IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 11, Issue 3 Ver. III (May - Jun. 2015), PP 66-77
www.iosrjournals.org

On The Use of Influence Function Matrix in Model Order Determination and Checking Of Outliers in Nigerian Economic Series

Rafiu, A.A[1], Oseni, B.A[2], Oladeji, A.O[3]
[1] *Department of Mathematics and Statistics, the Polytechnic, Ibadan. Adeseun Ogundoyin Campus, Eruwa, Nigeria.*
[2] *Department of Mathematics and Statistics, The Polytechnic, Ibadan, Nigeria.*
[3] *Department of Purchasing and Supply, the Polytechnic Ibadan, Adeseun Ogundoyin Campus, Eruwa, Nigeria.*

Abstract: *An influence function matrix for autocorrelation and partial autocorrelation were constructed for some economic series and it is used as a tool for the detection of outliers and model order, depending on the lagging effects of observation in the correlation coefficient and size of the observations. A suitable model was fitted for all the data judging from the behaviour of both the plots on ACF and PACF and the influence function matrix and forecast were made for model fitted for each data with the presence and removal of outliers from the model. Our findings reveals the detection of 4 outliers for Real GDP, 6 for GDP at current market prices and 1 for GDP at current factor cost (Nominal GDP), 6 for retail prices of diesel recorded by N.N.P.C. while there is no outlier detected for retail prices of PMS (petrol) it was also discovered that the confidence interval on the forecast values for the model without outliers overlap that of the model with outliers.*
Keywords: *Autocorrelation, Partial autocorrelation, Influence function matrix, Outliers and Autoregressive Moving Average (ARMA).*

I. Introduction

We are living in a world in which statistical data are so indispensable for planning economic and social development and for checking on current implementation of programmes and for assessing results. But data have to be understood and correctly handled and this is task of statistics. It is the task of statistics in the sense that if the data are wrongly collected or recorded, it will reflect in the conclusions of the researcher and any subsequent analysis based on such data will be, at best, a waste of time and possibly even disastrous, since it may mislead, with serious consequences. Most especially if the data is on economic; any economic policy made by the government on such data would not work even if an expert is invited from abroad to work on such problematic data. That is why it is advisable to check our data for any outliers or strange value before we start using it for analysis.

Then, what is an outlier from the statistical point of view? Outliers or Wild shots are observations that appear to be inconsistent with the rest of the collected data (Iglewiez and Hoaglin (1993)). That is, Outliers are extreme values of a set of data which are not typical of the rest of the data. Outliers can have deleterious effect on statistical analysis and results. These include increase in error variance, reduction in the power of the statistical tests and distributional assumptions and can seriously bias or influence parameter estimates that may be of substantive interest.

How could this case of outliers be tackled? That is what brought about the title of this paper. In this paper we used the influence function matrix for the autocorrelation and partial autocorrelation as a tool for detecting outliers and model order depending on the lagging effect of observation in the correlation coefficient and the size

II. Material and Method

The sources of data used in this paper were from the Central Bank of Nigeria (CBN) Ibadan Branch. The data on GDP at factor cost, current market prices and current factor cost were extracted from the statistical Bulletin of C.B.N. Other data used in this paper were also gotten through the secondary means. Data on retail prices of some petroleum products (e.g. PMS, DPK and AGO)from 1991 to 2000 were extracted from the Energy Correspondent Newspaper Libraries of N.N.P.C. Apata Depot Ibadan. These data were tagged series A to E, Series A to C for data on GDP and Series D and E for data on retail prices of PMS and AGO respectively.

The statistical tools adopted for detection of outliers and model order are autocorrelation function (ACF), partial autocorrelation function (PACF) and influence function matrix (IFM). Before explaining the concepts of ACF, PACF and IFM, there is a need to explain what autocovariance is.

Autocovariance

The covariance between an observation made at time t (i.e. X_t) and its value at time $t+k$ is called Autocovariance of lag k because it measures the covariance between observations which are k unit apart in time.

Autocovariance of lag k is defined as

$$\gamma_k = Cov(X_t, X_{t+k})$$
$$= E[(X_t - \mu)(X_{t+k} - \mu)]; k = 0,1,2,\ldots$$

Where

$\mu = E(X_t)$ is the theoretical or population mean, which is always assumed to be Zero.

$k = 0,1,2,\ldots \leq \frac{N}{4}$ (Box and Jenkins 1976)

$N =$ Population Size.

The set of γ_k is called Autocovariance function if it is a continuous function and Autocovariance sequence if it discrete.

