J. Basic. Appl. Sci. Res., 2(6)5964-5970, 2012
© 2012, TextRoad Publication

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Arima Fit to Nigerian Unemployment Data

Ette Harrison ETUK[1], Bartholomew UCHENDU[2], Uyodhu VICTOR-EDEMA[3]

[1]Department of Mathematics/Computer Science, Rivers State University of Science and Technology, Nigeria
[2]Department of Mathematics/Statistics, Federal Polytechnic, Nekede, Imo State, Nigeria
[3]Department of Mathematics/Statistics, Rivers State University of Education, Nigeria

ABSTRACT

Nigerian unemployment data is modelled by Box-Jenkins approach and the use of automatic model selection criteria Akaike Information criterion (AIC) and Schwarz Information Criterion (SIC). It is inferred that the most adequate model is autoregressive integrated moving average of orders 1, 2 and 1(ARIMA(1 ,2 ,1)). Forecasts are obtained on the basis of the model.
KEY WORDS: Unemployment data, ARIMA modelling, AIC, SIC, Nigeria.

INTRODUCTION

A time series is defined as a set of data collected sequentially in time. It has the property that neighbouring values are correlated. This tendency is called *autocorrelation.* A time series is said to be stationary if it has a constant mean and variance. Moreover the autocorrelation is a function of the lag separating the correlated values called *the autocorrelation function* (ACF).

A stationary time series $\{X_t\}$ is said to follow an *autoregressive moving average model of orders p and q* (designated ARMA(p,q)) if it satisfies the following difference equation

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + ... + \beta_q \varepsilon_{t-q} \quad (1)$$

or

$$\boldsymbol{\alpha}(B)X_t = \boldsymbol{\beta}(B)\varepsilon_t \quad (2)$$

where $\{\varepsilon_t\}$ is a sequence of random variables with zero mean and constant variance, called *a white noise process,* and the α_i's and β_j's constants; $\boldsymbol{\alpha}(B) = 1 + \alpha_1 B + \alpha_2 B^2 + ... + \alpha_p B^p$ and $\boldsymbol{\beta}(B) = 1 + \beta_1 B + \beta_2 B^2 + ... + \beta_q B^q$ and B is the backward shift operator defined by $B^k X_t = X_{t-k}$.

If p=0, model (1) becomes *a moving average model of order q* (designated MA(q)). If, however, q=0 it becomes *an autoregressive process of order p* (designated AR(p)). An AR(p) model of order p may be defined as a model whereby a current value of the time series X_t depends on the immediate past p values: $X_{t-1}, X_{t-2}, ..., X_{t-p}$. On the other hand an MA(q) model of order q is such that the current value X_t is a linear combination of immediate past values of the white noise process: $\varepsilon_1, \varepsilon_2, ..., \varepsilon_q$. Apart from stationarity, invertibility is another important requirement for a time series. It refers to the property whereby the covariance structure of the eries is unique [7]. Moreover it allows for meaningful association of current events with the past history of the series [2].

An AR(p) model may be more specifically written as

$$X_t + \alpha_{p1}X_{t-1} + \alpha_{p2}X_{t-2} + ... + \alpha_{pp}X_{t-p} = \varepsilon_t$$

Then the sequence of the last coefficients$\{\alpha_{ii}\}$ is called *the partial autocorrelation function*(PACF) of $\{X_t\}$. The ACF of an MA(q) model cuts off after lag q whereas that of an AR(p) model is a combination of sinusoidals dying off slowly. On the other hand the PACF of an MA(q) model dies off slowly whereas that of an AR(p) model cuts off after lag p. AR and MA models are known to have some duality properties. These include:

1. A finite order AR model is equivalent to an infinite order MA model.
2. A finite order MA model is equivalent to an infinite order AR model.
3. The ACF of an AR model exhibits the same behaviour as the PACF of an MA model.
4. The PACF of an AR model exhibits the same behaviour as the ACF of an MA model.
5. An AR model is always invertible but is stationary if $\boldsymbol{\alpha}(B) = 0$ has zeros outside the unit circle.
6. An MA model is always stationary but is invertible if $\boldsymbol{\beta}(B) = 0$ has zeros outside the unit circle.

