International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878
Volume 3 Issue 1, January 2015

Fitting an Arima Model to a Poisson Process

J. B. Ehimony

Department of Mathematics and Statistics, Kogi State Polytechnic, P.M.B. 1101 Lokoja, Kogi State Nigeria

Abstract: ***The Autoregressive Integrated Moving Average (ARIMA) is normally used to fit data that are collected over time space in a stochastic process. The univariate Box- Jenkins Arima model technique was used to fit an appropriate model to the data set from two independent stochastic processes observed from a Poisson experiment. The fitted model to the count data help us to understand on how to generate a series of counted events within a time space and also to study the similar pattern and behavior of the random process observed during the analysis.***

Keywords: ARIMA, Stationary, Non-Stationary, Difference, Poisson Process

1. Introduction

Autoregressive Integrated Moving Average (ARIMA) process is time series processes that behave as though they had no fixed mean. The behavior of this processes are homogeneity in nature in the sense that the trend in one part of the series behaves much like any other part of the processes. Thus, an ARIMA model helps to describe the non-stationary behavior that can be differenced to obtain a stationary process, by fitting data that are collected over time space. The main objective of this research of work is to draw inferences from the series of observation collected from a Poisson processes base on average customers waiting time observed from two automated teller machines (ATM) by comparing their average waiting time . For this to be achieve, we set up a hypothetical probability model to test the estimated parameters of the fitted model, check for adequacy of the model using goodness of fit to the data. In this paper, our focus is to find a good model that will be appropriate to represent the behavior of counting data. The univariate Box-Jenkins ARIMA modeling approach (Box and Jenkins, 1976) along with the model building approach of box and Hunter (1978) will be used to find an appropriate model to the count data series.

2. Univariate Box-Jenkens-Arima Model

The UBJ-ARIMA method applies only to stationary data series. A stationary time series has a mean variance and autocorrelation function that are essentially constant through time. The stationary assumptions helps to simplifies the theory of UBJ-ARIMA model, when one which to estimates the parameters from a moderate number of observation. If a time series is stationary, then the mean and variance of any major subject of these should not differ significantly from the mean and variance of any other major subject of the time. Most time series data are non-stationary nature; its mean does not vary about a constant mean. It exhibit homogenous behavior in nature.

A model that exhibit homogenous behavior non- stationary behavior can be defined as using Box-Jerkins and Reinsel (1994)

$$\alpha (B) (1-B)^d y_t = \theta B a_t \qquad (1)$$

Where y_t is the response variable at time t

$\alpha (B)^t$ represent the AR process operator,

$\theta(B)$ and the MA process operator, a_t represent the white noise,

B is the backward shift operator and d is the number of times the data series must be differenced to induce a stationary mean – equation (1) is often re-written as

$$\alpha (B) w_t = \theta B a_t \qquad (2)$$

Where $W_t = (1-B)^d y_t$ The operator α (B) and $\theta(B)$ are defined as

$$\alpha (B) = 1- \alpha_1(B) -\ldots\ldots- \alpha_2(B)^p \text{ and } \theta(B)=1-\theta_1(B) - \theta_2(B)^{2+}-\ldots\ldots-\theta_q(B)^q \qquad (3)$$

Therefore, homogenous non stationery behavior can sometimes be represented by a model that calls for the d^{th} hyference of the process to be stationary, variable a is usually 0, 1 or 2 (Pavkartz, 1983). From the above ARIMA model can be defined as an order p, d, q or ARIMA (p, d, and q) process

ARIMA (p, d. p) process can be defined by

$$W_t = \alpha_1 w_{t-1} + \alpha_2 w_{t-2} + \text{------------} + \alpha_1 w_{t-1} + a_2 - \theta_t a_{t-1} - \theta_2 a_{t-2} - \ldots\ldots\ldots\ldots - \theta_q a_{t-q} \qquad (4)$$

Where $W_t=(1-B)^d y_t$

The non-stationary series can sometime transferred to stationary series througt dylevenseing(d).

