Original Article

FORECASTING OF MONTHLY PATIENT VOLIUME AT RADIOLOGY DEPARTMENT COMING FOR ULTRASOUND: A TIME SERISE APPROCH.

Khubaib Shahid[*], Tariq Manzoor[**], Muhammad Ibrahim[***], Tanveer Ahmed[****], Muhammad Fiaz[*****]
[*]Professor, Head of Radiology Fatima Memorial Hospital, Shadman, Lahore.
[**]Assistant Professor Medical Imaging, Department of Allied Health Sciences, FMH college of Medicine &Dentistry Shadman Lahore
[***]Associate Professor, Department of Statistics, Govt. MAO College, Lahore.
[****]Assistant Professor Radiology LGH
[*****]Consultant Radiologist Alrazi Health care Diagnostic Centre Lahore

ABSTRACT:

BACKGROUND & OBJECTIVE: Prediction of clinical events/occurrences, patients' volume at OPD, emergency overcrowding, and stay length in hospital for treatment and bed occupancy in wards is considered an important feature for future planning of hospital/clinical management. Patients incoming at emergency and OPD has drawn considerable attention since last two decades. This study was conducted to forecast the patient's volume at Radiology Department of FMH hospital, Shadman Lahore, coming for Ultrasound. Time series approach was used to analyze the data.

METHOD: Time series monthly data of patients coming for ultrasound at radiology department, from January 2001 to May 2015 was used for fitting the best model.

RESULT & CONCLUSION: The ARIMA (12, 1, and 12) model for patient volume was found appropriate, after residuals diagnostic checks. The ARIMA model was used to forecast the patient's volume from June 2015 to May 2017 by applying software GRETL. The actual patient's volume for May 2015 is 5482 and the predicted value for same time is 5422. The predicted error is 1.1%. It is now concluded that the fitted ARIMA is adequate model and can be used to forecast the patient's incoming to Radiology Department for future planning and management.

KEYWORDS: Ultrasound, Medical Time series data, ARIMA model, Box-Jenkin Methodology, Forecasting.

ACKNOWLEDGMENTS: Author's team is thankful for the cooperation of Mr. Khawaja Azhar Mehboob, and staff of Record office, and Staff of Radiology Agha Zulfiqar, Mrs Safia and Transcriptionists for their cooperation and coordination in data collection.

INTRODUCTION:

Overcrowding at various hospitals' departments is reported frequently in Pakistan like around the world resulting in increasing care cost, developing stress in medical staff, lengthy waiting time in providing medical facilitation and overall poor health care massage from patient towards health care system. Prolonged waiting time for having health facility creates frustration and dissatisfaction in patients and probably patient leaves the clinic without treatment as waiting times increases (1, 2)
Radiological diagnostic system is an important part of patient/health care system in hospital and clinics like other departments; this department also faces the problem of overcrowding. Both patients and medical staff face a lot of problem as consequences of

Corresponding Author:
Tariq Manzoor, Assistant Professor, Medical Imaging,
Department of Allied Health Sciences,
FMH college of Medicine &Dentistry Shadman Lahore
Email: tariq_manzoor2001@yahoo.com

overcrowding. Generally hospitals and related wards, departments, diagnostic labs, offices, OPD are built to cover the present requirements but these requirements become insufficient in future. Forecasting of reliable future requirement may provide useful information to hospital administration for expansion to manage the patient/health care system in more efficient and effective way (3). Various researchers used traditional statistical methods like regression analysis to forecast the clinical variables/factors such
as bed occupancy, patient volume at OPD, number of daily surgical procedures in hospital. Box – Jenkin in 1976 developed (4,5) forecasting techniques which is more reliable than conventional regression analysis. Box-Jenkin methodology is used for forecasting when we have time series data .A time series data is set of observation regarding any phenomena which is collected over time at regular interval like hourly, daily, monthly, yearly data of patients . Time series modal like ARIMA, SARIMA are used to forecast the patient volume at various hospital wards and departments (6,7,8).
Joneses (2008) used SARIMA model to forecast the daily patient volume in emergency department and he concluded that SARIMA model is more reliable approach to forecast the patient volume (9) Boyle (2012) used ARIMA model to forecast the emergency department admission (10). Schweigler (2009) applied ARIMA model to forecast the hourly bed occupancy at emergency department (11), Reis (2003) Carried out a study to model the syndromic surveillance data and concluded that ARIMA model is an accurate and reliable method for prediction of medical data (12) This study aims to develop a time series model for forecasting the monthly volume of patients visiting radiological department for ultrasound at FMH Hospital Lahore.