Parametric parsimony consideration in model building entails preference for the mixed ARMA fit to either the pure AR or the pure MA fit. Stationarity and invertibility conditions for model (1) or (2) are that the equations $\boldsymbol{\alpha}(B) = 0$ and $\boldsymbol{\beta}(B) = 0$ should have roots outside the unit circle respectively.

***Corresponding Author:** Ette Harrison ETUK, Department of Mathematics/Computer Science, Rivers State University of Science and Technology, Nigeria. Email: ettetuk@yahoo.com

Often, in practice, a time series is non-stationary. Box and Jenkins [2] proposed that differencing of an appropriate data could render a non-stationary series $\{X_t\}$ stationary. Let degree of differencing necessary for stationarity be d. Such a series $\{X_t\}$ may be modelled as

$$(1 + \sum_{i=1}^{p} \alpha_i B^i)\nabla^d X_t = \beta(B)\varepsilon_t \qquad (3)$$

where $\nabla = 1 - B$ and in which case $\alpha(B) = (1 + \sum_{i=1}^{p} \alpha_i B^i)\nabla^d = 0$ shall have unit roots d times. Then differencing to degree d renders the series stationary. The model (3)
is said to be an autoregressive integrated moving average model of orders p, d and q and designated ARIMA(p, d, q). The purpose of this paper is to fit an ARIMA model to monthly unemployment rate data of Nigeria.

MATERIALS AND METHODS

The data for this work are monthly unemployment rate data from 1999 to 2008 obtainable from quarterly abstracts of the Central Bank of Nigeria. Unemployment rate in this context is the percentage of the workforce that are without jobs.

Determination of the differencing order d:

Preliminary analysis of time series involves the time-plot and the correlogram. A stationary time series exhibits no trend and the degree of variability is invariant with time. In addition the covariance is a function of the time lag. The time plot of a stationary time series shows no change in the mean level as well as the variance over time. The autocorrelation function should decay fast to zero.

Test for stationarity:

The ACF of a non-stationary time series starts high and declines slowly. Moreover to test for stationarity we shall be using the Augmented Dickey-Fuller (ADF) test. This involves testing for b=1 against b < 1 in $X_t = a + bX_{t-1} + \varepsilon_t$. The software Eviews 3.1 that we shall use has facility for the ADF test also.

Determination of the orders p and q:

As already mentioned above, an AR(p) model has a PACF that truncates at lag p and an MA(q)) has an ACF that truncates at lag q. In practice $\pm 2\sqrt{n}$ are the nonsignificance limits for both functions. We shall explore the range of models ARMA(a,b), $0 \le a \le p$, $0 \le b \le q$ for an optimum one. To do this we shall use the automatic model determination criteria AIC and SIC (e.g. [1], [3], [4] and[8]) defined by:

$$AIC(p+d+q) = n \ln \tilde{\sigma}^2_{p+d+q} + 2(p+d+q)$$

$$SIC(p+d+q) = n \ln \tilde{\sigma}^2_{p+d+q} + (p+d+q)\ln(n)/n$$

where $\widetilde{\sigma_k^2}$ is the maximum likelihood estimate of the residual variance when the model has k parameters.The optimum model corresponds to the minimum of the criteria within the explored range.

Model Estimation:

The involvement of the white noise terms in an ARIMA model entails a nonlinear iterative process in the estimation of the parameters, α_i's and β_j's. An optimization criterion like least error of sum of squares, maximum likelihood or maximum entropy is used. An initial estimate is usually used. Each iteration is expected to be an improvement of the last one until the estimate converges to an optimal one. However,for pure AR and pure MA models linear optimization techniques exist (See for example [2],[6]). There are attempts to adopt linear methods to estimate ARMA models (See for example, [3], [4], [5]).

