

SEASONAL MODELLING OF ROAD TRAFFIC ACCIDENT

I.A.Iwok

Department of Mathematics/Statistics, University of Port-Harcourt, P.M.B.5323, Port-Harcourt, Rivers State; Nigeria.

Abstract

This work modelled road traffic accidents that exhibit seasonal behaviour and non-statioarity in variance. The modelling procedure was demonstrated using monthly road traffic accident data (1999-2014) obtained from federal road safety commission in Port Harcourt, Nigeria. Seasonality and non-stationarity in variance were detected from the series raw plot, autocorrelation and partial autocorrelation functions. To obtain stability in variance and level of the series, some transformation techniques were applied. Seasonal component was incorporated into the Box and Jenkins (1970) ARIMA model to cater for the periodic nature of the series. The resulting estimated seasonal-ARIMA model was subjected to different diagnostic checks, and was found to be adequate. The proposed model was then used to generate forecasts of road traffic accidents for the next thirty months.

Keywords: Stationarity, Seasonality, Transformation, Autocorrelation and Partial autocorrelation.

INTRODUCTION/REVIEW

Globally, road traffic accidents (RTA) constitute one of the leading causes of deaths in human race. According to world health organisation (WHO), RTA are an emerging global epidemic (WHO, 2004). Statistics shows that approximately 1.2 million deaths occur yearly worldwide as a result of road traffic crashes (WHO,2004). A break down of the figure indicate, however, that about 70% of the deaths occur in developing countries of which Nigeria is one. Recent studies have shown that the proportion of deaths from RTA in Nigeria increased from 38.2% to 60.2% between 1991 and 2001 (Atubi, 2012). A publication of the Nigerian federal road safety commission (FRSC) revealed that Nigeria records the highest rate of deaths from motor accidents in Africa; leading 43 other nations in the number of deaths per 10,000 vehicle crashes (FRSC, 2006; Obinna, 2007). This is followed by Ethopia, Malawi and Ghana with 219, 183 and 178 deaths per 10,000 vehicles respectively (Daramola, 2004).

Despite the observed dramatic increase in road traffic deaths, little attention has been paid to its prevention in most developing countries. In developed countries, however, serious effort has been made to reduce the menace of RTA to a considerable level. For instance, the United State and United Kingdom records an average of 1.6 and 1.4 traffic deaths per 1000 people (Trinca et al, 1988). Far in excess, Nigeria records 32 traffic deaths per 1000 people (Filani and Gbadamosi, 2007).

Road traffic accidents are not common to Nigeria alone but a universal hazard. Downing et al (1991) noted that road crashes impact on the economy of developing countries at an estimated cost of 1.2% of the country's gross national product (GDP) per annum as a result of mobidity, mortality and property. It is a well known fact that no country is completely independent in terms of economy. The economic growth of one country is a function of the economic growth of others. Because of its economic importance and the resulting impacts on human lives, road traffic researches should not be over-emphasized. Many researchers have suggested ways of combating the road traffic accidents and some sensitive governments have implemented them. Few of the remedies are construction of good roads and establishment of government agencies in charge of traffic responsibilities.

However, this work differs a bit. In Nigeria, road traffic crashes seem to be seasonal. It appears to be rampant in most month of the year than others. This work therefore seeks to model the road traffic accidents using time series technique, taking seasons into consideration with the aim of generating forecasts for the future. Of course, if the future can be predicted, measures can be taken to avert the situation. The study area of this work is the Port-Harcourt city of Nigeria. Port-Harcourt is the Nigeria's major traffic centre in the south-south region and is considered most suitable for this study.

The ability to model complex seasonal time series greatly increases the applicability and usefulness of autoregressive integrated moving average (ARIMA) model building proposed by Box and Jenkins (1970).

Many time series such as climate, economic, accident etc. are observed to exhibit some periodic and recurrent nature. In such cases, the popular ARIMA models cannot provide good approximations for the true underlying process. Hence the need for a component in ARIMA model that caters for the periodic influences called seasonal ARIMA (SARIMA) model. Series that results from events that are periodic and recurrent are called seasonal.

However, in practice, it may not be reasonable to assume that the seasonality component repeats itself precisely in the same way cycle after cycle (Brockwell, 1986). Seasonal ARIMA models do allow for randomness in the seasonal pattern from one cycle to the next.

In general, we say that a series exhibits periodic behaviour with period s, when similarities in the series occur after s basic intervals.

In most cases, the magnitude of a seasonal swing in the series depends on the level of the time series X_t and thus exhibit an increasing seasonal variation. Time series of this form are said to be non-stationary in variance and requires some transformation techniques in order to produce a transformed series that displays constant seasonal variation, before being modelled. Non-stationarity problems often occur in two forms: either in level or variance. While level stationarity can be achieved by differencing, variance stationarity can be obtained by square root, quartic or logarithmic transformation of the series.

METHODS OF ESTIMATION

Let the time series under consideration be X_t .

(i) **ARIMA Model**

If d is a non-negative integer, then the $\{X_t\}$ is said to be an $ARIMA(p,d,q)$ process if

$Y_t = \nabla^d X_t$ is a causal autoregressive moving average (ARMA) process.

X_t satisfies a difference equation of the form:

$$\phi^*(B)X_t \equiv \phi(B)\nabla^d X_t = \theta(B)\varepsilon_t$$

where B is the backward shift operator, $\nabla = 1 - B$, $B^m X_t = X_{t-m}$, $\{\varepsilon_t\} \sim NIID(0,\sigma^2)$, $\phi(B)$ and $\theta(B)$ are polynomials of degree p and q respectively, $\phi(B) \neq 0$ for $|B| \leq 1$.

The polynomial $\phi^*(B)$ has a zero of order d at $B = 1$.

The process $\{X_t\}$ is stationary if and only if $d = 0$, in which case the $ARIMA(p,d,q)$ reduces to an $ARMA(p,q)$.