Thus an autocovariance of lag 2 is;

$$\gamma_2 = E[(X_t - \mu)(X_{t+2} - \mu)]$$

While an autocovariance of Lag zero is

$$\gamma_0 = E[(X_t - \mu)(X_{t+0} - \mu)]$$
$$= E\left[(X_t - \mu)^2\right]$$
$$= Variance\ (\sigma^2)\ of\ \{X_t\}$$

Hence,

$$\gamma_{(-k)} = E[(X_t - \mu)(X_{t-k} - \mu)]$$

Let $S = t - k$ this implies $t = s + k$ and

$$\gamma_{(-k)} = E[(X_{s+k} - \mu)(X_s - \mu)]$$
$$= E[(X_s - \mu)(X_{s+k} - \mu)]$$
$$= \gamma_k$$

This $\gamma_{(-k)} = \gamma_k$ i.e. γ_k is an even function.

Autocorrelation Function (ACF)

A measure of association or correlation between observations which are k units apart in time is called Autocorrelation coefficient of lag k and is defined as

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$
$$= \frac{E[(X_t - \mu)(X_{t-k} - \mu)]}{E\left[(X_t - \mu)^2\right]},\ k = 0,\ 1,\ 2,\ldots$$

If $k = 0,\ then$

$$\rho_0 = \frac{E[(X_t - \mu)(X_{t-k} - \mu)]}{E[(X_t - \mu)^2]} = \frac{\gamma_0}{\gamma_0} = 1$$

$$\therefore \ \rho_{-k} = \frac{\gamma_{-K}}{\gamma_0} = \frac{\gamma_K}{\gamma_0} = \rho_k$$

Hence ρ_k is also an even function

ρ_k = Autocorrelation function i.e. the sequence $\{\rho_k : k = 0,\ 1,\ 2,...\}$ is called Autocorrelation function (ACF).

Given Observations $X_1,\ X_2,\ ...,X_N$ the sample estimate γ_k of ρ_k is correspondingly defined as;

$$\gamma_k = \frac{C_k}{C_O},\ \ k = 0,\ 1,\ 2,...$$

Where

$$C_k = \frac{1}{N}\sum_{t=1}^{N-K}\left(X_t - \overline{X}\right)\left(X_{t+k} - \overline{X}\right)$$

and

$$\overline{X} = \frac{1}{N}\sum_{t=1}^{N} X_t$$

C_k denote the sample autocovariance of lag k. It can be shown that

$$E(C_k) = 1 - \frac{K}{N}\left[\gamma_k - Var\left(\overline{X}\right)\right]$$

So that C_k is a based estimator. But some authors prefer

$$C'_k = \frac{1}{N-K}\sum_{t=1}^{N-K}\left(X_t - \overline{X}\right)\left(X_{t+k} - \overline{X}\right)$$

Which is also biased but has a higher mean square error. Thus the estimate of autocorrelation function is given by

$$\rho_k = \frac{C_k}{C_0};\ \ k = 0,\ 1,\ 2,... \text{M}$$

Where $M \leq \frac{N}{4}$.

The Influence Function

The influence function of an estimate depends on the parameters being estimated, the observation vector whose influence is being measured and the distribution function of that observation vector.

Chernick et al (1982), considered the influence function for the estimation of the autocorrelation function of a time series. We present the influence function of autocorrelation (ρ_k) matrix and considered the extension for partial autocorrelation (ϕ_{kk}) matrix.

Let ρ_k denote the autocorrelation function at lag k, for stationary series $\{X_t\}_{t=1}^{N}$ with $\mu = E(X_t)$ and $\sigma^2 = Var(X_t)$ and without loss of generality on ρ_k let $Z_t = \frac{(X_t - \mu)}{\sigma}$ for each t, then the influence function.

$$I(F, T(F), x) = \lim_{c \to 0}\left[\frac{T(1-\varepsilon)F + c\partial x) - T(F)}{\ell}\right]$$

where T(F) is the estimator under consideration, and F the empiric distribution, become $I(H, \rho_K, z)$ that is $I(F, \rho_K, X) = I(H, \rho_K, Z)$ where H is the distribution for $I(Z_t, Z_{t+k})$

Let

$$U_{i,k,1} = \left\{ \frac{(Z_i + Z_{i+k})}{\sqrt{1+p_k}} + \frac{(Z_i - Z_{i+k})}{\sqrt{1+p_k}} \right\} / 2$$

$$U_{i,k,2} = \left\{ \frac{(Z_i + Z_{i+k})}{\sqrt{1+p_k}} + \frac{(Z_i - Z_{i+k})}{\sqrt{1+p_k}} \right\} / 2$$

Then we have

$$(\mathbf{1} - p_k^2)^2 U_{i,k,1} U_{i,k,2} = Z_i Z_{i+k} - \rho_k (Z_i^2 + Z_{i+k}^2)/2$$

Hence

$$I[H, p_k, (Z_i Z_{i+k})] = (1 - p_k^2) U_{ik1} U_{ik2}$$

Analysis Of Data

The available five economic series data sets presented below have been used to illustrate the procedure earlier discussed. The data were tagged series A to E.