Diagnostic Checking:

The model that is fitted to the data should be tested for goodness-of-fit. The automatic order determination criteria AIC and SIC are themselves diagnostic checking tools. Further checking can be done by the analysis of the residuals of the model. If the model is correct, the residuals would be uncorrelated and would follow a normal distribution with mean zero and constant variance.

RESULTS AND DISCUSSION

The time plot of the original series NUMP in Fig.1and the correlogram of Figure 2 clearly depict non-stationarity. Differencing the series once yields a still non-stationary process, DNUMP; the ADF test of Table 1 confirms the non-stationary nature. This necessitated second order differencing. The ADF test of Table 2 attests to the stationary nature of the second differences SNUMP. We note that in this table the dependent variable is the third difference TNUMP of the original series. From fig. 4, the ACF cuts off at lag 5 and the PACF at lag 4. Exploring the range of models {ARMA(p,q): $0 \le p \le 4$, $0 \le q \le 5$}for the optimal on the basis of AIC and SIC yields an ARMA(1,1) as summarized in Table 3.

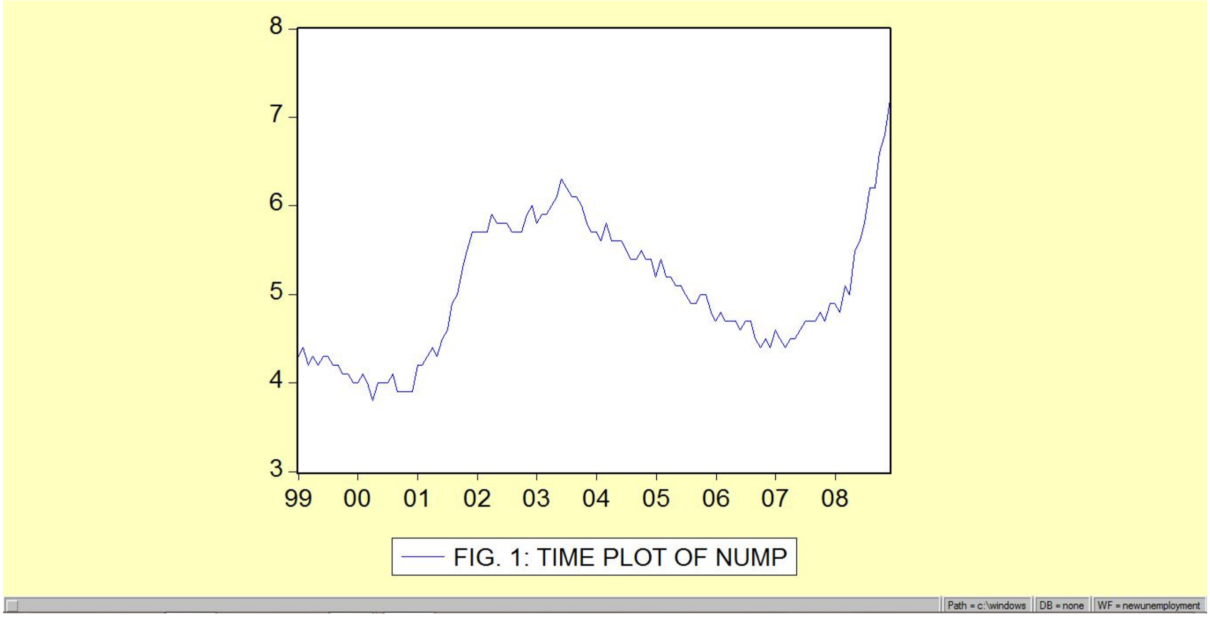
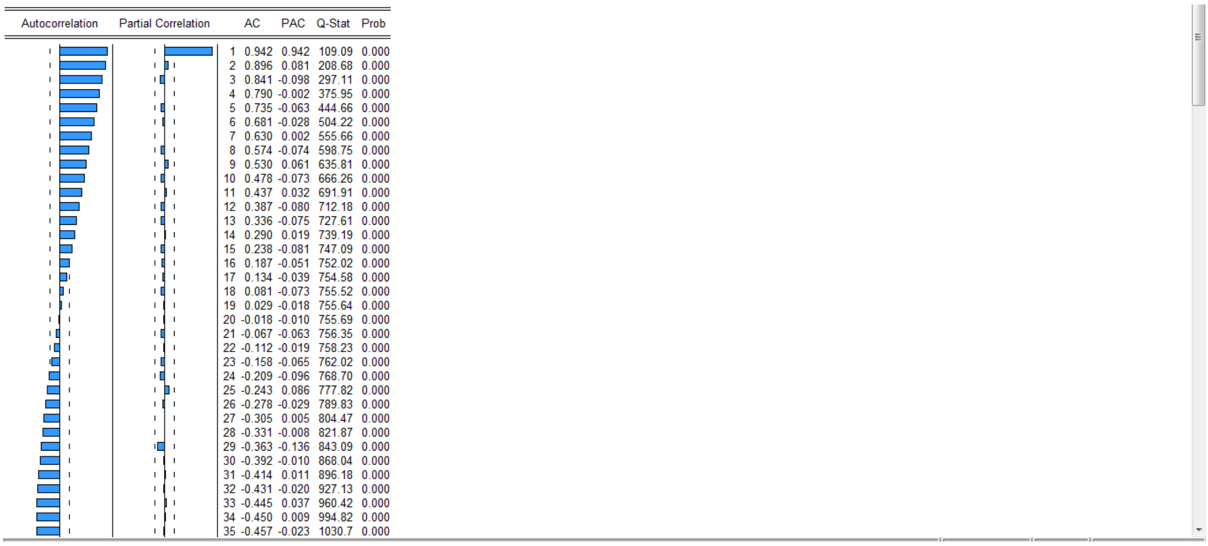