(ii) **The SARIMA model**

If d and D are non-negative integers, then $\{X_t\}$ is said to be a $SARIMA(p,d,q) \times (P,D,Q)_s$ process with period s if the differenced process $Y_t = \nabla^d \nabla_s^D X_t$ is a causal $ARMA$ process

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)\varepsilon_t \quad (1)$$

where $\nabla_s = 1 - B^s$,

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_P B^P \text{ and } \Theta(B) = 1 + \Theta_1 B + \Theta_2 B^2 + \cdots + \Theta_Q B^Q$$

are the seasonal autoregressive operator of order P and seasonal moving average operator of order Q added to the $ARIMA$ model $\phi^*(B)X_t$ to cater for the seasonal influence.

Because of the involvement of the seasonal component in the SARIMA model, the covariance structure for a SARIMA process seems to be quite complicated. However, to identify SARIMA models from the sample correlation function, we first find d and D so as to make the differenced observations $Y_t = (1-B)^d(1-B^s)^D X_t$ stationary in appearance.

If the variability of the time series increases as time advances, then we stabilize the variance of the series by using a pre-differencing transformation such as logarithm, quartic root etc. Next, we examine the sample autocorrelation and partial autocorrelation function of $\{Y_t\}$ at lags which are multiples of s in order to identify the orders P and Q. If $\hat{\rho}(.)$ is the sample autocorrelation function of $\{Y_t\}$, then P and Q should be chosen so that $\hat{\rho}(ks)$, $k = 1,2,\dots$ is compatible with the autocorrelation function of an $ARMA(P,Q)$ process. The orders p and q are then selected by attempting to match $\hat{\rho}(1),\dots,\hat{\rho}(s-1)$ with the autocorrelation function of an $ARMA(p,q)$ process.

RESULTS

(i) Detection of Seasonality

Figure 1 (see appendix) shows the raw plot of the accident data X_t . As observed in the figure, the series appears to be seasonal and non-stationary especially in variance. In figures 2 and 3, the spikes at seasonal lags (12, 24, 36, …) in autocorrelation function (ACF) plot and partial autocorrelation (PACF) plot confirms the presence of seasonality.

(ii) Stationarity in Level and Variance

As clearly seen in figure 1, the variance of X_t is unstable and thus requires some transformation to get stabilized. This was achieved by taking natural logarithm of X_t $(lnX_t = X_t^*)$ and the result plotted in figure 4. Close examination of figure 4 shows that the series is only near stationarity, thus demanding for another stabilizing technique. Finally, differencing of $X_t^*(= DX_t^*)$ was applied and stationarity was obtained both in variance and in level (see figure 5).

(iii) The SARIMA model

The ACF plot of the DX_t^* shows an exponential decline at seasonal lags while at non-seasonal lags, it is inconclusive (see figure 6).

Also, the PACF plot of DX_t^* cut off after the first seasonal lag, while at the non-seasonal lags, it is inconclusive (see figure 7). Thus, a seasonal model with non-seasonal differencing is obvious.

Combining the behaviours of ACF and PACF plots at both seasonal and non-seasonal lags, the multiplicative form of the seasonal model is obtained as: $SARIMA(0,1,0) \times (1,0,0)_{12}$.

Writing the model more explicitly, we take cognizance that $DX_t^* = (1-B)X_t^*$, $d = 1$ and $s = 12$.

Thus from (1), we have

$$\Phi_p(B^s)\nabla^d X_t^* = \varepsilon_t$$

$$\Rightarrow \left(1 - \Phi_{1,12}B^{12}\right)(1-B)X_t^* = \varepsilon_t$$

which on expansion gives

$$X_t^* = X_{t-1}^* + \Phi_{1,12}X_{t-12}^* - \Phi_{1,12}X_{t-13}^* + \varepsilon_t.$$

Minitab software was used in the analysis and the result (see table 1) provides the following estimates of the SARIMA model:

$$\hat{X}_t^* = X_{t-1}^* + 0.4988(X_{t-12}^* - X_{t-13}^*)$$

It should be recalled here that the series X_t^* was obtained from the original series X_t by applying natural logarithmic transformation (lnX_t). Hence the resulting estimated SARIMA model for generating forecast is obtained by:

$$\hat{X}_t = e^{\hat{X}_t^*}$$

The above model was used to generate forecasts presented in table 2.

DIAGNOSTIC CHECKS

(i) Raw and Estimate Plot

This is obtained by overlaying the plots of the original data on the plots of the values estimated by the SARIMA model as shown in figure 8. As can be seen clearly, the two plots closely agree with each other indicating a good fit of the model.

(ii) Residual Analysis

Variance and White noise

One of the conditions of model adequacy is that the residuals must follow the white noise process. In other words $\varepsilon_t \sim NIID(0, \sigma^2)$. The SARIMA model was fitted to the data and the calculated residual mean and variance are $0.0001 \approx 0.00$ and 0.2311. The small variance obtained is an indication of a good fit of the model. The plot of the residual is shown in figure 9. A mere inspection of the plot convinces that the residual follows a white noise process. Thus, the fitted model is adequate.

Residual autocorrelation checks

Figure 10 and table 2 illustrate the autocorrelation function pattern of the residuals. As shown, there are no patterns or statistically significant coefficients in the table and the figure. Again this is a confirmation of white noise residuals, indicating the adequacy of the SARIMA model.

Ljung-Box Staistic

Here, the first $M = 20$ autocorrelations of the ε_t's are being considered together for the adequacy of the overall model. Using the Minitab software, the Ljung-Box statistic (LBQ) given by

$$Q^* = n'(n'+2)\sum_{l=1}^{M}(n'-l)^{-1}r_l^2(\varepsilon_t)$$

provides the LBQ values (see table 2). It should be noted that if the fitted model is appropriate, then

$Q^* \sim \chi_\alpha^2(M - n_p)$, where n_p is the number of parameters in the model (Box and Jenkins,1970). On the other hand, if the fitted model is inappropriate, the average values of Q^* will be inflated.

