarati (2003) referred the publication of Time Series Analysis, Forecasting and Control by Box-Jenkins (1976) and introduced a new generation of forecasting tools. Popularly known as Box-Jenkins methodology, but technically known as ARIMA model, the emphasis of this method is not on constructing single-equation or simultaneous-equation models but on examining the probabilistic, or stochastic properties of an economic time series on their own thereby allowing the data to 'speak' for themselves. Unlike regression models, in which Yt is explained by K regressors, ARIMA models allow Yt to be explained by past values of Y itself and the stochastic error term. ARIMA model therefore constitute a flexible class of models for describing the behaviour of economic and financial time series.

Simons and Laryea (2004) analyze the efficient market hypothesis for Ghana, Mauritius, Egypt and South Africa. The findings reveal that the South African stock

market is efficient in weak-form, while the other markets are weak-form inefficient. The results also entail that the stock returns from the Ghana, Mauritius and Egypt are predictable from historical prices. Hence, this study fitted an ARIMA model (Egypt ARIMA (*1, 0, 1*); Ghana (*1, 0, 2*); and Mauritius (*2, 0, 1*)) to the excess return data of the three markets using the Box-Jenkins method. The study used the ARIMA model to generate one-period forecasts for the subsequent 12 periods for these three countries. The ARIMA forecasts in all the markets outperformed the naïve model.

In a similar study, Rahman and Hossain (2006) apply ARIMA (*1, 0, 1*) among other approaches to examine the weak form efficiency of the Dhaka Stock Exchange. The results do not accept the null hypothesis of random walk for all share index and general price index. On the basis of Akaike's Information Criterion, R^2, and autocorrelation functions, they find ARIMA (*3,0,2*) and ARIMA (*1,0,1*) respectively, as the best fitted model with all coefficients significant at 1% level of significance. This study divides the sample data into two sub-samples. The first sub-sample is named historical period, the second sub-sample is named validation period. Then they employ ARIMA (*3, 0, 2*) and ARIMA (*1, 0, 1*) to generate data for the validation period and test how the fitted value deviates from the actual values. They find that the fitted values derived from the models and actual data is all but well fitted.

Al-Shiab (2006) studies the univariate ARIMA forecasting model on the Amman Stock Exchange (ASE) applying general daily index from 4/1/2004 to 10/8/2004. Four diagnostic tests are executed to select the best model describing the data, namely, R-square, adjusted R-square, Akaike's Information Criterion (AIC) and Schwarz Information Criterion (SIC). On the basis of these four diagnostic tests, ARIMA (*4, 1, 5*) model is chosen as the best model that explains the data and is suitable for accurate forecasting. The selected model predicts that the ASE will continue to grow by 0.195% for seven days starting on 11/8/2004. This forecast, however, is inconsistent with actual performance during the period of the forecasting since the ASE went down by 0.003%. He concludes that the forecast error implies the ASE tends towards weak form efficiency.

Irfan *et al.* (2010) analyze weak form of efficiency in the Karachi Stock Exchange (KSE). The data set includes the daily and monthly prices of KSE-100 indices from 1st January1999 to 31st August 2009. Different tests such as ARIMA, unit root and autocorrelation function have been employed to examine the efficiency of the KSE. The parameters of AR (p) and MA (q) are compared according to the different criterion like AIC and SIC to select the best fitting model in both returns. Correlogram of ARIMA residuals show no autocorrelation and partial autocorrelation is left in both series, therefore, there is no need to search out another ARIMA model. ARIMA (1, 0, 1) for daily return series and ARIMA (0, 0, 1) for monthly return series are finally selected. All the tests reveal that both daily and monthly returns do not show random behavior and suggest that the KSE is inefficient in weak form.

O Emenike (2010) describes that it is no news that the global economic meltdown has led to shortage of financial resources and to a general downtrend in stock prices all over the world. Forecasting stock prices will give a way to predict and perhaps avert the risk

of a large adverse change in stock prices. This research study, therefore, models and forecasts prices of the Nigerian Stock Market using the autoregressive integrated moving average (ARIMA) methodology. Monthly data set of All-Share Indices of the NSE from 1985 to 2008 is taken as the fit sample; on the other hand January 2009 to December 2009 is taken as out-of-sample forecast period. Several diagnostic tests are performed to choose the *p, d, q* parameter that best fit the index. The ARIMA (1, 1, 1) model is selected to make forecasting of index points and growth rates. Predicted index points and growth rates indicate deviation from the actual indices and growth rates. However, theil U statistics shows that forecasting made by ARIMA (1,1,1) model outperformed the naïve model. Hence, the deviations show that the global economic crisis put down the correlation relationship existing between the NSE All-Share Index and its past.

A study by Al-Zeaud (2011) suggests ARIMA methodology as one of the forecasting approaches for the financial time series. ARIMA model has been used to make forecast the volatility for banking sector of Amman Stock Exchange (ASE). The weekly data set of indices has been used to predict volatility from 01/01/2005 to 01/04/2010. The research study uses different econometric techniques like co-integration test, unit root analysis, minimum mean square error (MSE) t-statistics and p-statistics to select the best ARIMA models at 95% confidence interval. All the models from ARIMA (0, 0, 0) to the model ARIMA (2, 2, 2) are tested. Finally, ARIMA (2, 0, 2) is selected as the best fitted model based on the minimum mean square error at 95% confidence interval. Hence, investors could take advantage using this model to quantify and predict about return volatility before making trading decisions.

Zahid *et al*. (2012) aim at analyzing weak form efficiency of the Karachi stock exchange. Different parametric and non-parametric techniques of random walk model are used. The bid and offer rates of Karachi Stock Exchange (KSE) of Pakistan over the period of March 2000 to October 2011 are considered for testing the weak form efficiency. Different parametric and non-parametric approaches like Kolmogrov-Smirnov goodness of fit test, run test, Q-Q probability plots, auto-correlation coefficient test, unit root test, correlogram test and ARIMA methodology are employed to analyze the efficiency of the KSE. The skewness and kurtosis measures for the stock return series declare that it is non-normal. The same conclusion is reached with Kolmogorov-Smirnov (K-S) approach. The parameters of AR (p) and MA (q) are compared according to the different criteria like AIC and SIC to select the best fitting model and ARIMA (1, 0, 8) is selected. All parametric and non-parametric methods employed in this study reject the hypothesis of normality of the series and randomness of stock returns in Karachi stock market and suggest that the Karachi stock market is inefficient in weak-form.

3. CONCEPTUAL FRAMEWORK

The efficient market hypothesis (EMH) consists of following statements about, (1) the theory that security prices reflect the true value of stocks; (2) the absence of arbitrage opportunities in an economy populated by rational, profit-maximizing agents; and (3) the market hypothesis that stock prices always fully reflect available information (Fama 1970).

The market efficiency can be categorized in the following ways, weak, semi-strong and strong-form of market efficiency. These classifications were originally suggested by Fama (1970). The *weak form efficiency* contains the information that reflects only the history of prices or returns. The *semi-strong form of market efficiency* includes most information known to all market participants. Finally, *strong form efficiency* consists of all information known to any market participant.

According to Jensen (1978), an efficient market is clearly characterized with respect to an information set Φ_t if it is impossible to make excess profits by trading on the basis of information set (Φ_t). Fama (1970) presented a general notation indicating how investors generate price expectations for stocks. This could be explained as (Cuthbertson, 1996):

$$E(p_{j,t+1} \mid \Phi_t) = [1 + E(r_{j,t+1} \mid \Phi_t)] p_{jt} \tag{1}$$

where E is the expected value operator, $p_{j,t+1}$ is the stock price j at time *t+1*, $r_{j,t+1}$ is the return on security *j* during period *t+1*, and Φ_t is the set of information fully available to investors at time *t*. The left-hand side of the equation (1) $E(p_{j,t+1} \mid \Phi_t)$ denotes the expected end-of-period price on stock *j*, given the information available at the beginning of the period Φ_t. On the right-hand side, $1 + E(r_{j,t+1} \mid \Phi_t)$ indicates the expected return over the forthcoming time period of securities having the same amount of risk as security *j*. Under the efficient market hypothesis (EMH), investors cannot earn abnormal profits on the available Φ_t other than by chance. The level of over value or under value of a particular security is defined as:

$$x_{j,t+1} = p_{j,t+1} - E(p_{j,t+1} \mid \Phi_t) \tag{2}$$

where $x_{j,t+1}$ indicates the extent to which the actual price for security *j* at the end of the period varies from the price expected by investors based on the information available Φ_t. Consequently, in an efficient market it must be true that:

$$E(x_{j,t+1} \mid \Phi_t) = 0 \tag{3}$$

This implies that the information is always reflected in stock returns. Therefore, the rational expectations of the particular security return according to the EMH may be represented as:

$$P_{t+1} = E_t P_{t+1} + \varepsilon_{t+1} \tag{4}$$

where P_t is the stock price and ε_{t+1} is the forecast error. $P_{t+1} - E_t P_{t+1}$ should therefore be zero on average and should be uncorrelated with any information Φ_t. Also

$E(x_{j,t+1} \mid \Phi_t) = 0$ when the random variable (good or bad news), the expected value of the forecast error is zero.

$$E_t \varepsilon_{t+1} = E_t (P_{t+1} - E_t P_{t+1}) = E_t P_{t+1} - E_t P_{t+1} = 0 \quad (5)$$

Underlying the efficiency market hypothesis (EMH), it is timely to mention that expected security returns are entirely consistent with randomness in stock returns. This position is supported by the *law of iterated expectations* (Campbell *et al.,* 1997; Samuelson 1965). The random walk hypothesis (RWH) holds, the efficient market hypothesis holds; the efficient market hypothesis must hold with respect to past returns.

4. ECONOMETRIC METHODOLOGY

4.1 Box-Jenkins (BJ) Methodology (ARIMA)

Box and Jenkins (1976) present a four-step model (ARIMA model) for selecting a best fitted model. The four-stage ARIMA methodology is defined as:

Step 1: Identification (choosing appropriate *p, d, q*)

Identification, the identification stage uses two graphical techniques to estimate correlation including an autocorrelation function (ACF) and a partial autocorrelation function (PACF) to find the appropriate values of p, d and q. Then the estimated ACF and PACF are used to choose one or more ARIMA models that appear appropriate. Efficient market hypotheses (EMH) assert that stock return indices are fundamentally random. The introductory model for volatility estimation in stock returns is the random walk model as:

$$R_i = \theta + \mu_t \quad (6)$$

Secondly, the easiest ways to express autoregressive of order one AR (1) model may also be estimated as:

$$R_i = \theta + \rho R_{i-1} + \mu_t \qquad -1 \leq \rho \leq 1 \quad (7)$$

where θ in both above equations is the constant parameter, ρ is the estimated parameter and u_t is white noise with zero mean and constant variance σ^2, an uncorrelated random error term. This model looks like the Markov first order autoregressive model. If $\rho \geq 1$, which means a unit root problem exists in the returns, R_t is non-stationary series. All three terms; non stationary, random walk and unit root can be dealt as identical. If $\rho \leq$, R_t can be changed into stationary series.

a. <u>Auto Regressive (AR) Model</u>

The first-order autoregressive model is the extensively used model of correlation. The AR (1) model is specified as:

$$R_i = \theta + \rho R_{i-1} + \mu_t \qquad -1 \leq \rho \leq 1 \quad (8)$$

where α is the vector of constant term, here the value of R at time t depends on its previous value and an error term; the R values are stated as deviations from their mean values. The higher order autoregressive model or autoregressive model of order "p" denoted by AR (p) is given as:

$$R_i = \theta + \alpha_i R_{t-1} + \alpha_i R_{t-2} + \ldots + \alpha_p R_{t-p} + \mu_t \qquad (9)$$

Then R_t is supposed to follow a random walk model (RWM) with drift due to the presence of its constant parameter α , α_p are the parameters of autoregressive coefficients and u_t is an uncorrelated random error variable.

b. <u>Moving Average (MA) Model</u>

Moving average process of order q can be formed easily by a weighted average of random error term and can be written in the following way:

$$R_i = \theta + \mu_t + \beta_0\mu_t + \beta_1\mu_{t-1} + \beta_1\mu_{t-2} \ldots . \beta_q\mu_{t-q} \qquad (10)$$

where θ is the intercept term, u_t is white noise error term having zero mean and variance σ^2_u and $\beta q's$ are unknown parameters.

Step 2: Estimation (parameter estimation of chosen model)

After identifying the appropriate p and q values, the next step is to estimate the parameters of autoregressive (AR) and moving average (MA). This calculation mostly is done by ordinary least square (OLS) method to choose appropriate ARIMA model at this stage. Thus, R_t (time series of stock prices) follows an ARIMA (p, d, q) process if it can be expressed as:

$$R_i = \theta + \alpha_i R_{t-1} + \ldots . + \alpha_p R_{t-p} + \mu_t + \beta_0\mu_t + \beta_1\mu_{t-1} \ldots . \beta_q\mu_{t-q} \qquad (11)$$

Because there are 'p' autoregressive terms and 'q' moving average terms, 'd', denotes numerical value of differenced time series before it goes stationary and θ is a constant term. Where the intercept parameter is related to the mean of R_t, the errors are assumed to be uncorrelated random variable with zero mean and constant variance, are the unknown parameters of autoregressive process and are the unknown parameters of moving average process. A simplest form of the autoregressive moving average model of order 1 of both p and q orders *ARIMA* (1, 1) can be written as:

$$R_i = \theta + \alpha_i R_{t-1} + \mu_t + \beta_1\mu_{t-1} \qquad (12)$$

where θ is an intercept term and is assumed to be uncorrelated random variables, is an unknown parameter of autoregressive model and is an unknown parameter of moving average process.

Step 3: Diagnostic checking

Having chosen particular ARIMA model, the next step is to see whether the selected model is appropriate. It depends on the estimated residuals series, it should be white noise. Box and Jenkins (BJ) methodology suggests some diagnostic checking to see whether selected model is statistically appropriate or not; if "Yes" then go for the last

stage i.e. forecasting; if "Not" then go back for the first stage and repeat the procedure from identification and estimation of parameters to diagnostic checking until there is a good final model.

Step 4: Forecasting

One of the reasons of the widely accepted ARIMA models is its success in forecasting. The forecasting made by this method is more authentic than those made from other econometric models particularly for short run forecasts (Gujarati, 2003).

5. EMPIRICAL RESULTS

The objective of this section is to test the efficiency of the Karachi stock market, to make forecasting regarding return volatility and to quantify volatility of returns on overall KSE weekly index from 2002 to 2012 using a parametric technique i.e. ARIMA methodology. Thus, R_t (time series of stock prices) follows an ARIMA (p, d, q) process as shown in equation (11):

$$R_i = \theta + \alpha_i R_{t-1} + \ldots + \alpha_p R_{t-p} + \beta_0 \mu_t + \beta_1 \mu_{t-1} \ldots \beta_q \mu_{t-q}$$

Return volatility for index is calculated by using the formula:

$$R_{At} = \{In (P_{At}/P_{A, t-1})\} * 100$$

The first step of this analysis is based on the statistics of KSE 100 Index.

Table 1
Descriptive Statistics of KSE 100 Index

Estimators	Mean	Median	Maximum	Minimum	Coef. of variation
KSE 100 Index	8920.34	9664.25	16943.19	1362.73	0.045
Estimators	Std.Dev	Skewness	Kurtosis	variance	
KSE 100 Index	4014.46	-0.241	2.079	16115889.09	

Source: Researcher's Calculation

Table 1 presents descriptive statistics of weekly index for 2002-12. A positive value of kurtosis shows that the series of stock return has flatter tail than that of the stationary distribution. The negative value of skewness suggest that time series has asymmetric distribution skewed to the left. Furthermore, the overall return series indicates low dispersion, which depicts the spread means too small around the mean.

Table 2
Descriptive Statistics of KSE 100 Index (Return Volatility)

Estimators	Mean	Median	Maximum	Minimum	Coef. Of variation
KSE 100 Index	0.441	0.819	10.92	-20.09	7.802
Estimators	Std.Dev	Skewness	Kurtosis	variance	
KSE 100 Index	3.438	-1.365	8.332	11.82	

Source: Researcher's Calculation

Table 2 contains the statistics for return volatility. Statistics indicate presence of volatility over the period of 2002-2012. The coefficient of standard deviation indicates that a significant change in volatility appears which depicts an association between socio-economic conditions of the country and stock market behaviour. The mean is less than the median means that distribution skewed to the left. The return series skewed to the left indicates asymmetric distribution. In addition, the large value of means distribution has a sharper peak, fatter and longer tails rather than normal distribution. This situation is favourable for risk taking investors. A unit root analysis in table 3 indicates that return series is stationary at levels. Both t-value and p-value drive us towards the rejection of null hypothesis of unit root.

Table 3
Unit Root Analysis of Weekly Return Series of KSE 100 Index

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-19.92	0.000
Test critical values, 1% level	-2.569	
5% level	-1.941	
10% level	-1.616	

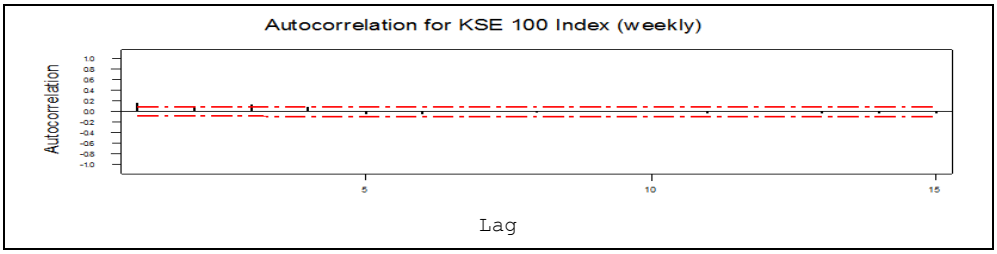
Source: Researcher's Calculation
*MacKinnon (1996) one-sided p-values

Table 4 shows the autocorrelation function (ACF) for the weekly returns of KSE 100 index, the autocorrelation has been calculated up to 15 lags along with t-statistics. The autocorrelation at both lags 1 and 3 is significant where the computed values (0.160 and 0.130) of autocorrelation are significantly greater than tabulated t-value i.e. 2. The remaining lags have smaller t-values than 2.

Table 4
Autocorrelation Function, Stock Market Volatility

Lag	ACF	T-Values
1	0.160	3.721
2	0.071	1.532
3	0.130	2.943
4	0.084	1.754
5	-0.051	-1.203
6	-0.060	-1.241
7	-0.000	-0.114
8	-0.020	-0.552
9	0.000	0.031
10	0.000	0.094
11	-0.040	-0.902
12	0.000	0.061
13	-0.044	-0.892
14	-0.032	-0.721
15	-0.041	-0.821

Source: Researcher's Calculation



Source: Researcher's Estimation

Figure 1: Autocorrelation Function, Stock Market Volatility

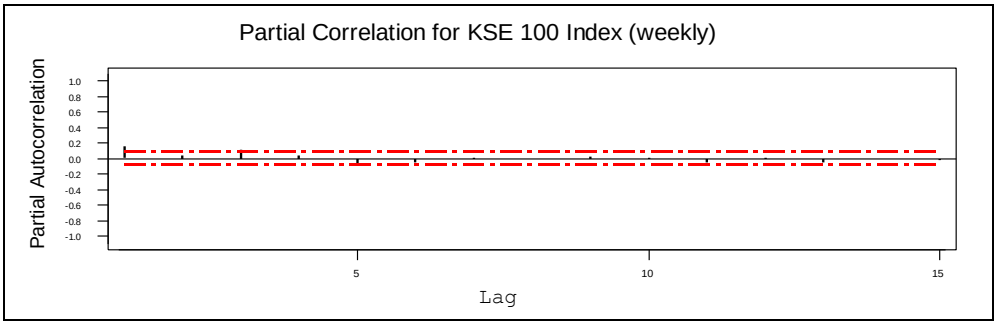
The figure 1 shows that the ACF has spikes at lags 1 and 3 and an indication of significant positive autocorrelation. In addition, it also means that autocorrelation of different observations is not within sampling error of zero. The autocorrelation of all other lags (2, 4-15) is within the confidence interval of 95%. It is called autoregressive (AR) process of order two.

Table 5 shows the partial autocorrelation function; the partial autocorrelation for lags 1 and 3 is greater in absolute value than 2. The remaining partial autocorrelations (2, 4-15) are insignificant. It is called moving average (MA) process of order two. Figure 2 shows the PACF for the weekly return series of KSE 100 index. The partial autocorrelation function indicates significant spikes at lags 1 and 3. All the other partial autocorrelations (2, 4-15) are insignificant and lie within the 95% confidence limits.

Table 5
Partial Autocorrelation Function, Stock Market Volatility

Lag	ACF	T-Values
1	0.160	3.721
2	0.041	1.012
3	0.110	2.703
4	0.040	0.964
5	-0.084	-1.990
6	-0.060	-1.354
7	0.000	0.073
8	-0.011	-0.182
9	0.030	0.701
10	0.000	0.121
11	-0.053	-1.162
12	0.010	0.234
13	-0.051	-1.083
14	-0.010	-0.290
15	-0.023	-0.491

Source: Researcher's Calculation



Source: Researcher's Estimation

Figure 2: Partial Autocorrelation Function, Stock Market Volatility

In this study, AIC and SIC along with the mean square error (MSE) are used to choose the best ARIMA model for the weekly KSE 100 Index data. Table 6 shows all the choices for the selection of best ARIMA model. For this purpose, all possible models have been tested on the basis of the mean square error (MSE), Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) and the minimum values of all above three criteria are considered best in selecting an appropriate ARIMA model. Finally, the best selected model is ARIMA (1, 0, 1) based on MSE which is 11.45 and minimum AIC and SIC values are 5.272 and 5.290 respectively, and then ARIMA (1, 0, 1) is chosen as the best model for forecasting.

Table 6
The Result of the Best ARIMA Model for Market Volatility

Model(ARIMA)	MSE	AIC	SIC
(1,0,0)	11.51	5.276	5.291
(0,0,1)	11.53	5.287	5.302
(1,0,1)	11.45	5.272	5.290
(1,0,2)	11.47	5.275	5.306
(2,0,1)	11.47	5.276	5.307
(0,0,2)	1154	5.289	5.312
(2,0,0)	11.51	5.275	5.298
(2,0,2)	11.47	5.276	5.314

Source: Researcher's Calculation

The ARIMA (1, 0, 1) model is finally selected. This task has been completed through Minitab program. The program has a speciality of using least squares algorithms through a combination of search routines and successive approximations to obtain final least square point estimates of the parameters. The final estimates are those that minimize the sum of squared errors to a point where no other estimates can be found that yield smaller sum of squared errors. This is known as convergence. Table 7 shows the estimates of the ARIMA (1, 0, 1). All the coefficients of AR (1) and MA (1) are significant along with constant. Both t-values and p-values reject the null hypothesis of market efficiency and declare the Karachi Stock Market as inefficient. Thus, the final equation before forecasting is expressed as,

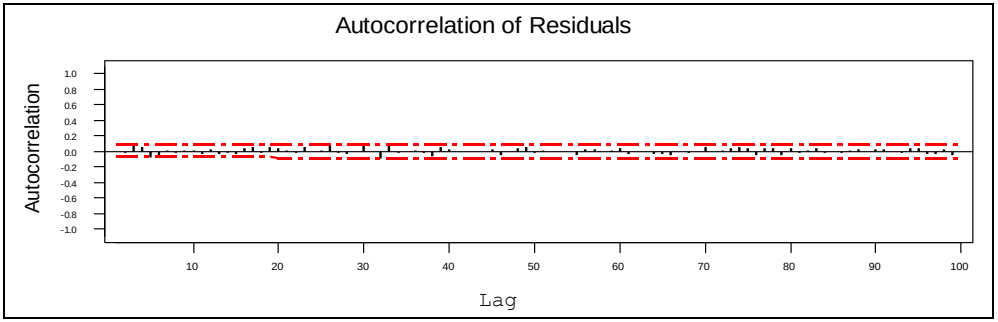
$$R_i = 0.15912 + 0.6288\, R_{t-1} + 0.4870\, \mu_{t-1}$$

Table 7
Final Estimates of Parameters for Best Fitted ARIMA (1, 0, 1) Model

Final Estimates of parameters				
Type	**Coefficients**	**SE Coefficients**	**T-values**	**P-values**
AR (1)	**0.629**	**0.161**	**3.890***	**0.000**
MA(1)	**0.487**	**0.181**	**2.692***	**0.007**
Constant	**0.159**	**0.073**	**2.183***	**0.029**
Mean	**0.429**	**0.197**		

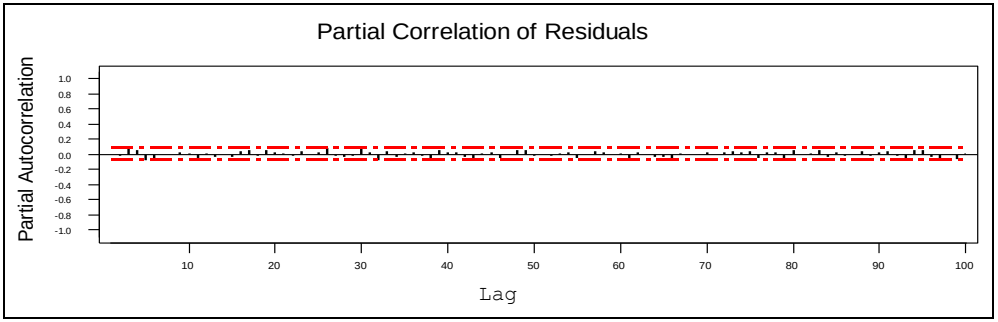
Source: Researcher's Calculation *Significant at 5%

A compulsory task which is necessary to be conducted is diagnostic checking. The aim of this task is to prove the validity the goodness of fit of the ARIMA model through the analysis of the residuals. If the residuals show stochastic behaviour means the computed ACF and PACF from are statistically equal to zero. Therefore, the residuals are indicating random behaviour and the model is showing a goodness of fit to the data undertaken. The spikes also lie in the confidence limits in the following figures.



Source: Researcher's Estimation

Figure 3: Autocorrelations of Residuals



Source: Researcher's Estimation

Figure 4: Partial Autocorrelations of Residuals

Table 8 shows the predicted 15 weeks a head of the volatility. Since, the whole data are 568 observations (weekly observation), so it is appropriate to choose the predicted value ahead for 13 to 15 observations, because ARIMA model adequate for the short term forecasts. The 95% prediction interval for the forecasts is also computed. The lower interval contains values with negative sign which can be ignored because the volatility is calculated by taking the absolute value into account. There are some periods of large and small fluctuations, an indication of clustering volatility.

Table 8
Forecasting Return Volatility 15 Weeks ahead at 95% Confidence Interval

Period	Forecast	Lower	Upper
566	0.121	-6.512	6.754
567	0.235	-6.464	6.935
568	0.307	-6.418	7.033
569	0.352	-6.383	7.088
570	0.381	-6.359	7.120
571	0.399	-6.343	7.139
572	0.409	-6.332	7.152
573	0.417	-6.325	7.159
574	0.421	-6.321	7.163
575	0.424	-6.318	7.166
576	0.426	-6.317	7.168
577	0.427	-6.316	7.169
578	0.427	-6.315	7.169
579	0.428	-6.314	7.170
580	0.428	-6.314	7.170

Source: Researcher's Calculation

6. CONCLUSION

The main aim of this paper is to predict the return for KSE 100 index using ARIMA model and to test the weak form of efficiency as well. Different parametric and non-parametric approaches like unit root analysis, correlogram test and ARIMA model are used to test the efficiency of the KSE market. The skewness and kurtosis measures for the stock return series declare that it is non-normal. The results reveal that Karachi stock exchange is inefficient in weak form. In this research study, all the models from ARIMA (0, 0, 0) to the model ARIMA (2, 0, 2) are tested, the best model is selected on the basis of the MSE, AIC and SIC. The convenient model that fitted the data reasonably well is ARIMA (1, 0, 1), at 95% confidence interval. The 95% prediction interval for the forecasts indicates that almost all periods show volatility in weekly index of Karachi stock exchange (KSE). There are some periods of large fluctuations like first three weeks and remaining periods show small fluctuations indicating clustering volatility, a peculiar characteristic of emerging markets. However, parametric tests strongly recommend that return series does not follow the random walk model and rejects the hypothesis of weak form efficiency. Overall, results from the empirical analysis powerfully propose that the KSE is inefficient in weak form. The observed findings reveal that the prevailed inefficiency in KSE could affect the decision making regarding the portfolios investment and asset allocation, trading in the financial markets, risk management conditions and information available to the investors. Also, the investors could use to predict and quantify about return volatility before making their trading decisions.

REFERENCES

1. Alnaa, S.E. and Ahiakpor, F. (2011). ARIMA (autoregressive integrated moving average) Approach to Predicting Inflation in Ghana. *Journal of Economics and International Finance*, 3(5), 328-336.
2. Al-Shiab, M. (2006). The Predictability of Amman Stock Exchange Using the Univariate Autoregressive Integrated Moving Average (ARIMA) Model. *Journal of Economic & Administrative Sciences*, 22(2), 17-35.
3. Al-Zeaud, Hussein Ali (2011). Modelling and Forecasting Volatility using ARIMA Model. *European Journal of Economics, Finance and Administrative Science*, 35, 109-125.
4. Assis, K., Amran, Y. Remali, A and Affendy, H. (2010). A Comparison of Univariate Time Series Methods for Forecasting Cocoa Bean Prices. *Trends in Agricultural Economics*, 3, 207-215.
5. Bartlett, M.S. (1964). On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series. J. Roy. Stat. Soc., B, 8, 27-41.
6. Beninga, S. (2008). *Financial Modeling*. Massachusetts London, The MIT Press Cambridge.
7. Box, G.P.E. and Jenkins, G.M. (1970). Time Series Analysis, Forecasting and Control. Holden Day, San Francisco.
8. Box, G.P.E. and Jenkins, G.M. (1976). *Time Series Analysis, Forecasting and Control*. 2nd ed. Holden-Day, San Francisco.
9. Brown, R.G. (1959). *Statistical Forecasting For Inventory Control*. New York, McGraw- Hill.
10. Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997). *The Econometrics of Financial Markets*. New Jersey, Princeton University Press, Princeton.
11. Cuthbertson, K. (1996). *Quantitative Financial Economics, Stocks, Bonds, Foreign Exchange*. John Wiley and Sons, London.
12. Elena. Kaloty and Storis K. StaiKuoras (2009). *Modeling Stock Market Volatility*. Greg N. Gregoriou Chapman Hall book, New York.
13. Fahimifard, S.M., M. Homayounifar, M. Sabouhi and A.R. Moghaddamnia (2009). Comparison of ANFIS, ANN, GARCH and ARIMA Techniques to Exchange Rate Forecasting. *Journal of Applied Science*, 9, 3641-3651.
14. Fama, E.F. (1970). Efficient Capital Markets, a Review of Theory and Empirical Work. *Journal of Finance*, 25,383-417.
15. Gujarati, D.N. (2003). *Basic Econometrics*. 4th ed Delhi, McGraw Hill Inc.
16. Holt, C.C., Modigliani, F., Muth, J.F. and Simon, H.A. (1960). *Planning, Production, Inventories, and Work Force*. Prentice Hall, Englewood Cliffs, NJ, USA.
17. Irfan, M., Irfan, Maria and Awais, M. (2010). Investing the Weak Form Efficiency of an Emerging Market Using Parametric Tests, Evidence From Karachi Stock Market. *Electronic Journal of Applied Statistical Analysis*, 3(1), 52-64.
18. Jenkins, G.M. and Watts, D.G. (1968). *Spectral Analysis and its Application*. Day, San Francisco, California, USA.
19. Ljunge, G.M. and Box, G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65, 297-303.

20. Makridakis, S., Anderson, A., Filds, R., Hibon, M., Lewandowski, R., Newton, J., Parzen E. and Winkler, R. (1982). The Accuracy of Extrapolation (time series) methods, Results of a Forecasting Competition. *Journal of Forecasting*, 1, 111-53.
21. Meese, R. and Geweke, J. (1982). A Comparison of Autoregressive Univariate Forecasting Procedures for Macroeconomic Time Series. Unpublished Manuscript, University of California, Berkeley, CA, USA.
22. Mostafaei, A. and Sakhabakhsh (2011). Modeling and Forecasting of OPEC Oil Prices with ARFIMA Model. *International Journal of Academic Research*, 3(1), 817-822.
23. Emenike, Kalu (2010). Forecasting Nigerian Stock Exchange Returns, Evidence from Autoregressive Integrated Moving Average (ARIMA) Model. Retrieved from *http://ssrn.com/abstract=1633006*
24. Pankratz, A. (1983). *Forecasting with Univariate Box-Jenkins Models*. Hon Wiley & Sons, New York.
25. Pindyck, R.S. and Rubinfeld, D.L. (1981). *Econometric Models and Economic Forecasts*. 2nd ed. New York, McGraw-Hill.
26. Purna, C.P. (2011). Forecasting International Tourists Footfalls in India, An Assortment of Competing Models. *International Journal of Business and Management*, 6(5), 190-202.
27. Quenouille, M.H. (1949). Approximate Tests of Correlation in Time- Series. *J. Roy. Stat. Soc.*, B11, 68-84.
28. Rahman, S. and Hossain, M.F. (2006). Weak Form Efficiency, Testimony of Dhaka Stock Exchange. *Journal of Business Research*, 8, 1-12.
29. Saeed, N. Saeed, A. Zakria, M. Tariq M.B. (2000). Forecasting of Wheat Production in Pakistan using ARIMA Models. *International Journal of Agriculture and Biology*, 2(4), 352-353.
30. Samnel, E. and Ferdinand, A. (2011). ARIMA (autoregressive integrated moving average) approach to predicating inflation in Ghana. *Journal of Economics and International Finance*, 3(5), 328-336.
31. Samuelson, P.A. (1965). Proof That Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, 6, 41-50.
32. Simons, D. and Laryea, S.A. (2004). Testing the Efficiency of Selected African Stock Markets. A Working Paper, 56.
33. Torben, D., Doberv, D. and Schaumbury, E. (2009). Jump-Robust Volatility Estimation Using Nearest Neighbor Truncation, *NBER Working Paper*, 15533.
34. Tseng, F.-M., Yub, H.-C. and Tzeng, G.-H. (2002). Combining Neural Network Model with Seasonal Time Series ARIMA Model. Technological Forecasting & Social Change, 69, 71-87.
35. Yule, G.U. (1926). Why Do We Sometimes Get Nonsence-Corrleations between Times Series. A Study in Sampling and the Nature of Series. *J. Roy. Stat. Soc.*, 89, 1-69.
36. Yule, G.U. (1927). On a Method of Investigation Periodicities in Disturbed Series, With Special Reference to Wolfer's Sunspot Number. *Phill. Trans.*, A, 226, 267-98.
37. Zahid, Faisal M., Ramzan, Shahla and Shumila, Ramzan (2012). A Parametric and Non-Parametric Approach for Testing Random Walk Behavior and Efficiency of Pakistani Stock Market. *International Journal of Science and Technology*, 2(5), 316-322.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 265-272

RELATIONSHIP BETWEEN THE STUDENT'S PREFERENCE AND THE DEMAND OF THE SUBFIELD IN MARKET

M. Ashar Burney[1], Syed Adil Imam[2], Wajahat Maqbool[3], Ammar bin Aamir[4]
Ali Kamal[5], Ali Raza[6], Ali Feroz[7] and **Waqas Ahmed[8]**
Department of Computer Science, National University of Computer
and Emerging Sciences, Karachi, Pakistan
Email: [1] k122056@nu.edu.pk
[2] k12261@nu.edu.pk
[3] k122198@nu.edu.pk
[4] k122074@nu.edu.pk
[5] k122017@nu.edu.pk
[6] k122173@nu.edu.pk
[7] k132311@nu.edu.pk
[8] k122193@nu.edu.pk

ABSTRACT

Our research aims to study the correlation between the preferences of the sub-fields of computer science graduates in Karachi city and currently the most widely employed subfield of computer science in the industry. Furthermore it examines and categorizes the subfields of applied and theoretical computer science on the basis of magnitude of its usage in the industries, banks, software houses and in technical institutes. The research is an extension of the ideas presented in the papers highlighting the imbalances between the preferences of students and the demanded fields. Previous researches have explored that out of 10 fields with most imbalances between graduates and the job openings, six of them relate to computer science. Another noticeable fact in the paper (Hsf-college to career in America) was that two third of the jobs created in next five years will be requiring high skilled workers such as college graduates to fill in the jobs. Our research is divided in two stages. First we interrogate the respondents from the corporate world and then we switch our focus on fresh graduates. We have a total of 15 respondents belonging to all age groups and genders from 10 reputed software houses categorized in gaming, networking, and development. Next we examined the trend of students in terms of their preference or interest in a particular subfield. We used survey based purposive sampling technique. The statistically collected data is analyzed using chi-square test. We conclude that market value of a sub-field has a significant impact on the student's preference of the field.

1. INTRODUCTION

The computer science is one of the fastest growing fields in the world with the evolution of faster and smaller designed chips, development of multi-tasking systems, the emergence of cloud computing. The rapid and widespread use of computers and information technology has generated a need for highly trained workers proficient in various job functions. These computer specialists include computer scientists, database

administrators, and network systems and data communication analysts. Job tasks and occupational titles used to describe these workers evolve rapidly and continually, reflecting new areas of specialization and changes in technology, as well as the preferences and practices of employers. Computer scientists work as theorists, researchers, or inventors. Their jobs are distinguished by the higher level of theoretical expertise and innovation they apply to solve complex problems and the creation or application of new technology. The areas of computer science field range from complex theory to hardware design to programming-language design. Some researchers work on multidisciplinary projects, such as developing and advancing uses of virtual reality, extending human-computer interaction, or designing robots. They may work on design teams with electrical engineers and other specialists. It is of great interest for the experts and the upcoming graduates to have a clue about the most popular sub-field of computer science. The job posts like search engine optimizer (SEO) is the result of recent developments in the field. The ever expanding world of computer science also is creating a lot of opportunities for fresh graduates to stream into the market. In the midst of the ever expanding field of computer science, students have been opened up with a lot of options to pursue a career in.

The research focuses to obtain data to study the relationship between the student's preference of the field and the most popular field in the market currently.

2. LITERATURE REVIEW

Chief Economists David Langdon, George, McKittrick, David Beede, Beethika Khan, and Mark Doms (July 2011) analyzed the STEM Education and STEM Jobs for the past and future 10 years by using the sample of public through population surveys. Growth rate of STEM and Non STEM Employment, Occupations in STEM, Availability of Jobs, Workers and their wages, STEM diploma holders, undergraduates, degree holders and their average earnings has been used as variables. Comparisons of wages, Regression analysis, American Community Surveys, calculations of the regression-adjusted earnings techniques are used. Analysis of data shows that the growing rates of STEM jobs are the main cause to economic validity. They are essential in the development of technological and scientific innovations.

In 2012 Hispanic Scholarship Fund (HSF) examined the STEM occupation; this report is an excerpt from HSF's full-length College to Career analysis that profiles science, technology, engineering, and math (STEM) occupations. The analysis looked at recent college graduates' chosen fields of study and how they relate to the career opportunities available in a range of STEM industries and occupations. This data was then compared to the projected number of entry-level job openings from 2012 to 2018 in the highest-demand occupations of this sector. College graduates and number of employments are used as the variables. Bar charts and pie charts are used as techniques for comparisons. According to HSF's research "computer systems analyst" will be in demand till 2018.

Daniel Costa (2012) investigated that Microsoft Corporation recently published report which examines the number of American college graduates in computer science and all of the available job openings in computer-related occupations between now and 2020 in

America. Shortage of workers in computer-related occupations, College graduates in Computer science, number of H1 visa allowed, number of STEM graduates has been used as variables. Linear regression approach has been applied which shows unemployment rate as the years progresses. Results have shown that in CS jobs opening less than one-fourth to less than one-half of workers in computing occupations have a computer science degree and the unemployment rate which Microsoft have provided is totally wrong as unemployment rate in America have never gone above 3.7. Once the unemployment rate for college-educated workers in computer-related occupations begins to approach the true measure of full employment for the occupation, it will make sense to debate whether new STEM green cards should be created and if H-1B levels should increase.

Ben Cover, John I. Jones, and Audrey Watson examines the STEM occupations its demands and employment occupation for guiding the youngsters about which field to choose by using the sample of Occupational Employment Statistics (OES) in May, 2009. Fields in demand, Employment and Occupation has been used as the variables. Bar Graphs used for simple comparisons, Bubble Chart and Data Mapping analysis have been applied. Result shows that the highest paying STEM occupation is "Natural Sciences Manager". The STEM occupations are increasing in the places where technology is evolving and these areas give high wages to their employees.

Michael Anft, Observed that the payment which technology workers get is fairly high then the Salaries which chemical, civil engineers get in America. This is due to the large number of job openings and less number of workers in IT department. Shortage of IT workers, salaries, number of jobs is the variables. Due to the shortage of IT workers private companies are forced to bring workers from different countries so that they can work on different projects but to bring them to America companies have to offer large sum of salaries and different packages which some times in the end becomes costly for companies. As a result companies now offer scholarships for high School graduates so that they pursue BS CS degree and companies expect government to do the same.

3. RESEARCH METHODOLOGY

We have used survey sampling method through structured questionnaire developed on the basis of our variables. Our Questionnaire is answered by a large number of respondents from different companies related with IT and also by fresh graduates who expressed their preferences. There is no discrimination between different age groups or gender. Our study variables are Student’s preferred or focused field and mostly widely used subfield, salary package, and most jobs available in a particular field. We have analyzed our variables using chi-square test. It is a statistical hypothesis test used to verify the hypothesis statements. In order to obtain the results of chi-square test, SPSS version 20 is used.

4. COMMENTS AND CONCLUSION

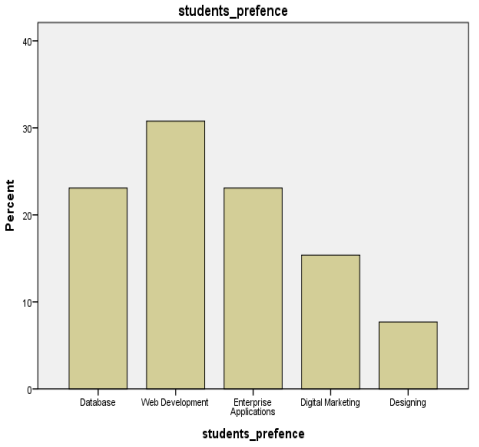
Database is the most demanded field till now our hypothesis was right and we obtain our result with the help of different charts and table which are describe below.

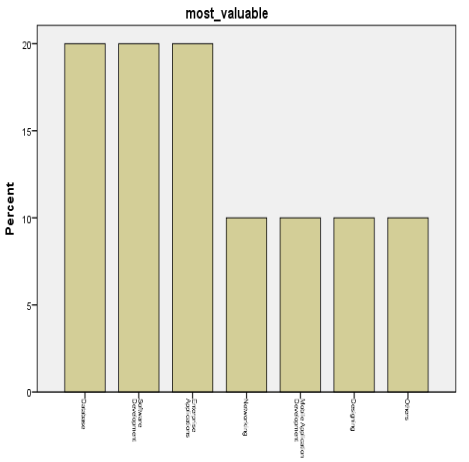
Student's Preference

	Frequency	Percent
Database	3	23.1
Web Development	4	30.8
Enterprise Applications	3	23.1
Digital Marketing	2	15.4
Designing	1	7.7
Total	13	100.0

The above table gives the Student's preference of the sub-fields. It can be seen from the table that most students prefer web development (30.8 %).

According to our survey student prefer Web Development more than Database, Enterprise Application, Marketing and designing. Above 30%+ of the student prefer Web Development and designing would prefer by few student. The most valuable field in the market is Database, Software Development and Enterprise Application which have 20% each respectively. Networking, Mobile Application, Designing and other fields are almost have same value that is 10% according to our survey graph.





Most Valuable Sub-field in Companies

	Frequency	Percent
Database	2	15.4
Software Development	2	15.4
Enterprise Applications	2	15.4
Networking	1	7.7
Mobile Application Development	1	7.7
Designing	1	7.7
Others	1	7.7
System	3	23.1
Total	**13**	**100.0**

The last table shows the frequency table of the sub-fields considered the most valuable by the companies. Database, Software Development and Enterprise Applications are equal in percentages. The above table also points out that digital marketing is considered the least valuable subfield by companies.

Jobs Available In Each Sub-field and Salary Package Offered By Companies

	Frequency	Percent
Database	4	40.0
Web Development	1	10.0
Software Development	3	30.0
Networking	1	10.0
Designing	1	10.0
Total	10	100.0

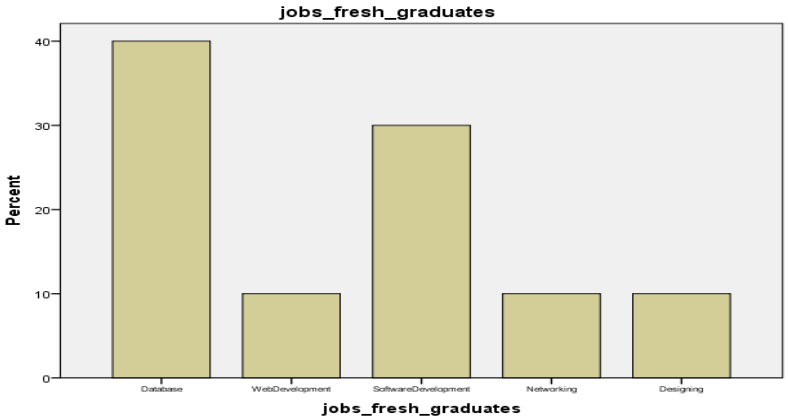
The above table gives the frequency table for the most jobs available in a particular sub-field in the industry. Table shows that the most jobs available are in fact in the database (40 %).

According to our survey, 40% of fresh graduates gets job in Database field and 30% gets job in Software Development. 10% of fresh graduates get job in Networking, Designing and Web Development. 60%+ companies preferred 25 to 30 thousand, 20% company preferred 31 to 40 thousand and less than 20% companies preferred 51 to 60 thousand salaries for newly IT graduated person according to our survey graph.

Salary Range

	Frequency	Percent
Rs.25-30 thousands	7	70.0
Rs.31-40 thousands	2	20.0
Rs.51-60 thousands	1	10.0
Total	**13**	**100.0**

It can be seen that most companies pay fresh graduates with in 25 to 30 thousand. 70% of the companies pay within this range.

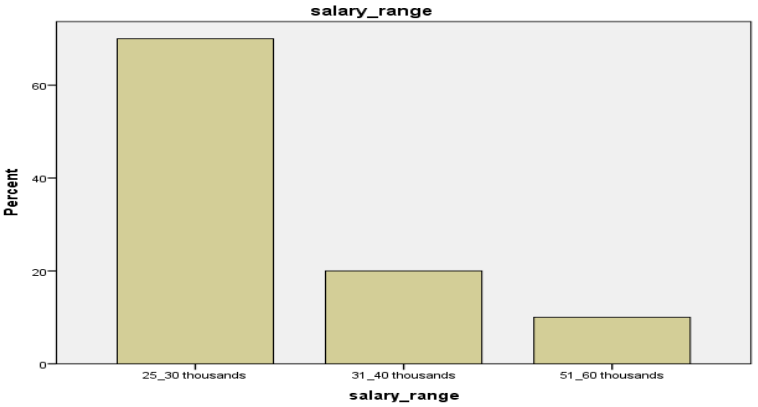


Hypothesis Assessment by Using Chi-squared Test

This table show the fields which are in demand in market and students preferences

students_prefence * most_valuable Crosstabulation

			most_valuable							
			Database	SoftwareDevelopment	EnterpriseApplications	Networking	Mobile Application Development	Designing	Others	Total
students_prefence	Database	Count	0	1	1	1	0	0	0	3
		% within students_prefence	0.0%	33.3%	33.3%	33.3%	0.0%	0.0%	0.0%	100.0%
		% within most_valuable	0.0%	50.0%	50.0%	100.0%	0.0%	0.0%	0.0%	30.0%
		% of Total	0.0%	10.0%	10.0%	10.0%	0.0%	0.0%	0.0%	30.0%
	Web Development	Count	1	1	1	0	0	1	0	4
		% within students_prefence	25.0%	25.0%	25.0%	0.0%	0.0%	25.0%	0.0%	100.0%
		% within most_valuable	50.0%	50.0%	50.0%	0.0%	0.0%	100.0%	0.0%	40.0%
		% of Total	10.0%	10.0%	10.0%	0.0%	0.0%	10.0%	0.0%	40.0%
	Enterprise Applications	Count	0	0	0	0	0	0	1	1
		% within students_prefence	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
		% within most_valuable	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	10.0%
		% of Total	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%	10.0%
	Digital Marketing	Count	1	0	0	0	0	0	0	1
		% within students_prefence	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
		% within most_valuable	50.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%
		% of Total	10.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%
	Designing	Count	0	0	0	0	1	0	0	1
		% within students_prefence	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%
		% within most_valuable	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	10.0%
		% of Total	0.0%	0.0%	0.0%	0.0%	10.0%	0.0%	0.0%	10.0%
Total		Count	2	2	2	1	1	1	1	10
		% within students_prefence	20.0%	20.0%	20.0%	10.0%	10.0%	10.0%	10.0%	100.0%
		% within most_valuable	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	20.0%	20.0%	20.0%	10.0%	10.0%	10.0%	10.0%	100.0%



	Value
Nominal by Nominal Phi	**1.671**
Cramer's V	**.835**
N of Valid Cases	**10**

CONCLUSION

In the light of the results of the chi-square tests, which shows a count of less than 5 suggests that there is a significant influence or impact of market value of field on students preference. Students generally are inclined towards the field which provides maximum opportunities to them. It also proves the point that students always try to remain connected with the market demands of the field even when they are studying. Through our survey we also came to know that many companies prefer fresh graduates to have expertise in database rather than any other field like networking, web development or other fields. However, most students were interested in both database and web development. Most companies were reluctant on paying the fresh graduates more than 40 thousand. We conclude that in the ever growing world of competition, students keep a track on the most popular field in the market to make their chances better of employability.

REFERENCES

1. Cover, B., Jones, J.I. and Watson, A. (2011). Science, technology, engineering, and mathematics (STEM) occupations: a visual essay. *Monthly Labor Review*, 134, 3-15.
2. Anft, M. (2013). The STEM Crises: Reality or Myth.? *The Chronicle of Higher Education*.
3. TIP Strategies (2012). *College to Career: High Demand Occupations: Science, Technology, Engineering and Math*. The Hispanic Scholarship Fund.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 273-278

A STATISTICALLY ANALYSIS ON THE EFFECTS ON STUDENT'S STUDIES USING SOCIAL NETWORKING SITES

Ali Shah[1], Maria Naveed[1], Ahmed Salim[1], Vishal Sharma[1], Vikash Kumar Batra[1] and **Fareeha Sultan[2]**

[1] Department of Computer Science, National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: k122211@nu.edu.pk k112046@nu.edu.pk
k122201@nu.edu.pk k122267@nu.edu.pk
k122072@nu.edu.pk

[2] Department of Humanities and Sciences, National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: fareeha.sultan@nu.edu.pk

ABSTRACT

Social media has undoubtedly managed to facilitate and promote learning all around the globe. However, as everything has its pros and cons, our aim is to therefore discover and understand the ultimate output consequences in view of the two factors respectively. Starting off with the analysis of social media's role in business enterprises and educational institutes, it indeed has a positive role in terms of promoting and speeding up business as well in being a tool for exchange of knowledge efficiently and massively: It increases peer to peer, instructor to instructor and instructor to student interaction in a virtual learning environment. In fact it is very systematic enough to encourage positive learning habits in students by facilitating them with easily accessible and updated knowledge. However, business in spite of flourishing is not in a position to be managed very efficiently as a lot of networking involved in it is not standardized yet. Similarly, in the domain of knowledge, it is playing a role as likely as in fulfilling the satisfaction of many people to be more social and increase their social contacts so that they become acceptable and recognized in the society. By taking the survey on the basis of convenience sampling, we also see that if there is a check and balance maintained in between the two areas i.e. socializing and academic orientation, and then it is possible to achieve a healthy output from students and consequently better conditions.

KEYWORDS

Impact on students; virtual learning environment; socializing and academic orientation.

1. INTRODUCTION

Ever since the old times, social media has remained as one of the important means of information sharing and communication around the globe. With the advancement of communication networks and technology, people have shifted from old primary methods of communicating to modern methods of communication. The real importance lies upon the fact that it is used to provide a communication link among masses nevertheless of

space, distance and time. One of the most commonly used social media sites are Twitter, Facebook, and LinkedIn which allow easy and instant networking means to people by providing file sending, information sharing and video and message sending facilities. However, provided these facilities, people are implementing ways of online learning and promotion of education and academics. Social media acts as a platform for teachers and students to connect with each other in a virtual learning environment and conduct proper online studies. In addition, social sites are playing a role in the shifting of the world in economic, political and other aspects.

Apart from the above mentioned benefits, social media is also playing a role as a way of making new acquaintances, sharing information and instant messaging. It allows people to connect with old and new friends and increase online socialization. People have therefore become addicted to its use with this regards as well. In other words, social media is currently a means of both socializing and learning. However, the direction that it's taking the youth towards is examined in this research paper. This is because; the young people are the ones who are most likely to get affected by its uses, both negatively and positively.

2. LITERATURE REVIEW

Many researches has been done on social media such as Raacke and Bonds-Raacke (2008) researched on the use of the internet as a communicating tool that has changed the way people interact nowadays. So they present a study to evaluate friend networking sites that what are the characteristic and what uses and gratification that are met by using such sites. This research was done by survey of questionnaire given to college students.

Catherine Dwyer, Starr Roxanne Hiltz and Katia Passerini (2007) investigate the influence upon trust of users on social sites and social interactions within them. The data collected regarding Facebook and MySpace showed comparisons of perceptions of trust and privacy issue along with willingness to trust information. Trust, privacy, social networking sites and hedonic information systems were used as variables along with Facebook and MySpace.

Christina Kakderi, Verena Hausmann and Aleksej Heinze (2013) examines that the popularity and use of social media is increasing in Europe, to help establish a new international business communication culture, which can aid in developing new business relations by the use of social media. Therefore, the idea of social network usage is based more on engagement, reciprocity and giving than direct selling or promoting of the ideas or products. Another characteristic of social media is interconnectivity which is important. As a multinational team, data was collected in all countries through a literature review. Focus groups with at least six SMEs in each of the eight countries, which led to the development of an online survey. The online survey was administered in 31 countries. The primary target audience of the survey was SMEs and students. The survey of students aimed to investigate uses of social media for non-professional purposes by younger generations of future managers.

Aine Dunne and Margaret-Anne Lawlor (2010) examine as to why young people use and participate in social networking sites (SNSs) with specific reference to Bebo. For source data a qualitative method is used here for exploring the uses and gratifications (U and G).

Lastly Waleed Mugahed Al-Rahmi and Mohd Shahizan (2013) Othman examine the use of social media of higher education as they are able to understand the purpose of education. The purpose of this study was to highlight the importance of social media in academic performance by peer to peer and collaborative learning. The research was carried out by mean of questionnaire. The result however was positive and it showed that the use of social media in academic performance is very high.

3. METHODOLOGY

The research data was collected by a questionnaire and interviews. The age range of the respondents is 17 or a little below till 35 and above. Also, the sample is comprised of 109 male and 41 female.

In this research the form was created on internet via Google docs and distributed to social sites users via internet. In addition, we used SPSS to analyses the data. Questions made were not only easy to understand but were also easy to answer.

This study proposes a discussion on the content in the framework for the impact on student's studies using social media sites. However, this research found that social media relates to the overall student's academic behavior. The variable used in this study is student's behavior on their studies using social media sites, impact of social media sites on students, and the ample time spent on social media. Similarly a conceptual framework for this study is applied for the student's activities of using such sites. We figured that the use of social media has increased communication with other students and teachers not only locally but has also globally enabled the student to create an interactive learning environment. Unlike the mentioned advantage, the use of social media has increased other activities of students thus distracting them from studying.

Social media has increased and made it easy to communicate with people whether it is just to contact with other people or to create a learning environment. It encourage student to have fun playing while creating a learning environment and it also creates a sustained behavior environment for students.

4. DATA ANALYSIS

Upon the questionnaire which was distributed the response given however concludes the results of our research paper. We build custom tables and charts (pie and bar) to analyze and conclude our results. The table (figure 1) shows the use of the social media with different student's field. The result shows that student mostly use social media for 1-2 hours. And major use for social media is from computer science student.

hours * Field Crosstabulation Figure 1

Count

		Field					
		Other	Computer science	Media science	Bussiness administration	Engineering	Total
hours	LESS THAN 1 HOUR	3	21	1	4	3	32
	1-2 HOURS	4	36	3	5	8	56
	2-3 HOURS	6	15	3	3	4	31
	MORE THAN 3 HOURS	2	23	2	2	1	30
Total		15	95	9	14	16	149

The results of the activities that are perform using social media sites are:

The table (figure 2) shows that the major use of the social media sites is to keep in touch with friends and moderately use for educational purposes.

Figure 2

		Gender	
		Male	Female
		Count	Count
Activities on social media	Making new friends	23	6
	keep in touch with friends	80	26
	downloading music and videos	25	12
	creating polls quizzes and survey	8	3
	sending suggestions and ideas to the websites	14	4
	time pass	57	13
	for educational purpose	40	17
	other activities	4	4

This shows that the use of social media regarding creating a learning environment is relatively low as we have seen the mean of educational purpose i.e. 38%

Furthermore, the results for the decline in spending time in extra-curriculum activities after the widespread of social networking sites show an increase. It is yet true as students are spending time in keeping in touch with friends and family, disconnecting them from the real world outside the social media. Thus it has cause a decline in time spent on extra curriculum activities. The result for the decline in spending time in extra-curriculum activities is 74.15% positive.

The social media can be useful or not. It totally depends upon the how a user uses it. As it can be used for 'time passes and other sorts of things but it can also be used for group discussion and other learning environment. Major of our response i.e. 105 out of 150 for this was that the use of social media depends upon the user.

Moreover majority of the students who are using social media sites are neglecting time while socializing the results in their overall academic performance as their important work are on pending. The result for yes was 61 and no is 21.

The table (figure 3) shows the cross tabulation of the age and the use of the social media for time pass. The result shows that almost 50% of the responses use it for time pass. Thus as mentioned above the students are using it for different purpose distracting them from studying and thus result major important work is left on pending.

Age * time pass Crosstabulation Figure 3

Count

		time pass		
		NO	YES	Total
Age	17 OR LESS	3	5	8
	18-20	56	40	96
	21-25	14	20	34
	26-30	2	1	3
	30-25	2	4	6
	36 OR HIGHER	3	0	3
Total		80	70	150

5. COMMENTS AND CONCLUSION

Social media no doubt started off as a means to ease and speed up communication around the world. With the increase in the trend of gaining knowledge and people's inclining attitude towards learning, it aided people as a tool for enchanting information and learning online: It increased interaction between students and teachers. However, as our research paper has a domain restricted to people and particularly students of Pakistan, we observe that overall there is a lacking of proper and serious approach toward using the social media. There is more inclination towards using it a means of time pass or free entertainment. In other words, the real purpose of using this facility is getting violated as we shall discuss further.

As it is evident from the above tables, social networking site are mostly in use by the computer science students. This is because such students' work is purely computer based. However, we observe that even such students in spite of their tremendous work load; spend an average of 1-2 hours on it. To explain it further, in other professions and domains, males are more likely to be dominant than females in our society. The table above shows that amongst males a numerical count of 80 uses the sites for making new friends, and the rest of the counts denote other purposes including educational purposes which has a count of 40. Therefore, after totaling up we find out that there is a more trend of using the sites for socializing and self-enjoyment purposes. Although the third table shows that people are 10 percent less likely to use social media for time pass purposes, still however this data does not put us into that much confidence as to people really are toward learning. This is because at the age of 18-20, people do tend to use it for both

slightly more for educational purposes as they are mostly students. However, this slight bigger difference of it doesn't inculcate us fully that people really use the social media for educational purposes because the overall trend is towards socializing and time pass.

REFERENCES

1. Dunne, A., Lawlor, M-A. and Rowley, J. (2010). Young People's Use of Online Social Networking Sites: a Uses and Gratifications Perspective. *J. Res. Interactive Mktg*., 4(1), 46-58.
2. Dwyer, C., Hiltz, S.R. and Passerini, K. (2007). Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. *Proceedings of the Thirteenth Americas Conference on Information Systems AMCIS* 2007, Paper 339.
3. Raacke, J., and Bonds-Raacke, J. (2008). MySpace and Facebook: Applying the Uses, and Gratifications Theory to Exploring Friend-Networking Sites. *Cyberpsychology & Behavior*, 11(2), © Mary Ann Liebert, Inc., DOI: 10.1089/cpb.2007.0056.
4. Aaltonen, S., Kakderi, C., Hausmann, V. and Heinze, A. (2013). *Social media in Europe: lessons from an online survey*, Worcester College, Oxford, UK.
5. Al-Rahmi, W.M. and Othman, M.S. (2013). The Impact of Social Media use on Academic Performance among university students: A Pilot Study. *Journal of Information Systems Research and Innovation*, 4, 1-10. ISSN: 2289-1358.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 279-284

STATISTICAL ANALYSIS ON THE ROLE OF GPA, QUALIFICATIONS AND SKILL SETS IN THE RECRUITMENT AND SELECTION OF AN APPLICANT (GRADUATE)

Humza Bin Javaid[1], Unais Qayyum[1], Ebad Nabi Hashmi[1], Arsal Asif[1], Akeed A. Farees[1] and **Fareeha Sultan[2]**

[1] Department of Computer Science, National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: k122302@nu.edu.pk k122323@nu.edu.pk k122274@nu.edu.pk k122277@nu.edu.pk k122281@nu.edu.pk

[2] Department of Humanities and Sciences, National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: fareeha.sultan@nu.edu.pk

ABSTRACT

The study is to show how the organizations commit the recruitment process and what factors play the biggest part. Nowadays, the students enrolled in educational organizations tend to lean off of their main objective, namely attaining skills that will serve the demands of the market. Instead, they get drawn towards the retention of Grade point average. To counter the numerous problems and loss of inclination that the student suffers due to this preposterous mentality, to call attention to what might be overlooked as the major source of social deprivation, drug abuse and even suicide and to understand and take advantage of the actual factors that matter whilst applying for a job at an organization. We are finding that organizations gravitate heavily on the skills, personality and other traits of an applicant rather than their GPA or qualifications alone. Moreover, these findings can be statistically manipulated to come up with a formulated outcome that may serve as a benchmark for the students. The aspects explored are the skills and techniques which are required to be a successful candidate in today's market environment. We hope to find how the many local organizations see and execute the induction process and what must be the actual objectives of the students in contrast to these findings. So, that the societal errors may perish and the market receives a brand of much more productive and effective employees

1. INTRODUCTION

This is our attempt to discriminate between the hiring criteria of the companies involved in today's market. Comparison of Grade Point Average requirements with respect to the companies' policies and or compulsions and any other skills and/or traits of an applicant. The main targets of our sample space were software houses and organizations that require employees of technical talents, mostly because 'we' ourselves are enrolled as computer science students. Therefore, finding statistical and numerical facts to counter the society's claims was our main objective.

To acquaint ourselves with the trends of today's market we made an excessive effort by studying published literature on recruitment strategies and techniques. Moreover, we visited as many as 50 organizations personally to conduct interviews so that our deductions may be fruitful and geographically accurate.

This research was conducted to foster the innovative mind-set of the students rather than allowing them to be ensnared in the thirst of obtaining a remarkable Grade Point Average. So, that they may emerge as the innovators of the new era.

2. LITERATURE REVIEW

Many researches have been done in this field by various authors, like Sheila M. Rioux, Ph.D. & Paul R. Bernthal, Ph.D. (1999). The Internet is the most popular advertising medium to recruit for managerial/ professional positions. So the authors headed on and studied organizations which are members of DDI HR Benchmark Group. They collected the data by asking organizations about their method of recruitment. Then they recorded this statistical data.

Jobs.ac.uk. (2013) the authors chose to publish a paper on the recruitment trends of last year. Their paper expressed the following as their findings in summary:

'lack of skilled/quality candidates, Cost reduction, Job boards continue to be the most effective recruitment tool for recruiters, Sector specialism, pricing and audience size, adoption of Applicant Tracking Systems in the sector is very low and cheap.' The researchers included the recruitment trends across academic, research, science and related sectors in their studies. Their survey included questions on recruitment priorities, recruitment challenges and the most effective candidate-attraction channels.

Ray French and Sally Rumbles (2010).56 per cent of organizations were focusing more on retaining than recruiting talent due to the inflation and/or economic regression. The others scrutinized various organizations based in UK. They used previously published researches and various surveys (CIPD, 2007c; IRS, 2008) to analyse their data.

Andrea Broughton Beth Foley Stefanie Ledermaier Annette Cox(2013).The framework of this research was based around the incidence of using social media tools for recruitment, Triggers for the use of social media in recruitment, Different tools and ways of using social media for initial candidate search. The authors studied the various methods by which Social Networking sites are used for the purpose of recruitment.

Kevin Lowden, Stuart Hall, Dr Dely Elliot and Jon Lewin (2011).The authors gathered qualitative insights from the employer case studies. They also scrutinized the graduate skills and attributes that are valuable to the employers. They also performed extensive research on the pre-existing literature.

3. METHODOLOGY

A sample size of 100 companies was chosen from local city. So, that the result generated may be geographically accurate and oriented towards the local trends. Consequently the observers of the research will be able to co-relate the results with the actual organizations and make better use of it. The organizations chosen as samples were

forwarded questionnaires via Google Docs on their respective email addresses and interviews have been taken from them. Indepth, 20 interviews were conducted with organizations. Where we met the HR heads of the organisations or personnel from the HR department to communicate with them and analyse their views on the current recruitment process of the company/organization.

We opted for convenience sampling and for the sake of finding inexpensive approximation of the facts we diverted towards the more convenient approach. We went forward and included Software houses, solution companies, universities, hospital etc. to study their recruitment methods. Questionnaires were made in contrast to the aspects that companies generally consider when hiring a graduate. These includes Grade Point Average (GPA), education, projects, skillset, age, gender, communication skills, and personality of an individual applying for a job.

Sampling frame was selecting a few companies of each type, for example, 5 software houses were selected, and 5 hospitals were selected. Descriptive statistics is used for statistical analysis.

4. DATA ANALYSIS

After the collection of data we translated the data into a state where it was readable by SSPS. Then we applied some functions and got the desired output in the form of pie charts and bar graphs. We can observe in **Fig 1.0**, the priority given to each facet of the criteria. The table demonstrates that management skills and basic skill set (as per the job's requirement) are preeminent compared to the other assets. As **Fig 1.0** suggests, management skills have a large a larger portion of the percentage in regards to priority. Whereas, qualifications rank the second last in terms of priority, with a meagre 8.2 share of the total percentage. We can gather that an ideal employee or candidate for a job would be a person with splendid management skills and basic skill set (as per the job's requirement).

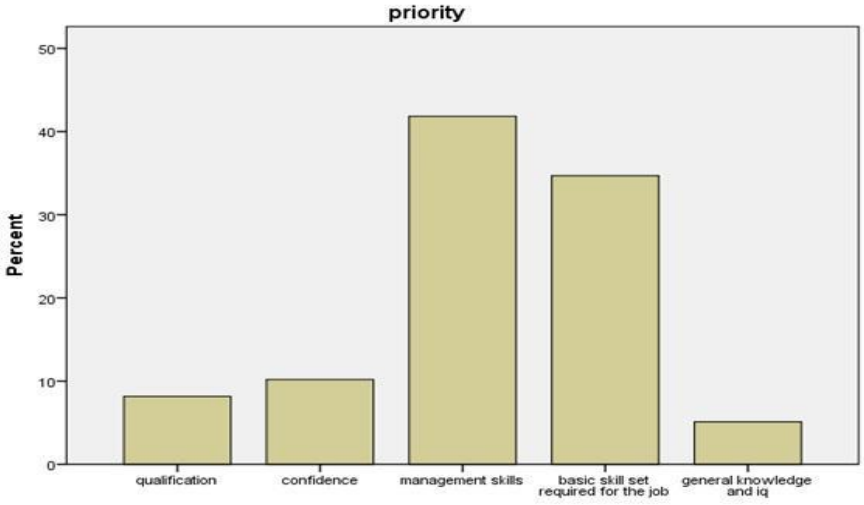


Fig. 1.0

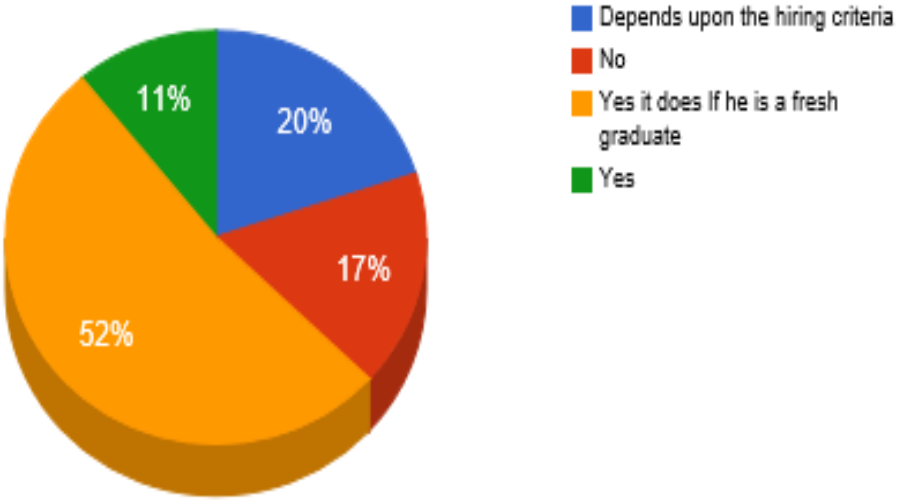

11%
20%
17%
52%
Depends upon the hiring criteria
No
Yes it does If he is a fresh graduate
Yes

Fig. 2.0

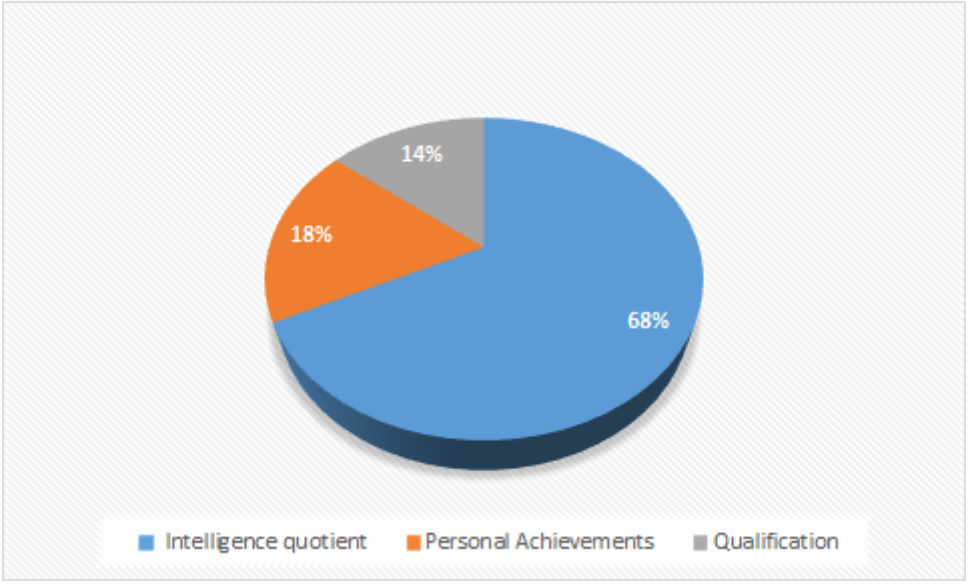

14%
18%
68%
Intelligence quotient
Personal Achievements
Qualification

Fig. 3.0

The preference of GPA in the hiring criteria is displayed by **Fig 2.0**. We can observe that GPA matters most when recruiting a fresh graduate. This is mainly because there is no other credible source to express the skills and/or experience, as well as the intelligence quotient of the candidate. The claim that is made by **Fig 2.0** is also supported by **Fig 3.0**. Intelligence quotient is of the highest priority in the recruitment factor according to **Fig 3.0**. This is so because the organizations need quick learners to grasp the concepts. So, they could readily participate and take initiative in their projects or activities.

Surprisingly enough, we found out through personal interviews that if the applicant is a dropout, but is surprisingly skilled and fulfills the job requirements, they may be willing to consider him if there is no fixed hiring criteria from the officials of the organization. But, the applicant will only have a future in this particular organization. Although the experience he gains from this particular job might increase his eligibility for vacancies at other organizations.

5. CONCLUSION

The mechanisms used to select graduates for various positions vary from one recruiter to the other and it depends on the type of job being offered. Commonly used mechanisms and tools include face-to-face interviews, internet, agencies, and tests, among others. The interests of employers during the graduate recruitment and selection process vary from employer to employer, and depend on the type of vacancy. Within these factors are varied aspects such as skills, qualifications, knowledge, experience, organizational capabilities, team spirit, and resilience, among others. To this purpose, graduates should prepare extensively when called for interviews by being committed to the needs and requirements of the recruitment and selection process. Any misunderstanding should be communicated and cleared on spot.

It is very important for a graduate to concentrate on improving his skill set along with maintaining his grades, because in many organizations it is often observed that your skill set grabs more attention than your grade point average, in past we surveyed some companies and software houses, hospitals have also been the part of our survey.

What we have gathered from the survey of hospitals is that they paid a bit more attention on the grades rather than the skills. This was mainly because the skills of the employees are fostered under the direct supervision and command of experienced doctors and staff members. We have also visited software houses, and what we have gained from them is that they were paying more attention on applicant's skill set and experience. Most of them consider the applicant according to the working skill set and past experience. They also highlighted that applicant's university project would be more attractive for the HR department. No gender discrimination was observed in their organizational culture and recruitment methods. Once employed, all candidates are treated the same way, regardless of their skill-set or GPA.

6. ACKNOWLEDGEMENTS

We’d like to be thankful to all the contributors who ever so willingly aided our research by providing us data and detailed insight on the matter. Moreover, we are thankful to our teacher for providing us the proper support and helping us throughout the

process of the research and analysis phase. Lastly, we are thankful to DOW University for providing a platform where we can present the result of our hard work.

7. REFERENCES

1. Broughton, A. Foley, B., Ledermaier, S. and Cox, A. (2013). *The use of Social Media in the recruitment process*, Acas research publications. ISBN 978-1-908370-32-7.
2. jobs.ac.uk(2013), *2013 Recruitment Trend Survey Results*.
3. Lowden, K. Hall, S. Elliot D. and Lewin, J. (2011). *Employers' perceptions of the employability skills of new graduate*. Published by Edge Foundation. ISBN 978-0-9565604-3-8.
4. French, R. and Rumbles, S. (2010). *Recruitment and Selection*. CIPD, Publications, London, pp. 169-190. ISBN 9781843982579.
5. Rioux, S.M. and Bernthal, P.R. (1999). *Recruitment and Selection Practices Survey Report*, Volume 2: (1998–1999) Issue 1–Succession Management Practices Survey.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 285-292

A STATISTICAL ANALYSIS ON THE REASONS OF CHANGING THE TELECOMMUNICATION NETWORK BRANDS

M. Abeer-ur-Rehman[1], Shayan Ahmad[1], M. Imran Rafique[1], M. Maaz Shams[1], Syed Usman Nizami[1], Wasif Ali Khan[1] and **Fareeha Sultan[2]**

[1] Department of Computer Science, National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: k122130@nu.edu.pk k122170@nu.edu.pk
k122134@nu.edu.pk k122142@nu.edu.pk
k122132@nu.edu.pk k122152@nu.edu.pk

[2] Department of Humanities and Sciences, National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: fareeha.sultan@nu.edu.pk

ABSTRACT

Telecommunication networks are one of those needs without which, one may find it pretty difficult to connect with his/her circle of family, friends, work or office. People of Pakistan use Telecommunication Networks as a very fundamental tool to connect with the world. But on the same hand what are the possible problems because of which people switch their respective brands. We hold the regard to understand and evaluate the main, current and exclusive status of the telecommunication companies who provide their services in different forms to its users. Some hypotheses through which we conducted our research are when People choose their telecommunication network according to their field of work. On the other hand people watch advertisements and get attracted towards a Telecommunication Network. As another option that people find disturbances in the quality of service that compels them to change to another favorable network. The main objective of our research is to give the analytical reasons and possibility of a choice when one switches the brand to another and if not as in a non-common case on the other hand, then what leads to their satisfaction for such a brand advertisements in different forms are observed every day. For those reasons of either staying behind the table or turning one against the other, can extract the general thinking of the public where the companies such as Warid, Zong, Telenor, Ufone, & Jazz (Mobilink) persist to have their user banks. For the major cities of Pakistan that we know, this research and its survey, which is based on questionnaire is conducted from more than 200 convenient sample, in the atmosphere of the biggest city of Pakistan, Karachi, where brands tend to serve the best service and facility to their users. This research will provide a statistical overview of the used networks in different fields, ages and interest.