FIG. 1: TIME PLOT OF NUMP



	AC	PAC	Q-Stat	Prob
1	0.942	0.942	109.09	0.000
2	0.896	0.081	208.68	0.000
3	0.841	-0.098	297.11	0.000
4	0.790	-0.002	375.95	0.000
5	0.735	-0.063	444.66	0.000
6	0.681	-0.028	504.22	0.000
7	0.630	0.002	555.66	0.000
8	0.574	-0.074	598.75	0.000
9	0.530	0.061	635.81	0.000
10	0.478	-0.073	666.26	0.000
11	0.437	0.032	691.91	0.000
12	0.387	-0.080	712.18	0.000
13	0.336	-0.075	727.61	0.000
14	0.290	0.019	739.19	0.000
15	0.238	-0.081	747.09	0.000
16	0.187	-0.051	752.02	0.000
17	0.134	-0.039	754.58	0.000
18	0.081	-0.073	755.52	0.000
19	0.029	-0.018	755.64	0.000
20	-0.018	-0.010	755.69	0.000
21	-0.067	-0.063	756.35	0.000
22	-0.112	-0.019	758.23	0.000
23	-0.158	-0.065	762.02	0.000
24	-0.209	-0.096	768.70	0.000
25	-0.243	0.086	777.82	0.000
26	-0.278	-0.029	789.83	0.000
27	-0.305	0.005	804.47	0.000
28	-0.331	-0.008	821.87	0.000
29	-0.363	-0.136	843.09	0.000
30	-0.392	-0.010	868.04	0.000
31	-0.414	0.011	896.18	0.000
32	-0.431	-0.020	927.13	0.000
33	-0.445	0.037	960.42	0.000
34	-0.450	0.009	994.82	0.000
35	-0.457	-0.023	1030.7	0.000