Series A: GDP at 1984 Factor Cost (Real GDP)

1990	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**
90,342.1	34,614.1	97,431.1	100,015.2	101,330.0	103,510.0	107,030.0	110,400.0	112,000.0	116,000.0

Series B: GDP at Current Market Prices

1990	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**
260,636.7	324,010.0	549,808.8	697,090.0	914,940.0	1,977,740.0	2,833,170.0	2,939,500.0	2,837,200.0	2,224,796.9

Series C: GDP at Current Factor Cost (Nominal GDP)

1990	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**
222,157.6	257,873.0	320,247.3	544,330.7	691,600.0	911,070.0	1,960,690.0	2,749,720.0	2,834,800.0	2,721,510,0

Series D: Retail Prices of PMS (Premium Motor Spirit) Petrol

1991	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**2000**
0.70	0.70	3.25	11.00	11.00	11.00	11.00	11.00	20.00	22.00

Series E: Retail Prices of AGO (Automotive Gas Oil) Diesel

1991	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**2000**
0.55	0.55	3.00	9.00	9.00	9.00	9.00	9.00	19.00	21.00

The plot of Autocorrelation was done with the help of MINITAB, the autocorrelation function and the influence function matrix as well as partial autocorrelation function with its corresponding influence function matrix were given in this papers and interpretation based on the result were given for each data.

We shall compute the influence function with the critical value based on the standard error given as

$$\left\{ \frac{n-k}{n(n+2)} \right\}^{\frac{1}{2}}$$

Given n observation with k lag, we compute the influence function matrix.

Autocorrelation Function

- Real GDP

Series A (Auto)

	K=1			K=2			K=3		
t	$U_{i,1}$	$U_{i,2}$	1( )	$U_{i,1}$	$U_{i,2}$	I( )	$U_{i,1}$	$U_{i,2}$	I( )
1.	0.036	1.015	0.015	0.294	1.247	0.342	0.599	1.431	0.818
2.	0.497	1.173	0.256	0.789	1.273	0.932	0.940	-0.033	-0.029
3.	0.907	1.084	0.430	1.323	-0.298	-0.370	1.170	-1.050	-1.214
4.	1.965	-1.035	-0.888	1.553	-1.407	-2.045	1.251	-1.154	-1.381
5.	0.575	-1.580	-0.395	0.035	-1.290	-0.036	-0.238	-1.062	0.251
6.	-0.859	-1.012	0.381	-1.094	-0.901	0.933	-1.272	-0.617	0.800
7.	-1.117	-0.628	0.310	-1.240	-0.360	0.425	-1.279	0.028	-0.026
8.	-1.107	-0.062	0.033	-1.079	0.297	-0.290	0	0	0
9.	-0.811	0.541	-0.190	0	0	0	0	0	0
10.	0	0	o	0	0	0	0	0	0

Series A Lag k=1,2,3

	1	2	3
1	BL	+	+
2	BL	+	BL
3	+	-	-
4	-	-	-
5	-	BL	+
6	+	+	+
7	+	+	BL
8	BL	-	BL
9	BL	BL	BL
10	BL	BL	BL

Thus the above sign gives the influence function matrix.
BL =Blanks.

Series A (Partial)

	K=1			K=2			K=3		
t	$U_{i,1}$	$U_{i,2}$	1( )	$U_{i,1}$	$U_{i,2}$	I( )	$U_{i,1}$	$U_{i,2}$	I( )
1.	0.036	1.015	0.015	1.351	1.978	4.096	0.377	1.345	0.499
2.	0.497	1.173	0.256	2.062	2.384	7.538	0.949	-0.180	-0.168
3.	0.907	1.084	0.430	1.638	-0.566	-1.428	1.347	-1.274	-1.692
4.	1.965	-1.035	-0.888	1.127	-0.819	-1.418	1.439	-1.355	-1.924
5.	0.575	-1.580	-0.395	0.933	-1.798	-2.571	-0.080	-1.029	0.081
6.	-0.859	-1.012	0.384	-2.227	-2.100	7.213	-1.176	-0.449	0.533
7.	-1.117	-0.628	0.310	-2.024	-1.445	4.517	-1.289	0.230	-0.285
8.	-1.107	-0.062	0.033	-1.302	0.403	-0.800	0	0	0
9.	-0.811	0.541	-0.190	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	0

Series A Lag k=1,2,3

	1	2	3
1	BL	+	+
2	BL	+	BL
3	+	-	-
4	-	-	-
5	-	-	BL
6	+	+	+
7	+	+	-
8	BL	-	BL
9	BL	BL	BL
10	BL	BL	BL

Thus the above sign gives the influence function matrix.