1. INTRODUCTION

It is important to understand the basic information about the switching process to understand why customers switch from one brand to another. Customers have very important place for the companies. Market conditions have become very competitive and companies are trying to capture the shares of each other. Due to the intense competition

in telecom industry customer frequently switch from one brand to another. This research is going to investigate the brand switching phenomenon in perspective of telecommunication industry.

Success of organization depends on satisfaction of customers over the brand that is why customers are very important player for the market. Customers have very different attitudes, perceptions and behavior towards the brand, customers often switch their brand depending upon various factors. Brand switching occurs when customer switch their loyalty from one brand to another this shift from one product to another is called brand switching behavior of customers.

Brand switching can be of two types that is, "temporary" and "permanent". If cigarette A is not available in a particular shop then the customer can shift to cigarette B, this is temporary brand switching it can be long lasting if the customer shift to particular brand permanently. Permanent brand switching is harder to change. Brand switching is basically a process by which the customers choose to switch a product or brand of daily use to a product or brand of similar nature.

Brand switching behavior of customers differs in different industries. Some industries are characterized by high brand switching rates and some are low brand switching rates. Telecom industry is in one that possesses the high rate of brand switching from customers. This is mainly due to the intense competition in telecom industry. This frequent switch from one brand to another of customers has become very important for the service providers to investigate the switching behavior. The study is focused on telecom industry of Pakistan. Five major players are playing the roles in this highly market switching brand competitions Mobilink, Ufone, Zong, Telenor and Warid.

2. LITERATURE REVIEW

Muhammad Asif Khan (2010) examines the dimensions of users' perceived service quality of cellular mobile telephone operators in Pakistan from survey based form data. Tangible, reliability, responsiveness, assurance, convenience, empathy, network quality, and perceived service quality are the variables that have been used. Instrument Development, Pilot Testing of Instrument, Sample and Data Collection, Reliability and Validity of Instrument, Data Analysis, Results and Analysis are the methods used. The results of this study concur with the findings of earlier researches in mobile phone industry that the dimensions discussed in the study have positive and significant effect on mobile phone users' perception of service quality.

Jawaria Fatima Ali, Imran Ali, Kashif-ur-Rehman, Ayse Kucuk Yilmaz, Nadeem Safwan and Hasan Afzal (2010), analyzes the determinants of consumer retention in cellular industry of Pakistan. Call clarity, User friendliness, Value-added services, Customer support services, Customer complaints, Variety of price, Customer satisfaction are the variables of this very study. Data and analysis, Sample, Instrumentation and measurement are the research methodology used. The study has found significant relationship between price reasonability and consumer satisfaction, call clarity is also found to be having significant relationship with customer satisfaction and consumer retention, significant and strong relationship was observed among customer satisfaction and consumer retention.

Rashid Saeed, Rab Nawaz Lodhi, Muhammad Saad Saeed, Zeeshan Fareed, Amna Sami, Fareha Dustgeer, Zahid Mahmood and Moeed Ahmad (2013), analyzes the factors behind brand switching in Telecom industry of Pakistan, in the city of Sahiwal. Price, Inconvenience, Service Failure, Brand Switching are variables used. Data Collection, Data Analysis are the methodologies that has been used. The study of this research shows that people within Sahiwal city switched their brands when they found lower price rates of services used, moreover, it is also concluded or predicted about the people of Pakistan to switch their networks if they find inconvenience in the provision of services while they won't switch to another network if there is a service failure.

Muzammil Hanif, Sehrish Hafeez, Adnan Riaz (2010), examines the factors affecting the customer satisfaction in the telecom sectors to conclude the generalized point of customer satisfaction for any brand. Age, Gender, Type of connection, network, experience with network, are the variables used for this research. Methodologies such as population, procedures, and measures through questionnaires are used. This respective study seeks through the factors, that independent variable influences as well as compliment dependent variables, price fairness plays a very immense role in the customer satisfaction, which concludes that any brand in the market can be successful with fair tariffs and memorable customer services which will eventually make them reign over the market for a longer period of time.

3. METHODOLOGY

3.1 Population

Karachites were treated as the population of this study. By using the convenience sampling technique we have taken the sample size of 356 users of telecommunication services and all the GSM mobile phone operators such as U-fone, Telenor, Mobilink (Jazz), Warid and Zong were taken into consideration.

3.2 Procedure

Questionnaire in the form of e-form (Google form), hard copy and interview survey was considered to be effective enough in the research. Data was analyzed using Descriptive Analysis and Chi Square test in SPSS. Questionnaires were floated among offices, students of University and College, teachers, workers, family and friends. Data analysis and Chi square tests are the statistical techniques used to evaluate the end result.

3.3 Measures

A brief description about the purpose and significance of the study was put into a questionnaire which was used for this analysis. The first part of the questionnaire consisted of questions which would analyze the demography of the customer such as age, occupation, type of connections used and how many reserved networks and for what purpose they are used.

The second part observed the customers satisfaction, fairness price and service quality. Special emphasis was put upon the decisions of why the customers choose that network and why was it necessary for network switching. Many samples show that the candidates are not aware of all the hidden charges against the network but still carry on it. At the end of this part, the customers about the reason of network switching and about the problems/irritation of the current network.

Third part considers the suggestion which the audience prefers in a network and which network is the best for their field in their opinions.

3.4 Hypothesis

Hypotheses-1: People choose their telecommunication network according to their field of work.

Hypotheses-2: People find disturbances in the quality of service that compels them to change to another favorable network.

4. DATA ANALYSIS

Through statistical research it was known that among 356 users, the users used Zong 12%, Mobilink 15%, Telenor 22%, Ufone 37%, Warid 14%.

Customers preferred qualities that should be in a good network and the most emphasized quality is "service quality" which is 41%, "call packages" are 13% as it is the most attracting and appealing thing in a network, "call charges" is found to be 12% as it is something that is considered essential while choosing a network. "Sms packages" is known to be 9% whereas "sms charges" is 6%. Value added services play an important role to maintain a good customer relation, therefore, it was known to be 6%, whereas GPRS charges 2% and GPRS packages were evaluated as 5%.

The graph shows that Ufone has the highest number of users of age 14-18 which is 37%, while Telenor has 19%, zong 17%, warid 16% and mobilink 12%.

Users of different ages use different SIM's. Chi Square test value was 0.00 which proves the hypothesis of age and usage dependency. Customers of the age limit 19-25 have the highest users of Ufone that is 45%, Telenor has 28%, warid has 12%, mobilink has 10% and zong has 5%. Customers of the age limit 26-35 has the highest users of Ufone which is 32%, while Zong as 20%, Warid 19%, Telenor 14%, and Mobilink 15%. 30% of the customers in the age limit 36-50 used Ufone, 17% Zong, 27% Mobilink, 15% Telenor, 11% Warid. 23% of the individuals of age above 50 used Warid and Moblink while 18% of the same population used Zong, Telenor and Ufone.

Customers, who use telecommunication network, belong from different occupation and field of work. This research shows that people choose their networks according to their field which shows the dependency factor and the value of chi square test calculated is 0.00. The following table 1.1 proves our first hypotheses.

Table 1.1
Usage of Networks According to the Field

Network	OCCUPATION						
	Student (College)	Student University	Teacher	Office	Business-man	Other Occupation	Total
Mobilink (JAZZ)	3	9	4	12	6	19	53
Warid	4	20	6	11	7	3	51
Ufone	13	65	10	18	10	17	133
Telenor	9	39	5	10	5	9	77
Zong	5	6	7	9	4	11	42
Total	34	139	32	60	32	59	356

Hidden charges create an important impact on the user's decision to switch to another. This was supposed to be dependent upon the network currently in use, but these factors are found to be independent on each other, the value of chi square test is 0.34 and the bar chart as shown in the chart 1.1 shows its in-dependency.

The Chi Square test showed result as 0.00 % when was equated between best services known by the people about their respective telecommunication networks and the networks they currently use, chart 1.2 shows the relationship.

Observing the surveys, people tend to use their reserved/alternate SIM's very much for different reasons. Such that it was found that Mobilink (JAZZ) network has 32% of the people who use it because of personal privileges, 2% because of business convenience, 8% TV Advertisements, 4% when fed up with the one in use, 17% because they found attractive packages, 4% use it because of Free Winnings, 2% Lucky Draws, 8% Others. However, 25% have no reserved/alternate SIM's at all, they used Mobilink (JAZZ) and are comfortable with it.

Telenor users were known using its network as 24% with personal privileges, 0% for business convenience, 0% because of TV Advertisements, 2% when fed up with existing one, 27% found attractive packages, 7% with Free Winnings, 2% for lucky draws, 7% for Other reasons. However, 29% of them used Telenor only.

Ufone is used by 28% because of personal privileges, 0% for Business Convenience, 9% because of TV advertisements, 8% when fed up with existing one, 6% found attractive Packages, 0% free winnings, 6% for lucky draws, 0% for Other reasons. However, 43% used Ufone only.

Warid users were found 16% with personal privileges, 2% with business convenience, 14% for TV Advertisements, 2% when fed up with existing one, 20% found attractive packages, 5% because of free winnings, 2% with lucky draws, 7% Others. However, 32% of them only used Warid.

Zong has 20% users with personal privileges, 2% with Business Convenience, 9% because of TV advertisements, 7% when fed up with existing one, 20% found attractive Packages, 2% with Free Winnings, 2% for lucky draws, 7% for Other reasons. However, 31% used Zong only.

People were also found with the reasons because of which they put on their alternative SIMs. Majority of the surveys present that Ufone network is the most preferred reservation SIM for it has a 25% statistics, while Mobilink (JAZZ) has 15%, Telenor has 12%, Warid has 12% and Zong has 15%. The remaining 21% are those who do not keep reserved/alternate SIM(s) and are very much comfortable with the one they are using.

A lot of people have their own choice for using which network's SIM. Many of them used SIMs because of their family and friends while others use according to their specific field. Businessmen tend to use 22% Mobilink (JAZZ), 34% Ufone, 16% Telenor, 16% Warid and 12% Zong.

23% Officers tend to use Mobilink (JAZZ), 34% Ufone, 12% Telenor, 13% Warid and 18% Zong. Teachers tend to use 12% Mobilink (JAZZ), 34% Ufone, 19% Telenor, 13% Warid and 22% Zong. 12% University Students use Mobilink (JAZZ), 47% Ufone, 27% Telenor, 11% Warid and 3% Zong.

12% College Students preferred to use Mobilink (JAZZ), 41% Ufone, 23% Telenor, 9% Warid and 15% Zong while 24% of others with some different occupation or field used Mobilink (JAZZ), 39% Ufone, 13% Telenor, 7% Warid and 17% Zong.

Table 1.3 shows the relationship between the first network used by the users and current network used by them. Table 1.4 shows the relationship between the first network used by the networks and switching reasons, the dependency of the table can be observed from table 1.4 which shows chi-square value as 0.01.

Table 1.3

First Network	Current Network					
	Mobilink(Jazz)	Warid	Ufone	Telenor	Zong	Total
Mobilink(Jazz)	44	7	22	9	7	89
Warid	2	31	12	3	6	54
Ufone	3	6	83	15	10	117
Telenor	3	4	14	47	6	74
Zong	1	3	2	3	13	22
Total	53	51	133	77	42	356

Table 1.4

First Network	Switch Reasons									
	Call Charges	SMS Charges	GPRS Charges	SMS Packages	Call Packages	Internet Packages	Service Packages	Free Winnings	Null	Total
Mobilink (Jazz)	24.0	5.0	4.0	4.0	16.0	2.0	11.0	5.0	18.0	89.0
Warid	6	3	2	2	5	1	13	3	19	54
Ufone	19	14	6	5	11	1	15	3	43	117
Telenor	7	13	4	6	7	0	9	4	24	74
Zong	4	3	2	0	0	3	4	1	5	22
Total	60	38	18	17	39	7	52	16	109	356

5. FINDINGS AND CONCLUSION

It was clearly known after the survey, interview and evaluation that most of the people did choose their telecommunication network because of their field of work or occupation. Some of them used it for different other reasons known by themselves. Most of them also gave a general idea why they liked their SIMs they are using currently and the most known from them were SMS packages and Service quality. People switched because of different disturbances they felt, considering the top two parameters responsible for switching were known to have 37% of the entire telecommunication network users switching because of the hidden charges and 23% of them switching because of the irrelevant messages from the companies for different promotions. Service

quality was therefore preferred as the reason for switching brands by the people. That proves our hypotheses through analysis.

This paper is prepared in the most populated and one of the most important cities of Pakistan, Karachi, to analyze the usage of different networks by with respect to different factors by comparing them with one another. Many researches have been made of this study and have succeeded in a lot of valid points.

We used questionnaires and quantitative approach for the collection of data from different occupants and areas within the city. Age group and various towns were taken into serious notice, due to the fact of having many perspectives from different people. After all the hard work put into this research, a maximum number of 356 surveys could be made due to time constraint and study issues. Multiple regression and correlation were used to test hypothesis. Our calculations and evaluations of all the surveys conducted show that Service Quality is the most common choice selected by the audience as the best feature in their current network which is in use. Call charges was an average feature among the users. On the other hand, GPRS charges and Free Winnings were the worst features in the audiences' view, as some of them said that these features were not the facilities most needed for communication among family and friends.

6. LIMITATION OF THE STUDY

This study has been taken from Karachi, southern Sindh, Pakistan and the population is telecommunication network users in Karachi city. The research covers only 350 subscribers to get the conclusion. The size of sample is very small due to time constraint, but a larger sample than the current result will be more significant.

CHART 1.1

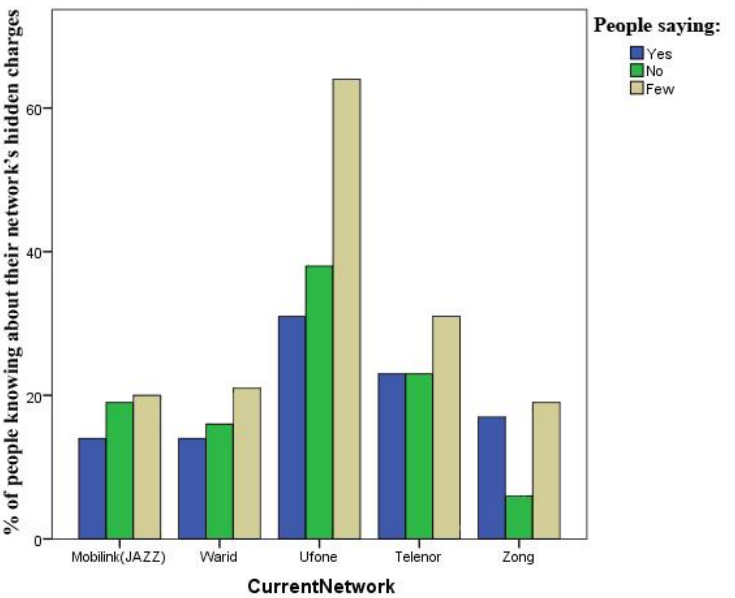
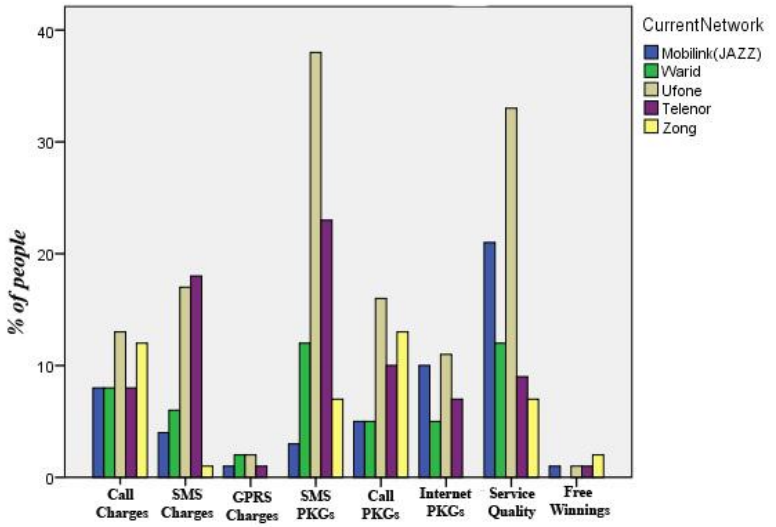


CHART 1.2



BEST THINGS KNOWN BY THE PEOPLE ABOUT THEIR NETWORK

7. ACKNOWLEDGEMENT

We are highly thankful to our teacher Ma'am Fareeha Sultan who lead us through the guidance of this respective research of Probability and Statistics, the management of the DUHS that gave us the opportunity to participate into the conference, and last but not the least, our respectful university, FAST-NUCES, that gave us the permission to participate into the conference and provided us with necessary facilities.

8. REFERENCES

1. Anderson, E.W., Fornell, C. and Mazvancheryl, S.K. (2004). Customer Satisfaction and Shareholder Value. *Journal of Marketing*, 68, 172-185.
2. Herrmann, A., Xia, L., Monroe, K.B. and Huber. F. (2007). The Influence of Price Fairness on Consumer Satisfaction An Empirical Test in the Context of Automobile Purchases. *J. Prod. & Brand Mgmt.*, 16(1), 49-58.
3. Ali, J.F., Ali, I., Kashif-ur-Rehman, Yilmaz, A.K., Safwan, N. and Afzal, H. (2010). Determinants of consumer retention in cellular industry of Pakistan. *Afr. J. Bus. Mgmt.*, 4(12). 2402-2408.
4. Khan, M.A. (2010). *An Empirical Assessment of Service Quality of Cellular Mobile Telephone Operators in Pakistan*. Canadian Center of Science and Education.
5. Hanif, M., Hafeez, S. and Riaz, A. (2010). Factors Affecting Customer Satisfaction. *International Research Journal of Finance and Economics*. 60, 44-52.
6. Saeed, R., Lodhi, R.N., Saeed, M.S., Fareed, Z., Sami, A., Dustgeer, F., Mahmood, Z. and Ahmad, M. (2013). Factors Affecting Brand Switching in Telecommunication, A Case Study of Sahiwal District, Pakistan. *World Applied Sciences Journal*, 26(5). 667-671.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 293-300

IMPACT OF SMARTPHONE'S ON UNIVERSITY STUDENTS: A STATISTICAL ANALYSIS

Vikash Lal Dodani[1], Avinash Kumar[1], Paras Kumar[1], Ahsan Islam[1]
Muhammad Noman[1], Sarfraz Ahmed[1] and **Fareeha Sultan[2]**

[1] Department of Computer Science, National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: k122118@nu.edu.pk k122128@nu.edu.pk
k122220@nu.edu.pk k122112@nu.edu.pk
k122120@nu.edu.pk k122164@nu.edu.pk

[2] Department of Humanities and Sciences, National University of Computer and Emerging Sciences, Karachi, Pakistan
Email: fareeha.sultan@nu.edu.pk

ABSTRACT

Smartphone is one of the best inventions of this era. It enables people to bring the world in their pocket. People can remain in touch anywhere through this small device. No need to have a monitor screen and then connect it with a CPU, use mouse and keyboard to open a browser for browsing or social networking etc. Instead you can do this with a single touch of your finger just because of Smartphone. Smartphone has a great impact on our society which can be categorized as positive and negative. Smartphone has changed the cultures and living style of people across the world. Few researches have been carried out investigations about the impact of Smartphone on society and several articles have been written in this regard but this study will primarily focus on impacts of Smartphone on students and will find that how students use this technology and alarm the society about the negative aspect of smart phones. Convenience Sampling was used to carry out this research in which a survey was conducted through questionnaires, interviews and other sources to Smartphone Students and they were asked to answer some questions. From this research we analyzed that smart phones have both positive and negative impacts but it has a great negative impact on the students than the positive.

KEYWORDS

Smartphone; invention; impact.

1. INTRODUCTION

The Smartphone's are the leading devices taking the front end and playing the role of universal mobile terminal. A Smart phone is a phone with more advanced technology, capability of connectivity and computing than the basic features of normal phones, like it is capable of photo shoot, play games, play videos, navigation, built-in camera, audio/video playback, recording, send/receive e-mail, built in apps for social web sites, wireless Internet and much more. Smart phones were not in the thoughts till earlier this century; the first Smartphone was discharged by IBM in 1993. The first true smart phone was made in 2002 named as BlackBerry 5810. This mobile phone/PDA had e-mail, net

capabilities, calendar, address book and different applications. While, Apple's iPhone, discharged in 2007, and use of Smartphone is increasing day by day. From a simple calculation to social, Smartphone is a leading device in this modern period, while Smartphone is a useful technology and it help people in everyday life but it also has negative effects especially on students. This research paper consists of introduction, literature review, methodology, data analysis, conclusion, acknowledgement and references.

2. LITERATURE REVIEW

To know about the development, traffic and positive and negative impact of smart phones, researches have been done by various authors, like, Mohammad Sarwar and Tariq Rahim Soomro (2013), analyzes the key contextual factors for the supporting innovations in the use of smart phones by using the sample of the six major countries (USA, France, Germany, Italy, Spain, and United Kingdom) on line surveys. Smartphone, Mobile Applications, Social Impact, Addiction, Health has been used as variables. The positive and negative impact comparison of using the Smartphone has been used in this research paper. There are several ways to control and minimize the negative impacts of Smartphone in society by Education and Guidance. It is recommended that, in order to understand the positive and negative impact of Smartphone it is very important to educate the users on how to use Smartphone smartly. The education should emphasis to enhance the positive impacts and highlight the negative impacts clearly so that the users can take advantages of this exciting technology.

In addition, Hossein Falaki, Dimitrios Lymberopoulos, Ratul Mahajan, Srikanth Kandula, Deborah Estrin (2010) analyzes the detailed look at smartphone traffic. Browsing, media, maps and email has been used as variables. Sniffers were employed on the devices to record all data sent and received. Results show that browsing contributes most traffic and packet loss is the primary bottleneck for traffic throughput. They analysis points at several simple mechanisms that can improve the efficiency, performance and power consumption of Smartphone communication. Aggregating multiple small transfers through an in-cloud proxy can reduce the overhead of small transfers; increasing the socket buffer sizes at servers can improve throughput. In the future, the author had a plan to investigate in the detail the efficacy of these mechanisms.

Furthermore, Hendrik Mueller, Jennifer L. Gove, John S. Webb (2012) research on the use of the tablets and the purpose of their research was to find the use of the tablet in today's I for that they interviewed many people from different I houses and they found out that the use of tablet is very vast and they concluded that those who have tablets use it very frequently and everywhere in bed, while eating food.

Moreover, Andrew Bibby (2011) analyzes the development of tablet from researches and surveys. Tablets, iphones II, ipad, Apple, Samsung, Dell, Motorola, Blackberry and Android have been used as variables. The survey of tablet use have been conducted in the US by Change Wave Research that which brand has more demand They included top brands and survey on them. There survey shows that Apple has 80%, Samsung has 7%, Dell has 5%, Motorola has 4% and Blackberry has 3% demand of tablets. So Apple has more demand among all.

3. METHODOLOGY

This research tried to find the negative impact on students. The population of this research includes all Smartphone users. Research paper is based on views of 118 people about Smartphone. People have different professions like students, teachers, doctors etc. Most of the sample was between ages of 15 to 30. Convenience Sampling was used to prove our hypothesis i.e. the Smartphone usage has more negative impacts than positive impacts and to carry out this research the survey was conducted through questionnaires, interviews and other sources to Smartphone user. They were asked about their age, profession and Smart phone brand. There was also a question that how much time they spend using Smartphone in a day and for what purpose they use Smart phone. Moreover, to investigate about the impact of Smartphone on communication skills and health because of Smart phone were requested to answer some questions about communication skill and book reading with which of smart phone users are affected or they feel insecure to take their smart phones outside. They use it in school hours and they find Smartphone a useful technology as related to above described problems which were included in the questionnaire. Since every technology has some pros and cons, so, to find out the advantages of this technology users were asked to rank smart phones from 1 to 10 and then the results were analyzed through a software named Statistical Package for the Social Sciences (SPSS).To find out the more accurate results and analyze the data we have used descriptive statistics and chi-square test of independence.

4. DATA ANALYSIS

After the survey we got the data which shows that most of the users were students which were 76% of the sample size. About 30% users use Smartphone for social networking purpose, 21% for internet surfing, a reasonable population i.e. 25% own it for study purpose which shows the brighter side of smart phones, 21% play games on Smart phone and few use it for other purposes. The most favorite Smartphone of the participants is Samsung as 34% have this brand, the second popular brand is HTC because of 17% users have this, 15% have Apple brand and 25% uses other brands. The 49% population buy Smartphone for entertainment purpose and 8% have smart phones just to show off which shows the negative side of technology. Then 34% people agree on that Smartphone affects their oral communication with friends while 44% did not feel any effect on their oral communication. The survey reveal that Smartphone has now became a symbol of danger for the users outside home, 54% users feel insecure while using Smartphone outside while 22% do not feel any danger and 24% become insecure sometimes. Book reading has now become an old fashion due to technology as the survey shows that Smartphone is affecting the 41% people's book reading hobby. On the other hand same percent of the people did not feel any affect in their book reading in daily life. 37% people think that this technology helps them in their daily life and studies whereas 19% of the people find this wastage of time technology. But 16% of the people get benefit from this technology and they find that this technology saves their time. The 26% of the people are not affecting by this technology. Due to getting constant updates on smart phones people might get disturb, 47% of the people lose their concentration by getting constant updates on smart phones while, 31% of people didn't lose their concentration

from other works and 21% of the people sometimes lose their concentration because of the smart phones. Having a Smartphone people can become more social with the outside of the world, 49% of the people think that their life becomes more social due to Smartphone, on the other hand, 26% of the people did not become more social due to having a Smartphone whereas 25% people's might think that their life become social. Smart phones have also affected the health of the people, 21% of the people think that their eyes are affected due to brightness of the Smartphone, 16% people feel headache due to Smartphone, and 61% of the people's health is not affected by the smart phones. The advance technology of Smartphone has also many disadvantages and we start the ranking of disadvantage from (1 to 10). 3% of the people think that Smartphone has many disadvantages while 26% of the people find it having fewer disadvantages.

The figure-1 shows the relationship of age and effect on health that majority of the people are from the age group of 20-24 who feel no effect while using smart phones. Whereas, some of the people feel eye pain due to brightness of Smart phone. Some of the people between age of 15 to 19 result same i.e. most of the users feel no any effect on their health. Some of also have headache due to extra use of Smartphone.

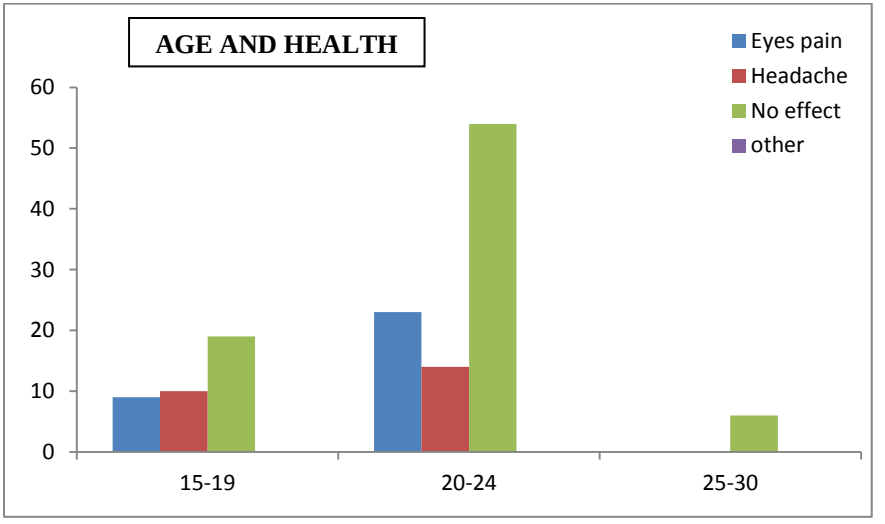


Figure-1

The figure-2 shows the relationship of age and time that most of the user is between 20 and 24 majority of which uses smart phone less 2 hours in a day. Almost equal percent of users uses 2 to 4 hours and few uses till the battery end up, while others use it all the time. The second highest user is i.e. some uses it till the battery ends up.

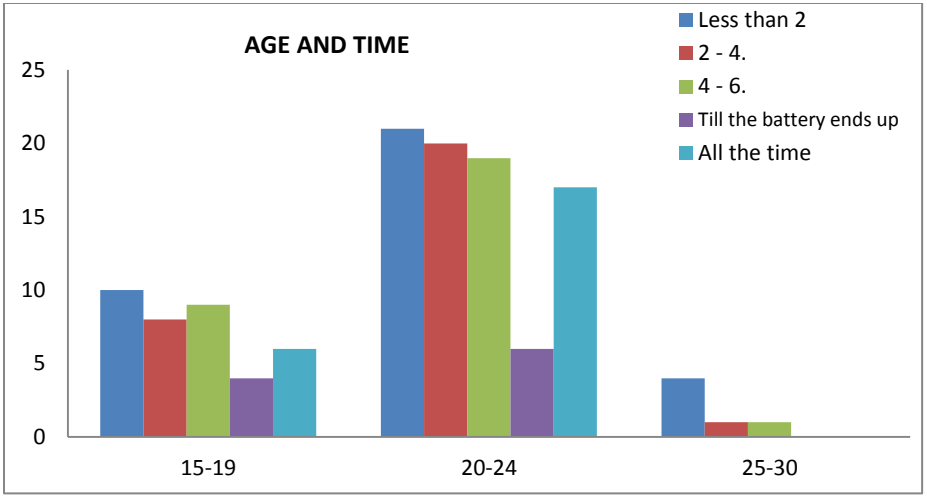


Figure-2

The question whether they use smart phones during school hours or not? Most of the students responded positively. By analyzing this through P-value of Chi-square test of independence we got the value, which is 0.002 shows the dependency between usage of Smartphone during schools hours and their profession as the chart is shown in the figure-3.

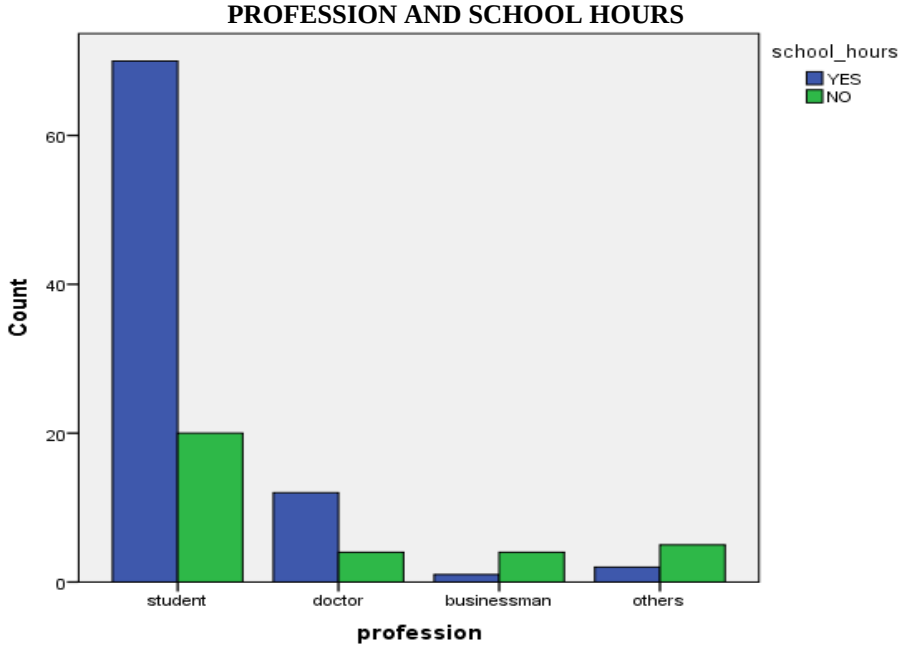


Figure-3

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14.828[a]	3	.002
Likelihood Ratio	13.140	3	.004
Linear-by-Linear Association	13.042	1	.000
N of Valid Cases	118		

The figure-4 below shows the relation between Profession and Health Effect where the P-value of Chi-square test of independence was applied on participants to investigate the effects of Smartphone on health, the value is 0.001 concludes that impacts on health are dependent on profession because most of users were student.

PROFESSION AND HEALTH EFFECT

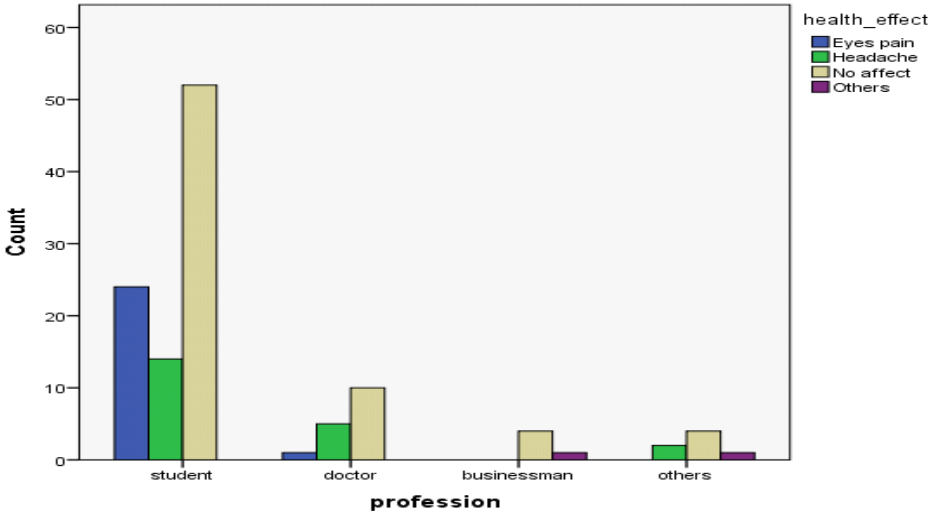


Figure-4

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	27.389[a]	9	.001
Likelihood Ratio	21.949	9	.009
Linear-by-Linear Association	6.927	1	.008
N of Valid Cases	118		

The cross table-1 of profession and time shows that majority of the students use Smartphone less than 2 hours in a day which were 25% of the total, equal number of student use it 2-4 hours and 22% of the total students use the Smartphone 4-6 hours.

Table-1

		time					
		less than 2 hours	2-4 hours	4-6 hours	till battery ends up	All the time	Total
profession	student	23	23	20	9	15	90
	doctor	4	3	4	1	4	16
	businessman	2	1	0	0	2	5
	others	2	0	2	2	1	7
Total		31	27	26	12	22	118

5. CONCLUSION

This is true that Smartphone plays various roles in our lives and in our society. It plays with our demands, moods, and needs and entertains us. There is no doubt that the Smartphone is bringing very joyful apps and time passing games. As by the survey we have taken, it has been concluded that half of the people use smart phones just to entertain their selves and one-third people use it for the study purpose. Most of them feel insecure when they are out of their home but don't feel any kind of effects in the oral communication with their friends. Three-fourth people use the smart phones in their school hours. Half of the people responded that smart phones help people with their studies but there are some who think that it is just wastage of time and does not help them in anything, in fact they lose their concentration by getting constant updates. People get easily socialized after having them. There was a funny part which came out after questioning people that 80% people get tensed with the less battery life of smart phones and charging them again and again. So after all that it has been concluded that Smartphone does play very important role in our life but, according to the result this study concludes that smart phones has more negative impact than the positive on the students as they use mostly in negative ways than to get the positive benefits.

6. ACKNOWLEDGEMENT

We would like to thank all the participants who took part in this research to make this successful, our batch mates, teachers, and those who helped us to make the survey possible. A special thanks to our instructor Miss Fareeha Sultan for helping us potentially on every step of our research.

REFERENCES

1. Bibby, A. (2011). *The Rapid Development of Tablet Computing*. A white paper for TCO Development by Andrew Bibby.
2. Muller, H., Gove, J.L., Webb, J.S. (2012). Understanding Tablet Use: A Multi-Method Exploration. *MobileHCI'12*, September 21-24, 2012, San Francisco, CA, USA.
3. Falaki, H., Lymberopoulos, D., Mahajan, R., Kandula, S. and Estrin, D. (2010). A First Look at Traffic on Smart phones. *IMC'10*, November 1-3, 2010, Melbourne, Australia.
4. Sarwar, M. and Soomro, T.R. (2013). Impact of Smart phones on Society. *European Journal of Scientific Research*, 98, 216-226.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 301-306

POWER CRISIS: A RESULT OF MISMANAGEMENT RATHER THAN UNAVAILABILITY OF RESOURCES

Waleed Arshad[1], Muhammad Naveed Sharif[2], Sheharyar Mohsin[3]
Abeeha Sattar[4] and **Fareeha Sultan[5]**
Department of Computer Science, FAST-National University
of Computer and Emerging Sciences, Karachi, Pakistan
Email: [1] k122021@nu.edu.pk
[2] k122018@nu.edu.pk
[3] k122030@nu.edu.pk
[4] k122034@nu.edu.pk
[5] fareeha.sultan@nu.edu.pk

ABSTRACT

The intense power crisis has an overall adverse effect on the development and progress of any country. The purpose of this paper is to determine the REAL and actual reasons for the power crisis in Karachi; the reasons, NOT those which are commonly known by the people through common unreliable sources, but the reasons which are actually the cause of power crisis in Karachi, the production level and industrial level reasons which are the real major cause of the crisis. Hypothetically this research revealed that the factor of insolvability of power crisis comes not under the authorities of the power generation or distribution plants, but it is majorly a result of mismanagement and lack of governance. The data collected is majorly on exploratory basis from personal investigation of technical personnel related to the power sector. The findings revealed expected results; the Karachi Electric is capable of producing enough power for the whole of Karachi but they are being given a limited amount of resources, like gas or oil. According to 70-80 percent of technical people related to power sector, the crisis is majorly due to a mere mismanagement and lack of governance at the upper hand. Along with the alternate sources, Pakistan should rely on some permanent and long lasting sources of power production such as nuclear power. Despite of its heavy investment, nuclear power is cheaper once installed and repays the investment in a very short period of time.

KEYWORDS

Power crisis; mismanagement; nuclear power.

1. INTRODUCTION

a) Background of study

Power crisis has always been defined in terms of lack of energy resources, and unavailability of proper facilities. The area of management is mostly ignored and overlooked in such prior cases where every single cause of the power crisis means a lot. The question is how much is the role of management in relation to the crisis currently faced by Pakistan. The purpose of this paper is to look upon those causes which have

always been overlooked and never emphasized upon, which prove to be one of the major reasons for the crisis, i.e. management.

b) Objectives

- According to chairmen of WAPDA, the annual requirement of electricity was increasing by the rate of 4.8% per annum till the year 2013, which has now increased to 8-10%. This shows that the strategy which was being used to overcome crisis will not be able to work now.
- The problem lies with the fact that sufficient capacity is completely available for production but it is not being used. Power stations all over Pakistan produce a lot lesser than they can actually produce. Objective is to find out what is the reason behind those STOPS that are hurdling the production despite of the fact that there is complete availability of resources plus equipment.
- What do we recommend to fully utilize those capacity which are under working condition but still are not being utilized?

2. LITERATURE REVIEW

Afia Malik (2012) analized the crisis in Pakistan in terms of crisis in Governance. She found that the power crisis is majorly a result of internal crisis in governance. According to her study the demand supply gap resulted in a regular load shedding for up to twenty hours in several rural areas. Quick increase in demand, high system losses, and lack of generation lead to this gap. She found out that poor decisions, and bad governances lead to the downfall of the power sector in Pakistan.

Power crisis is considered to be a core hurdle in the development of any country. Muazam Rashid Dar (2013) and Muhammad Azeem (2013) examined the impact of power crisis on the economy of Pakistan. They specifically determined and came up with the facts that power crisis adversely affects the economy of Pakistan to such a great extent. As economy is one of the building blocks of any nation's development, so is the case here. According to them, proper exploration of energy resources could be one of the optimal solutions to the power crisis.

Masnoon S. Ahmed (2005) studied the importance of nuclear power generation in the development. The reviews show that there has been consistent shortage of power since 2006-07 in Pakistan. And consistently the shortage of power is increasing every year. The reasons suggest that this is due to the improper and lack of infrastructure facilities. Masnoon S. Ahmed (2005) emphasized on the dependency on nuclear power as a solution for the power crisis.

As the population is rapidly growing, so is the demand of energy is increasing. Pakistan explores and exploits different energy resources such as gas, coal, petroleum and Nuclear energy, in order to fulfill its power requirements.

3. METHODOLOGY

This research is basically exploratory survey based study. The primary data was collected through personal interviews of technical personnel plus collection through questionnaires related to the power sector mainly from K-Electric (formal K.E.S.C), KANNUP and HUBCO. Purpose sampling technique has been used to formulate a

prescribed specific population. The search was done using the keywords: Power, Crisis, Mismanagement and nuclear power. The secondary data was collected through specific research journals; descriptive Statistical data has been used throughout. The paper is divided into the following sections: Background, Methodology, Literature Review, Objectives, Findings, Conclusion, Acknowledgment and Recommendations.

4. FINDINGS

The past few years give a clear view of the failure to take actions to fulfill the energy demands. A threefold rise in energy demand over the last two eras has been responded to with an ill-proportioned increment at the supply end. Pakistan's energy demand had grown at an annual consumption growth rate of 4.8 percent in the past five years, but now it is expected to grow at 8 to 10 percent per annum. Consequently, with the advent of 2013 the gap between demand and supply grew to 5,500MW indicating a 40% deficit of electricity. The dominant energy crisis has not appeared overnight — the signs were evident for a number of years but the authorities failed to react in time.

Current Installed (MWe)	Projected Demand		Gap (MWe)
	Year	(MWe)	
23538	2013	29038	5500
	2018	46766	23228
	2023	75317	51779
	2028	121299	97761
	2033	195353	171815
	2038	314618	291080
	2043	506696	483158
	2048	816038	792500
	2050	987407	963869

Fig. 1: Rate of demand and supply of electricity (MWe) from 2013 to 2050 (Every five years)

Pakistan has been facing an unmatched energy crisis since the last few years. The problem becomes more severe during summers. Primarily, during the peak crisis there was a power outage of 4-5 hours daily but it keep on increasing as the demand of electricity rises. According to the reports of WAPDA, the electricity needs of Pakistan increases more quickly than it can be produced which cause a major increase in the duration of load sheading. The precise rate of demand and supply is given in the figure.

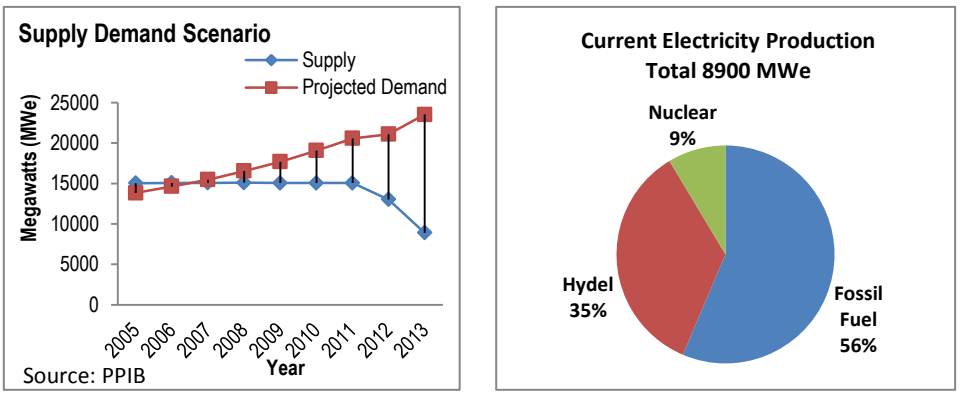


Fig. 2: Existing Installed Power Generation in Pakistan
Current Demand: 11200 (MWe)

TYPE OF GENERATION	Installed Capacity (MWe)	Current Production (MWe)
WAPDA (Hydel & Thermal)	13300	4593
K Electric	2316	872
IPP (Independent Power Producers)	7070	2670
Pakistan Atomic Energy Commission (NUCLEAR)	852	765
TOTAL	23538	8900

Name of Plant	Expected Production (MWe)	Starting Year
K2	1100	2019
K3	1100	2021
C3	330	By the end of 2014
C4	330	2016 (Max. two years from present)

Fig. 3: Expected Power Production of Upcoming Projects by PAEC

The total power production as predicted by Dr. Ansar Pervaiz, Chairman PAEC, is 8800 MWe by the year 2030. Five more Power production plants are to be installed before the year 2030 each will be expected 1100 MWe power production. In a recent interview Dr. Ansar Pervaiz said, "Pakistan would install more nuclear power plants to generate 42,000 MWe of electricity under Vision 2050. Till 2030, we will build seven plants of 1100 MWe capacity each besides building four 300-340 MWe plants"

According to him the average price of nuclear energy is Rs. 7 per kwh and if all plants are brought to working conditions by 2050 then the issue of circular debt will be resolved immediately. This average price includes all the fuel costs and operational costs as well.

Dr. Ansar Pervaiz also said, "C-1 and C-2 were running on 80 percent capacity and if all power plants including thermal, hydel and others run on such capacity there would be no load shedding."

5. COMMENTS AND CONCLUSION

The power sector of Pakistan is facing a serious era of mismanagement and unreal decisions; i.e. wrong decisions at the wrong time under wrong conditions. A better governance system will ensure proper decision making strategies which will hold the progress of the power sector. Unfortunately, our system lacks this major affecting entity. In order to control and finally eliminate the crisis factor we need to build up better governance strategies along with quick decision-making organizations, which will accordingly decide how to manage and control things under different circumstances.

The research that we have done showed some expected results, which ensure us that there are some serious hurdles that have to be taken into consideration in order to solve this Power crisis issue. Below are the hurdles that were found as a result of our research and study:

- We have the adequate power capacity installed but most of them are not working as most of our power plants rely on thermal systems and fossil fuels.
- Unavailability of funds, we are running a small part of our thermal systems.
- Mismanagement in governance as discussed in the National Policy 2013.
- Political interference.
- The IPPs are under intense circular debt which is currently unpaid by the Government. Consequently, the IPPs are unable to produce an optimum amount of Power. For the IPPs to work properly and produce Power at a maximum level the Government have to clear the debts.
- Line loses in shape of theft of electricity is increasing which also results in the increase in per unit rate. The burden is also beard by the sincere citizens.
- Electricity producing companies have lower quality grids that are used to supply electricity from power stations to consumers. These grids are affected by uncertain change in the climate and shutdown at very high temperature or stops working.
- K-Electric (formal K.E.S.C) has been privatized which has enhanced its production capacity but Govt. is not providing it the appropriate distribution network. As K-Electric is responsible to provide electricity to two of our provinces Sindh and Balochistan, in spite of the fact supply of fuel is uncertain.
- We are not concerned towards using the cheaper source of electricity i.e. hydel energy.
- Govt. is not taking interest in making large dams.
- Ill management and poor governance is involved in every matter of our country whether social, economic or country's sovereignty, e.g. the issue of Kala Bagh Dam.
- Mostly people do not have awareness of conserving power and some do not care.

6. RECOMMENDATIONS

- The goal of the National Policy 2013 should be strictly implemented i.e. Align the ministries involved in the energy sector and improve the governance of all related federal and provincial departments as well as regulators.
- Funds must be produced on high priority basis which is a big hurdle in power generation.
- Govt. should pay IPP's on regular basis instead of spending on Rental Power.
- The supply of fuel especially thermal and furnace oil must be made certain to use the existing capacity.
- Line loses in shape of theft of electricity should also be decreased by making new and harder laws.
- During the management process, electricity from neighboring countries may be used.
- Mega projects like construction of big dams should be started and should not be made a victim of politics.
- Along with big dams, small hydro project should be started as a short term solution.

- Alternative energy sources like solar, tidal and wind must be encouraged. Private sector should also be involved.
- The strategies of the National Policy 2013 should be implemented.
- Make use of Nuclear energy for power production as it is not influence by politics and a cheaper source of energy. No doubt installation requires funds but after installation, it is the best and reliable for power production.
- Electricity saver appliances should be introduced in the country.

REFERENCES

1. Javaid, M.A., Hussain, S., Abdul Maqsood, Arshad, Z., Arshad, M.A., Idrees, M. (2011). Electrical Energy Crisis in Pakistan and Their Possible Solutions. *International Journal of Basic and Applied Sciences IJBAS-IJENS*. 11(05).
2. Boureston, J. and Sultan, M. (2008). Understanding Pakistan's Energy Security Needs and the Role of Nuclear Energy. *Research Report 16*, SASS South Asian Strategic Stability Institute, London.
3. Dar, M.R., Malik, M.A. and Ramzan, M. (2013). Impact of Energy Consumption on Pakistan's Economical Growth. *Int. J. of Humanities and Social Science Invention*, 2(6), 51-60.
4. Malik, M.H. (2012). Chairman Special Committee on Energy Crisis in Pakistan, "*Energy Crisis in Pakistan*" (YP16-KPK03).
5. Thranert, O. and Wagner, C. (2009). Pakistan as a Nuclear Power. *SWP Research Paper*, Stiftung Wissenschafft and Politik German Institute for international and Security Affairs.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 307-314

PAKISTAN IN IT AFTER 10 YEARS: A STATISTICAL ANALYSIS

M. Faraz[1], M. Raza Devjani[2], Manisha Kumari[3], Wajeeha Ihsan[4]
Saba Mohammad Ali[5] and **Ayesha Mahmood[6]**
Department of Computer Science, FAST-National University
of Computer and Emerging Sciences, Karachi, Pakistan
Email: [1] k122326@nu.edu.pk
[2] k122077@nu.edu.pk
[3] k122314@nu.edu.pk
[4] k122101@nu.edu.pk
[5] k122032@nu.edu.pk
[6] k122065@nu.edu.pk

ABSTRACT

Information Technology (IT) has revolutionized the entire world, doesn't matter of which field you are talking about, it has grabbed each and everyone's attention to it. From dawn to dusk we are surrounded by multiple application of IT. This study was designed to serve as comprehensive development report on the future of IT in Pakistan. In this context, the study aims to give small attention to important changes that occurred in IT since independence and a lot to what is expected in next 10 years from Pakistan, especially in terms of usage. IT is concerned with the improvements in a variety of human and organizational problem-solving endeavors through the design, development, and use of technologically based systems and processes that enhance the efficiency and effectiveness of information in a variety of strategic, tactical, and operational situations. We have surveyed many individuals related to IT which includes users, IT professionals, IT teachers, and IT students to collect the data using convenient sampling. The study is a qualitative and quantitative analysis based on information taken from both secondary and primary information. The discussion concentrates on structure, access, quality, and future prospects of information technology in Pakistan. Statistical analysis of this study predicts that in next 10 years Pakistan is going to be one of the most promising centers of IT in the world and will be valued more than now.

KEYWORDS

Information Technology, convenient sampling, development.

1. INTRODUCTION

Information technology has been around for several years. Communicating through technologies is very common nowadays. Our world today has changed with aid of IT. Tasks that were once done manually and were time-consuming are now very easy to execute. Information technology is concerned with the improvements in a variety of human and organizational problem-solving that came through design, development and

use of technology and processes that enhance the efficiency and effectiveness of information in a variety of strategic, operational situations.

When Pakistan got independent in 1947, it was far away from technology. The first computer-based company was found in 1957 and then slowly many other companies and departments including many universities started installing mainframe computers for their uses. Prices were so high that even government budgets were not enough for buying computers. Starting from late 60's to 70's, Pakistani companies started realizing the importance of the field of IT. Things became more useful when personal computers were introduced in late 80's; soon they replaced mainframes and mini computers. From 1995, ISP started providing internet servicing and from 2000, IT was in everybody's mind as from then, IT professionals were hired and a lot of awareness was made regarding IT.

Pakistan is a rising force in IT and will improve more in the next ten years considering the improvements it has made in the past five years. Information technology in Pakistan is a growing and rising industry that has a large potential. Matters relating to the IT industry are overseen and regulated by the Ministry of Information Technology of the Government of Pakistan. The IT industry is regarded as a successful sector of Pakistan. The government of Pakistan has given numerous favors to IT investors in the country since last decade that resulted in the development of the IT sector. In the years 2003-2005 the country's IT exports saw a rise of about fifty percent and amounted a total of about 48.5 million USD. As of 2011, Pakistan has over 20 million internet users and is ranked as one of the top countries that have registered a high growth rate in internet penetration. Overall, it has the 27th largest population of internet users in the world. In the fiscal year 2012-2013, the Government of Pakistan aims to spend Rs. 4.6 billion on information technology projects, with emphasis on e-government, human resource and infrastructure development. According to a document prepared by the Ministry of Information Technology and Telecom, Pakistan's information technology exports rose 97.34 percent during last five years. Also, a document that was presented in the National assembly said that total IT exports valued around $ 1.250 billion during past five years. A total of seventeen private companies of Pakistan achieved the targets of exceeding one million US dollars in IT service exports. There are over 100 countries that depend on Pakistan IT companies for IT goods and services. As a result such rapid development in past, It is highly expected that Pakistan's will get close to the advancement that IT will possess in next 10 years.

This research paper statistically investigates that what expected scope information technology has in Pakistan in the next ten years. Pakistan has been developing in IT but it will be predicted out how it will bring changes in its IT industry in the future and whether it will improve or deteriorate based on people view. This paper includes literature review of previous studies in this regards, methodology and discussion on the results obtained by using statistical techniques.

2. LITERATURE REVIEW

Bert Amor (2002) examines the future direction of Information Technology for construction. A significant focal point for this research in terms of its dissemination and the derivation of a shared research agenda has been the working commission concerned

with IT for construction within the international council for innovation and research in building and construction.

M.I. Afreen (2004) examines the impact of E-commerce on the economy of Pakistan. It also focuses on the future of banking sector as well as the Governance policy and law enforcement details that is available by the government itself. The sample was from the public sector, private sector that are moving to adapt information technology. SWOT analysis was done to analysis the strength and weakness, opportunities and threats for the E-governance revolutionary change. The findings they made were that electronic technology is changing each aspect of society generally and gradually and in near future all banking sector will be utilizing this source.

S. Mukhtar & K. Rashid (2012) investigates the impact of information technology on the technical education of Pakistan. The study also briefs the importance changes and implication of information technology. The study is based on the qualitative and quantitative analysis based on secondary data. The variables that were witness the importance of Information technology in last 10 years and for upcoming era were the increase in IT Scholarship about 515 scholarship a year, increase in IT engineering, workshops, E-commerce, IT education for managers and future aspects of IT. The findings were that many students are attracted and scope of the IT is increasing rapidly. The Information technology benefits seem around the society.

M. Ali, A. Ahmed, A.W. Shaikh and A.H.S. Bukhari (2011) investigate the impacts of information technology trends on e-learning process in Pakistan by using the sample of institutes and campuses of AIUP and VUP (n=219). The uses of IT resources, cellular phone network, and video conferencing have been used as variables. It is concluded that Improper and incomplete use of information technology creates problems for e-learners as well as e-teachers and e-trainers in Pakistan. It is recommended that the lectures and presentations are created using the Multimedia tools, in which animation and games should be used; video conferencing should be used; cellular phone network can also be used to improve e- learning system.

Z. Mahmood and R. Siddiqui (2000) analyze state of Technology and productivity in Pakistan's manufacturing industries by using the sample of industries growth factor. Uses of IT resources and some strategic directions to build technological competence have been used as variables. This study concludes that How Pakistan's Technology and productivity are co related factors in manufacturing industries and how these industries impacts the economy of the country. It is recommended that the development of an efficient and effective technology apparatus is of utmost importance for the growth of industries. Given a liberal economic environment in the country, which is essential to improve efficiency and productivity, the paper offers the following four strategic directions in order to improve the status of the S & T system in Pakistan:

1. Augment the public sector Science and Technology apparatus with the private sector funding and oversight.
2. Take measures to upgrade research institutions to the international standard.
3. Streamline the technology creation, absorption, and diffusion system.
4. Enhance the demand for S & T in industries.

3. METHODOLOGY

The methodology used in this paper is not only about the views of people for the present and future but also to analyze the changes in information technology of Pakistan in the previous years.

We have surveyed three categories of people on what they think about the future of IT in Pakistan, considering the present situation. The sample was divided into the following categories:

- General users of all technology.
- Students and teachers (related to the field of IT)
- IT professionals.

General users include all those people that use any kind of technology whether online banking, online education, computers, or smart phones, etc.

Students and teachers are all those that are related to the field of IT and possess critical views and observation related to their respective field.

IT professionals includes individuals who provide IT services or are at jobs in IT industries.

The sample size for our survey was 80 per category. We surveyed through online questionnaires and paper questionnaires.

The statistical technique used was a statistical software package named SPSS 12.0.

The method of sampling used was convenient sampling.

Questions asked from the sample population of our different categories were all made by keeping their current status in view, whether they are students, simple users or IT professionals.

Some important variables we discussed from our categories included rate of technology use, rapid IT involvement in surroundings, graduates career scope, university's plan, youngster's interest in IT, Pakistan IT companies domination in next 10 years, etc. The most important of them all which is used in all categories is youngster's interest in IT, from this it was meant that are individuals shifting to IT rather than other available fields?, and the answers were all positive in every category.

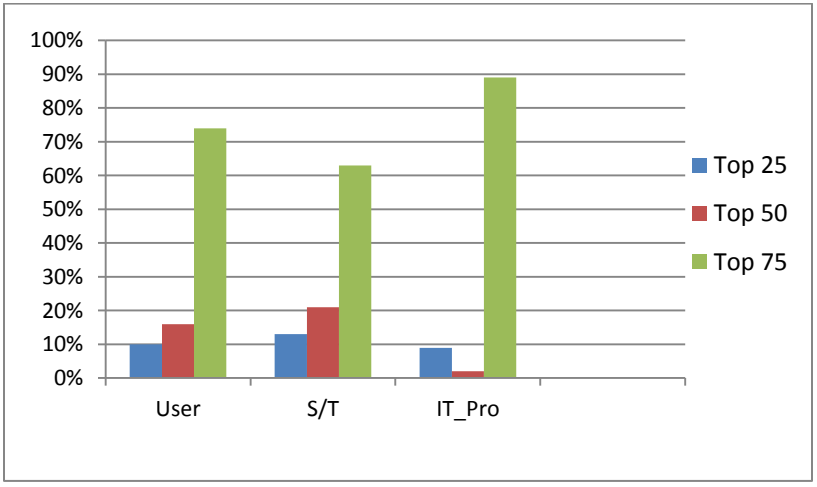
4. RESULTS AND DISCUSSIONS

It was found from surveys that 72% of IT professionals and 55% of teachers and students think that Pakistan will be in the top 75 countries in the next ten years, which is a not at all a impropriate expectation as Pakistan is ranked 105^{th} now. 99% of IT professionals and students and teachers in the IT field think that they have a promising future in the field of IT and 88% believe there will be an increase in the IT industries in Pakistan. 90% of the professionals think their company is using the latest technologies of IT. 82% of IT professionals think that standard of IT in Pakistan is good and foreign companies are interested in Pakistan for their IT goods and services which is true according to our hypothesis based on the past five years. 77% of IT professionals think that they will be in foreign countries after ten years which is not good because Pakistan

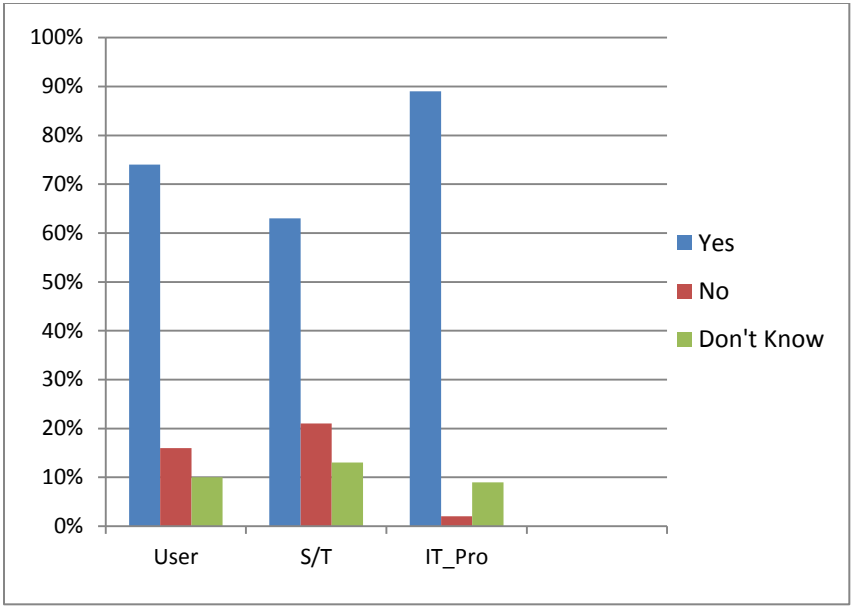
needs good IT professionals to improve in the field of IT. But 97% of the professionals also answered affirmative to whether there will be more IT professionals in the future or not. This will be a good thing for Pakistan IT industry. 71% of professionals believe they get good projects or will get projects from foreign in future. Our hypothesis was also based on the fact that 100 countries depend on Pakistan's IT goods and services. 89% of the professionals find that people are getting involved in IT more than ever. It is true because people of all ages are now directly or indirectly using any IT product nowadays.

The students and teachers of the IT industry were asked a different set of questions. Their responses showed that 21% of them are in IT because of interest but 34% are in the field because of the scope of the field so it is true that people believe that IT has a good scope ahead. 60% of the people of this category believe that Pakistan is among the best IT industries of the world. Our hypothesis also stated that Pakistan is ranked 105 in the field of IT. Only 7% of the people think that their university will plan for more advance courses in the future. 79% people think that more people are shifting to IT as a profession. Pakistan's ranking on other side also depends upon how many foreign IT companies establishes their business in Pakistan.

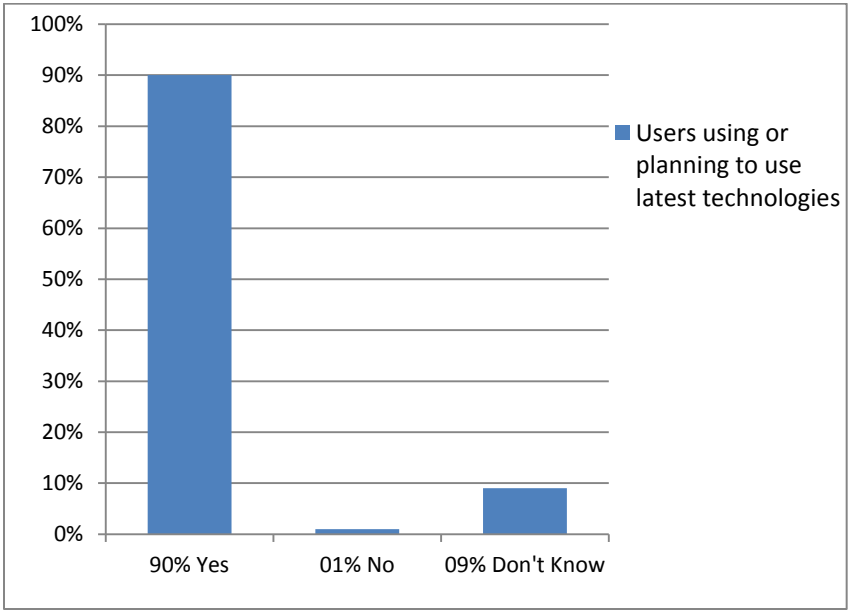
As for the users of IT services and products, 99% of the users think that IT has made their lives easier and fast. 90% of users plan to use latest IT technology. 72% of users will use online shopping in future. 61% of users think that online voting system will be available in Pakistan in the next ten years which in our hypothesis was rue because the government of Pakistan is taking steps to bring E-governance in Pakistan. 78% of the users are using or are planning to use GPS technology in the future. Only 31% of the users think that government is taking steps to promote IT in Pakistan. 91% of the users keep themselves updated with the latest technologies joining the world of IT. 74% of users think that people are joining the field of IT more than any other field like medical or engineering.



A graph presenting the results of views from multiple samples when asked, "Where do you see Pakistan's ranking in IT after 10 Years? Which is 105th now".



A graph representing the results of views from multiple samples when asked "Do you see people around you shifting towards the field of IT rather than other available fields".



A graph representing the results of views from single samples when asked "Are you using or planning to use latest technology devices in present or near future".

6. CONCLUSIONS

Right from the rise of IT in the world, Pakistan has been involving itself in each and every IT field that can be beneficial for its cause. Even in today's world, Pakistan is rated one big investor for IT worldwide when it comes to quality engineers.

Taking analysis of all the results/view/data collected, it can clearly be said and believed that Pakistan is in the very right direction in IT. People from every walk of life believes that IT is taking Pakistan under its shade day by day and that Pakistan will soon be among best involving countries in the world for IT. Doesn't matter from whom you ask about the current and near future of it, either it be a simple user, a IT student or teacher, or a IT field related professional, Pakistan is very high among all of their view. Almost in every case whenever asked about a positive outcome from Pakistan, 90%+ said **YES.**

People are using latest computer accessories, even a small kid knows and understand all related stuff. Youngster are planning and getting admitted to high class IT universities for making IT as their profession. Engineering/medical is no more the only choice left. More universities purely related to IT are being introduces in Pakistan and more are accepted in near future. Students are using/accessing advance course related websites,.

Apart from what is about to come, even today IT professionals are highly paid in Pakistan, and the ration is expected increase in near future. IT companies are equipped with high computing devices allowing them to deal with development of complex softwares. According to professionals, more IT companies will be developed as Pakistan is going to have a very number of IT professionals. Apart from all positives, the expected ranking of Pakistan in IT can never be ignored. 80%+ IT professionals stated that Pakistan will shift from 105th position to Top 50, which without a doubt is a huge improvement.

The only that seems to be worrying the people is the govt. of Pakistan. A big ratio suggested that Pakistan is slow in adopting latest technology related to Internet communication like 3G and 4G, though Pak govt. has shown interest in such cases but still if anything stopping Pakistan to one of the most shinning nation in IT is the .govt.

5. ACKNOWLEDGEMENT

We would like to acknowledge the support of management of our respective university FAST-NUCES Karachi who gave us this opportunity to make a research and also to permit us collecting required surveys with their reference. We are also grateful to our tutor Madam Faqiha Sultan who encouraged us to pursue this topic and spend some extra time helping us to achieve a clearer structure.

REFERENCES

1. Bert, A. (2002). *Future Direction of Information Technology for Construction*: October at http://www.itcon.org/2002/15, Newzealand.
2. Afreen, M.I. (2004). *Impact of E-Governance on the Economy of Pakistan – The Revolutionary Plan*, Center for Advanced Studies in Engineering Islamabad.

3. Rashid, K. (2012). *Impact of Information technology in Pakistan*, Head of Department, CIIT, Islamabad.
4. Ali, M., Ahmed, A., Shaikh, A.W. and Bukhari, A.H.S. (2011). Impacts of information technology on E-learning system in Pakistan, *Newsletter,* 60 - *Scribd*, Balochistan University, Balochistan.
5. Zafar, M. and Siddiqui, R. (2000). The Pakistan Development Review: PDR.- Islamabad, ISSN 0030-9729, ZDB-ID 2075544. 39(1), 1-21

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 315-332

RETROSPECTIVE AND PROSPECTIVE SAMPLING APPROACHES AND MULTIVARIATE ANALYSES FOR THE ASSESSMENT OF GENETIC VARIATION ACROSS A HETEROGENEOUS POPULATION

Nuzhat A. Akram[1,2] and **Shakeel R. Farooqi**[2§]
[1] Department of Genetics, University of Karachi,
Karachi, Pakistan.
[2] Department of Forensic Medicine, Hamdard University
Karachi, Pakistan
[§] Corresponding Author: Email: farooqis@uok.edu.pk

ABSTRACT

Objectives: Prospective and retrospective sampling approaches and multivariate analyses were employed for the assessment of genetic variation in a heterogeneous population. Karachi the largest cosmopolitan city of Pakistan was chosen as the model population.

Methods: Population heterogeneity was assessed by calculating ethnic fractionalization. Three measures of variation were used to assess genetic variation. These included blood groups ABO and Rh, anthropometric measures, and autosomal STRs. Sample populations for each measure were independent of each other. Each individual was investigated for his/her self-declared and parental ethnicities using a questionnaire. In retrospective approach, simple random sampling method was used for blood group study (n=3000). Five ethno-linguistic groups were identified on the basis of self-declared ethnicities. These groups were taken as population strata. Multivariate analysis of variance (MANOVA) was used for testing hypothesis of homogeneity. In prospective approach stratified random sampling for autosomal STRs (n=175) and stratified convenient sampling methods were used for anthropometric studies (n=514). Analyses of variance (ANOVA) for the number of STR alleles and anthropometric data were performed to show statistically significant differences across the population strata. Coefficients of variation across the three measures were calculated. Hierarchical cluster analyses based on complete linkage method and squared Euclidean distance were performed for each of the three measures of variation.

Results: Analyses of variances showed non homogeneity across the population strata for the three measures of variation. Coefficients of variation (CV) across blood groups, anthropometric measures and STR markers were 35.4%, 14.6% and 16.1% respectively. The clustering of the five population strata was different in dendrograms of retrospective and prospective approaches.

Conclusions: Significant differences across the five population strata were found in both retrospective and prospective approaches. However, significantly different coefficients of variations and different clustering patterns in dendrograms for the two sampling approaches indicate the importance of sampling method in a heterogeneous population.

1. INTRODUCTION

South Asia has always been the focus of interest for geneticists, anthropologists, linguists and historians due to its incredible population diversity (Petraglia and Allchin, 2007). Pakistan is an ethnically and linguistically diversified state. Its population is said to be sub-structured, with dozens of constituent ethnic and linguistic groups or subpopulations (Blood, 1994; Rahman *et al.*, 2001, Overall, 2009, Akram and Frooqi, 2014).

The census population of Karachi is composed of five major ethno-linguistic groups each sharing more than three percent in census population (Pakistan Census Organization, 1998). These groups are designated as Punjabi, Pathan, Sindhi, Balochi and Muhajir (Table 1). Two of the ethno-linguistic groups Sindhi and Balochi are the older residents of the city (Gankovsky, 1964). After partition of 1947, millions of people migrated from India and settled in Karachi. These immigrants are known as Muhajir. Later, people from different areas of Pakistan migrated to Karachi as it is Pakistan's most industrialized city and the commercial and financial capital. Hence, Karachi's inhabitants constitute a cosmopolitan population making the city a melting pot of diversified ethno-linguistic groups. Mating among these population groups is usually based on ethnicity, language, *bradri*, and caste (Shami *et al.*, 1989, 1990; Hussain and Bittles, 1998; Hussain, 1999). Hence consanguinity rate is as high as 60 to 80 percent.

The pattern of genetic variation within a population is best described by the relative frequencies of different genetic variants (Chakraborty *et al.*, 1988). The choice of the genetic markers for population studies depend on their polymorphism and distribution across the populations. Polymorphic markers that show variation across populations are best suited for population studies. Blood groups ABO and Rh have been used for many decades for identification purposes as well as to delineate anthropological differentiation between different races (Hirszfeld and Hirszfeld, 1919; Fisher and Race, 1946; Mourant, 1954; Beckman, 1959; Mourant, 1970; Kalabushkin *et al.*, 1982; Majumdar and Roy, 1982). The first systematic application to anthropology of the ABO blood groups was made by Hirszfeld and his wife in 1918-19. They showed a decrease in frequency of blood group A from Western Europe to India, and of blood group B from India to Western Europe. Hirszfeld concluded that higher frequencies of blood group A or blood group B across human races had developed by mutation from O. Frequencies of these blood groups were changed by human migration and subsequent intermarriages (Allan, 1963). It is worth mentioning that a higher incidence of blood group B is found across Indian and Mongolian races thus supporting Hirszfeld hypothesis about geographical distribution of blood groups. The study of the blood groups of human populations has become a planned and integrated investigation for various purposes (Grunnet *et al.*, 1994; Larsen *et al.*, 2005; Lee *et al.*, 2012; Daniels, 2013; Akram and Farooqi, 2014).

Anthropometric measurements are important for determining age, sex, racial and ethnic groups (Williams *et al.*., 1995). Anthropometry of cranio-facial measurements plays a vital role in anthropological research and forensic investigation (Oladipo, 2007, 2010a, 2010b; Krishan, 2008; Olotu, 2009, Mahdi and Anibeze, 2012; Ewunonu and Anibeze, 2013). The most important cranio-facial measurements are height (or length) and width of head that are used in cephalic index determination. Genetic,

nutritional as well as environmental factors have been attributed to variations in cephalic indices across human populations (Kasai *et al.*., 1993; Williams *et al.*, 1995; Shah and Jadhav, 2004). Nasal index is also a very useful tool in anthropology in delineating racial and ethnic variations (Aung, 2000; Franciscus and Long, 2001; Porter and Olson, 2003). Importance of nasal index has also been documented in forensic science (Zhang *et al.*, 1990; Xu *et al.*, 2001).

Short Tandem Repeat (STR) markers are tandemly connected units each consisting of 2-6 base pairs (bp) (Polymeropoulos *et al.*, 1990; Goodwin *et al.*, 2011). STR sequences account for 3% of total human genome (Lander *et al.*, 2001). STRs do not have a stable probability of being present across the populations that is their frequency estimates vary among the populations (Budowle and Moretti, 1999; Tvedebrink, 2010). Hence their utility as identification markers varies from population to population (Butler, 2006; Haasl and Payseur, 2011; Curic *et al.*, 2012). In the past two decades a number of STRs have been explored for application to human identification (Butler and Hill, 2012).

The main objective of the present study is to provide a valid sampling method to investigate heterogeneous population using different sets of polymorphic markers so that population strata can be identified as distinct entities. This is particularly useful for case control and forensic studies where inference is greatly influenced by the method of sampling.

2. MATERIALS AND METHODS

Sampling Area

Karachi the largest urban agglomeration of Pakistan was chosen as the sampling area for the present study.

Ethnic Fractionalization

Ethnic fractionalization was calculated according to Fearon (2003). If the population share of ethnic groups (% in census population) are denoted by P_1, P_2, P_3.....Pn, the fractionalization will be 1- $\sum P^2$.

Samples

In retrospective approach the city was divided into five areas i.e. East, West, Center, North and South. Hospitals, colleges and community centers were selected from each of these areas. Three thousand unrelated individuals were selected through simple random sampling. Self-declared and parental ethnicities of each individual were investigated as confounding variables. Each individual was categorized into one of the population groups on the basis of self-declared ethnicity. Five major population groups were identified that accounted for 77.96% of the total samples of 3000 individuals. These population groups were Balochi, Muhajir, Pathan, Punjabi and Sindhi.

In prospective approach six hundred individuals of ages more than 20 – 22 years were selected across the five population groups namely Balochi, Muhajir, Pathan, Punjabi and Sindhi namely through stratified convenient sampling for anthropometric study. Each population group was identified on the basis of self-declared ethnicity.

In prospective approach of STR study one hundred and seventy five individuals were selected across the five population groups namely Balochi, Muhajir, Pathan, Punjabi and

Sindhi through stratified random sampling. Each population group was identified on the basis of self-declared and parental ethnicities, namely both the parents of individuals belonged to the same population group.