MATERIAL & METHODS:

Retrospective time series data from January 2001 to May 2015 of patients who visited hospital for ultrasound assessment / evaluation / diagnosis was taken from department of radiology FMH Lahore. FMH is a tertiary care hospital which is located at Shadman Lahore and about half Million patients annually come here for treatment.
Time series ARIMA model was used to forecast the number of patients at Radiology Department for Ultrasound. The ARIMA model which was developed by Box-Jenkin in 1976 and still it is very popular method both in Economies/Econometrics and Medical Sciences. In the beginning it was developed to predict economic data. But with the passage of time it is also used in other fields and
now it is also very popular to predict Patient data. The ARIMA model is abbreviated from Auto Regressive Integrated Moving Average. The ARIMA model is set of three processes auto regression (AR) moving average (MA) and difference (I) to eliminate the trend in data. The general ARIMA model is written as ARIMA (p,d,q) where "p" is order of AR, "q" is order of MA and "d" is order of difference (4,5) The ARIMA model is written mathematically as when d=0

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots\ldots\ldots\ldots + \Phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots\ldots\ldots\ldots + Z_q \theta_{t-ij}$$

The ARIMA model is used for integrated stationary process and is best when time series is non-stationary, the non-seasonal ARIMA is classified as ARIMA(p,d,q), the process Xt is said to be ARIMA if

$$W_t = \nabla^d X_t = (1-B)^d X_t$$

$$W_t = \alpha_1 W_{t-1} + \alpha_2 W_{t-2} + \ldots\ldots\ldots + \alpha_p W_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots\ldots\ldots + \theta_q Z_{q-\dot{C}}$$

We can say the process is stationary after differencing the process "d" times, normally in practice the "d" is taken one. The ARIMA model building process which also known as Box-Jenkin methodology involves the following steps:

a) Model identification
b) Parameter estimation
c) Model checking for adequacy
d) Forecasting
In first step, the order of respective time series model determined is based on the behavior or trend of the data. The pre-condition for building of model is that, the data is stationary or not. To check the stationarity of data we have different statistical tool like line chart, correlogram and

ADF/PP unit root test. The ARIMA model is best considered when data is stationary (this means that arithmetic mean, variance an auto correlation are constant over time at appropriate lag). Correlogram and ADF (unit root test) test used to check the stationarety of data (13,14)

The correlogram also provides more information about the trend/ behavior in time series data. ACF and PACF provide us tentative orders of ARIMA model. The correlogram provides many orders of the ARIMA model in the form of suggestions but final model is selected on the basis of various criteria such as AIC,BIC and R2 (15,16,17) . The AIC and BIC are the measures of goodness of fit of an ARIMA model. For different models the values of AIC and BIC are ranked and the model is selected with lowest value of AIC and BIC with highest value of R2. The next step after identification of model is the estimation of parameters. The method of maximum likelihood is used to estimate the parameters. After estimation of parameter next is to check the adequacy of model this is also known as model diagnostic. A model is adequate if the R2 is high and un-explained part of data or residuals is small or least. A Q-statistic is then used for such purpose which was introduced by Ljung-Box is 1979 (18). The ultimate goal at ARIMA model is to forecast the variable of interest. As all diagnostic checks have passed, then model is now adequate for forecasting.

The data of monthly patient's volume at Radiology Department FMH Hospital, Lahore from January 2001 to May 2015 was acquired for modeling. Computer software E-view 5 (Quantitative Micro Software (QMS), USA) and GRETL (Freeware software) was used to analyze the time series data.

RESULT

The model identification procedure begun by investigating data plot, correlogram, ACF and PACF, the original time series ultrasound data of patients were plotted to observe the behavior or trend in the data, the graph of data at level is shown in Fig 1 and at first difference in Fig 2. The Fig 1 showed a clear trend in data and non-stationary behavior of data, this non-stationary behavior can be observed in correlogram of data in Fig.3 where all spikes of ACF are beyond the 95% confidence interval. The time plot of data at first difference (d=1) showed that the data became stationary it means that there is no trend in data, also correlogram in Fig.4 supported that data is

stationary at difference 1 (d=1). A very popular statistical test to confirm the stationarity in data is ADF unit root test (13). Results are shown in Table.1 confirmed that data is non-stationary at level (p-value >0.05) and data is stationary at first difference (p value <0.001) under the null hypothesis the data is non-stationary. Fig.4 consists of ACF and PACF (Correlogram) for monthly Ultrasound data of patients from January 2001 to May 2015, which helps us in determining the order of autoregressive term (p) and "q" lagged error term, the suggested tentative models at different order "q" and "p" were compared using AIC, BIC and R2 then most appropriate model was selected for forecasting. The criteria for best model is lowest AIC & BIC and highest R2.