BL=Blank.

According to Shangodoyin (1998), the lag (k) with the highest number of blanks and for which $\rho(k+1)$ or $\phi(k+1)$ cut off the possible order of the model. Therefore, from the influence function matrix for ACF and PACF we deduce that the possible order of the model for this series are 1 or 3 (since it is lag 1 and 3 that has the highest number of blank of 5). We can then conclude that the best model for this series is ARMA(1,1) or ARMA(3,3). The influence function matrix on Autocorrelation identifies 3 outliers (at t=3,4,6,), and the influence function matrix computed on the partial autocorrelation identifies 4 outliers (at t= 3,4,6,7).We shall delete the observation appropriately.

- GDP at Current Market Prices

Series B (Auto)

	K=1			K=2			K=3		
t	$U_{i,1}$	$U_{i,2}$	1( )	$U_{i,1}$	$U_{i,2}$	I( )	$U_{i,1}$	$U_{i,2}$	I( )
1.	0.717	0.727	0.351	0.789	0.795	0.569	0.889	1.533	1.354
2.	0.727	0.722	0.354	0.800	1.362	0.988	0.900	-0.890	-0.796
3.	0.722	1.238	0.601	0.795	-0.792	-0.569	0.895	-0.888	-0.789
4.	1.238	-0.721	-0.599	1.362	-0.790	-0.978	1.533	-0.887	-1.358
5.	-0.721	-0.719	0.349	-0.792	-0.789	0.579	-0.890	-0.887	0.804
6.	-0.719	-0.718	0.348	-0.790	-0.789	0.578	-0.888	0	0.802
7.	-0.718	-0.718	0.348	-0.789	-0.789	0.577	-0.887	0	0
8.	1.237	0.727	-0.597	0.727	0.720	0.355	0	0	0
9.	0.720	1.237	0.601	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	0

Series B lag k=1,2,3

	1	2	3
1	+	+	+
2	+	+	-
3	+	-	-
4	-	-	-
5	+	+	+
6	+	+	+
7	+	+	BL
8	-	+	BL
9	+	BL	BL
10	BL	BL	BL

Thus the above sign gives the influence function matrix
BL=Blank.

*GDP at Current Market Price
Series B (Partial)

	K=1			K=2			K=3		
t	$U_{i,1}$	$U_{i,2}$	1( )	$U_{i,1}$	$U_{i,2}$	I( )	$U_{i,1}$	$U_{i,2}$	I( )
1.	0.717	0.727	0.351	0.971	0.923	0.844	1.065	1.663	1.827
2.	0.727	0.722	0.354	0.942	1.571	1.475	0.836	-0.826	-0.711
3.	0.722	1.238	0.601	0.888	-0.886	-0.784	0.830	-0.824	-0.704
4.	1.238	-0.721	-0.599	1.535	-0.871	-1.338	1.486	-0.760	-1.169
5.	-0.721	-0.719	0.349	-0.921	-0.918	0.859	-1.004	-1.001	1.057
6.	-0.719	-0.718	0.348	-0.919	-0.918	0.857	-1.002	-1.001	1.055
7.	-0.718	-0.718	0.348	-0.918	-0.918	0.856	-0.912	-0.084	0.089
8.	1.237	0.727	-0.597	2.733	0.202	0.844	0	0	0
9.	0.720	1.237	0.601	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	0

Series B lag k=1,2,3

	1	2	3
1	+	+	+
2	+	+	-
3	+	-	-
4	-	-	-
5	+	+	+
6	+	+	+
7	+	+	BL
8	-	+	BL
9	+	BL	BL
10	BL	BL	BL

Thus the above sign gives the influence function matrix.
BL=Blank.

The highest number of blanks is recorded at k=3 and both ACF and PACF cut off after this lag 3, which implies that ARMA model of order 3[ARMA (3,3)] could be fitted.

The influence function matrix (IFM) based on autocorrelation function identifies 6 outlier at t=1,2,3,4,5, and 6 while the influence function matrix calculated on the partial autocorrelation function also identifies 6 outliers at the same position of t i.e. 1,2,3,4,5,6. The observations that are outlier shall be deleted appropriately depending on the correlation function to be used.