From table 2, the first 20 autocorrelations gives Q^* = LBQ = 29.32.

Setting $\alpha = 0.01$, $\chi^2_{0.01(19)} = 36.19$. Then, under the null hypothesis that the model is adequate, there is no evidence to query the SARIMA model, since $Q^* < \chi_\alpha^2(M - n_p)$. Hence, the model is adequate.

APPENDIX

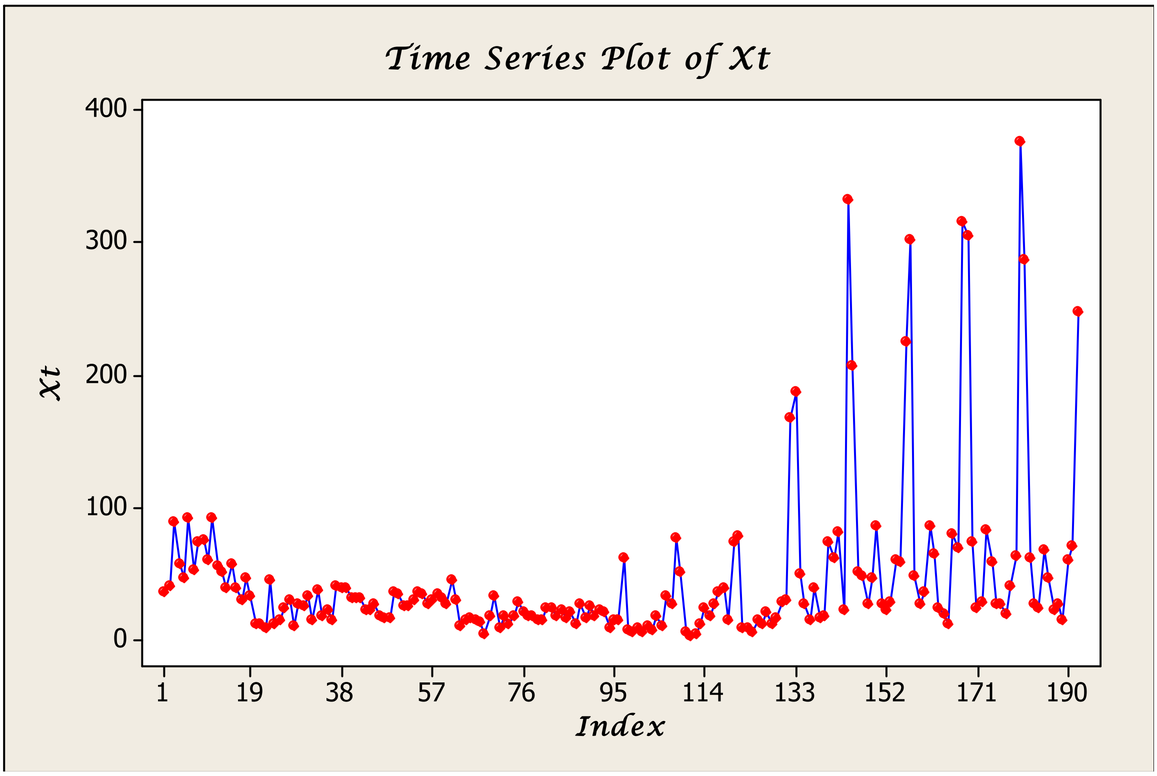

figure 1

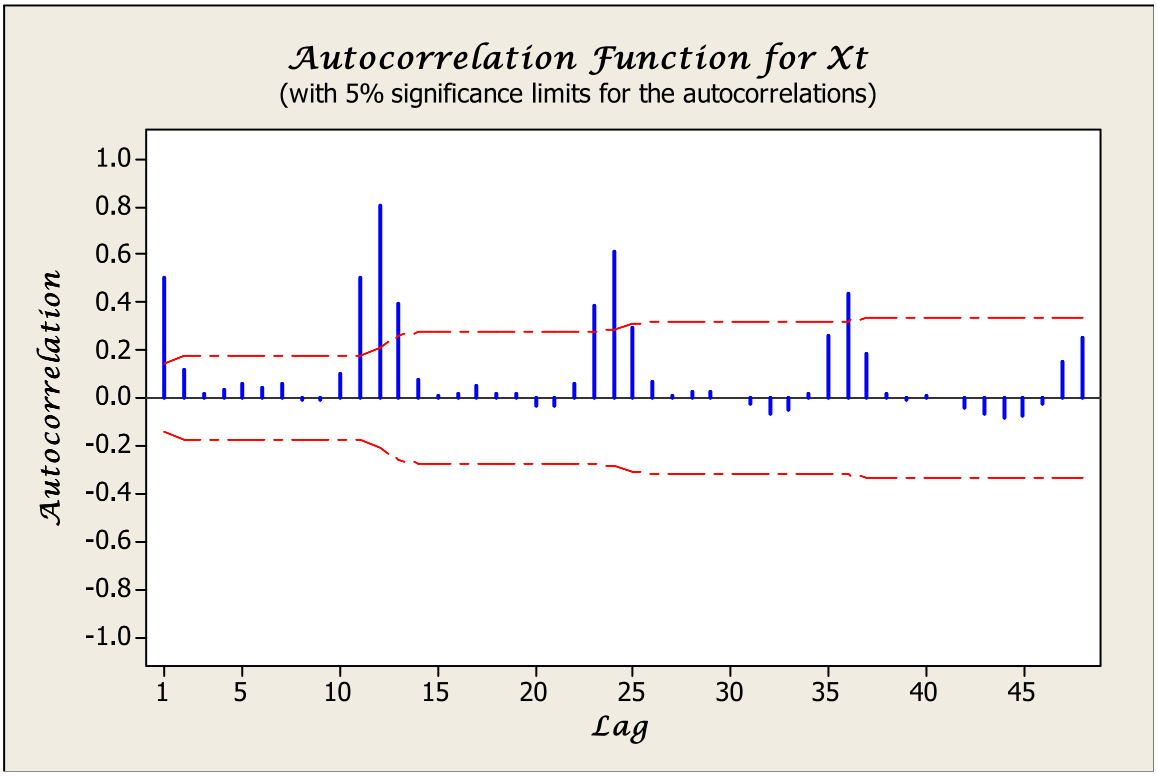