The backshift operator B can also used in homogenous non- stationary model

$$\alpha (B) w_t = \theta(B) a_t \qquad (5)$$

Where

$W_t=(1-B)^d y_t,$

y_t is the response variable at time t, θ(B) represent the MA process operator and is given by

$\theta(B) = 1- \theta_1 B - \theta_2 B^2$---------- $\theta_q B^q$ and α (B) represent the AR process operator and it is given by

$\alpha(B)= 1- \alpha_1 B - \alpha_2 B^2$-----$\alpha_p B^p$, a_t represents the white noise and d is the number of times the data services must be differenced to induced a stationary mean (Box et al. 1994)

3. Steps in Modeling

The classical method of model identification as described by Box and Jenkins (1970) is judge the appropriate of the plotted autocorrelation function (ACF) and partial autocorrelation function (PACF) The ACF at lag K is the correlation between the observed data, Y_t say$_1$ and Y_{t+k}. this is given by:

$$\rho_k = \frac{E[(Y_t-\mu)(Y_{t+k}-\mu)]}{\sqrt{E[(Y_t-\mu)^2[(Y_t+_k-\mu)^2]}} \quad (6)$$

Where K is the time lag and can take value from 0, 1, 2----- and $\mu = E(Y_t) = E[Y_{t+k}]$

One important use of the ACF in modeling its use to determining whether a series is stationary or not. If the mean of a series is stationary then the estimated ACF's of the series will drop all slowly toward zero.

The partial Autocorrelation function (PACF) of a process Y at lag K, is devoted by $\propto_{kk}$, is defined as the correlation between the adjusted value of Y_t and Y_{t+k}(Box et al. 1994). This can be defined as

$$\alpha_{kk}=\frac{E[(Y_t - Y_t)(Y_{t+k} - Y_{t-k})]}{E[(Y_{t+k} - Y_{t-k})^2]}$$

Where $y_t = \alpha_{k-1.1}y_{t-1} + \alpha_{k-1.2}y_{t-2} +$ --------- $+ \alpha_{k-1.k-1}y_{t-k-1}$ and $y_{t-k} = \alpha_{k-1.1}y_{t-k+1} + \alpha_{k-1.2}y_{t-k-1} + \ldots\ldots\ldots\ldots + \alpha_{k-1.k-1}y_{t-1}$

Note that at identification stage, the estimated ACF and PACF are compared to each other base on their theoretical characteristics to the common the time series model to find a match.

The stage two is checking the models are the estimate coefficient significantly different from zero and the randomness of the residuals. The significance of the ARMA coefficients can be evaluated by comparing estimated parameters with the standards deviations. In this research work, the maximum likelihood approach which has been proved to reflect useful information about the parameter contained in the data is used.

The third stage in model building is diagnostic check. The residual ACF is used as device for testing the independent assumption of the random shock can be defined as

$$\rho_x(a) = \frac{e[(a_i - a)(a_{t+k} - a)]}{E[(a_t - a)^2]} \quad (8)$$

4. Testing of Adequacy of the Fitted Model

When one test for adequacy of the fitted model, the chi-squared test for goodness of fit id used. This is called Ljung-Box test in the literature; see Ljung and Box (1978). The test is based on all the residual ACF as a set.

Given K residual autocorrelations, the hypothesis to be tested is

Statement of Hypothesis

$H_0 : \rho_1(a) = \rho_2 = - - - \rho_k(a) = 0$ vs H_1 not H_0

Test statistics

$$Q = n(n+2)\sum_{k-1}^{R}(n-k)^{-1} r_k^{\,2} \ (a)$$

Or

$$Q = N\sum_{k-1}^{R} r_k^{\,2}$$

This statistic is referred to as the Portmanteau statistics, It follows a χ^2 distribution with (k – p- q) degree of freedom, where p and q are the AR and MA orders of the model and N is the length of the time series. If the computed Q exceeds the value from the χ^2 table for some specified significance level, we null hypothesis that the series of autocorrelation represents a random series is rejected at the level. The P- value gives the probability of exceeding the computed Q by chance alone, given a random series is rejected at that level. This non-random residual s give high Q and small P- value. The significance level is related to the p- value by:

Significance level (%) = 100 (1- P)

Application

The data sets from Poisson process on waiting time of student using automated teller machines (ATM) in kogi state polytechnic Lokoja (Ehimony, 2014) this was conducted to determine the trend of average waiting time in a process.