*GDP at Current Factor Cost (Nominal GDP)
Series C (Auto)

	K=1			K=2			K=3		
t	$U_{i,1}$	$U_{i,2}$	1( )	$U_{i,1}$	$U_{i,2}$	I( )	$U_{i,1}$	$U_{i,2}$	I( )
1.	-0.319	-0.117	0.119	0.317	2.726	1.051	-0.325	0.378	-0.123
2.	-0.685	2.806	-1.924	-0.347	0.329	-0.137	-0.536	-1.330	0.747
3.	2.731	0.200	0.542	2.471	-0.696	-2.110	2.565	-0.142	-0.382
4.	0.553	-1.417	-0.779	0.342	-0.296	-0.121	0.406	-0.014	-0.004
5.	-1.356	-0.282	0.387	-1.434	-0.415	0.738	-1.306	-0.152	0.215
6.	0.399	-0.022	0.012	-0.419	-0.143	0.079	-0.376	-0.054	0.027
7.	-0.053	0.035	0.004	-0.063	-0.038	0.006	-0.051	-0.001	0.000
8.	0.726	0.721	0.353	1.228	-0.721	-0.599	0	0	0
9.	0.716	0.726	0.350	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	0

Series C lag k = 1, 2, 3,

	1	2	3
1	BL	+	BL
2	-	BL	+
3	+	-	-
4	-	BL	BL
5	+	+	BL
6	BL	BL	BL
7	BL	BL	BL
8	+	-	BL
9	+	BL	BL
10	BL	BL	BL

Thus the above sign gives the influence function matrix.
BL=Blank.

Series C (Partial)

	K=1			K=2			K=3		
t	$U_{i,1}$	$U_{i,2}$	1( )	$U_{i,1}$	$U_{i,2}$	I( )	$U_{i,1}$	$U_{i,2}$	I( )
1.	-0.319	-0.117	0.119	0.321	2.728	1.066	-0.323	0.377	-0.122
2.	-0.685	2.806	-1.924	-0.347	0.329	-0.137	-0.544	-1.335	0.763
3.	2.731	0.200	0.542	2.472	-0.693	-2.105	2.567	-0.128	-0.345
4.	0.553	-1.417	-0.779	0.342	-0.296	-0.121	0.406	-0.012	-0.003
5.	-1.356	-0.282	0.387	-1.435	-0.417	0.738	-1.308	-0.160	0.226
6.	0.399	-0.019	0.012	-0.419	-0.144	0.079	-0.376	-0.056	0.028
7.	-0.053	0.035	0.004	-0.063	-0.039	0.006	-0.051	-0.001	0.000
8.	0.726	0.721	0.353	1.228	-0.721	-0.599	0	0	0
9.	0.716	0.726	0.350	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	0

Series C lag k = 1, 2, 3,

	1	2	3
1	BL	+	BL
2	-	BL	+
3	+	-	-
4	-	BL	BL
5	+	+	BL
6	BL	BL	BL
7	BL	BL	BL
8	+	-	BL
9	+	BL	BL
10	BL	BL	BL

The highest number of blank for both the ACF and PACF is recorded at lag 3. Both ACF and PACF cut off at this lag, indicating that ARMA (3,3) could be fitted i.e. the order is 3. The influence function identifies just one outlier at t=3. And when the outlier t=3 is screened out, ARMA model of order 3 ARMA (3,3) will be suitable for this series. It is at t=3 (1992) that the GDP started to increase astronomically and when we compared t1 and t2 and t2 and t3 we discover that the difference between the latter and former nearly triple

Retail Prices of PMS (Petrol)
Series D (Auto)

	K=1			K=2			K=3		
t	$U_{i,1}$	$U_{i,2}$	1( )	$U_{i,1}$	$U_{i,2}$	I( )	$U_{i,1}$	$U_{i,2}$	I( )
1.	1.197	-0.396	-0.301	1.826	-0.275	0.307	1.896	-0.067	-0.093
2.	2.209	-0.734	-1.035	2.062	-0.311	0.391	-0.148	0.010	-0.006
3.	3.250	-1.082	-2.246	0.621	-0.090	0.039	0.515	-0.014	-0.002
4.	0.862	-0.284	-0.155	0.664	-0.097	0.044	0.625	-0.019	-0.006
5.	-1.781	0.599	-0.676	-1.572	0.245	0.231	-1.751	0.072	-0.079
6.	-1.905	0.640	-0.774	-1.901	0.295	0.335	-1.929	-0.079	-0.097
7.	-2.201	0.739	-1.033	-2.014	0.394	0.376	-0.810	0.036	-0.013
8.	0.817	0.828	0.674	-0.775	0.677	-0.621	0	0	0
9.	-0.677	0.570	0.396	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	0

Series D lag k = 1, 2, 3

	1	2	3
1	-	+	BL
2	-	+	BL
3	-	BL	BL
4	BL	BL	BL
5	-	BL	BL
6	-	+	BL
7	-	+	BL
8	+	-	BL
9	+	BL	BL
10	BL	BL	BL

Series D (Partial)