Blood Grouping

Anti A, anti- B and anti –D sera were used to confirm ABO blood group and Rh factor of each individual.

Anthropometric Measurements

Head length was calculated from summit of glabella to inion (the most prominent occipital point). Head breadth is taken across maximum horizontal diameter.

Nasal length was calculated from the nasion to the tip of the nose and nasal height was measured across the nasal bridge.

STR Genotyping

All the three markers i.e. CSF1PO, TPOX, and TH01 were amplified simultaneously using CTT Multiplex kit in 2400 thermal cycler following the protocols provided by the Promega Geneprint STR system technical manual. PCR samples (amplicons) were run on denaturing 4% and 6% PolyAcrylamide Gel. DNA bands on polyacrylamide gel were visualized through silver staining performed according to the protocol provided by the Promega Geneprint STR system manual (Bassam *et al.*, 1991). Alleles for each of the three loci were determined by comparison of the sample fragments with those of allelic ladder provided with the Promega kit.

Allele Frequency Estimation

ABO gene frequencies were estimated through adjusted Bernstein method (Yasuda and Kimura, 1968; Nam and Gart, 1976). Allele frequencies for all the three STR markers were estimated through maximum likelihood method (Hedrick, 2011).

Cephalic and Nasal Indices

Cephalic index was calculated using the formulae *cephalic index= (head length/head breadth)* × *100.* Nasal index was calculated using the formulae *nasal index= (nasal length/nasal breadth)* × *100*

Assessing Genetic Variation

Analysis of Variance (MANOVA) was performed for the distribution of blood groups ABO and Rh across the population groups identified on the basis of self-declared ethnicities.

i) Analysis of Variance (ANOVA) was performed for anthropometric measurements across the population strata based on self-declared ethnicities.
ii) Analysis of variance (ANOVA) was performed for the number of alleles of the three STR markers across the population strata based on self-declared and parental ethnicities.

Coefficients of Variation

Coefficients of variation were calculated for blood groups ABO and Rh, anthropometric measurements and the three STR markers.

Cluster Analyses

Hierarchical cluster analyses based on complete linkage method and squared Euclidean distance were performed across the five population groups (strata) for the three measures of variations i.e. blood groups ABO and Rh, anthropometric measurements and the three STR markers.

3. RESULTS

Ethnic Fractionalization

Ethnic fractionalization was 0.725 that means there is 72.5 percent chance for randomly selected individuals that they would come from different population groups (Table 1).

Allele Frequency Estimation

Distribution of ABO blood groups across the five population groups is shown in Table 2. Sample sizes for anthropometric measures and STR markers are shown in Table 3. Allele frequencies for all the three STR markers namely CSF1PO, TPOX and TH01 are shown in Fig. 1 through Fig. 3.

Cephalic and Nasal Indices

Cephalic and nasal indices are shown in Fig. 4 and Fig. 5 respectively.

Genetic Variation

Analysis of variance (MANOVA) showed statistically significant (p value <0.05) difference across the five population strata for the distribution of blood groups ABO and Rh (Table 3). Cephalic indices across the five population strata also showed statistically significant difference (p value <0.05) (Table 4). Distribution of the two STR markers TPOX and TH01 were also significantly different across the population strata (Table 5).

Coefficients of Variation

Coefficient of variation was the highest for the blood group ABO and lowest for the STR locus CSF1PO (Table 6). On average the variation was high across the sample population of simple random sampling and low across the sample population of stratified random sampling.

Cluster Analyses

Hierarchical cluster analyses showed the relationship across the five population groups by constructing dendrograms based on similarities between them (Fig. 6 through Fig. 8). Dendrograms of the prospective approach showed the same clustering pattern across the five population groups (Fig 7 and Fig. 8).

4. DISCUSSION

Ethnic diversity does not necessarily confer genetic diversity to a population. It merely indicates the presence of population clusters or subpopulations within a larger population based on self-declared ethnicity. The extent of ethnic diversity can be estimated through a statistic called ethnic fractionalization ‘F’ as described by Fearon (2003). Karachi showed a higher ethnic fractionalization. Therefore the city can be taken as a model heterogeneous population. However the statistics does not satisfy the need to investigate population stratification. The issue of population stratification is highly

crucial for case control and forensic studies. Once the population heterogeneity is known (as indicated by high ethnic fractionalization 'F' value) attempts should be made to identify genetic variation across these ethnic groups. For this purpose those polymorphic genetic markers or phenotypic traits are suitable that are likely to be population specific or at least show variation across population groups. In the present study high value of 'F' was taken as an evidence for the heterogeneity of the population. For validation two sets of polymorphic genetic markers namely blood groups ABO and Rh and STR markers and two phenotypic traits namely cephalic and nasal indices were used. Both genetic markers have been used as identification markers for many decades. Phenotypic traits are considered to be specific to major human races.

Choice of sampling method appropriate for a heterogeneous population is another important issue in population studies (Akram and Farooqi, 2014). Simple random sampling is considered the most unbiased method for a homogeneous population. When it is employed for a heterogeneous population the variable responsible for heterogeneity (for example self-declared ethnicity) is considered as confounding variable. However the individuals are selected through randomization regardless of their self-declared ethnicity. For such data we can use statistical test like ANOVA or non-parametric Kruskal Wallis to find significant difference across these self-declared ethnic groups or population strata. This approach is called retrospective. On the other hand in prospective approach stratified random sampling is used in which each stratum of the population is identified right from the beginning of the study. Individuals are chosen from each stratum either through randomization or convenient sampling. In the present study both the approaches were employed with different sets of polymorphic markers. Coefficients of variation (CV) were more than two folds for retrospective approach than prospective approach. The larger variation of retrospective approach may be the results of sampling method, choice of markers and sample sizes. Sample size of the retrospective approach is much larger than retrospective approach (n= 3000 for blood groups, n = 175 for STRs, and n=473 for anthropometric measures). We know that variation tends to be larger for smaller sample sizes. Moreover, markers with higher degree of polymorphism are likely to produce larger variation. Set of markers for retrospective approach (blood groups ABO and Rh) exhibit smaller degree of polymorphism than STR markers of prospective approach. Therefore the difference in sample sizes or choice of markers cannot be accounted for larger variation of retrospective method. The value of CV favours the prospective approach for sampling purposes in a heterogeneous population as it shows a conservative estimate of the variation across various population strata or ethnic groups. Therefore, stratified random sampling with a prospective approach can be considered more appropriate for a heterogeneous population than simple random sampling with a retrospective approach. This is important for case control and cross sectional studies where approaches are directed to find out a cluster/s that is the focus of attention in disease prevention/control. An inflated variation is likely to increase the incidence of false positive or false negative cases.

For prospective approaches a slightly different identity criteria were set for anthropometric measures and STR markers. Self-declared ethnicity was taken as identity criterion for population strata for anthropometric measures. On the other hand self-declared and parental ethnicities were taken as identity criteria for population strata for

STR markers. However, coefficient of variation does not differ significantly for the measures based on two different identity criteria. It may indicate the high consanguinity rate or marriages within population stratum. In other words self-declared ethnicity can be taken as proxy for parental ethnicities.

Dendrograms of the two prospective approaches also showed the same clustering pattern of the five population groups. That is Pathan, Punjabi and Sindhi showed a higher degree of similarity than other two population groups Balochi and Muhajir. Dendrogram of the retrospective approach showed a different clustering pattern of the five population groups. In principle, it is not justified to infer aspects of the history or structure of an entire population on the basis of variability of one or a few markers that account for only a partial component of overall genetic variability (Tarazona-Santos *et al.*, 2001). However it is helpful in planning future research particularly sampling strategy and study design for population studies.

REFERENCES

1. Akram, N. and Farooqi, S. (2014). A model-based 'varimax' sampling strategy for a heterogeneous population. *Ann. Hum. Biol.*, 41(2), 116-124.
2. Allan, T.M. (1963). Hirszfeld and the ABO blood groups. *Brit. J. Prev. Soc. Med.,* 17, 166-171.
3. Aung, S.C., Foo, C.L. and Lee. S.T. (2000). Three dimensional laser scan assessment of the Oriental nose with new classification of oriental nasal types. *Br. J. Plast. Surg.*, 53(2), 109-116.
4. Bassam, B.J., Caetano-Anolles G. and Gresshoff, P.M. (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.*, (196), 80-83.
5. Beckman, L. (1959). A contribution to the physical anthropology and population genetics of Sweden. Variations of the ABO, Rh, MN and P blood groups. *Hereditas*, 42, 1-189.
6. Blood, P.R. (1994). *Pakistan: A country study.* Library of Congress. Federal Research Division. Washington DC.
7. Budowle, B. and Moretti, T.R. (1999). Examples of STR population database for CODIS and for casework. *9th International Symposium on Human Identification*, 1, 64-73.
8. Butler, J.M. (2006). *Statistics and population genetics*. New Jersey State Police Training Workshop. Hamilton NJ.
9. Butler, J.M. and Hill, C.R. (2012). Biology and genetics of new autosomal STR markers useful for forensic DNA analysis. *Forensic Sci. Rev.*, 24, 15-26.
10. Chakraborty, R., Smouse, P.E. and Neel, J.V. (1988). Population amalgamation and genetic variation: Observation on artificially agglomerated tribal populations of Central and South America. *Am. J. Hum. Genet.*, 43, 709-725.
11. Curic, G., Gasic, V., Pluzaric, V. and Smiljcic, D. (2012). Genetic parameters of five new European Standard Set STR markers (D10S1248, D22S1045, D2S441, D1S391) in the population of Eastern Croatia. *Croat. Med. J,* 53(5), 409-415.
12. Daniels, G. (2013). *ABO blood groups*. John Wiley and Sons Ltd. The Atrium Southern Gate, Chechester, West Sussex, PO198SQ, UK.

13. Ewunonu, E.O. and Anibeze, C.I.P. (2013). Anthropometric study of the facial morphology in a South Eastern Nigerian population. *Hum. Bio. Rev.,* 2(4), 314-323.
14. Fearon, J.D. (2003). Ethnic and cultural diversity by country. *J. Economic Growth*, 8, 195-222.
15. Fisher, R.A. and Race, R.R. (1946). Rh gene frequencies in Britain. *Nature Lond.*, 157, 48-51.
16. Franciscus, R.G. and Long, J.C. (1991). Variation in human nasal height and breadth. *Am. J. Phys. Anthropol.,* 85(40), 419-427.
17. Gankovsky, Yu. V. (1964). *The peoples of Pakistan an ethnic history*. 2nd ed. Peoples Publishing House. Lahore, Pakistan.
18. Goodwin, W., Linacre, A. and Hadi, S. (2011). *An introduction to forensic genetics*. 2nd ed. Wiley Blackwell. UK.
19. Grunnet, N., Steffensen, R., Bennett, E.P. and Clausen, H. (1994) Evaluation of histo-blood group ABO genotyping in a Danish population: frequency of a novel O allele defined as Oz. *Vox. Sang*., 67, 210-215.
20. Haasl, R.J. and Payseur, B.A. (2011). Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity*, 106(1), 158-171.
21. Hedrick, P.W. (2011). *Genetics of populations*. 4th edition Jones and Bartlett Publishers. USA.
22. Hirszfeld, L. and Hirszfeld, H. (1919). Serological differences between the blood of different races. *Lancet*, 194(5016), 675-679.
23. Hussain R. (1999). Community perceptions of reasons for preference for consanguineous marriages in Pakistan. *J. Biosoc. Sci.,* 31(4), 449-61.
24. Hussain, R. and Bittles, A.H. (1998). The prevalence and demographic characteristics of consanguineous marriages in Pakistan. *J. Biol. Sci.,* 40(2), 261-75.
25. Kalabushkin B.A., Prokhorovskaia, V.D., Khil'chevskaia, R.I. and Altukhov, I. (1982). Migration as a factor in the genetic heterogeneity of the population of large cities. *Genetika*, 18(5), 833-838.
26. Kasai, K.L.C., Richard, T. and Brown, T. (1993). Comparative study of craniofacial morphology in Japanese and Australian aboriginal population. *Hum. Biol.,* 65, 821-834.
27. Krishan, K. (2008). Estimation of stature from cephalo-facial anthropometry in north Indian population. *Foren. Sci. Int.,* 181(1-3), 52e1-52e6.
28. Kumar, D. (2004). *Genetic disorder of the Indian subcontinent.* Springer. Kluwer Academic Publishers. Dordrecht. The Netherlands.
29. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.
30. Larsen, T.B., Johnsen, S.P., Gislum, M., Møller, C.A.I., Larsen, H., Sørensen, H.T. (2005). ABO blood groups and risk of venous thromboembolism during pregnancy and the puerperium. A population-based, nested case–control study. *J. Thromb. Haemost.,* 3, 300-04.

31. Lee, B., Zhang, Z., Wikman, A., Lindqvist, P. and Reilly, M. (2012). ABO and RhD blood groups and gestational hypertensive disorders: a population based cohort study. *BJOG*, 19, 1232-1237.
32. Mahdi, E. and Anibeze, C.I.P. (2012). An investigation on cephalometric parameters in Iranian population. *J. Developental Biology and Tissue Engineering*, 4(1), 8-11.
33. Majumdar, P.P. and Roy, J. (1982).Distribution of ABO blood groups on the Indian subcontinent: A cluster analytic approach. *Current Anthropology*. 23(5), 539-566.
34. Mourant, A.E. (1954). The distribution of human blood groups. Blackwell Scientific Publications, Oxford, England.
35. Mourant, A.E. (1970). The use of blood groups in the study of populations. *Ann. N. Y. Acad. Sci.,* 169(1), 194-198.
36. Nam, J.M. and Gart, J.J. (1976). Bernstein and gene counting methods in generalized ABO like systems. *Ann. Hum. Genet. London*., 39, 361-373.
37. Oladipo, G.S., Olabiyi, A.O., Oremosu, A.A. and Noronha, C.C. (2007). Nasal indices among major ethnic groups in Southern Nigeria. *Scientific Res. Ess.,* 2(1), 20- 22.
38. Oladipo, G.S., Chinagorom, E. and Iruoghene, G.O. (2010a). Craniofacial dimensions of Ijaw children of Nigeria. *Biomed. Int.,* 1(2), 25-29.
39. Oladipo, G.S., Okoh, P.D., Akande, P.A. and Oyakhire, M.O. (2010b). Anthropometric study of some craniofacial parameters: head circumference, nasal height, nasal width and nasal index of adult Omoku indigenes of Nigeria. *Am. J. Sci. and Indus Res.,* doi: 10.5251/ajsir.2011.2.54.57.
40. Olotu, J.E., Eroje, A., Oladipo, G.S. and Ezon-Ebidor, E. (2009) Anthropometric study of the facial and nasal length of adult Igbo ethnic group in Nigeria. *The Int. J. Biol. Anthropol.,* 2(2).
41. Overall, A.D.J. (2009). The influence of the wahlund effect on the consanguinity hypothesis: consequences for recessive disease incidence in a society structured Pakistani population. *Human Heredity*, 67(2).
42. Pakistan Census Organization. (1998). http://www.census.gov.pk/.
43. Petraglia, M.D. and Allchin, B. (2007). The evolution and history of human populations in South Asia. *Springer. The Netherlands.*
44. *P*olymeropoulos, M.H., Xiao, H., Rath, D.S. and Merril, C.R. (1990). Trinucleotide repeat polymorphism at the human intestinal fatty acid binding protein gene (FABP2). *Nucleic Acid Res.,* 18(23), 7198.
45. Porter, J.P. and Olson, K.L. (2003). Analysis of the African American female nose. *Plast. Reconst. Surg.,* 111(2), 620-626.
46. *R*ahman, Z., Afroze, T. and Weir, B.S. (2001). DNA typing results from two urban subpopulations of Pakistan. *J. Forensic Sci.,* 46, 111-115.
47. Shah, G.V. and Jadhav, H.R. (2004). The study of cephalic index in student of Guyarat. *Biol. J. Med. College*, 153, 25-26.
48. Shami, S.A., Qaisar, R. and Bittles, A.H. (1991). Consanguinity and adult morbidity in Pakistan. *Lancet.,* 338, 954-55.
49. Shami, S.A., Schmitt, L.H. and Bittles, A.H. (1989). Consanguinity related prenatal and postnatal mortality of the populations of seven Pakistani Punjab cities. *J. Med. Genet.,* 26, 267-71.

50. Tarazona-Santos, E., Carvalho-Silva, D.R., Pettener, D., Luiselli, D., De Stefano, G.F., Labarga, C.M., Rickards, O., Tyler-Smith, C., Pena, S.D.J. and Fabrício, R.S. (2001). Genetic differentiation in South Amerindians is related to environmental and cultural diversity: Evidence from the Y chromosome. *Am. J. Hum. Genet.,* 68(6), 1485-1496.
51. Tvedebrink T. (2010). *Statistical aspects of forensic genetics*. Ph.D. Dissertation. Aalborg University. Denmark.
52. Williams, P., Dyson, M., Dussak, J.E., Bannister, L.H., Berry, M.M., Collins, P., Ferguson, M.W.J. (1995). Skeletal system. *Gray's Anatomy,* 38th ed; Churchll Livingstone, Edinburgh. pp. 609-612.
53. Xu, B., Wang, Y., Ma, J., Li. M. and Xu, L. (2011). A computer-aid study on the craniofacial features of Archang race in Yunnan province of China. *Hua Xi Kou Qiang Yi Xue Za Zhi*, 19(6), 394-396.
54. Yasuda, N. and Kimura, M.A. (1968). Gene-counting method of maximum likelihood for estimating gene frequencies in ABO and ABO like systems. *Ann. Hum. Genet.*, 31, 409-420.
55. Zhang, X.T., Wang, S.K., Zhang, W. and Wang, X.F. (1990). Measurement and study of the nose and face and their correlations in the young adult of Han nationality. *Plast. Reconstr. Surg.,* 85(4), 532-536.

Table 1

Demographic information of the samples (population strata) of the present study. Population strata included Balochi, Muhajir, Pathan, Punjabi and Sindhi. Ethnic fractionalization 'F' was calculated according to Fearon (2003).

S. No.	Population Groups (m)	%Population in Karachi**	Ethnic Fractionalization 'F'
1	Balochi	4.34	0.725
2	Muhajir*	48.52	
3	Pathan	11.42	
4	Punjabi	13.94	
5	Sindhi	7.22	

* The term is used for the people who have migrated from India after 1947 Partition. Also synonymous with Urdu speaking population group.

** According to 1998 census http://www.census.gov.pk/.

Table 2

Distribution of blood group ABO across the five population groups (strata).

S. No.	Populations groups	Blood group A	Blood group B	Blood group AB	Blood group O	Total
1	Balochi	33	35	26	10	104
2	Muhajir	243	429	288	114	1074
3	Pathan	71	112	49	32	264
4	Punjabi	128	249	181	65	623
5	Sind	56	126	77	15	274
6	Agglomerated population	531	951	621	236	2339

Table 3

Sample sizes across the five population groups (strata) selected through stratified sampling for anthropometric measures and STR study

S. No.	Subpopulations	Sample size (anthropometric measures)	Sample Size (STR study)
1	Balochi	62	32
2	Muhajir	122	47
3	Pathan	154	30
4	Punjabi	112	37
5	Sindhi	64	29
6	Total	514	175

Table 4

Multivariate Analysis of Variance (MANOVA) for blood groups ABO and Rh across the five population groups (strata). Population strata included Balochi, Muhajir, Pathan, Punjabi and Sindhi. Significance value was predefined at 0.05. P value shows a highly significant difference across the five population groups (strata).

Multivariate Tests		Effect Value	F	Hypothesis Degree of Freedom	Error Degree of Freedom	P value
Intercept	Pillai's Trace	.904	10961.027	2.000	2333.000	.000
	Wilks' Lambda	.096	10961.027	2.000	2333.000	.000
	Hotelling's Trace	9.397	10961.027	2.000	2333.000	.000
	Roy's Largest Root	9.397	10961.027	2.000	2333.000	.000
Population groups	Pillai's Trace	.009	2.764	8.000	4668.000	.005
	Wilks' Lambda	.991	2.767	8.000	4666.000	.005
	Hotelling's Trace	.010	2.770	8.000	4664.000	.005
	Roy's Largest Root	.009	4.968	4.000	2334.000	.001

Table 5

Analysis of Variance (ANOVA) for the distribution of anthropometric measures. Anthropometric measures included skull length, skull breadth, nasal length and nasal breadth, cephalic index and nasal index. Population strata included Balochi, Muhajir, Pathan, Punjabi and Sindhi. Significance value was predefined at 0.05. P value shows a highly significant difference across the five population groups (strata) for all the anthropometric measures except nasal index.

		Sum of Squares	Degree of Freedom	Mean Square	F	P Value
Skull length	Between Groups	116.430	5	23.286	4.568	.000
	Within Groups	2589.524	508	5.097		
	Total	2705.954	513			
Skull breadth	Between Groups	970.013	5	194.003	10.144	.000
	Within Groups	9715.281	508	19.125		
	Total	10685.295	513			
Nasal length	Between Groups	15.668	5	3.134	7.395	.000
	Within Groups	215.269	508	.424		
	Total	230.937	513			
Nasal height	Between Groups	34.532	5	6.906	6.745	.000
	Within Groups	520.162	508	1.024		
	Total	554.694	513			
Cephalic Index	Between Groups	7007.963	5	1401.593	7.647	.000
	Within Groups	93104.608	508	183.277		
	Total	100112.571	513			
Nasal Index	Between Groups	604.214	5	120.843	1.368	.235
	Within Groups	44866.686	508	88.320		
	Total	45470.899	513			

Table 6

Analysis of Variance (ANOVA) for the distribution of STR alleles across the five population groups (strata). Population strata included Balochi, Muhajir, Pathan, Punjabi and Sindhi. Significance value was predefined at 0.05. P value shows a highly significant difference across the five population groups (strata) for the STR markers TPOX and TH01.

		Sum of Squares	Degree of Freedom	Mean Square	F	P Value
CSF1PO	Between Groups	2.009	4	.502	.551	.698
	Within Groups	314.280	345	.911		
	Total	316.289	349			
TPOX	Between Groups	18.525	4	4.631	2.511	.042
	Within Groups	636.244	345	1.844		
	Total	654.769	349			
TH01	Between Groups	74.640	4	18.660	4.254	.002
	Within Groups	1513.257	345	4.386		
	Total	1587.897	349			

Table 7

Coefficients of Variation (CV) across blood groups ABO and Rh, anthropometric measures and STR markers.

Population groups	Blood groups		Anthropometric measure		STR markers		
	ABO	Rh	Cephalic index	Nasal index	CSF1PO	TPOX	TH01
Balochi	45.75	33.56	15.61	13.46	8.90	15.74	28.81
Muhajir	41.02	29.18	18.32	15.82	8.89	14.40	26.19
Pathan	44.36	29.16	21.65	25.73	11.14	14.84	25.89
Punjabi	39.67	27.85	24.24	15.22	8.04	15.94	18.83
Sindhi	37.50	32.56	22.09	17.07	7.43	14.65	26.47
Total	40.89	29.97	22.20	19.34	8.94	15.29	26.70

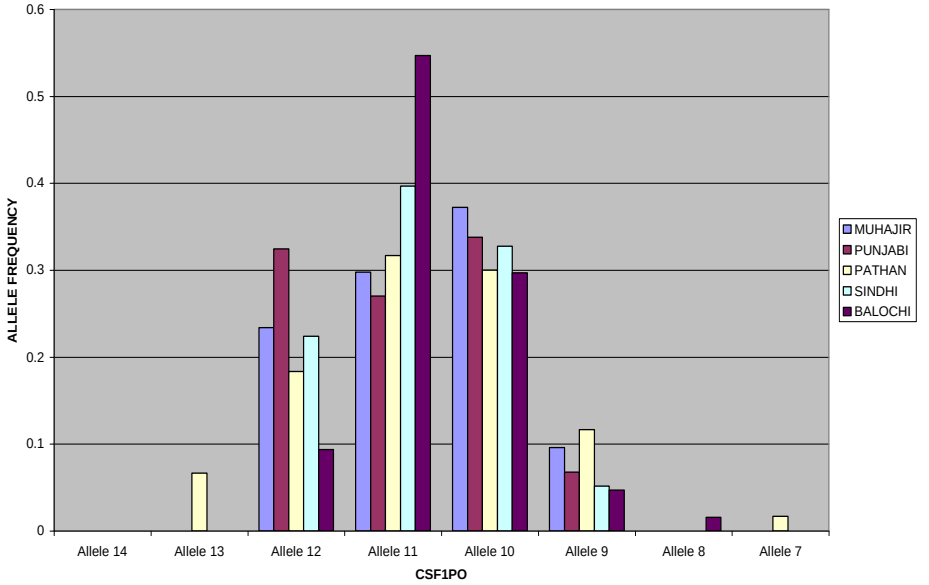


Fig. 1: Allele frequencies for the locus CSF1PO across the five population groups (strata)

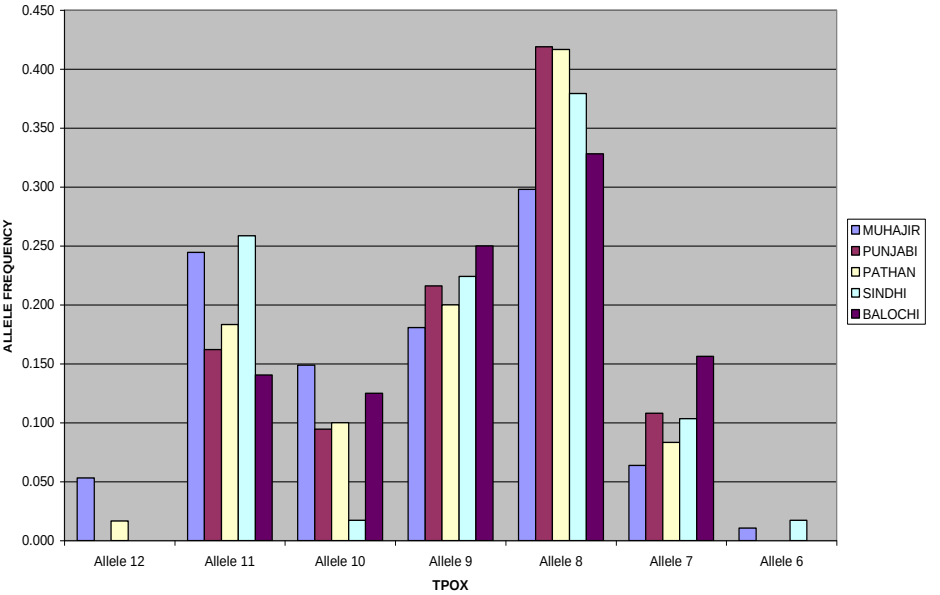


Fig. 2: Allele frequencies for the locus TPOX across the five population groups (strata)

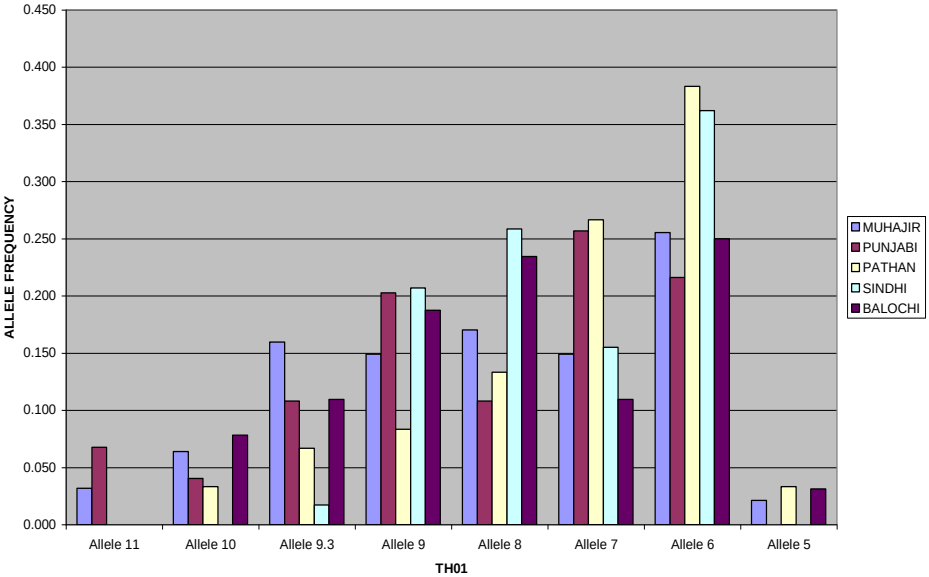


Fig. 3: Allele frequencies for the locus TH01 across the five population groups (strata).

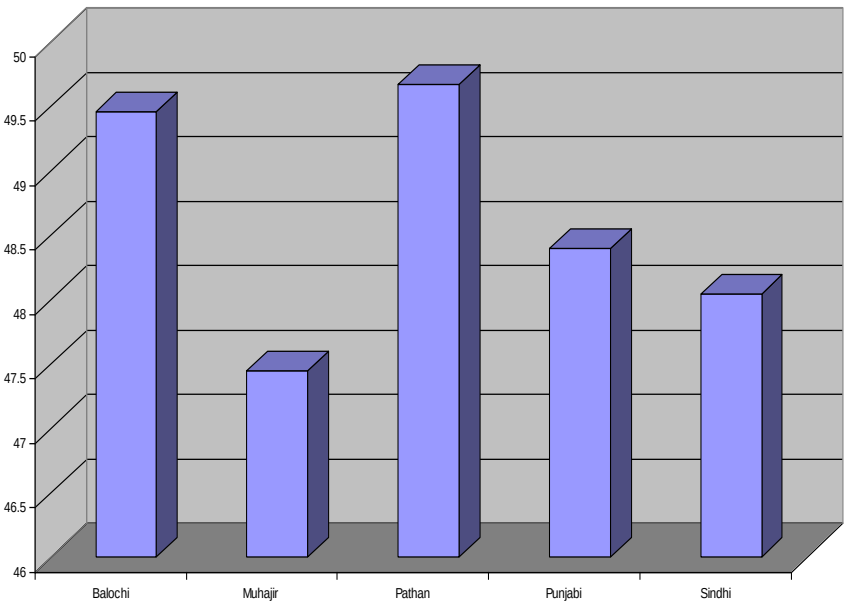


Fig. 4: Cephalic indices across the five population groups (strata)

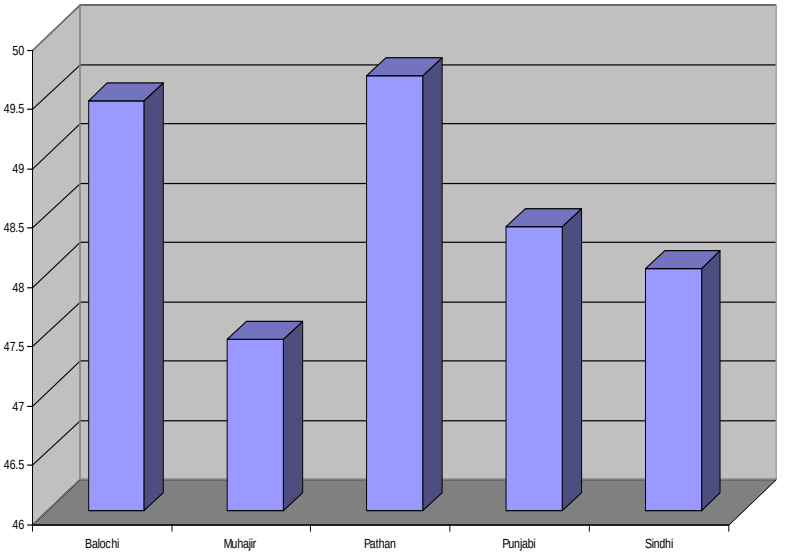


Fig. 5: Nasal indices across the five population groups (strata)

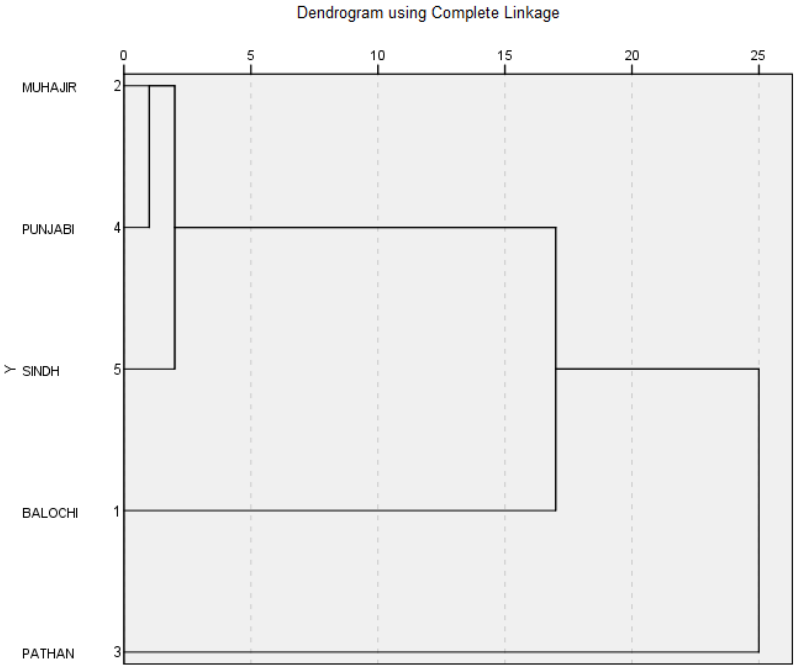


Fig. 6: Dendrogram of cluster analysis for the blood groups ABO and Rh distribution showing relationship across the five population groups (strata) of the present study. Hierarchical cluster analysis was based on complete linkage method and squared Euclidean distance.

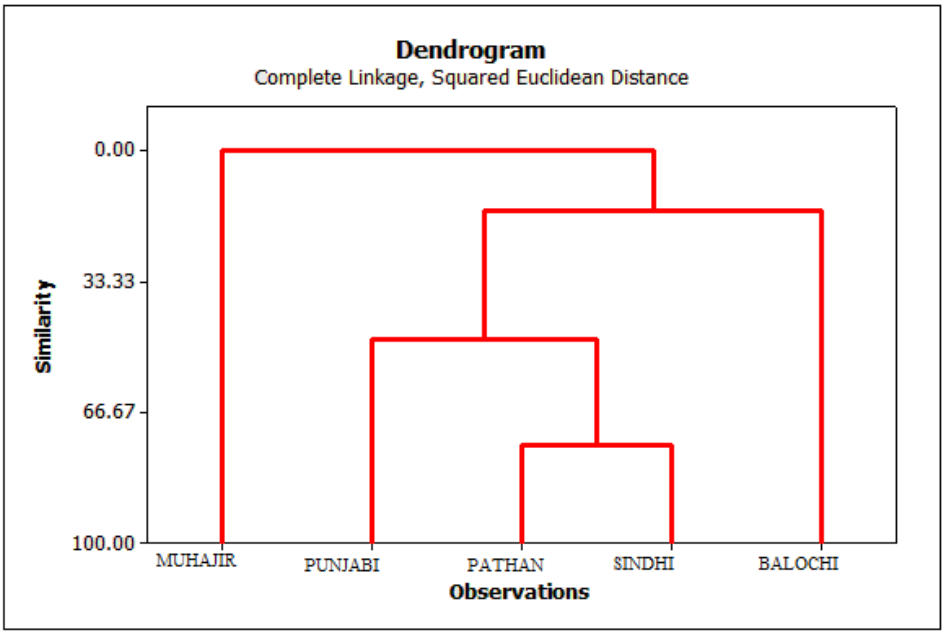


Fig. 7: Dendrogram of cluster analysis for the STR markers CSF1PO, TPOX and TH01 distribution showing relationship across the five population groups (strata) of the present study. Hierarchical cluster analysis was based on complete linkage method and squared Euclidean distance.

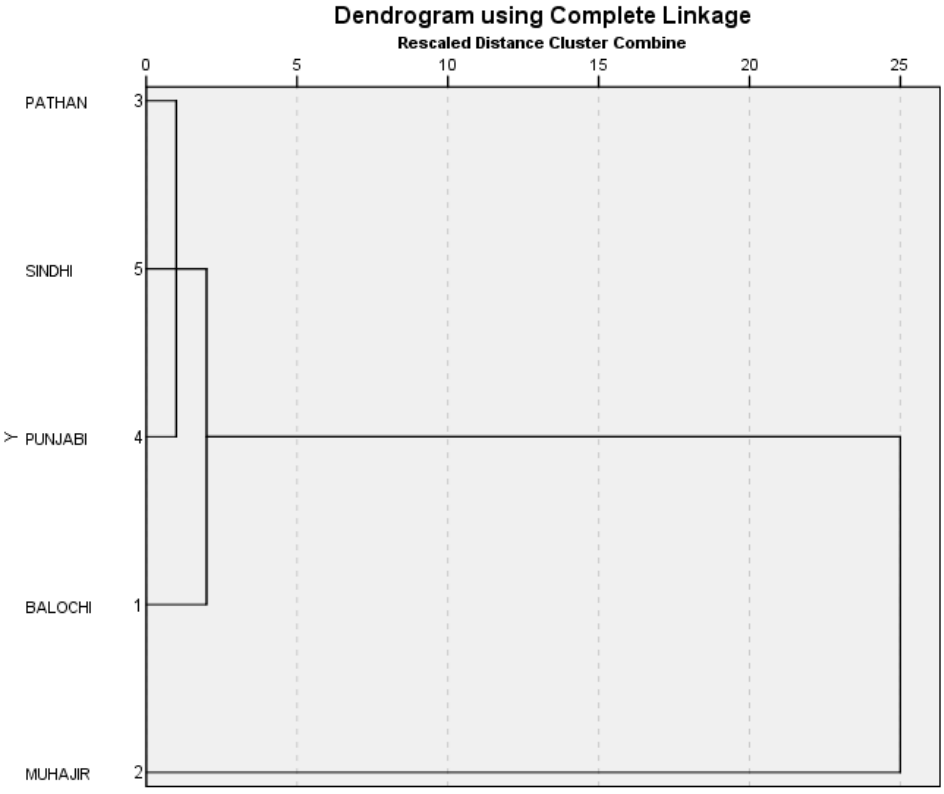


Fig. 8: Dendrogram of cluster analysis for cephalic and nasal indices showing relationship across the five population groups (strata) of the present study. Hierarchical cluster analysis was based on complete linkage method and squared Euclidean distance.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 333-340

CONSUMER BUYING BEHAVIOUR: A FACTOR OF COMPULSIVE BUYING PREJUDICED BY WINDOWSILL PLACEMENT

Irfan Hameed
Business Administration Department, Iqra University, Karachi, Pakistan
Email: irfan_h144@yahoo.com

ABSTRACT

This empirical research investigates the impact of windowsill placement on the compulsive buying behavior of consumers on three different types of products i.e., convenience products, shopping products, and specialty products. Positive effect of windowsill placement on all three types of product categories has been hypothesized. The categorical regression (Optimal scaling) was used to test the hypotheses. The data was collected via self-administered questionnaire from Pakistan through systematic random sampling, and the sample consisted of 500 respondents. The results of data analysis supported only the 1st hypothesis which highlighted that placement of products in shopping centers has an impact of unplanned buying of consumers for convenience products. While rest of the two hypotheses regarding shopping and convenience products were not supported by the data. This research is helpful for those companies which believe in classical conditioning. This is perhaps one of the first study in non-western (Pakistani) context.

KEY WORDS

Windowsill Placement, Compulsive Buying, Convenience Products, Shopping Products, Specialty Products.

INTRODUCTION

This study is an insight into the compulsive buying pattern of individuals on the basis of maneuvers used by marketers. The relationship between windowsill placement and the demand of customers have been studied with respect to different consumer product categories. Windowsill placement has been used as independent variable and types of consumer products have been used as dependent variables namely convenience products, shopping products, and specialty products. Unsought products have not been used for research purpose because of its very nature. Operational definitions of the variables have been given below under the heading of theoretical background.

THEORETICAL BACKGROUND

Consumer Products:

Consumer products are products and services bought by final consumers for personal consumption. Marketers usually classify these products and services further based on how consumers go about buying them. Consumer products include *convenience products*, *shopping products*, *specialty products*, and *unsought products* (Kotler, Armstrong, 2012).

Windowsill placement:

Windowsill placement refers to the placement of the product in the shopping center also known as shelf placement. In this study word windowsill placement denotes favorable position in the shopping center like placing it just behind the cashier, at eye level, or at the entrance of the store. All eye-movement studies of advertising or catalog displays show that visual area strongly increases attention (Janiszewski, 1998; Lohse, 1997). Several shopper surveys (Inman, Winer, & Ferraro, 2009) and field experiments (Chevalier, 1975; Curhan, 1974; Inman & McAlister, 1993; Wilkinson, Mason, & Paksoy, 1982) have shown that large increases in shelf space increase brand sales even when the price and location of the products remain unchanged (Campo & Gijsbrechts, 2005). There is a broad consensus on the following aspects of how people visually process scenes (Henderson & Hollingworth, 1999; Pieters & Wedel, 2007; Rayner, 1998; Wedel & Pieters, 2008).

Compulsive Buying Behavior:

When individuals buy compulsively, they purchase excessive quantities of products that they don't need and cannot afford (Roberts & Roberts, 2012). It is conceptualized as a response to deal with unpleasant life experiences, inner deficiencies or negative feelings (Faber & O'Guinn, 1992; O'Guinn & Faber, 1989). This feeling makes consumers purchase products to help alleviate negative feelings of stress, disappointment, frustration or lack of self-esteem (Scherhorn, 1990).

Hypotheses Construction:

Three hypothesized statements have been made to check the relationship between dependent and independent variable (s). Firstly the relationship of shelf placement on sales of convenience products, secondly the relationship of shelf placement on sales of shopping products, thirdly the relationship of shelf placement on sales of specialty products. The impact of placement has been studied with the help of the statistical test by using Statistical Package for Social Science (SPSS). The test used is categorical regression (Optimal scaling) to interpret the findings that whether placement of products in shopping centers has an impact of unplanned buying of consumers or not. If yes then it is on which types of products.

H_1: *There is a positive impact of windowsill placement in supermarkets on compulsive buying behavior of consumers for convenience products.*

$$\text{Convenience products} = \alpha + \beta\,(\text{Windowsill placement}) + \xi$$

H_2: *There is a positive impact of windowsill placement in supermarkets on compulsive buying behavior of consumers for shopping products.*

$$\text{Shopping goods} = \alpha + \beta\,(\text{Windowsill placement}) + \xi$$

H_3: *There is a positive impact of windowsill placement in supermarkets on compulsive buying behavior of consumers for specialty products.*

$$\text{Specialty goods} = \alpha + \beta\,(\text{Windowsill placement}) + \xi$$

RESEARCH METHODS

Participants and design:

Data has been collected from the public of the most crowded city of the Pakistan (i.e. Karachi). Probability based sampling technique (i.e. systematic random sampling technique) has been used for selecting the sample out of the entire population. Sampling friction have been calculated with the help of this formula

Sampling friction = sample size / population

On average almost one thousand individuals visit each superstore in evening and the target was to select one hundred respondents from each store,

Sampling friction = 500 (100 respondents from each store * 5 stores in total) / 5000 (1000 individuals in each store * 5 stores)

=> 500 / 5000 => 1 / 10 (every tenth respondent was be selected)

Hence every tenth respondent have been targeted for the purpose of data collection. A sample size of five hundred respondents has been used for the research purpose. Hundred respondents were selected from every supermarket altogether five supermarkets were targeted.

Measures:

Self-administered closed ended questionnaires have been used for the purpose of data collection. All questionnaires were standardized and each questionnaire was having sixteen questions in total which were succeeding likert scale having 5 choices and they were coded in this phenomenon. 1 for strongly disagree, 2 for Disagree, 3 for Neutral, 4 for Agree, 5 for strongly agree.

RESULTS & DISCUSSION

Findings and Interpretation of the Results

H_1: *There is a positive impact of windowsill placement in supermarkets on compulsive buying behavior of consumers for convenience products.*

Model Summary

Multiple R	R Square	Adjusted R Square	Apparent Prediction Error
.406	.165	.104	.835

Dependent Variable: Convinience_Products
Predictor: Shelf_Placement

Correlation Coefficient (R) is 0.406 which shows that the relationship between windowsill placement in supermarkets and compulsive buying behavior of consumers for convenience products is moderate. Moreover value of Correlation Coefficient (R) is above 0 hence the relationship between the variables is direct. Coefficient of Determination (R^2) is 0.165 which shows that the 16.5% model is being explained by

the windowsill placement and remaining 83.5% is being explained by unknown variables (which are not taken in to account for the purpose of this research.

Coefficients

	Standardized Coefficients		df	F	Sig.
	Beta	Bootstrap (1000) Estimate of Std. Error			
Shelf_Placement	.406	.133	499	9.294	.000

Dependent Variable: Convenience_Products

The significance value is 0.000 which is less than 0.05 so on the basis of p-value null hypothesis that there is no relationship has been rejected and alternate/research hypothesis that there is positive relationship between windowsill placement in supermarkets and compulsive buying behavior of consumers for convenience products has been accepted.

H_2: *There is a positive impact of windowsill placement in supermarkets on compulsive buying behavior of consumers for shopping products.*

Model Summary

Multiple R	R Square	Adjusted R Square	Apparent Prediction Error
.368	.136	.089	.864

Dependent Variable: Shopping_Products
Predictor: Shelf_Placement

Correlation Coefficient (R) is 0.368 which shows that the relationship between windowsill placement in supermarkets and compulsive buying behavior of consumers for shopping products is weak. Moreover value of Correlation Coefficient (R) is above 0 hence the relationship between the variables is direct. Coefficient of Determination (R^2) is 0.136 which shows that the 13.6% model is being explained by the windowsill placement and remaining 86.4% is being explained by unknown variables (which are not taken in to account for the purpose of this research.

Coefficients

	Standardized Coefficients		df	F	Sig.
	Beta	Bootstrap (1000) Estimate of Std. Error			
Shelf_Placement	.368	.283	3	1.691	.179

Dependent Variable: Shopping_Products

The significance value is 0.179 which is greater than 0.05 so on the basis of p-value null hypothesis that there is no relationship has been accepted and alternate/research hypothesis that there is positive relationship between windowsill placement in supermarkets and compulsive buying behavior of consumers for shopping products has been rejected.

H_3: There is a positive impact of windowsill placement in supermarkets on compulsive buying behavior of consumers for specialty products.

Model Summary

Multiple R	R Square	Adjusted R Square	Apparent Prediction Error
.466	.217	.190	.783

Dependent Variable: Speciality_Products
Predictor: Shelf_Placement

Correlation Coefficient (R) is 0.466 which shows that the relationship between windowsill placement in supermarkets and compulsive buying behavior of consumers for specialty products is moderate. Moreover value of Correlation Coefficient (R) is above 0 hence the relationship between the variables is direct. Coefficient of Determination (R^2) is 0.217 which shows that the 21.7% model is being explained by the windowsill placement and remaining 78.3% is being explained by unknown variables (which are not taken in to account for the purpose of this research.

Coefficients

	Standardized Coefficients		df	F	Sig.
	Beta	Bootstrap (1000) Estimate of Std. Error			
Shelf_Placement	.466	.556	2	.702	.500

Dependent Variable: Speciality_Products

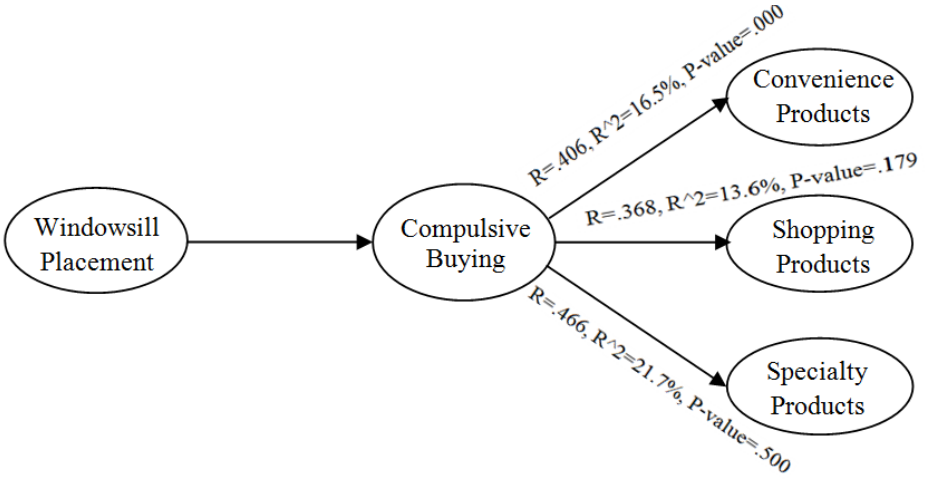
The significance value is 0.500 which is greater than 0.05 so on the basis of p-value null hypothesis that there is no relationship has been accepted and alternate hypothesis that there is positive relationship between windowsill placement in supermarkets and compulsive buying behavior of consumers for specialty products has been rejected.

CONCLUSION, IMPLICATIONS AND FUTURE RESEARCH DIRECTIONS

Conclusion:

There is a positive relationship between convenience products and the independent variable which is windowsill placement. Whereas there is no relationship of windowsill placement of shopping, as well as specialty products in supermarkets. It was proved that people prefer buying specialty products such as, electronics from their original stores.

The points I came across through this research were how the manufacturers promote their product (whether it be by windowsill placement or by salesperson or TV commercials). Customers should know what they want to and do not want to buy. Customers should be aware of the strategic techniques of the manufacturers.



Implications and Future Research Directions:

In order to attract more consumers to buy their products, manufacturers must put in a lot of effort to make their product eye catching by windowsill placement. Stores should be well maintained. Racks should be checked to time and the products where people make decision without much thinking they should be placed at the front. Frequent promotions can be offered in order to enhance awareness about the products. This research can further be carried in other countries. This research has been done by taking in to account a city of Pakistan. Furthermore it can be done on unsought products. In which ways we can enhance the sales of unsought products because of their very nature.

REFERENCES

1. Campo, K. and Gijsbrechts, E. (2005). Retail Assortment, Shelf and Stock out management: Issues, Interplay and Future Challenges. *Applied Stochastic Models in Business & Industry*, 21(4-5), 383-392.
2. Chevalier, M. (1975). Increase in Sales Due to In-Store Display. *Journal of Marketing Research*, 12, 426-431.
3. Curhan, R.C. (1974). The Effects of Merchandising and Temporary Promotional Activities on the Sales of Fresh Fruits and Vegetables in Supermarkets. *Journal of Marketing Research*, 11, 286-294.
4. Faber, R.J. and O'Guinn, T.C. (1992). A Clinical Screener for Compulsive Buying. *Journal of Consumer Research*, 19(3), 459-469.
5. Henderson, J.M. and Hollingworth, A. (1999). High-Level Scene Perception. *Annual Review of Psychology*, 50, 243-271.

6. Inman, J.J. and McAlister, L. (1993). A Retailer Promotion Policy Model Considering Promotion Signal Sensitivity. *Marketing Science*, 12(4), 339-356.
7. Inman, J.J., Winer, R.S. and Ferraro, R. (2009). The Interplay among Category Characteristics, Customer Characteristics, and Customer Activities on In-Store Decision Making. *Journal of Marketing*, 73, 19-29.
8. Janiszewski, C. (1998). The Influence of Display Characteristics on Visual Exploratory Search Behavior. *Journal of Consumer Research*, 25(3), 290-301.
9. Kotler, P. and Armstrong, G. (2012). *Principles of Marketing*. Pearson Education, Inc.
10. Lohse, G.L. (1997). Consumer Eye Movement Patterns on Yellow Pages Advertising. *Journal of Advertising*, 26(1), 61-73.
11. O'Guinn, T.C. and Faber, R.J. (1989). Compulsive Buying: A Phenomenological Exploration. *Journal of Consumer Research*, 16(2), 147-157.
12. Pieters, R. and Wedel, M. (2007). Informativeness of Eye Movements for Visual Marketing: Six Cornerstones. in *Visual Marketing: From Attention to Action*, Michel Wedel and Rik Pieters, eds. Mahwah, NJ: Lawrence Erlbaum Associates, 43-72.
13. Rayner, K. (1998). Eye Movement in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, 12(3), 372-422.
14. Roberts, J.A. and Roberts, C. (2012). Stress, gender and compulsive buying among early adolescents. *Young Consumers*, 13(2), 113-123.
15. Scherhorn, G. (1990). The addictive trait in buying behavior. *Journal of Consumer Policy*, 13(1), 33-51.
16. Wedel, M. and Pieters, R. (2008). A Review of Eye-Tracking Research in Marketing, in *Review of Marketing Research*, Vol. 4, Naresh K. Malhotra, ed. Armonk, NY: M.E. Sharpe, 123-147.
17. Wilkinson, J.B., Mason, J. and Paksoy, C.H. (1982). Assessing the Impact of Short-Term Supermarket Strategy Variables. *Journal of Marketing Research*, 19, 72-86.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 341-346

IMPACT OF MODERN WEB MEDIA ON TRADITIONAL MASS MEDIA

Hina Gul[1], Muhammad Asrar-ul-Haque[2] and **Samar Ikram[3]**
[1] Prontolinks, Karachi, Pakistan. Email: gul.hina08@gmail.com
[2] Allama Iqbal Open University, Islamabad, Pakistan
Email: asrarishere@gmail.com
[3] Freelance Researcher, Karachi, Pakistan
Email: samarmasood@gmail.com

ABSTRACT

In this, paper we explained the impact of web media on mass media and print media specially by using modified basic statistics. The main objective of this research is how we use and transform the measure of scale in social sciences according to our data.

There were two types research methodology used for this purpose first was online data collection and 2nd was face to face pen paper interview but same questionnaire was used in both method, the reason was that we want to captured the actual scenario. A sample of 90 respondents were selected, 50 official and student were interviewed by online method while 30 housewives were interviewed by F2f method. Questionnaire consists on close ended questions. Convenient sampling techniques used in this research, editable soft copy of questionnaire had send to our friends and officials while for the face-to-face randomly selected the housewives.

In this current era, the usage of internet is increasing day by day, therefore almost all the entertainment, sports and news channels are providing web facility of their programs specially news for news channel most of the people who are using internet they also browsing any news web site or other web site related to media during their work. But, females still prefer the mass media and but the print media going back-ward.

1. INTRODUCTION

The very nature of technological evolution dictates the replacement of existing mediums and systems with their modified versions. Sometimes the process of modification results in completely new forms; as happened in the mid-nineties when Television replaced radio as the preferred media for both; businesses and consumers. It can correctly be stated as a tectonic shift that happens once a generation and the shift of old brand to new, is painful and usually fateful to old brands (Marlatt, 1999). A 1999 report in Electronic Markets noted that "advertisers will increasingly migrate to online business–to–business sites." Today, the World Wide Web leads the scenario with its wide reach and convenience. With the popularity and interactivity offered by the Internet and World Wide Web, media organizations see the Web as a medium they must conquer if they are going to survive (Villano, 1999).

TV channels, radio, newspapers and magazines maintain their versatility and have not lagged in their efforts to provide effective information and entertainment but Web media has one undefeatable advantage; adaptability and availability to the consumers' schedule and not vice versa.

2. PURPOSE

It is crucial for a business to reach its consumers at a time, place and state of mind when the consumer is most pliable to the presented information. If the information is self-searched, the chances of its positive impact on the consumers multiply further. Similarly, the deepest desire of an individual is to get what he or she wants at own convenience. Hence, both are highly sensitive to the mediums which offer the most. This research paper focuses on addressing the constantly changing answers to the above questions. The results of this survey are especially useful for the advertising agencies and media buying houses.

3. MATERIAL & METHODS

The research methodology comprises of online data collection and face-to-face pen and paper interviews via random sampling; first was used to collect relevant information from active web media users while the latter targets mass media enthusiasts. The actual scenario was captured by serving all participants with the same pre-structured questionnaire. The total sample of 80 respondents includes working men and women, students, housewives and other stay-at-homes.

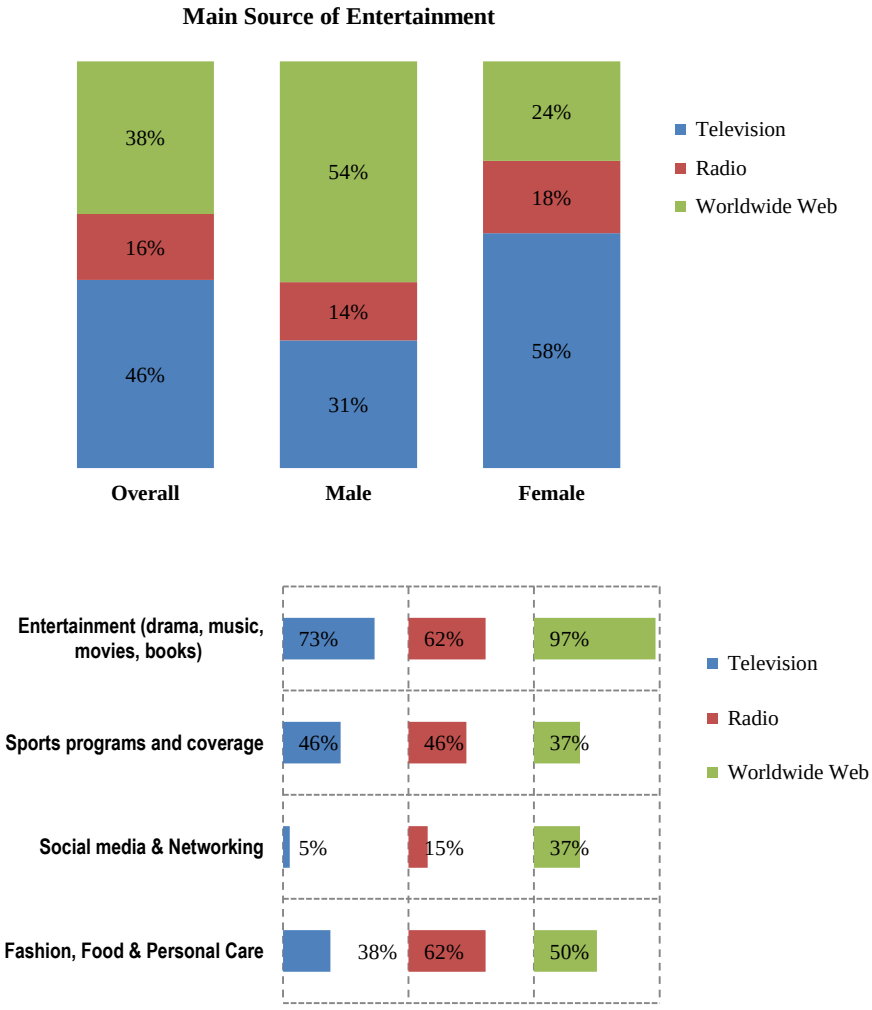
The prospects were contacted via email, social media and in-person visits. Pen and paper interviews were conducted in randomly selected geographical areas. All questionnaires follow self-administered approach therefore respondents educated below intermediate or aged below 18 years have not been selected.

After collecting the data from different media audiences, we have used modified basic statistics to transform the relevant measures of scale in social sciences to suit the research objectives.

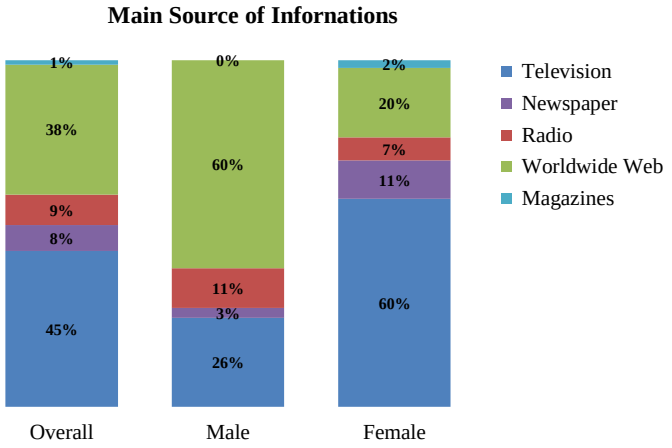
Regression analysis has been run to identify the main attribute/s affecting the web media score. For this regression we have transformed the 5 point rating scale into 100 points (much more~100, a little more~75, about the same~50, a little less~25, much less~0).

4. CALCULATION AND RESULTS

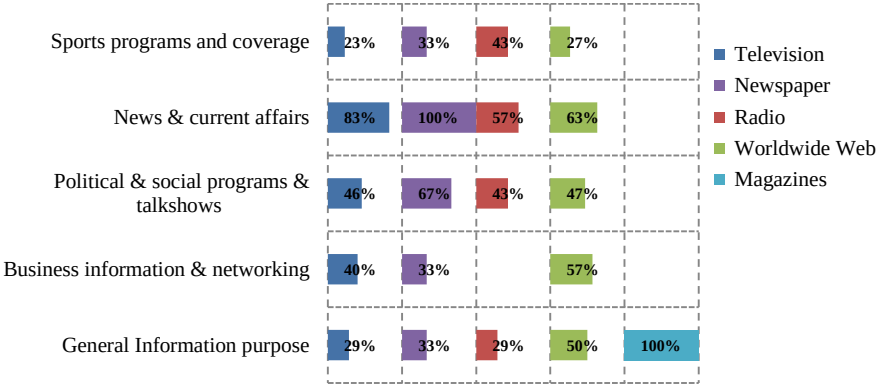
While the television was introduced in mid-nineties, it took around 20 years before its impact hit full force when it started taking consumer time and advertising revenues away from radio and to some degree, newspapers and magazines. Today, Television is an absolute presence in all households. Our research attests to that fact and also shows that 45% of the participants usually spend more than 3 hours watching TV daily. Web media seekers spending a similar amount of time browsing the internet are still at a relatively low 26% and even then males comprise the majority of this percentage (42.9%). This is quite interesting especially when we consider that 86.3% of our research participants do have web browsing and other internet facilities. Radio mostly comes built in with different mobile phones so its availability is high (81.3%). It must be noted that this availability does not encourage consumption which is a quite low 16% on overall basis. Newspaper and magazines are at 63.8% and 26% respectively.



Television still holds the main position for entertainment at an overall 46% but Web is constantly minimizing the gap (38%) particularly among the male entertainment seekers. Approximately 58% females are still considerably more oriented towards television for drama, music and movies because of the convenience of watching it while they take care of their household chores. Also, many females lack the required knowledge to use the web effectively especially where troubleshooting is concerned. Males have, though, shifted heavily towards Web entertainment (54%) but when it comes to sports, television is the preference. Radio is at an all the time low for entertainment at only 16.3%.

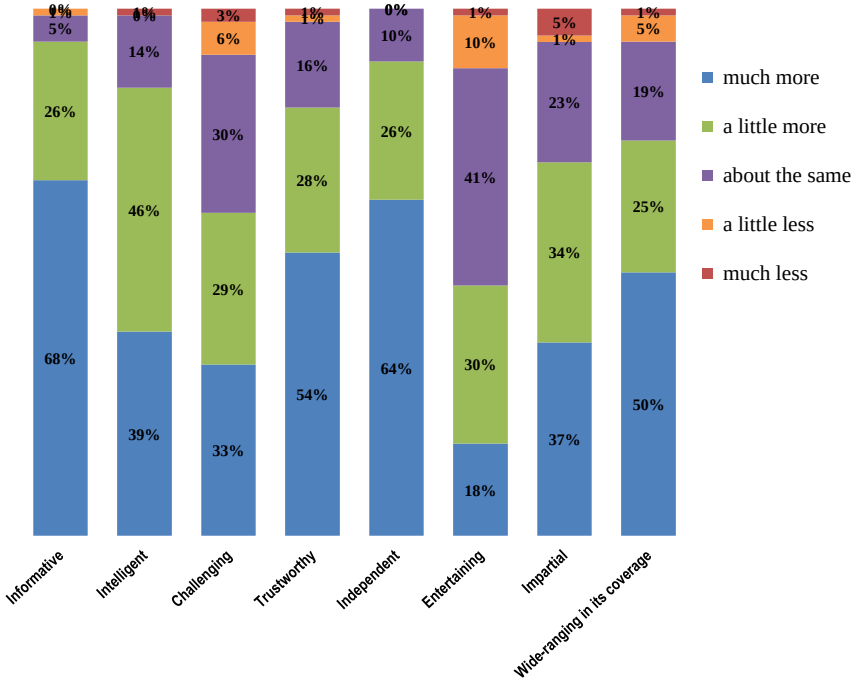


Web media takes second place to television for providing information. Only 38% participants seek this medium for information and 60% of these are males. This situation completely reverses when applied to TV (45%) where 60% of the information seekers happen to be women. News & current affairs (82.9%), political and social programs/talk shows (45.7%) on Television drive this high interest. Web is on par for these talk shows due to availability of video channels and torrents like YouTube, daily motion etc. Web media offers an advantage over all other mediums where business info and networking is concerned (57%).



People consider the web media much more informative, intelligent, challenging, trustworthy, entertaining, impartial and wide-ranging.

The last years have witnessed a steady incline in the time spent on web media. Running the regression analysis has further confirmed that the challenge and independence aspect of web media is actually what makes it so much more attractive and important.



	Overall	Gender		Survey Type	
		Male	Female	online	F2F
Print Media	15.6	15.6	15.6	14.1	18.0
Electronic Media	42.6	31.4	51.2	36.0	53.5
Web Media	41.9	53.0	33.2	*49.9*	28.5

	Unstandardized Coefficients		Standardized Coefficients	Sig.
	B	Std. Error	Beta	
Informative	-0.205	0.148	-0.186	0.169
Intelligent	-0.133	0.131	-0.144	0.310
Challenging	0.163	0.096	0.235	0.096
Trustworthy	0.015	0.116	0.018	0.898
Independent	-0.285	0.151	-0.268	0.063
Entertaining	0.024	0.105	0.032	0.816
Impartial	-0.010	0.097	-0.015	0.915
Wide-Ranging in its Coverage	-0.079	0.091	-0.109	0.389

6. COMMENTS AND CONCLUSION

Web media penetration is rising steadily in Pakistan and Television is the only main competitor blocking its way to the lead position. So far, the convenience of leaving the TV on as one goes through the daily routine and sensational programmes and talk shows are the main barriers for the Web. However, it is considered as more trustworthy and on-the-spot as compared to all other media which is why advertisers are quickly moving large slices of their advertising budgets to interactive websites and video channels like facebook, twitter, google plus, instagram, youtube and daily motion etc. It will not be far-fetched to say that it will definitely be the most preferred medium for both information and entertainment in the next 2 years in Pakistan.

REFERENCES

1. Mediacentral (1999). *First day: Study blames Internet for sharp audience fall off*. http://www.mediacentral.com.
2. Marlatt, A. (1999). *Mark Walsh, CEO of Web only b–b start–up VerticalNet, tells why he believes that print–centric trade publishers are vulnerable in their own markets*. http://www.mediacentral.com/magazines/folio/9971990302.htm.
3. Villano, M. (1999). Mating game: How newspaper in education programs strive to link worlds of print and Web. *Editor & Publisher*, (6 February), 28-31.
4. Timmers, P. (1998). Business models for electronic markets. *Electronic Markets*, 8(2), 3-8.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 347-358

CORRELATION BETWEEN ENERGY TO NOISE RATIO AND BIT ERROR RATE IN CHAOS BASED COMMUNICATION SYSTEM

Nida Ashraf[1], Faisal Afzal Siddiqui[2], Shiekh M. Zeeshan Iqbal[3]
and **Shabana Rizvi[3]**

[1] Department of Applied Physics, University of Karachi
Karachi, Pakistan. Email: brc.nida@gmail.com
[2] Department of Statistics, University of Karachi
Karachi, Pakistan.
[3] Department of Physics, University of Karachi
Karachi, Pakistan.

ABSTRACT

The performance of any communication system is assessed by its bit error rate (BER). Energy to Noise ratio plays an important role in evaluation of a communication system. The application of chaos based communication system is very popular. To check the relationship between energy to noise ratio and BER, WCDMA downlink simulation model as per 3gpp specification is used with chaos codes along OVSF spreading codes. Simulation is done in the presence of an additive white Gaussian noise (AWGN) channel along with QPSK modulation. The simulations and Bit Error Rate (BER) evaluation with respect to Energy-To-Noise ratio are performed using MATLAB.

KEYWORDS

WCDMA; QCSK; OVSF; Chaos map; BER; QPSK; 3gp; correlation.

1. INTRODUCTION

Multi-megabit internet access, communications using voice over internet protocol(VOIP), voice activated calls, unparalleled network capacity, video and high quality image transmission with the same quality as the fixed networks are provided by the 3G systems [1,2]. Direct sequence spread spectrum technology DS-CDMA (Direct-Sequence Code Division Multiple Access) is used in WCDMA system [3]. Having multicode connections by 2 layered coding using OVSF (Orthogonal Variable Spreading Factor) codes for channelization and using pseudo noise code for scrambling. It provides variable data rate [4]. Hence the signals are spread over the complete bandwidth. Supporting bandwidth on demand (BoD) service [5]. Hence a narrow band (data) code modulates a wide band code (PN-sequence), resulting into a noise like wide band signal and then undergoes phase shift keying[1]. Since WCDMA is a 3G GSM standard, it suffers from the same security issues as GSM does. As there is no network authentication. It can face problems with attacks from false base stations (BS) i.e., intrusion resulting in channel or call hijack [6].

This can be overcome by using chaos based communication. The chaotic systems are very sensitive to the initial conditions. Chaotic signals exhibit rapid de-correlation.

Chaotic signals have mixed deterministic / stochastic nature. The receiver should have a duplicate of the transmitter's chaotic signal or synchronize with the transmitter [7]. Direct sequence (DS) spread spectrum communication systems and Chaotic communication systems resemble in a way such that the data is spread over a wide transmission bandwidth and then de-spreaded by the receiver having a time-synchronized spreading sequence .Chaos based communication systems offer potential advantages such as ease of generation of broadband signals, high security, good multiple access capability, resistance to jamming and immunity to interference. The advantage of chaotic system in communications technology is its avoidance of intrusion; it is very difficult to differentiate a chaotic system generated signal from noise or a random signal **[**8].chaotic sequences permit more users in the system and the system obtains a greater security because of the difficulty they present to be reconstructed. Chaotic codes have very low values of the cross correlation function among them [9].

Previously detailed studies have been done by [10, 11] using different aspects of the communication system focussing only the BER performance as it was the key feature of any communication system. In this paper it is the first time that existing WCDMA system is modified by using OVSF codes along with chaotic code generated by different maps

There are several types of mappings, some of them which shall be used are:

1. Logistic map.
2. Quadratic map.
3. Gauss map.
4. Tent map.

A. *Logistic* Map

The logistic map is a mapping of polynomial of degree 2. The logistic map equation is [12]:

$$x_{n+1} = r x_n (1 - x_n) \quad (1)$$

x denote state vector and r denotes the iterative function chaotic map.

B. *Quadratic* Map

Quadratic map equation is [13]:

$$x_n = f_c(x_n) = {x_n}^2 + c \quad (2)$$

x denote state vector and $f(.)$ denotes the iterative function.

C. *Gaussion* Map

Gauss map equation is [14]:

$$x_{n+1} = exp\,(-\alpha {x_n}^2) + \beta \quad (3)$$

x denote state vector and $exp\,(.)$ denotes the iterative function where α and β are real parameters. For some values of those parameters the solution *xi* is chaotic.

D. *Tent* Map

Tent map equation is

$$x_{m+1} = \mu(1 - x_m) \quad (4)$$

x denote state vector and μ $(.)$ denotes the iterative function. Using these map equations we will overcome the short comings of WCDMA system.

In this research QCSK modulation will be done. Where for the symbol s we transmit for half Symbol period the chaotic reference chip

$$r(t) = \sqrt{E_b} C_x(t) \quad (5)$$

and in the second half the information-bearing chip

$$i(t) = \sqrt{E_b} m_s(t) \quad (6)$$

where C_x the chaotic reference is signal (segment of a chaotic waveform) and m_s is message signal and E_b is the energy per bit. [15].

In the present study, Simulation of WCDMA downlink model was done by using the specifications as referred by the third generation partnership project (3gpp). First QCSK is applied then channelization and scrambling is done. Then after QPSK modulation it is transmitted through AWGN channel. It is then received and BER with respect to E_O/N_O is determined.

2. SIMULATION MODEL

The architecture of the system is presented by generalized block diagrams of the transmitter and receiver of the communication system.

A. Transmitter

The block diagram of transmitter of WCDMA downlink simulation model as per physical layer specifications by IMT2000 is shown in Figure 1. After QCSK the Channelization and scrambling is performed. Channelization/spreading is done by a short spread sequence (OVSF code). Scrambling is done by long gold sequence [16]. Then QPSK modulation is performed. Where the in-phase and quadrature components are filtered by a RRC filter of roll off factor equal to 0.22, and then after multiplication with sinusoidal waves having 90^0 phase reversal are summed as $Icos\ \omega t + jQsin\ \omega t$ and then transmitted.

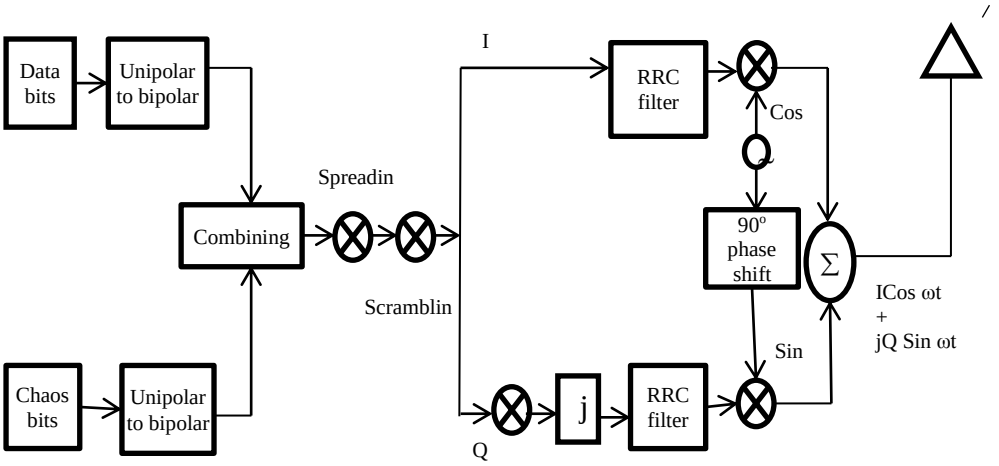


Figure 1: WCDMA Simulator Transmitter Block Diagram

B. *Receiver*

The block diagram of receiver of the simulation model is shown in figure 2. After descrambling and de-spreading. Then data and chaos bits are separated and the data bits are obtained. For different signal to noise ratios the energy to noise ratio (Eb/NO) is calculated later on used to find the bit error rate (ber).