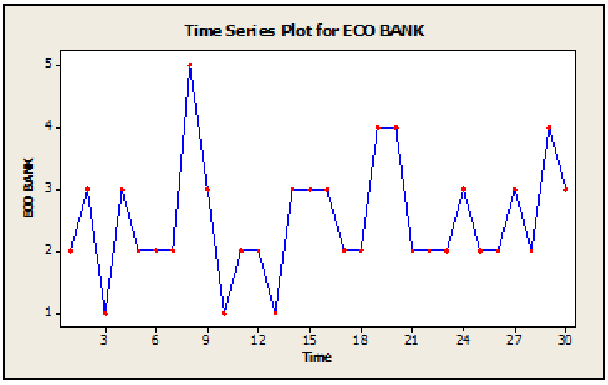
The data were collected over 30 runs from two automated teller machines (ATM) for Eco Bank and United Bank for African (UBA). We observed the average waiting time (in second) for the two machines see table below.

	The observed average waiting times in (sec)
Eco –Bank	2, 3, 1, 3, 2, 2, 2, 5, 3, 1, 2, 2, 1, 3, 3, 3, 2, 2, 4, 4, 2, 2, 2, 3, 2, 2, 3, 2, 3, 2, 4, 3,
UBA average waiting time (in sec)	2, 2, 1, 2, 2, 2, 2, 4, 4, 1, 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 1, 2

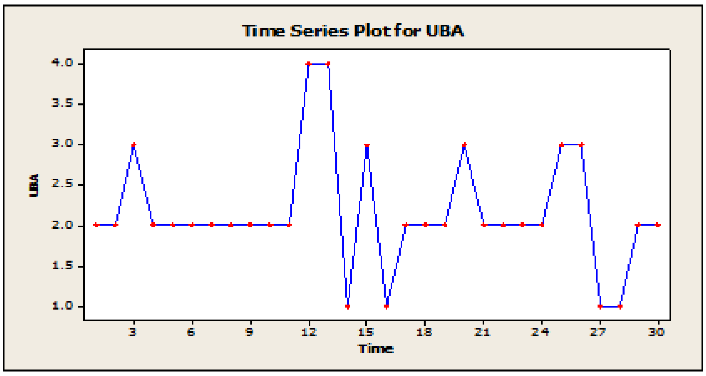
Source: Ehimony (2014)

International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878
Volume 3 Issue 1, January 2015