figure 2

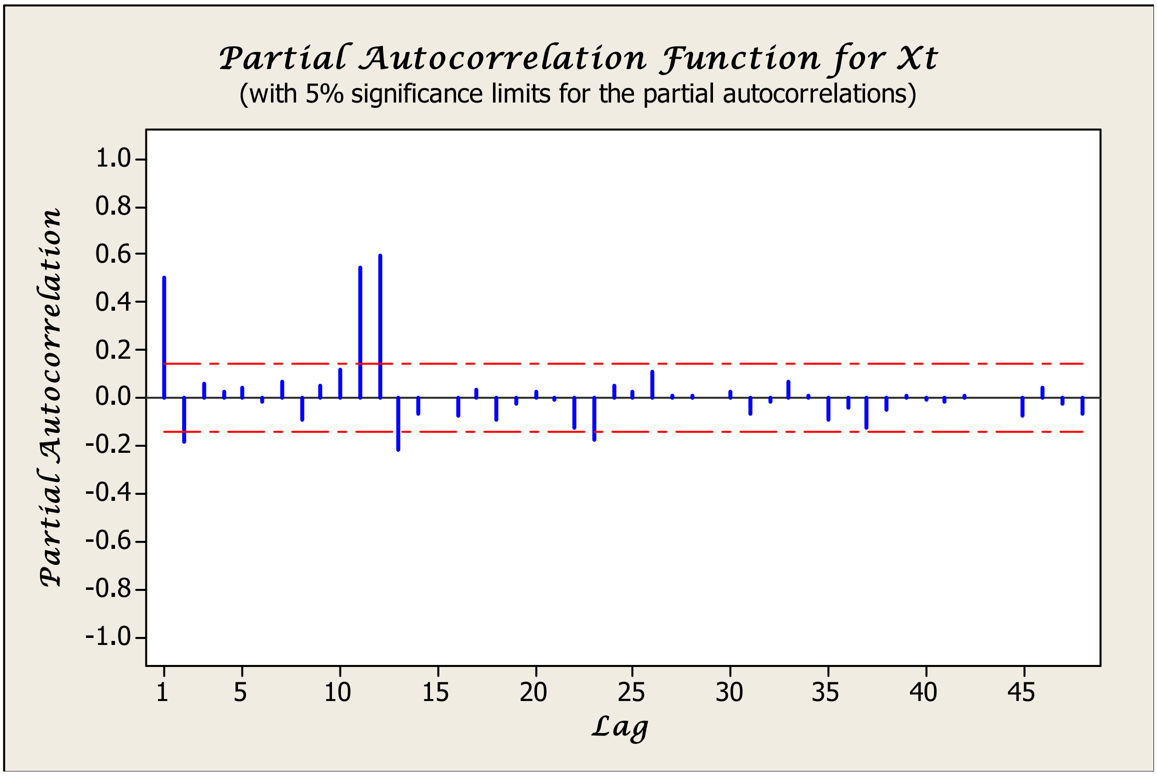

figure 3

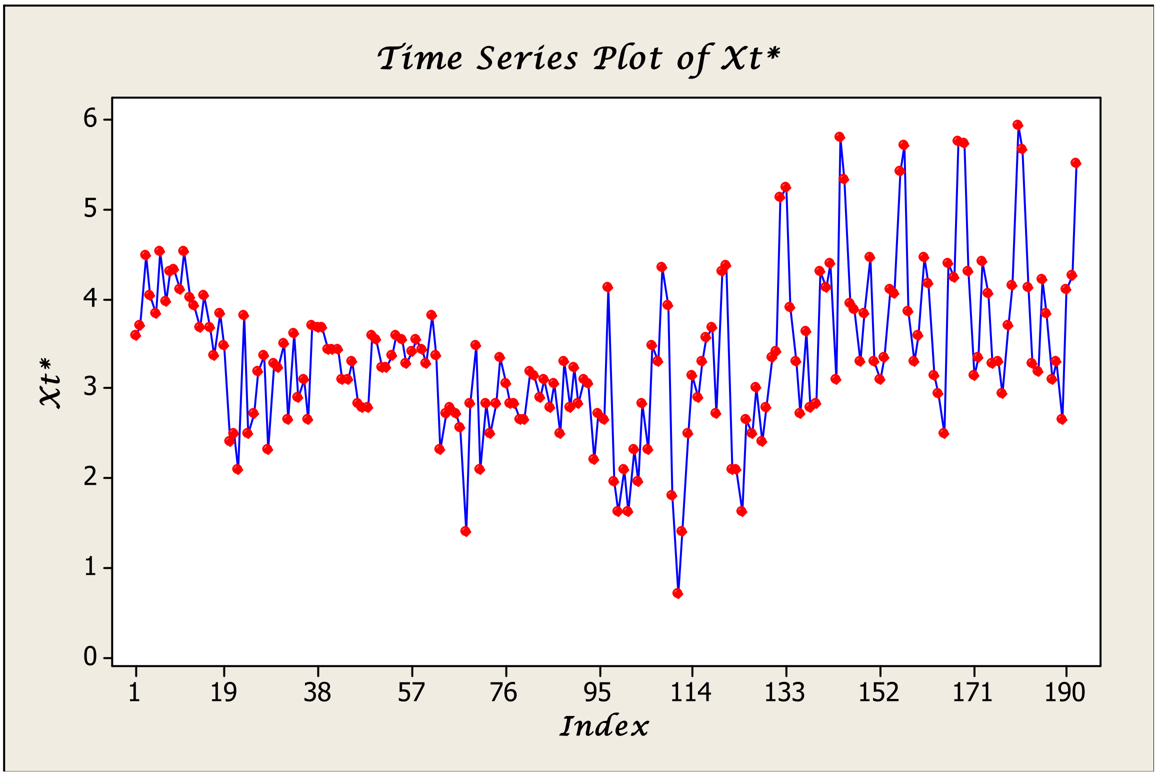

figure 4

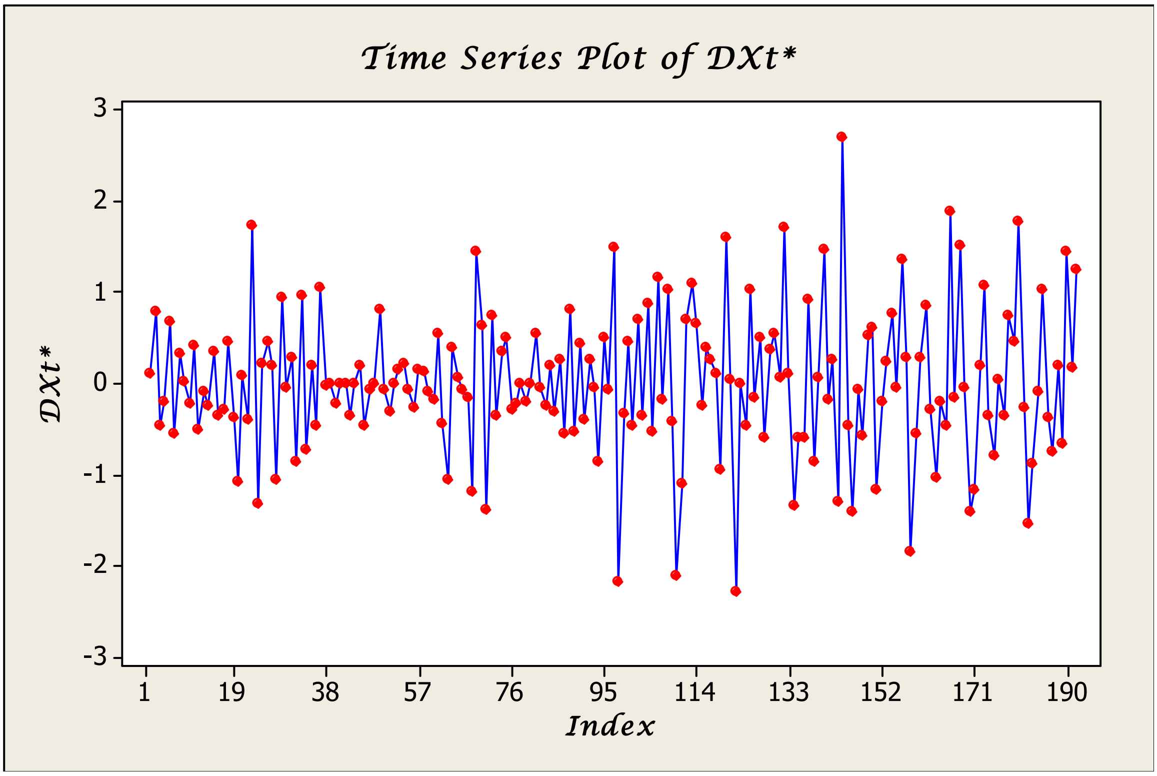

figure 5