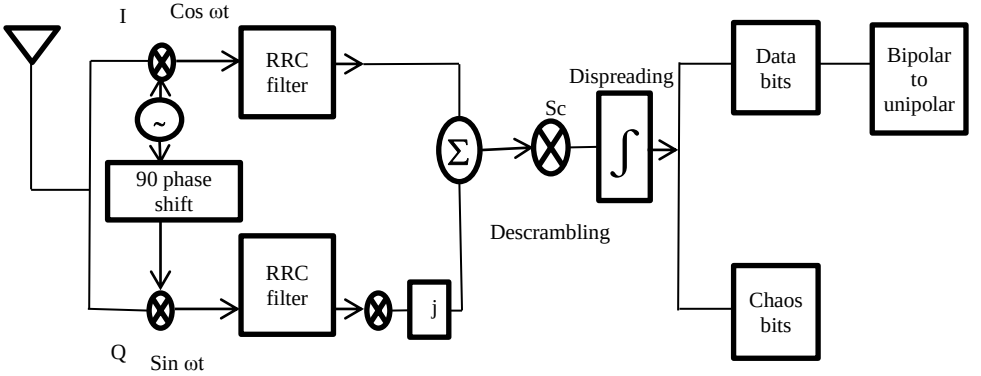


Figure 2: Block Diagram of Receiver

3. SIMULATION RESULTS & DISCUSSION

A. Transmitter simulation results

After performing QPSK modulation the finally obtained signals are represented by Scatter plot

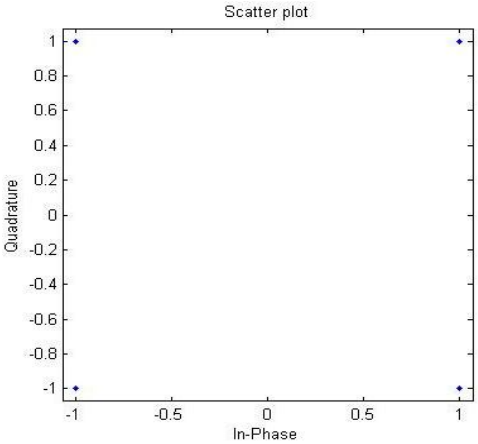


Figure 3: For Logistic Sequence

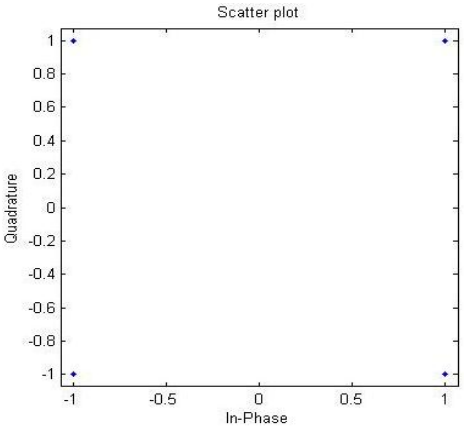


Figure 4: For Quadratic Sequence

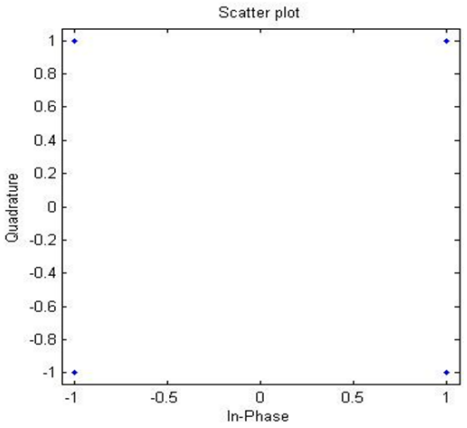


Figure 5: For Gaussian sequence

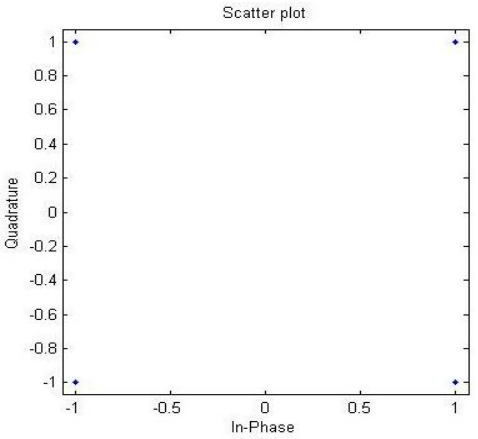


Figure 6: For tent sequence

Then the signal is filtered by a root raised cosine filter with roll off factor of 0.22. This filter is designed by the following equation [17]

$$h(t) = \frac{\frac{4\alpha}{\pi\sqrt{Ts}}\left[\cos\left\{\left(1+\alpha\right)\pi\frac{t}{Ts}\right\}+\frac{1}{4\alpha\frac{t}{Ts}}\sin\left\{\left(1-\alpha\right)\frac{\pi}{Ts}.t\right\}\right]}{\left[1-\left(4\alpha\frac{t}{Ts}\right)^2\right]} \quad (7)$$

where α= roll off factor, then I and Q channels of discussed sequences after filtering are multiplied by sinusoidal waves 90^0 out of phase are transmitted as $Icos\ \omega t + jQsin\ \omega t$ signals.

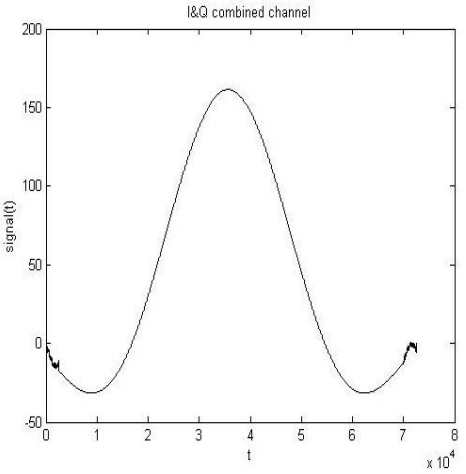


Figure 7: Filtered logistic sequence signal

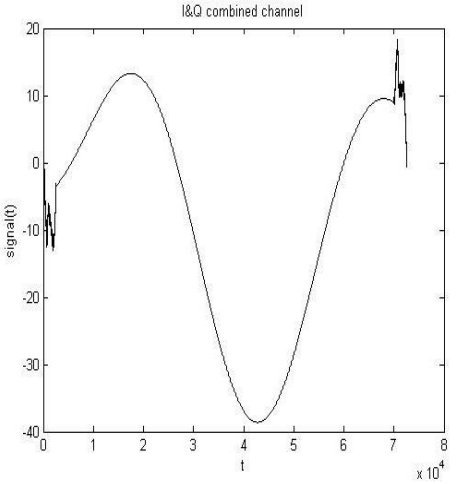


Figure 8: Filtered Quadratic sequence signal

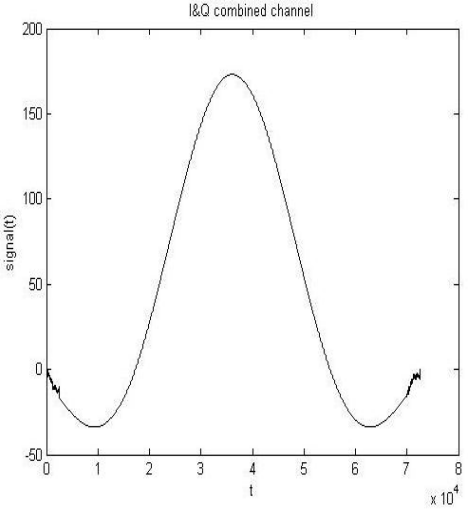


Figure 9: Filtered Gaussian sequence signal

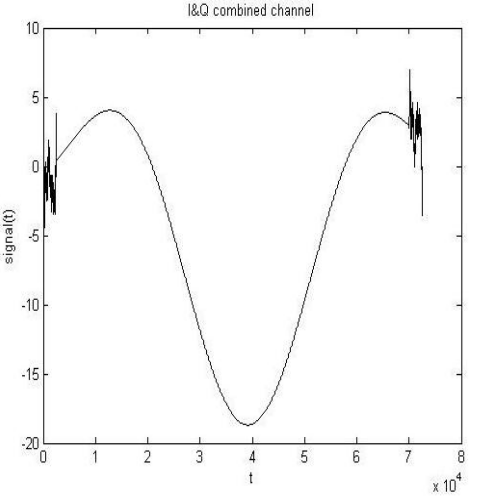


Figure 10: Filtered Tent sequence signal

B. *Receiver simulation Results*

After passing through an additive white Gaussian noise channel (AWGN), the signals received by the receiver are,

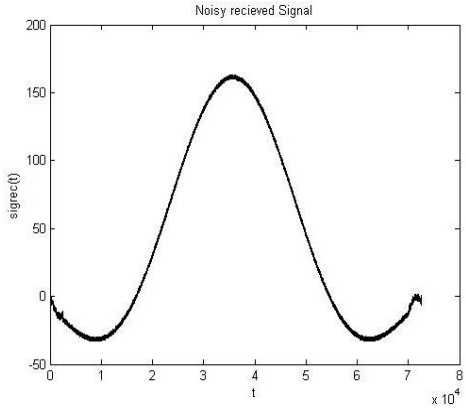


Figure 11: Received Logistic Sequence Signal

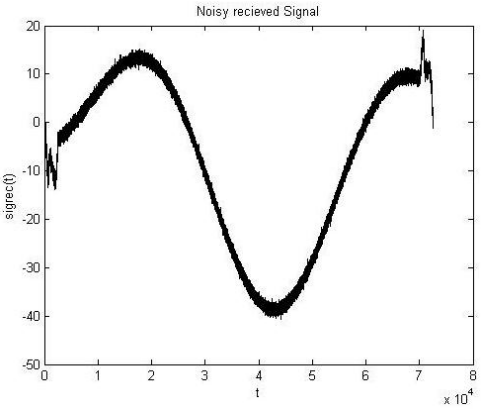


Figure 12: Received Quadratic Sequence Signal

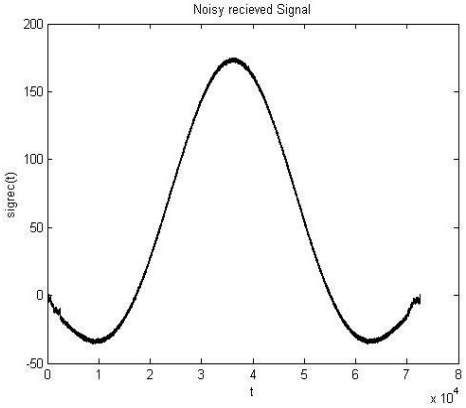


Figure 13: Received Gaussian sequence signal

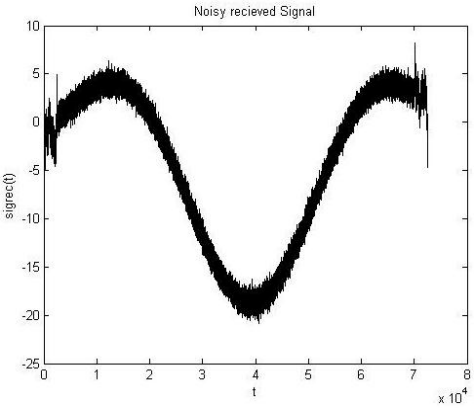


Figure 14: Received Tent sequence signal

As seen from the figures the received signals are distorted due to noise effects. Then I & Q channels of obtained signals are thus multiplied by sinusoidal waves 90^0 out of phase. At the receiver a matched filter is placed, to remove the effects of ISI. The resultant I& Q signals are filtered. The I & Q signals thus obtained then undergo I+iQ,

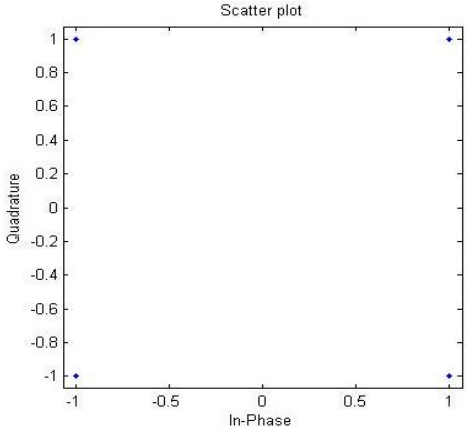


Figure 15: For Logistic sequence signal

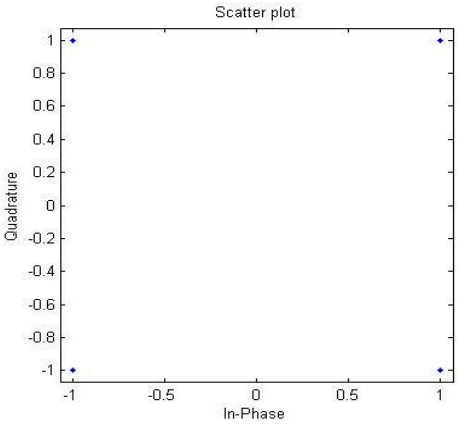


Figure 16: For Quadratic sequence signal

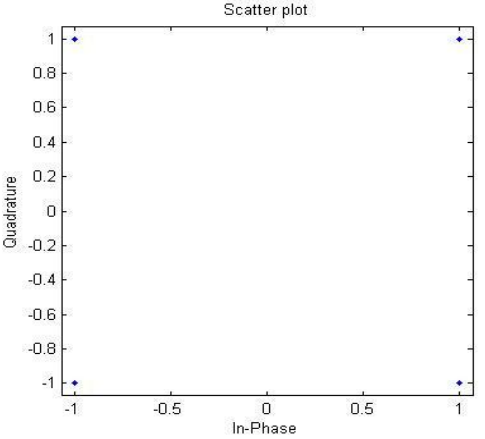


Figure 17: For Gaussian sequence signal

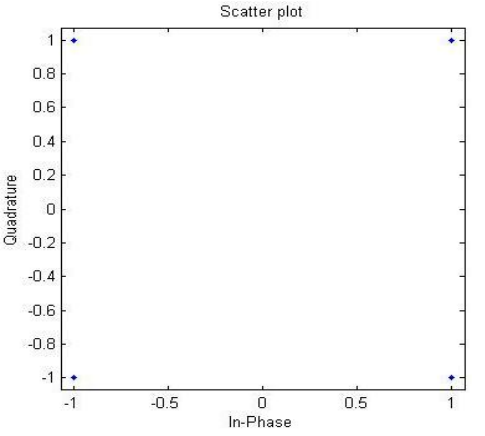


Figure 18: For tent sequence signal

A comparison of scatter plots of transmitted and received signals show that the received code is the same as the transmitted code .i.e., the code security is maintained. And then they are de-spreaded and descrambled then the data and chaos bits are separated, the data bits then undergo bipolar to unipolar conversion to get the data sequence originally transmitted.

C. *Bit Error Rate* Calculation

By using the following universal formula,

$$\frac{\mathrm{Eb}}{\mathrm{NO}}\ \mathrm{Db} = \log\left(\frac{S}{N}\right)_{dB} + \log\left(\frac{BW}{bit\ rate}\right) \tag{8}$$

The energy to noise ratio is calculated,where the BW (bandwidth) is 5MHz for WCDMA,the bit rate is 3.84Mcps, and signal to noise ratios of 20dB,30dB,40dB and 50dB are used.

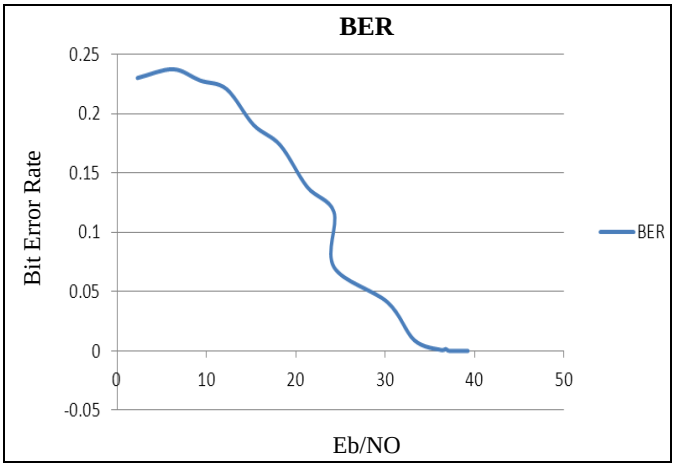


Figure 19: BER Graph with Eb/NO on x-axis

Table 1: Eb/NO Vs. BER

Eb/NO	BER
2.2639	0.2301
6.2639	0.2375
9.2639	0.2281
12.2639	0.2207
15.2639	0.1906
18.2639	0.1734
21.2639	0.1383
24.2639	0.1172
24.2639	0.0707
30.2639	0.041
33.2639	0.009
36.2639	7.81E-04
36.7639	0.0016
37.0139	3.91E-04
37.2639	0
39.2639	0

The bit error rate graph is plotted with the Eb/NO on the x-axis and BER on the y axis. As everything has been developed in Matlab, both the GUI of BERTOOL and the normal methods of checking bit by bit are used.

4. CORRELATION

The correlation between noise and bit error rate is negative 0.9790. The coefficient of determination (r^2) is 95.86 approximately 96%.

5. CONCLUSION

With the increase of energy to noise ratio the bit error rate decreases and the antenna coverage is limited i.e., cell size is limited. The simulations show that bit error rate of Chaos codes gives same performance as the orthogonal OVSF code [18], but with enhanced security. Hence these codes have more security as the chaotic sequence is unique to a particular system. This makes these codes robust. The proposed system provides properties of a Non-coherent system: robust receiver; Spread spectrum system: resistance to interference; Chaotic signals: easy to generate, good correlation properties. The correlation between noise and bit error rate is negative 0.9790. The negative sign shown that both are inversely proportional i.e. one increases other decreases with 97.90 %. The coefficient of determination (r^2) is 95.86 approximately 96%. It shown that we have control the variation till 96% and only 04% are uncontrollable factors.

6. ACKNOWLEDGEMENT

I would like to present a cordial thanks to Miss Fareeda Farrukh in charge department of Applied Physics University of Karachi, for providing lab facilities and thanks to professor doctor Tehseen Rahim chairman department of Physics for expert opinion and sincere advises.

REFERENCES

1. Rappaport, T.S. (2002). *Wireless communications principles and practice, India* 2nd ed, Pearson Education Inc.
2. Oliphant, M.W. (1999). *The Mobile Phone meets the Internet*. *IEEE Spectrum*, 20-28.
3. Tse, D. and Viswanath, P. (2005). *Fundamentals of Wireless Communication*, Cambridge University Press.
4. Adachi, F., Sawahashi, M. and Okawa, K. (1997). Tree-structured generation of orthogonal spreading codes with different lengths for forward link of DS-CDMA mobile radio. *Electronics Letters*, 33(1), 27-28.
5. Ojanpera, T. and Prasad, R. (1998). An Overview of Air Interface Multiple Access for IMT-2000/UMTS. *IEEE communications Magazine*, 36, 88-95.
6. Helsinki University of Technology, S-38.153 *Security of Communication Protocols*. k p.perttula@hut.fi (Accessed April 15, 2013).
7. Feng, J.C. and Tse, C.K. (2008). *Reconstruction of Chaotic Signals with Applications to Chaos-Based Communications*, Singapore, World Scientific Publishing Co Pte Ltd.
8. Michaels, A.J. (2009). Digital *Chaotic Communications*, Ph.D. thesis, Georgia Institute of Technology.
9. El-Khamy, S.E. (2004). New Trends in Wireless Multimedia Communications Based on Chaos and Fractals, *21st National Radio Science Conference (NRSC2004)*, (NTI).
10. Kondo, S. and Milstein, B. Performance of multicarrier DS-CDMA systems. *IEEE Trans. Commun.*, 44(2), 238-246.
11. Kaddoum, G., Gagnon, F., Charge, P. and Roviras, D. (2012). A generalized BER prediction method for differential chaos shift keying system through different communication channels. *Wireless Personal Commun.*, 64, 425-437.
12. Ausloos, M. and Dirickx, M. (2006). *The Logistic Map and The Route to Chaos from the Beginning to Modern Applications*. Springer-Verlag Berlin Heidelberg.

13. Yampolsky, M. and Zakeri, S. (2000). *Mating Siegel Quadratic Polynomials*. The Journal of the American mathematical Society, 14(1), 25-78.
14. Hilborn, R.C. (2004). Chaos *& Nonlinear Dynamics: An Introduction for Scientists and Engineers*, 2nd ed, Oxford, Univ. Press, New York.
15. Galias, Z. and Maggio, G.M. (2001). Quadrature Chaos-Shift Keying: Theory and Performance Analysis, *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, 48(12), 1510-1519.
16. Nagi, P. "*WCDMA Overview*", Network system technologies, LLC. Accessed on February 14, 2012 from www.nstecs.net/yahoo_site_admin/assets/.../WCDMA.15780843.pdf
17. Third Generation Partnership Project Technical Specification Group Radio Access Network Working Group 1, "*Base Station (BS) radio transmission and reception (FDD)*," TS 25.104 V6.8.0 (2004-12).
18. Masud, M.A., Samsuzzaman, M. and Rahman, M.A. (2010). Bit Error Rate Performance Analysis on Modulation Techniques of Wideband Code Division Multiple Access. *Journal of Telecommunications*, 1(2).

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 359-364

DEVELOPMENT STRATEGIES IN OIC COUNTRIES

Khateeb Ullah Khan
Govt. Degree Science and Commerce College, Landhi-Korangi, Karachi, Pakistan
Email: kahteebkhan924@hotmail.com

ABSTRACT

Based on the pioneering work of Clark, Kuznets, and Lewis, Chenery and his associates developed the concept of development strategies calculated from normal and actual values of production and trade orientation. This study deals with the development strategies adopted by the OIC countries, in the period 1990-2012. Chenery and Syrquin in different studies divided the strategies into four division viz. primary specialization, balanced production and trade import substitution and industrial specialization. The criteria for division into different strategies is based upon TO (trade orientation) and PO (production orientation). EL (export level) that shows openness of an economy and foreign capital inflow are also taken into consideration. Islamic countries that follow Chenery (1975) criteria of income and population size for the calculation of development strategies are included. The year 2005 has been taken as base year for the above mentioned criteria. Positive values of TO and PO are termed as high and likewise negative values are divided into normal and low. Width of normal range according to Chenery's criteria, includes roughly 40 percent of the observations for each measure except capital inflow i.e. categorized as high (over 4%) normal (0 to 4 %) and negative values.

1. INTRODUCTION

Building on the foundations of the work of Simon Kuznets, Colin Clark and Arthur Lewis, Chenery explained the problems of economic growth and an econometric framework was chalked out that eased the analysis of range of socio economic indicators. Average patterns of development provide a good starting point to find out the diversity in the country patterns of development. The typology of development strategies found on the departure of each country from the average pattern of production and exports.

Exports include total exports, primary exports and manufacturing exports (calculated from the formulae given in the appendix). All exports are taken as share of GDP. Trade orientation (TO) is the difference between actual and normal trade bias for the year 2005. Production orientation (PO) is similar to TO and this is difference of actual and estimated production bias. The production and trade biases are calculated from the deviations between primary and manufacturing sectors of production and trade.

Apart from TO and PO, openness measured by export level and also the net resource inflow have been taken into consideration. Chenery (1975) criteria for the division of values of the aforementioned measures are used with definition of categories of strategies. The four fold strategy is adopted in the above mentioned work of Chenery.

The aim of this paper is find out the patterns of allocation or development strategies in organization of Islamic cooperation (OIC) that comprises 57 member states and is regarded as world second largest after UN. This division into different categories of allocation patterns or strategies would be an instructive instrument for the policy maker to see as to what extent their execution is the line of their development plans. It would also try to regulate the resources of OIC for the beneficial use of the community as a whole.

2. METHODOLOGY, ECONOMETRIC FRAMEWORK AND DATA

We follow Chenery (1975) methodology to analyze the development strategies with the help of actual and estimated variables and calculated values like TO (trade orientation), PO (production orientation), EL (export level) and F (capital inflow). The average or 'normal' values are obtained with the help of regression model by substitution of country parameters like income per capita and population size expected values are obtained for the calculation of above mentioned values to be compared for individual country strategy.

Trade orientation (TO) is the difference between the expected and actual trade bias that is

$$TO=T-\hat{T}$$

where $\hat{T}$ is the normal or expected trade bias,

$$\hat{T}=(\hat{E}p-\hat{E}m) / \hat{E}$$

and likewise T is calculated from actual values.

$$TO=(Ep-Em) / E - (\hat{E}p-\hat{E}m) / \hat{E},$$

PO is analogous to trade orientation with the difference that no divisor is used in it.

$$PO=P-\hat{P} \text{ and}$$
$$\hat{P}= \hat{V}p -\hat{V}m \text{ and } P=(Vp – Vm)$$
$$PO=(Vp – Vm) – (\hat{V}p -\hat{V}m)$$

and openness measured by exports level is the ratio of actual exports and normal exports, symbolically,

$$EL= E / \hat{E}$$

and also F=capital inflow= M (imports) – E(exports)

Conceptually growth is better represented by logistic curve but for simplicity regression model is used by economist, Chenery (1975) and onwards used, a semilog model for this purpose. The squared terms have been beneficial for capturing non-linearities. Following model has been used for the purpose of calculating the expected values of exports and value added sectors,

$$X= \alpha+\beta_1 \ln Y+\beta_2 (\ln Y)^2 +\gamma_1 \ln N +\gamma_2 (\ln N)^2$$

where

X=dependent variable (here this corresponds to value added production in different sectors and total exports and exports in different sectors),

Y=GNI per capita (US$2005),
N=population in millions of inhabitants,
α, β and γ are the parameters to be estimated.

Pooled regression has been carried out for cross section analysis. For OLS the model can be directly run on stacked (or concatenated) groups. The countries are divided into large and small according to their population size above 20 million or less in the year 2005. Regression models are run separately on both the divisions of population and the required normal values, for the income and population size of the year 2005, are obtained. Table 1 shows the categories of TO, PO and F values. TO and PO values have been categorized under Chenery (1975) criteria that the normal range contains roughly 40 % of the observations and high and low categories are above and below normal. For F, the net resource inflow, the criteria is bit different, high is greater than 4%, negative values are in low group and in between is normal range of capital inflow.

Table 1
Categories of TO, PO and Capital Inflow (F) Values

	High	Normal	Low
TO	> +1.20	+1.20 to 0.20	< 0.20
PO	>+0.40	+ 0.40 to -0.04	< -0.04
F	>4%	0% to 4%	-ve values

On the basis of the above criteria the countries are divided into one of the following four categories of strategies for development.

Primary specialization
1. High TO
2. High PO
3. High EL (usually)

Balanced production and trade
1. Normal TO
2. Normal PO

Import substitution
1. High TO
2. Normal to low PO
3. Low total exports

Industrial specialization
1. Low TO
2. Low PO

Data in this study corresponds to the period 1990-2012. The year 2005 is the base and reference year. All variables are collected from the official website of the World Bank WDI (data.worldbank.org). All variables (dependent) are taken as shares of GDP. Exports of primary and manufacturing as share of GDP are calculated with the help of formulae (see appendix). Value added primary corresponds to ISIC rev 3 top level A, B, C, E and F, whereas value added manufacturing corresponds to top level D. Capital inflow is also taken as share of GDP. OIC members list comprises 57 countries of which 49 pass the Kuznets (1971) criterion of one million population. Bahrain, Brunei-Darussalam, Comoros, Djibouti, Guyana, Maldives, Qatar, and Suriname are excluded. (Palestine is not included in WDI list of countries). Further eleven OIC members are rejected due to want of data. These are Afghanistan, Chad, Guinea-Bissau, Iraq, Kuwait, Lebanon, Libya, Oman, Somalia,

Turkey, and Uzbekistan. Only twenty out of remaining 38 countries, due to data discrepancies, are presented in table 2.

4. RESULTS AND DISCUSSION

Table 2 depicts all the relevant data required for portraying the strategies adopted by the OIC member states and classification of countries into one of the strategies. The countries with asterisk are borderline cases for that strategy. The countries for which strategy was ambiguous in the year 2005 value of terminal year (or near to that year) was taken. Last column shows strategy year if it is other than 2005

Primary specialization

As evident from the values of capital inflow that three out of four countries included in this category show capital outflow i.e. the values are with negative signs The Gambia is an exception. The capital inflow is recorded low in this category. Countries are small in size of population. Strategies for all except Turkmenistan are for 2005. In UAE the expected or normal exports are much higher than that of the actual one for the reference year resulting in the very low level of exports. Income per capita shown by GNI per capita (US$2005) is highest in UAE of the magnitude of more than 44000 USD 2005, the Gambia in 2005 earns just above 400 USD 2005. Other two countries Azerbaijan and Turkmenistan are above 1300 USD 2005.

Primary specialization is referred to as the strategy of deferred industrialization; initial primary strategy provides a basis for rapid industrialization (Chenery (1979)). But growth concentrates in few sectors only and also depend on foreign investment. Development of manufacturing sector for exports with already risen wage and income scenario becomes problematic.

Balanced trade and production

This category contains nine countries of which the richest is Saudi Arabia with almost 14000 GNI per capita (US$ 2005)and Senegal the poorest one earning below 800 per capita US$ 2005. Majority of the countries in this group is small. The most populous is Pakistan and Algeria is the second highest with about 180 million and 31 million population size respectively. Mauritius earning moderately at almost 7000 US$2005 level is limited with very tiny population of just over one million inhabitants. Cameroon and Cote D' Ivoire show greater openness in this class of countries. Mauritius has been expected to have more exports than actually it does showing the meager value for openness or exports level. Capital inflow shows mix picture with more positive values in this category. Countries with asterisk are borderline cases.

Chenery (1979) points out that countries with poor resource base or external demand required for sustainable expansion of primary exports and the deliberate choice of these countries to pursue industrialization to avoid the undesirable features of primary specialization may be the causes of the countries to be less specialized and falling in the normal category of TO and PO.

Import Substitution

Both the countries included in this group, Kyrgyz Republic and Sierra Leone earn less than 500 US$2005 and are small countries. Openness is moderate to low due to low

actual exports compared with the normal values. Capital inflow is around 20 percent of GDP. Strategy for Sierra Leone corresponds to the year 2012.

Table 2:
Development Strategies of OIC 2005
(Other Years Indicated in the Last Column)

COUNTRY	Y	N	TO	PO	E^	E	EL	F	Year
PRIMARY SPECIALIZATION									
Azerbaijan	1382.167	8.392	1.450	0.518	0.648	0.629	0.972	-0.100	
Gambia, The	418.877	1.437	1.398	0.483	1.474	0.328	0.222	0.168	
Turkmenistan	1310.942	4.501	1.804	0.506	0.922	0.955	1.036	-0.146	2000
UAE	44229.577	4.149	1.451	0.487	3.146	0.676	0.215	-0.156	
BALANCED TRADE AND PRODUCTION									
Algeria	2378.604	31.719	1.043	0.385	0.468	0.412	0.880	-0.198	2000
Jordan	2387.662	5.411	0.593	0.002	1.024	0.528	0.515	0.416	
Kazakhstan	3404.504	15.147	1.074	0.132	0.776	0.535	0.690	-0.088	
Saudi Arabia	13540.450	24.690	1.086	0.389	1.484	0.571	0.384	-0.321	
Tunisia	3055.680	10.029	0.218	0.007	0.862	0.449	0.521	0.004	
Cameroon	898.289	19.098	0.806	-0.035	0.283	0.310	1.095	0.019	
Cote d'Ivoire*	894.322	18.260	1.028	-0.055	0.293	0.465	1.585	-0.077	2008
Mauritius*	6558.643	1.291	0.912	0.420	2.418	0.548	0.227	0.111	
Senegal*	787.338	13.726	0.616	-0.069	0.350	0.243	0.694	0.154	
Pakistan	811.149	179.160	0.294	0.060	0.220	0.123	0.561	0.079	2012
IMPORT SUBSTITUTION									
Kyrgyz Republic	459.396	5.163	1.418	0.137	0.682	0.387	0.568	0.190	
Sierra Leone	291.215	4.493	1.526	0.365	0.749	0.086	0.116	0.215	2002
INDUSTRIAL SPECIALIZATION									
Malaysia	5309.309	25.843	-0.182	-0.097	0.875	1.129	1.290	-0.219	
Morocco	1923.214	30.125	-0.099	-0.053	0.398	0.323	0.813	0.056	
Egypt, Arab Rep.*	1521.895	80.722	0.035	0.097	0.270	0.174	0.646	0.023	
Uganda*	378.100	35.148	-1.207	-0.016	0.084	0.237	2.825	0.106	

Industrial specialization

In OIC member countries four are classified as following industrial strategy . Malaysia's income per capita tops the category with above 5000 GNI per capita US$ 2005 and the least income country is Uganda with about 400 per capita income level. Two countries are border line cases (Egypt and Uganda). All the industrial countries are large with population ranging from 25 to 80 million. Actual exports in Malaysia and Uganda exceeds that the normal one suggesting openness in these countries, whereas the remaining two Morocco and Egypt show less than unity value of exports level. Only in Malaysia capital inflow is negative while in Morocco and Uganda it is high of the order of 6 and 11 percent.

5. CONCLUSION

Majority of the countries are import surplus. Primary specialization countries are mainly exporting countries with negative capital inflow and their high TO and PO values

hint towards production and exports of the primary goods. Industrial specialization and import substitution countries are import surplus with more imports than exports.

The classification of development patterns into different strategies may provide a basis for evaluating development performance as the plans and implementation on them and results there from, as evident in table 2, may differ.

The relationship between patterns of development and growth rates and income and income distribution may be established subsequently as was done in Chenery and Taylor (1968) growth rates and income distribution analyzed in Adelman and Morris (1971). Chenery (1975) provides some evidence of more equal distribution in industrial specialization countries and lesser equality in primary specialization.

REFERENCES

1. Adleman, I. and Morris, C. (1971). An Anatomy of Income Distribution Patterns in Developing Nations. A summary of findings. *IBRD Staff Working Paper No. 116*. IBRD.
2. Chenery, H.B. and Taylor, L. (1968). Development Patterns: Among countries and Over Time. *The Review of Economics and Statistics*, 50, 391-416.
3. Chenery, H.B. and Syrquin, M. (1975). *Patterns of Development 1950-70*. London: Oxford University Press.
4. Chenery, H.B. (1979). *Structural Change and Development Policy*. London: Oxford University Press.
5. Clark, C. (1940). *Conditions of Progress*. London: MacMillan.
6. *Data.worldbank.org* official website of the World Bank for data collection.
7. Kuznets, S. (1971). *Economic Growth of Nations: Total Output and Production Structure*. Cambridge, MA: Harvard University Press.
8. Kuznets, S. (1956-67). Quantitative aspects of Economic Growth of Nations. *Economic Development and Cultural Change*. (A series of 10 articles).
9. Lewis, W.A. (1954). Economic Development with Unlimited Supplies of Labor. *The Manchester School*, 22, 139-191.
10. Syrquin, M. and Chenery, H.B. (1989). *Patterns of development 1950-83*. Washington D.C: The World Bank.

APPENDIX

Ep=(1-ME).MR.XPR

Em=XPR.ME.MR

XPR= Exports of goods and services as share of GDP. It is obtained by dividing the exports of goods and services (% of GDP) by 100.

MR=merchandise exports as share of exports of goods and services. It is obtained by dividing merchandise exports (current US$) by exports of goods and services (current US$).

ME= manufactured exports as percentage of all merchandise exports is taken as share by dividing it by 100.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 365-380

POPULATION PROJECTION USING MODIFIED MARKOV CHAIN

Muhammad Zakria[1] and **Faqir Muhammad**[2]
[1] Department of Mathematics & Statistics, Allama Iqbal Open University, Islamabad, Pakistan. Email: zakria@aiou.edu.pk
[2] Department of Management Sciences, Air University, Islamabad, Pakistan. Email: aioufsd@yahoo.com

ABSTRACT

Modelling the drastic increase in the population is the major interest area of the social scientists, particularly of the developing countries. Usually, the exponential and regression models as well as time series models are used to project the population of any territory for some years ahead. These methods, due to their limitations, can't project the population of different age groups. The proposed Modified Markov Chain method is introduced to project the population of different age groups. On the other hand, to examine the disparity of the population, Gini coefficient and Lorenz curve are calculated. These inequality measures also showed some disparity within the population census 1972; 1981; 1998. The pattern of age distribution of projected population from 2001 to 2021 seems different as compared to the pattern of age distribution of population censuses 1981 and 1998, whereas the overall population size has also increased. The Component growth model was also used to project the population of different age groups of the population.

KEY WORDS

Modified Markov Chain; Disparity; Gini coefficient; Population census.

1. INTRODUCTION

Population studies focused on projection have become one of the most interesting research areas for demographers. It is fact that all the policies and managements are based on the existing and projected population. Furthermore, the age and sex distribution has significance effects on the socio economic development of the country. Currently, Pakistan was ranked 6th populous country in the world with population 193.24 million (International Data Base [IDB], 2013) whereas in 1950, total population of Pakistan was 34 million population (Statistics division, 1972). Overall, the population of Pakistan is increased more than five times during the last sixty six years. Naturally, such drastic increase in population size complicates the future planning and projects of Pakistan. Loh and George (2007) scrutinized the effect of net international migration on the Canadian population and its age sex distribution during the next 50 years using the base population of 2005. It was concluded that if net international migration continues, the size of the population may increase but the age sex distribution would not be increased as desired by Canadian Government. Moreover, it was also reported that if international migration continues, the percentage of the dependent population (<15 and 65+) is expected to be 44.3%, 61.2% and 68.7% for the years 2005, 2031, 2056 respectively. Similarly, these

percentages would be 68.0% and 80.9% in 2031 and 2056 respectively without considering the net international migration. Keyfitz (1964) projected the Norwegian brown rats female population as well as for some hypothetical beetle's population using the matrix approach. The main requirement of this technique was the probabilities of survival of group of objects. These probabilities were obtained from the female population of aged 0-45 during the years 1940-1955. Later on, the whole population was divided mainly into three groups i.e. 0-14, 15-29. 30-44 and projection was made for the next 60 years up-to the 2000. One million population of an age 0-14 was used as an initial vector. Nichols, Sauer, Pollock and Hestbeck (1992) estimated the transition probabilities for stage-based population projection matrices using capture-recapture data. Fujiwara and Caswell (2002) used the multistage mark recapture method (MSMR) on the North Atlantic right whale (Eubalaena glacialis) data and transition probabilities were converted into matrix form using MSMR method. In this article, the total population of Pakistan as well as population within different age groups are projected for next 20 years using modified Markov chain. In addition to, the age sex distribution of population of Pakistan within different age groups particularly; childhood (0-4) and dependent population (0-15 & 64+), working group (15-64) and ageing population (64+) are also projected. Age Inequality is the most important aspect for population needs of any country, particularly of developing countries like Pakistan. The population inequality overtime is also discussed examine the disparity within age distribution of population of Pakistan.

This article is organized as follows. In § (2) some basic and methodological background is given. The following § (3) includes a result's discussion. In § (4) some conclusions are drawn.

2. MATERIAL AND METHODS

The most authentic and reliable empirical data about last three population census of Pakistan i.e. 1972, 1981 and 1998 have been used in this study (Statistics Division, 1972; 1984; 2001). To project the population of different age groups as well as total population of Pakistan upto the year 2021, the proposed augmented transition probability matrix has been used.

Ross (2008) described the one step transition probability matrix (P_m) with one step transition probabilities p_{ij} i.e. the probability of an object of moving from state *i* to state *j*. Since the probabilities are nonnegative and the process must make a transition into some state, such that

$$p_{ij} \geq 0,\ i,j \geq 0;\ \sum_{j=0}^{n} p_{ij} = 1,\ i = 0,1,\ \ldots,n$$

$$P_m = \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} & \cdots & \cdots & \cdots & \cdots & p_{0n} \\ p_{10} & p_{11} & p_{12} & p_{13} & \cdots & \cdots & \cdots & \cdots & p_{1n} \\ p_{20} & p_{21} & p_{22} & p_{23} & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & p_{ij} & \cdot & \cdot & \cdot \\ p_{10} & p_{11} & p_{12} & p_{13} & \cdots & \cdots & \cdots & \cdots & p_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{n0} & p_{n1} & p_{n2} & p_{n3} & \cdots & \cdots & \cdots & \cdots & p_{nn} \end{bmatrix} \quad (2.1)$$

and the following $X^{(o)}$ represents an initial row vector

$$X^{(o)} = [x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad \cdots \quad x_{n-1} \quad x_n] \tag{2.2}$$

A proposed augmented transition probability matrix of order 9x9 similar to that of 2.1 was developed to project the age distribution of population of Pakistan (see matrix 2.3). Each object has only two options, an object may go to the next state but can never be returned to previous state or may remain in the same state (only if object passed away). In real life population phenomena, the new born babies are always more in numbers than passed away individuals from age 0 to age 70+. Actually, these new born babies compensate the declined population due to any reason and cause to increase in total population. Although, the census population in different age groups belongs to different birth cohorts but the new born babies are always positioned in first age group. For this purpose, a one-step transition probability matrix (P_m) of order n x n was created. The n^{th} column consists of the probability of dying of each state/ age group whereas the n^{th} row named source/sink state. The first and last entry of source/sink state are "r" and "1-r" respectively where "r" represents the relative proportion of total population considered as new born babies.

$$P_m = \begin{bmatrix} 0 & p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p_1 \\ 0 & 0 & p_2 & 0 & 0 & 0 & 0 & 0 & 1-p_2 \\ 0 & 0 & 0 & p_3 & 0 & 0 & 0 & 0 & 1-p_3 \\ 0 & 0 & 0 & 0 & p_4 & 0 & 0 & 0 & 1-p_4 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{n-2} & 1-p_{n-2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-r \end{bmatrix} \tag{2.3}$$

where p_1 is the transition probability of moving from State 1 (babies, age 0-9) to State 2 (adolescents, age 10-19) during the time period of ten years?

p_2 is the transition probability of moving from State 2 (adolescents, age 10-19) to State 3 (adults, age 20-29) during the time period of ten years?

In general, p_k is the transition probability of moving from state k to State k+1 during the time period of ten year?

State *n*-1 is the state of peoples of an age 70+.

Here "r" is the average growth rate of population of all provinces of Pakistan including Islamabad during 1972-1998 multiplied by the length of age group i.e. 10 years.

The resulting Modified Markov chain is irreducible/regular, aperiodic and hence a stationary distribution (limiting distribution). The population of different age groups is projected as under

$X^{(1981)} = X^{(1972)}. P_m$
$X^{(1991)} = X^{(1981)}. P_m$
$X^{(2001)} = X^{(1991)}. P_m$
and so on

Empirical example: The initial row vector (2.4) i.e. $X^{(1972)}$ is obtained by substituting the population of different age groups of 1972 population census in equation (2.2). However, the last element of this initial vector presets the total population of Pakistan as per population census 1972 excluding the population of FATA (Statistics division, 1972).

$$x^{(1972)} = [19544973 \quad 13168302 \quad 9208929 \quad 7289623 \quad 5390373 \quad 3497042 \quad 3497042 \quad 4362641 \quad 1876447 \quad 62309340\,] \qquad (2.4)$$

The proposed augmented transition probability matrix P_m using 1972 and 1981 population census is obtained as under

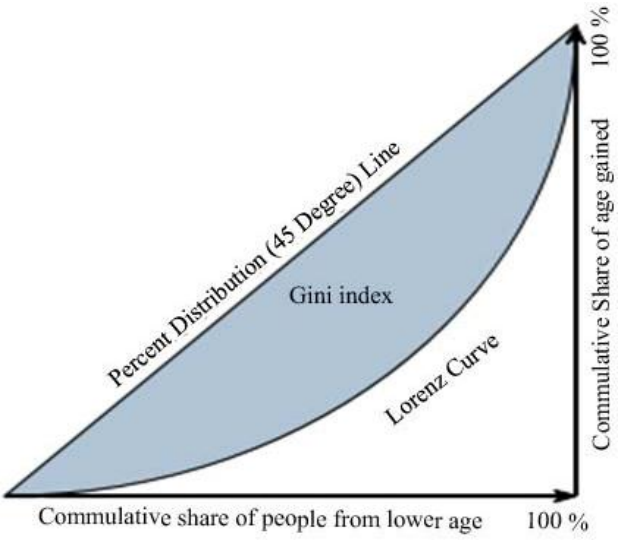
$$P_m = \begin{bmatrix} 0 & 0.9499 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0501 \\ 0 & 0 & 0.889 & 0 & 0 & 0 & 0 & 0 & 0.1110 \\ 0 & 0 & 0 & 0.9572 & 0 & 0 & 0 & 0 & 0.0428 \\ 0 & 0 & 0 & 0 & 0.9522 & 0 & 0 & 0 & 0.0478 \\ 0 & 0 & 0 & 0 & 0 & 0.9160 & 0 & 0 & 0.0840 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.8490 & 0 & 0.1510 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.580 & 0.4200 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6000 \end{bmatrix} \qquad (2.5)$$

To project the population of different age groups of 1981, simply the initial row vector $x^{(1972)}$ has to be multiplied with that of the matrix P_m given in equation 2.5. The resultant vector would be considered as the projected population of different age groups. The overall projected population of the year 1981 is obtained by adding all the elements of resultant vector. Later on, this resultant vector will be considered as the next initial vector for the population projection of 1991. Similarly, to project the population of different age groups of 1991, the resulting vector $x^{(1981)}$ of order 1x9 will multiply with that of the same matrix P_m given in equation 2.5 and so on. So for each next 10 years projection, only the initial row vector has to be updated as per above mentioned procedure before the multiplication with that of proposed augmented one step transition probability matrix P_m given in 2.5.

Disparity Measures

It is unanimously acceptable that population of different age groups have different requirements like food, medicine, industry, housing, marriages etc. The precise projection of the population of different age groups determines the population needs which are inevitable for planning and administration of the country from micro to macro level. It is universally acceptable that without the optimum knowledge of age sex distribution of population of a country; the best and outstanding planning of that state is impossible. Gini coefficients and Lorenz Curves are the most popular tools to discuss the disparity of the data.

The graph of the inequality/disparity (Lorenz curve) is on the left hand side: The less concave in the Lorenz curve away from the line of equality indicates less degree of age disparity. The Gini coefficient presents the area of concentration between the Lorenz curve and the line of perfect equality as it expresses a proportion of the area enclosed by the triangle defined by the line of perfect equality and the line of perfect inequality. The



Gini coefficient (G) ranges from 0 to 1, zero means all measurements are equal, on the other hand, the Gini coefficient (G) means an infinitely large set of observations which is the ultimate inequality. Cowell (1995, 2000) illustrated that different scientists applied these measures into other areas than income and wealth, but mostly within Economics. According to Stuart and Ord (1994), the closer the value of the coefficient is to 1, the more unequal the distribution. Pan American Health Organization (2001) computed the biased and unbiased Gini coefficients and their confidences intervals to measure the inequality of different data sets e.g. GNP per capita, infant mortality rate, live births and infant deaths etc. Rodrigue et al. (2013) measured the traffic inequality at different terminals by computing the index of dissimilarity (ID), Gini coefficient (G) and Gini's means difference (GMD). Brown (1994) used the Gini-style indices to evaluate the spatial patterns of health practitioners and used Alberta data for theoretical considerations as an application. Gini coefficients along with their confidence intervals are calculated and Lorenz Curves are also drawn for male, female and for both sexes separately for the census population as well as projected population of Pakistan. The population of different age segments e.g. babies, adolescents, working, dependent and ageing population was also projected.

3. RESULTS AND DISCUSSION

The population of different age groups was projected by proposed Modified Markov chain for the period of 1981-2021 and is given in Table 1. The overall projected population for the years 1981, 1991, 2001, 2011 and 2021 are 83.59 million, 107.52 million, 135.80 million, 176.18 million and 228.77 million respectively. The projected populations are almost close to the population projections (Total Population by Country, 2012; Country Comparison Population, 2012; International Data Base [IDB], 2010; National Institute of Population Studies [NIPS] 2011; Finance Division, 2013; World population prospects [WPP], 2012). The projected population in different age groups from 0-9 age group to 70 + are approximately close to the actual 1981 population census of Pakistan (Statistics Division, 1984). The proposed Modified Markov Chain Model for population projection is recommended only upto 30 years projections from which the probabilities of survival are derived. The accuracy of projected population depends only on the reliability of probabilities of survival, the correctness of average population growth rates and the initial row vector. The more accurate the data, the more accurate the projections would be. On the other hand, Population Projections more than 30 years are

usually not encouraging and recommended due to the instability of growth rate of population. Klosterman (1990) described the twenty years forecasting as long term and two to three years as short term forecasting.

To examine the dissimilarities, Gini coefficients and confidence intervals of age distribution of projected populations were also calculated for the years 1981, 1991, 2001, 2011, 2021 (see Table 2). The coefficient for the year 2021 is 0.4715 which is slightly less than the other projection years except 1981. On the other hand, the coefficient for the population 1998 is exactly same to that of dissimilarity index of age distribution of projected population 2001. These measures indicated the medium level of concentration among different age groups of projections population. Moreover, these coefficients do not specify the lower and higher side of more concentration among different age groups of population. This is an indication that if the growth rate of 1998 as well as associated social and cultural values continues; the shape of population distribution will remain stable but the population size would surely be increased provided the Total Fertility Rate, life expectancy, sex ratio and international migration do not change drastically.

Table 3 also reveals the Gini coefficients and their confidence intervals of three population census 1972, 1981 and 1998. There is a significance disparity in age distribution of 1998 as compared to the age distribution of population census 1972 and 1981.

The percentages of projected population for some specific years and significant age segments were also estimated and given in Table 4. According to the 1998 population census, approximately 14.80%, 28.60%, 53.10% and 3.50% population were of ages 0-4, 5-14, 15-64 and 64+ respectively. The corresponding estimated population percentages for the year 2023 are 10.12%, 19.86%, 65.01% and 5.00% respectively. According to Population by country (2012), Pakistan was ranked at 46th position in the world with that of about 37.20% population of age 0-14 as compared to approximately 43% in 1998. It is a good indication about the decreasing trend of growth rate of population of Pakistan. The projected population percentage of working age group (15-64) is about 58.60% and ranked the Pakistan 98th position? It implies that there are 97 countries in the world that have more working and healthier population than Pakistan. On the other hand, Pakistan was ranked 94th position according to the aged population percentage 4.20% of age 65+. It indicates the increasing trend in older population. The results indicate that the babies and adolescent population will decrease in future whereas population of age (15-64) will increase approximately 9% and 12% respectively as compared to 1998. This is a good indication that the population of Pakistan will be healthier and vigorous in future. Ryan and Willits (2007) also reported the increase in aging population of United States. In 1900, the percentage of elderly population (75+) was 29% which is expected about 56% in 2040. It seems that in future globally, there would be a problem of ageing population and would become a big hindrance in the development of the country. No doubt, Pakistan is a religious and Islamic state where most of the people feel pleasure to look after their elders. Even though, due to the increased literacy rate and busy schedule in the modern era of life, the ageing population would be affected definitely. In the light of these estimates, the Government of Pakistan should thresh out the expected social problems and needs of ageing population as well as their management well in time. It is indispensable to overcome unavoidable circumstances of ageing population in near future.

Table 5 reveals the Gini coefficients and their confidence intervals of significant age segments of component projected population during 1998-2023. These estimates indicate the larger disparity into significant age segments of the population during 2013-2023 as compared to 1998.

The Lorenz curves of actual population of 1972 and 1981 census for both sexes were drawn in Figures 1 and 2 separately. These Figures indicate the medium level of concentration among 1972 and 1981 population censuses. The Lorenz difference for 1998 population census was also given in Figures 3. It indicates the slightly high level of concentration as compare to 1972 and 1981 population census. The Lorenz curves of projected population for the years 201 and 2021 were also drawn and given in Figures 4, 5 respectively. These plots are slightly different with that of the plots of 1998 population census.

The Lorenz curves of the projected population of significant age segments of the years 1998, 2013, 2018 and 2023 were drawn and given in Figures 6, 7, 8 and 9 respectively. These plots indicated the greater inequality in population age segments during the next 10 years as compared to the population age segments of 1998.

Figure 10 presents the comparison of age distribution for the years 1998 to 2023. Over all the population of Pakistan would definitely be increased during the next ten years with almost similar increase in each age group except the age groups 5-14, 15-64. Figure 11 and Figure 12 present the age sex distribution for the years 2018 and 2023 as compared to the population pyramids of 1998. It indicates the increase in working and ageing population, although the increase in working group is more than ageing group. The decrease in babies and adolescent population seems to be because of increased literacy rate whereas increase in ageing population because of improved medication and better domestic hospitality.

4. CONCLUSION AND RECOMMENDATIONS

The Proposed Modified Markov chain modelling was found to be parsimonious to project the age distribution of population upto 30 years projection only. The overall projected population by proposed augmented transition probability matrix is consistent with that of the population projections of Pakistan (Total population by country, 2012; Country Comparison Population, 2012; IDB, 2010; NIPS, 2011; Finance Division, 2013; WPP, 2012). To examine the age disparity of the population censuses 1972, 1981 and 1998, the Gini coefficients and their confidence intervals indicate the medium level of concentration within the age distribution of the projected population.

The projected population by component growth model indicated the decrease in new born babies (0-4) and adolescent population (5-14) whereas the increase in working and ageing population during the year 2023. These projections are approximately close to the population projections (WPP, 2012).

Gini coefficients and Lorenz curves were also premeditated for projected population of different age segments. The results demonstrate high level of concentration and increase in ageing population. Pakistani, particularly, Muslims as a followers of Islamic ideology, feel proud to look after their fore-fathers. Due to this reason, it may be supposed that ageing population would not become a significant problem. On the other

hand, the aged population may not be feeling secure due to the busy schedule of their offspring in this modern era of life. Anyhow, it seems that ageing population would become an inevitable problem in the future international competition era. The Government should think about well in time on the socio-economic problems of ageing population from all aspects; explicitly, the management and the needs of aged population. Otherwise, it would become difficult to facilitate the elders and handling their problems efficiently. Advanced countries handled this problem by increasing the earning hands, establishment of old people home and medical insurance etc. The Government should also subsidize the science education, in particular the technical education as well as invite the foreign investor for small industry to cater the job requirements of qualified and technical graduates.

REFERENCES

1. Brown, M. (1994). Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on the Alberta data. *Social Science and Medicine*, 38(9), 1243-1256.
2. Cowell, F.A. (2011). *Measuring Inequality*. 3rd ed. Oxford University Press Inc., New York.
3. Country Comparison Population (2012). Country Comparison Population. Retrieved on November 9, 2013, from http://www.indexmundi.com/g/r.aspx?c=pk&v=24.
4. Fujiwara, M. and Caswell, H. (2002). Estimating population projection matrices from multi stage mark-recapture data. *Ecology*, 83(12), 3257-3265.
5. Government of Pakistan (2013). *Pakistan Economic Survey 2012-13*. Economic Adviser's Wing, Finance Division, Government of Pakistan, Islamabad, Pakistan.
6. International Data Base (IDB) (2010). International Data Base (IDB), information gateway, U.S. Census Bureau. Retrieved on November 9, 2013 from http://www.census.gov/idb/ranks.html.
7. Keyfitz, N. (1964). The population projection as a matrix operator. *Demography*, 1(1), 56-73.
8. Klosterman, R.E. (1990). Community analysis and planning techniques. Lanham, MD: Rowman & Lettlefield.
9. Loh, S. and George, M.V. (2007). Projected population size and age structure for Canada and provinces: With and without international migration. *Canadian Studies in Population*, 34(2), 103-127.
10. National Institute of Population Studies (2011). *Population Growth and its Implications.* Government of Pakistan, Islamabad.
11. Nichols, J.D., Sauer., J.R., Pollock, K.H. and Hestbeck, J.B. (1992). Estimating transition probabilities for stage-based population projection matrices using capture-recapture data. *Ecology*, 73(1), 306-312.
12. Pan American Health Organization (2001). Measuring health inequalities: Gini coefficient and concentration index. *Epidemiological Bulletin of PAHO*, 22(1), 3-4.
13. Ross, M. (2007). *Introduction to Probability Models*. Academic press, USA.
14. Rodrigue, J.P., Comtois, C. and Slack, B. (2013). *The Geography of Transport Systems. 3rd ed.*, Routledge. New York.
15. Ryan, A.K. and Willits, F.K. (2007). Family ties, physical health, and psychological wellbeing. *Journal of Aging and Health*, 19(6), 907-920.

16. Statistics Division (1972). Statistical report of Pakistan. Population Census Organization, Islamabad. Government of Pakistan.
17. Statistics Division (1984). *1981 Census report of Pakistan*. Population Census Organization, Government of Pakistan, Islamabad.
18. Stuart, A. and Ord, J.K. (1994). Kendall's advanced theory of statistics. 6th ed. London: Edward Arnold.
19. Statistics Division (2001). *1998 Census report of Pakistan*. Population Census Organization, Government of Pakistan, Islamabad.
20. Coutsoukis, P. (2012). Total Population by Country. Retrieved on October 5, 2013, from http://www.photius.com/rankings/world2050_rank.html.
21. United Nation (2012). *World Population Prospects: The 2012 Revision*, United Nation, Department of Economics & Social Affairs, Population Division, Population Estimates & Projection Section. Retrieved November 2, 2013, from http://esa.un.org/wpp/unpp/p2k0data.asp.

Table 1:
Projected Population by Proposed Modified Markov Chain during 1981-2021

Age Group	1981	1991	2001	2011	2021
00-09	24,923,736	32,555,598	41,842,713	54,320,283	70,472,116
10-19	18,565,770	23,675,057	30,924,562	39,746,393	51,598,836
20-29	11,706,620	16,504,969	21,047,126	27,491,936	35,334,543
30-39	8,814,787	11,205,577	15,798,557	20,146,309	26,315,281
40-49	6,941,179	8,393,440	10,669,951	15,043,386	19,183,315
50-59	4,937,582	6,358,120	7,688,391	9,773,675	13,779,741
60-69	2,968,989	4,192,007	5,398,044	6,527,444	8,297,850
70+	2,530,332	1,722,013	2,431,364	3,130,865	3,785,918
Fata	2,199,000	2,909,685	38,35,103.21	46,88,737.54	56,04,459.08
Total	83,587,994	107,516,466	135,800,706	176,180,289	228,767,600

Table 2:
Estimates and 95% Confidence Interval of Gini Coefficient of Projected Population during 1981-2021

Projection Year		Coefficient	95% Confidence Interval
1981	Both sex	0.4509	(0.3770, 0.5475)
1991	Both sex	0.4732	(0.4064, 0.6170)
2001	Both sex	0.4720	(0.4011, 0.6293)
2011	Both sex	0.4725	(0.4055, 0.6166)
2021	Both sex	0.4715	(0.3931, 0.6195)

Table 3:
Estimates and 95% Confidence Interval of Gini Coefficient of Population Census during 1972-1998

Population Census	Sex	Coefficient	95% Confidence Interval
1972	Both sex	0.4504	(0.3895, 0.5512)
	Male	0.4337	(0.3711, 0.5553)
	Female	0.4695	(0.4063, 0.5728)
1981	Both sex	0.4552	(0.4137, 0.5408)
	Male	0.4400	(0.3891, 0.5319)
	Female	0.4720	(0.4071, 0.5873)
1998	Both sex	0.4720	(0.3957, 0.5927)
	Male	0.4667	(0.3971, 0.5744)
	Female	0.4777	(0.3954, 0.5834)

Table 4:
Percentage of Component Projected Population of Significant Age Segments during 1998-2023

Projection Year	0-4	5-14	15-64	65+
1998	14.80	28.60	53.10	3.50
2003	12.86	27.17	56.17	3.81
2008	12.29	23.41	60.50	3.80
2013	11.67	21.56	62.50	4.24
2018	10.93	20.79	63.76	4.52
2023	10.12	19.86	65.01	5.00

Table 5:
Estimates and 95% Confidence Intervals of Gini Coefficients during 1998-2023

Population Census	Coefficient	95% Confidence Interval
1998	0.5418	(0.4068, 0.7799)
2013	0.6158	(0.5337, 0.7745)
2018	0.6253	(0.5472, 0.7662)
2023	0.6326	(0.5531, 0.7500)

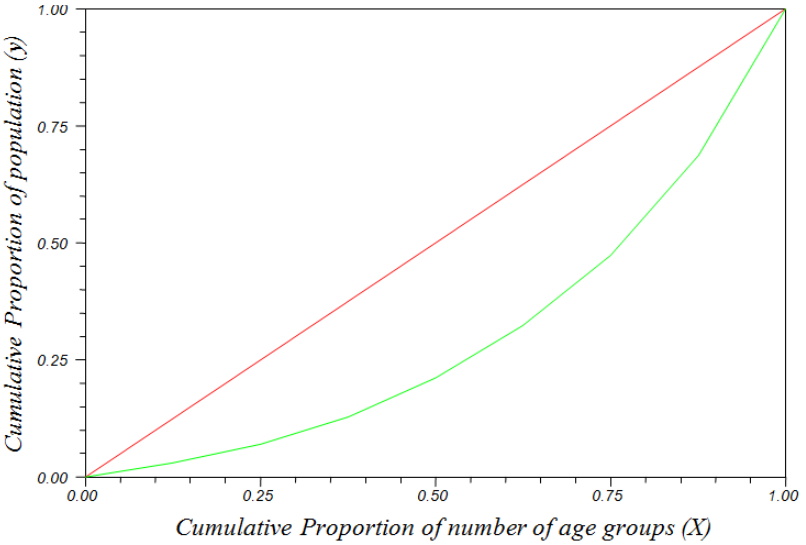


Figure 1: Lorenz Curve of Population Census-1972 (Both Sex)

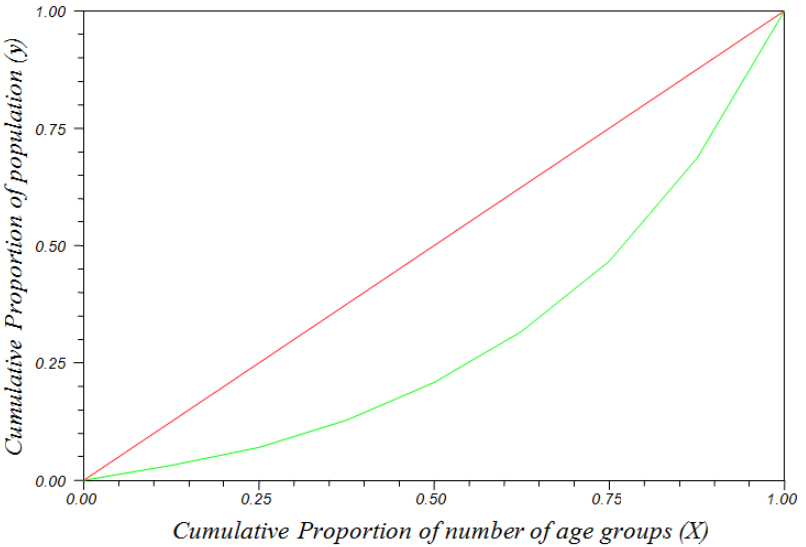


Figure 2: Lorenz Curve of Population Census-1981 (Both Sex)

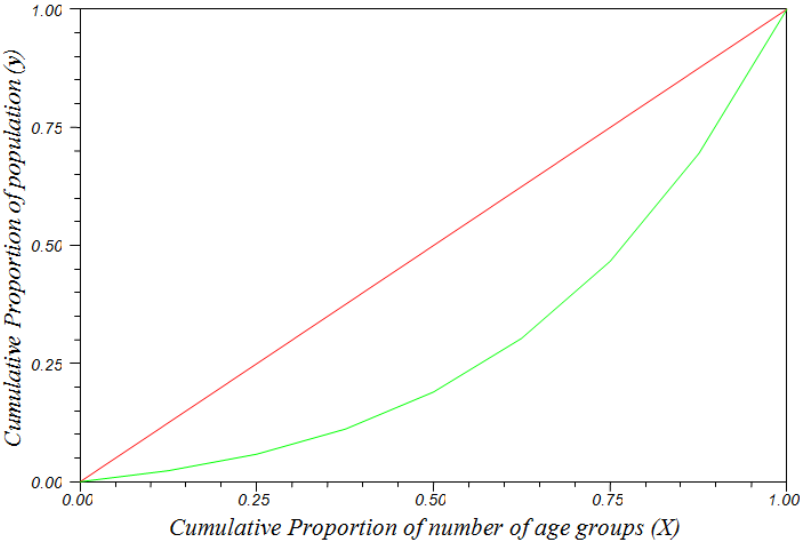


Figure 3: Lorenz Curve of Population Census 1998 (Both Sex)

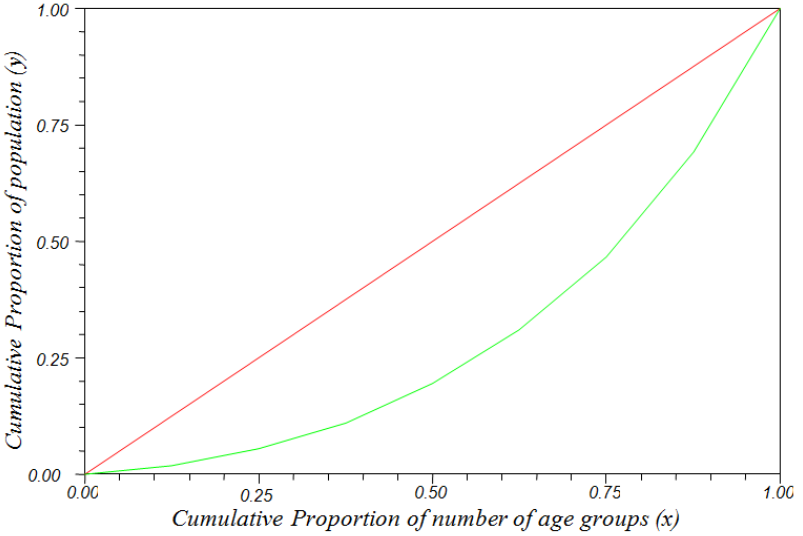


Figure 4: Lorenz Curve of Projected Population 2011 (Both Sex)

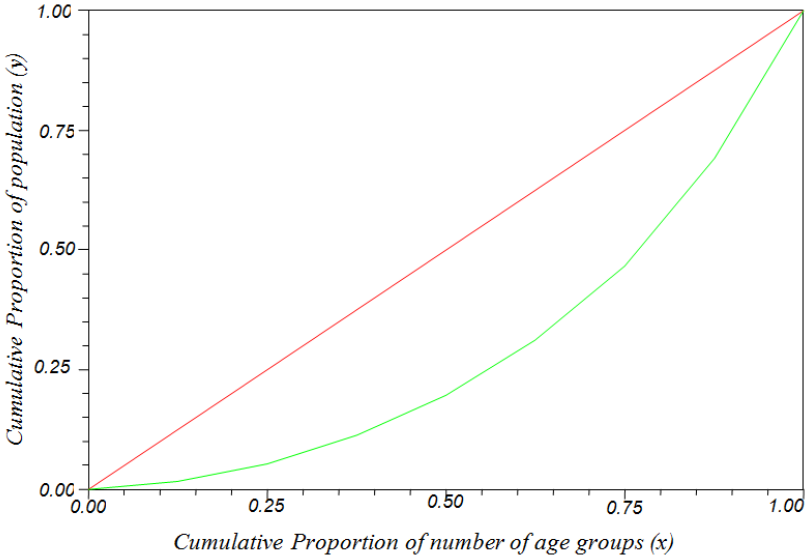


Figure 5: Lorenz Curve of projected Population 2021 (Both Sex)

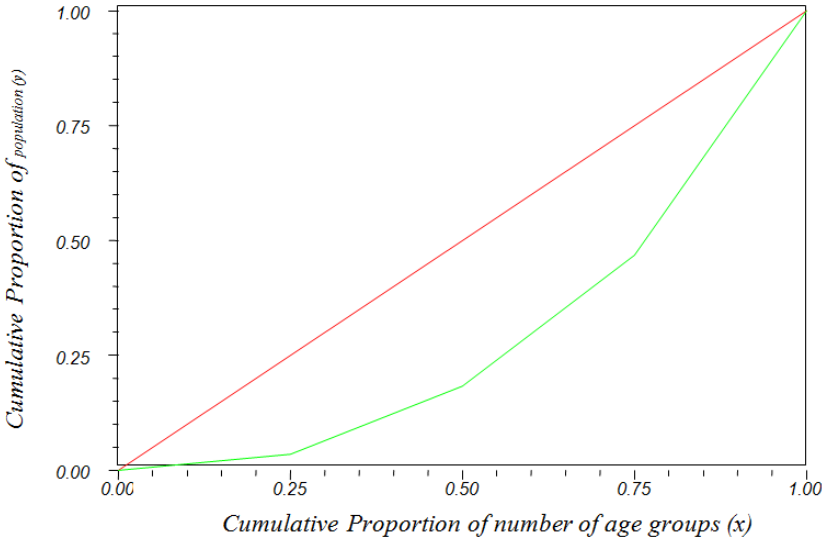


Figure 6: Lorenz Curve of Component Projected Population 1998 (Both Sex)

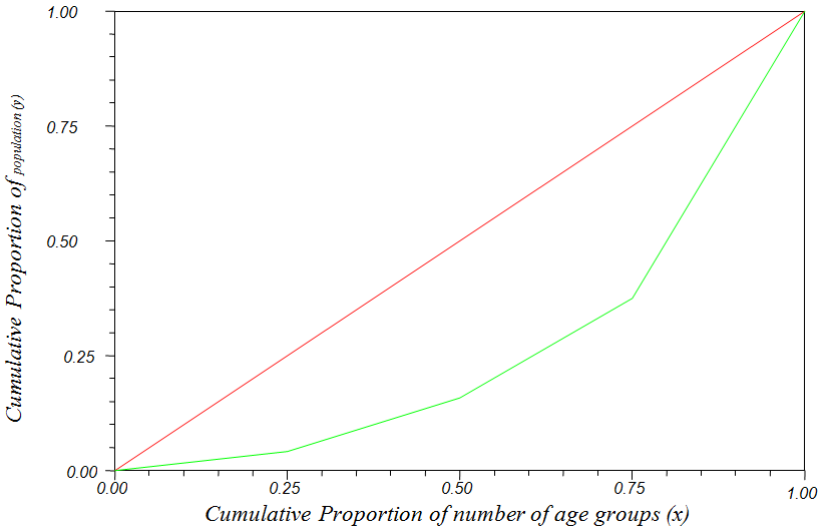


Figure 7: Lorenz Curve of Component Projected Population 2013 (Both Sex)

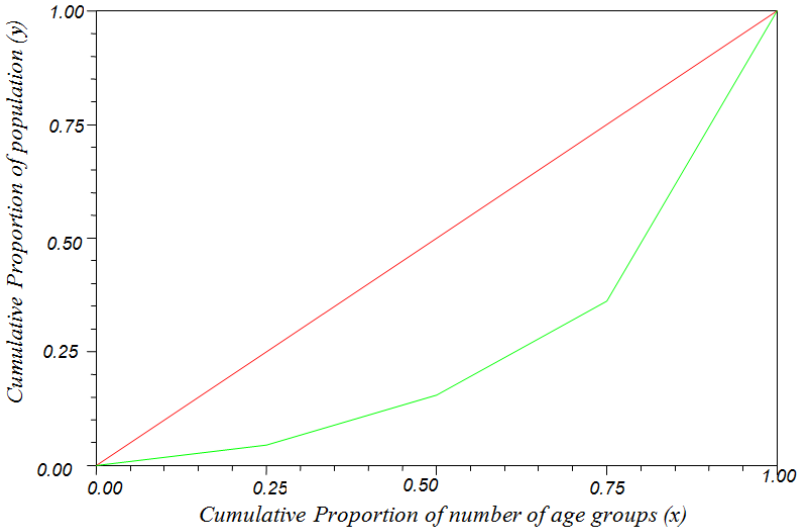


Figure 8: Lorenz Curve of Component Projected Population 2018 (Both Sex)

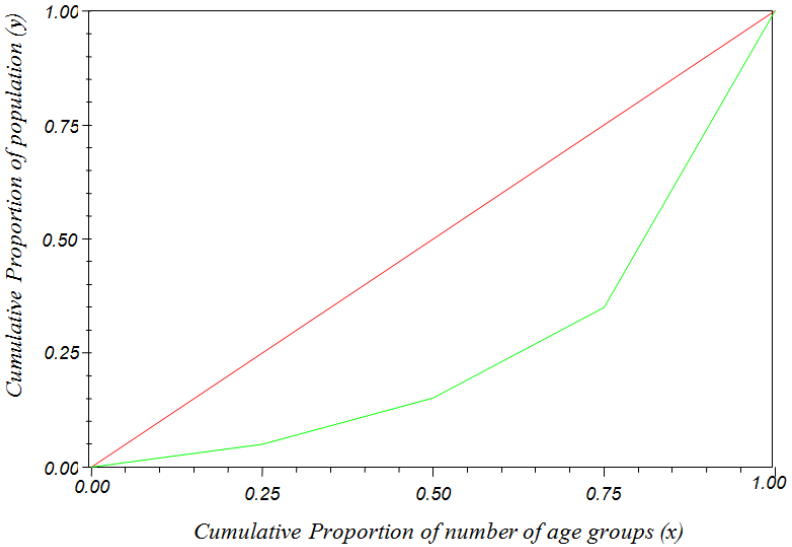


Figure 9: Lorenz Curve of Component Projected Population 2023 (Both Sex)

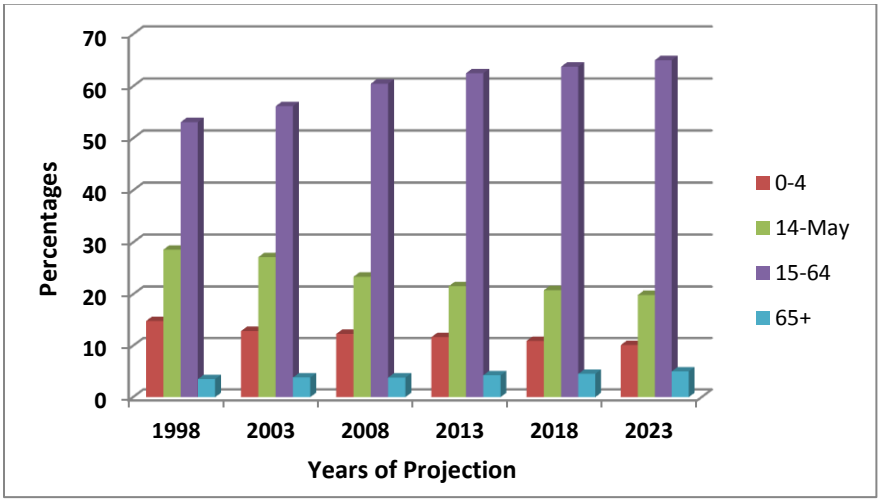


Figure 10: Comparison of Age Distribution Percentages during 1998-2023

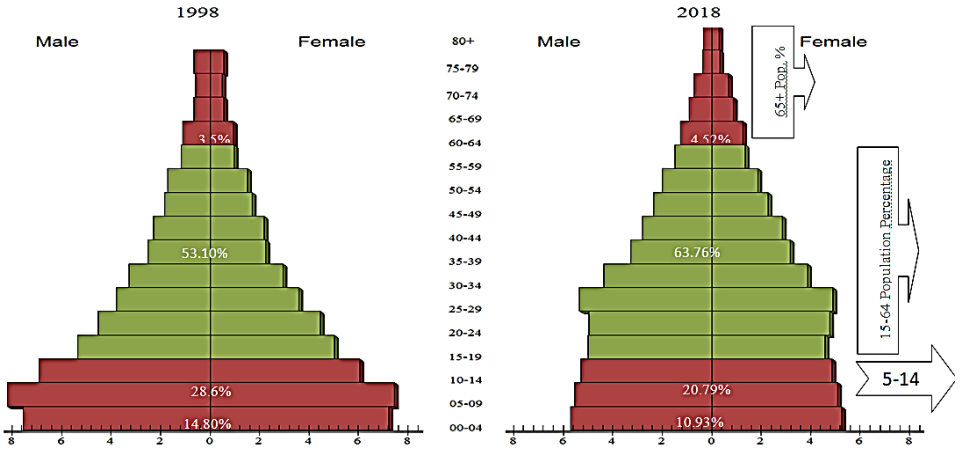


Figure 11: Comparison of Population Pyramids of 1998 & 2018

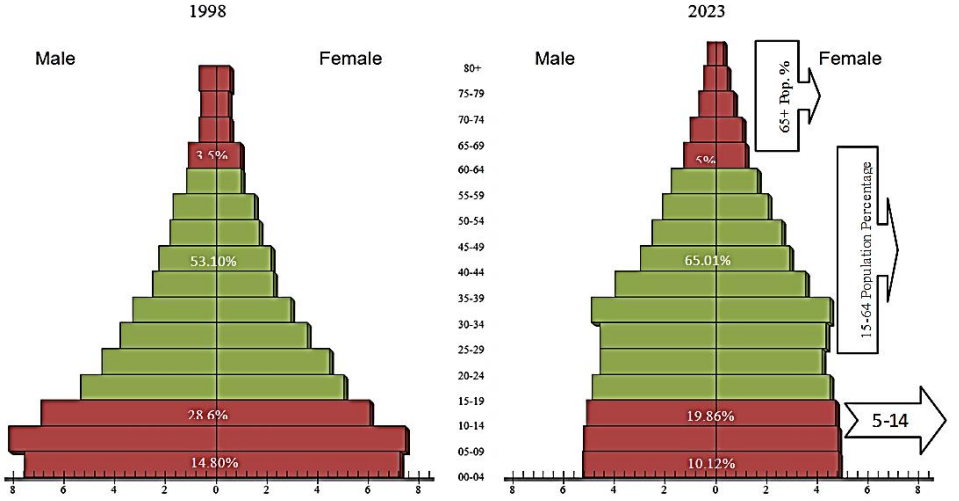


Figure 12: Comparison of Population Pyramids of 1998 & 2023

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 381-390

FRAMEWORK TO COMPREHEND THE IMPACT OF INSTITUTIONS QUALITY AND INDUSTRIAL COMPETITIVENESS ON ECONOMIC GROWTH OF PAKISTAN IN SAARC REGION: AN EMPIRICAL INVESTIGATION

Yasir Ali Soomro
Iqra University, Karachi, Pakistan
Email: yasir.soomro@hotmail.com

ABSTRACT

Pakistan's economy continues to remain exposed to the impulses of international developments as well as internal security environment. The intensity of the global financial crisis has further added to Pakistan's difficulty. Despite support from the IMF and other bilateral and multilateral donors, Pakistan's external account remains exposed to a number of uncertainties. Pakistan's problems are by now well-known and they are multi-sectorial. They affect not only stability of the Pakistani state itself, but also its immediate neighbours as well as the West. The relationship between economic performance and the quality of domestic institutions has emerged recently as a major subject of interest in Social Sciences. The literature shows that the higher the quality of domestic institutions the better the effects on the Human development and growth of a country but empirical evidence has not been given yet so far for single country specifically Pakistan. In this study authors has aggregate variables to judge Pakistan's institutional quality along with recently developed index of institutionalized social technologies and its sub-indices namely Risk reducing technologies in a proper theoretical framework with empirical results.