FIG. 2. CORRELOGRAM OF NUMP

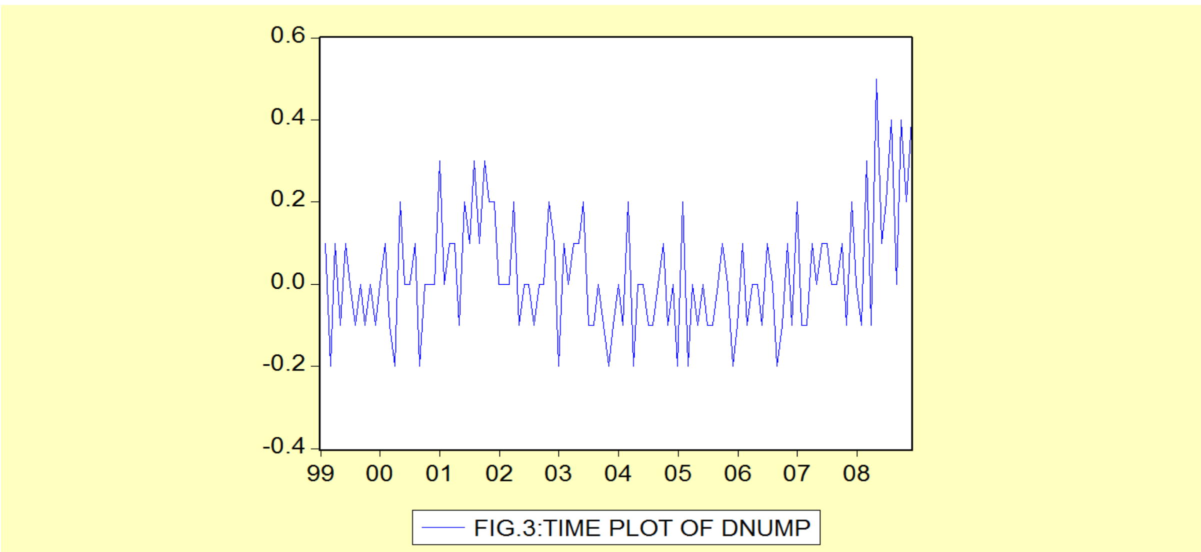


FIG.3:TIME PLOT OF DNUMP

TABLE 1: Augmented Dickey Fuller Test on DNUMP

ADF Test Statistic	**-1.400118**	**1% Critical Value***		**-3.4885**
		5% Critical Value		-2.8868
		10% Critical Value		-2.5801
***MacKinnon critical values for rejection of hypothesis of a unit root.**				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: SDNUMP				
Method: Least Squares				
Date: 12/24/11 Time: 17:16				
Sample(adjusted): 1999:07 2008:12				
Included observations: 114 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(X(-1))	-0.222120	0.158644	-1.400118	0.1643
D(X(-1),2)	-0.830500	0.161518	-5.141845	0.0000
D(X(-2),2)	-0.519311	0.166142	-3.125705	0.0023
D(X(-3),2)	-0.276272	0.147540	-1.872524	0.0638
D(X(-4),2)	-0.243273	0.099133	-2.454007	0.0157
C	0.012374	0.012392	0.998626	0.3202
R-squared	0.596501	Mean dependent var		0.002632
Adjusted R-squared	0.577821	S.D. dependent var		0.197981
S.E. of regression	0.128639	Akaike info criterion		-1.212419
Sum squared resid	1.787181	Schwarz criterion		-1.068408
Log likelihood	75.10788	F-statistic		31.93177
Durbin-Watson stat	2.014234	Prob(F-statistic)		0.000000