ECO ATM DATA



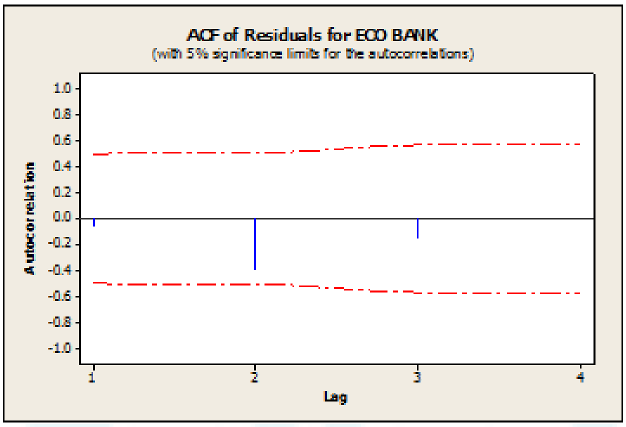
run number

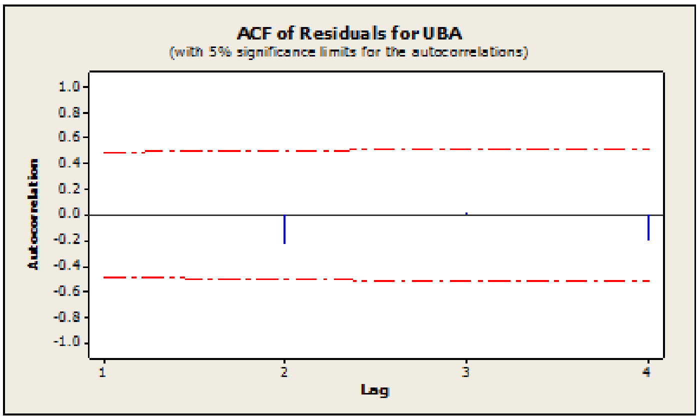


run number

The graphical presentation describe a steady increase in the waiting time which is shown in figure I. the behavior of this series can be model as a time series model, since the behavioral pattern does not exhibit homogenous behavior of any kind, then the series us a non-stationary one.

The current value, say Y_t is been influenced by the lag value Y_{t-1}.there is continuous increase in the waiting time observed from both machine due to uncontrollable factor known as while noise.





Since the original time series $Y_1, Y_2, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots Y_n$ are stationary, we can now look at the sample autocorrelation function (ACF) and partial autocorrelation function(PACF) for a particular behaviors that indicate a non seasonal theoretical Box- Jenkins model. The figures I and II shows the behaviors of the ACF and PACF which out off quickly after lag q for ACF while PACF cuts off quickly after lag p. The cutting off of both p and q shows that the average waiting time of both machines were not same due to uncontrollable force that influences the servers to operate at a steady rate. We need to note that every series that are not stationary are tend to follow first order autoregressive moving average of order p and q with first order series, which can be defined as

$$Y_t = \alpha_t + \alpha_1 y_{t-1} + e_t$$

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	-0.1987	0.2481	-0.80	0.435
Constant	0.3392	0.2560	1.33	0.204

$$Y_t = -0.1987 + 0.3392 y_{t-1} \quad \text{Equation (i)}$$

SE	(0.2481)	(0.256)
T	(-0.8)	(1.33)

From the test of the significant of the estimated parameters of the model, we discovered that the average waiting time of ECO- BANK ATM were not stationary, there were significant difference between the time taken by each customer to withdrawn using ATM card see the model above.

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 30, after differencing 18

Residuals: SS = 18.8598
MS = 1.1787 DF = 16

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	9.8	*	*	*
DF	10	*	*	*
P-Value	0.461	*	*	*

Testing of adequacy of fitted model using Ljung-Box chi-square, there was an evident that the model was sufficient enough to describe the average waiting the customer spend before it can be serve, since $\chi^2(9.8)$>p-value(0.461)see the table of modified Box-pierce above.

ARIMA Model: UBA

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	-0.4296	0.2223	-1.93	0.071
Constant	-0.2137	0.2380	-0.90	0.383

$$Y_t = -0.4296 - 0.2137y_{t-1} \quad \text{Equation (2)}$$

SE (0.2223) (0.2380)
T (-1.93) (-0.90)

The test of significant difference between the parameters in model, indicate that the estimated parameters in the model describe the effectiveness of the UBA ATM machine, the average waiting time were less i.e customer were served in quickly.

Differencing: 0 regular, 1 seasonal of order 12
Number of observations: Original series 30, after differencing 18
Residuals: SS = 16.3076
MS = 1.0192 DF = 16

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	6.6	*	*	*
DF	10	*	*	*
P-Value	0.764	*	*	*

The model was sufficient enough to described fitted model on average waiting time of the customers to be served, since $\chi^2(6.6)$ > p-value(0.764).See modified Box-pierce above.

5. Conclusion

The fitted model for two experimental data from the Poisson process it indicate that the series there were steady increase in the average waiting time due to uncontrollable forces (white noise)in the process. The information will helps to reducing queuing of customers to be served. The model fitted can be used to predict the average number of customers to be served within a short period of time.