KEYWORDS

Institutional Quality; Macroeconomic Factors; GDP; Stock Exchange and SAARC.

INTRODUCTION

Background

The term stock market is abstract for the mechanism that facilitates the exchange of corporate shares. It can be explained by other terms as the sum of all stocks, specially within a an economy, However Stock market is not as same as the stock exchange, The stock exchange is defined as a body that could be a corporation or mutual organization which facilitates the buyers and sellers of stock. For example, the stock market in the Pakistan includes the trading of shares listed on the Karachi Stock Market in Pakistan, Bombay Stock exchange in India, Dhaka Stock Exchange in Bangladesh, Colombo Stock Exchange in Sri Lanka and Nepal Stock Exchange in Nepal.

Stock Market Participants and Trading

In the early history only the wealthy people were the investors in multiple businesses but now the markets have been institutionalized officially with sellers and buyers into institutions e.g. investment firms, hedge funds companies, banking industry, mutual

funds companies, insurance companies and pension funds firms. Due to the institutionalized investing, stock markets have become more interesting in its operations

Stock Markets in SAARC countries

Country	Stock Markets
Pakistan	Islamabad Stock Exchange
	Karachi Stock Exchange
	Lahore Stock Exchange
Bangladesh	Chittagong Stock Exchange(CSE)
	Dhaka Stock Exchange (DSE)
India	Bombay Stock Exchange
	Calcutta Stock Exchange
	National Stock Exchange
	Cochin Stock Exchange
	Pune Stock Exchange
Nepal	Nepal Stock Exchange
Sri Lanka	Colombo Stock Exchange
Maldives	Maldives Stock Exchange
Bhutan	Royal Securities Exchange of Bhutan

Problem Statement

In the current literatures there have been a lot of researches on the macroeconomic factors affecting stock market development like the relationship of economic growth and stock market development (Levine and Zervos, 1996), other researchers have found out rate of saving, gross domestic income, domestic credit, value traded, interest rate, inflation, remittances, consumer deposits, money supply, trade deficit, industrial production and many other macroeconomic variables as key determinants of stock market development. It has been also found that there is a unidirectional cause and effect in stock prices from macroeconomic factors (Hussain and Mehmood, 2001). However the importance of institutional quality has also been considered as one of the important factor contributing to stock market development especially in developing countries (Billmeier and Massa, 2007; Yartey, 2008; Burhop et al, 2011). But there has not been much interest directed towards institutional quality impacting the stock market development in the SAARC region economies. So this research tries to add empirical evidence in the context of importance of institutional quality impacting stock market development in the SAARC region. This could help in understanding better the factors affecting the stock market development and can contribute significantly in making better decisions and devising policies more effectively.

Research Question

What is the impact of institutional quality and the macroeconomic variables on the development of stock market in the SAARC region countries specifically Pakistan's Economy?

LITERATURE REVIEW

In current years, equity markets have shown a crucial role in the economic growth and financial development and there have been large fluctuations in stock markets as well. The measure of a stock market capitalisation to determine stock market is considered to be a critical measurement in the domain of financial development. Likewise domestic lending to private sector or the monetization is considered critical measure for the development in financial intermediation. It is a well-known fact that stock markets and commercial banking have major contribution in the transformation of savings into investment, and also boost economic growth and financial development of the economy. However, it is argued that these two kinds that are commercial banking and stock market should be analysed as substitutes or complements.

In recent years, there has been a rise in the world stock markets, and the major role has been played by the emerging markets in this boom. The very important issue in domestic financial liberalization programs has been Stock market development in most of the developing countries of the world. It is regarded as any program of financial liberalization is not complete without the establishing and developing of stock markets. From a geographical perspective in the South Asian countries, In the SAARC countries relatively to other countries there has been a little interest directed towards the research of stock market development of the SAARC region. SAARC region is considered to be comprising of the fastest emerging economies in the world. SAARC countries include Nepal, Sri Lanka country of Tamil Tigers, Islamic Republic of Pakistan, India, Bangladesh, Afghanistan, Bhutan and Maldives.

This research focuses on institutional and macroeconomic factors that contribute in stock market development in SAARC region countries with dataset of 5 countries using the period from 1996 to 2010. Rahman and Rahaman (2011) concluded that usually the development of banking domain is strongly related with the equity market variables. Further it is investigated that how significantly Institutional quality has impacted on stock market capitalisation in SAARC region in the period of 15 years. It is generally considered that strong rights of property protection and minimisation of political risk encourage individual's confidence thus enhances stock market activities. Thus, stock market investment is encouraged with the development of good quality institutions which leads to stock market development ultimately.. Raza, Iqbal, Ahmed, Ahmed and Ahmed (2010) also concluded that companies in economies with better effectiveness of legal rights are generally able to get financial help from external sources easily. In this research the institutional quality as general measure of economic freedom has been examined in relation to stock market development in SAARC region countries.

RESEARCH METHODOLOGY

Research Approach

This research is quantitative because the hypothesis were tested by analyzing the sample data of different variables and impact of each variable on the dependent variable will be identified i.e. in positive way or in negative way and significant or insignificant.

Research Purpose

The basic purpose of this research is explanatory. The theory has been tested on the sample of data. Therefore it has been investigated here systematically the impact of certain variables that contribute on the dependent variable.

Research Design

The research has been quantitative and correlation in nature. The relationship between dependent variable and multiple independent variables will be tested. Based on the findings the hypothesis will be accepted or rejected and final implications will be suggested on the bases of research results.

Data Source

Data for institutional quality has been taken from Heritage Foundation's 2012 Index of Economic Freedom and data for macroeconomic variables has been taken from World Bank's online data publication.

Target Population

The research focuses on the regions of SAARC economies in particular countries Pakistan, Bangladesh, India, Nepal and Srilanka.

Sample Size

The dataset of 5 economies has been considered. The year 1995 to 2010 that is for 15 years of 5 countries of the SAARC region making it a total observations of 75 has been taken as the sample size.

Statistical Technique

The sample data of dependent variable and independent variables will be analyzed by ordinary least square method of multiple regression analyses technique.

Research Model

$$MCP = \alpha + \beta_1 IQ + \beta_2 INC + \beta_3 INV + \beta_4 SVT + \beta_5 DC + e$$

where

MCP: Market Capitalisation: This can be defined as aggregate current market value of the shares outstanding of the listed companies that is the sum of prices of shares times no of outstanding shares. In this research this proxy has been used as a market capitalisation of locally listed companies as a percentage of GDP.

IQ: Institutional Quality: Institutions can be defined as the formal governing structure that forms regulations to shape the political, economic and social affairs of the society. Better efficient and effective institutions represent better institutional quality. It has been measured by Heritage Foundation's 2012 index of economic freedom by scoring 0 to 100 of different components of the economy representing quality of institutions. The different components include freedom from corruption, property rights, government spending, fiscal freedom, trade freedom, financial freedom, investment freedom, business freedom, labour freedom and monetary freedom[1]. Higher the score the higher will be the institutional quality.

[1] 2012 Index of Economic Freedom published by Heritage Foundation and The Wall Street Journal

INC: Income: This has been measured as the Gross Domestic Production at constant 2000 US $ in Billions.

INV: Investment: This can be defined as the total value of additions to the fixed assets plus any net inventories of that particular economy. In this research it has been measured in the form of capital formation as a percentage of GDP.

SVT: Stock value traded: This is defined as the value of the all shares traded in any particular period. This has been measured as the total percentage of share value traded over GDP.

DC: Domestic credit: This is defined as the domestic lending to the private sector and has been measured as the domestic lending to the private sector as a percentage of GDP.

Research Hypotheses

H_1: Institutional quality has a significant impact on stock market capitalization.

H_2: GDP has a significant impact on stock market capitalization.

H_3: Domestic credit has a significant impact on stock market capitalization.

H_4: Stock traded value has a significant impact on stock market capitalization.

H_5: Gross capital formation has a significant impact on stock market capitalization.

DATA ANALYSIS

Data run was performed to find about the relationship between the dependent variable that is stock market capitalization and independent variables that are institutional quality, remittances, investment, domestic credit and stock market liquidity in terms of their averages, deviations, overall model fitness and the results of OLS method. On the basis of which it will be decided how the independent variables (Institutional Quality, Gross Income, Gross Capital Formation, Domestic Credit and Stocks Traded) are affecting the stock market development in the SAARC region.

Descriptive Statistics

Table 4.1
Reports the Descriptive Statistics for each Variable

Descriptive Statistics			
	Mean	**Std. Deviation**	**N**
Market capitalisation	23.933	24.801	75
Institutional Quality	54.119	4.772	75
Gross Capital Formation	24.279	4.864	75
Gross Domestic Income	153.655	241.013	75
Domestic Credit	31.061	8.358	75
Stock Value Traded	22.626	31.598	75

This table depicts the descriptive statistics of all variables analysed in this research. The 15 years of data of each variable from 5 different countries of SAARC region has been considered. It was found that the average Market Capitalization of the SAARC

region is 23.933 percent as a percentage of respective GDPs and there is a standard deviation of 24.801, this shows that there is high variation in the stock market capitalization of the SAARC region this could be due to the varying size of the economies considered in this research, specially India because it has relatively larger market size than rest of the SAARC economies. While mean for the Institutional Quality in the SAARC region is 54.119 out of the score of 100 as measured by the Heritage Foundation and there is a standard deviation of 4.772. This shows that institutional quality is almost similar across the SAARC region economies because it's not fluctuating too much. However it can be improved significantly if the proper measures are taken to improve the score from the average of 54.119 to higher score. The independent variable gross capital formation has the mean value of 24.279 percent respective GDPs and the standard deviation of 4.864 percent; this shows that an average gross capital formation is representing 24 percent of gross domestic income of SAARC region, which is quite good for the stock market development. Domestic credit has the mean value of 31.061 and the standard deviation of 8.358. Stock traded has the mean value of 22.626 and the standard deviation of 31.598. While an average GDP of the SAARC region representing five countries in this research, is found to be 153.655 US billion dollars with the standard deviation of 241.013 US billion dollars. The standard deviation of the GDP of SAARC region is very high because in this research sample contains different size economies. The countries that were considered in the SAARC region are Pakistan, India, Bangladesh, Srilanka and Nepal.

Pooled Ordinary Least Square (OLS) Estimations

This part explains the impact of different independent variables on the stock market capitalization in the SAARC region with estimation and direction.

Model Summary

Model		R	R Square	Adjusted R Square	Std. Error of the Estimate
Dimension	1	.882[a]	.777	.761	12.12253

a. Predictors: (Constant), GDP in billions, INSTQ, DCRDT, STOCKTR, GCAP

ANOVA b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	35375.645	5	7075.129	48.145	.000[a]
	Residual	10139.948	69	146.956		
	Total	45515.593	74			

a. Predictors: (Constant), GDP in billions, INSTQ, DCRDT, STOCKTR, GCAP

b. Dependent Variable: MCAP

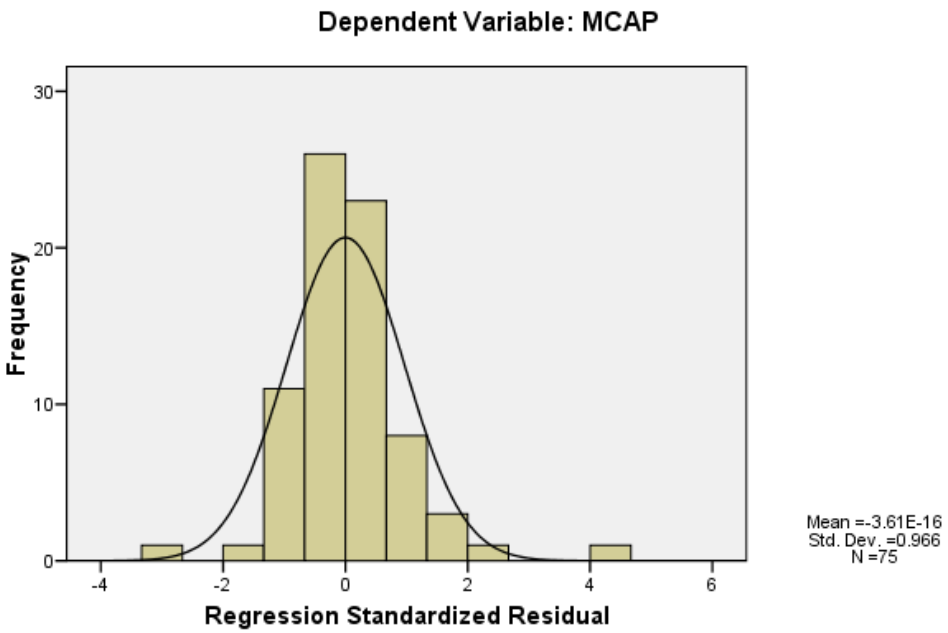
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-74.380	17.929		-4.149	.000
	GCAP	1.298	.474	.255	2.736	.008
	DCRDT	.713	.223	.240	3.198	.002
	STOCKTR	.266	.065	.339	4.107	.000
	INSTQ	.615	.311	.118	1.976	.052
	GDP in billions	.035	.011	.338	3.261	.002

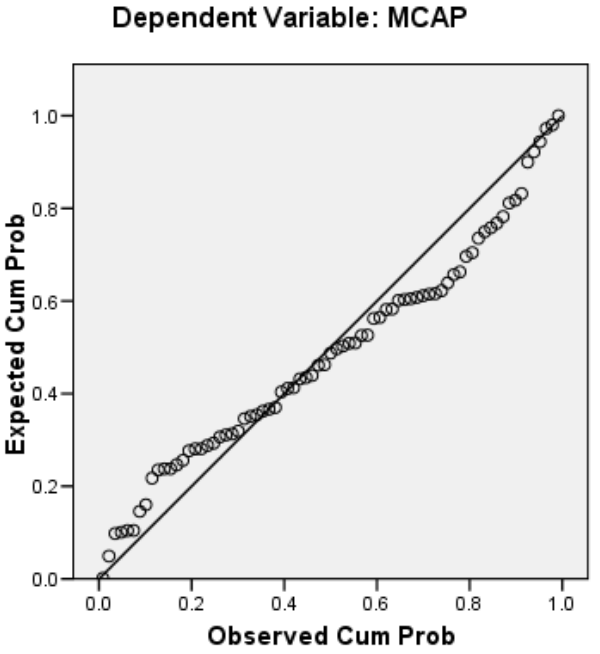
a. Dependent Variable: MCAP

Variables	Coefficient	t-stats	Prob.	VIF
Constant	-74.380	-4.149	0.000	
Institutional Quality	0.615	1.976	0.052	1.112
Gross Capital Formation	1.298	2.736	0.008	2.681
Domestic Credit	0.713	3.198	0.002	1.748
Stock Value Traded	0.266	4.107	0.000	2.110
GDP	0.035	3.261	0.002	3.326
Adj. R2	0.761			
D.W stats	1.784			
F-stats (prob)	48.145 (0.000)			

HISTOGRAM



Normal P-P Plot of Regression Standardized Residual



The results shows that stock market capitalization have significant impact of institutional quality in a positive direction and it is third most important factor affecting stock market development in the model. This means that there will be better stock market development if the institutions in that country are working effectively and efficiently in facilitating the business in the country and if the institutions are restricting the business through different laws and regulations, taxation policies, license permits and trade restrictions, resulting in decrease in the quality of institutions will have negative effects on the stock market development in the country.

Moreover the table depicts that all the other macroeconomic variables considered in this research like capital formation, domestic credit, stock value traded and gross domestic product have been found, having significant impact in a positive direction on the stock market development in the SAARC region.

Capital formation have a significant and positive impact and have been found most important factor contributing in stock market development in SAARC region. This measure explains the countries expenditure in the formation of new capital in the economy, if people in the country are carrying out their businesses and want to expand it then more capital will be required which can be raised through stocks which ultimately develops the stock market. Domestic Credit have significant positive impact because the money lending to the public will make people to invest more in the stocks, which will contribute significantly in the stock market development. Stock Value traded represents

the liquidity of the stock market, hence more liquid the stock market, the more will be the investments in the stocks because investors will be able to turn their stocks into the liquid for other consumptions, and this also contributes significantly in the stock market development.

R Sq is the value of variation in the dependent variable that is Market Capitalization as explained by the independent variables (Institutional Quality, Gross Income, Gross Capital Formation, Domestic Credit, Stocks Traded and Inflation). This value explains the association strength between dependent and independent variables. On which basis model fitness is suggested. The value of adjusted R square 0.761 means that the 76.1 percent variation in the Stock Market Capitalization can be explained by the independent variables (Institutional Quality, Gross Income, Gross Capital Formation, Domestic Credit and Stocks Traded). The f value found is 48.145 and is highly significant because the p value is less than 0.005; therefore the model fitness is enough to explain the market capitalization. Durbin Watson value is 1.784 which approximates 2 therefore it shows that there is no autocorrelation among the variables. This has been also confirmed by LM test with lag 1 which show the p value of 0.504451 means alternative hypothesis that autocorrelation exist is rejected.

CONCLUSION

The data were collected from the online publications that are for Macroeconomic data the World Bank economic indicators were used and for the Institutional Quality the average score in 2012 Index of Economic Freedom by The Heritage Foundation. After applying the statistical technique it was found that Institutional Quality like other macroeconomic factors has a significant role to play for the development of the stock market. However poor economic structure has caused inefficient institutions and bureaucratic delays. Law and order situation has been very volatile causing delays in institutional performance. There is not much independence of judiciary and the judicial procedures are slow, costly and are quite influential by political pressure. Intellectual Property rights like Copyrights, patents are not well protected. Corruption level is on rise and poorly administration of taxation has seriously affected the Institutional Quality. Entrepreneurs are not properly facilitated to start a new business this is due to the heavy costs of licensing authorities. The majority of the workforce is unemployed and inflation rate is raising causing devaluation of money. The investments especially investments from overseas in the form of Foreign Direct Investment is affected because of the political instability, sectarian riots and heavy bureaucracy.

Recommendations

The institutional quality and macroeconomic variables can be enhanced by the following actions

- The system of judiciary should be strong and free from powerful influence especially political influence. There should be proper system for the protection of property rights through strong laws. Corruption in the public segment can be controlled through strong laws.

- The proper taxation system which does not overburden the tax payers. Government expenditure should be utilised in the most effective manner in order to facilitate both consumers and businesses.
- Regulatory system for initialising business should be developed to facilitate the business. The requirements for the licensing should be eased for the individuals to start a business.
- Trade tariffs and barriers should be imposed to provide maximum benefit to the local producers. Proper infrastructure and the minimization of bureaucracy can enhance the growth in private investments. The proper credit facilities should be available for the borrowers to use it in most effective manner and unnecessary borrowings should be controlled.

Future Recommendations

How institutional quality can be improved in respective countries of SAARC region and to find out the ways to economic reformation.

REFERENCES

1. Billmeier, A. and Massa, I. (2009). What drives stock market development in emerging markets – institutions, remittances, or natural resources? *Emerging Markets Review*, 10(1), 23-35.
2. Burhop, C., Chambers, D. and Cheffins, B. (2011). Is Regulation Essential to Stock Market Development? Going Public in London and Berlin, 1900-1913. *Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str.* 10, D-53113 Bonn.
3. Husain, F. and Mahmood, T. (2001). The Stock Market and the Economy in Pakistan. *The Pakistan Development Review*, 40(2), 107-114.
4. Levine, R. and Zervos, S. (1996). Stock Market Development and Long Run Growth. *The World Bank Economic Review*, 10(2).
5. Rahman, A and Rahaman, M.M. (2011). Performance of Stock Market of Bangladesh: An Empirical Study on Chittagong Stock Exchange (CSE). *Asia Pacific Journal of Research in Business Management*, 2(1).
6. Raza, A., Iqbal N., Ahmed, Z., Ahmed, M. and Ahmed, T. (2010). The Role of FDI on Stock Market Development: The Case of Pakistan. *Journal of Economics and Behavioural Studies*, 4(1), 26-33.
7. Yartey, C.A. and Adjasi, C.K. (2007). Stock market development in Sub-Saharan Africa: Critical issues and challenges. *IMF Working Paper* 07/209.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 391-396

SPATIAL ANALYSIS OF HEPATITIS C EPIDEMIOLOGY IN FAISALABAD DISTRICT

Samaira Batool and **Muhammad Yaseen**[§]
Department of Math. and Stat., University of Agriculture, Faisalabad, Pakistan
Corresponding Author: [§]myaseen208@gmail.com

ABSTRACT

Hepatitis C, caused by the Hepatitis C virus (HCV), is an infectious disease affecting mostly the liver. The infection is often without symptoms. However, long lasting infection can lead to scarring of the liver. HCV is spreading mostly by blood-to-blood contact, drug use, poorly sterilized medical equipment and transfusions. The disease is becoming a major health problem for developing countries including Pakistan. World Health Organization rated Pakistan as second country in the world having high rates of chronic liver infections. About 8.6 million Pakistanis are affected with Hepatitis C. Spatial epidemiology is a discipline to understand the relationship between disease risk and exposures of interest. In this study the incidence data on Hepatitis C collected from the Allied Hospital Faisalabad over the period 2010–2012 were analyzed and the spatial dependency of the Hepatitis C risk was explored using spatial statistics besides risks were used to create surface maps of Faisalabad over the period 2010–2012 to identify areas at high risk in the Faisalabad city. It has been found that more than half of the rural population of Faisalabad city is suffering from Hepatitis C.

1. INTRODUCTION

Hepatitis C is an infectious disease affecting mostly the liver, caused by the Hepatitis C virus (HCV). An estimated 150-200 million people worldwide are infected with Hepatitis C (Hashmi et al., 2010). Only limited information is available for the epidemiology of Hepatitis C. Hepatitis C is generally without symptoms but long lasting infection can cause serious liver disease which leads to liver cancer or failure of the liver.

The symptoms of Hepatitis C are Jaundice, Abdominal pain, Loss of appetite, Nausea and Fatigue. Hepatitis C is transmitted by blood to blood contents and also by poorly sterilized medical equipment, and by the persons who are addicted of drugs.

The disease is becoming a major health problem of developing countries, including Pakistan that has the second highest prevalence rate of hepatitis C ranging from 4.5% to 8% (Khattak et al., 2002). World Health Organization (WHO) has rated Pakistan as second country in the world having high rates of chronic liver infections. About 10 million population of Pakistan is infected with Hepatitis C virus (HCV) (Umar and Bilal, 2012). Hepatitis C is the major problem of the developing countries where the medical facilities are not as much as required.

Studies in Pakistan on small targeted groups including blood donors, health professionals, drug abusers and chronic liver disease patients indicate that the prevalence of Hepatitis C is as high as 40 percent (Abbas et al., 2010).

Several studies indicate that the rate of Hepatitis C is much more in rural areas rather than the urban areas in Pakistan. It is worrisome to note that 66% population of Pakistan is living in the rural areas (Shaikh and Hatcher, 2005) where the general health facilities are not properly available. The public carries the burden of disease or they are at high risk due to untrained persons, unscreened blood transfusions, lack of education, poverty, and lack of medical facilities (Jiwani and Gul, 2011).

Epidemiology is the study of distribution, cause and control of the disease in human population. Spatial statistics has been widely used in epidemiology for the study of the distribution of diseases. Spatial Epidemiology is the study of spatial (geographical) prevalence of disease incidence and its relationship to the potential risk factors. The knowledge of spatial and temporal disease is essential for the epidemiologists to understand the population interaction with the environment.

The majority of epidemiological studies are observational in nature. The observational studies are Cohort studies, Case-control studies, Cross-sectional studies, Ecologic studies and Semi-ecological studies (Elliot et al., 2000). Spatial epidemiology at small-area scale can be divided into three main areas such as disease mapping, geographic correlation studies, and clustering, disease clusters, and surveillance (Elliott and Wartenberg, 2004). In this study we modeled the incidence data on Hepatitis C collected from the Allied Hospital Faisalabad over the period 2010–2012 to exploit any spatial dependency of the Hepatitis C in the Faisalabad city.

2. MATERIAL AND METHODS

The focus of this study is the Faisalabad city which constitute of eight towns. Faisalabad is the third largest city of Pakistan which stands in the rolling flat plains of northeast Punjab between the longitude $73^{\circ}74$ East, latitude $30^{\circ}31.5$ North, with an elevation of 184 meters above the sea level.

In this study the incidence data on Hepatitis C collected from the Allied Hospital Faisalabad over the period 2010-2012 were analyzed and the spatial dependency of the Hepatitis C risk was explored using spatial statistics besides risks were used to create surface maps of Faisalabad over the period 2010-2012 to identify areas at high risk in Faisalabad city. The form of the data is observed and expected cases which allow the calculation of standardized morbidity ratio, the ratio of observed to expected cases. The projection of longitude and latitude of each town is also taken with their adjacent areas.

Disease mapping provides a simple description of visual summary of geographical risk and also provide the estimates of the risk by area to inform the public health resources. Disease mapping is generally carried out by using the count data. Disease maps have a long history. A survey in 1991 identified 49 international, national, and regional disease atlases (Walter and Birnie, 1991). Disease maps typically show standardized mortality or morbidity (e.g., incidence) ratios (SMR_i) for geographic areas such as countries, counties, or cities. The rate in area i is estimated by the standardized mortality (or morbidity) ratio (SMR_i), calculated as O_i/E_i, where O_i is the observed number of deaths or incident cases of disease in the area (assumed to follow an independent Poisson distribution) and E_i is the expected counts of incidents (Elliott and

Wartenberg, 2004). Considering that O_i's are independent and drawn from the Poisson distribution whose mean is $\theta_i E_i$ where θ_i is the relative risk. Then the maximum likelihood estimator of θ_i is (SMR) and represented by $\widehat{\theta_i}$ with the variance $var(\theta_i) = {\theta_i}/{E_i}$ which is estimated by $\widehat{var(\theta_i)} = {\theta_i}/{E_i}$. It is assumed that there is no interaction between the risk and the population strata, i.e., the relative risk θ_i depends only on the region. Assumption about the equivalence of mean and variance for the Poisson distribution is very stringent and does not hold always as mostly the counts are over-dispersed. To model the counts where the variance is higher than the mean the negative binomial distribution is used instead of Poisson distribution. The negative binomial distribution can be regarded as the mixed model with the random effect follow the Gamma distribution. This model is known as Poisson-Gamma model,

$$O_i|\theta_i, E_i \sim Poisson(\theta_i E_i)$$

$$\theta_i \sim Gamma(\nu, \alpha)$$

The relative risk θ_i is considered as the random variable drawn from the Gamma distribution with mean α and variance ${\upsilon}/{\alpha^2}$ (Hahne et al., 2010).

If a disease exhibits clustering then the areas of high residual risk will lead (in expectation) to an access of cases in those areas such a collection of cases is what we define as a cluster (Elliot et al., 2000). Measuring the correlation among the neighboring observations in the pattern is done using the spatial autocorrelation statistic Moran's I. Moran's I is defined as a measure of the correlation among neighboring observations in a pattern (Rock, 1988). Moran's autocorrelation coefficient (often denoted as I) is an extension of Pearson product-moment correlation coefficient to a univariate series(Paradis, 2009; Besag and Moran, 1975).

Moran's I is given by the formula

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}(X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

where W_{ij} is the weight between observation i and j, and S_0 is the sum of all weights W_{ij}'s, that is, $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}$. Small p-value associated with the test statistics of the Moran's I gives an indication of spatial variability.

3. RESULTS AND DISCUSSION

About 600 patients were reported for HCV in the Allide Hospital, Faisalabad during the period of 2010-2012. It was observed that the rural population was suffering more from Hepatitis C disease than the urban population. Jaranwala town was the most effected region during 2010-2012 as 183 HCV patients were reported in that area. The geographical disparity in the occurrence of the disease was observed in the Faisalabad district (Fig. 1). The expected count of patients for each town of Faisalabad are shown in Fig. 2.

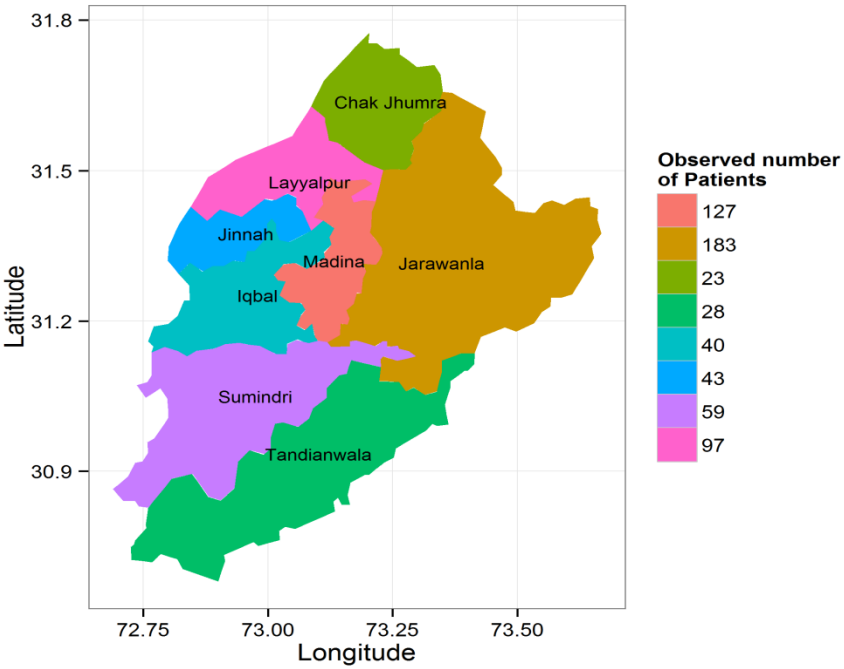


Fig. 1: Observed counts of HCV patients during the period 2010-2012 in Faisalabad district

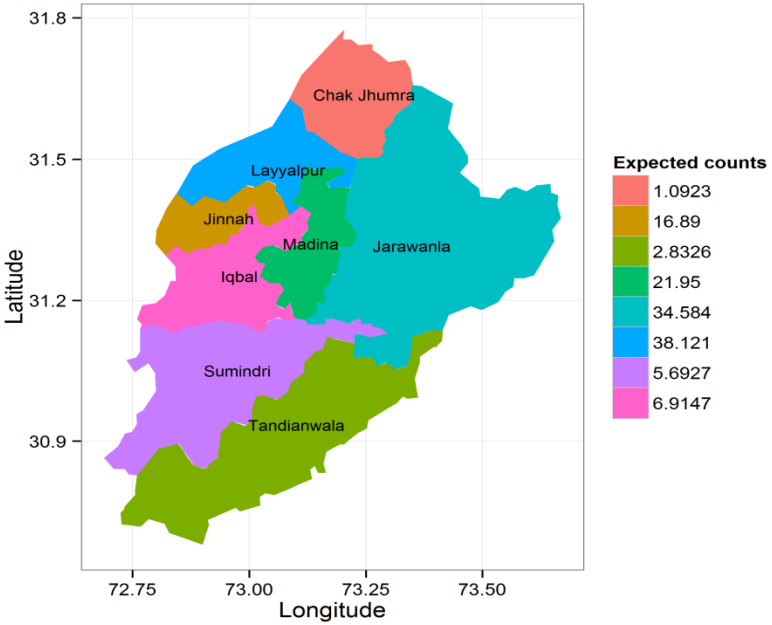


Fig. 2: Expected counts of HCV patients during the period 2010-2012 in Faisalabad district

The Moran's I test revealed that there is a spatial dependence of the disease (p-value = 0.0054) and some spatial clusters exist in the region. The Moran's I test identified Jarawanla, Madina, and Layyalpur towns as primary cluster with high incidence rate of disease and other towns as temporary cluster with low incidence rate of disease.

The Poisson-Gamma model estimated α and υ as 4.3956 and 4.6307 respectively which gives a prior mean of the relative risks of 1.0535 (very close to 1). The Poisson-Gamma model also identified the same towns Jarawanla, Madina and Layyalpur towns as the regions with high relative risk of disease (Table 1).

Table 1
The Poisson-Gamma model estimates with areas of high risk during 2010-2012

Towns	Observed	Expected	Relative Risk
Jarawanla	183	34.584	112.95274
Madina	127	21.950	123.48311
Layyalpur	97	38.121	54.31607

The results of both Moran's I test and the Poisson-Gamma model confirm geographical heterogeneity of HCV disease in Faisalabad district. The governmental and non-governmental organizations can use this information to better target health interventions and choose areas to implement control measures against HCV in a more efficient way.

REFERENCES

1. Abbas, Z., Jeswani, N., Kakepoto, G., Islam, M., Mehdi, K. and Jafri, W. (2010). Prevalence and mode of spread of hepatitis B and C in rural Sindh, Pakistan. *Tropical Gastroenterology*, 29(4), 210-216.
2. Besag, J.E. and Moran, P.A. (1975). On the estimation and testing of spatial interaction in Gaussian lattice processes. *Biometrika*, 62(3), 555-562.
3. Elliot, P., Wakefield, J.C., Best, N.G. and Briggs, D. (2000). *Spatial Epidemiology: Methods and Applications*. Oxford University Press.
4. Elliott, P. and Wartenberg, D. (2004). Spatial Epidemiology: Current approaches and future challenges. *Environmental Health Perspectives*, 998-1006.
5. Hahne, F., Huber, W., Gentleman, R. and Falcon, S. (2010). *Bioconductor Case Studies*. Springer.
6. Hashmi, A., Saleem, K. and Soomro, J.A. (2010). Prevalence and factors associated with hepatitis C virus seropositivity in female individuals in Islamabad, Pakistan. *International Journal of Preventive Medicine*, 1(4), 252.
7. Jiwani, N. and Gul, R. (2011). A Silent Storm: Hepatitis C in Pakistan. *J. Pak. Med. Stud.*, 1(3), 89-91.
8. Khattak, M., Salamat, N., Bhatti, F. and Qureshi, T. (2002). Seroprevalence of hepatitis B, C and HIV in blood donors in northern Pakistan. *J. Pak. Med. Assoc.*, 52(9), 398-402.

9. Paradis, E. (2009). *Moran's Autocorrelation Coefficient in Comparative Methods.* R Foundation for Statistical Computing, Vienna.
10. Rock, N.M. (1988). Selective bibliography of numerical geology. *Numerical Geology: A Source Guide, Glossary and Selective Bibliography to Geological Uses of Computers and Statistics*, 329-427.
11. Shaikh, B.T. and Hatcher, J. (2005). Health seeking behaviour and health service utilization in Pakistan: Challenging the policy makers. *Journal of Public Health*, 27(1), 49-54.
12. Umar, M. and Bilal, M. (2012). Hepatitis C, A Mega Menace: A Pakistani Perspective. *J. Pak. Med. Stud.*, 2(2), 68-72.
13. Walter, S. and Birnie, S. (1991). Mapping Mortality and Morbidity Patterns: An International Comparison. *International Journal of Epidemiology*, 20(3), 678-689.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 397-402

MODELING SURVIVAL RATES OF CARDIOVASCULAR PATIENTS IN FAISALABAD

Sidra Saleem and **Muhammad Yaseen**[§]
Department of Math. and Stat., University of Agriculture, Faisalabad, Pakistan
Corresponding Author: [§]myaseen208@gmail.com

ABSTRACT

Cardiovascular diseases have become common problem in developing countries including Pakistan. Survival analysis is a statistical technique used to model the time duration to until one or more events happen, such as death in biological organisms and failure in mechanical systems. In this study, we modeled the survival rates of cardiovascular patients who had heart attacks using survival analysis and log-rank test was used to compare the survival rates across gender. It was observed that the survival rates were not the same across gender. Male patients had less chances of survival as compared to female patients. The data have been collected from cardiology center, District Headquarter (DHQ) Hospital, Faisalabad included 106 cardiovascular patients who had heart attacks were observed to check survival rate from which 60 were males and 46 were females.

1. INTRODUCTION

Survival analysis is a statistical technique used to model the time duration to until one or more events happen, such as death in biological organisms and failure in mechanical systems. It has been showing to be increasingly important since the 1950s (Langova, 2008). The term survival analysis came into being from early studies, where the event of interest was death. Now the scope of the survival analysis has become wide. These days, scientists are using it for time until onset of disease, time until stock market crash, time until equipment failure, time until earthquake, and so on (Smith and Smith, 2001). The common events can be death, occurrence of disease, marriage, divorce etc. Few examples of studies where tools of survival analysis were used like leukemia patients and time in remission, time to develop a heart disease for normal individuals, elderly population and time until death, and heart transplants and time until death (Kleinbaum and Klein, 2011). Survival analysis is also known as time to event analysis. The statistical software R supports a wide variety of tools for the analysis of survival data (R Core Team, 2014; Therneau and Grambsch, 2000).

In this article, we modeled the survival rates of cardiovascular patients who had heart attacks using survival analysis and log-rank test was used to compare the survival rates across gender.

2. MATERIAL AND METHODS

Survival analysis is a statistical technique used to model the time duration to until one or more events happen. The survival time can be measured in days, weeks, years etc. This

can be affected when some of the subjects may not the experience of event of interest or death before the completed of the study, even though they died if observation continued. We label these conditions as censored observations. The data have been collected from cardiology center, DHQ Hospital, Faisalabad included 106 cardiovascular patients who had heart attacks were observed to check survival rate from which 60 were males and 46 females. The age and gender of the cardiovascular patients were considered as covariates to identify their main or interaction effect on the survival time of the patients. A coded variable was created with values of zero and one, the value of zero represents a censored value or survival and the value of one represents the death of the patient.

Survival function describes the probability that an individual alive longer than a specified time t since the start of the observation, i.e., $S(t) = P(T \geq t)$ where T is a non-negative random variable which times from the start of an observation to the occurrence of the observed event. In practice the survival function is estimated using the Kaplan-Meier method (Kaplan and Meier, 1958). The Kaplan-Meier, non-parametric procedure, is a method of estimating time-to-event models in the presence of censored cases.

In most clinical studies the comparison of two or more survival functions is also of interest. The log rank test which is also known as the Mantel log-rank test, the Cox Mantel log-rank test, or the Mantel-Haenszel test can be used to make comparison between two or more survival functions (Singh and Mukhopadhyay, 2011).

The Cox proportional hazard model was introduced by Cox (1972). It is similar to the logistic regression, but the Cox model is preferred over the logistic model which ignores survival time and censoring information (Singh and Mukhopadhyay, 2011). The Cox proportional hazard model is a statistical technique to access the relationship between survival of patient and any covariates. It allows to estimate the hazard of event of interest for individuals given their predictive variables. The formula for Cox proportional hazard model is

$$h(t, X) = h_o(t) \exp\left(\sum_{i=1}^{n} \beta_i X_i \right)$$

where $h(t, X)$ is the hazard function for an individual whose independent variables had the value X, $h_0(t)$ is the baseline or hazard function and corresponds to the probability of reaching an event when all the explanatory variables are zero, β_i's are the regression coefficients and X_i's are explanatory variables.

3. RESULTS AND DISCUSSION

The collected data had 106 cardiovascular patients from which 5 out of 46 female and 20 out of 60 male patients could not survive. On the very first day of their heart attack, the patients had the 86.8% survival chance with 95% confidence interval of 80.6-93.5% (Table 1). The survival probabilities with their 95% confidence intervals are depicted in the Figure 1. Longer stay of patients in hospitals was an indication of severity of the

disease and reduced the survival probabilities considerably especially stay of more than ten days had seen substantial decline in survival probabilities (See Table 1 & Figure 1).

Table 1
Survival Analysis Estimates

Days	Risk	Event	Survival Estimates	S.E	95% CI	
					LCL	UCL
1	106	14	0.868	0.0329	0.806	0.935
2	84	2	0.847	0.0352	0.781	0.919
3	73	3	0.812	0.0391	0.739	0.893
4	58	1	0.798	0.0408	0.722	0.883
6	29	2	0.743	0.0534	0.646	0.856
7	19	1	0.704	0.0634	0.590	0.840
10	8	1	0.616	0.0993	0.449	0.845
30	1	1	0.000	NaN	NA	NA

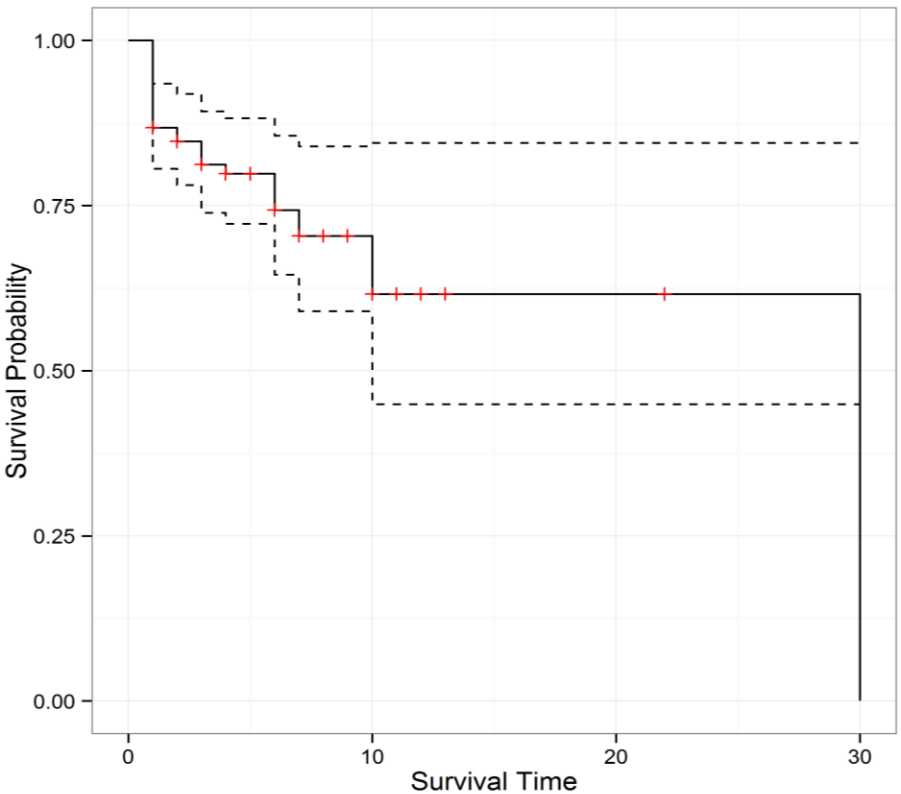


Fig. 1: Survival Probabilities with 95% Confidence Interval

Even though the survival probabilities for both male and female patients decreased with longer stay in hospitals but the survival probabilities of female patients always stayed above that of male patients (See Tables 2, 3 & Figure 2). The difference in survival probabilities was found to be statistical significant in favor of female patients $\left(\chi^2 = 5.8, df = 1, p = 0.016\right)$.

Table 2
Survival Analysis Estimates for Female

Days	Risk	Event	Survival Estimates	S.E	95% CI	
					LCL	UCL
1	46	2	0.957	0.0301	0.899	1.000
2	41	1	0.933	0.0373	0.863	1.000
3	33	1	0.905	0.0456	0.820	0.999
6	14	1	0.840	0.0753	0.705	1.000

Table 3
Survival Analysis Estimates for Male

Days	Risk	Event	Survival Estimates	S.E	95% CI	
					LCL	UCL
1	60	12	0.800	0.0516	0.705	0.908
2	43	1	0.781	0.0537	0.683	0.894
3	40	2	0.742	0.0577	0.637	0.864
4	33	1	0.720	0.0602	0.611	0.848
6	15	1	0.672	0.0728	0.543	0.831
7	11	1	0.611	0.0882	0.460	0.810
10	5	1	0.489	0.1300	0.290	0.823
30	1	1	0.000	NaN	NA	NA

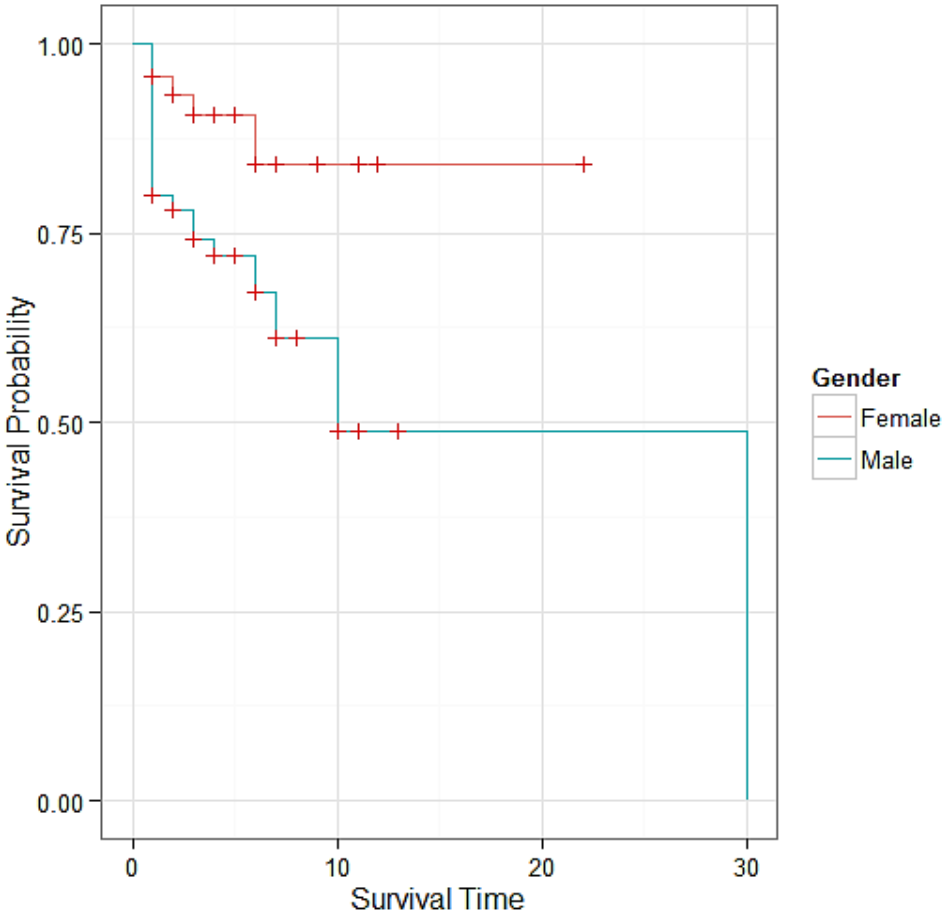


Fig. 2: Survival Probabilities for Males and Females

The estimates of the Cox proportional hazard model showed that the gender had statistically significant effect on the survival probabilities of Cardiovascular patients in favor of female patients $\left(\beta_{Male} = 1.15, Z = 2.286, p = 0.022\right)$.

4. CONCLUSION

Longer stays of Cardiovascular patients in hospitals could be an indication of severity of the disease and can reduce the survival probabilities considerably. It has been found that with longer stay, the survival probabilities could reduce substantially for both genders. However, female Cardiovascular patients had high survival probabilities than male patients.

REFERENCES

1. Cox, D.R. (1972). Regression Models and Life-Tables. *J. Roy. Statist. Soc., Series B*, 34(2), 187-220.
2. Kaplan, E.L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *J. Amer. Statist. Assoc.*, 53(282), 457-481.
3. Kleinbaum, D.G. and Klein, M. (2011). *Survival Analysis: A Self-Learning Text*. Springer, 3-rd Edition.
4. Langova, K. (2008). Survival Analysis for Clinical Studies. *Biomedical Papers of the Medical Faculty of Palacky University in Olomouc*, 152(2), 303-307.
5. R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
6. Singh, R. and Mukhopadhyay, K. (2011). Survival Analysis in Clinical Trials: Basics and must know areas. *Perspectives in Clinical Research*, 2(4), 145.
7. Smith, T. and Smith, B. (2001). Survival analysis and the application of Cox's proportional hazards modeling using SAS. In *Proceedings of the twenty-sixth annual SAS users group international conference. SAS Institute Inc., Cary, NC*, 244-246.
8. Therneau, T.M. and Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 403-420

FACTORS AFFECTING YIELD AND TASTE OF POPCORN: A PLANNED APPROACH

Laiq Muhammad Khan and **Ejaz Ahmed**
College of Computer Science and Information System
Institute of Business Management, Karachi, Pakistan
Email: laiq.muhammad@iobm.edu.pk; eahmed @iobm.edu.pk

ABSTRACT

Design of experiment is a powerful statistical tool. It helps in identifying factors (input) that results in variation in the output. We received complaints from a local food center indicating that they were observing serious quality issues at their popcorn selling point. Specifically they identified two major problems, namely (i) many a time the proportion of un –popped was more than usual (ii) few customers complained about the taste of popped popcorn.

We were approached to address the problem and identify factors responsible for complaints and Taguchi's L_8 orthogonal array was used and experiment was repeated under different conditions. This article describes complete protocol of the experiment and identifies the factor, level combination for optimum results.

The focus of this experiment was to decide the factors which affect the yield and taste of popcorn.

This study provides a recipe of popping popcorn using the commonly used pots of the kitchen, stove or burner of the cooking range in the kitchen and kernels (the experiment was run in a home environment).

In this study effect of six main parameters Type of pot, Brand of popcorn, Duration of heating, Type of fat, Quantity of fat and Quantity of kernels over the yield and taste of popcorn were studied. Effect of each main factor was studied at two levels" Low" and "High".

The study finds an optimum factor-level combination of all five factors giving yield of popcorn with good taste.

KEY WORDS

Popcorn; Design of experiment; Fractional factorial Design; Replication.

JEL Classification: C 90

1. INTRODUCTION

Design of experiments (DoE) is a powerful tool that helps experimenters to optimize yield identifying most suitable factor-level combination. In practice, however it is not taught as it deserves. In many sessions that we attended DoE is very soon backed by

analysis of variance (ANOVA) and students are taught how to test hypothesis under different design situations. As such students fail to appreciate the strength of DoE as they perceive it to be nothing but a quantitative problem solving technique. We were planning to teach DoE in a different style when we had an interaction with a local popcorn seller who told us about few of his problems. He was worried about the losses he was facing due to bad quality of kernels resulting in losses. Lesser amount of popcorn ready for sale was an issue. We found after discussion that customers complained about the taste of popcorn. Majority of customers were of the view that taste was not consistent and that they experienced taste variation a lot. The vendor was worried as he used almost same recipe and adopted similar procedure still was unable to find why taste was not consistent.

We then arranged a meeting with few other popcorn sellers to have a better idea about the problem and to know the practices they usually adopted. Later it was decided to develop a proper design layout, identify the factors and their levels, establish input-output relationship and run a properly designed experiment. It was further decided to document protocols of the experiment in a professional way and analyze findings to communicate a practical solution.

The first author observed that one vendor was using cooking oil while another vendor was using vegetable ghee (a local form of oil) for popping the popcorn. Vendors told that the quality and taste of popcorn gets better by using ghee instead of the cooking oil. When we asked what they meant by quality, they replied that popped popcorn should be of large size and unpopped popcorn should be very few with no burns.

1.1 What is Popcorn?

Popcorn is a popular snack among people of all ages. People like to eat popcorn particularly when they are watching movie, drama or football match, for example. Nowadays kernels are easily available in packets in area shops, general stores, shopping malls in major cities all around the world an of Pakistan. These kernels are popped in microwave oven. But these types of microwave popcorns are expensive for people in developing countries as such as Pakistan where a vast majority cannot afford this.

Popcorn was probably an important part of life in the ancient Americans. On a 1,700-year old painted funeral urn found in Mexico, a corn God is shown wearing a headdress of popcorn. *(Popcorn, Ingrained in America's Agricultural History, http://specialcollections.nal.usda.gov/popcorn-exhibit)*

United States is the largest popcorn producing country of the world. Maize is being cultivated in 25 states of United States. Americans today consume 17.3 billion quarts of popped popcorn each year. An average American eats about 68 quarts of popcorn. *("United States Department of Agriculture – 1982 Popcorn Report")*

Barnaby Cobo a missionary in Peru between 1609 and 1629 observed that Peruvian Indian toasted a certain kind of corn until it bursts. They called it pisancalla and used that as a confection *("Popcorn was popular in Ancient Peru, Discovery Suggests". History.com.20 January 2012)*

Popcorn and candy stand in a park at Topeka, Kansas, July 31, 1912

Popcorn vendor at Paris, Illinois, August 3, 1912

1.2 Popcorn in Pakistan

In the year 2010-11 Pakistan produced an estimated 3.707 million tons of Maize and total cropped area of Maize was 974.2 thousand hectares (Statistics Division, Pakistan Bureau of Statistics). The Punjab province accounted for 79.8% and Khyber Pakhtunkhwa Province of Pakistan accounted for 19.7 % of national production. Very little Maize is produced in the other two provinces Sindh and Baluchistan of Pakistan.

Table 1
Production and Area under Cultivation of Maize, Year 2010-2011

Provinces	Production (000 tons)	Area(000 hectares)
Punjab	2959.2	543.6
Sind	1.6	2.5
KPK	740.5	422.9
Baluchistan	5.7	5.2

(Source: Statistics Division, Pakistan Bureau of Statistics)

Figure 1: Local Setup of Popping Popcorn in Karachi

1.3 How a kernel changes into Popcorn

A kernel of popcorn contains a small amount of water stored inside a circle of soft starch. The soft starch is surrounded by the kernel's hard outer surface. As the oil around the kernels heats up, the water expands building pressure against the hard starch surface. Eventually, this outer layer gives way, causing the popcorn to explode. As it explodes, the soft starch inside the popcorn becomes inflated and bursts, turning the kernel inside out. The steam inside the kernel is released, and the popcorn is popped, hot and ready to eat.

2. LITERATURE REVIEW

Anderson and Anderson (1993) conducted a study about the microwave popcorn to decide the vital factors which affect the yield of popcorn. They used five factors namely Price, Time, Power, Preheat and Elevate, each two levels. Their response variables were yield and taste. They found that there is no effect of preheating over the taste and yield. Factors price, time, power, and elevation were found significant. Anderson and Anderson (1993) concluded that taste response varies significantly depending on other factors such as temperature and settings of microwave oven.

Ort (2010) performed an experiment "Great Popcorn Experiment" to compare the method of stovetop popcorn with preheated and cold oil and to decide whether the viscosity of the oil causes problems in popping or not. He opted following two procedures to conduct the experiment.

Method 1 (The Ruhlman method): Corn and oil were placed together into a cold pan and heated, shaking pretty constantly and lid on tightly. The pan was removed from heat after fifteen seconds when popping stops.

Method 2 (The David Ort method): Placing corn oil in a pan and heating over medium heat. Once oil was heated well, the corns were placed in the pan and heated more till the popping stops and remove the pan from heat.

Ort (2010) found that the second method when the oil is first heated results better yield.

The current study was planned to modify and if possible improve the results of previously conducted studies. It was also of interest to formalize the procedure necessary to adopt while conducting properly designed experiment.

In next section the complete plan and execution of the experiment is explained. In section 4, the analysis and conclusions are presented. Confirmatory test based on optimum factor-level combination is described in section 5.

3. EXPERIMENTAL PLAN

In the current study we brainstormed a lot and selected six factors, as listed in Table 2

The Four noise factors that could not be easily controlled were.

a. Temperature of pot at the time of starting of each experiment
b. Time between runs
c. Cleaning pot after each experiment
d. Flame of stove

It was observed that stove flame was also depending on gas pressure that also does not remain a constant.

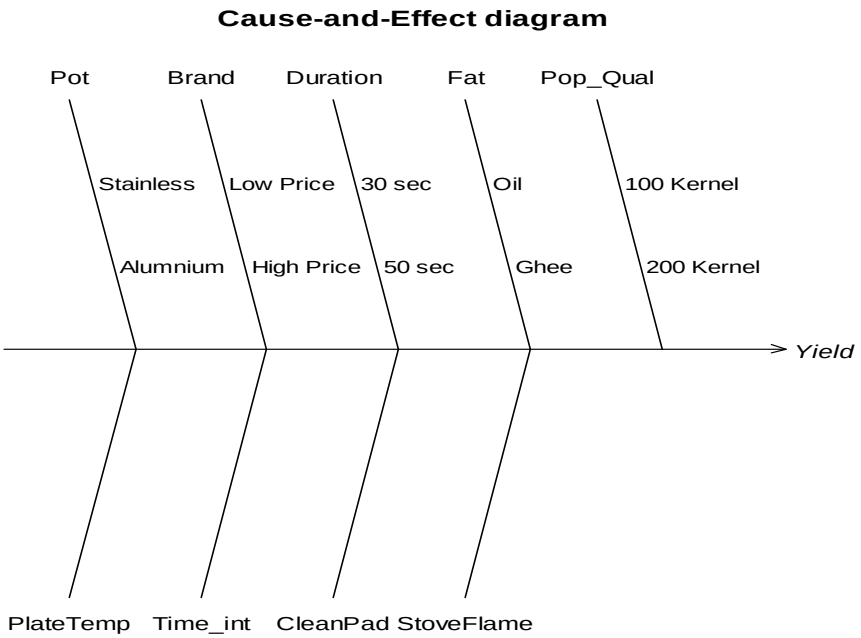


Figure 2: Fishbone Diagram Displaying Cause and Effect Procedure

Table 2:
Factors selected for the experiment and their levels

Code	Factor	Low (-1)	High (+1)
A	Type of saucepan	Steel	Aluminum
B	Brand of kernels	Low price	High price
C	Time of heating	30 seconds	45 seconds
D	Type of fat	Cooking oil	Vegetable ghee
E	Quantity of fat	15 mgs	24 mgs
F	Quantity of Kernels (#)	100	200

A 2^6 full factorial designed would have required 64 experimental runs. To make the experiment simple and executable it was decided to run 2^{6-3} fractional factorial experiment. The experiment was replicated in three replicated. The response variables in this study were yield and taste of popcorn.

Eatable Popcorns Unpoped kernels, Bullets, Burn Kernels

Aluminum Pot Stainless Steel Pot

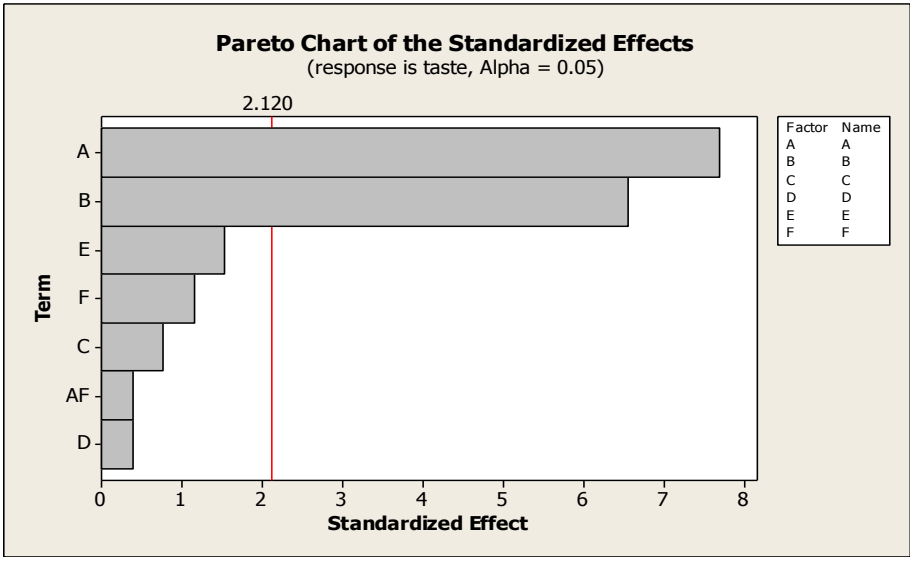


Figure 3: Resources required for the Experiment

3.1 Protocol of the Experiment

The experiment was performed in our small kitchen. One of the five burners of the cooking range was used. In order to judge the taste of popcorn a panel of judges consisting of first author's wife and daughter in law was arranged. The taste of popcorns in each run was recorded on a rank 1 to 10 with 10 being highest. The ranks assigned in different runs were averaged. The protocol of the experiment is outlined below:

Step	Details
1.	Arranged the required item.
2.	Ignited one of the five burners.
3.	Placed the pot selected according to run order over the flame and heated for ten seconds only in first run.
4	Selected fat in required quantity, as of run was put in the pot and heated for designated time.
5.	Kernels of selected brand were heated for designated time.
6.	Started counting time when the 4^{th} kernel popped after the designated time.
7.	a) Pot moved from the flame. b) Eatable popcorns counted. c) 10 popcorns selected randomly by the judges to judge the taste.

Instructions:

a) The pot was cleaned after each run.
b) The pot was heated for 10 seconds with fat in each run before putting the kernels in the pot.

4. ANALYSIS

The recorded observations of yield and taste of popcorn are given in Table 3

Table 3
Design Matrix for six factors and two levels for yield of popcorn

Run	Type of pot	Brand of Pop corn	Time	Fat	Qty of fat	Qty of popcorn	Yield of popcorn			Taste of popcorn		
	A	B	C	D	E	F	Y_1	Y_2	Y_3	Y_1	Y_2	Y3
1	1	1	-1	1	-1	-1	83	92	76	8.5	8.5	9.0
2	-1	1	-1	-1	1	-1	81	77	65	7.5	7.0	7.0
3	1	1	1	1	1	1	96	84	81	9.0	9.0	9.0
4	1	-1	1	-1	1	-1	81	76	82	7.5	7.0	7.5
5	-1	-1	1	1	-1	-1	67	62	82	5.5	5.5	5.0
6	1	-1	-1	-1	-1	1	85	75	78	7.5	7.0	7.5
7	-1	-1	-1	1	1	1	59	56	73	5.0	7.5	6.0
8	-1	1	1	-1	-1	1	59	79	87	7.5	6.5	7.0

4.1 Statistical Analysis of the Yield of Popcorn

The normal probability plot Figure 4 shows that the yield of popcorn follows normal distribution

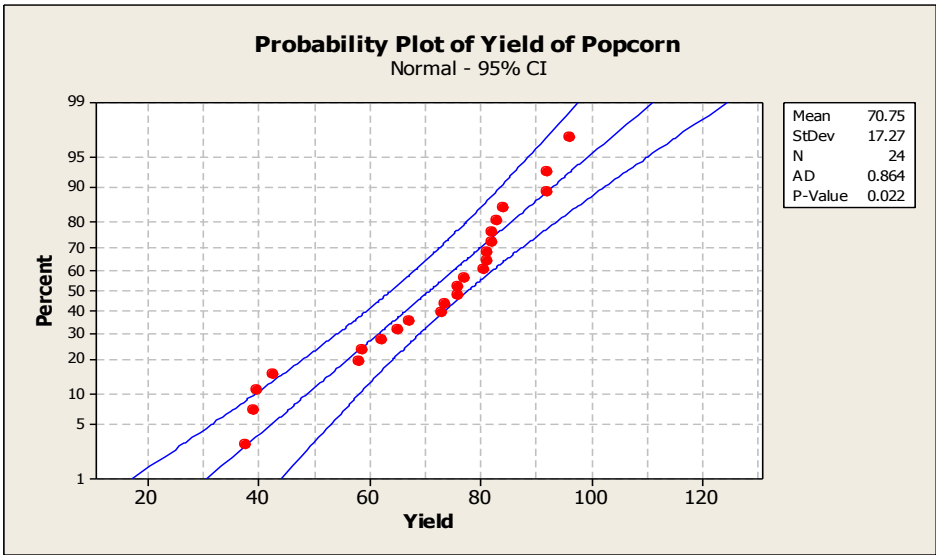


Figure 4

ANOVA Table for Yield of Popcorn

Table 4
Estimated Effects and Coefficients for Yield (Coded Units)

Term	Effect	Coef	SE Coef	T	P
Constant		70.750	2.404	29.44	0.000
A	3.417	1.708	2.404	0.71	0.487
B	12.583	6.292	2.404	2.62	0.019
C	8.583	4.292	2.404	1.79	0.093
D	13.000	6.500	2.404	2.70	0.016
E	13.000	6.500	2.404	2.70	0.016
F	-12.500	-6.250	2.404	-2.60	0.019
A*F	-5.917	-2.958	2.404	-1.23	0.236

S = 11.7748 PRESS = 4991.25
R-Sq = 67.64% R-Sq(pred) = 27.20% R-Sq(adj) = 53.49%

Analysis of Variance for Yield (coded units)

Source	DF	Seq SS	AdjSS	AdjMS	F	P
Main Effects	6	4427.63	4427.63	737.94	5.32	0.003
A	1	70.04	70.04	70.04	0.51	0.487
B	1	950.04	950.04	950.04	6.85	0.019
C	1	442.04	442.04	442.04	3.19	0.093
D	1	1014.00	1014.00	1014.00	7.31	0.016
E	1	1014.00	1014.00	1014.00	7.31	0.016
F	1	937.50	937.50	937.50	6.76	0.019
2-Way Interactions	1	210.04	210.04	210.04	1.51	0.236
A*F	1	210.04	210.04	210.04	1.51	0.236
Residual Error	16	2218.33	2218.33	138.65		
Pure Error	16	2218.33	2218.33	138.65		
Total	23	6856.00				

Alias Structure

I + A*B*D + A*C*E + B*C*F + D*E*F + A*B*E*F + A*C*D*F + B*C*D*E
A + B*D + C*E + B*E*F + C*D*F + A*B*C*F + A*D*E*F + A*B*C*D*E
B + A*D + C*F + A*E*F + C*D*E + A*B*C*E + B*D*E*F + A*B*C*D*F
C + A*E + B*F + A*D*F + B*D*E + A*B*C*D + C*D*E*F + A*B*C*E*F
D + A*B + E*F + A*C*F + B*C*E + A*C*D*E + B*C*D*F + A*B*D*E*F
E + A*C + D*F + A*B*F + B*C*D + A*B*D*E + B*C*E*F + A*C*D*E*F
F + B*C + D*E + A*B*E + A*C*D + A*B*D*F + A*C*E*F + B*C*D*E*F
A*F + B*E + C*D + A*B*C + A*D*E + B*D*F + C*E*F + A*B*C*D*E*F

Table 4 shows that the factors Brand of popcorn (B), Duration of popping(C), Type of fat (D), Quantity of Fat (E) and Quantity of Kernels(F) being popped affects the yield of popcorn significantly. This is also supported by Figure 5 where normal probability plot indicates the significance effect of factors B, D, E and F

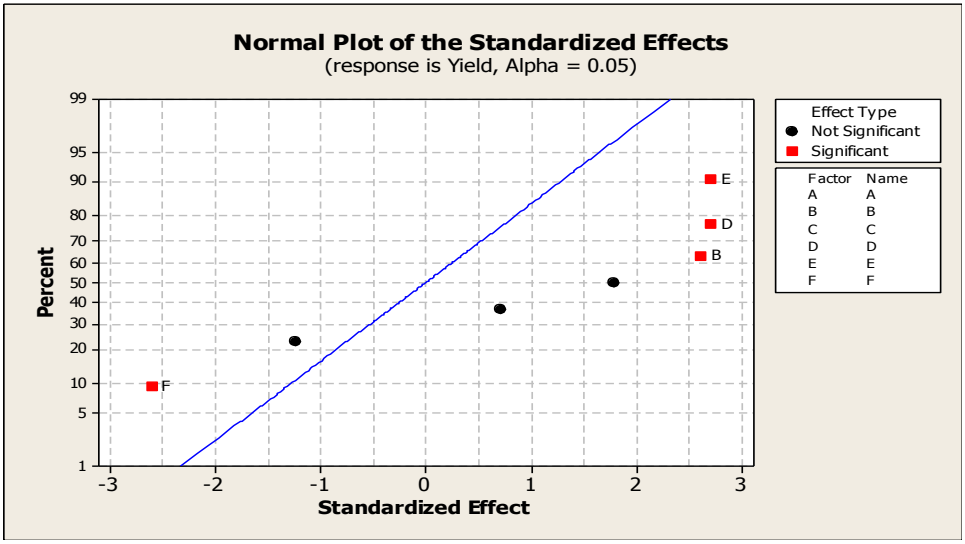


Figure 5

By performing residual analysis we observed that residuals are normally distributed, Figure 6.

It is also evident that residuals behave well initially but becomes volatile later.

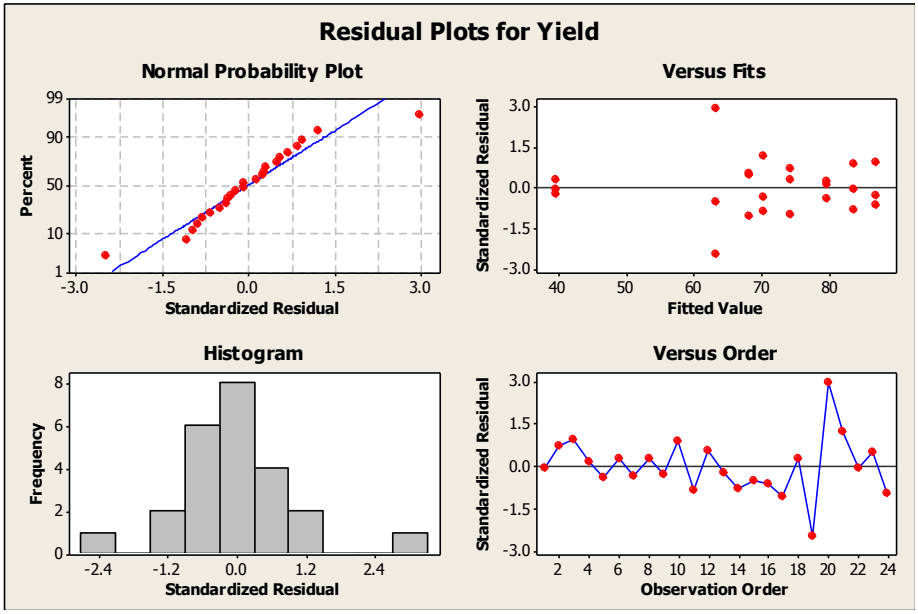


Figure 6

Figure 7 shows the contribution of each factor among the factors affecting the yield of popcorn. It can be seen that among the affecting factors, the Factor ”Quantity of Fat ” (E) has the maximum affect over the yield and the factor “Quantity of Kernels” (F) has the minimum effect over the yield of popcorn

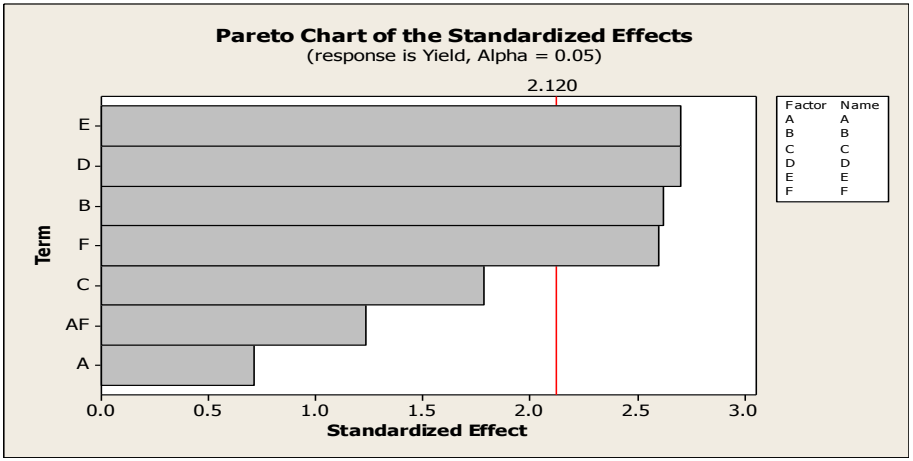


Figure 7

The main effect plot given in Figure 8 reveals, in addition that factors B, D and E affects the yield at their high level, whereas factor F affects at its lower level.

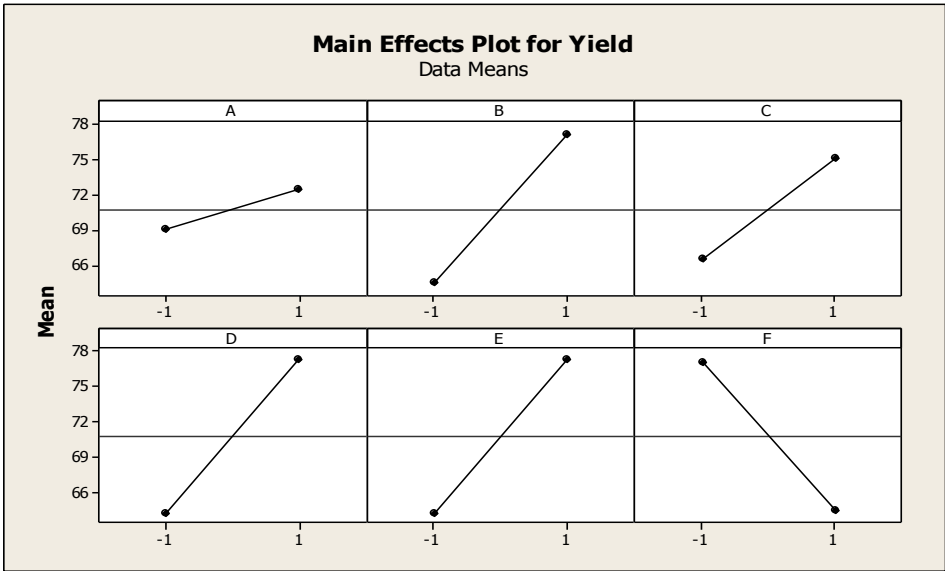


Figure 8

Analysis of variance Table 4 does not indicate any significant interaction between the factors under study, but Figure 9 indicates that there is interaction between the factors A and B, A and C, and A and D but these interaction effects are not significant.