TABLE 2: Augmented Dickey Fuller Test on SDNUMP

ADF Test Statistic	**-7.831665**	**1% Critical Value***		**-3.4890**
		5% Critical Value		-2.8870
		10% Critical Value		-2.5802
*MacKinnon critical values for rejection of hypothesis of a unit root.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: TNUMP				
Method: Least Squares				
Date: 12/24/11 Time: 17:10				
Sample(adjusted): 1999:08 2008:12				
Included observations: 113 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(X(-1),2)	-3.783728	0.483132	-7.831665	0.0000
D(X(-1),3)	1.730148	0.424541	4.075343	0.0001
D(X(-2),3)	1.016548	0.331170	3.069566	0.0027
D(X(-3),3)	0.555296	0.215904	2.571960	0.0115
D(X(-4),3)	0.130074	0.097086	1.339781	0.1832
C	0.009594	0.012215	0.785434	0.4339
R-squared	0.878016	Mean dependent var		0.002655
Adjusted R-squared	0.872316	S.D. dependent var		0.361905
S.E. of regression	0.129319	Akaike info criterion		-1.201433
Sum squared resid	1.789405	Schwarz criterion		-1.056616
Log likelihood	73.88096	F-statistic		154.0331
Durbin-Watson stat	2.003449	Prob(F-statistic)		0.000000

FIG. 4: CORRELOGRAM OF SNUMP

Table 3. Comparison of models within the range of exploration using AIC and SIC

p	q	AIC	SIC
0	1	-0.997	-0.974
0	2	-1.234	-1.187
0	3	-1.223	-1.153
0	4	-1.116	-1.022
0	5	-1.208	-1.091
1	0	-0.999	-0.975
1	1	-1.250	-1.203
1	2	-1.235	-1.164
1	3	-1.224	-1.129
1	4	-1.215	-1.097
1	5	-1.150	-1.009
2	0	-1.174	-1.127
2	1	-1.231	-1.159
2	2	-1.220	-1.126
2	3	-1.234	-1.116
2	4	-1.217	-1.075
2	5	-1.215	-1.049
3	0	-1.158	-1.086
3	1	-1.212	-1.117
3	2	-1.225	-1.105
3	3	-1.208	-1.065
3	4	-1.194	-1.027
3	5	-1.190	-0.999
4	0	-1.225	-1.129
4	1	-1.229	-1.109
4	2	-1.211	-1.067
4	3	-1.238	-1.070
4	4	-1.222	-1.030
4	5	-1.218	-1.002

TABLE 4: Model Estimation

Dependent Variable: SDNUMP				
Method: Least Squares				
Date: 12/24/11 Time: 18:32				
Sample(adjusted): 1999:04 2008:12				
Included observations: 117 after adjusting endpoints				
Convergence achieved after 6 iterations				
Backcast: 1999:03				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.388391	0.099793	-3.891961	0.0002
MA(1)	-0.716548	0.077618	-9.231771	0.0000
R-squared	0.587306	Mean dependent var		0.005128
Adjusted R-squared	0.583717	S.D. dependent var		0.199069
S.E. of regression	0.128440	Akaike info criterion		-1.249770
Sum squared resid	1.897123	Schwarz criterion		-1.202554
Log likelihood	75.11156	F-statistic		163.6567
Durbin-Watson stat	1.966983	Prob(F-statistic)		0.000000
Inverted AR Roots	-.39			
Inverted MA Roots	.72			

The chosen model as summarized in Table 4 is ARIMA(1, 2, 1) and is given by

$$SDNUMP_t = -0.388391SDNUMP_{t-1} - 0.716548\varepsilon_{t-1} + \varepsilon_t$$
$$(\pm 0.099793) \qquad\qquad (\pm 0.077618)$$

Clearly non-linear techniques used by Eviews 3.1 involved an iterative process that converged after six iterations. We observe that the coefficients are both highly significant, each being more than twice its standard error. The roots of $\alpha(B) = 0$ and $\beta(B) = 0$ are -2.56 and 1.39, both outside the unit circle indicating stationarity and invertibility respectively. Besides the residual plot of Fig. 5 confirms that the residuals follow the normal distribution with zero (actually 0.01) mean. The kurtosis is 2.8 which compares favourably with the normal distribution standard of 3. 0.