	K=1			K=2			K=3		
t	$U_{i,1}$	$U_{i,2}$	1( )	$U_{i,1}$	$U_{i,2}$	I( )	$U_{i,1}$	$U_{i,2}$	I( )
1.	1.197	-0.396	-0.301	1.792	0.190	0.482	1.894	-0.047	-0.131
2.	2.209	-0.734	-1.035	2.024	0.214	-0.617	-0.149	0.009	0.004
3.	3.250	-1.082	-2.246	0.610	0.067	-0.052	0.515	-0.009	-0.005
4.	0.862	-0.284	-0.155	0.651	0.071	-0.060	0.625	-0.012	-0.010
5.	-1.781	0.599	-0.676	-1.542	-0.157	-0.360	-1.749	0.053	-0.112
6.	-1.905	0.640	-0.774	-1.865	-0.191	-0528	-1.927	-0.058	-0.137
7.	-2.201	0.739	-1.033	-1.976	-0.202	0.593	-0.809	0.027	-0.020
8.	0.817	0.828	0.674	-0.775	0.677	0.616	0	0	0
9.	0.677	0.570	0.396	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	0

Series D lag k = 1, 2, 3

	1	2	3
1	-	-	BL
2	-	-	BL
3	-	BL	BL
4	BL	BL	BL
5	-	-	BL
6	-	-	BL
7	-	-	BL
8	+	-	BL
9	+	BL	BL
10	BL	BL	BL

Thus the above sign gives the influence function matrix.
BL = Blank

The highest of blanks for both ACF and PACF is recorded at lag 3 as could be seen from the influence function matrix (IFM) for both correlation. Both ACF and PACF cut off at this lag and this implies that the possible order of the model for this series is 3. We can then conclude that the best model for this series is ARMA (3, 3); and as could be seen from the IFM table it is an ARMA model of full order 3. The influence function matrix constructed for both ACF and PACF does not identify any observation as outlier. Meaning that N.N.P.C is justified with all the increment made so far on the prices of petrol bearing in mind the prices of petrol all over the world and the standard of living in the country.

- Retail Prices of AGO (Diesel)

Series E (Auto)

	K=1			K=2			K=3		
t	$U_{i,1}$	$U_{i,2}$	1( )	$U_{i,1}$	$U_{i,2}$	I( )	$U_{i,1}$	$U_{i,2}$	I( )
1.	0.817	0.828	0.674	1.215	1.432	2.184	0.930	1.090	1.041
2.	0.825	1.081	0.888	1.202	1.336	1.958	0.807	-0.417	-0.344
3.	1.079	0.988	1.088	1.067	-0.236	-0.308	1.044	-0.599	-0.643
4.	1.009	-0.503	-0.505	0.912	-0.483	-0.537	0.963	-0.494	-0.488
5.	-0.480	-0.685	0.335	-0.699	-0.773	0.674	-0.607	-1.321	0.842
6.	-0.683	-0.570	0.396	-1.112	-1.583	2.172	-0.831	-1.554	1.349
7.	-0.562	-1.255	-0.708	-1.047	-1.786	2.309	-0.584	-0.048	0.037
8.	1.066	-0.837	-0.309	0.863	-0.283	-0.154	0	0	0
9.	-1.209	1.433	2.132	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	0

Series E lag k= 1,2,3

	1	2	3
1	+	+	+
2	+	+	-
3	+	-	-
4	-	-	-
5	+	+	+
6	+	+	+
7	-	+	BL
8	-	BL	BL
9	+	BL	BL
10	BL	BL	BL

Thus the above signs gives the influence function matrix.
BL= Blank.

Series E (Partial)

	K=1			K=2			K=3		
t	$U_{i,1}$	$U_{i,2}$	1( )	$U_{i,1}$	$U_{i,2}$	I( )	$U_{i,1}$	$U_{i,2}$	I( )
1.	0.817	0.828	0.674	1.217	1.435	2.133	0.937	1.097	1.059
2.	0.825	1.081	0.888	1.204	1.359	1.969	0.805	-0.413	-0.342
3.	1.079	0.988	1.088	1.067	-0.234	-0.307	1.042	-0.594	-0.638
4.	1.009	-0.503	-0.505	0.911	-0.482	-0.537	0.962	-0.489	-0.484
5.	-0.480	-0.685	0.335	-0.781	-0.775	0.677	-0.615	-1.326	0.860
6.	-0.683	-0.570	0.396	-1.114	-1.585	2.184	-0.841	-1.556	1.376
7.	-0.562	-1.255	-0.708	-1.050	-1.789	2.323	-0.585	-0.051	0.039
8.	1.066	-0.837	-0.309	1.324	-1.053	-2.045	0	0	0
9.	-1.209	1.433	2.132	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	0

Series E lag K = 1,2,3

	1	2	3
1	+	+	+
2	+	+	-
3	+	-	-
4	-	-	-
5	+	+	+
6	+	+	+
7	-	+	BL
8	-	-	BL
9	+	BL	BL
10	BL	BL	BL

Thus the above sign gives the influence function matrix.
BL=Blank.