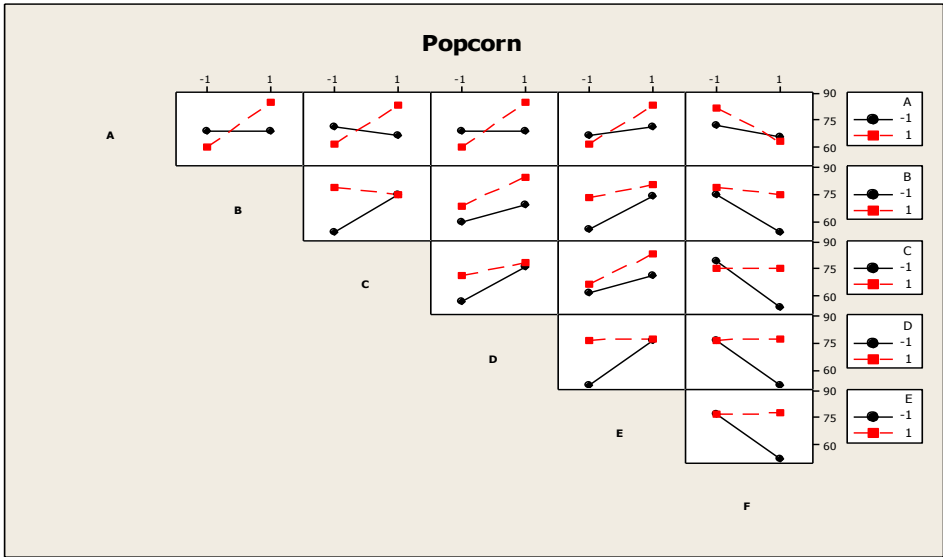


Figure 9 Interaction between the factors under Study affecting the Yield of Popcorn

The fitted model relating yields with significant factors turn out to be

$$\hat{y} = 70.750 + 6.292\ B + 4.292\ C + 6.5\ D + 6.5\ E - 6.25\ F$$

4.2 Statistical Analysis of the Taste of Popcorn

In this section we perform similar analysis on the taste data of popcorns, collected during the experiment presented in Table 3. Figure 10 displays that taste data follows normal distribution and behaves well.

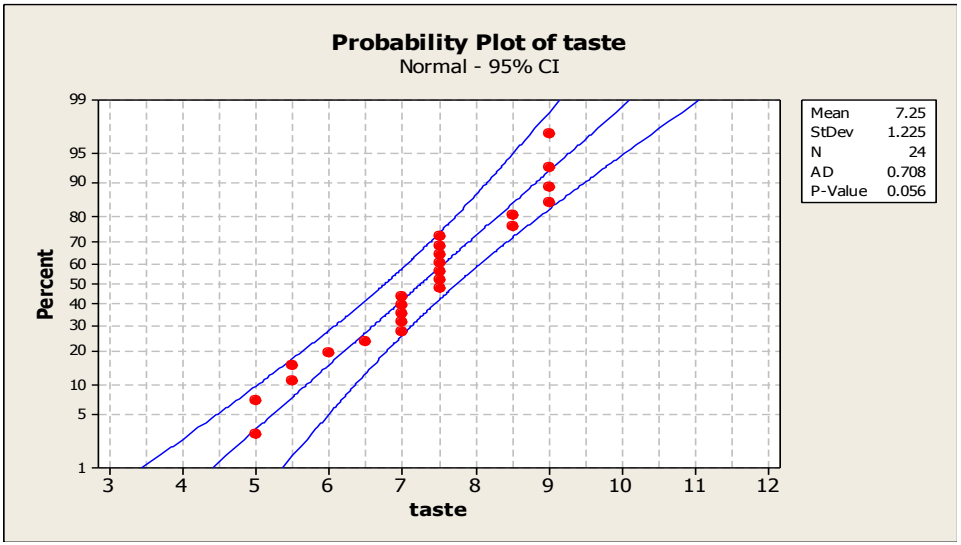


Figure 10

Analysis of Variance (ANOVA) Table 5 indicates that the effect of the factors (Type of pot) A and (Brand of popcorn) B are significant over the taste of popcorn.

ANOVA Table
Table 5
Estimated Effects and Coefficients for taste (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		7.25000	0.1083	66.97	0.000
A	1.66667	0.83333	0.1083	7.70	0.000
B	1.41667	0.70833	0.1083	6.54	0.000
C	-0.16667	-0.08333	0.1083	-0.77	0.453
D	0.08333	0.04167	0.1083	0.38	0.705
E	0.33333	0.16667	0.1083	1.54	0.143
F	0.25000	0.12500	0.1083	1.15	0.265
A*F	-0.08333	-0.04167	0.1083	-0.38	0.705

S = 0.530330 PRESS = 10.125
R-Sq = 86.96% R-Sq(pred) = 70.65% R-Sq(adj) = 81.25%

Analysis of Variance for taste (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	6	29.9583	29.9583	4.9931	17.75	0.000
A	1	16.6667	16.6667	16.6667	59.26	0.000
B	1	12.0417	12.0417	12.0417	42.81	0.000
C	1	0.1667	0.1667	0.1667	0.59	0.453
D	1	0.0417	0.0417	0.0417	0.15	0.705
E	1	0.6667	0.6667	0.6667	2.37	0.143
F	1	0.3750	0.3750	0.3750	1.33	0.265
2-Way Interactions	1	0.0417	0.0417	0.0417	0.15	0.705
A*F	1	0.0417	0.0417	0.0417	0.15	0.705
Residual	Error	16	4.5000	4.5000	0.2813	
Pure	Error	16	4.5000	4.5000	0.2813	
Total	23	34.5000				

Unusual Observations for taste

Obs	StdOrder	taste	Fit	SE Fit	Residual	St Resid
12	9	5.00000	6.16667	0.30619	-1.16667	-2.69R
17	17	7.50000	6.16667	0.30619	1.33333	3.08R

R denotes an observation with a large standardized residual.

Half Normal probability plot, Figure 11 for taste shows that two factors (Type of pot) A and (Brand of popcorn) B have significant effect over the taste and are responsible for the variation in taste of popcorn.

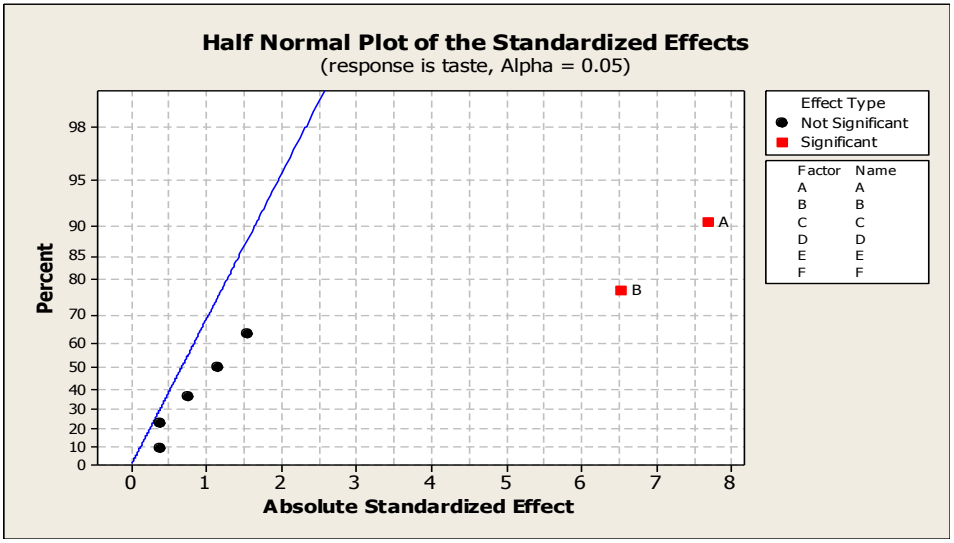


Figure 11

The fitted model shows that residuals behave well except at two points. Looking at the run plot of residuals in Figure 12, it reveals that residuals display a volatile behavior at run number 12 and 16; otherwise the residual performance is quite good.

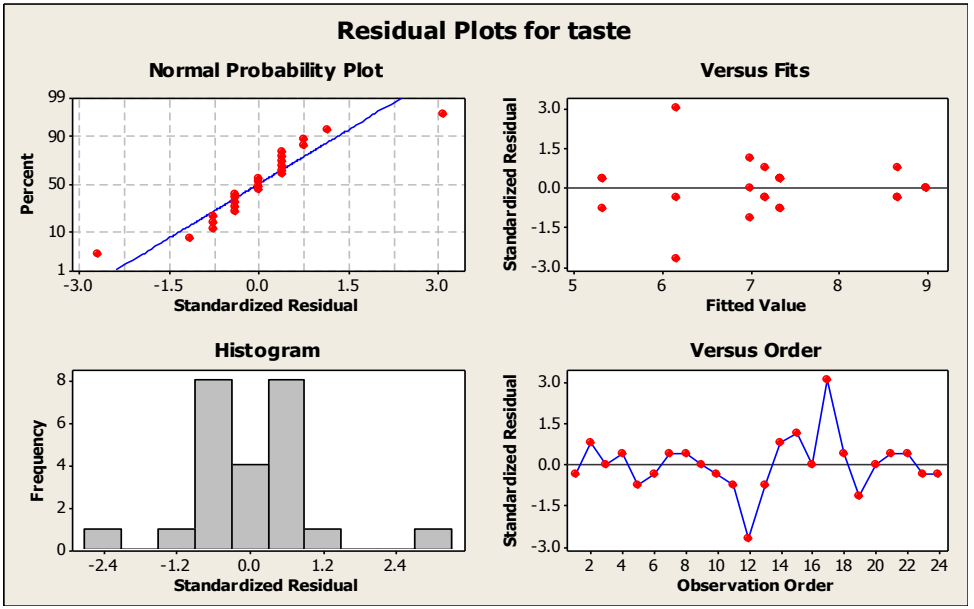


Figure 12

The main effect plot given in Figure 13 reveals that the factors A and B show better taste at higher levels i.e. when good quality of popcorns are popped in Aluminum pot the taste of the popcorn is better.

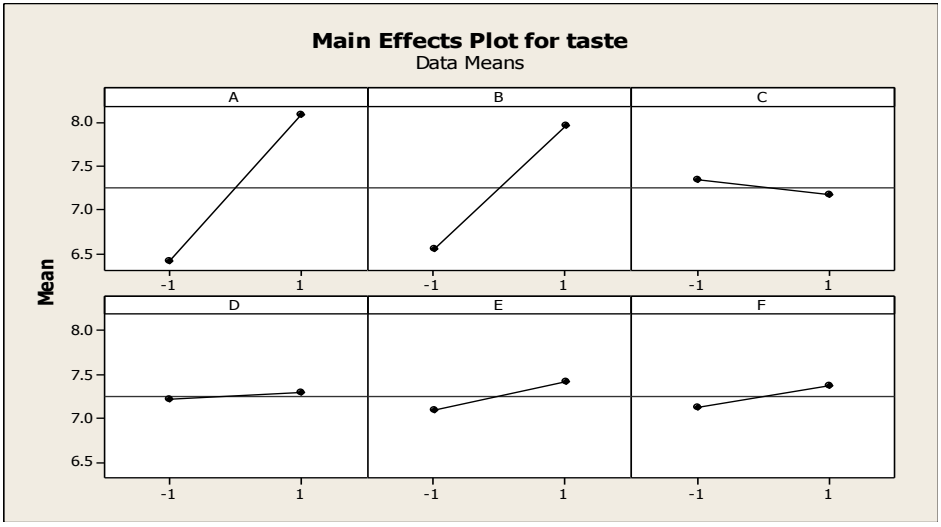


Figure 13 Aluminum pot and using the high price Kernels.

Pareto Chart given in Figure 14 indicates that the contribution of the factors A and B is significant among all factors. It shows that the good quality popcorn popped in Aluminum pot gives better taste

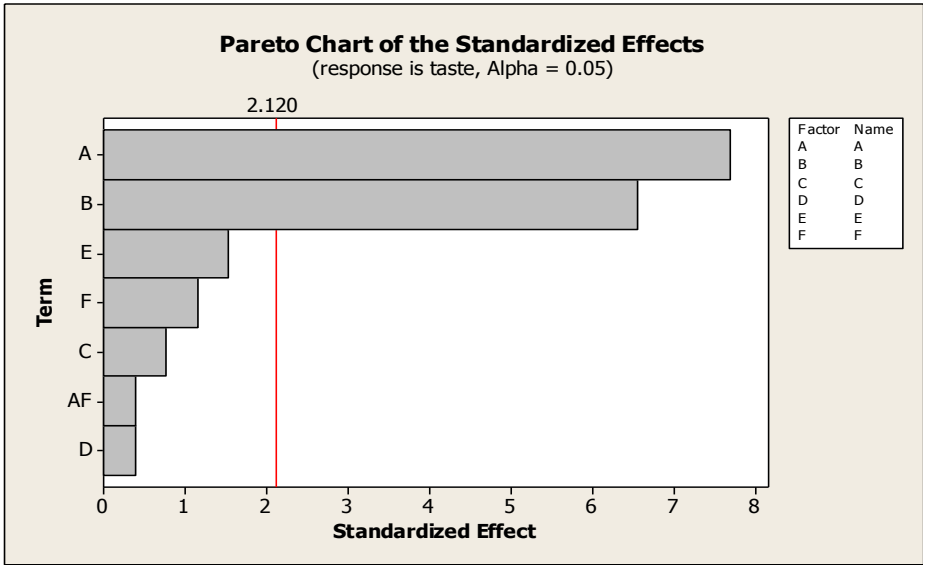


Figure 14

Interaction plot given in Figure 15 shows an interaction effect between C and E, and between factors C and F for the taste of popcorn data. However we found that both of these interaction effects are not significant

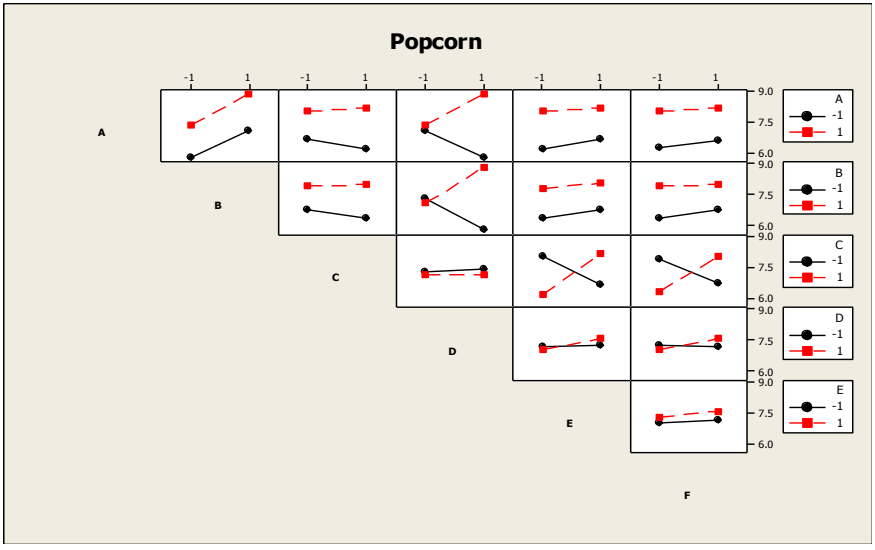


Figure 15

5. CONFIRMATORY EXPERIMENT

Based on the findings of this experiment and results obtained through various procedures, optimum factors level combinations were identified. A confirmatory test was then planned and five run, using optimum factor level combination were conducted as given in table 6.

Table 6

Optimum Functional Level Contributor					Yield of Popcorn					Taste of Popcorn				
A	**B**	**C**	**D**	**E**										
+	+	+	+	-	94	97	92	92	93	8.5	8.5	8	8	8.5

6. CONCLUSION

A study was conducted to demonstrate the effectiveness of the design of experiment. Popcorn were popped and two out puts yield and taste Six factors were identified as input (controllable) factors. It was further decided to use each factor at two levels. Since complete 2^6 experiment was difficult to run a 2^{6-3} design with three replicates was conducted. Properly designed experiment (supported by quality Control tools such as PDCA cycle was used and specific protocol was followed.

Analysis of this planned experiment showed that Type of pot(A) Brand of popcorn (B),Type of fat (D), Quantity of fat (E), and Quantity of Kernels (F) are significant factors affecting the yield and taste of popcorns.

Optimum levels were determined and confirmatory taste was conducted to find out the validity of our conclusion. The confirmatory test supported our findings.

REFERENCES

1. Anderson, M.J. and Anderson H.P. (1993). Applying DOE to Microwave Popcorn. *Process Ind. Quality*, 3, 1-3.
2. Anderson, M.J. and Whitcomb, P.J. (2007). *Practical Tools for Effective Experimentation*, 2nd Edition, Productivity Press.
3. Antony J. and Capon N. (1998). Teaching Experimental Design Techniques to Industrial Designers. *International Journal of Engineering Education*, 14(5), 335-343.
4. Azeredo M.B.V., Silva, S.S. and Rekab K. (2003). Improve Molded Part Quality. *Quality Progress*, 72-76.
5. Box, G.E.P and Bisgaard, S. (1987). The Scientific Context of Quality Improvement, A look at the use of scientific method in quality improvement. *Quality Progress*, 54-61.
6. Hunter W.G. (1977). Some Ideas about Teaching Design of Experiments, with 2^5 Examples of Experiments Conducted by Students. *The American Statistician*, 31, 12-17.

7. Joaquin E.A. and Bestard A. *Factorial Experiment Design to Analyze Fuel Consumption of a Vehicle*, Online available at www.mdc.edu/hialeah/.../AnArticle_ ExpDesign_JoaqimBestard.pdf
8. *Popcorn was Popular in Ancient Peru*, Discovery Suggests. History.com.20 January 2012
9. *Popcorn, ingrained in America's Agriculture History*, http://specialcollections.nal.usda.gov/popcorn-exhibit
10. United States Department of Agriculture – *1982 Popcorn Report*.
11. Vining, G. (2013). Technical Advice: Experimental Protocol and the Basic Principles of Experimental Design, *Quality Engineering*, 25, 307-311.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 421-432

STUDY OF PROBLEMS FACED BY THE PRIVATE SECONDARY SCHOOLS SUBJECT TEACHERS IN THE INTEGRATION OF ICT DURING TEACHING AND LEARNING PROCESS

Rubina Usman Ali and **Tayyaba Zarif**
Shaheed Benazir Bhutto University, Shaheed Benazirabad, Pakistan
Email: dr.tayyabazarif@gmail.com

ABSTRACT

This research is aimed to observe the problems faced by the private Secondary subject teachers in the integration of ICT during teaching and learning process. The main focus of this research is to examine the effects of various factors like infra-structure, IT trainings, management support and teacher interest on integration of ICT. The study used a survey method to collect the data. The questionnaire comprised 23 questions and response was solicited against 5 point scale. 10 Private Schools located in 10 towns of Karachi were included in the study. 10 teachers (60% of staff) from one school were selected randomly on the basis of subjects from each 10 towns for this current research, so 100 subject teachers are selected as a target sample who were administered the questionnaire. Rate of return of the questionnaire was 100%. For analysis primary data has been used. Regression and correlation are used as analytical tools for the empirical estimation. This study finds that the management support and teachers interest in ICT are statically significant factors with a positive sign. Moreover, the Infra-structure and teachers IT training has also be found to be statistically significant factors with negative sign. Results suggest that in order to implement ICT in subject teaching, the management should ensure the proper infra-structure of computer lab with quarterly preplanned teachers trainings regarding ICT, so the teachers will show positive attitude towards technology aided learning. This study is likely to be a good guide for those secondary schools in both private and public sectors, which have ICT infra-structure. It could help them integrate ICT in teaching and learning process.

KEY WORDS

ICT; Teachers attitude; Infra-structure; Integration of ICT.

1. INTRODUCTION

In any institute management support is a key factor in the integration of ICT in subject teaching. In Lyari town the lack of interest of management observed during the last 5 years. This study examines why teachers do not incorporate and use computer technology as a resource, in teaching practices. The study specifically investigated teachers' problems towards ICT resources use in their instruction. The study also explored the factors that affect teachers' use or non-use of computer technology resources in teaching. The study finally examines whether and to what extent opportunities, facilities and training are provided to teachers and to what extent they contribute. This research study is based on the reasons that why teachers are not using ICT as a teaching tool in secondary level teaching.

The ubiquitous nature of Information and Communication Technologies (ICTs) has transformed almost every facet of our society and radically changed how people live, work and play (Komza 2005). As part of that transformation, the pervasiveness of ICT in classrooms has become essential, as educational institutions must prepare students for living in 'a knowledge society' as part of this 'information age'. To achieve this preparation, the need to integrate ICT in classrooms has become critical to the work of schools. It was hoped that in conjunction with preparing students for this technologically driven society, teachers would also adopt new and innovative teaching methods. schools will have little positive effect without proper broadband connections and training of teachers to use computers effectively". Teachers therefore, are seen as key players in relation to this infusion of ICT. However, despite considerable funding, more powerful equipment and extensive support structures, it appears that ICT has not brought about the widespread changes in teaching methodologies that was initially hoped for. ICT it seems, has been merely superficially embraced by many teachers and many schools. A substantial body of research asserts that teachers have difficulty in integrating ICT because of obstacles or barriers. These barriers, if not addressed, can have a profound influence on the integration of ICT in both mainstream secondary and classrooms. Literacy and numeracy levels amongst students are often weak and they often have very negative attitudes towards education.

This study examines why teachers do not incorporate and use computer technology as a resource, in teaching practices. The study specifically investigated teachers' problems towards ICT resources use in their instruction. The study also explored the factors that affect teachers' use or non-use of computer technology resources in teaching. The study finally examines whether and to what extent opportunities, facilities and training are provided to teachers and to what extent they contribute. This research study is based on the reasons that why teachers are not using ICT as a teaching tool in secondary level teaching.

ICT skills are improving teaching and learning environments, students are better able to develop the required skills. But the teachers who have trained are not implementing ICT in teaching successfully they have some problems this study based on those problems.

The purpose of this study is to examine the problems to the integration of ICT in teaching. In this study all the schools have ICT infra-structure and the whole staff are ICT trained by Intel Teach Program, so the basic problems faced by those teachers are the main focus of this study. Based on this study the research provides recommendations on improving ICT integration in the class rooms.

The need of this study is to investigate and identify the problems that prevent the successful integration of ICT in those private secondary schools of Karachi which have ICT infra-structure and their staff are ICT trained. This study presents a portrayal of the experiences of teachers in relation to technology integration in schools of private sector. It examines how teachers are using ICT and what factors prevent ICT use. Attention is given to internal problems, within teachers themselves and external problems, over which the teacher may have little or no control.

The main aims of this study was

- To evaluate the teachers competency of integrating computers.
- To investigate the problems that prevents the integration of ICT in secondary schools of Karachi.
- To analyze the factors lead to successful ICT integration in the classroom.
- To understand the importance of infra-structure in the implementation of ICT in teaching.
- To record the computer accessibility and its impact in the integration of ICT in the classrooms.
- To show the impact of management support in the integration of ICT in class rooms.
- To understand the importance of time management of integration of ICT in teaching.
- To know the teachers competency of integrating computers.
- To analyze the importance of computer trainings for integration of ICT.

This research study has the potential to contribute to existing research in relation to the obstacles preventing ICT integration in the learning process. This research is expected to benefit educators by extending the knowledge base that exists already, as it presents empirical evidence in relation to these problems.

The study may be of significance to other teachers, since many of them may also experience the same difficulties, as those encountered by teachers that participated in this research. The findings from this research could be generalized to other same kinds of secondary schools of Karachi.

This study may help to raise awareness among management and teachers, about the problems to ICT integration that exist. A thorough understanding of problem, will inform educators, in deciding how to address them, in the hope that they can be eliminated entirely from the teaching and learning process.

2. LITERATURE REVIEW

The use of ICT in the classroom has become important, as it provides opportunities for students to learn how to operate in an information age. The study of obstacles to the use of ICT in education may assist educators in overcoming barriers and support students in becoming successful technology adopters in the future. This literature review analyses some relevant literature and aims to identify the perceived barriers to technology integration in education. The availability of computer equipment does not in itself guarantee ICT integration in education. Successful implementation is a complex process, determined by pedagogical values, attitudes, curricular needs and physical infrastructures (Granger *et al.* 2002, p.480). Akbaba-Altun (2006), in a study of eighteen Turkish schools, concluded that successful integration of technology is not simple, because it depends on such interlinking variables. ICTs are radically transforming the curriculum in a number of ways, demanding that teachers reflect on traditional methodologies.

Educators themselves accepted that the integration of ICT into the classroom will greatly enhance the learning experience (Sutherland *et al.* 2004). The growth of ICT help

students to adjust in the global economy, the way in which they are taught and what they are taught, requires adjustments to and around ICT (Watson 1999). However, Balanskat *et al.* (2006) argued that teachers appear to acknowledge the value of ICT, difficulties continue to be encountered in adopting and integrating such technologies, while Mueller *et al.* (2008), concluded that although many teachers are comfortable with technology in general, they still may not be ready or capable to integrate such technology, in their classrooms. The following section provides a brief overview of what ICT integration means.

In 1980 computers were very cheap and easily available for consumers. Governmental policies are going to be change in 21st century these policies motivated the mass production of computers for schools. Many researchers asserted that ICT will be an essential component of the education process for future generations (Bransford *et al.* 2000; Grimus 2000; Yelland 2001). Towards the end of the 1980s, the term 'information technology' began to replace the word 'computer' (Pelgrum and Law 2003). This term referred to the computer's processing ability, indicating a shift from computing technology to the capacity to store and retrieve information. Later, the term ICT emerged, signalling the introduction of e mail and electronic messaging with computer technology (Pelgrum and Law 2003).

It became accepted during this period, that educational systems needed to prepare students to adjust to and survive in this 21st century. This meant preparing students for "lifelong leaning in an information society" (Pelgrum and Law 2003, p.20). ICT integrated education worked as a catalyst for change, fostering 21st century skills in problem solving and critical thinking, as well as the development of student centered learning (McGrail 2005, p.6). Kozma (2008, pp.1085-1087) identified three rationales for the introduction of ICT into education. Firstly, the *economic rationale* refers to the role it can play in preparing students as future workers and in supporting economic development.

Secondly, the *social rationale* where ICT investment aims to: increase knowledge sharing, encourage cultural creativity, increase civic participation, make government services more accessible and finally enhance social cohesion. Finally, the *educational or pedagogic rationale,* where ICT can advance educational reform and improve educational management structures. Similarly, Hepp *et al.* (2004) broadly concur, identifying three reasons for the use of ICT in education: the development of new skills for the information age, increased productivity and the development of quality learning.

Drent and Meelissen (2008), identified three objectives for the integration of ICT in education. They are: the use of ICT as a 'discipline or profession'; ICT as a 'teaching or learning medium' and the use of ICT as an 'object of study' (Drent and Meelissen, 2008, p.187). It can be gleaned from these objectives that integration involves aiding the teaching and learning process (apart from the third objective which is a discipline in itself). Successful integration of ICT in education can lead to a number of benefits and some of these benefits will be explored in the following section.

The use of technology in the 21st century education environment has become an unstoppable. (Cohen *et al.* 2004; Laubsch 2006). ICT impacts on a large section of education, from record keeping and school websites to the creation of online learning communities (Bishop 2007).

21st century institutions do not need students to be physically present. Virtual classrooms with improved internet accessibility provide a great opportunity and facilitation. (Stennes 2008). However, the usefulness of ICT use in subject teaching depends on the success with which it has been integrated (Condie and Munro 2007). Dawes (2001) asserted that new technologies could support education across the entire curriculum, providing innovative opportunities for effective communication. ICT in education has undoubted potential, to be influential in changing teaching methodologies.

Many researches and Studies showed that technology use can result in effective literacy gains with skills development. Many weak reader can be motivated and engaged through the use of ICT (Lynch *et al.* 2000; Ó Murchú 2000; Segers and Verhoeven 2002). Condie and Munro (2007, p.5) concluded that the use of ICT has had positive effects in a number of subjects. Schofield and Verban 1988 (cited in Parr 1995), concluded that using technology Aided Instruction (TAI) diverts the teacher's focus to weaker students.

3. THEORETICAL FRAMEWORK

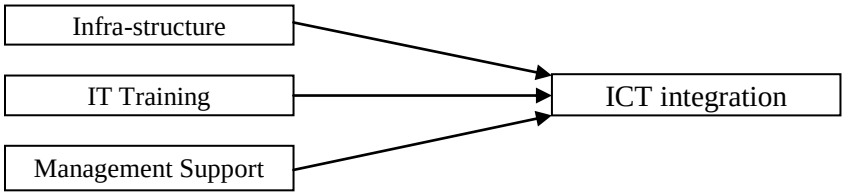


Figure I

4. HYPOTHESES AND DESCRIPTION OF VARIABLE

4.1 Dependent Variable

ICT integration is taken as the dependent variable. ICT integration in education in 21st century environment has become an unstoppable. (Laubsch 2006). ICT integration depends on many factors, so It is difficult to integrate ICT in Education.

4.2 Independent Variables

4.2.1 Infra-structure

Infra-structure is play a vital part in the integration of ICT in Education. A successful ICT integration is impossible without equipment, its maintenance and availability. This variable is very important in Education, in fact a good infra-structure supports an educator to communicate.

H1: There is a positive impact of infra-structure on the integration of ICT in education.

4.2.2 IT Trainings

A teacher should have some knowledge of IT for 21st century learning environment; this variable is taken as an independent variable. More IT trainings develop more active learning environment for learners and lack of training facilities create hurdles in ICT integration.

H2: There is a positive impact of IT Trainings on integration of ICT in Education.

4.2.3 Management Support

A teacher should have support of management; this variable is taken as an independent variable. More support and positive attitude of management and top administration develop more active friendly, problem solving and learning environment for learners and lack of support of management create many problems for educators, learners and management in ICT integration.

H3: There is a positive impact of management support on integration of ICT in Education.

5. DATA DESCRIPTION AND METHODOLOGY

This study is based on 100 teachers responses, who are MTs (Master Trainers) trained by Intel ® Education. The main reason behind using this data is to explore the impact of the independent variable on the dependent variable . This data will enable us to minutely monitor the relationship between these variables. Majority of previous studies used different case studies and other variables but in our opinion our variables like infra-structure, IT trainings and Management support have an immediate impact on ICT integration in education. These variables are important factors for successful integration of ICT in education. Data is taken from 10 different schools of Lyari town due to fact that these 10 schools have trained teachers, infra-structure and management support in different scenario. Hence, this study will enables us to determine very important factors of ICT integration in education. variables.

Data used in this study has been obtained from the questionnaire distributed in 10 schools and each school filled 10 questionnaires by 10 different subject teachers. Linear regression model and correlation analysis is used to analyze the impact of the independent variables.

Descriptive Statistics					
	N	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
	Statistic	Statistic	Statistic	Statistic	Statistic
Grade	100	1.00	3.00	2.4000	.86457
Age	100	1.00	3.00	1.1500	.50000
Gender	100	1.00	2.00	1.7800	.41633
Valid N (listwise)	100				

6. RESULTS AND DISCUSSION

6.1 Reliability Test

The value of Cronbach's Alpha showed that tools are 60 to 75 % reliable. All the items supported the research hypothesis and they support to analyze the data and results. A complete summary of reliability results is reported in table 2.

Table 2: Reliability Test

Variable	**Cronbach's Alpha**	**No of items**
ICT Integration	0.821	08
Infra-structure	0.722	09
ICT Integration	0.551	03
Management support	0.709	03

6.2 Correlation Analysis

The results of correlation analysis shows that ICT Integration is correlated 67.1% with Infra-structure, 55.6% with IT Trainings and 67.6% with Management support. This shows the fact that ICT integration is successful due to the strong relationship with independent variables. A complete summary of correlation analysis is reported in table 3.

Table 3: Correlation Analysis

Correlation	ICT Integration	Infra structure	ICT Training	Management Support
ICT Integration	1			
Infra-structure	.671**	1		
ICT Training	.556**	.585**	1	
Management Support	.676**	.776**	.576**	1

6.3 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square	F Chang	Sig. F Change	Durbin Watson
1	.727[a]	.528	.513	1.80492	.528	35.113	.000	.455

6.3.1 Explanation

The results of model summary are given in the table 4. The dependent variable is ICT integration. R squared is the magnitude of determinants. In our study it has been found that we have R square value of 0.528 which means that 52.8% model has explained parts. Adjusted R squared is 0.513 which is close to R square value. This shows that model and data is fit to use. In our analysis, F value is 35.11 at significant level of .000 less than 1% which shows that independent variables explained variation in the dependent variable very well.

6.4 Regression Analysis

Table 5: Regression results

Hypotheses	Variable	Coef.	t-value	p value	Decision
	Constant		2.299	.024	
H1: Infra-structure has positive impact on ICT integration	IS	.289	2.389	.019	Accepted
H2: IT training has positive impact on ICT integration	ICTT	.186	2.052	.043	Accepted
H3: Management support has positive impact on ICT integration	MS	.342	2.915	.004	Accepted

Dependent Variable: ICT integration

EXPLANATION AND DISCUSSION OF RESULTS

6.4.1 Hypothesis H1

The results of regression show that t- value is 2.389 with positive sign at significant level of .019. As p-value is less than .05, results are significant.

Hence, our first hypothesis i.e. H1: Infra-structure has positive impact on ICT integration is accepted. This study found that the impact of infra-structure is positively significant. There may be many explanations. The more reasonable explanation may be that due to good infra-structure , accessibility of equipments and proper time tabling of computer lab, the integration of ICT can be successful in education.

6.4.2 Hypothesis H2

The results of regression show that t- value is 2.052 at significant level .043. As p-value is less than .05, results are significant. Hence, our second hypothesis i.e. H2: IT training has positive impact on ICT integration is accepted. Teachers training is an important factors to create an e-learning environment. Trainings help educators to develop their skills and self-direction in this era. Our results confirm that more IT trainings are compulsory for successful integration of ICT. Hence, there is a need of quality and continuous trainings of teachers in education sector.

6.4.3 Hypothesis H3

The results of regression show that t- value is 2.915 at significant level .004. As p-value is less than .05, results are significant. Therefore, our third hypothesis i.e. H3: Management support has positive impact on ICT integration is accepted. The results show a positive relationship between the FDI inflow and stock exchange which is consistent with literature on the subjects. This fact shows that support of management always play a role of catalyst and create a positive learning environment in education, in fact an educator always depends on the support of management. In the results, management supports show a strong correlation of 62.64% and confirm that management support is a vital factor in the process of ICT integration.

7. CONCLUSIONS AND FINDINGS

Information and communications technology has become an integral part of our personal and working lives. The students of today have been born into this age of technology. The prevalence of ICT requires educators to ensure that all students are capable of full participation in this digital world. This final chapter outlines the key conclusions reached from the research and offers some recommendations to address the issues raised.

The findings from this study point to a number of barriers which are hindering the integration of ICT in secondary schools. These are a lack of time for teachers to engage with ICT. In addition a lack of training also emerged as a significant barrier. Finally, while teachers in the schools expressed satisfaction with ICT resources, there was a difficulty in accessing such resources.

The results from this study demonstrate that teachers, whether ICT capable or not, are generally positive in relation to ICT's role in education. They are clear about the potential benefits of an ICT permeated education process. A number of teachers in this study

acknowledged the potential of ICT as a teaching resource, while a number of teachers have undertaken training in their own time and at their own expense. The factors that lead to successful ICT integration appear to be infra-structure, IT trainings and management support regarding integrated learning.

Informal in-house training is a worthwhile Endeavour as it brings positive experiences, leading to greater knowledge and competence. Greater knowledge inevitably leads to an increase in confidence and the alleviation of fear and apprehension. Full ICT integration will only be achieved when problems are eliminated or substantially alleviated. There needs to be in place a vision for ICT. This far reaching vision, articulated in a collaboratively designed and regularly reviewed plan, needs to be proactively driven, particularly by management. This drive to achieve integration needs to be supplemented by modern resources, skilled and enthusiastic educators, consistent and reliable support and unhindered access.

This study was carried out within a 10 schools of Karachi and it dealt exclusively with teachers. Students did not form part of this particular study, as it dealt solely with the problems faced by secondary teachers. The findings from such studies might indicate if other schools had similar or entirely different issues, in relation to ICT integration. Such a comparative study could influence and inform any potential strategies, formulated to tackle these problems. Furthermore, the findings might provide a contrast, between the difficulties that private school experience and the difficulties that mainstream secondary schools experience.

REFERENCES

1. Ager, R. (2003). *Information and Communications Technology in Primary Schools* (2nd Ed), David Fulton Publishers, London.
2. Allen, Potter, Sharp & Turvey (2007). *Achieving QTS: Primary ICT. Knowledge, Understanding, Practice* (3rd ed). Learning Matters, Exeter.
3. Barber, D., Cooper, L. and Meeson, G. (2007). *Learning and teaching with Interactive Whiteboards. Primary and early Years*. Learning Matters, Exeter.
4. Bennett, Hammill, Pickford (2006). Progression in Primary ICT. David Fulton Publishers, London.
5. Bennett, R. (2004). *Using ICT in Primary English Teaching*. Learning Matters, Exeter.
6. Bennett, R. (2006). Learning ICT with Maths. David Fulton Publishers, London.
7. Briggs, M. and Pritchard, A. (2002). *Using ICT in Primary Mathematics Teaching.* Learning Matters, Exeter.
8. Buckingham, D. (2003). *Media Education - literacy, learning and contemporary culture.* Polity, Cambridge.
9. Buckingham, D. (2007). *Beyond Technology: Children's learning in the age of digital culture.* Polity Press, Cambridge.
10. Byrne, J. and Sharp, J. (2002). *Using ICT in Primary Science Teaching*. Learning Matters, Exeter.
11. Carrington, V. and Robinson, M. (2009). (Eds.) *Digital Literacies. Social Learning and Classroom Practice.* Sage, London.
12. Duffty, J. (2006). *Extending Knowledge in Practice: Primary ICT*. Learning Matters. Exeter.

13. Easingwood, N. and Williams, J. (2004). *ICT and Primary Mathematics: A Teacher's Guide*. David Fulton Publishes, London.
14. English, R. (2006). *Maths and ICT in The Primary School. A Creative Approach*. David Fulton Publishers, London.
15. Ermann, M.D. and Shauf, M.S. (2003) (3rd Ed). *Computers, Ethics and Society.* OUP, Oxford. *Interactive: Managing ICT in Schools*, Questions Publishing
16. Evans, J. (2004). (Ed.) *Literacy Moves On: Using popular culture, new technologies and critical literacy in the primary classroom.* David Fulton Publishers, Chiswick.
17. Ferrigan, C. (2005). *Passing The ICT Skills Test* (2nd Ed). Learning Matters, Exeter.
18. Fox, B. (2003). *Successful ICT Leadership in Primary Schools.* Learning Matters, Exeter.
19. Fox, B., Montague-Smith, A. and Wilkes, S. (2000). *Using ICT in Primary Mathematics*. Practices and Possibilities. David Fulton Publishers, London.
20. Gage, J. (2005). *How to use an Interactive Whiteboard really effectively in your primary classroom.* David Fulton Publishers, London.
21. Gerver, R. (2010). *Creating Tomorrow's Schools Today*. Continuum, London.
22. Gillespie, H. (2006). *Unlocking Learning and Teaching with ICT. Identifying and overcoming barriers.* David Fulton Publishers. Chiswick.
23. Gillespie, H., Boulton, H., Hramiak, A.J. and Williamson, R. (2007). *Learning and Teaching with Virtual Learning Environments*. Learning Matters, Exeter.
24. Hall, D. (2010). *The ICT Handbook for Primary Teachers - A guide for Students and Professionals*. Routledge, Oxford.
25. Hammill (2006). Learning ICT in the Arts. David Fulton Publishers, London. .
26. Hammill (2006). Learning ICT with Science. David Fulton Publishers, London.
27. Hill, C. (2003). *Teaching using Information and Learning Technology in Further Education.* Learning Matters, Exeter.
28. Jarvis, G. (2003). *Using ICT in Primary Humanities Teaching*. Learning Matters, Exeter.
29. Lankshear, C. and Knobel M. (2006). *New Literacies: Everyday practices & classroom Learning* (2nd Ed). OUP, Maidenhead.
30. Lankshear, C. and Knobel, M. (2003). *New Literacies: Changing Knowledge and Classroom Learning.* Open University Press, Maidenhead.
31. Lankshear, C. and Knobel, M. (2008). (Eds.) *Digital Literacies: Concepts, policies ans Practices*. Peter Lang, New York.
32. Meadows, J. (2004). *Science and ICT in the Primary School*. David Fulton Publishers, London.
33. Monteith, M. (2002). *Teaching Primary Literacy with ICT*. Open University Press, Buckingham.
34. Papert, S. (1990). *Mindstorms: Children, Computers and Powerful Ideas.* Harvester Wheatsheaf, New York.
35. Papert, S. (1993). *The Children's Machine*. Basic Books.
36. Papert, S. (1996). *The Connected Family*. Longstreet Press, Atlanta.
37. Pickford (2006). Learning ICT in the Humanities. David Fulton Publishers, London, Prensky, M. (2006). Don't Bother Me Mom, I'm Learning. How Computer and Video Games Are Preparing Your Kids For Twenty-first century. Paragon House.
38. Pritchard, A. (2004). Learning on the Net: A Practical Guide to Enhance Learning in Primary classrooms. David Fulton Publishers, London.

39. Rudd, A. and Tyldesley, A. (2006). *Literacy and ICT in the Primary School.* David Fulton Publishers, London.
40. Selwyn, N., Potter, J. and Cranmer, S. (2010). *Primary Schools and ICT: Learning from pupil perspectives*. Continuum, London.
41. Sharp, J. and Byrne, J. (2002). *Using ICT in Primary Science Teaching.* Learning Matters, Exeter.
42. Smith, A. and Willcocks, S. (2005). *Creative ICT.* David Fulton Publishers.
43. Somekh, B. and Davis, N. (1997). *Using IT effectively in teaching and learning*. Underwood, J. (ed.) (1994) *Computer based learning: Potential into Practice*. David Fulton Publishers, London.
44. Underwood, J. and Underwood, G. (1990). *Computers and Learning: Helping Children acquire Thinking Skills.* Basil Blackwell, Oxford.
45. Unsworth, L., Thomas, A., Simpson, A. and Asha, J. (2005). *Children's Literature and Computer Based Teaching.* Open University Press, Maidenhead.
46. Useful background reading.
47. Warlick, D. (2005). Classroom Blogging: A teacher's guide to the blogosphere. Lulu.
48. Wegerif, R. and Dawes, L. (2004). *Thinking and Learning with ICT*. Routledge Falmer, London.
49. Wheeler, S. (2005). (Ed.) *Transforming Primary ICT.* Learning Matters, Exeter.
50. Wheeler, S. (2009). (Ed.) Connected Minds, Emerging Cultures. Cybercultures in Online Learning. Information Age Publishing, Inc, Charlotte, NC.
51. Wray, D. (2006). Teaching Literacy Across the Primary Curriculum. Learning Matters Ltd, Exeter.

WORK CITED PAGE

1. *ICT* in *Education*/Key Challenges in *Integrating ICTs* in *Education* ...
   en.wikibooks.org/.../ICT...Education/Key_Challenges_in_Integrating...
2. *ICT Integration in Education* - What do we mean? | GESCI | ICT in ...
   www.gesci.org › Media › Eye on ICT Blog
3. *ICT Integration in Education* - African Virtual University.
   www.avu.org/Courses-Tables-List/ict-integration-education.html
4. *ICT Integration* | VVOB vzw Kenya.
   www.vvob.be/kenya/?q=ictintegration
5. Status of *ICT Integration in Education* in Southeast Asian Countries.
   www.icde.org/Status+of+ICT+Integration+in+Education+in+Southeas..
6. *ICT integration in Education* | Scoop.it.
   www.scoop.it/t/ict-integration-in-education
7. CEMASTEA - *ICT Integration. cemastea.*
   wikispaces.com/ICT+Integration
8. *ICT Integration in Education* - International Symposium - ePals ...
   www.epals.com/forums/t/31404.aspx
9. The MacGyvers of *ICT integration in education* | Singapore ...
   educononline.com/.../the-macgyvers-of-ict-integration-in-education/
10. Conference Models of *ICT Integration in Education*.
    www.ite.educacion.es/congreso/modelostic/index.php?lang=en

11. *Implications for the disciplines of ICT integration in education.*
herkules.oulu.fi/isbn9514272463/html/x1834.html
12. Recent developments in *ICT integration in education*.
www.mdg-review.org/...ict...education-/237-recent-developments-in-...
13. *ICT integration in Education*.
www.slideshare.net/vasudhak2000/ict-integration-in-education
14. *ICT Integration in Education* in West and Central Africa : Studies on ...
www.ernwaca.org/panaf/spip.php?article46
15. Kenya MOEST: *ICT* in *Education* OPTIONS Paper.
ict.aed.org/kenya/ex_summary.htm
16. *ICT Integration in Education* En. oer.avu.org › English
17. 2iE, A Leader in *ICT Integration in Education* in Sub Saharan Africa.
www.2ie-edu.org/index.php?...lintegration...
18. *ICT INTEGRATION* INTO *EDUCATION* - UNESCO Bangkok.
www2.unescobkk.org/education/ict/resources/.../Singapore_ICT.doc
19. *ICT* transforming *education*: a regional guide; 2010.
unesdoc.unesco.org/images/0018/001892/189216e.pdf
20. More videos for *ICT integration in education* »
21. *Education* in Tunisia - Wikipedia, the free encyclopedia.
en.wikipedia.org/wiki/Education_in_Tunisia
22. SEAMEO - *Education* high officials discuss *ICT integration* in 30th ...
www.seameo.org/index.php?option=com_content&task...
23. Capacity Building for *ICT* in *Education* - infoDev.
www.infodev.org/en/Document.888.pdf
24. Conclusions Models of *ICT Integration in Education*.
www.ite.educacion.es/congreso/modelostic/index.php?...
25. *INTEGRATING* TECHNOLOGY IN THE CLASSROOM.
www.pliroforiki.org/joomla/index.php?option=com...gid...
26. *ICT Integration in Education* - Option Mathematics | Certificate/Diploma.
www.avu.org/.../ict-integration-in-education-option-mathematics.htm...
27. Examining (f)actors influencing *ICT integration* Examining (f)actors ...
www.oecd.org/dataoecd/38/44/41713737.pdf
28. *ICT Integration* and Pedagogical Engineering.
www.un-gaid.org/tabid/884/Default.aspx
29. The African *Education* Summit Sommet Africain de l'*Education*.
africanbrains.net/edusa/wp-content/uploads/.../Kenya-Presentation.pd...
30. Distance *Education* for Professional Development in *ICT Integration* ...
www.ijbs

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 433-440

REGIONAL CLOUD COVER CHANGES AND ITS IMPACT ON WINTERTIME TEMPERATURE VARIABILITY USING SINGULAR VALUE DECOMPOSITION ANALYSIS

Syed Muhammad Fahad Riaz[1,2] and **Danish Mumtaz**[2]

[1] Department of Mathematical Sciences, University of Karachi, Karachi, Pakistan. Email: fahad.riaz@nu.edu.pk

[2] Department of Electrical Engineering, National University of Computer & Emerging Sciences (FAST), Karachi, Pakistan. Email: k122414@nu.edu.pk

ABSTRACT

In previous literature emphasis was given on various climatic indices and its possible influences on climate variability over India and Pakistan region. El-Nino Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO) were considered to be the major contributors of climate variability over South Asia. There are also some other high/low semi-permanent pressure systems (Icelandic Low and Siberian High etc) discussed in previous literatures that contribute the climate variability over the region. The region under study has very complex topography. The winters are cold and lasts for a long period in Northern part of the region under study whereas in Southern part the winters are cold but comparatively winter season has short period. The region receives enough rainfall during winters. Cloud cover (or cloudiness) is fraction of sky covered by clouds (usually measured in percentage). In this study we investigate the cloud cover changes over India and Pakistan for the period 1950-2010. We find the associations of the wintertime mean minimum temperature variability with the changes in cloud cover over India and Pakistan region.

The method of Singular Value Decomposition (SVD) is generalization of diagonalization of square symmetric matrix. We have employed SVD analysis to investigate the relation between the coupled fields (cloud cover and wintertime mean minimum temperature). We have then calculated the squared covariance fraction (SCF) values for the dominant modes because it explains covariance of each mode. Correlations have also been used to analyze the possible relation between the changes in cloud cover and wintertime mean minimum temperature. We have shown that the changes in cloud cover have dominant impact on wintertime temperature variability over the region. We used MATLAB software to calculate temporal cross-covariance matrix between the de-trended fields and SVD have been calculated of the temporal cross-covariance matrix.

KEYWORDS

Cloud cover, wintertime temperature, Correlation, Cross-Covariance and Singular Value Decomposition.

1. INTRODUCTION

El-Nino Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO) are considered to be major contributors in climate change of South Asia. Syed et al. (2006) discussed the contribution of the two on precepitation over Central South West Asia. There are also some semi-permanent centers of action (COA) that are considered to be major contributors (Icelandic Low, Siberian High etc.) of the variability of climate change as disscussed by Iqbal et al. (2011) and Iqbal et al. (2012) and dominant impact of South Asia Low on summer monsoon over the region (Iqbal et al., 2011). In this paper we discussed the influence of cloud cover changes in wintertime temperature variability over India & Pakistan (IP) region. There have been several studies of cloud-cover changes between warm and cold periods (Croke, Cess, & Hameed, 1998).

We have used SVD analysis in our investigation of the impact of regional cloud cover changes on wintertime mean minimum temperature. The SVD analysis method identifies pairs of coupled spatial patterns and their temporal variation, with each pair explaining a fraction of the covariance between the two fields. The SVD analysis had been found to give robust results compared with other methods for extracting coupled signals, it has a small systematic bias and is relatively easy to implement (Bretherton et al., 1992).

2. MATERIALS

In our analysis the data used is taken from NCEP/NCAR Reanalysis data that is available from 1948 to 2012. The minimum temperature data (1.9o latitude x 1.875o longitude) is spanned over the period January 1948 to January 2014 and can be seen from http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.html and the total cloud cover data (1.9o latitude x 1.875o longitude) is available from January 1948 to September 2012 and can be seen from http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.html.

3. METHODOLOGY

The method of Singular Value Decomposition (SVD) analysis is employed to observe the possible relationship between the amounts of total cloud cover (measured in percentage) and wintertime (December-January-February-March; DJFM) mean minimum temperature (measured in Kelvin) over India and Pakistan (IP) region for the period 1950 to 2010. Bretherton et al. (1992) and Wallace et al. (1992) popularized the method for observing the patterns in coupled fields. The reader interested in details of SVD analysis is suggested to go through Bretherton et al. (1992). To perform the SVD analysis the temporal cross-covariance matrix between the seasonal (DJFM) amounts of cloud cover and mean minimum temperature for the period 1950 to 2010 for IP region has been calculated. The SVD of the cross-covariance matrix has been calculated which gives the set of two orthogonal singular vectors (called right and left singular vectors) and set of singular values associated with each pair of singular vectors. The squared covariance fraction (SCF) is measured by dividing the squares of each singular value by the sum of all squared singular value. We then calculated the expansion coefficients time series of each mode which is calculated by taking the projection of the original data onto the singular vectors. Correlation coefficients between the expansion coefficients for the leading modes have also been calculated to observe the strength of coupling of two fields.

The high value of correlations does not guarantee that the two fields are strongly coupled. To guarantee that the coupling between the two fields (cloud cover and minimum temperature) is strong we calculated the normalized root-mean squared covariance (RMSC) which is the sum of all squared covariance normalized by diving it by the product of variances of the two fields and is given by:

$$RMSC = \left[\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(\overline{x_i y_j})^2}{(\sum_{i=1}^{M}\overline{x_i}^2)(\sum_{j=1}^{N}\overline{y_j}^2)}\right]^{\frac{1}{2}} \quad (1)$$

The value of RMSC is of order 0.1 or greater than 0.1 for well correlated fields (Wallace et al., 1992).

4. RESULTS & CONCLUSION

The SVD of the temporal cross-covariance between the total cloud cover and mean minimum temperature for the region under study has been calculated. We find the SCF value of the first leading mode to be 76.7%. The total squared covariance explained by the first three leading modes of the coupled cloud cover and minimum temperature variations is 98.1%.

Correlations between the expansion coefficients are then calculated for the first five leading modes. Table 1 depicts the SCF, cumulative SCF and correlations for the five leading modes. The high values of the correlation coefficients between the expansion coefficients time series shows that the two fields are strongly coupled. Although the SCF values for the second, fourth and fifth mode are low but they have high correlation values which show the importance of these modes.

Table 1
First five leading modes. SCF, cumulative SCF and correlation between the expansion coefficients time series

Modes	SCF (%)	CSCF	Correlations
1	76.7	76.7	0.6859
2	19.5	96.2	0.7799
3	1.9	98.1	0.5036
4	0.94	99.04	0.7142
5	0.43	99.47	0.6743

To guarantee the strong relation between the coupled fields we have also calculated the RMSC value. We find the RMSC value to be 0.35 which shows the two fields (cloud cover and wintertime mean minimum temperature) are strongly coupled.

This provides the evidence that the wintertime mean minimum temperature is strongly influenced by the amount of cloud cover over the IP region.

5. COMPOSITE ANALYSIS & PHYSICAL JUSTIFICATION

To give evidence in more details we here also discussed the physical justifications of the association between the regional total cloud cover and wintertime minimum temperature. Composite maps have been plotted using NCEP/NCAR Reanalysis data (see figures 5 & 6).

Figure 5 depicts the air temperature for the years from 1950 to 2010. In figure 6 we have plotted the air temperature anomalies for the years when the cloud cover data averaged over the IP region were highest minus the years when cloud cover data averaged over the IP region were lowest. We see that there is decrease in air temperature when the cloud cover data were highest. Figure 7 shows the amount of total cloud cover for the period 1950 to 2010 over in the decrease of temperature during the season proving evidence that the temperature during winter season (DJFM) is strongly influenced by the cloudiness. the region. The region under study receives enough snow/rainfall during winter seasons (DJFM) which in turn results in the decrease of temperature during the season proving evidence that the temperature during winter season (DJFM) is strongly influenced by the cloudiness.

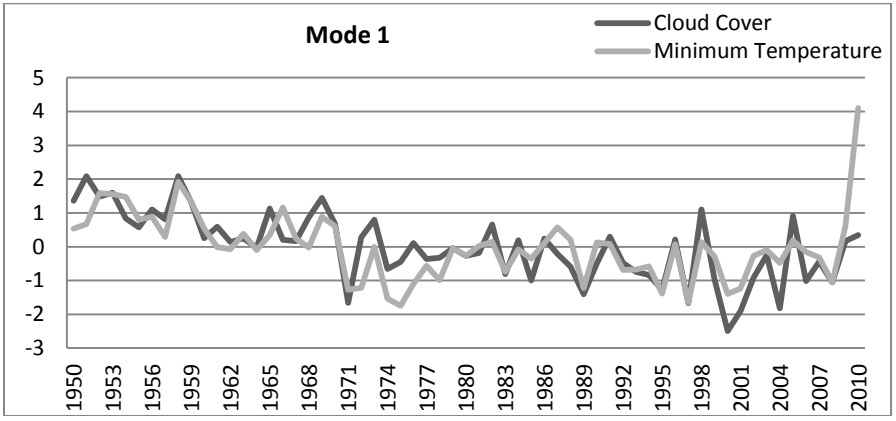


Fig. 1: Squared Covariance Fraction for First Five Leading Modes

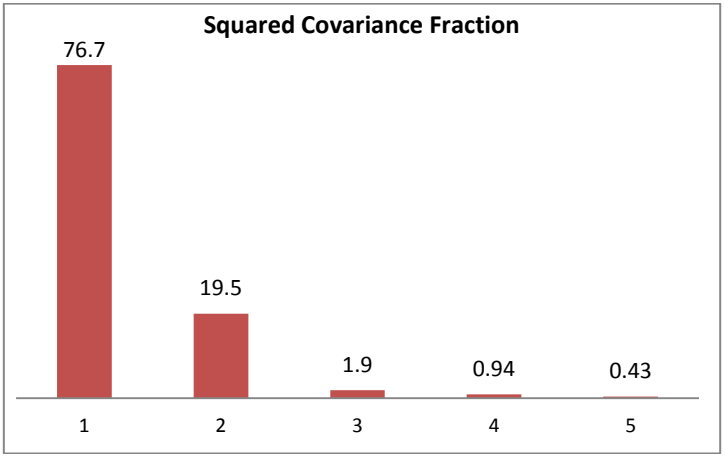


Fig. 2: Comparison between Normalized Expansion Coefficients Time Series of the cloud Cover and Wintertime Minimum Temperature of the First Leading Mode

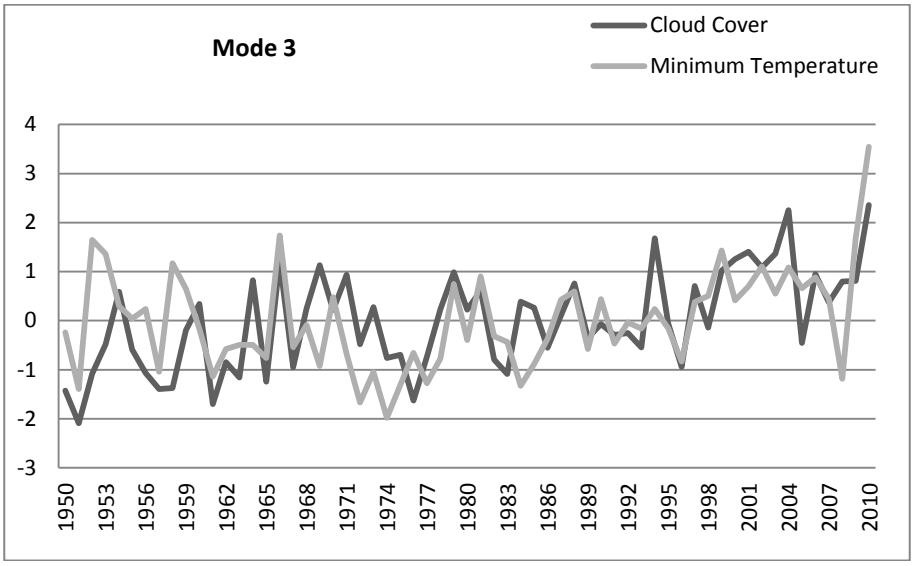


Fig. 3: Comparison between Normalized Expansion Coefficients Time Series of the Cloud Cover and Wintertime Minimum Temperature of the Second Leading Mode.

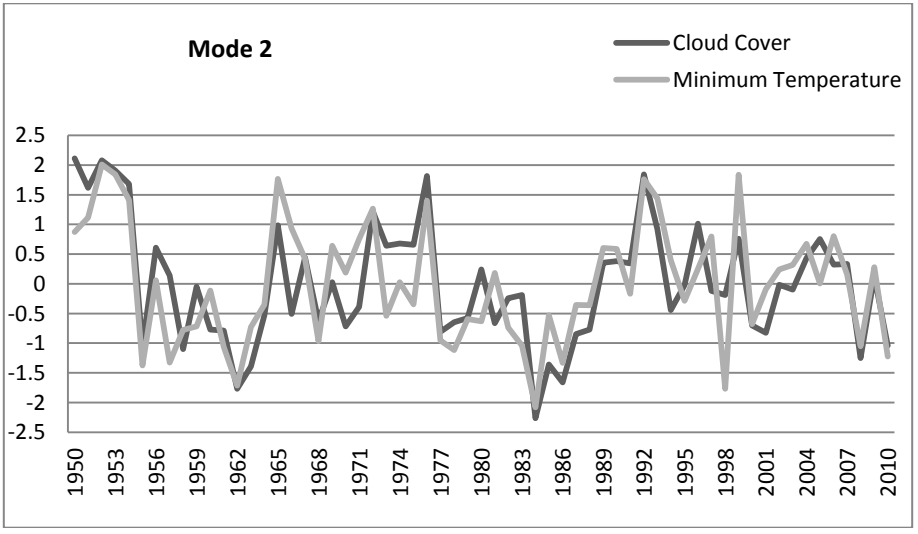


Fig. 4: Comparison between Normalized Expansion Coefficients Time Series of the Cloud Cover and Wintertime Minimum Temperature of the Third Leading Mode.

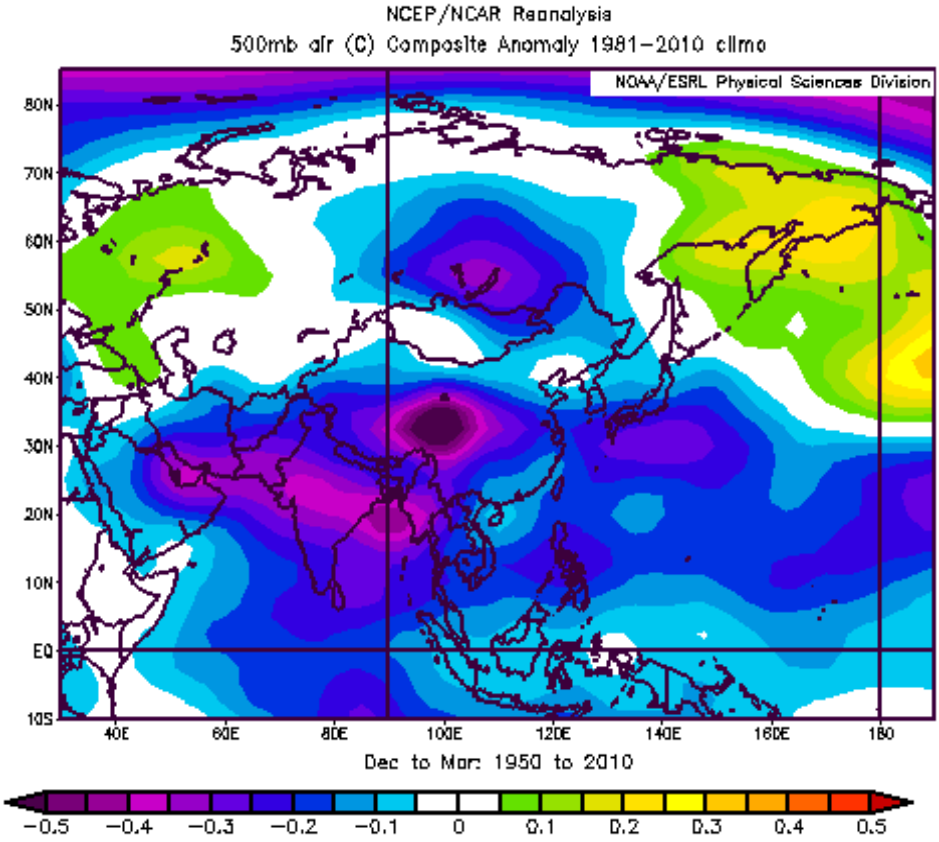


Fig. 5: Air Temperature Anomaly for the Years from 1950 to 2010

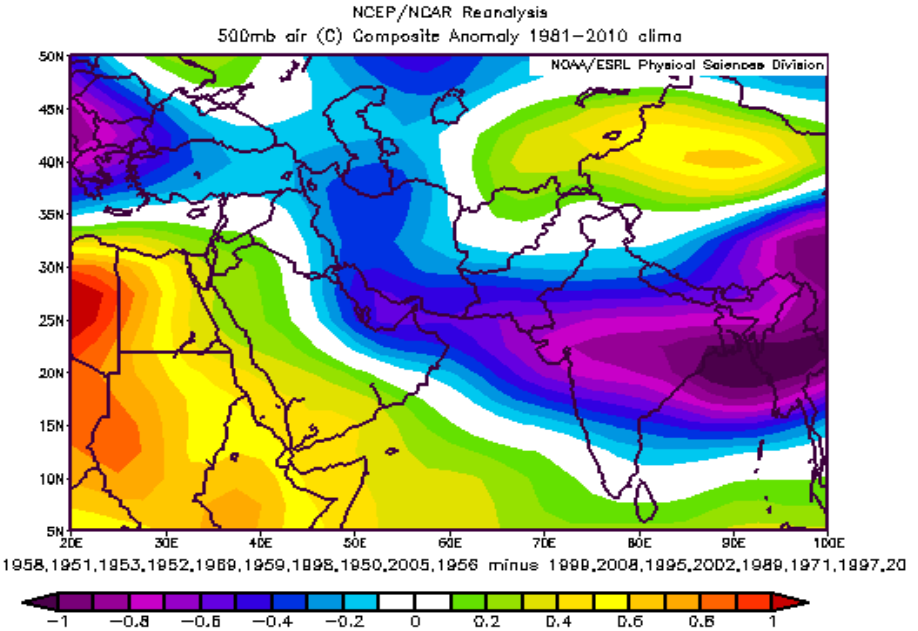


Fig. 6: Air Temperature Anomaly for the Years when Cloud Cover Data Averaged Over the Region was Highest Minus the Years when it was Lowest

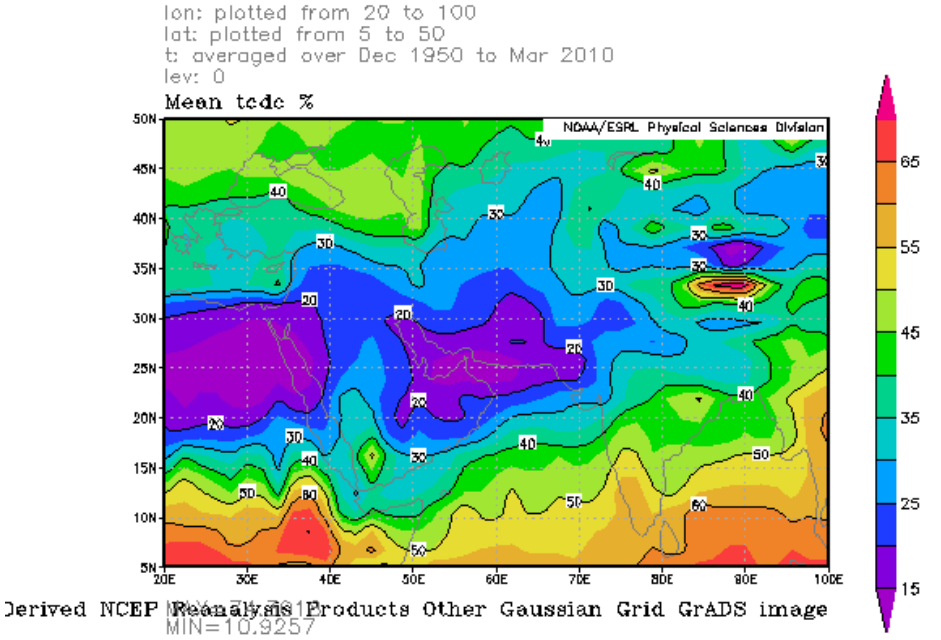


Fig. 7: Total cloud cover (measured in %) from 1950 to 2010 for the season DJFM

REFERENCES

1. Bretherton, C., Smith, C. and Wallace, J. (1992). An intercomparasion of methods for finding coupled patterns in climate data. *Journal of Climate*, 541-560.
2. Cherry, S. (1997). Some Comments on Singular Value Decomposition Analysis. *Journal of Climate*, 10, 1759-1761.
3. Croke, M.S., Cess, R.D. and Hameed, S. (1998). Regional Cloud Cover Change Associated with Global Climate Change: Case Studies for Three Regions of the United States. *Journal of Climate*.
4. Iqbal, M.J. and Ilyas, K. (2011). Influence of Icelandic Low pressure on winter precipitation variability over northern of Indo-Pak region. *Arabian Journal of Geosciences*.
5. Iqbal, M.J., Baig, M.J. and Naz, S. (2011). Dominant impact of South Asian low heat on summer monsoon rainfall over Central India. *Arabian Journal of Geosciences*.
6. Iqbal, M.J., Riaz, S.M. and Ghauri, M.B. (2012). Impact of Siberian High on rainafall variability over Northern part of Indo-Pak region. *Arabian Journal of Geosciences*.
7. Newman, M. and Sardeshmukh, P. (1995). A caveat concerning singular value decomposition. *Jornal of Climate*, 352-360.
8. Syed, F., Giorgi, F., Pal, J. and King, M. (2006). Effect of remote forcings on the winter precipitation of central southwest Asia part 1: observations. *Theoritical and Applied Climatology*, 147-160.
9. Wallace, J., Smith, C. and Bretherton, C. (1992). Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *Journal of Cliamte*, 561-576.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 441-444

COMPUTER AND TRANSPORTS

Irshad Ullah
Department of Computer Science, CMJ University Shilong, India
Irshadullah79@gmail.com

ABSTRACT

As transport is way by which peoples and resources are travelled from one place to the other. So computer also play a vital role in this regard. Here it will be discussed that what are the facilities provided by the computer in this sector.

KEYWORDS

Transport, Sector, Facility, Role, Computer.

1. INTRODUCTION

i. Transport

Transport or transportation is the movement of people, animals and goods from one location to another. Modes of transport include air, rail, road, water, cable, pipeline and space. The field can be divided into infrastructure, vehicles and operations. Transport is important because it enables trade between people, which is essential for the development of civilizations.

Transport infrastructure consists of the fixed installations including roads, railways, airways, waterways, canals and pipelines and terminals such as airports, railway stations, bus stations, warehouses, trucking terminals, refueling depots (including fueling docks and fuel stations) and seaports. Terminals may be used both for interchange of passengers and cargo and for maintenance.

Vehicles traveling on these networks may include automobiles, bicycles, buses, trains, trucks, people, helicopters, watercraft, spacecraft and aircraft. Operations deal with the way the vehicles are operated, and the procedures set for this purpose including financing, legalities and policies.

Although humans are able to walk without infrastructure, the transport can be enhanced through the use of roads, especially when using the human power with vehicles, such as bicycles and inline skates. Human-powered vehicles have also been developed for difficult environments, such as snow and water, by watercraft rowing and skiing; even the air can be entered with human-powered aircraft.

Animal-powered transport is the use of working animals for the movement of people and goods. Humans may ride some of the animals directly, use them as pack animals for carrying goods, or harness them, alone or in teams, to pull sleds or wheeled vehicles.

A fixed-wing aircraft, commonly called airplane, is a heavier-than-air craft where movement of the air in relation to the wings is used to generate lift. The term is used to

distinguish this from rotary-wing aircraft, where the movement of the lift surfaces relative to the air generates lift. Two things necessary for aircraft are air flow over the wings for lift and an area for landing. The majority of aircraft also need an airport with the infrastructure to receive maintenance, restocking, refueling and for the loading and unloading of crew, cargo and passengers. While the vast majority of aircraft land and take off on land, some are capable of takeoff and landing on ice, snow and calm water.

The aircraft is the second fastest method of transport, after the rocket. Commercial jets can reach up to 955 kilometers per hour (593 mph), single-engine aircraft 555 kilometers per hour (345 mph). Aviation is able to quickly transport people and limited amounts of cargo over longer distances, but incur high costs and energy use; for short distances or in inaccessible places helicopters can be used. [14] As of April 28, 2009 The Guardian article notes that, "the WHO estimates that up to 500,000 people are on planes at any time."[17]

Rail transport is where a train runs along a set of two parallel steel rails, known as a railway or railroad. The rails are anchored perpendicular to ties (or sleepers) of timber, concrete or steel, to maintain a consistent distance apart, or gauge. The rails and perpendicular beams are placed on a foundation made of concrete or compressed earth and gravel in a bed of ballast. Alternative methods include monorail and maglev.

A train consists of one or more connected vehicles that run on the rails. Propulsion is commonly provided by a locomotive that hauls a series of unpowered cars that can carry passengers or freight. The locomotive can be powered by steam, diesel or by electricity supplied by trackside systems. Alternatively, some or all the cars can be powered, known as a multiple unit. Also, a train can be powered by horses, cables, gravity, pneumatics and gas turbines. Railed vehicles move with much less friction than rubber tires on paved roads, making trains more energy efficient, though not as efficient as ships.

Intercity trains are long-haul services connecting cities; [18] modern high-speed rail is capable of speeds up to 350 km/h (220 mph), but this requires specially built track. Regional and commuter trains feed cities from suburbs and surrounding areas, while intra-urban transport is performed by high-capacity tramways and rapid transits, often making up the backbone of a city's public transport. Freight trains traditionally used box cars, requiring manual loading and unloading of the cargo. Since the 1960s, container trains have become the dominant solution for general freight, while large quantities of bulk are transported by dedicated trains.

A road is an identifiable route, way or path between two or more places. [3] Roads are typically smoothed, paved, or otherwise prepared to allow easy travel; [12] though they need not be, and historically many roads were simply recognizable routes without any formal construction or maintenance. [13] In urban areas, roads may pass through a city or village and be named as streets, serving a dual function as urban space easement and route. [1]

The most common road vehicle is the automobile; a wheeled passenger vehicle that carries its own motor. Other users of roads include buses, trucks, motorcycles, bicycles and pedestrians. As of 2002, there were 590 million automobiles worldwide. Road transport offers a complete freedom to road users to transfer the vehicle from one lane to

the other and from one road to another according to the need and convenience. This flexibility of changes in location, direction, speed, and timings of travel is not available to other modes of transport. It is possible to provide door to door service only by road transport.

Automobiles offer high flexibility and with low capacity, but are deemed with high energy and area use, and the main source of noise and air pollution in cities; buses allow for more efficient travel at the cost of reduced flexibility.[8] Road transport by truck is often the initial and final stage of freight transport.

Water transport is movement by means of a watercraft—such as a barge, boat, ship or sailboat—over a body of water, such as a sea, ocean, lake, canal or river. The need for buoyancy is common to watercraft, making the hull a dominant aspect of its construction, maintenance and appearance.

In the 19th century the first steam ships were developed, using a steam engine to drive a paddle wheel or propeller to move the ship. The steam was produced in a boiler using wood or coal and fed through a steam external combustion engine. Now most ships have an internal combustion engine using a slightly refined type of petroleum called bunker fuel. Some ships, such as submarines, use nuclear power to produce the steam. Recreational or educational craft still use wind power, while some smaller craft use internal combustion engines to drive one or more propellers, or in the case of jet boats, an inboard water jet. In shallow draft areas, hovercraft is propelled by large pusher-prop fans. (See Marine propulsion.)

Although slow, modern sea transport is a highly efficient method of transporting large quantities of goods. Commercial vessels, nearly 35,000 in number, carried 7.4 billion tons of cargo in 2007. [6] Transport by water is significantly less costly than air transport for transcontinental shipping; [4] short sea shipping and ferries remain viable in coastal areas. [5][10]

2. COMPUTER AND TRANSPORT

Computer is used in all of part of life. It is also used to manage and facilitate the transport system. Now a day's all transport systems used computer in some shape. The tax system of transport is managed by the computer. Also air reservation systems computers are used. Computer is used from water transport to air, roads and railways. The data is gathered for the analysis purpose through observations and questionnaire technique. The data was analyzed and the results produced shows that if the computer systems extend to more parts then it will be more beneficent. Also the maintenance systems must be improved for the better performance.

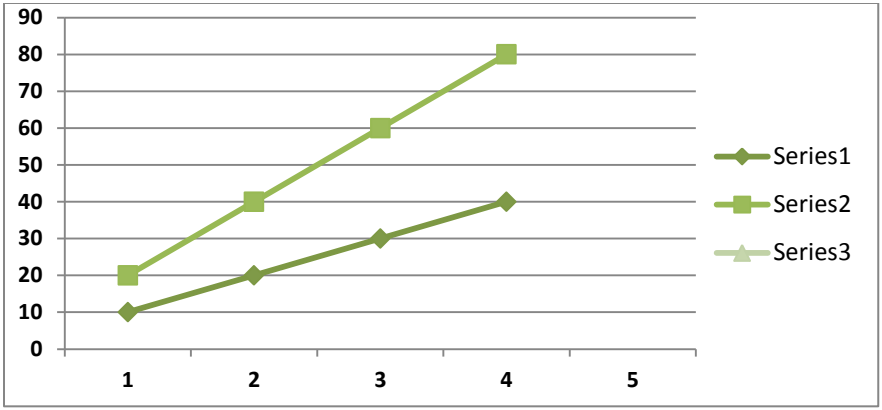


Fig. Experimental Analysis

3. CONCLUSION AND FUTURE WORK

In this research it is studied that how computer works in transport. Its performance also analyzed. Further this study must be extended to obtain better performance.