The proposed models are reported in Table.2 with their corresponding values of R2, AIC and BIC. Those are ARIMA(1,1,1), ARIMA(3,1,3), ARIMA (7,1,7) and (12,1,12). On the given criteria of best model, the model ARIMA (12, 1, 12) was chosen as most appropriate model with lowest AIC & BIC and highest R2. After selection of most appropriate model, next is to test whether residuals analysis is done well. One important assumption of ARIMA is that residuals follow normal distribution with zero mean, in this case mean of residuals is almost zero (0.01) and follows normal distribution (Fig .5) and graph of residuals is showed that residuals are wondered about zero (Fig .6). As the model is finalized after residual check, next is the estimation of model parameters, we used Maximum Likelihood Estimation (MLE) method, the final model ARIMA (12,1,12) is written as

$$W_t = 25.50 + 0.51 W_{t-12} + 0.15 Z_{t-,,}$$

Using the above fitted model, the patient's volume from June 2015 to May 2017 using software GRETL was forecasted which is summarized in Table . 3. The actual value of patient's volume for May 2015 is 5482 and the

predicted value for same time is 5422 with predicted error 1.1%. We may observe that predicted value is close to true value and the error is adjusted using most advanced statistical technique, however we may observe that the values of next predicted period are reported with 95% confidence interval and all predicted values fall within the confidence interval.

CONCLUSION:

A best model for Patient's data is developed for forecasting on the basis of available tools. The empirical analysis indicated that ARIMA (12,1,12) is best fitted model for patients data for short run and long run forecasting. The estimated number of patients for next 24 months is given in Table .3. The estimated results of model are close to real in sample values which showed that model is fitted well. The explained 63% variation is also accounted for model. This is good indicator of the model. The finding of this study may be useful for future planning, clinical researchers, hospital administration, and Government also, for the welfare of patients and health care system.

1. ACKNOWLEDGMENT

We are highly thankful to Mr. Khawaja Azhar Mehboob and staff of Record office and Radiology Department FMH Lahore, who helped us in providing data.

2. ETHICAL CONSIDERATION

This study was carried out on the patient's data available at Radiology Department records and there was no direct interaction with patient, no patients personal identification was used, so the study has not reviewed by Institutional Review Board (IRB).

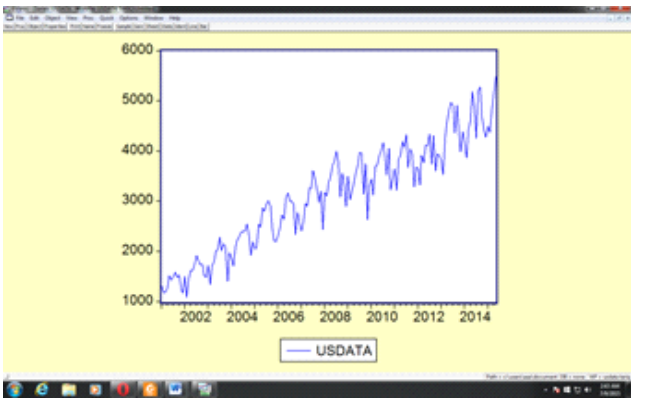


Fig.1 Time plot of the data at level (original data and d=0)

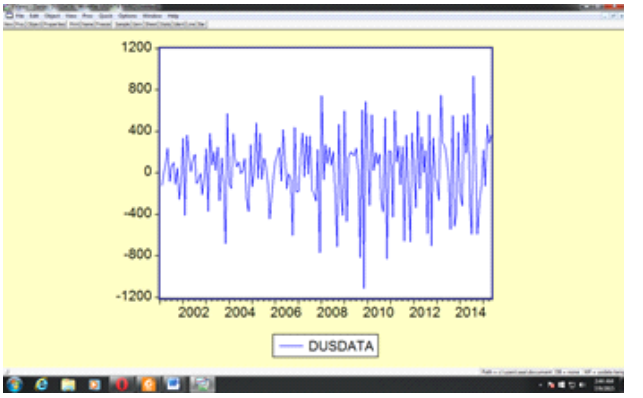


Fig.2 Time plot of the data at first difference (d=1)