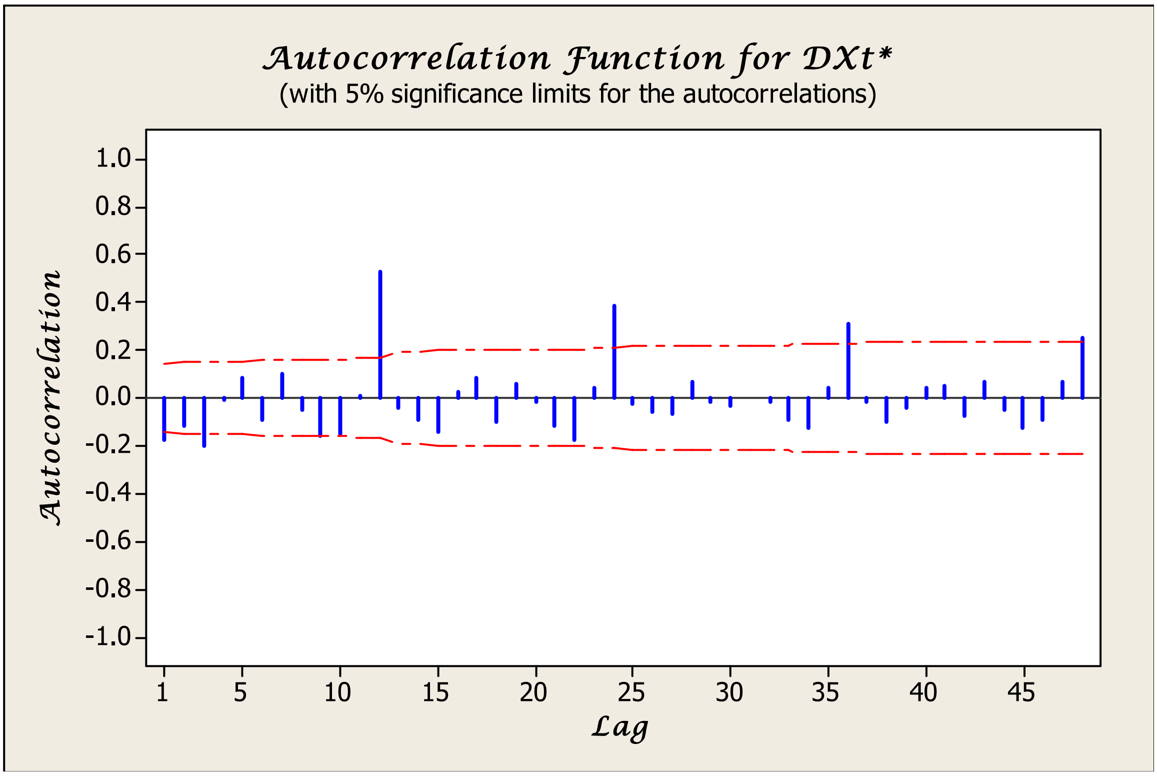

figure 6

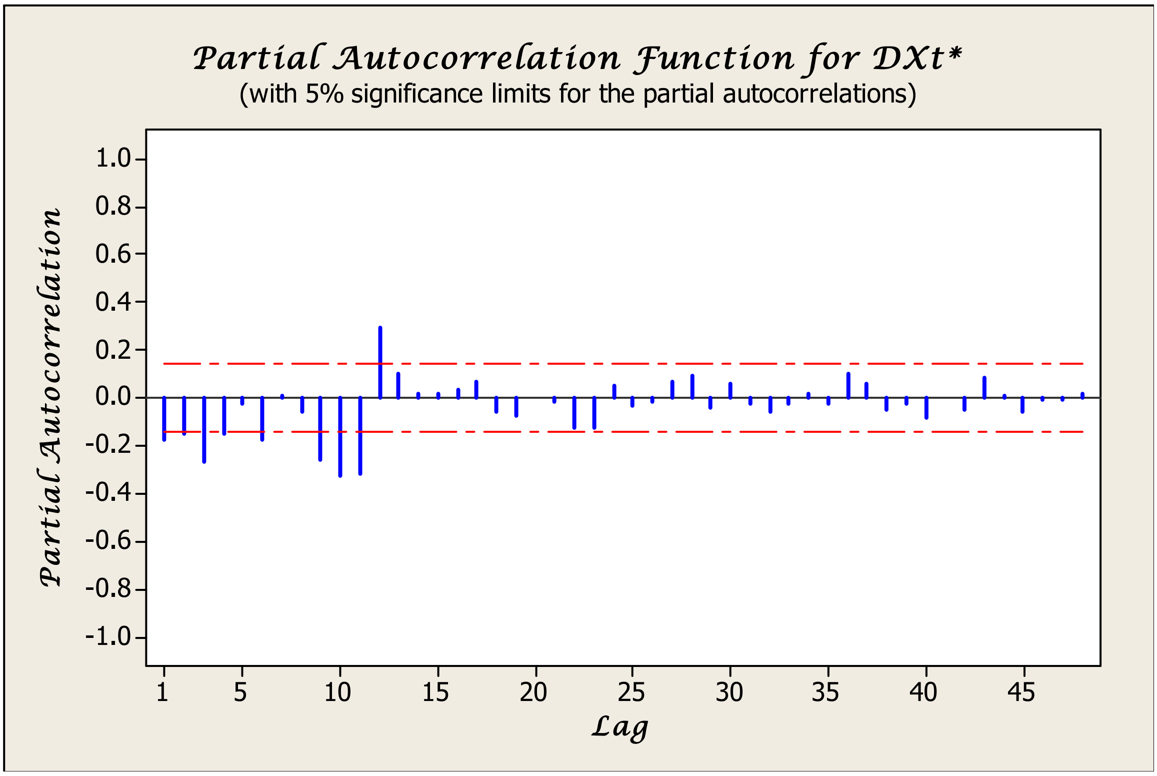

figure 7

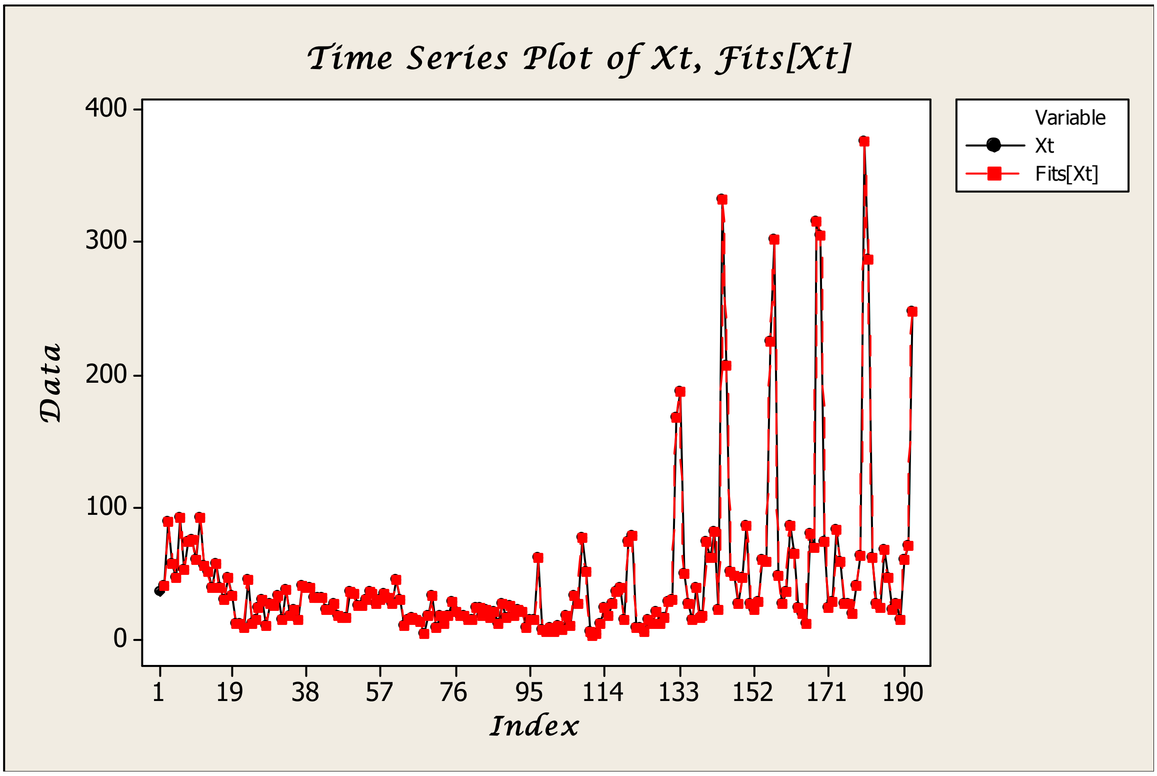

figure 8

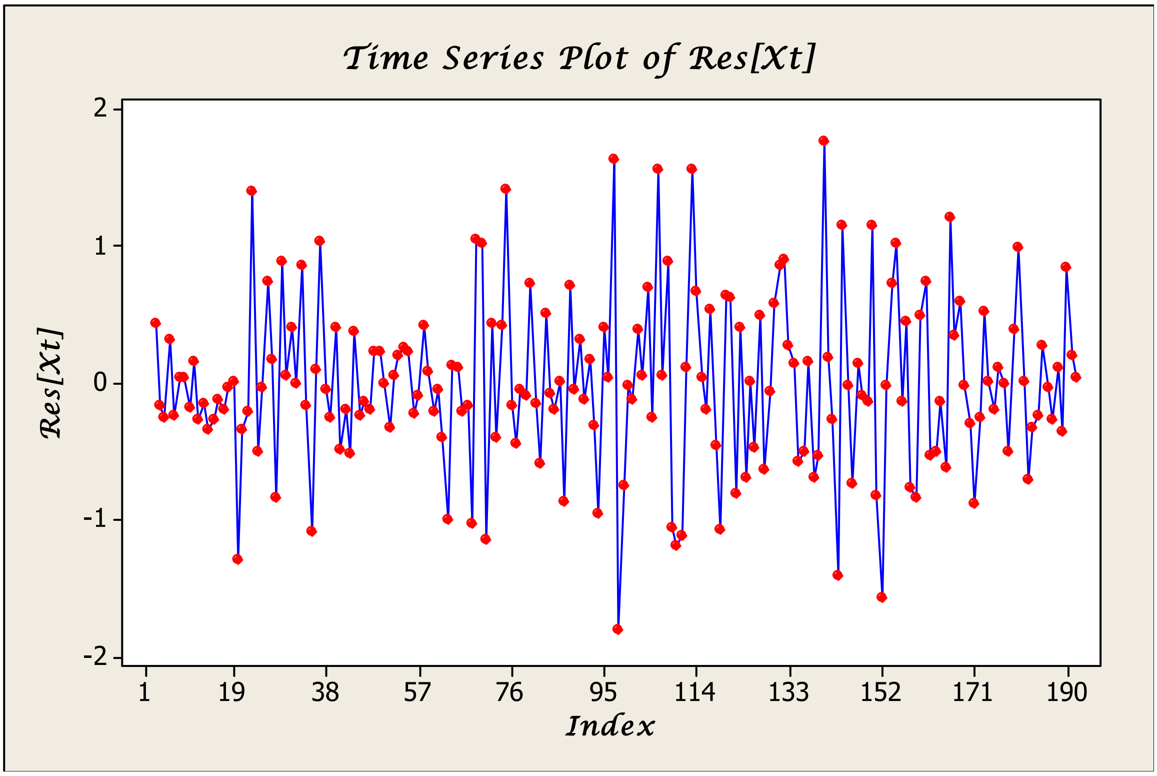

figure 9