REFERENCES

1. Bent, F., Mette, K.S.H., and Søren L.B. (2005). How (In)Accurate Are Demand Forecasts in Public Works Projects. *Journal of the American Planning Association* 71(2), 131-146.
2. Chopra and Meindl, 2007: 3, Retrieved 2014.
3. Chopra and Meindl, 2007: 63-64, Retrieved 2014.
4. Cooper et al., 1998: 279, Retrieved 2014.
5. Cooper et al., 1998: 280, Retrieved 2014.
6. Cooper et al., 1998: 275-76, Retrieved 2014.
7. Cooper et al., 1998: 281, Retrieved 2014.
8. Cooper et al., 1998: 278, Retrieved 2014.
9. Lay, 1992: 6-7, Retrieved 2014.
10. *Major Roads of the United States*. United States Department of the Interior. 2006-03-13. Retrieved 24 March 2007.
11. *Road Infrastructure Strategic Framework for South Africa*. National Department of Transport (South Africa). Retrieved 24 March 2007.
12. Stopford, 1997: 4-6, Retrieved 2014.
13. Stopford, 1997: 8-9, Retrieved 2014.
14. Stopford, 1997: 422, Retrieved 2014.
15. Stopford, 1997: 29, Retrieved 2014.
16. *Swine flu prompts EU warning on travel to US*. The Guardian. April 28, 2009.
17. *The United Nations Conference on Trade and Development* (UNCTAD) 2007, p. x and p. 32.
18. *What is the difference between a road and a street*? Word FAQ. Lexico Publishing Group. 2007. Retrieved 24 March 2007.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 445-450

COMPUTER AND STUDENTS MANAGEMENT

Irshad Ullah
Department of Computer Science, CMJ University Shilong, India
Email: irshadullah79@gmail.com

ABSTRACT

Computer systems are very power full and efficient. Students may be managed by analyzing and evaluating student's data. Student's data may be gathered easily with the help of computer systems. The results obtained will be accurate and used for the purpose of student management.

KEYWORDS

Manage; Data; Analysis, System; Purpose.

1. INTRODUCTION

1.1 Computer

A computer is a general purpose device that can be programmed to carry out a set of arithmetic or logical operations automatically. Since a sequence of operations can be readily changed, the computer can solve more than one kind of problem.

Conventionally, a computer consists of at least one processing element, typically a central processing unit (CPU), and some form of memory. The processing element carries out arithmetic and logic operations, and a sequencing and control unit can change the order of operations in response to stored information. Peripheral devices allow information to be retrieved from an external source, and the result of operations saved and retrieved.

In World War II, mechanical analog computers were used for specialized military applications. During this time the first electronic digital computers were developed. Originally they were the size of a large room, consuming as much power as several hundred modern personal computers (PCs).[18]

Modern computers based on integrated circuits are millions to billions of times more capable than the early machines, and occupy a fraction of the space. [13] Simple computers are small enough to fit into mobile devices, and mobile computers can be powered by small batteries. Personal computers in their various forms are icons of the Information Age and are what most people think of as "computers." However, the embedded computers found in many devices from MP3 players to fighter aircraft and from toys to industrial robots are the most numerous.

Etymology

The first use of the word "computer" was recorded in 1613 in a book called "The young mans gleanings" by English writer Richard Braithwait I haue read the truest

computer of Times, and the best Arithmetician that euer breathed, and he reduceth thy dayes into a short number. It referred to a person who carried out calculations, or computations, and the word continued with the same meaning until the middle of the 20th century. From the end of the 19th century the word began to take on its more familiar meaning, a machine that carries out computations.[17]

History

History of computing hardware

Rudimentary calculating devices first appeared in antiquity and mechanical calculating aids were invented in the 17th century. The first recorded use of the word "computer" is also from the 17th century, applied to human computers, people who performed calculations, often as employment. The first computer devices were conceived of in the 19th century, and only emerged in their modern form in the 1940s.

First general-purpose computing device

A portion of Babbage's Difference engine.

Charles Babbage, an English mechanical engineer and polymath, originated the concept of a programmable computer. Considered the "father of the computer",[5] he conceptualized and invented the first mechanical computer in the early 19th century. After working on his revolutionary difference engine, designed to aid in navigational calculations, in 1833 he realized that a much more general design, an Analytical Engine, was possible. The input of programs and data was to be provided to the machine via punched cards, a method being used at the time to direct mechanical looms such as the Jacquard loom. For output, the machine would have a printer, a curve plotter and a bell. The machine would also be able to punch numbers onto cards to be read in later. The Engine incorporated an arithmetic logic unit, control flow in the form of conditional branching and loops, and integrated memory, making it the first design for a general-purpose computer that could be described in modern terms as Turing-complete.[19][29]

The machine was about a century ahead of its time. All the parts for his machine had to be made by hand - this was a major problem for a device with thousands of parts. Eventually, the project was dissolved with the decision of the British Government to cease funding. Babbage's failure to complete the analytical engine can be chiefly attributed to difficulties not only of politics and financing, but also to his desire to develop an increasingly sophisticated computer and to move ahead faster than anyone else could follow. Nevertheless his son, Henry Babbage, completed a simplified version of the analytical engine's computing unit (the mill) in 1888. He gave a successful demonstration of its use in computing tables in 1906.

During the first half of the 20th century, many scientific computing needs were met by increasingly sophisticated analog computers, which used a direct mechanical or electrical model of the problem as a basis for computation. However, these were not programmable and generally lacked the versatility and accuracy of modern digital computers.[3]

The first modern analog computer was a tide-predicting machine, invented by Sir William Thomson in 1872. The differential analyzer, a mechanical analog computer

designed to solve differential equations by integration using wheel-and-disc mechanisms, was conceptualized in 1876 by James Thomson, the brother of the more famous Lord Kelvin.[15]

The art of mechanical analog computing reached its zenith with the differential analyzer, built by H.L. Hazen and Vannevar Bush at MIT starting in 1927. This built on the mechanical integrators of James Thomson and the torque amplifiers invented by H.W. Nieman. A dozen of these devices were built before their obsolescence became obvious.

The modern computer age begins

Alan Turing was the first to conceptualize the modern computer, a device that became known as the Universal Turing machine.

The principle of the modern computer was first described by computer scientist Alan Turing, who set out the idea in his seminal 1936 paper,[9] He also introduced the notion of a 'Universal Machine' (now known as a Universal Turing machine), with the idea that such a machine could perform the tasks of any other machine, or in other words, it is provably capable of computing anything that is computable by executing a program stored on tape, allowing the machine to be programmable. Von Neumann acknowledged that the central concept of the modern computer was due to this paper.[27] Turing machines are to this day a central object of study in theory of computation. Except for the limitations imposed by their finite memory stores, modern computers are said to be Turing-complete, which is to say, they have algorithm execution capability equivalent to a universal Turing machine.

1.2 Students

Pakistan

In Pakistan 12-year school is categorized in three stages: Primary school, Secondary school and Higher Secondary school. It takes 5 years for a student to graduate from Primary school, 3 years for Secondary school and 2 years for Higher Secondary school and 2 years for College/pre-university college. On average Bachelor's degree spans over three years with exception to engineering degree that takes 4 years and 2 years Master's degree.

India

In India also, 12-year school is categorized in three stages: Primary school, Secondary school and Higher Secondary school. In some states like Maharashtra and Andhra Pradesh, 2 year Higher Secondary is Junior College and the Institute is different from regular school. For College its average 3 years for Bachelor's degree (exception B. Tech. and B.E. i.e. 4 years) and 2 years Master's degree.

Iran

An Iranian Student going to her classmate house for study together at evening, in Nishapur.

In Iran 12-year school is categorized in three stages: Elementary school, Junior high school and High school. It takes 5 years for a student to graduate from elementary school, 3 years for junior high school and 3 years for high school and 1 year for pre-university

college(in depended on your field). After graduating from high school(or pre-university if needed), students acquire a diploma. Having a diploma, a student can participate in the Iranian University Entrance Exam or Konkoor. The university entrance exam is conducted every year by National Organization of Education Assessment, an organization under the supervision of the Ministry of Science, Research and Technology.[1] Members of the Baha'i religion, a much-persecuted minority are officially forbidden to attend university, in order to prevent members of the faith becoming doctors, lawyers or other professionals.

Australia

Students of Stony Creek State School, Queensland, 1939.

In Australia, after preschool (which is not compulsory), children begin infants school, usually after the fifth birthday. In New South Wales the first year is called 'kindergarten', in Victoria 'prep', and in South Australia 'reception'; students then continue on to 'year one' through to 'year three', the move to primary school and start 'year four' through to 'year six', except in Western Australia, South Australia and Queensland, where primary school currently continues to 'year seven'. However, by 2014 all states will conclude primary schooling at 'year six'. Students begin secondary school (also known as high school) in 'year seven' (or 'year eight' in WA and Qld)and continue to 'year twelve'. Year seven/eight to nine/ten is referred to as 'middle school' and year ten/eleven through to twelve as 'senior school'. After year twelve, students may pursue tertiary education at university or TAFE (technical and further education). The general term for people in infants, primary, secondary and tertiary education is 'students'.

Germany

In Germany, the German cognate term "student" is reserved for those attending a university. University students in their first year are colloquially called Ersties ("firsties").

2. COMPUTER AND STUDENT'S MANAGEMENT

Students may be managed by the use of computer systems. Students management may be differ. Software's may be used to organize and computerize students data. The examination systems may be computerized to speed up the process. The admission systems through computer secure more than manual systems. In this research data was gathered from different organization mostly in Frontier province of the Pakistan. Results obtained were analyzed shows that computer systems play a key role in students management. The results analysis shows that outcome will be more if advancement made in different fields like medium of instruction, Courses and staff training and a proper respectable Government, Parents, Teachers and students involvement.

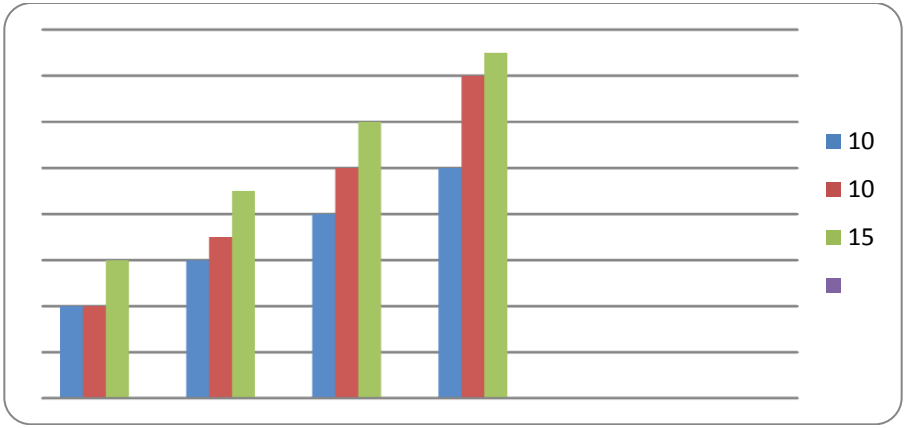


Fig 2.1: Results Analysis

The results show that the management is improved with well arrangement and by increasing the information obtained from the students.

3. CONCLUSION AND FUTURE WORK

This research shows that computer help to manage the students from different sides. In future the research may be strengthening to obtained more results for the betterment of the society.

4. REFERENCES

1. a b c d Copeland, J. (2006). *Colossus: The Secrets of Bletchley Park's Codebreaking Computers*, Oxford: Oxford University Press, 101-115, ISBN 0-19-284055-X.
2. *A Computer Pioneer Rediscovered, 50 Years On*. The New York Times. April 20, 1994.
3. 17 Arthur W. Burks. *The First Electronic Computer*. Retrieved 2014.
4. *Babbage*. Online stuff. Science Museum. 2007-01-19. Retrieved 2014.
5. Bletchley's code-cracking Colossus. *BBC News*, 2 February 2010. Retrieved 2014.
6. Computer Conservation Society, *Our Computer Heritage Pilot Study: Deliveries of Mark I and Mark I Star Computers*. Retrieved 9 January 2010.
7. Computer, N. (1989). *Oxford English Dictionary* (2 ed). Oxford University Press. 1989. Retrieved 2014.
8. *Crash! The Story of IT: Zuse at the Wayback Machine*. (Archived March 18, 2008).
9. ENIAC (1946). *Approximate Desktop & Notebook Power Usage*. University of Pennsylvania. Retrieved 20 June 2009.
10. *Early Electronic Computers (1946-51)*, University of Manchester. Retrieved 16 November 2008.
11. Enticknap, N. (1998). *Computing's Golden Jubilee*, Resurrection (The Computer Conservation Society) (20), ISSN 0958-7403. Retrieved 19 April 2008.
12. *Early Computers at Manchester University*, Resurrection (The Computer Conservation Society) 1(4), Summer 1992, ISSN 0958-7403. Retrieved 7 July 2010.
13. Fensom, J. (2010). *Harry Fensom Obituary*. Retrieved 17 October 2012.

14. Generations of Computers. Retrieved Feb 2014.
15. Girvan, R. (2003). *The Revealed Grace of the Mechanism: Computing after Babbage*, Scientific Computing World.
16. Halacy, D.S. (1970). *Charles Babbage, Father of the Computer*. Crowell-Collier Press. ISBN 0-02-741370-5.
17. Intel Corporation. *Intel Core 2 Duo Mobile Processor: Features*. Retrieved 20 June 2009.
18. January 15, 1941. *Notice in the Des Moines Register*. Retrieved 2014.
19. John Presper Eckert Jr. and John W. Mauchly (1947). *Electronic Numerical Integrator and Computer*, United States Patent Office, US Patent 3,120,606, filed 26 June 1947, issued 4 February 1964, and invalidated 19 October 1973 after court ruling on Honeywell v. Sperry Rand.
20. Jump up to: a b c d "*The Modern History of Computing*". Stanford Encyclopedia of Philosophy.
21. *Let's Build Babbage's Ultimate Mechanical Computer*. Opinion. New Scientist. 23 December 2010. Retrieved 2012-08-01.
22. Lavington, S. "*A Brief History of British Computers: The First 25 Years (1948-1973)*." British Computer Society. Retrieved 10 January 2010.
23. Napper, R.B.E., *Introduction to the Mark 1*, The University of Manchester. Retrieved 4 November 2008.
24. *Proceedings of the London Mathematical Society*.
25. Randell, B.; Fensom, H. and Milne, F.A. (1995). *Obituary: Allen Coombs*, The Independent. Retrieved 18 October 2012.
26. *The Colossus Rebuild*, http://www.tnmoc.org/colossus-rebuild-story
27. Von Neumann ... 1972, quoted in Jack Copeland (2004) The Essential Turing, p22.
28. Zuse, H. "*Part 4: Konrad Zuse's Z1 and Z3 Computers*". The Life and Work of Konrad Zuse. EPE Online. Archived from the original on 2008-06-01. Retrieved 2008-06-17.
29. Zuse, Konrad (2010). [1984], *The Computer – My Life Translated* by McKenna, Patricia and Ross, J. Andrew from: *Der Computer, Mein Lebenswerk* (1984) (in English translated from German), Berlin/Heidelberg: Springer-Verlag, ISBN 978-3-642-08151-4.
30. zen.wikipedia.org/wiki/Computer. Retrieved 2014.
31. Zuse, K. (1993). *Der Computer*. Mein Lebenswerk. (in German) (3rd ed.). Berlin: Springer-Verlag. p. 55. ISBN 978-3-540-56292-4.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 451-456

COMPUTER AND ENTERTAINMENT

Irshad Ullah
Department of Computer Science, CMJ University Shilong, India
Email: irshadullah79@gmail.com

ABSTRACT

Computer is a modern device which is used by the people for different purposes. It is used an any part of life. Different organization using the computer for different purpose. In this study it will be discussed that how computer may be used for the purpose of entertainment. Data will be collected and analyzed. The results produced will be discussed.

KEYWORDS

Entertainment, Data, Computer, Life, Study.

1. INTRODUCTION

Entertainment is a form of activity that holds the attention and interest of an audience, or gives pleasure and delight. It can be an idea or a task, but is more likely to be one of the activities or events that have developed over thousands of years specifically for the purpose of keeping an audience's attention.[30] Although people's attention is held by different things, because individuals have different preferences in entertainment, most forms are recognizable and familiar. Storytelling, music, drama, dance, and different kinds of performance exist in all cultures, were supported in royal courts, developed into sophisticated forms and over time became available to all citizens. The process has been accelerated in modern times by an entertainment industry which records and sells entertainment products. Entertainment evolves and can be adapted to suit any scale, ranging from an individual who chooses a private entertainment from a now enormous array of pre-recorded products; to a banquet adapted for two; to any size or type of party, with appropriate music and dance; to performances intended for thousands; and even for a global audience.

The experience of being entertained has come to be strongly associated with amusement, so that one common understanding of the idea is fun and laughter, although many entertainments have a serious purpose. This may be the case in the various forms of ceremony, celebration, religious festival, or satire for example. Hence, there is the possibility that what appears as entertainment may also be a means of achieving insight or intellectual growth.

An important aspect of entertainment is the audience, which turns a private recreation or leisure activity into entertainment. The audience may have a passive role, as in the case of persons watching a play, opera, television show, or film; or the audience role may be active, as in the case of games, where the participant/audience roles may be routinely

reversed. Entertainment can be public or private, involving formal, scripted performance, as in the case of theatre or concerts; or unscripted and spontaneous, as in the case of children's games. Most forms of entertainment have persisted over many centuries, evolving due to changes in culture, technology, and fashion. Films and video games, for example, although they use newer media, continue to tell stories, present drama, and play music. Festivals devoted to music, film, or dance allow audiences to be entertained over a number of consecutive days.

Some activities that once were considered entertaining, particularly public punishments, have been removed from the public arena. Others, such as fencing or archery, once necessary skills for some, have become serious sports and even professions for the participants, at the same time developing into entertainment with wider appeal for bigger audiences. In the same way, other necessary skills, such as cooking, have developed into performances among professionals, staged as global competitions and then broadcast for entertainment. What is entertainment for one group or individual may be regarded as work by another.

The familiar forms of entertainment have the capacity to cross over different media and have demonstrated a seemingly unlimited potential for creative remix. This has ensured the continuity and longevity of many themes, images, and structures.

Entertainment can be distinguished from other activities such as education and marketing even though they have learned how to use the appeal of entertainment to achieve their different goals. The importance and impact of entertainment is recognized by scholars[17][16] and its increasing sophistication has influenced practices in other fields such as museology.[34][33]

Psychologists say the function of media entertainment is "the attainment of gratification".[28] No other results or measurable benefit are usually expected from it (except perhaps the final score in a sporting entertainment). This is in contrast to education (which is designed with the purpose of developing understanding or helping people to learn) and marketing (which aims to encourage people to purchase commercial products). However, the distinctions become blurred when education seeks to be more "entertaining" and entertainment or marketing seek to be more "educational". Such mixtures are often known by the neologisms "edutainment" or "infotainment". The psychology of entertainment as well as of learning has been applied to all these fields.[23] Some education-entertainment is a serious attempt to combine the best features of the two.[37][18]

An entertainment might go beyond gratification and produce some insight in its audience when it skillfully considers universal philosophical questions such as: "What is the meaning of life?"; "What does it mean to be human?"; "What is the right thing to do?"; or "How do I know what I know?". Questions such as these drive many narratives and dramas, whether they are presented in the form of a story, film, play, poem, book, dance, comic, or game. Dramatic examples include Shakespeare's influential play Hamlet, whose hero articulates these concerns in poetry; and films, such as The Matrix, which explores the nature of knowledge[4] and was released world-wide.[15] Novels give great scope for investigating these themes while they entertain their readers.[22] An example of a creative work that considers philosophical questions so entertainingly that it

has been presented in a very wide range of forms is The Hitchhiker's Guide to the Galaxy. Originally a radio comedy, this story became so popular that it has also appeared as a novel, film, television series, stage show, comic, audio book, LP record, adventure game and online game, its ideas became popular references (see Phrases from The Hitchhiker's Guide to the Galaxy) and has been translated into many languages.[36] Its themes encompass the meaning of life, as well as "the ethics of entertainment, artificial intelligence, multiple worlds, God, and philosophical method".[16]

Imperial and royal courts have provided training grounds and support for professional entertainers, with different cultures using palaces, castles and forts in different ways. In the Maya city states, for example, "spectacles often took place in large plazas in front of palaces; the crowds gathered either there or in designated places from which they could watch at a distance."[3] Court entertainments also crossed cultures. For example, the durbar was introduced to India by the Mughals, and passed onto the British Empire, which then followed Indian tradition: "institutions, titles, customs, ceremonies by which a Maharaja or Nawab were installed ... the exchange of official presents ... the order of precedence", for example, were "all inherited from ... the Emperors of Delhi".[13] In Korea, the "court entertainment dance" was "originally performed in the palace for entertainment at court banquets."[8]

Court entertainment often moved from being associated with the court to more general use among commoners. This was the case with "masked dance-dramas" in Korea, which "originated in conjunction with village shaman rituals and eventually became largely an entertainment form for commoners".[1] Nautch dancers in the Mughal Empire performed in Indian courts and palaces. Another evolution, similar to that from courtly entertainment to common practice, was the transition from religious ritual to secular entertainment, such as happened during the Goryeo dynasty with the Narye festival. Originally "solely religious or ritualistic, a secular component was added at the conclusion".[27] Former courtly entertainments, such as jousting, often also survived in children's games.

In some courts, such as those during the Byzantine Empire, the genders were segregated among the upper classes, so that "at least before the period of the Komnenoi" (1081–1185) men were separated from women at ceremonies where there was entertainment such as receptions and banquets.[19]

Court ceremonies, palace banquets and the spectacles associated with them, have been used not only to entertain but also to demonstrate wealth and power. Such events reinforce the relationship between ruler and ruled; between those with power and those without, serving to "dramatise the differences between ordinary families and that of the ruler".[30] This is the case as much as for traditional courts as it is for contemporary ceremonials, such as the Hong Kong handover ceremony in 1997, at which an array of entertainments (including a banquet, a parade, fireworks, a festival performance and an art spectacle) were put to the service of highlighting a change in political power. Court entertainments were typically performed for royalty and courtiers as well as "for the pleasure of local and visiting dignitaries".[31] Royal courts, such as the Korean one, also supported traditional dances.[31] In Sudan, musical instruments such as the so-called "slit" or "talking" drums, once "part of the court orchestra of a powerful chief", had

multiple purposes: they were used to make music; "speak" at ceremonies; mark community events; send long-distance messages; and call men to hunt or war.[29][7][2]

2. COMPUTER AND ENTERTAINMENT

Computer may be an important environment for entertainment. Its provide the opportunity in modern era to the society. The main thing is that it gives some sort of indoor entertainment. It has a separate world of entertainment for the people of every age.

Some example may include video games, Movies, knowledge games. Such things help to improve and make sharp the mental health of the society. In this research the data was gathered and analyzed from different part of the society. From lower age to old age people. The data was analyzed and practical experiments were performed. The results show a very good improvement in the development of a society. From the results it may be analyzed that to involve individual an entertainment it refresh their mind and also increase the competition power.

3. CONCLUSION AND FUTURE WORK

In future this work may be extended to make it more authentic. This research shows that entertainment is necessary for society. Furthermore that computer entertainment is easier and less time consuming.

4. REFERENCES

1. a b Garland, L. (2006). *Byzantine Women: Varieties of Experience 800–1200*. Aldershot, Hampshire: Ashgate Publishing Limited. pp. 177–8. ISBN 978-0-7546-5737-8.
2. a b Carlsen, S. (2009). *A Splintered History of Wood*. New York: Harper Perennial. p. 170. ISBN 978-0-06-137356-5.
3. a b c Kuhns, R.F. (2005). *Decameron and the Philosophy of Storytelling: Author as Midwife and Pimp*. New York; Chichester West Sussex: Columbia University Press. p. 7. ISBN 0-231-13608-0.
4. a b Van Zile (2001), p. 6. Retrieved 2014.
5. a b Walthall (2008). Retrieved 2014.
6. Allen, C. (1984). *Lives of the Indian Princes*. London: Century Publishing. p. 210. ISBN 0-7126-0910-5.
7. Arthur, M. (2001). *When this bloody war is over: soldiers' songs from the First World War*. London: Piatkus. ISBN 0-7499-2252-4.
8. *British Museum catalogue image of Sudanese slit drum*.
9. Dunkle, R. (2008). *Gladiators: violence and spectacle in ancient Rome*, Harlow, England; New York: Pearson/Longman, ISBN 9781405807395
10. Frost, W. ed. (2011). *Conservation, Education, Entertainment*? Channel View Publication. ISBN 978-1-84541-164-0.
11. Gakhar, S. (2007). *The Influence of Storytelling on Pre-Service Students' Attitudes and Intentions* (MSc Thesis). Iowa State University.

12. Heinrich, A. (2007). Meech, Tony, ed. *Heinrich, Entertainment, propaganda, education: regional theatre in Germany and Britain between 1918 and 1945*. Hatfield, England: University of Hertfordshire Press. ISBN 978-1-902806-74-7.
13. Holbrook, P. (1998). Bevington, David, ed. *The Politics of the Stuart Court Masque*. Cambridge University Press. pp. 42-3. ISBN 0-521-59436-7.
14. *IMDb The Matrix Worldwide Release Dates*. Retrieved 2014.
15. Irwin, W., ed. (2002). *The Matrix and Philosophy*. Peru, Illinois: Carus Publishing Company. p. 196. ISBN 0-8126-9502-X.
16. Joll, N., ed. (2012). *Philosophy and the Hitchhiker's Guide to the Galaxy*. Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan. ISBN 978-0-230-29112-6.
17. Jones, P. (1975). *Philosophy and the Novel*. Oxford, Clarendon. Retrieved 2014.
18. Laing, D., Oliver, P. and Wicke, P. (2003). Horn, David, ed. *Continuum Encyclopedia of Popular Music of the World Part 1 Media, Industry, Society*. Continuum. ISBN 0-8264-6321-5.
19. McDonald Walton (2007), p. 51. Retrieved 2014.
20. McGregor, N. "Episode 94: Sudanese Slit Drum (Transcript)". *History of the World in 100 Objects*. BBC Radio 4/British Museum. Retrieved 6 February 2013.
21. McReynolds, L. (2003). Russia at Play: Leisure Activities at the End of the Tsarist Era. Cornell University. ISBN 0-8014-4027-0.
22. Macleod, S. and Watson, S. (2007). Knell, Simon J., ed. *Museum Revolutions*. Oxon, New York: Routledge. ISBN 0-203-93264-1.
23. Maharg, P. et al (2011). de Freitas, Sara, ed. *Digital Games and Learning*. London, New York: Continuum International Publishing Group. ISBN 978-1-4411-9870-9.
24. McGregor, N. (2010). *A History of the World in 100 Objects*. London: Allen Lane. pp. 613 ff. ISBN 978-1-84614-413-4.
25. Potter, D.S. and Mattingly, D.J. (1999). *Life, Death, and Entertainment in the Roman Empire*. University of Michigan Press. p. 308. ISBN 0-472-10924-3.
26. Roshwald, A. and Stites, R. (2002). *European Culture in the Great War: The Arts, Entertainment and Propaganda, 1914–1918*. Cambridge University Press. ISBN 0-521-57015-8.
27. Sayre, S. and King, C. (2010). *Entertainment and Society: Influences, Impacts, and Innovations*. (Google eBook) (2nd ed.). Oxon, New York: Routledge. ISBN 0-415-99806-9. p. 22.
28. Simpson, M.J. (2005). *The Pocket Essential Hitchhiker's Guide* (2nd ed.). Pocket Essentials. p. 120. ISBN 1-904048-46-3.
29. Singhal, A., Cody, M.J., Rogers, E. and Sabido, M. ed. (2008). *Entertainment-Education and Social Change: History, Research, and Practice*. Taylor & Francis. ISBN 1-4106-0959-6.
30. *The Oxford English Dictionary* (1971). Oxford University Press, Vol 1 pp. 213-4.
31. Walthall, A. ed. (2008). *Servants of the Dynasty: Palace Women in World History. London, England*: University of California Press. ISBN 978-0-520-25443-5. pp. 4-5.
32. Wiseman, D.C. (1977). Medieval Sport: Quest for Survival, Distributed by ERIC Clearinghouse Microfiche.
33. Van Zile (2001), p. 9. Retrieved 2014.
34. Van Zile (2001), p. 69. Retrieved 2014.

35. Van Zile, J. (2001). *Perspectives on Korean Dance*. Middletown, Connecticut: Wesleyan University Press. ISBN 0-8195-6494-X. p. 36.
36. Videos and Games. Shrum, L.J.J. (2012). The Psychology of Entertainment Media (2nd ed.). Routledge. ISBN 978-1-84872-944-5.
37. Vorderer, P. (2006). *Psychology of Entertainment*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc. pp. 367–434. ISBN 0-8058-5238-7.
38. Zillmann, D. and Vorderer, P. (2000). *Media Entertainment – the psychology of its appeal*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc. Taylor & Francis e-library (2009). pp. vii. ISBN 0-8058-3324-2.
39. zn.wikipedia.org/wiki/Entertainment_Computer_System. Retrieved 2014.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 457-

UNIVERSITY BRANDING, ITS EFFECT ON INDUSTRY FOR FRESH GRADUATED STUDENT'S JOB PLACEMENT

M. Ibrahim Shamsi[1], **Syed Shuja Hasan**[2] and **M. Salman Samad**[2]
[1] National University of Computer & Emerging Science - FAST,
Department of Sciences and Humanities, Karachi, Pakistan.
Email: Ibrahim.shamsi@nu.edu.pk
[2] National University of Computer & Emerging Science - FAST,
Department of Computer Sciences, Karachi, Pakistan.
Email: K122172@nu.edu.pk; K122370@nu.edu.pk

ABSTRACT

University branding plays a vital role when it comes to hunting for a job at any firm. The name of University where a person graduates from, matters a lot in hunting for a suitable job placement. A person is only addressed by his or her institution in his complete professional life. Students graduated from a branded University receive more benefits and attention from the society. The benefits are in terms of social image, economic edge, and handsome pay at job, recognition and respect. In short they get all the perks of being a graduate of an elite class university.

Our assumption was that the graduates of branded universities are getting employed earlier as compared to the graduates of other universities. To validate this assumption a survey of 100 respondents was conducted from IT graduates presently working on IT related companies. A second survey was also conducted from 100 IT employers (Human Resource Persons) regarding the criteria and preferences of job offerings to new candidates. The classification of university whether it is branded or not was threefold. First is the Academic Prestige (Recognition by authorities), second is the Academic History, Tradition and Culture, and the third is Public Ethos (A commitment to diversity, social equity and investment for the common good in the state, the nation and around the globe). Job placement was measured through various dimensions; Initial salary offered by the employer, time gap between graduation and first job placement, initial job status and company profile. Results showed that there is a significant effect of university name on job placement. Although universities are striving to get recognize by higher authorities, the role of alumni plays a significant part in job placement.

1. INTRODUCTION

What ***brand*** is all about?

A ***brand,*** term which can be defined as "a name, term, sign, symbol, or design, or combination of them which is intended to identify the goods and services of one seller or group of sellers and to differentiate them from those of competitors. Building a brand driven culture is a lifelong commitment to a mindset and a way of life that takes time, planning and perseverance that produces intangible outputs which include greater customer satisfaction, reduced price sensitivity.

***Branding Education*:**

As a result of increased national and international competition within Higher education, universities (once hesitant to use modern marketing techniques) are now undertaking branded marketing campaigns. This renewed interest in branding and strategy has also triggered researchers to study universities as brands.

Branding Universities has direct influence on job placement. And institutions compete with each other as a ***brand*** and the criteria to win this market war is enrollment of students, and their employment rate once they graduate. The desire of each institution is to grow and increase their market share by enrolling more and more students especially in undergraduate programs, so they eventually brand themselves.

2. LITERATURE REVIEW

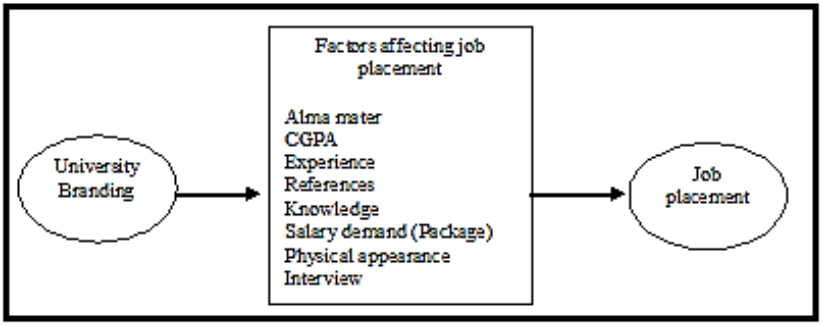
(Aaker, Brand equity in higher education, 1991) told that brand is a name or a platform that distinguishes between the products or services offered by various providers. Though, (Marconi, 1993) expressed that a brand is made to enhance value of the product and give it a character whereas, a name is used to recognize the product. Other researchers too have put up similar descriptions and stated that the buyer must recognize a different appearance and extra value for each brand (Chernatony D. , Brand equity in higher education, 1993) and (Riley, 1999); (Dumas, 1997); (Styles & Ambler, 1996). (McWilliam & De, 1989) tried to hypothesize a brand by presenting two scopes, namely, performance or functionality and personal representation. Then, (Chernatony D. , Brand equity in higher education, 1993) tested this model in both markets and proved that a brand can only be conceptualized by these two dimensions in both kinds of market. Usually, as written by (Chernatony & Harris, 1993), there is a change in the branding works from the emphasis on the idea of brand image, which transmits to the customers' insights of brand diversity, to brand identity which stresses the uniqueness of the brand (Kapferer, 1997). Understanding the concept of brand is the initial phase to understand the concept of brand equity. Brand equity as defined by (Aaker, Brand equity in higher education, 1991): a set of resources such as name recognition, reliable customers, apparent quality, and association that is linked to the brand and increases value of the product or service offered. (Keller, Brand equity in higher education, 1993) on the contrary, defines it as the outcome of the brand name on the customers response to the advertising related to a particular product. It is shown from the above text that "brand equity is a multi-dimensional concept"((McDonald & De, 1998) and can be studied from a lot of different viewpoints, comprising the financial markets, consumers, the company, the staff and the way of interaction (al & et, 2003); (al V. e., 2002) (al S. e., 2007). The description of brand equity from the financial point of view explains the brand the name as an asset having worth for a company due to its ability to create increase earnings (Weitz, 1988);(McDonald, 1989); (al K. e., 2003).

For a customer, brand equity signifies a better performance, lower risk and a good image of the product. Consumer-based brand equity demonstrations the added value of the brand to the consumer ((Farquhar, 1989) and can be described as "The general value that a buyer relates with the usage of the brand", From an organization's position, brand equity embodies qualities such as lesser financial risk, more cash, higher rent, lower marketing cost ((Mooij, 1993). Moreover, a brand can establish stronger customer

loyalty, less price bound, improved marketing efficiency, openings for licensing contracts as well as company extensions and a therefore, good competitive position (Keller, Brand equity in higher education, 2001). Lastly, employee-based brand equity (EBBE) is completely another dimension of brand equity which emphasizes on the employees' opinions toward the brand. EBBE imitates "uniqueness of a company, brand consistency, brand name and brand transparency" (Supornpraditchai, 2007). If we look at the branding impact on service industry we may find out some results similar to out topic. (Muller, 1998) Inspected the brand equity in service industry and found out that the standard of the product, product delivery are proven to be the core points in brand equity. Higher Education embody the idea of brand image which plays a vital role in judgment of quality as said by (Byron, 1995), (Ekwulugo, 2003).So it can be perceived that a big brand is critical as a prospect reliever which helps in reducing the decision-making process (Swait, 1998), (Chen, 2008) (Aaker, Branding education, 1997) Aaker defines branding impact on an individual as the set of individual's characteristics which seems to be connected with the university to see they are connected. (Aaker, Branding education, 1997) Believes that the ones who are branded properly possess five personality dimensions and these are the qualities which lead to differ them in the market with others. The Environment Of university also play a vital role in branding of an individual as par as (Rhoades, 1998) proposed, that faculty have become "managed professionals," where as the level administrators have become "managerial professionals". But According to (Clark, 1996) The Educational institutions in common expand, they tend to flourish into categories of Knowledge which leads to determine that most of the knowledge which exist are correct Educational structures are in fact a theory of knowledge, in which they help define what currently counts as knowledge. A brand position is part of the brand identity and value proposition that is to be actively communicated to the target audience and that demonstrates and advantage over competing brands (Aaker, Brand equity in higher education, 1991). The thing which matters the most about is the Brand Equity, what is does it displays the end user's view on the brand. (Aaker, Brand equity in higher education, 1991) was the first one to propose brand model Equity, what he did is he pinpointed various dimensions of brand equity they are brand name awareness, brand perceived quality, brand loyalty, brand associations and the last one is brand proprietary.

3. MODEL FRAMEWORK

The name of the University of a Candidate has become an important issue while applying for a job. Employers usually seek a candidate from a branded university. To prove this we conducted a survey both from employees and employers in which we asked about them to rate the factors affecting job placement. These factors are the name of the university from where the candidate graduated (alma mater), CGPA of the candidate, how much experience earned, contacts and references of the candidate in the firm where applied for job, how much knowledge he/she has, salary demanded by the candidate, the physical appearance and personality of the candidate and lastly the interview. The figure shown below describes a theoretical model framework of our study.



4. RESEARCH HYPOTHESES

Based on the earlier research studies and the theoretical explanation the hypothesis came into being that the impact of university branding is positively related with job placement in current era. Or it can be said that the job placement greatly depends on the university the candidates graduated from. The attributes of job placement were measured through the factors which were discussed in the model framework that is (1) Alma mater, (2) CGPA, (3) Experience, (4) References, (5) Knowledge, (6) Salary demand (Package), (7) Physical appearance and (8) Interview.

We proposed the following research hypotheses on the basis of those attributes.

- **H_1**: Alma mater is considered most by the more experienced employer than less experienced one while offering a job to a candidate.
- **H_2**: CGPA is considered most by the more experienced employer than less experienced one while offering a job to a candidate.
- **H_3**: Experience is considered most by the more experienced employer than less experienced one while offering a job to a candidate.
- **H_4**: Reference is considered most by the more experienced employer than less experienced one while offering a job to a candidate.
- **H_5**: Knowledge is considered most by the more experienced employer than less experienced one while offering a job to a candidate.
- **H_6**: Salary demanded is considered most by the more experienced employer than less experienced one while offering a job to a candidate.
- **H_7**: Appearance is considered most by the more experienced employer than less experienced one while offering a job to a candidate.
- **H_8**: Interview is considered most by the more experienced employer than less experienced one while offering a job to a candidate.

5. RESEARCH METHODOLOGY

Two questionnaires were constructed to survey the impact of university branding on job placement, one questionnaire for employer and one for employee. The questionnaire for employer had two sections. The first section was about

demographics variables of the respondent; those were gender and age group. The second was about the preferences of employers about the factors affecting job placement. The questionnaire for employee had three sections. The first section was about demographics variables of the respondent; those were gender, age group and CGPA of the candidate. The second section was about the job placement and salary of the employee. The third was about the preferences of employers (from employee point of view) about the factors affecting job placement.

The sampling frame was people of Karachi. We targeted 200 people including 100 employees and 100 employers working in different industries. Positive response was received from 103 employees and 57 employers altogether 160 were responded positively. The survey was done through one to one meetings; therefore, we can say that the response rate was high. Data was analyzed through the statistical package SPSS. There were three demographic variables Age and Gender, including CGPA for employee questionnaire. There were 8 factors which were asked in questionnaire so these 8 factors were treated differently for employer and employee. The factors for employers were treated in a way that we differentiated the employers in two ways that is one those who have less experience and those who have more experience. The mean of the work experience of two employers is 17.95, so the employers who have experience above 17.95 had more experience and they were taken into consideration. The descriptive statistics of all demographics are shown in table 1 and table 2. The research variables (8 factors) were analyzed through T- Test of Independence. The descriptive statistics of all eight factors and result of T- test is shown in the table 5.

Table 1: Demographic information of employers

		FREQUENCY	PERCENTAGE
GENDER	Female	2	3.5
	Male	55	96.5
AGE GROUP	20-29	2	3.5
	30-39	20	35.1
	40-49	21	36.8
	Above 49	14	24.6

Table 2: Demographic information of employees

		FREQUENCY	PERCENTAGE
GENDER	Male	72	69.9
	Female	31	30.1
AGE GROUP	20-29	79	76.7
	30-39	13	12.6
	40-49	5	4.9
	Above 49	6	5.8
CGPA	2.00-2.49	7	6.8
	2.50-2.99	28	27.2
	3.00-3.49	53	51.5
	3.50-4.00	15	14.6

6. RESULTS AND DISCUSSION

To test the reliability of the questionnaire all variables were analyzed using SPSS. The survey was conducted from 160 respondents in which 57 were employers and 103 employees. Out of 57 employers 55 were male. And out of 103 employees 72 were male. The statistic for employer were divided into two one with more experience and other less experience. The minimum experienced employer was 4 years and the maximum was 40 years. The average experience was 1.95 years as shown table 3.

Table 3: Experience (in years) of the employer

Mean	Standard Deviation	Minimum	Maximum
17.95	8.651	4	40

Then we took the mean of all the 8 factors that were collected from the employers and numbered them according to highest mean. From table 4 we got to a point that the employers give priority to the factors as following: 1) Interview, 2) Knowledge, 3) Alma mater that is the name of the university is in 3rd place, 4) Experience, 5) CGPA, 6) Salary demanded, 7) Appearance and lastly 8) Reference.

Table 4: Descriptive Statistics of the factors that employer feel to be considered most while offering job for a candidate (from employers point of view)

Factors	Mean	Standard
Alma mater[3]	3.8772	1.16604
CGPA[5]	3.6667	0.95119
Experience[4]	3.7018	0.94425
References[8]	2.5789	1.10109
Knowledge[2]	4.1228	1.01893
Salary Demand Rate[6]	3.6316	0.99340
Appearance[7]	3.4912	1.01955
Interviews[1]	4.2807	0.77354

The T-Test was applied on the employers data to get the statistics about a particular set of employers that is we targeted the employers with more experience and from the T test we get the results that alma mater is insignificant factor as it is considered by both less experienced and more experienced employers. Same are the statistics about CGPA, references, knowledge and the factor of salary demanded. But this test also shows that the factors which more experience employers require are experience and appearance and 10% of the more experienced employers seek for candidates who have good interview skills. The descriptive statistics of all eight factors and result of T-test is shown in the table 5 below.

Table 5: Independent Sample T-test factors v/s Experience of that employer

Factors	t	df	Sig-value	95% Confidence Interval		Comments
				lower	upper	
Alma mater	1.323	55	0.191	-0.21677	1.05803	Insignificant
CGPA	-0.864	55	0.391	-0.75080	0.29842	Insignificant
Experience	2.515	55	0.015	-0.12648	1.11955	Significant
References	-0.535	55	0.515	-0.77251	0.44711	Insignificant
Knowledge	-0.155	55	0.878	-0.60930	0.52200	Insignificant
Salary Demanded	0.622	55	0.536	-0.72030	0.37903	Insignificant
Appearance	-2.023	55	0.048	-1.09790	-0.00527	Significant
Interviews	-1.770	55	0.082	-0.78683	0.04873	Significant 10%

The same procedure was carried out with the employees' data as the employers'. We took the mean of all the 8 factors that were collected from the employees' and numbered them according to highest mean. From table 6 we got to a surprising conclusion that the priorities and opinions of the employees are same as that of the employers. And also there was no difference in opinion of the employees of branded universities and unbranded universities that is the statistics of the data were same for both employees from branded and unbranded universities.

Table 6: Descriptive statistics of the factors to be considered most by the employer (from employee point of view)

Factors	Mean	Standard
Alma mater[3]	3.8554	1.1954
CGPA[5]	3.5922	1.1327
Experience[4]	3.7282	1.1563
References[8]	2.6990	1.2821
Knowledge[2]	4.3592	0.8147
Salary Demand Rate[7]	3.5631	0.9967
Appearance[6]	3.5825	1.27974
Interviews[1]	4.4367	0.84795

7. COMMENTS AND CONCLUSION

This study primarily focused on the different factors affecting the job placement for candidates in different firms. This study also focuses on the importance of different factors that should be kept in mind by the candidate while applying for a job. Job placement was measured through various dimensions; Initial salary offered by the employer, time gap between graduation and first job placement, initial job status and company profile. Based on the results and finding it can be concluded that the top 3 factors that we found were 1) Interview, 2) Knowledge and 3) Alma mater. The candidates should focus on these factors as they were the opinions of the employers. Results showed that there is a significant effect of university name on job placement. Although universities are striving to get recognize by higher authorities, the role of alumni plays a significant part in job placement.

REFERENCES

1. Aaker (1996). Measuring brand equity across products and markets. *California Management Review*, 102-20.
2. Byron (1995). Brand equity and market-based assets of professional service firms. *Journal of Professional Services Marketing*, 3-7.
3. Chen (2008). Internationalization or international marketing? Two frameworks forunder standing international students' choice of Canadian universities. *Journal of Marketing for Higher Education*, 1-33.
4. Chernatony, D. (1993). New direction – understanding the dynamic of brands as molecules. *Admap*, 21-23.
5. Ekwulugo, B. (2003). International marketing of British education: research on the students' perception and the UK market penetration. *Marketing Intelligence & Planning*, 318-327.
6. Keller (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing*, 1-22.
7. Kim (2006b). Effects of supply chain management practices, integration and competition capability on performance. *Supply Chain Management: An International Journal*, 11(3), 241-8.
8. Kim (2006a). The effect of supply chain integration on the alignment between corporate competitive capability and supply chain operational capability. *International Journal of Operations & Production Management*, 26(10), 1084-1107.
9. Marconi (1993). *Beyond Branding: How Savvy Marketers Build Brand Equity to Create*. Probus Publishing Company, Chicago, IL, 404-13.
10. McDonald, D.C. (1998). *Creating Powerful Brands in Consumer, Services and Industrial Markets*, 2nd ed. Butterworth-Heinemann, Oxford.
11. Styles, and Ambler (1996). Brand development versus new product development: towards a. *Marketing Intelligence & Planning*, 10-20.
12. Supornpraditchai (2007). *Brand equity in higher education Aaker*. (1991). Building Strong Brands. The Free Press, New York, NY.
13. Weitz, S. (1988). *A perspective on brand equity principles and issues*. Marketing Science Institute, Cambridge, MA.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 465-474

BRAND EQUITY TRENDS AMONG TOP AUTOMOBILE BRANDS

Kamran Siddiqui and **Maliha Khaqan**
DHA Suffa University, Karachi, Pakistan
Email: kamran_siddiqui@hotmail.com
maliha.khaqan@gmail.com

ABSTRACT

This paper aims to present the region-wise brand equity trends among top automobile brands during the last decade. The research uses the data published by Interbrand. During 2001-2013, a total of 12 automobile brands from five countries appeared in the top 100 global brands list belonged to only three sub-continents i.e., America, Asia and Europe. Analyses were made on the basis of cumulative brand equity, average brand equity and growth patterns. Some major trends presented in this paper are; (a) global economic recession effects of the automobile sector; (b) Relegation of American auto brands; (c) Dominance of European and Asian auto brands; (d) Clusters of auto brands among top 100 global brands.

INTRODUCTION

Brand equity measurement has been identified as one of the significant topics for future research in brand management (Keller & Lehman, 2006). Much has been written about brand equity measurement but most of literature on this issue is still evolving and it might evolve further (Berthon, Capon, Hulbert, Murgolo-Poore, Pitt, & Keating, 2001). Interbrand's Brand Valuation methodology has been considered as the most prominent method for brand equity measurement. Their estimates are based on the basis of projected branded earnings / profits discounted to a present value amount (Perrier, 1997). Based on their valuation they have generated the annual ranking of the world's top brands in conjunction with Business Week (Chu & Keh, 2006). This annual brand equity ranking is well accepted by both academicians and corporate users and emerged as a reliable benchmark for brand equity measurements (Ward, & Perrier, 1998). The automobile market requires a global brand equity index which provides practical applications for benchmarking (Hsieh, 2004). Brand equity ranking is an important element in contemporary marketing literature and different brand equity models should provide bases for developing marketing strategies in automobile brands also (Siddiqui, 2011). Top 100 global brands ranking list by Interbrand requires four major criteria to be met; (a) the brand has to be solely global with at least one-third of earnings comes from outside of its parent country; (b) the brand must be a market-facing brand and must not have a purely monopsonic condition with no wider awareness; (c) the parent firm of the brand must be a publicly held and there must be substantial publicly available financial data; (d) the Economic Value Added (EVA) must be positive and overall brand value must be greater than US $ 1 billion (Interbrand, 2011).

METHODOLOGY

The data has been taken from the top 100 global brands list published by Interbrand from 2001 to 2013. During 2001-2013, a total of 12 automobile brands from five countries (Table 2) appeared in the top 100 global brands list belonged to only three sub-continents i.e., America, Asia and Europe (Table 1); no brand from Australia and Africa was qualified for the list. The following brands appeared in Top 100 global brands over the last 13 years.

ANALYSES

Analyses were made on the basis of Cumulative Brand Equity (CBE), number of brands, Average Brand Equity (ABE) and Growth Patterns in Cumulative Brand Equity for five countries and/or three regions. Cumulative Brand Equity (CBE) represents summation of brand equities for all brands in a particular country while ABE represents CBE divided by number of brands in a particular country. Interbrand's 13-year longitudinal data provides enormous opportunities for researchers to analyze short and long trends in it. Major trends among top automobile brands are as follows:

Effects of Global Economic Recession: Automotive sector was the worst effected by global economic recession 2009-10. All the global brands have gone through rough time during 2008-09 (Figure 1). However all auto brands have responded proactively to this challenge (Figure 1). Cumulative brand equity of auto sector declined from $138.9 billion in 2008 to $124.7 billion in 2010 and again regained the level of $156.56 billion in 2012 (Figure 1). The rising trend is clearly visible all throughout 2013.

Relegation of American Auto Brands: The regional trends have shown slight disparity and in places have been less encouraging. American brands seem to be silent in this sector. They were represented only by Ford, consistently appeared in Top 100 global brands list (Figure 2). The most recent American brand to be added is Chevrolet to the toppers' list. The relegation of American Auto brands can be seen as in 2001 its brand equity stood at $30.0 Billion and declined to $7.1 Billion in 2010. During last two years a revival of Ford's brand equity has been seen as it has been increased to $7.9 Billion in 2012 (Figure 3). The slight yet encouraging pick up of the American equity trend in 2013, may be attributed to the recent addition of Chevrolet.

Dominance of European & Asian Auto Brands: European brands are always dominant in this sector posting an increase over the previous year. German auto brands are consistent in their brand equity practices; CBE for German autos rise from $42.9 Billion in 2001 to $80.7 Billion in 2012. Asian brands and European brands were very competitive in 2007 with cumulative brand equity of $60.9 Billion and $60.7 Billion respectively. But it seems that Asian brands were unable to cope with the recession period and their cumulative brand equity had declined further to $49.7 Billion in 2010. Toyota seemed to be the major contributor to it as its brand equity decreased by $5.1 Billion in 2010 and Lexus vanished. Asian region has few strong brands in auto sector. Slight impetus has also been provided by Korean brands. Europe's cumulative trend has surpassed all others. In 2012 and 2013, there is clear dominance of the European brands. Germany posted a robust growth and is largely and perhaps singularly responsible for pulling up the European trend in 2013. However, interestingly, Asian region's average brand equity has always been greater as compared to other regions. In 2010, the average brand equity American Automotive brands were $5.2 Billion, European Automotive brands are $11.3 Billion and Asian Automotive brands are

$16.6 Billion with the trend extrapolating. This reflects that Asian Automotive brands are the strongest among all other brands from America and Europe.

Clusters of Auto Brands among Top 100 global brands: In the Interbrand's top 100 global brands list (2010) numerous clusters of auto brands can be seen (Figure 5). First cluster with Brand Equity more than $26 billion holds Toyota as the only member in this cluster with an annual growth rate of -16.4%. It's a unique case with a huge brand equity and negative growth rate. Second cluster holds Mercedes ($ 25.5 billion equity and 5.5% growth rate), BMW ($ 22.3 billion equity and 3.0% growth rate) and Honda ($18.5 billion brand equity and 3.95% growth rate). All three brands have $15 - $26 billion equity and moderate growth rate of 3 - 6% per year. Third cluster include Volkswagon ($6.9 billion brand equity and 6.3% growth rate), Ford has clearly lost its hold ($7.2 billion brand equity and 2.7% growth rate). Fourth cluster posses Audi ($18.5 billion brand equity and 3.95% growth rate) and Hyundai. Both Auto brands posses above $5 billion equity and above 9% growth rate in brand equity. Fifth cluster hold Ferrari and Porsche, having less than $5 billion equity and less than 5% growth rate. In the Interbrand's top 100 global brands list (2012) auto industry sector dynamics changed. First cluster with Brand Equity holds Toyota ($ 27.64 Billion Equity and 6.0% growth rate), Mercedes ($ 27.4 billion equity and 9.0% growth rate), BMW ($ 24.4 billion equity and 10.0% growth rate) and Honda ($19.4 billion brand equity and 5.0% growth rate). All four brands have $19 - $28 billion equity and moderate growth rate of 5 - 10% per year. Second cluster include Volkswagon ($7.8 billion brand equity and 14.0% growth rate), Audi ($6.7 billion brand equity and 13.0% growth rate) and Hyundai ($6.0 billion brand equity and 19.0% growth rate). Third Cluster includes expensive brands like Porche ($4.5 billion brand equity and 4.0% growth rate), Ferrari ($3.5 billion brand equity and 0.8% growth rate) and only American Brand Ford ($7.8 billion brand equity and 4.0% growth rate). Brands like Nissan have not been too consistent. Chevrolet is too recent an addition to make a significant impact on a persistent trend.

CONCLUSION

The cumulative brand equity of European Brands is the highest in the sector, while it is the Asian Brands that surprisingly have much higher average brand equity. Equally noteworthy is the elimination of the myth about the dominance of American brands that have slid consistently on the graph. Moreover, it is the unique case of the Asian Brand Toyota that rides the brand equity curve yet poses a negative growth rate, a phenomenon that challenges the brand equity measurement model. The clusters show significant patterns as well as highlight gaps in the market. The analysis of brand equity trends with respect to growth rate requires further in depth study.

BIBLIOGRAPHY

1. Ambler, T. (1995). Brand equity as a relational concept. *Journal of Brand Management*, 2(6), 386-397.
2. Arvidsson, A. (2006). Brand value. *Journal of Brand management*, 13(3), 188-192.
3. Berthon, J.P., Capon, N., Hulbert, J., Murgolo-Poore, N.J., Pitt, L. and Keating, S. (2001). *Organizational and Customer Perspectives on Brand Equity: Issues for Managers and Researchers.* Auckland: ANZMAC, Massey University.
4. Chu, S. and Keh, H.T. (2006). Brand value creation: Analysis of the Interbrand-Business Week brand value rankings. *Market Letters*, 17, 323-331.

5. Erik, B. Nes, E.B. and Gripsrud, G. (2010). Country of origin image and brand image congruence: The impact on brand equity. *9th Iternational Marketing Trends Congress*, Venice, January 21-23.
6. Hsieh, M. (2004). Measuring Global Brand Equity Using Cross-National Survey Data. *Journal of International Marketing*, 12(2), 28-57.
7. Ille, F.R. (2009). Building Chinese global brands through soft technology transfer, *Journal of Chinese Economic and Foreign Trade Studies*, 2(1), 47-61
8. Interbrand (2011). *Best global brands*, Interbrand Group, London, UK. available at: http://www.interbrand.com/en/best-global-brands
9. Interbrand (2011). *Best global brands: Methodology*, Interbrand Group, London, UK. Available at: http://www.interbrand.com/en/best-global-brands/best-global-brands-methodology/Overview.aspx
10. Kapferer, J. and Bastien, V. (2009). The specificity of luxury Management: Turning marketing upside down. *Journal of Brand Management*, 16(5/6), 311-322.
11. Keller, K.L. and Lehmann, D.R. (2006). Brands and branding: research findings and future priorities. *Marketing Science*, 25(6), 740-759.
12. Keller, K.L. and Lehmann, D.R. (2009). Assessing long-term brand potential. *Journal of Brand Management*, 17(1), 6-17.
13. Keller, K.L. (2003). *Strategic Brand Management*, 2nd edition, Upper Saddle River, NJ: Prentice Hall.
14. Liu, X. and Hu, M.Y. (2011). Umbrella brand price premiums: effects of compatibility, similarity, and portfolio size. *Journal of Product & Brand Management* 20/1, 58-64.
15. Perrier, R. (1997). *Brand Valuation*, 3rd edition, London Premier Books.
16. Quinn, J. (2009). *Britain overtakes US as top financial centre*, The Telegraph, 08 October 2009.
17. Raggio, R.D. and Leone, R.P. (2009). Drivers of brand value, estimation of brand value in practice and use of brand valuation: Introduction to the special issue. *Journal of Brand Management*, 17, 1-5.
18. Sciolino, E. (2009). *In the Lap of Luxury, Paris Squirms*, New York Times, January 14, 2009.
19. Siddiqui, K. (2011). *Brand Equity Trends among Top 100 Global Brands: Region-wise Analysis*, 11th South Asian Management Forum (SAMF), Association of Management Development Institutions in South Asia (AMDISA), May 16-17, 2011, Kathmandu, Nepal.
20. Silverstein, B. (2009). *Small country, big brands*. Brandchannel.com. 24 August, 2009.
21. Simon, C.J. and Sullivan, M.W. (1993). The Measurements and Determinants of Brand Equity: A Financial Approach. *Marketing Science*, 12(Winter), 28-52.
22. Wang, H.M.D. (2010). Corporate social performance and financial based brand equity. *Journal of Product & Brand Management*, 19/5, 335-345.
23. Ward, R. and Perrier, R. (1998). Brand valuation: The times are changing. *Journal of Brand Management*, 5(8), 83-189.
24. Yeung, M. and Ramasamy, B. (2008). Brand value and firm performance nexus: Further empirical evidence. *Journal of Brand Management*, 15, 322-335.

Table 1
Region-wise Summary of Brands included in Top 100 Global Brands (2001-13)

American Brands	Asian Brands	European Brands
Ford	Toyota	Mercedes
Chevrolet	Honda	BMW
	Hyundai	Volkswagen
	Lexus	Audi
	Nissan	Porsche
	Kia	Ferrari

Table 2
Country-wise Summary of Brands included in Top 100 Global Brands (2001-13)

Country	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Germany	3	3	3	5	5	5	5	5	5	5	5	5	5
Japan	2	2	3	3	3	4	4	3	3	2	3	3	3
Korea					1	1	1	1	1	1	1	2	2
US	1	1	1	1	1	1	1	1	1	1	1	1	2
Italy								1	1	1	1	1	1
Total	**6**	**6**	**7**	**9**	**10**	**11**	**11**	**11**	**11**	**10**	**11**	**12**	**13**

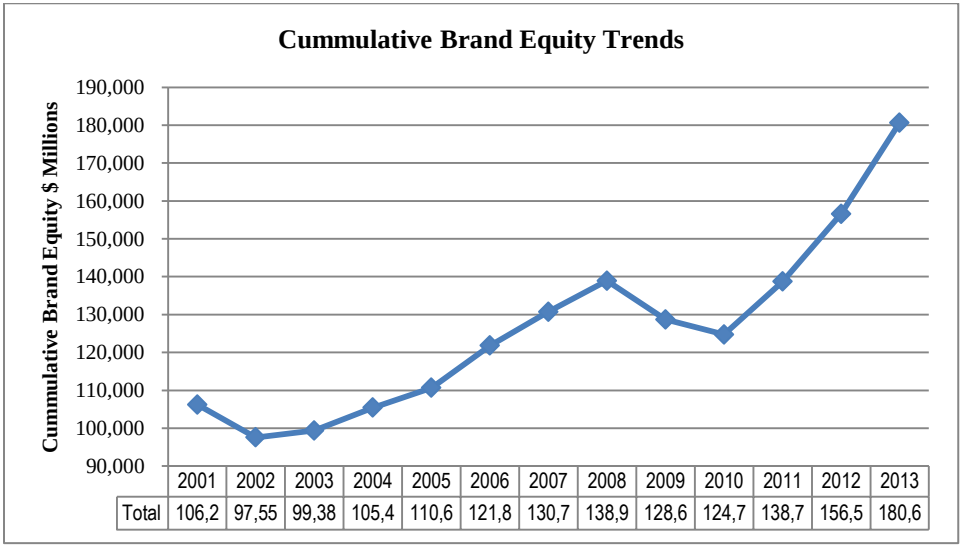
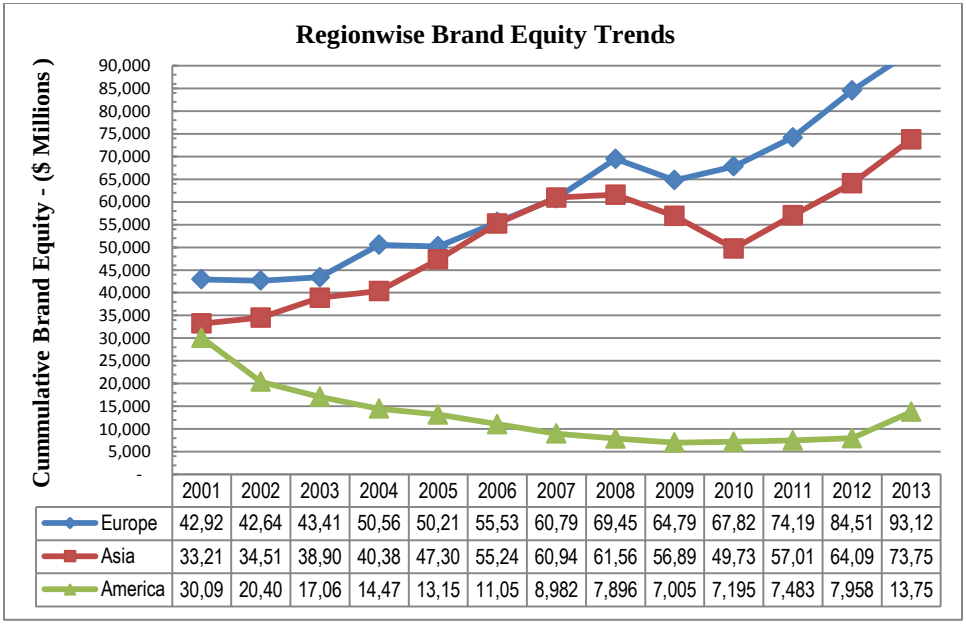
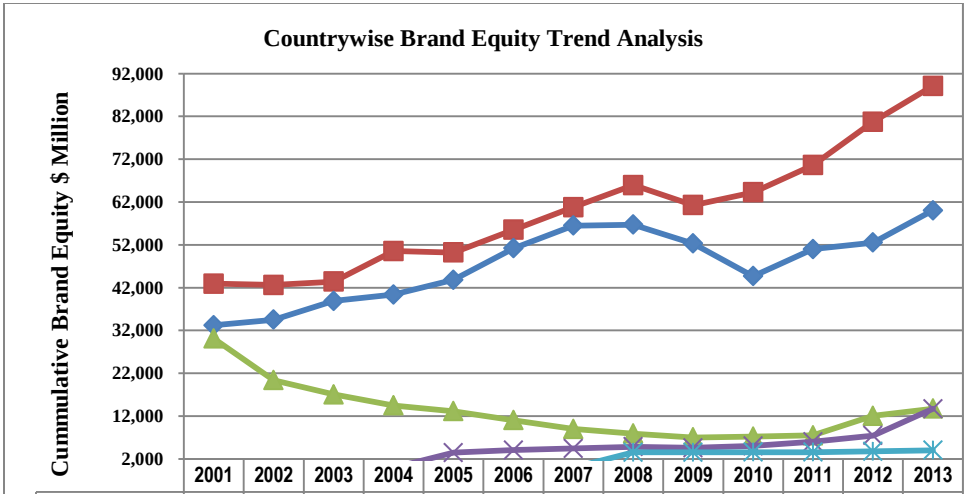


Fig. 1: Cumulative Brand Equity - Automobile Sector



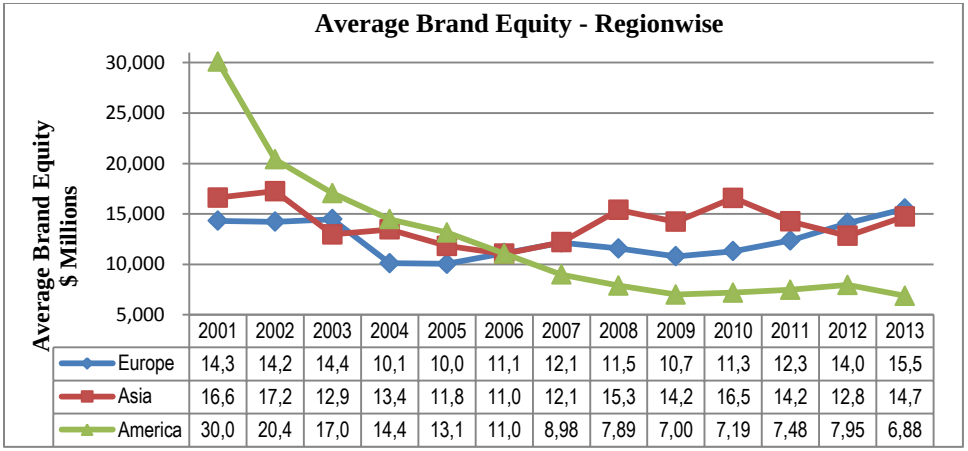
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Europe	42,92	42,64	43,41	50,56	50,21	55,53	60,79	69,45	64,79	67,82	74,19	84,51	93,12
Asia	33,21	34,51	38,90	40,38	47,30	55,24	60,94	61,56	56,89	49,73	57,01	64,09	73,75
America	30,09	20,40	17,06	14,47	13,15	11,05	8,982	7,896	7,005	7,195	7,483	7,958	13,75

Fig. 2: Cumulative Brand Equity Trends Analysis (Region-wise)



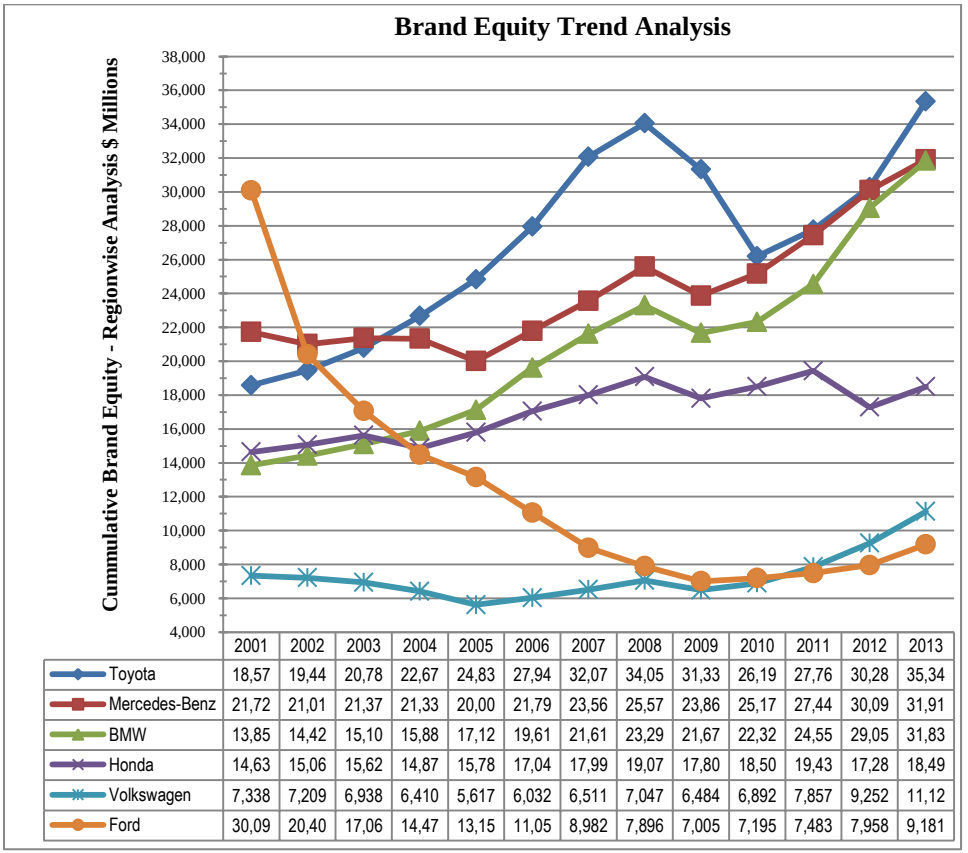
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Germany	42,92	42,64	43,41	50,56	50,21	55,53	60,79	65,93	61,26	64,25	70,60	80,74	89,11
Japan	33,21	34,51	38,90	40,38	43,82	51,16	56,49	56,71	52,29	44,69	51,01	52,52	60,03
USA	30,09	20,40	17,06	14,47	13,15	11,05	8,982	7,896	7,005	7,195	7,483	12,04	13,75
Korea	0	0	0	0	3,480	4,078	4,453	4,846	4,604	5,033	6,005	7,473	13,71
Italy	0	0	0	0	0	0	0	3,527	3,527	3,562	3,591	3,770	4,013

Fig. 5: Cumulative Brand Equity Trends Analysis (Country-wise)



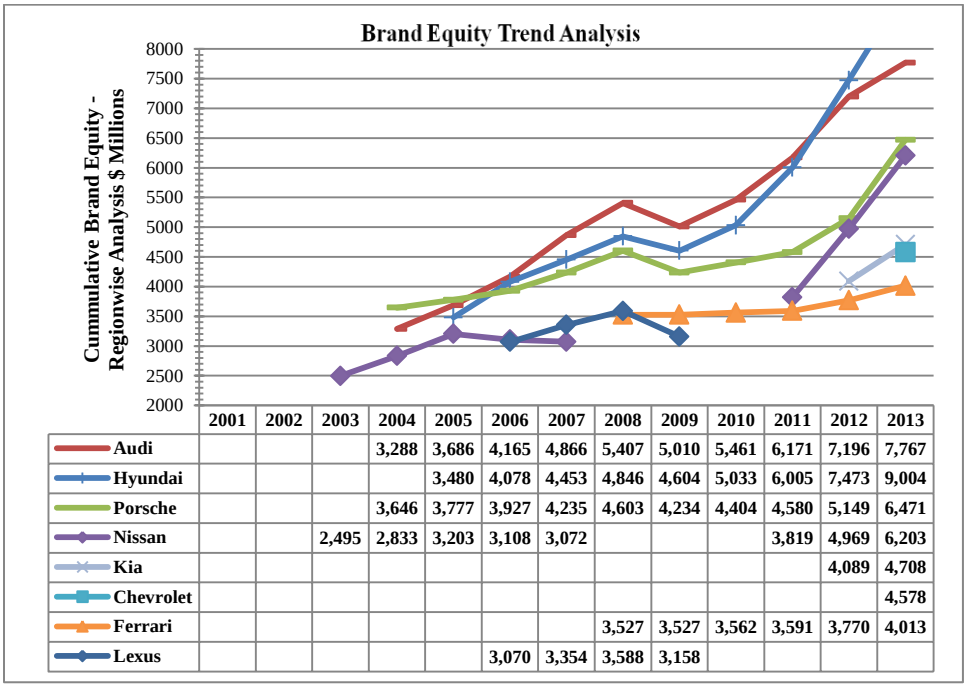
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Europe	14,3	14,2	14,4	10,1	10,0	11,1	12,1	11,5	10,7	11,3	12,3	14,0	15,5
Asia	16,6	17,2	12,9	13,4	11,8	11,0	12,1	15,3	14,2	16,5	14,2	12,8	14,7
America	30,0	20,4	17,0	14,4	13,1	11,0	8,98	7,89	7,00	7,19	7,48	7,95	6,88

Fig. 4: Average Brand Equity – Region-wise



	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Toyota	18,57	19,44	20,78	22,67	24,83	27,94	32,07	34,05	31,33	26,19	27,76	30,28	35,34
Mercedes-Benz	21,72	21,01	21,37	21,33	20,00	21,79	23,56	25,57	23,86	25,17	27,44	30,09	31,91
BMW	13,85	14,42	15,10	15,88	17,12	19,61	21,61	23,29	21,67	22,32	24,55	29,05	31,83
Honda	14,63	15,06	15,62	14,87	15,78	17,04	17,99	19,07	17,80	18,50	19,43	17,28	18,49
Volkswagen	7,338	7,209	6,938	6,410	5,617	6,032	6,511	7,047	6,484	6,892	7,857	9,252	11,12
Ford	30,09	20,40	17,06	14,47	13,15	11,05	8,982	7,896	7,005	7,195	7,483	7,958	9,181

Fig. 5: Brand Equity Trends - Top 6 Automobile Brands



	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Audi				3,288	3,686	4,165	4,866	5,407	5,010	5,461	6,171	7,196	7,767
Hyundai					3,480	4,078	4,453	4,846	4,604	5,033	6,005	7,473	9,004
Porsche				3,646	3,777	3,927	4,235	4,603	4,234	4,404	4,580	5,149	6,471
Nissan			2,495	2,833	3,203	3,108	3,072				3,819	4,969	6,203
Kia												4,089	4,708
Chevrolet													4,578
Ferrari								3,527	3,527	3,562	3,591	3,770	4,013
Lexus						3,070	3,354	3,588	3,158				

Fig. 6: Brand Equity Trends - Other Automobile Brands

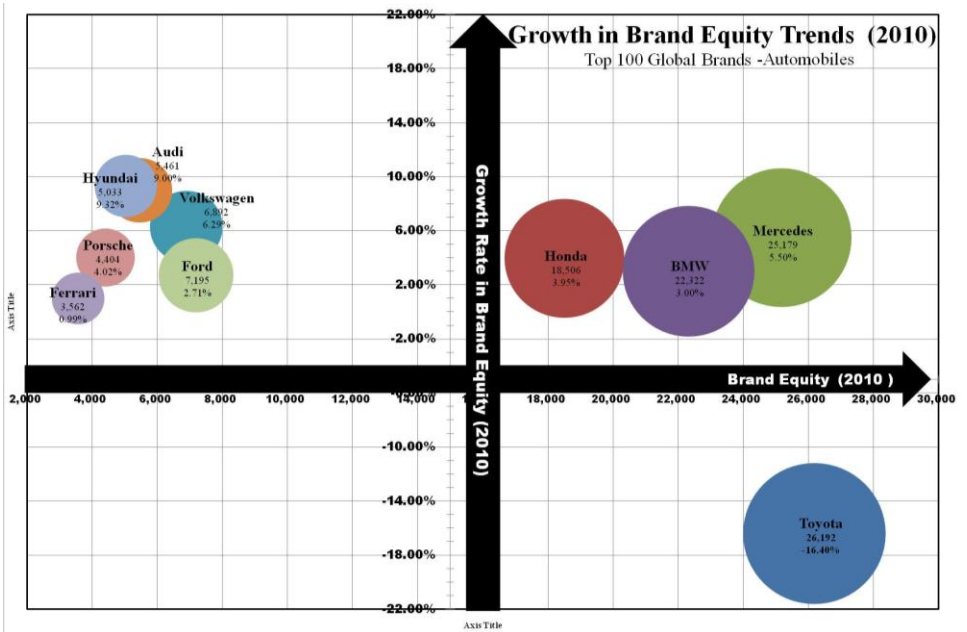


Fig. 7: Clusters of Top Global Brands (2010)

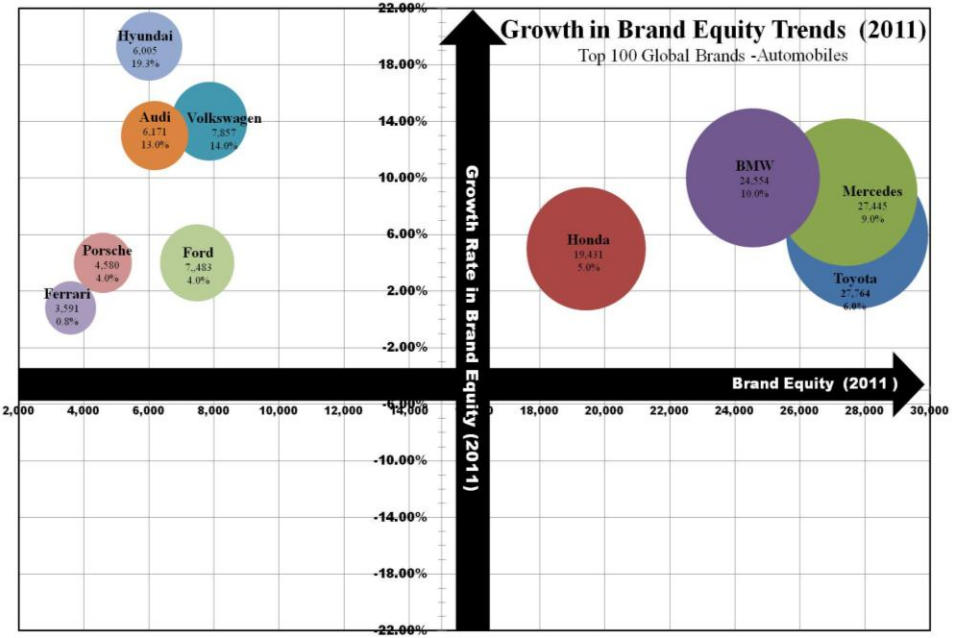


Fig. 8: Clusters of Top Global Brands (2011)

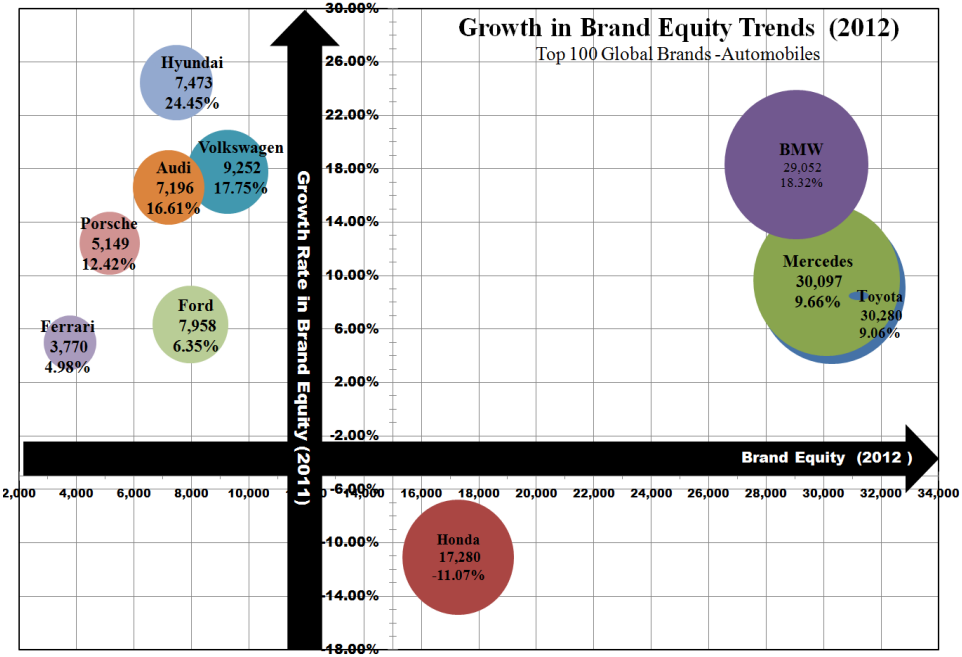


Fig. 9: Clusters of Top Global Brands (2012)

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 475-490

ROLE OF STATISTICAL ANALYSIS IN SELF ASSESSMENT FOR PROGRAM'S IMPROVEMENTS: A STUDY OF AN ENGINEERING SCHOOL OF PAKISTAN

Farrukh Idrees
Quality Enhancement Cell, FAST National University of Computer and Emerging Sciences, Islamabad, Pakistan
Email: shadrees@gmail.com

ABSTRACT

Higher education is a linchpin of the Pakistani economy. The paper addresses the call in literature for demonstrating the phenomenon of quality in higher education of Pakistan for continuous improvements. Various initiatives taken by the sector for enhancing performance have been presented corroborated with empirical evidences. The unit of analysis of the study is FAST School of Engineering. Quality management has facilitated in the assessment of program's objective for continuous improvements through alumni survey and employers' survey. Sample of 60 alumni and 25 employers is selected. Descriptive technique is used to analyze data through measuring frequencies. Report writing skills, ability to link theory with practice, ability to design system components etc. are found to be the areas for improvements. Various steps are taken by the School to make up the deficiencies. More rigorous quantitative techniques e.g. regression analysis, factor analysis etc. may be used to triangulate the findings. Further, sample size may be enhanced to generalize the findings of the study.