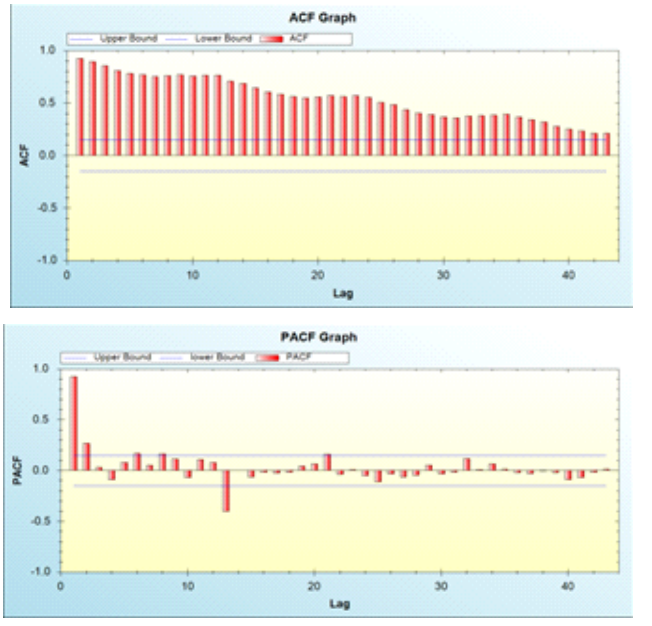
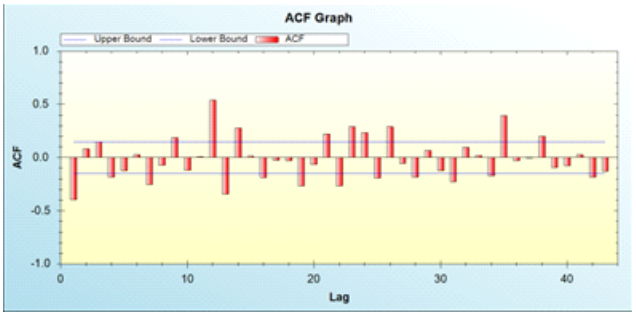


Fig.3 correlogram of the data at level (d=0)



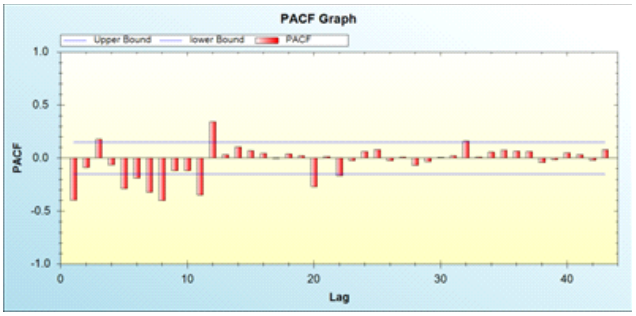


Fig.4 correlogram of the data at first difference (d=1)

Table.1 ADF test confirming stationary behavior of time series data at different level

Variable	ADF test statistic (p value) d=0	ADF test statistic (p value) d=1
US data	-0.365890 (0.918)	-6.299574 (0.000)

Table.2 the summary of criteria for proposed models

Parameters	Proposed models for US Data			
	ARIMA(1,1,1)	ARIMA(3,1,3)	ARIMA(7,1,7)	ARIMA(12,1,12)
R^2	0.16	0.02	0.08	**0.63**
BIC	14.42	14.59	14.55	**14.24**
AIC	14.48	14.64	14.61	**14.30**