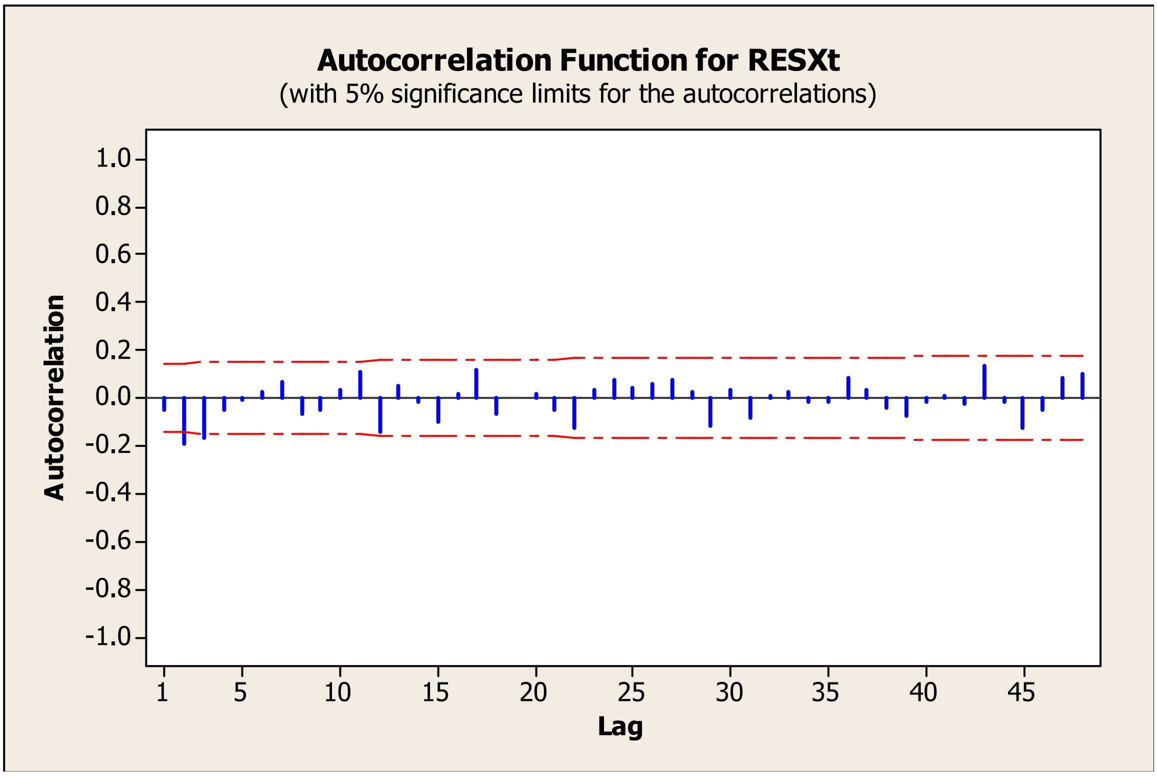

figure 10

Table 1

Iteration	SSE	Parameters
0	248.406	0.100
1	224.659	0.250
2	211.912	0.400
3	209.553	0.488
4	209.525	0.498
5	209.525	0.499
6	209.525	0.499