1. INTRODUCTION

Quality management concept came in to practice in 20th century in manufacturing sector with emphasis on customers' satisfaction. American Society for Quality defines Quality Management as a management approach to long–term success through customer satisfaction. In a quality management effort, all members of an organization participate in improving processes, products, services, and the culture in which they work. Since the inception of the quality management philosophy various frameworks have emerged to cope up with various challenges to enhance productivity e.g. Six sigma, process management, process reengineering, excellence models, balanced score card etc. The purpose of all these models is produce high quality products.

Since 1990s after tremendous success of quality management in manufacturing sector it started diffusing in services sector e.g. hospitality, healthcare, education etc. 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). Quality education is a prerequisite to gain access to knowledge which guarantees higher productivity (Maire 2002). For example: through various and sundry certifications, accreditations and frameworks, quality management system is acting as a check point to identify and constrain those institutions and their practices which are operational with profit incentive merely. Such institutions are "Degree Mills" with objective of producing graduates with limited or no vision about where these outputs will be consumed. Therefore there should be a mechanism of quality assurance to measure and manage meticulous conformance of institutions' practices with standards and norms of higher education.

The 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 favours 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 (Harden 2002; Harden et al. 1999; Davis 2003; Luitel, 2002; Schwarz et al. 1994). 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, Association to Advance Collegiate Schools of Business etc.).

There are also some challenges which cause hurdle in transferring the best practices from industries to education e.g. clear definition of an institution's customer is required (Jaideep et al. 1997; Brown et al. 1993; Marchese, 1993); rigid culture of the universities; unfocused improvement efforts without specific product or process improvement; expectations to see results immediately, and this is usually not the case (Bassi et al. 1994).

The thrust of this applied study to explain the extent to which quality assurance has contributed in continuous improvement of an engineering school. The paper has been organized in a way that till now introduction of quality in higher education with global context has been presented. **Section 2** provides panoramic view of quality management in the framework of Pakistani higher education industry. **Section 3** deals with the challenges being faced by Pakistani higher education sector. **Section 4** summarizes the background of unit of analysis i.e. FAST School of Engineering. **Section 5 and 6** share the findings and their analysis of the surveys for measuring program's outcomes **Section 7**, the last part of this study presents summary of the work, which is followed by limitations of the study.

2. QUALITY IN HIGHER EDUCATION SECTOR OF PAKISTAN

Pakistan's transition to a middle-income knowledge economy is critically dependent on the scale and performance of its higher education sector. Institutions of Higher Learning must play a leadership role in this transition through the production of skilled, innovative and enterprising knowledge workers.

In 2006 Quality Assurance Agency of **HEC (Higher Education Commission) of Pakistan** introduced Quality Enhancement Cells (QECs) as its operational units in universities to support internal quality assurance mechanism of institutes of higher education both in public and private sectors. Quality management system in higher education institutions of Pakistan is divided in to two tiers:

a. **External tier** that encompasses accreditation, external assessment, monitoring and evaluation, quality based ranking etc.
b. **Internal tier** which consists of curriculum revision, peer observation, student feedback, faculty course review, self-assessment etc. QEC are mandated to look after internal tier of the system.

Internal tier is more powerful than the external one because it ensures the operational working of the system more rigorously. Moreover, consolidated role of internal tier/internal quality assurance system is to create a quality oriented culture by

a. Bringing staff members together to share and learn from each other.
b. Creating a sense of responsibility among stakeholders.
c. Inculcating a process oriented approach throughout the institution.

In institutes of higher education, quality management is playing its role to improve programs through a quality assurance model, which encompasses various tools e.g. self-assessment, departmental review, external accreditations, benchmarking etc. for performance improvements. This framework is derived from EFQM Excellence Model ®.

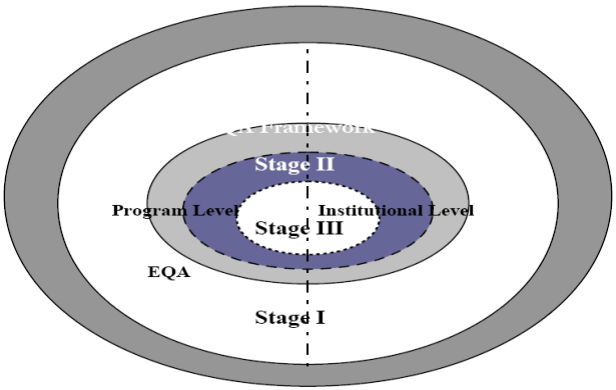


Figure 1: HEC Quality Assurance Framework in Pakistan

There are three stages in this framework

1. External quality assurance by accreditation bodies e.g. Pakistan Engineering Council, National Computing Education Accreditation Council, and National Business Education Accreditation Council etc.
2. Program level Quality Assurance through self-assessment.
3. Institutional level quality assurance through HEC institutional performance evaluation framework etc.

The research revolves around stage 2 in the context of self-assessment. The framework encourages justifying the role of continuous learning and improvement in the organisation through self-assessment. This justification is the adoption of an approach of management by processes and facts; which in turn is one of the basic building blocks of the model. Following are some pertinent areas of the program's self-assessment criteria which are conformed for performance improvements through quality management mechanism:

a. **Objectives of Programs**: Quality management ensures that all programs should have documented and measurable objectives and outcomes. These objectives facilitates in navigating programs in the right direction for producing high quality graduates who will be able to demonstrate the right set of skills and aptitudes..

b. **Curriculum standardization**: Quality management mechanism measures the extent to which curriculum is aligned with objectives and outcomes of programs through structured survey instrument(s) e.g. course review report, student feedback on courses etc.

c. **Laboratories and computing facilities**: Quality management system measures the extent to which laboratories and computing facilities are available to students and faculty to support teaching and research activities. Further it also measures the extent to which students and faculty have adequate and timely access to the manuals/documentation and instructions pertinent to laboratories to achieve programs' objectives.

d. **Students' advisory services**: Quality management ensures student have adequate support to complete the program in a timely manner and must have ample opportunity to interact with their instructors and receive timely advice about program requirements and career alternatives. This is done by taking feedback from students on a structured questionnaire periodically.

e. **Faculty:** Quality assurance ensures that faculty members must be current and active in their discipline and have the necessary technical depth and breadth to support the program. There must be enough faculty members to provide continuity and stability, to cover the curriculum adequately and effectively, and to allow for scholarly activities. To measure this, a structured questionnaire "Faculty Resume" is operationalized. Further, to measure and manage faculty satisfaction, a periodic satisfaction survey is launched regularly.

f. **Processes**: Quality assurance ensures that processes by which major functions are delivered (e.g. admission processes, examination process, placement process, registration process, faculty recruitment process, curriculum delivery process etc.) must be in place, controlled, periodically reviewed, evaluated and continuously improved.

g. **Institutional support**: Quality system measures the extent to which the institution's support and the financial resources for the program are available to provide an environment in which the program can achieve its objectives and retain its strength

h. Institutional facilities: Quality management measures the extent to which institutional facilities, including library, classrooms and offices are adequate to support the objective of the program through a structured mechanism.

In essence, strategic quality review on all above mentioned 8 dimensions is conducted for enhancing productivity of an institute of higher education as per the set objectives.

3. PROBLEM AREAS IN QUALITY MANAGEMENT IMPLEMENTATION

There are various challenges to the implementation of TQM in Pakistan:

- Quantification of various dimensions of quality or self-assessment e.g. extent to which programs have achieved their stated objectives and outcomes, measurement of students' satisfaction, measurement of faculty satisfaction etc.
- Further any big challenge is the transformation of functional orientation in to cross functional thinking and working is one of the biggest challenges in successful implementation of quality management in institutes of higher education.
- Moreover, adoption of student focused approach in designing program's objectives is another area that requires demonstration of top management commitment for implementing quality management initiatives to generating promised results of performance excellence.

4. QUALITY MANAGEMENT IN FAST SCHOOL OF ENGINEERING

All the above self-assessment criteria have been implemented in FAST School of Engineering through the introduction of Quality Enhancement Cell (QEC). The school is based in FAST National University of Computer and Emerging Sciences Islamabad which was constituted in year 2000. The school offers BS, MS and PhD in Electrical Engineering. 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 by aligning with objectives of programs.

The present study deals with the utilization of survey results of two crucial questionnaires i.e. Alumni Survey and Employers' Survey. These surveys are conducted at BS (Electrical Engineering) level on periodic basis to measure the extent to which program is successful in achieving its stated outcomes. The outcomes of the program revolve around imparting four basic attributes in the graduates:

a. Sound technical skills
b. Profound communication skills
c. Good interpersonal skills
d. High class work management skills

1. Methodology:

Table 1: Methodology

What?	How?
Measurement approach	Alumni Survey, Employers' Survey
Dimensions measured	Knowledge, Communication Skills, interpersonal Skills, Management and Leadership Skills
Population surveyed	Alumni Survey: 45 Alumni of School of Engineering Employers' Survey: 12 Employers of School of Engineering's graduates
Measurement instrument	On line likert scaled questionnaire
Number of statements	20 (for overall satisfaction)
Mode	Anonymous
Sample size	45 out of population of about 65
Response rate	Approx 70% in Alumni Survey Approx 30% in Employers' Survey
Data analysis software	SPSS
Data analysis technique	Frequency tables/bar charts
Communication of findings to questioned people	Online (charts plus text)
Research Assumption	All Satisfied students and employers are scoring at No. 5 of likert scale of questionnaire.

5. FINDINGS AND ANALYSIS

Following section summarizes the findings of the data collected along with analysis. Findings are summarized under four heads i.e. Ability to demonstrate acquired knowledge; Communication Skills; Interpersonal Skill; Work Management Skills.

A. Ability to Demonstrate Acquired Knowledge

That construct was measured through following six variables:
Ability of graduates of the program to demonstrate

I) Maths, Science and Humanities and Professional Disciple Knowledge
II) Problem formulation and solving skills
III) Collecting and analyzing appropriate data
IV) Ability to link theory with practice
V) Information technology knowledge
VI) Ability to design a system component

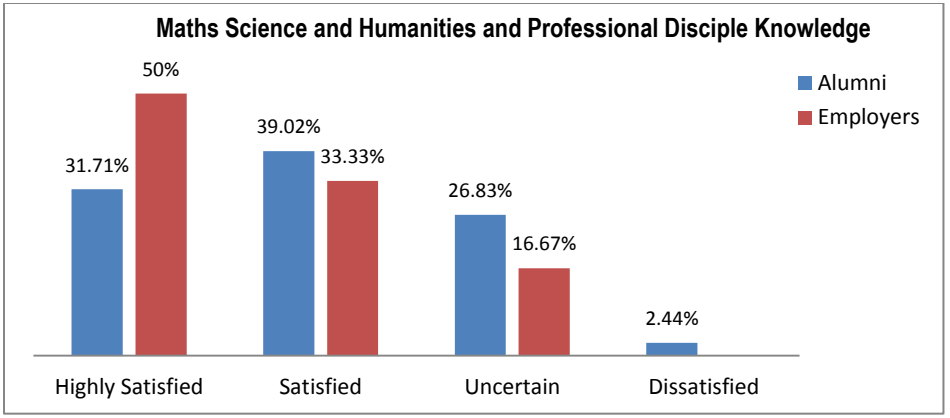


Figure 2

Most of the employers are satisfied with alumni's ability to demonstrate technical knowledge.

Most of the alumni are satisfied with their ability to demonstrate technical knowledge. However, about 29% of alumni are either uncertain/dissatisfied with their ability to demonstrate this attribute through program of study.

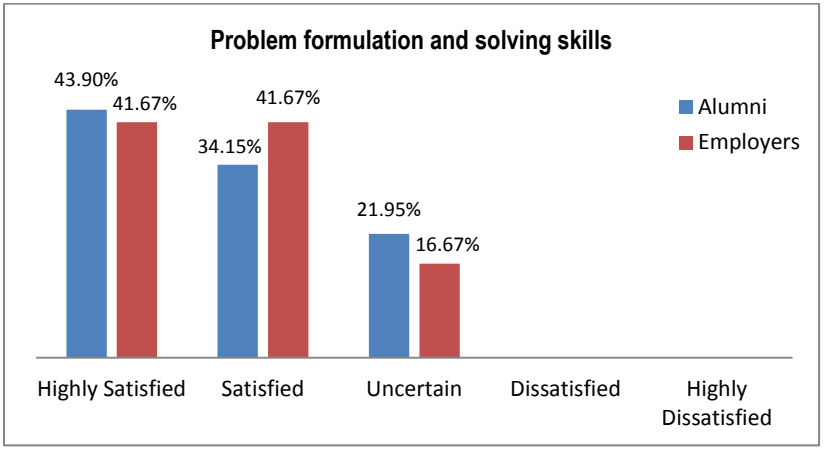


Figure 3

Most of the employers are satisfied with alumni's ability to demonstrate problem formulation and solving skills.

Most of the alumni are satisfied with their ability to demonstrate problem formulation and solving skills. About 22% of alumni are uncertain about the demonstration of this attribute through attending the program of study.

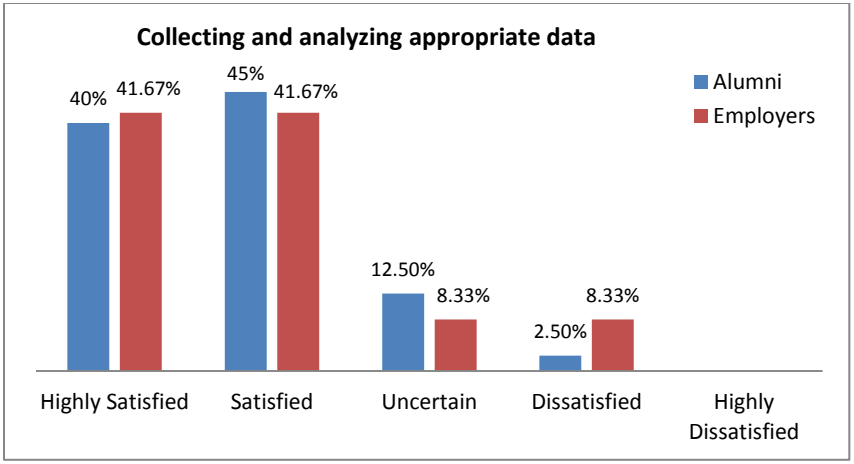


Figure 4

Most of the employers are satisfied with alumni's ability to demonstrate skills of collecting and analyzing data.

Most of the alumni are satisfied with their ability to demonstrate skills of collecting and analyzing data.

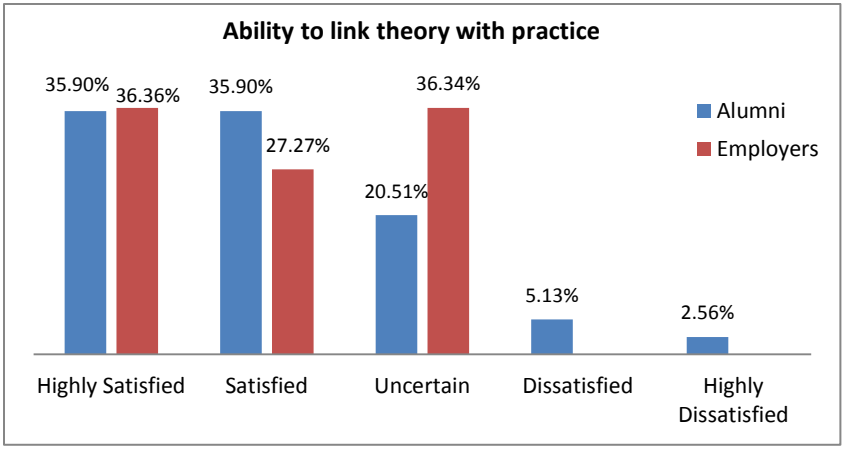


Figure 5

Most of the employers are satisfied with alumni's ability to link theory with practice. However, about 40% of the employers are either uncertain/dissatisfied with this outcome of graduates.

Most of the alumni are satisfied with their ability to demonstrate technical knowledge.

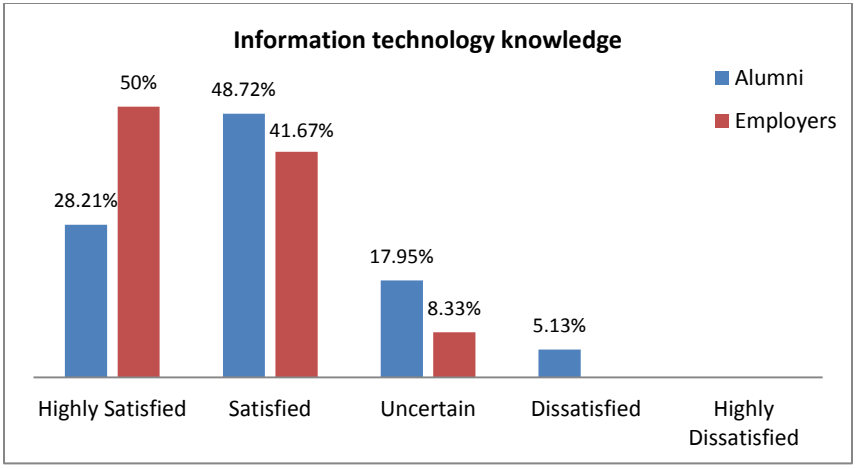


Figure 6

Most of the employers are satisfied with alumni's ability to demonstrate technical knowledge.

Most of the alumni are satisfied with their ability to demonstrate technical knowledge. However, about 25% of alumni are not satisfied with the demonstration of this outcome through the program of study.

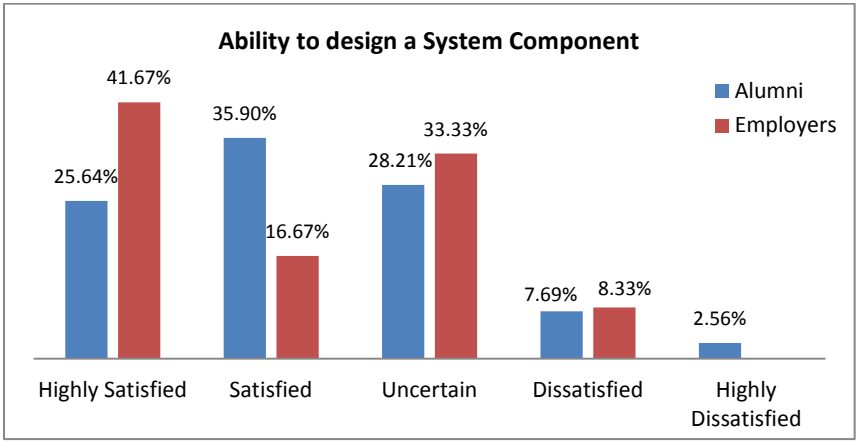


Figure 7

Most of the employers are satisfied with alumni's ability to design a system component. However, about 40% of employers are either uncertain/dissatisfied about the demonstration of this outcome.

Most of the alumni are satisfied with their ability to design a system component. However, about 38% of alumni are either uncertain/dissatisfied with the demonstration of this outcome.

B. Communication Skills

This construct was measured through three variables:

1. Oral communication
2. Report writing
3. Presentation skills

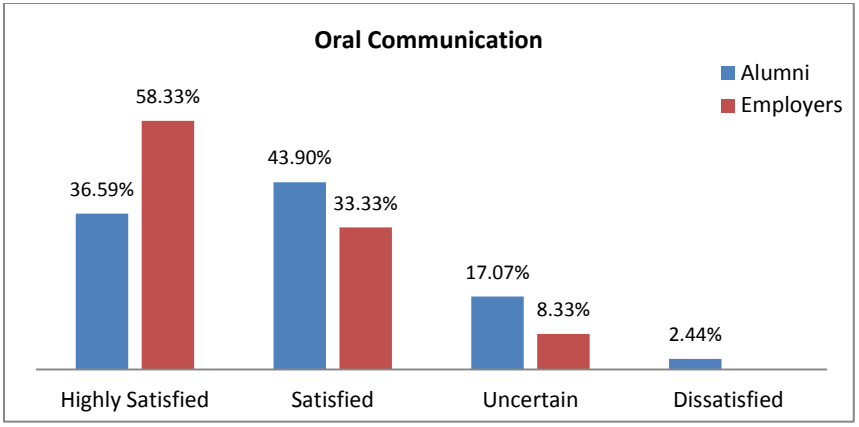


Figure 8

Most of the employers are satisfied with alumni's ability to communicate orally.

Most of the alumni are satisfied with their ability to communicate orally. However, about 20% alumni are either uncertain/dissatisfied about the demonstration of the outcome through the program of study.

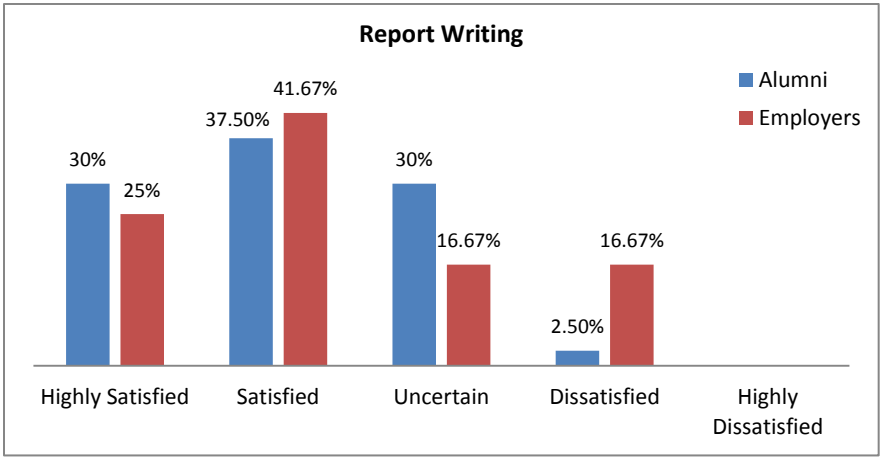


Figure 9

Most of the employers are satisfied with alumni's ability to communicate through report writing. However, about 30% of employers are either uncertain/dissatisfied with the demonstration of this outcome.

Most of the alumni are satisfied with their ability to communicate through report writing. However, about 30% of alumni are either uncertain/dissatisfied with the demonstration of this outcome.

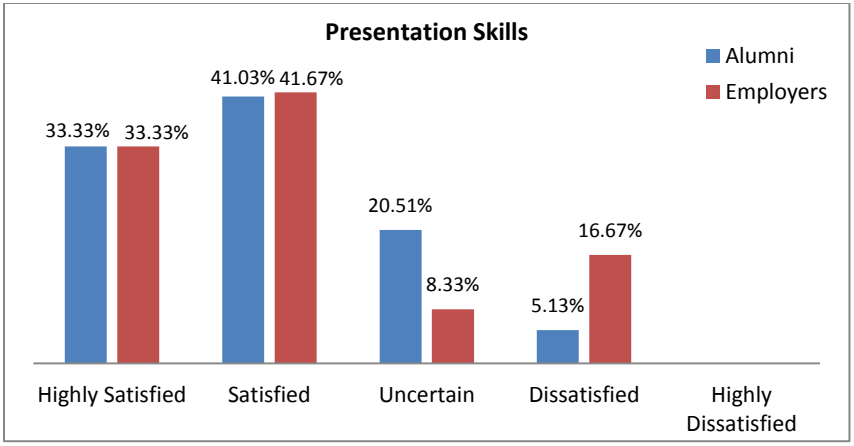


Figure 10

Most of the employers are satisfied with alumni's ability to demonstrate presentation skills. However, about 25% are either uncertain/dissatisfied with the demonstration of this attribute of graduates.

Most of the alumni are satisfied with their ability to demonstrate presentation skills. However, about 25% of alumni are either uncertain/dissatisfied with the demonstration of this attribute.

C. Interpersonal Skills

The construct was measured through following two variables:

1. Ability to work in teams
2. Ability to demonstrate ethical values

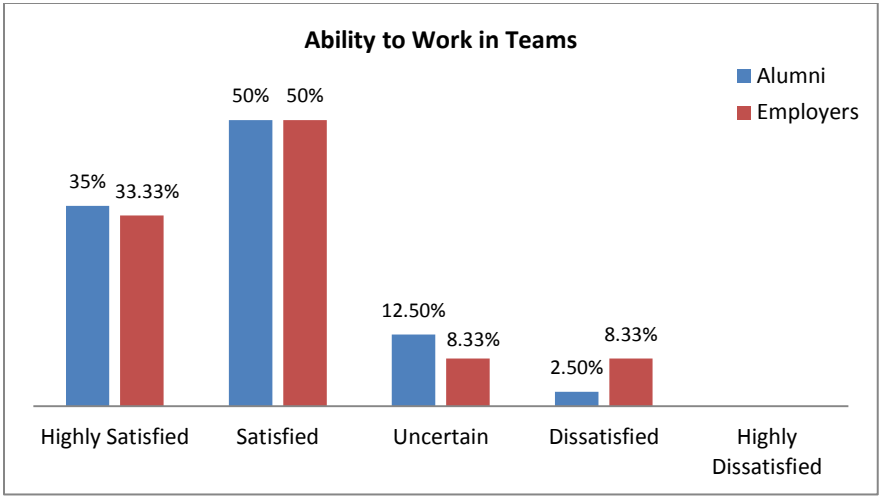


Figure 11

Most of the employers are satisfied with alumni's ability to design a system component.

Most of the alumni are satisfied with their ability to design a system component.

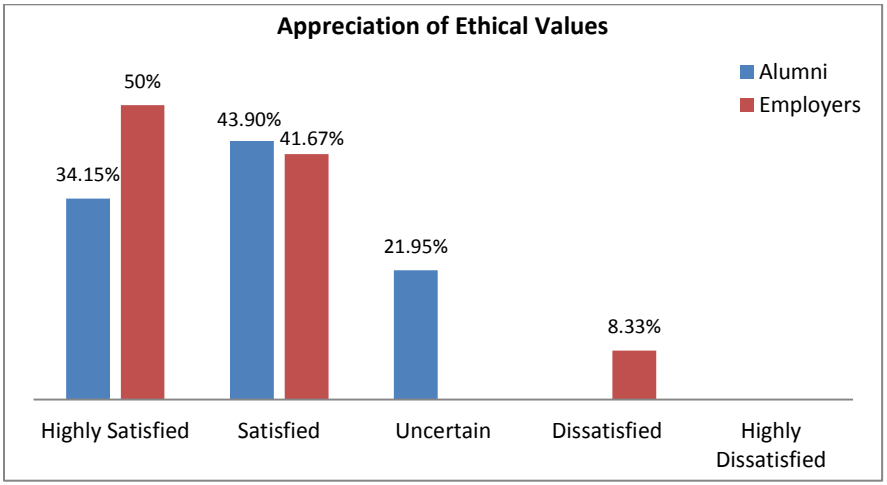


Figure 12

Most of the employers are satisfied with alumni's ability to demonstrate appreciation of ethical values.

Most of the alumni are satisfied with their ability to demonstrate appreciation of ethical values. However, about 22% of alumni are uncertain about the demonstration of this attribute through the program of study.

D. Work Skills

This construct is measured through three variables;

1. Time management skills
2. Ability to demonstrate discipline
3. Ability to make judgments

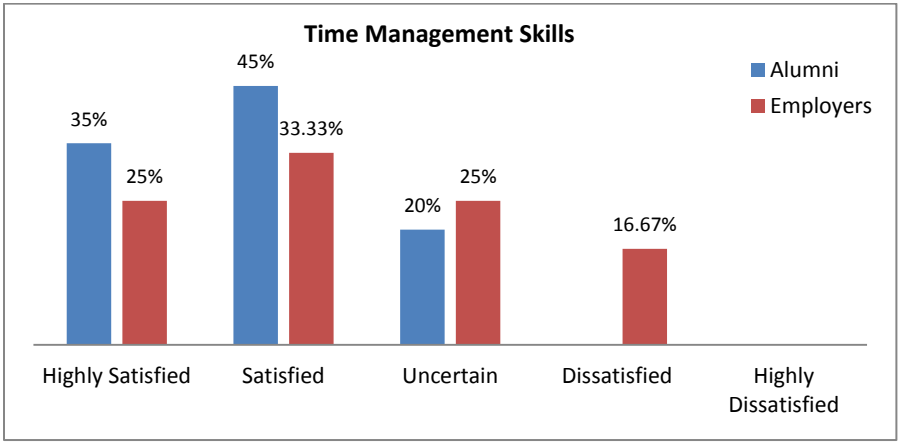


Figure 13

Most of the employers are satisfied with alumni's ability to demonstrate time management skills. However, about 37% of employers are either uncertain/dissatisfied with this attribute of alumni.

Most of the alumni are satisfied with their ability to design a system component. About 20% of alumni are uncertain about the demonstration of this attribute through attending the program of study.

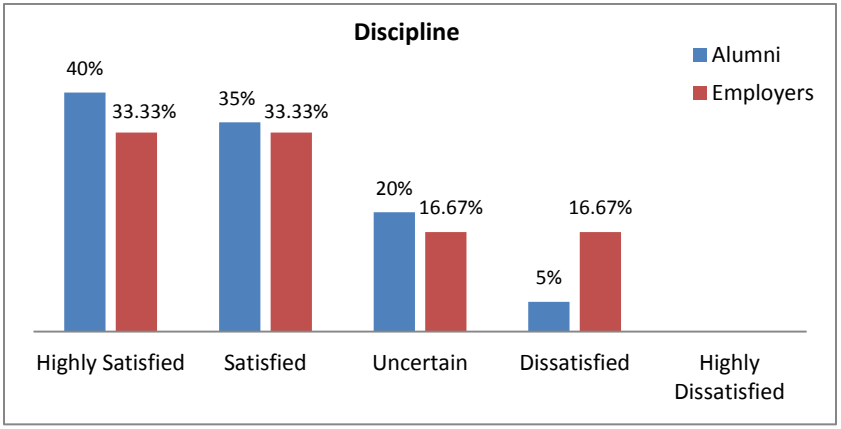


Figure 14

Most of the employers are satisfied with alumni's ability to demonstrate discipline. However, about 33% of employers are either uncertain/dissatisfied with the demonstration of this attribute in alumni.

Most of the alumni are satisfied with their ability to demonstrate discipline. However, about 25% of alumni are either uncertain/dissatisfied with the demonstration of this attribute through program of study.

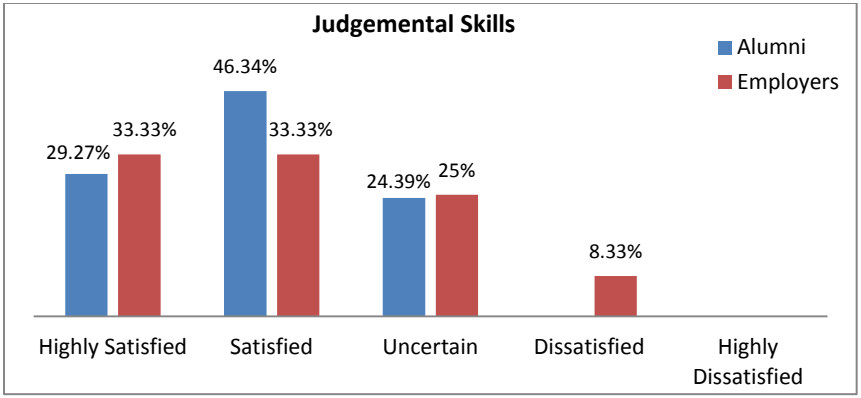


Figure 15

Most of the employers are satisfied with alumni's ability to demonstrate judgmental skills. However, about 33% of employers are either uncertain/dissatisfied with the demonstration of this skills in alumni.

Most of the alumni are satisfied with their ability to demonstrate judgmental skills. However, about 25% are uncertain about the ability to demonstrate this skill through program of study.

6. ACTIONS TAKEN FOR IMPROVEMENTS BY FAST SCHOOL OF ENGINEERING

Most of the quality parameters are conformed to the set standards in a way that both employers and alumni majority are satisfied with the demonstration of required skill. However, about all areas i.e. Knowledge, communication, interpersonal skill, and work management skills require improvements to further strengthen the quality of program in imparting stated outcomes. Various and sundry steps have been taken by the School management in improving the quality of academic programs. Some of the steps are mentioned below:

1. Developing technical skills of graduates:
   a. Alignment of labs objectives with that of theory courses
   b. Equipment Up gradation of labs
   c. Emphasis on hands on experience instead of simulation
2. Curriculum Up gradation:
   a. Revision in course outlines (In progress) based on input from faculty to make it more agile as per market requirements

3. Nurturing soft skills of students:
   Launching of IEEE Week to collaborate with industry and to polish soft skills of students e.g. Leadership, project management, time management, communication, and presentation etc.
4. Launching of new electives:
   Industrial Process Control
   (Others are in progress based up on the input from industry)

7. 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 objectives of programs, curriculum delivery, laboratories quality, faculty development, processes management, institutional support etc. Various and sundry initiatives have been taken by QEC as per the guidelines provided by HEC of Pakistan for performance improvements. The effective deployment of quality may facilitate FAST School of in gaining/sustaining a competitive position in the academic sector nationally and ultimately globally by producing high quality graduates that are able to demonstrate required outcomes i.e. Strong technical knowledge, interpersonal skills, communication skills, and work management skills. Further, strong and focused internal quality management system of a university also paves the way for rigorous self-assessment of the programs. That may in turn facilitate in finding strengths and areas for improvements. As a part of research limitations, more rigorous quantitative techniques e.g. regression analysis, factor analysis etc. may be used to triangulate the findings. Further, sample size may be enhanced to generalize the findings of the study.

REFERENCES

1. Al-Lawatia, D. (2006). *A Study of TQM Implementation in Higher Education*, MA Dissertation, Bradford; School of Management, University of Bradford.
2. Antony, J. and Preece, D. (2001). *Understanding and Implementing Quality*. Routledge.
3. Atikinson, P. (2012). *Creating Cutlure Change*, accessed at iomnet.org.uk on June 10, 2013
4. Bassi, J. and McMurrer, P. (1998). Training Investment Can Mean Financial Performance. *Training & Development*, 52(5), pp. 40.
5. Brown, J. and Koenig, F. (1993). Applying total quality management to business education. *Journal of Education for Business*, 68(6), 325-9.
6. Bruun, P. and Mefford, R.N. (1996). A framework for selecting and introducing appropriate production technology in developing countries. *International Journal of Production Economics*, 197-209.
7. Cecily, R. and Denih, P. (1996). TQM Just what the ethicist order. *Journal of Business Ethics*, 15(9), pp 963.
8. Cole, E. (1979). *Made in Japan — quality control circles*, Across the Board, November, 72-78
9. Davis, M.H. (2003). Outcome-based education. *Journal of Veterinary Medical Education*, 30(3), 258-263.

10. Dexelar, J. and Kliensorge, I. (2000). Using total quality processes and learning outcome assessment to develop management curricula. *Journal of Management Education*, 24(2).
11. Elmuti, D., Yunus, K. and Matthew, M. (1996). Are TQM Programmes in Higher Education Worth the Effort? *International Journal of Quality and Reliability Management*, 52.
12. Garbutt, S. (1996). Transfer of TQM from industry to education. *Education and Training*, 38(7), pp 16-22.
13. Harden, J.R., Crosby, M.H., Davis, M. and Friedman, R.M. (1999). AMEE Guide No. 14: Outcome-based education: Part 5-From competency to meta-competency: a model for the specification of learning outcomes. *Medical teacher*, 21(6), 546-552.
14. Harden, R.M. (2002). Developments in outcome-based education. *Medical teacher*, 24(2), 117-120.
15. Holusha, J. (1993). W. Edwards Deming, expert on business management, dies at 93. *The New York Times*, 21 December, p. A13.
16. Jaideep, M., Michael, R. and Sower, V. (1997). A comprehensive model for implementing Total Quality Management in Higher Euducation. *Benchmarking for Quality Management and Technology*, 4(2), 104-120.
17. James, R.E. (1994). Total Quality Management. *Organizations and Strategy*. 2005: Thomson/South-Western.
18. Jarrar, Y. and Aspinwall, E. (1999). Integrating TQM and BPR, Is it enough? *Total Quality Management*, 10(4/5), 584-593.
19. Luitel, B.C. (2002). *Outcome-based education*. Retrieved June, 9, 2005.
20. Marchese, T. (1992). TQM at Penn: A report on first experiences. *AAHE Bulletin*, 45(3), 3-14.
21. Norman, R. (1991). Xerox on the move. *Forbes*, 10 June, pp. 70-71.
22. Overman, S. (1990). Leader helps improve competitiveness. *HR Magazine*, May 1990, 58-60.
23. Phillips, W., Chang, D. and Buzzell, D. (1983). Quality, cost position and business performance: A test of some key hypotheses. *Journal of Marketing*, 47(2), 26-43
24. Schwarz, G. and Cavener, L.A. (1994). Outcome-Based Education and Curriculum Change: Advocacy, Practice, and Critique. *Journal of Curriculum and Supervision*, 9(4), 326-38.
25. Smith, D.H. and Lewis, R.G. (2006). *Total Quality in Higher Education*. Delray Beach, Fla: Pentagon Press.
26. Taylor, W.A. and Wright, G.H. (2003). A Longitudinal Study of TQM Implementation: Factors Influencing Success and Failure. *Omega*, 97-111.
27. Taylor, W.A. and Wright, G.H. (2006). The Contribution of Measurement and Information Infrastrustrue to TQM Success. *Omega*, 372-384.
28. Tenner, A. and Detoro, J. (1992). *Total Quality Management: Three Steps for Continuous Improvement*, MA; Reading Addison-Wesley.
29. Wollner, E. (1992). The Law of Producing Quality. *Quality Progress*, (January), 35-40.
30. Zairi, M. (1993). *Total Quality Management For Engineers*. London: Gulf Publishing Company.
31. Zari, M. and Jarrar, Y. (2005). *Excellence Toolkit: Delivering Sustainable Performance, Activity Based Costing: What does it mean*? Oxford: Information Press.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 491-500

ON FIRM PERFORMANCE THE EFFECT OF SPAN OF SUPPLY CHAIN INTEGRATION

Muhammad Ibrahim Shamsi[1], Syed Taha Ali[2] and **Uzair Ahmed[3]**
Department of Computer Science, FAST-National University
of Computer and Emerging Sciences, Karachi, Pakistan
Email: [1] ibrahim.shamsi@nu.edu.pk
[2] k122062@nu.edu.pk
[3] k122046@nu.edu.pk

ABSTRACT

This study aims to explore different styles of supply chain practices and performance of firms. The firms are classified with respect to the span of their integration in the supply chain. Two types of integration of firms are discussed in this study – one is the firms with shorter group of supply chain partners (Narrow Span of Integration) and second is the firms with larger group of supply chain partners (Broad Span of Integration). The focus of past researcher was to significantly differentiate the supply chain integration. We have focused in this study the importance of supply chain that should incorporate in the span of integration. Throughout the supply chain, this survey puts up an experimental and observational support, for the call for to involve associates.

The assumption of the study is that the performance of the firms that are integrated with larger group of their supply chain partners is higher than those that are integrated with shorter supply chain partners. To formalize and validate this assumption a survey of 100 firms was conducted from various firms of Karachi. Independent sample T-Test was used to compare the mean performance of both types of firms.

Firms with larger group of supply chain partners hold far intense attention and clarity on coalition with its associates than those with a shorter group. Consequently, the larger group, show more elevated amounts of performance attribute in relation to supply chain. The outcomes foreground the significance to supply chain professionals by accepting a larger perspective of the supply chain instead of concentrating exclusively on first tier customers and suppliers.

1. INTRODUCTION

Supply chain integration (SCI) can be explicitly stated to be a level at which supply chain members work mutually to achieve goals (Flores, 2002)). SCI plays a significant role in creating ever lasting value. Porter's value chain model can be traced back to the notion of leveraging linkages. This highlights the importance of exploiting both intra and inters firm linkages (Porter, 1985). Nowadays interest has taken a rise as firms seek to establish a competitive advantage in an increasingly competitive, dynamic global Marketplace. Increasingly academics too are drawing upon a rich theoretical basis with base in piles of literature in economics, organizational theory, and lean systems/

reengineering (Droge, 2004) to have a sound understanding of what encourages integration and how to implement it with desired results.

When it comes to SCI more is better than less many authors have showed disguised evidence to this, and to the results of not properly integrating e.g. (Westbrook, 2001); (Vickery, 2003). An important question here would be what does it mean to be more or fully integrated? Back in the days, focus of research was on either direct upstream or downstream integration, digging into either the integration of immediate suppliers or customers into decision making processes. Froehlich and (Westbrook, 2001) added to the discussion by side by side examining upstream and downstream integration in relation to forward physical flows and backward information flows. Although studies afterward have highlighted the significance of taking a broader view of integration, the issue of the breadth of the supply has not been addressed.

The need to integrate explicitly across the supply is not new e.g. (Hamel, 1994); (Hammer, 2001). In fact, many definitions of supply chain management state openly to the entire supply chain, from acquiring of essential raw materials to the utilization of finished products. For example, Lambert (2004) explains supply chain management as the "cross functional integration within the firm and across the network of firms that comprise the supply chain". (Mentzer, 2001) cleared the ambiguity between direct supply chains which involve a focal firm, supplier and Customer, advanced supply chains that 2nd tier suppliers and customers, and absolute supply chains that bypass the entire supply chain. Even after descriptions such as these, the literature has showed little progress to evaluate the affect of integration across a wider gap of the supply chain. Indeed, using (Mentzer, 2001) characterization, it is obvious that the literature has failed to put in words expressively or had its attention towards direct supply chain.

Our motive is to further discuss of integration to explicitly acquaint with firms beyond 1st tier suppliers and customers. In more detailing, we overview if the performance benefits accrue to firms that use extended supply chain partners, & to see if the integration pattern of these firms is different from those that affiliate only with immediate supply chain partners. We provide a different translation of what "more" integration means by pondering over the breadth of the supply chain involved instead of the intensity of integration over a narrow span of the supply chain.

2. LITERATURE REVIEW

Span of Integration:

According to a recent survey of the performance implications of supply chain integration Donk (2008) uncovered 2 key problems. Firstly, focus of most literature was on relationship development and/or firm orientation towards holistic, long-term decision making processes rather than integration per se. Indeed, this is due to their classification of previous research based on ally's of inter firm interaction, attitudes of firms towards partners, and involvement of collaboration practices. Secondly, while there is sufficient evidence of the positive effect of integration on performance, it is difficult to state specifically from the literature due to confusion in how integration is defined, made into operation, and measured.

In light of sufficient bolstering evidence of a good relationship between integration and performance, the present direction of research is on the relationship between extent of integration among supply chain members and its performance benefits. Extent of integration is operationalized in terms of "more versus less" integration as defined by different authors of pertinent research is summarized in (Westbrook, 2001) coined the expression "arcs of integration" to characterize the extent to which firms integrated with upstream and downstream supply chain partners. Firms with greatest arcs were taken under observation, also those with vast integration with both suppliers and customers, showed large improvements in several measures of financial productivity, and non-productivity based performance. Taking this work as base, (Rosenzweig, 2003) explored the connection between integration intensity, competitive capability, and performance. Their evaluation was that integration intensity is surely related to quality, delivery, process flexibility, and cost capability. In addition to this they found positive relationships between integration intensity and new product revenues, return on assets, and customer satisfaction. In applying the "arcs of integration" concept to a comparison of US and East Asian firms, (Rajagopal, 2005) not only proved the results of (Westbrook, 2001), they demonstrated that they held in an East Asian context.

Narasimhan and (Kim, 2002) researched on result of internal, supplier, and customer integration on product & international market diversification. Their conclusion was integration with suppliers and customers enhance & regulates the flow between international market diversification and performance. Precisely, they evaluated that integration could benefit the point of diminishing returns of product and market diversification over time. Their demonstration also included the moderating effect of international market diversification on the relationship between product diversification and performance becomes negligible as the level of internal customer, and supplier integration rises.

(Droge, 2004) pondered over internal and external integration in relation to design processes and their effect on product development time, product cycle time, and responsiveness. Their observation was internal and external integration exhibit positive relationships with all estimations of time based performance. (Bagchi, 2005) defined integration in terms of the frequency of participation with main suppliers and customers and the breadth of functional groups involved. Not only they pointed out differences in level of integration with suppliers versus customers, but they also distinguished changes within the organization integration. Furthermore, they discovered proof of positive relationships between integration with suppliers in connection with supply chain design, inventory control, sales administration, research and development, & procurement, as well as with customers with regard to production, and several measures of performance.

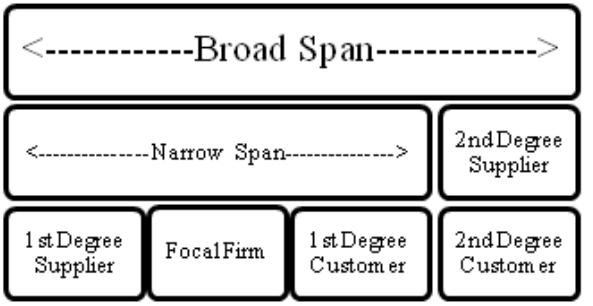
(Kim, 2006a) explored the effect of the degree of integration on the communication between corporate competitive capability and supply chain operational capability, & their impact on performance. Corporate competitive capability was made into operation in terms of cost, leadership, customer service, innovative marketing technology, and differentiation, while supply chain working capacity was evaluated in terms of technical, structural, and logistical capability. Proof for the hypothesis was found that as the level of integration advances, significant interactive effects diminish. They observed that internal and external sharing of resources and technological knowledge lowers the need for firms

to internalize supply chain operational capability. (kim, 2006b) made observations that among large firms, the level of integration has a good effect on corporate competitive capability and firm performance, but among small firms, this is true only with respect to performance. (Lee, 2007) explored the connection between strong internal, customer, and supplier linkages, & supply chain cost containment, performance reliability, and overall performance. Cost containment was measured in terms of logistics related costs on the other hand performance reliability was evaluated in terms of inventory performance. They derived that all three linkages were surely connected to overall performance and Performance reliability, and that strong internal and supplier linkages were related to cost containment. By making use of (Fisher, 1997) evaluation of functional versus innovative supply chains, Ramdas and Spekman (2000) concluded that innovative supply chains have strong integration than functional supply chains with respect to planning and control, quality management, service & after sales support. More of their observation was that high performing innovative supply chains are more of a stimulus to suppliers than those in functional supply chains. In addition to that firms with high performance record in innovative supply chains make of integrating mechanisms much more.

Narrow Span of Integration & Broad Span of Integration:

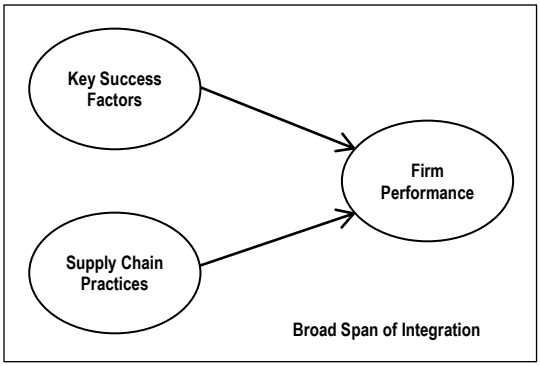
The outcomes give significant understandings into what it intends to be more integrated amongst the supply chain. Unlike prior analyzes that establish that the strength of integration has a positive consequence on performance, the outcomes here exemplify that how often of the supply chain is touched in integration efforts also plays a role. While the value of concerning participants from throughout the supply chain has been alluded to in both the management (e.g. (Hamel, 1994); (Hammer, 2001) and supply chain (e.g. (Power, 2005); (Rajagopal, 2005) literature, so far there has been no defending empirical prove. The outcomes establish that firms with a broad span of integration vary in their integration efforts from firms that focus on narrow span on multiple proportions. This is uniform with (Mentzer, 2001) distinction among extended, direct and ultimate supply chains. Not only do firms with a broad span of integration connect greater stress to facilitating chain wide integration efforts, they do so with regard to integration with customers and suppliers alone. The particular activities looked into with respect to customer and supplier integration didn't make a difference as to whether partners were first tier or beyond. Nevertheless, to the degree that firms generally focus on immediate customers and suppliers, the result is substantial in that it proposes that firms with a broad span are more centered on even these suppliers and customers than those with a narrow span. The outcomes concerning performance are crucial in that they manifest the positive change of a broad span of integration on relationship based results as well as betterments in customer service. The observation that firms with a broad span of integration outmatch those with a narrow span may be an indication that there is value in terms of reduced product development time, and higher quality and sales from building relationships with suppliers beyond the first tier. Or else, it may be an indication that reaching out to supply chain associates beyond the first tier indicates a bigger commitment and willingness to engaging partners in a collaborative, cooperative manner rather than reaching out to first tier customers and suppliers alone. It may also convey that firms with a broad span of integration have more laid down mechanisms to leverage supply chain relationships than those with a narrow span of integration. Whatever the outcome, the net outcome is that there is value linked with broader rather than narrower participation. This complements

the outcomes of (Westbrook, 2001). The significance of span of integration can be exemplified using the so-called "bullwhip" effect (Lee, 1997).At each interface in the supply chain, useless exchanges of data direct to data distortion and thus inadequate decision making. Firms with a broad span of integration can get over these challenges because they are not only better positioned to experience signals directly from the source rather than through intermediaries, they have deeper context with which to correctly understand and act on these signals. Furthermore, they experience signals in a well-timed manner. In contrast, firms with a narrow span of integration generally not only experienced less complete data, they must understand it in a less timely manner. The managerial implications of the outcomes can be illustrated by Wal-Mart's Remix initiatives to exploit supply chain integration (Hoffman, 2006). In an effort to improve in stock position and inventory turnover, Wal-Mart fixed inventory growth to one-half of sales growth. They cut down order sizes and demanded suppliers to deliver more often to Wal-Mart distribution centers. Nevertheless, smaller order sizes heightened the number of less-than-truckload shipments from vendors, thereby making over-crowding at the distribution centers. Wal-Mart lengthened its span of integration by integrating third party logistics suppliers. This eased over-crowding at the distribution centers as well as yielded extra shipping efficiencies.



3. MODEL FRAME WORK

This survey contributes to the writing by analyzing the patterns of integration practice, span of supply chain integration, and implications for performance. It allows for important understandings into what it intend to be more integrated, and highlights the motive for firms to employ the supply chain vastly. The outcomes also bring out chances for more exploration of the concept of supply chain integration. While they demonstrate the positive impact of a vast or broad span of integration, the results don't provide particular cues as to what tells these variations. The aim of the current survey was in part to investigate whether span of integration adds to performance differences. Based on this, the logical following step is to earn better insight into what it is regarding firms with a vast span of integration that states these performance differentials. The following model frame work was designed to validate our assumption.



4. RESEARCH HYPOTHESES

A business with a shorter span of integration will incline to include only itself and immediate supply chain partners to manage the supply chain, while a business with a broader span will incline to also include firms beyond immediate partners seems to be the exclusively study to date that clearly distinguished where in the supply chain proportional to the focal firm responding firms lay. Coherent with this, the underlying assumption of the current survey is that a firm's dedication to enlisting firms from a broad rather than narrow scope of the supply chain implicitly conveys differences in their loyalty to building inter firm connections. The following four hypotheses were proposed to validate the assumption of our research.

H1: The firm performance depends on span of integration
H2: The success factor depends on span of integration
H3: The supply chain practice depends on span of integration
H4: The firm performance depends on the key success factors and the supply chain practice.

5. RESEARCH METHODOLOGY

Data were collected using a survey of practicing managers. To assess the dimensions of supply chain integration, survey respondents were asked to indicate the importance of various practices in their integration efforts using a five-point Likert scale (1 not at all important, 5 extremely important). Practices included in the survey were identified from the literature and from discussions with industry professionals. Relationship performance was assessed by asking respondents to evaluate the success of their firm's partnerships with suppliers with respect to four dimensions of performance. The survey also asked about the participants in the responding firm's supply chain management efforts. Participants were classified as first & second tier suppliers and customers, other service providers. Respondents were asked to identify each participant using this scheme. After pre-testing and revision, the survey instrument was distributed to 102 respondents by going door to door. These 102 respondents included 95 of those which were integrated with their supply chain. Twenty one of the respondents were integrated broad span while 74 of them had narrow span of integration.

Group Statistics

	span	N	Mean	Std. Deviation	Std. Error Mean
Firm Performance	Broad Span	21	3.6071	1.02948	.22465
	Narrow Span	74	1.9595	.62487	.07264
Success Factors	Broad Span	21	3.5119	1.09965	.23996
	Narrow Span	74	2.0068	.61233	.07118
Supply Chain Practices	Broad Span	21	3.3535	.95013	.20734
	Narrow Span	74	1.9189	.24833	.02887

6. RESULTS AND DISCUSSIONS

To test the reliability of the questionnaire all 26 Likert scaled variables were analyzed using SPSS. The Cronbach's Alpha Value is 0.948 indicated that the data tool was reliable. Three statistical test of hypothesis named 'Independent samples T-test' were used to validate first three proposed hypotheses. The variables '1-firm performance', '2-success factors', and '3-supply chain practice' were hypothesized that there was a difference of opinions between the firms linkage with narrow span or integrated with first tier suppliers or first tier customers and the firms integrated with broad span or 2nd or higher tier of suppliers or customers. The Levene's Test was run to check whether the population variances of broad spanned firms were same as the variances of narrowly spanned firms. The significant values less than 5% indicated that in all the three cases the variances are not same. Therefore T-test for equality of means was checked assuming the variances are not same. All the three significant values of independent sample T test were significant. The firm performance 'market share', 'return on assets', 'overall competitive position' and 'overall customer service level' of broadly spanned firms were significantly higher than narrow spanned firms. It is therefore evidence that the mean are not same for both the population, or the average firm performance of broadly integrated firm was much higher than the firms who integrated narrowly. The key success factors 'sales improvement', 'quality improvement', 'cost reduction', and 'new product development' were emphasized more importantly by the broadly integrated firms. The sig value 0.000 strongly evident of this finding. The significant value 0.000 of the test variable key success factor indicated that broadly spanned firms considered those success factors more important than the firms with narrowly integrated. The output and results generated from SPSS are shown below.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. 2-tailed	Mean Difference	Std. Error Difference	Lower	Upper
Firm Performance	Equal variances assumed	12.048	.001	9.116	93	.000	1.64768	.18075	1.28875	2.00661
	Equal variances not assumed			6.979	24.328	.000	1.64768	.23610	1.16074	2.13463
Success Factors	Equal variances assumed	19.707	.000	8.176	93	.000	1.50515	.18409	1.13958	1.87072
	Equal variances not assumed			6.013	23.625	.000	1.50515	.25030	.98812	2.02217
Supply Chain Practices	Equal variances assumed	116.028	.000	11.781	93	.000	1.43456	.12177	1.19275	1.67637
	Equal variances not assumed			6.853	20.781	.000	1.43456	.20934	.99894	1.87018

To validate the fourth hypothesis a regression model was created using SPSS, using the variable 'firm performance' as dependent variable and the independent variables 'supply chain practice' and 'success factors' as predictors. The ANOVA is significant, that indicate the model is acceptable. The regression coefficients are significant while the intercept is not. This concludes that firm performance cannot be predicted without the supply chain practice and success factors.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.868[a]	.753	.748	.51148

a. Predictors: (Constant), Supply Chain Practices, Success Factors

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	79.008	2	39.504	151.001	.000[a]
	Residual	25.900	99	.262		
	Total	104.907	101			

a. Predictors: (Constant), Supply Chain Practices, Success Factors

b. Dependent Variable: Firm Performance

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.005	.156		0.032	.974
	Success Factors	.587	.085	.581	6.933	.000
	Supply Chain Practices	.422	.107	.330	3.935	.000

a. Dependent Variable: Firm Performance

The equation of the regression model is

Firm Performance = 0.587 (Success Factors) + 0.422 (Supply Chain Practice)

The R-square value is 0.753, showed that the predictors can explain 75.3% variation in firm performance. The coefficients of the predictors are significant and positively associated with firm performance. On one unit change in the success factors the firm performance increase by 0.587 units. Similarly on one unit change in supply chain practice the firm performance increase by 0.422 units.

6. COMMENTS AND CONCLUSION

This study focused mainly to explore different styles of supply chain practices and performance of firms. The firms were classified with respect to the span of their integration in the supply chain. The focus of past researcher was to significantly differentiate the supply chain integration. We have focused in this study the importance of supply chain that should incorporate in the span of integration on the firms' integration in supply chain. Based on the results and finding we can conclude that majority of the firms are not linked with their supply chain broadly. While the firm performances were significantly higher for the firms linked with their supply chain broadly than the firm who link narrowly. The four hypotheses were significant and conclude that Firms with larger group of supply chain partners hold far intense attention and clarity on coalition with its associates than those with a shorter group. Consequently, the larger group, show more elevated amounts of performance attribute in relation to supply chain. The outcomes foreground the significance to supply chain professionals by accepting a larger perspective of the supply chain instead of concentrating exclusively on first tier customers and suppliers.

REFERENCES

1. Anderson, B.A. (1999). Buyer-supplier relations in industrial markets: when do buyers risk making idiosyncratic investments? *Organization Science*, 10(4), 460-81.
2. Bagchi, H.S.-L. (2005). Supply chain integration: a European study. *International Journal of Logistics Management*, 16(2), 275-94.

3. Droge, C.J. (2004). The effects of internal versus external integration practices on time based performance and overall performance. In *Journal of Operations Management* (22, 557-73).
4. Fisher. (1997). *What is the right supply chain for your product.*
5. Flores, O.-K.A. (2002). The integration of manufacturing and marketing/sales decisions: impact on organizational performance. *Journal of Operations Management*, 20(3), 221-40.
6. Hamel, P.A. (1994). Strategy as a field of study: why search for a new paradigm. *Strategic Management Journal*, 15, 5-15.
7. Hoffman (2006). Mixing it up: Wal-Mart's latest strategy for delivering the goods promises to shake up other supply chains. *Journal of Commerce*, 22, 64-8.
8. Kim (2006a). The effect of supply chain integration on the alignment between corporate competitive capability and supply chain operational capability. *International Journal of Operations & Production Management*, 26(10), 1084-107.
9. Lee (2007). Relationship between supply chain performance and. *Supply Chain Management: An International*, 12(06), 444-452.
10. Lee, C., Kwon, I. and Severance, D. (2007). Relationship between supply chain performance and degree of linkage among supplier, internal integration, and customer. *Supply Chain Management, An International Journal,* 12(6), 444-452.
11. Lee, P.A. (1997). The bullwhip effect in supply chains. *Sloan Management Review*, 38(3), 93-102.
12. Mentzer, J. (2001). *Supply Chain Management*. London: Sage Publications.
13. Porter (1985). *Competitive Advantage: Creating and Sustaining Superior Performance.* Free Press, New York, NY.
14. Power. (2005). Supply chain management integration and implementation: a literature review. *Supply Chain Management: An International Journal,* 10(4), 252-63.
15. Rosenzweig, R.A. (2003). The influence of an integration strategy on competitive capabilities and business performance: an exploratory study of consumer products manufacturers. *Journal of Operations Management*, 21, 437-56.
16. Sezen. (2008). Relative effects of design, integration and information sharing on supply chain performance. *Supply Chain Management: An International Journal*, 13(3), 233-40.
17. Tan, K.A. (2006). Buyer-supplier relationships: the impact of supplier selection and buyersupplier engagement on relationship and firm performance. *International Journal of Physical Distribution & Logistics Management*, 36(10), 755-75.
18. Van De Vaart, T. and Van Donk, D. (2008). A critical review of survey-based research in supply chian integration. *Internation Journal of Production Economics*, 111, 42 - 55
19. Vickery, S., Jayaram, J., Droge, C. and Calantone, R. (2003). The effects of an integrative supply chain strategy on customer service and financial performance: an analysis of direct versus indirect relationships. *Journal of Operations Management*, 21, 523-39.
20. Vijay, R. and Tan, K. (2010). Cluster analysis of the impact of span of integration. *Supply Chian Management - An International Journal*, *15*(3), 207-15.

Proc. 12th International Conference on Statistical Sciences
Karachi, Pakistan – March 24-26, 2014, Vol. 26, pp. 501-506

STATISTICAL ANALYSIS OF VIOLENCE AGAINST WOMEN IN SOUTHERN PUNJAB

Rashda Qazi
Indus International Institute, D.G. Khan, Pakistan
Email: rashida_qazi@yahoo.com

ABSTRACT

This study belongs to sex discrimination, torture and sensitive issues of women. This is made cleared in this study that torturing attitude and illiteracy for women (special case as in southern Punjab) is not according to teachings of Islam. In spite of this, Islam has revealed the dignity of women in surah Nisa and Surah Maryam. The burning issues like "Kala Kali", acid attack, "Nikah with Quran", Sexual, mental and physical torture are also part of this report. As a sample the area of Southern Punjab and its report of 2012 is also presented.

KEYWORDS

Torture, sex discrimination, literacy, dignity, Kala Kali.

INTRODUCTION

Pakistan is a patriarchal society where men are the primary authority and women are subordinate. This has serious implications on women's health and social development as well as on men's health, behavior, interpersonal relationship and violence. Men role rests in perception that they are responsible for family honor and they are considered accountable for women's action in family. Man is associated to be bread earner and protector of the family. In a family unit, father plans future of his children and make major decisions of their education, profession and marriage. In Pakistan, men as a caregiver are never given attention to ensure health promotion and gender equality.

Women are 52 % of total population of Pakistan. In which 60 % women live in rural areas. Women can play an important role in country's development. However rural women are still living below poverty line as they are not being paid as per the labors standards of Pakistan. They do not have any awareness regarding their basic and civil rights. They are deprived of their basic rights. In villages, early age marriage is very common. Early age marriage of girls is not only violation of their right to education but at the same time a girl is deprived of her right to childhood , health and choosing husband of their own choice. Child marriage is a violation of human rights and is prohibited by a number of international conventions and other instruments naming Universal Declaration of Human Rights-1948, Convention on the Rights of the Child-1989, Child marriage impedes the Millennium Development Goals. The world bank reports that the average Pakistani boy receives only five years of schooling; the average girl just 2.5 years.

LITERATURE REVIEW

Violence is not only a women's issue, or even a health issue, but an issue that influences the development process as a whole. Domestic violence costs development, obstructs participation in development processes and contradicts the goals of development. Islam demands women's active participation in every field of life & gender balance society.

Women are three times more likely to die violently if there is a gun in the house. Usually the perpetrator is a spouse or partner, often with a prior record of domestic abuse [1].

In Pakistan, domestic violence is considered a private matter, as it occurs in the family, and therefore not an appropriate focus for assessment, intervention or policy changes. [3] Women have to face discrimination and violence on a daily basis due to the cultural and religious norms that Pakistani society embraces.[4] According to an estimate, approximately 70 to 90% of Pakistani women are subjected to domestic violence.[5] Various forms of domestic violence in the country include physical, mental and emotional abuse. Some common types include honor killing, spousal abuse including marital rape, acid attacks and being burned by family members. Spousal abuse is rarely considered a crime socially unless it takes an extreme form of murder or attempted murder which could range from driving a woman to suicide or engineering an accident (frequently the bursting of a kitchen stove).

Male dominance is frequently mentioned as a determinant of the domestic violence. [2] Decision making authority makes the man more dominant in the family and society and increases the likelihood of violence against women.[7] Pakistani society is a patriarchal society in which, male members who bear the decision-making authority, head the families. Women are usually not included in making decisions and are considered socially and economically dependent on men.[8] Women "consider themselves insecure, incomplete, ineffective and inefficient without males".[9] Therefore, the male dominance becomes one of the significant predictor of the violence against women in the country. Marital conflict has been found to be related with the domestic violence.[2]

Feminist theory views social phenomena as determined by the patriarchal structure of most societies. According to this theory woman abuse is one of the outcomes of a structure that allows prostitution and other sexist restrictions to keep women in servile positions. The feminist view also holds that until women are seen as other than subservient, compliant victims, little will change. It is a deeply embedded social problem that has to be addressed by social change.[6]

By examining the literature on the dynamics of violence against women it appears that factors like ideology of patriarchy, culture and society, religion, media and individual characteristics come together to explain violence and discrimination against women.

METHODOLOGY

In this study, secondary data is used for analysis. The data is obtained from various survey reports of AWAZCDS (an NGO working for women rights etc.). The focused area is D.G. Khan and Rajanpur. The variables (violence and year) are used for regression analysis. The related data is obtained by survey Reports of famous NGO AWAZCDS. The data is represented in the following table:

Case Type	Number of Cases				
	2008	2009	2010	2011	2012
Rape Case	384	279	520	286	329
Abduction Case	384	416	1041	687	606
Acid Attack Case	31	13	28	22	53
Burning Case	17	14	30	14	17
Violence Case	30	35	178	90	216
Divorce Case	67	20	19	7	88
Honour Killing Case	52	56	84	124	122
Kala Kali Case	19	18	17	8	27
Punchait Case	3	4	4	2	44
Vani Case	13	17	17	12	47
Watta Satta	5	7	16	6	26
Total Cases	**903**	**879**	**1954**	**1258**	**1575**

Statistical Analysis Report (Correlation, Mean and Standard deviation)

Correlations

		Rape	Abduction	Acid Attack	Burn Case	Violence	Divorce	Honour Kill	Kala Kali	Vani	Punchait	Watta Satta
Rape	Pearson Correlation	1	.904*	.130	.616	.980**	-.198	.070*	.073	-.061**	-.032	.433
	Sig. (2-tailed)		.035	.835	.268	.003	.750	.911	.908	.923	.959	.467
	N	5	5	5	5	5	5	5	5	5	5	5
Abduction	Pearson Correlation	.904*	1	.099	.675	.830	-.377*	.418	-.203	.045	-.037	.396*
	Sig. (2-tailed)	.035		.874	.211	.082	.532	.483	.744	.943	.953	.509
	N	5	5	5	5	5	5	5	5	5	5	5
Acid Attack	Pearson Correlation	.130	.099	1	.731	.134	.848	.490	.719	.816	.882	.824
	Sig. (2-tailed)	.835	.874		.161	.829	.069	.402	.171	.092	.048	.086
	N	5	5	5	5	5	5	5	5	5	5	5
Burn Case	Pearson Correlation	.980**	.830	.134	.523	1	-.123**	-.066	.102	-.160	-.094	.356**
	Sig. (2-tailed)	.003	.082	.829	.366		.844	.915	.870	.797	.880	.557
	N	5	5	5	5	5	5	5	5	5	5	5
Violence	Pearson Correlation	.616	.675	.731	1	.523	.309	.678	.442	.740	.712	.931
	Sig. (2-tailed)	.268	.211	.161		.366	.613	.208	.456	.153	.178	.021
	N	5	5	5	5	5	5	5	5	5	5	5
Divorce	Pearson Correlation	-.198	-.377	.848	.309	-.123	1	.040	.851	.602	.759	.551
	Sig. (2-tailed)	.750	.532	.069	.613	.844		.949	.068	.283	.137	.335
	N	5	5	5	5	5	5	5	5	5	5	5
Honour Kill	Pearson Correlation	.070	.418	.490	.678	-.066	.040	1	-.092	.716	.530	.528
	Sig. (2-tailed)	.911	.483	.402	.208	.915	.949		.883	.173	.358	.360
	N	5	5	5	5	5	5	5	5	5	5	5
Kala Kali	Pearson Correlation	.073	-.203	.719	.442	.102	.851	-.092	1	.601	.783	.721
	Sig. (2-tailed)	.908	.744	.171	.456	.870	.068	.883		.284	.117	.169
	N	5	5	5	5	5	5	5	5	5	5	5
Vani	Pearson Correlation	-.061	.045	.816	.740	-.160	.602	.716	.601	1	.966	.851
	Sig. (2-tailed)	.923	.943	.092	.153	.797	.283	.173	.284		.007	.067
	N	5	5	5	5	5	5	5	5	5	5	5
Punchait	Pearson Correlation	-.032	-.037	.882*	.712	-.094	.759	.530	.783*	.966	1	.884
	Sig. (2-tailed)	.959	.953	.048	.178	.880	.137	.358	.117	.007		.046
	N	5	5	5	5	5	5	5	5	5	5	5
Watta Satta	Pearson Correlation	.433	.396	.824	.931*	.356	.551	.528	.721	.851	.884*	1
	Sig. (2-tailed)	.467	.509	.086	.021	.557	.335	.360	.169	.067	.046	
	N	5	5	5	5	5	5	5	5	5	5	5

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

DESCRIPTIVE STATISTICS

	Mean	Std. Deviation	N
Rape	339.2000	103.10044	5
abduction	626.8000	264.10547	5
acid_attack	29.4000	14.87616	5
burn_case	18.4000	6.65582	5
violence	109.8000	84.09043	5
divorce	40.2000	35.22357	5
honour_kill	87.6000	34.59480	5
kala_kali	17.8000	6.76018	5
vani	23.2000	13.68211	5
punchait	11.4000	18.24281	5
watta_satta	12.0000	8.97218	5

CONCLUSION

The results show that violence cases in Pakistan are increasing yearly. Although Pakistan is a Muslim country and Islam strictly prohibits sex discrimination and violence. Still there is prevailing violence and aggression for women. There is a need to formulate society and empower women to save the country and society. The Pakistani laws must continue to discriminate against women directly, which include Citizenship Act, Qanun-e-Shahdat and indirectly PPC, Crpc. The harmful anti-women customary practices continue unabated despite existing laws, besides there is lack of knowledge, understanding of new laws and positive interpretations of older provisions which makes it difficult to effectively implement the law on the ground level. The different modes of trafficking including forced marriages, child marriages, bonded labor, child labor are not recognized and dealt with in any law, besides data collection on trafficking is weak and not integrated into other institutions, There is little or no focus on protecting, rehabilitating and repatriation of victims of trafficking. The weak legislation, inadequate resources, lack of political will and compartmentalization has been a major barrier in combating human trafficking in Pakistan.

RECOMMENDATIONS

There are some recommendations which can be applied to stop violence:

1. Govt. should make laws and enforce laws against women trafficking.
2. There should be local level programs for awareness of women about their rights and laws.

3. Govt. should give full protection to effected women and should develop industries to provide skills to such women.
4. The interface between law reforms to punish violence and reform of the penal codes criminal procedure and evidence laws and family laws must be recognized.
5. Human rights education should be compulsory for every citizen.
6. Enforcement of minimum age of marriage legislation and prohibition of forced and child marriage must receive priority.

REFERENCES

1. AWAZ Foundation Pakistan's Policy Research & Advocacy Unit (2009). Survey Report on "*Disarming Domestic Violence Campaign 2009*" with the International Action Network on Small Arms (IANSA) in 4 Districts of Southern Punjab i.e. Multan, Muzaffar Garh, Jampur and DG Khan.
2. Heise, L.L. (1998). Violence against women: an integrated, ecological framework. *Violence against Women*, 4, 262-90.
3. Fikree, F.F. and Bhatti, L.I (1999). Domestic violence and health of Pakistani women. *International J. Gynaecol. Obstet.*, 65, 195-201.
4. Bettencourt, A. (2000). Violence against women in Pakistan. Human Rights Advocacy Clinic; *Litigation Report Spring* 2000.
5. Human Rights Watch (1999). Crime or Custom? Violence against Women in Pakistan, *Report of Human Rights Watch*. Available from URL: http://www.hrw.org/ reports/1999/ Pakistan /index.htm
6. Gondolf, E.W. and Fisher, E.R. (1988). *Battered women as survivors: An alternative to treating learned helplessness*. Lexington, MA: Lexington Books.
7. Jejeebhoy, S. (1998). Associations between wife beating and fetal and infant death: Impressions from a survey in rural India. *Stud. Fam. Plann.*, 29, 300-8.
8. Fikree, F.F. and Pasha, O. (2004). Role of gender in health disparity: the South Asian context. *BMJ*, 328, 823-6.
9. Niaz, U. (2003). Violence against women in South Asian countries. *Arch Women's Ment. Health*, 6, 173-84.