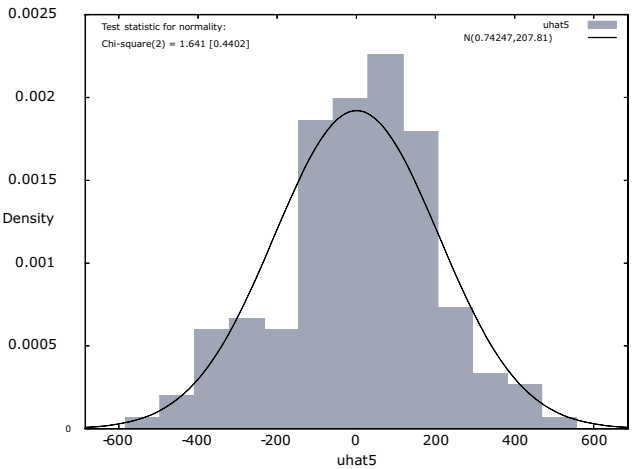


Fig.5 Normal curve of residuals

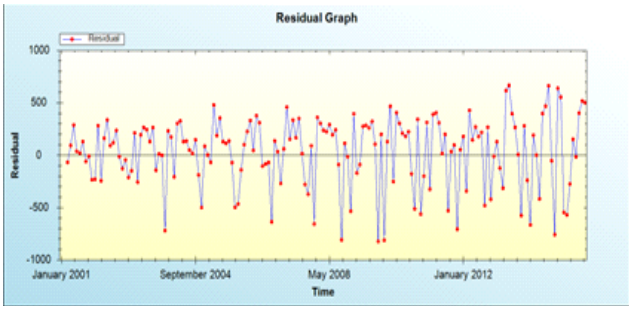


Fig.6 Graph of residuals

REFERENCES

1. Batal H, Tench J, McMillan S, Adams J, Mehler PS. Predicting patient visits to an urgent care clinic using calendar variables. Acad Emerg Med. 2001;8: 48–53.
2. Green LV, Soares J, Giglio JF, Green RA. Using queuing theory to increase

effectiveness of ED provider staffing. Acad Emerg Med. 2006; 13:61–8.
3. Tandberg D, Qualls C. Time series forecasts of emergency department patient volume, length of stay, and acuity. Ann Emerg Med. 1994;23:299– 306.
4. BOX, G.E.R. – JENKINS, G.M., Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco 1976.
5. Box GEP, Jenkins GM, Reinsel GC: Time Series Analysis: Forecasting and Control 3rd edition. Englewood Cliffs, NJ: Prentice-Hall; 1994.
6. FLEMING, D. M. – CROSS, K .W. – SUNDERLAND, R. – ROSS, A.M., Comparison of the Seasonal Pattern of Asthma Identified in General Practitioner Episodes, Hospital Admissions and Deaths. Thorax, 8, (2000), pp. 662–665.
7. SAYNAJAKANGAS, P. – KEISTINEN, T. – TUUPONEN, T., Seasonal Fluctuations in Hospitalisation for Pneumonia in Finland. Int J Circumpolar Health, 60, Num. 1 (2001), pp. 34–40.
8. LANI, L. – RIOS, M. – SANCHEZ, J., Meningococcal Disease in Spain: Seasonal Nature and Resent Changes. Gac Sanit, 15, Num. 4 (2001), pp. 336–340.
9. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. Acad Emerg Med. 2008;15:159–70.
10. Boyle J, Jessup M, Crilly J, et al. Predicting emergency department admissions. Emerg Med J. 2012;29:358–65.
11. Schweigler LM, Desmond JS, McCarthy ML, et al. Forecasting models of emergency department crowding. Acad Emerg Med 2009;16:301e8.
12. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. BMC Med Inform Decis Mak 2003;3:2
13. Dickey, D.A., Fuller, W.A.(1979). Distribution of estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. 74: 427–431
14. Pankratz, Forecasting with Univariate Box-Jenkins Models: Concepts and Cases, Wiley, New York, NY, 1983.
15. Sakamoto, Y., Ishiguro, M., and Kitagawa G. (1986). Akaike Information Criterion Statistics. D. Reidel Publishing Company.
16. Akaike H. On the likelihood of a time series model. Statistician 1978; 27: 217-35.
17. Schwarz, G. E. (1978). Estimating the Dimension of a Model. Annals of Statistics 6 (2): 461–464.
18. Ljung G.M, Box G.E.P. On a measure of lack of fit in time series models. Biometrika 1978; 65: 297-303

SR #	AUTHOR NAME	CONTRIBUTION
1	Dr. Khubaib Shahid	Main idea of study and instructions on plot of study
2	Dr. Tariq Manzoor	Prepared over all write up of study
3	Dr. Muhammad Ibrahim	Data analysis and interpretation over all
4	Dr. Tanweer Ahmad	Assisted in write up and interpretation of data
5	Dr. Muhammad Fiaz	Assisted in write up and interpretation of data

Submitted for publication: 08.10.2015
Accepted for publication: 20.06.2016