From the influence function matrix (IFM) constructed for ACF and PACF, the highest number of blank could be seen in lag 3and both ACF and PACF cut off at this lag. This implies that the possible order of the model for this series is 3, we can then conclude that the best model for this series is ARMA of order 3,i.e, ARMA (3,3) The influence function matrix for both ACF and PACF identifies 6 outlier (at t =1, 2, 3, 4, 5 and 6) We shall model all these series with and without (after deleting the outlier) outliers using ARMA model and then make forecast to see how the series will perform with the presence and absence of outliers.

Arma Modelling With And Without Outliers

We fit suitable ARMA model based on our findings on the series. Four of the series (Series A,B ,C and E) shall be fitted with and without outliers because no outlier was detected for series on retail prices of PMS (petrol). This model shall be fitted using MINITAB package.

Series A

With Outliers

Fitting ARMA (1,) model as

$$\frac{X_t + 0.934 X_{t-1}}{(0.159)} = \frac{\sum_t - 0.934 \sum_{t-1}}{(0.264)}$$

With standard error in bracket, the 3 steps ahead forecast as computed is

Lower 95	Forecast	Upper 95
87290.25	102821.25	118352.25
85127.072	102281.619	119436.262
82895.659	101514.874	120214.089

Without Outliers

The observation at t = 3, 4, 6 and 7 have been detected and we fitted ARMA (1,1) model as

$$\frac{X_t + 0.902 X_{t-1}}{(0.312)} = \frac{\sum_t + 0.014 \sum_{t-1}}{(0.338)}$$

The 3 steps ahead forecast as computed is

Lower 95	Forecast	Upper 95
84183.268	101899.415	119615.562
81497.128	101154.342	120811.556
78354.791	100402.405	122450.019

Our observation from the above values is that the estimates of the models with and without outliers are significant. And the confidence intervals on the forecast values for the model without outlier overlap that of the model with outlier.

Series B

With Outliers

Fitting ARMA (3, 3) model is

$$\frac{X_t + 0.12 X_{t-1}}{(0.002)} + \frac{0.11 X_{t-2}}{(0.001)} + \frac{0.18 X_{t-3}}{(0.009)} = \frac{e_t + 0.14 e_{t-1}}{(0.008)} + \frac{0.11 e_{t-2}}{(0.0017)} + \frac{0.14 e_{t-3}}{(0.0017)}$$

The 3 steps ahead forecast as computed is

Lower 95	Forecast	Upper 95
1321236.897	1355068.135	1388899.373
1303630.448	1354619.975	1405609.502
1290706.01	1354253.241	1417800.472

Without Outliers

The observation at t = 1 to 6 have been deleted and we then fit the model

$$\frac{X_t + 0.197 X_{t-1}}{(0.002)} + \frac{0.118 X_{t-2}}{(0.032)} + \frac{0.219 X_{t-3}}{(0.113)} = \frac{e_t + 0.15 e_{t-1}}{(0.008)} + \frac{0.326 e_{t-2}}{(0.157)} + \frac{0.28 e_{t-3}}{(0.113)}$$

The 3 steps ahead forecast as computed is

Lower 95	Forecast	Upper 95
1286182.009	1354723.398	1423261.787
1280722.315	13553867.525	1427012.735
1241215.713	1335341.034	1429466.355

Series C
With outliers
Fitting ARMA (3, 3) model as

$$\frac{X_t + 0.934\,X_{t-1}}{(0.131)} + \frac{0.858\,X_{t-2}}{(0.112)} + \frac{0.509\,X_{t-3}}{(0.103)} = \frac{e_t + 0.185\,e_{t-1}}{(0.012)} + \frac{0.178\,e_{t-2}}{(0.134)} + \frac{0.137\,e_{t-3}}{(0.159)}$$

The 3 step ahead forecast as computed is

Lower 95	Forecast	Upper 95
1276268.31	1321253.881	1366239.452
1262424.299	1320533.725	1378643.451
1243292.127	1319831.85	1396371.573

Without Outliers
The observation at t = 3 have been deleted and we then fit the model

$$\frac{X_t + 0.798\,X_{t-1}}{(0.127)} + \frac{0.667\,X_{t-2}}{(0.108)} + \frac{0.516\,X_{t-3}}{(0.089)} = \frac{e_t + 0.178\,e_{t-1}}{(0.011)} + \frac{0.165\,e_{t-2}}{(0.118)} + \frac{0.117\,e_{t-3}}{(0.148)}$$