Relative change in each estimate less than 0.0010

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
SAR 12	0.4988	0.0660	7.56	0.000

```
Differencing: 1 regular difference
Number of observations:  Original series 191, after differencing 190
Residuals:    SS =  208.254 (backforecasts excluded)
              MS =  1.102  DF = 189
```

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

Lag	12	24	36	48

Table 2

Autocorrelation Function: RESXt

Lag	ACF	T	LBQ
1	-0.048036	-0.66	0.45
2	-0.191180	-2.63	7.54
3	-0.164337	-2.18	12.81
4	-0.049013	-0.64	13.28
5	-0.010587	-0.14	13.30
6	0.022771	0.29	13.40
7	0.064038	0.83	14.22
8	-0.067178	-0.86	15.12
9	-0.053421	-0.69	15.70
10	0.033477	0.43	15.93
11	0.112804	1.44	18.52
12	-0.145155	-1.84	22.84
13	0.053170	0.66	23.42
14	-0.018487	-0.23	23.49
15	-0.099228	-1.23	25.54
16	0.020185	0.25	25.63
17	0.113426	1.39	28.34
18	-0.066490	-0.81	29.28
19	-0.002855	-0.03	29.28
20	0.013550	0.16	29.32
21	-0.053328	-0.65	29.93
22	-0.124536	-1.51	33.30
23	0.033742	0.40	33.55
24	0.072177	0.86	34.70

25	0.040599	0.48	35.06
26	0.059871	0.71	35.86
27	0.074002	0.88	37.08
28	0.027447	0.32	37.25
29	-0.119213	-1.41	40.47
30	0.029582	0.35	40.67
31	-0.086941	-1.01	42.41
32	0.004826	0.06	42.41
33	0.022675	0.26	42.53
34	-0.016126	-0.19	42.59
35	-0.014989	-0.17	42.65
36	0.087718	1.02	44.47
37	0.034421	0.40	44.75
38	-0.039431	-0.45	45.12
39	-0.077872	-0.90	46.59
40	-0.012693	-0.15	46.63
41	0.011833	0.14	46.66
42	-0.028051	-0.32	46.86
43	0.135558	1.55	51.42
44	-0.017523	-0.20	51.49
45	-0.128425	-1.45	55.64
46	-0.051766	-0.58	56.32
47	0.084255	0.94	58.13
48	0.100451	1.12	60.72

Table 3

Forecasts from period 192

Period	Forecast	95 Percent Limits Lower	Upper
193	254.537	198.249	310.825
194	52.450	-7.320	112.220
195	20.076	-40.123	80.276
196	23.633	-36.621	83.888
197	71.550	11.288	131.811
198	49.916	-10.346	110.178
199	23.300	-36.962	83.562
200	26.972	-33.291	87.234
201	15.714	-44.548	75.977
202	53.099	-7.164	113.361
203	68.240	7.977	128.502
204	291.144	230.882	351.406
205	265.733	195.083	336.382
206	55.744	-16.123	127.610
207	22.119	-49.901	94.140
208	23.760	-48.280	95.800
209	69.980	-2.062	142.023
210	48.565	-23.478	120.608
211	22.852	-49.191	94.895
212	26.981	-45.062	99.024
213	15.123	-56.920	87.166
214	55.479	-16.564	127.522
215	69.192	-2.851	141.235
216	275.920	203.877	347.963
217	261.871	177.678	346.065
218	54.608	-31.012	140.227
219	21.415	-64.385	107.215
220	23.716	-62.107	109.539
221	70.522	-15.304	156.347

```
222     49.031  -36.795  134.857
```

References

Box, G.E.P. and Jenkins G.M. (1970). Time series analysis. Forecasting and control. Holden Day, San Francisco, C.A. New York.

Brockwell, P. J. (1986). Time series. Theory and methods.Springer-Verlag, New York, Berlin Heidelberg, London. QA 280.B 76 1987 519.5'5 86-22047.

World health organisation (WHO) (2004). World report on road traffic injury prevention. Summary.World Health organisation. Geneva. Switzerland.(http://www.who int/world health day/2004/informaterials/world-report/en/summary_en_rev.pdf).

Atubi, A.O. (2012). A monthly analysis of road traffic accident in selected local government areas of Lagos State, Nigeria. Mediterranean Journal of Social Sciences. Vol.3(11). pp.47-62. Doi: 10.5901/mjss.2012.v3n11p47.

Federal road safety commission (FRSC) (2006). Annual report on road traffic, 31st December; Abuja.Obinna, C. (2007). Road traffic crashes kill 0.4 million youth every year. Nigerian Vanguard, April, 24. pp. 35.

Daramola, A.Y.(2004). Innovative options for financing transport infrastructure in Nigeria. The magazine of the Nigerian institute of social and economic research. No. 485, December.Ibadan.

Trinca, G.W. et al (1988). Reducing traffic injury. A global challenge. A.H. Massina and Co.Melboume.

Filani, M.O. and Gbadamosi, K.T. (2007). Spartial and temporal pattern of road traffic accidentoccurrences in Nigeria: 1970-1995. Nigerian Geographical Journal., Vol.5, No.1, pp. 55-70.

Downing, A.J. et al (1991). Road safety in developing countries. An overview, crowthome: Transport and road research laboratories.