The 3 steps ahead forecast as computed is

Lower 95	Forecast	Upper 95
1241491.743	1320403.997	1399316.251
1239755.813	1319601.125	1399446.437
1237093.797	1318970.156	1400846.515

Series D
The series has not outlier, we then fit ARMA (3,3) as

$$\frac{X_t + 0.834\,X_{t-1}}{(0.217)} + \frac{0.726\,X_{t-2}}{(0.135)} + \frac{0.328\,X_{t-3}}{(0.145)} = \frac{e_t + 0.153\,e_{t-1}}{(0.015)} + \frac{0.187\,e_{t-2}}{(0.133)} + \frac{0.46\,e_{t-3}}{(0.014)}$$

The 3 step ahead forecast as computed is

Lower 95	Forecast	Upper 95
2.274	9.952	17.63
0.651	9.22	17.789
-1.792	8.402	18.596

Series E
With Outliers
We fitted ARMA (3, 3) as

$$\frac{X_t + 0.726\,X_{t-1}}{(0.001)} + \frac{0.356\,X_{t-2}}{(0.114)} + \frac{0.824\,X_{t-3}}{(0.312)} = \frac{e_t + 0.147\,e_{t-1}}{(0.012)} + \frac{0.127\,e_{t-2}}{(0.153)} + \frac{0.578\,e_{t-3}}{(0.019)}$$

and we computed the 3 step ahead forecast as

Lower 95	Forecast	Upper 95
4.231	8.372	12.513
1.154	7.548	13.942
-0.683	7.291	15.265

Without Outliers
The observation at t = 1 to 6 have been deleted and we fitted the model ARMA (3, 3) as

$$\frac{X_t + 0.606\,X_{t-1}}{(0.012)} + \frac{0.471\,X_{t-2}}{(0.024)} + \frac{0.554\,X_{t-3}}{(0.018)} = \frac{e_t + 0.111\,e_{t-1}}{(0.103)} + \frac{0.123\,e_{t-2}}{(0.005)} + \frac{0.514\,e_{t-3}}{(0.113)}$$

The 3 step ahead forecast is

Lower 95	Forecast	Upper 95
1.775	7.864	13.953
-1.068	7.285	15.638
-1.328	6.588	17.504

III. Findings

We observed that the lag with lesser influence is the possible order of the model if the value of ACF and/or PACF is significant at this point and model for any series depends on the nature of the plots of ACF and PACF values. Also, the plots of the ACF and PACF confirm the identification of aberrant observation deduced by the Influence Function Matrix because at the point where the outliers are identified the plot of ACF and PACF indicate that the cut-off also takes place at that same point. The ARMA models fitted for both series with and without outliers are reliable because of the significance of the model estimates. The confidence interval on the forecast values for the model without outliers overlap that of the model with outliers.

The influence function matrix method is very easy and straight forward in determining the order of models compared with others method of model order determination.

IV. Conclusion

Outlier significantly affects the estimates of the model, apart from this, the model residual is affected, and these will have a combine effect on the precision of output generated. So checking for outliers and removing it make the economic series suitable for use. The ARMA modeling which we have used could also be used for Autoregressive modeling in general. All the subsets of the models are analysed using the likelihood estimates of the residual variance as criterion.The ARMA modeling works efficiently in practice and the method is reasonably efficient in computer time provided K is not too large. The only problem or limitation of this model is that it cannot be used to obtain the estimate of the coefficient of the full order model. But this cannot be regarded as a serious problem as the computer (statistical packages) can always be used to fit a full order of any autoregressive model. On a final note, it is observed that our procedure of detecting outlier and model order through (IFM) has limitation as it cannot be comfortably applied to large series for which the series of influence function matrix is not manageable by the working sheet, however, a breakdown of IFM into manageable size based on the series size and length of the lag can alleviate this problem. It is hereby recommended that further research into the contribution of IFM with the inverse autocorrelation function to be carried out.

References

[1]. Anderson, OD. (1971). The Statistical Analysis of Time Series. Wiley, New York.
[2]. Box G.E.P. and Jenkins, G.M. (1976). Time Series Analysis, Forecasting and control. Holden-day Inc.
[3]. Chernick, M.R.; Dowring, D.J. and Piks, D.K (1982) "Detecting Outliers in Time Series Data ". Journal Amer. Stat. Vol. 77, p. 743-747.
[4]. Durbin, J. (1960). "The fitting of time series Models" Rev. Int. Inst. Of Stat. Vol. 28, p. 223-244.
[5]. Granger, C.W.J. and Morris, M.J.(1976). Modelling and Interpretation.
[6]. Shangodoyin, D.K. (1998). "The use of Corrrelation function for detecting outlier and model order determination". Journal of the Nigerian Statisticians. (Page 24-28).