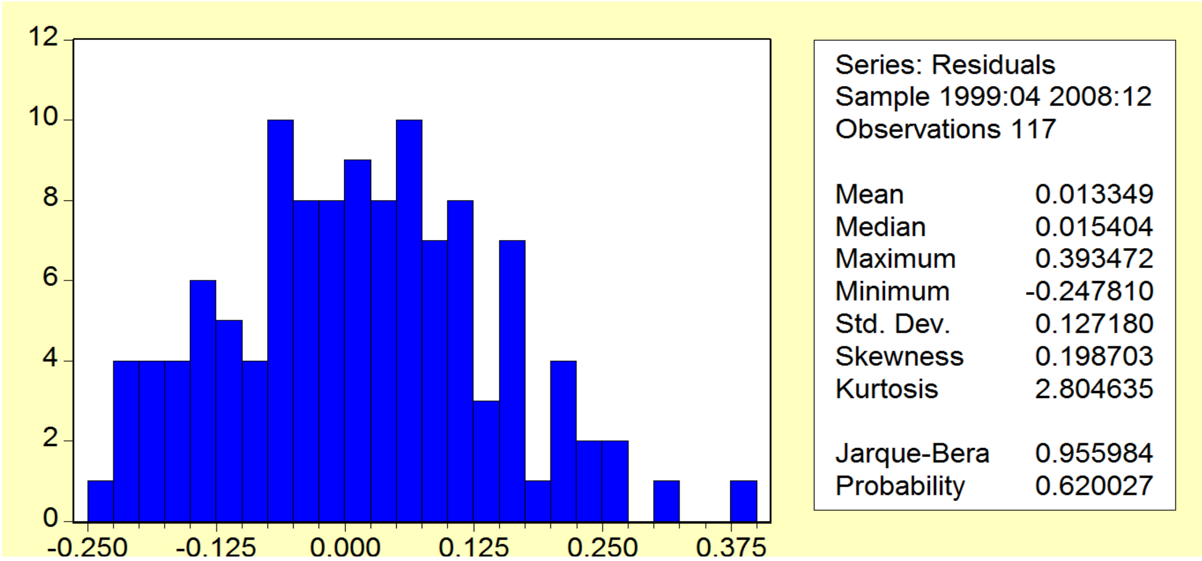


FIG. 5. HISTOGRAM OF THE MODEL RESIDUALS

Forecasting:

An ARIMA(1, 2, 1) model may be written as

$$\nabla^2 X_t = \alpha_1 \nabla^2 X_{t-1} + \beta_1 \varepsilon_{t-1} + \varepsilon_t$$

This translates into

$$X_t - 2X_{t-1} + X_{t-2} = \alpha_1 (X_{t-1} - 2X_{t-2} + X_{t-3}) + \beta_1 \varepsilon_{t-1} + \varepsilon_t$$

That is,

$$X_t = (\alpha_1 + 2) X_{t-1} - (1 + 2\alpha_1) X_{t-2} + \alpha_1 X_{t-3} + \beta_1 \varepsilon_{t-1} + \varepsilon_t$$

At time t+k, the model may be written as

$$X_{t+k} = (\alpha_1 + 2) X_{t+k-1} - (1 + 2\alpha_1) X_{t+k-2} + \alpha_1 X_{t+k-3} + \beta_1 \varepsilon_{t+k-1} + \varepsilon_t$$

Taking conditional expectations at time t, we have

$$\hat{X}_t(1) = (\alpha_1 + 2) X_t - (1 + 2\alpha_1) X_{t-1} + \alpha_1 X_{t-2} + \beta_1 \varepsilon_t$$
$$\hat{X}_t(2) = (\alpha_1 + 2) \hat{X}_t(1) - (1 + 2\alpha_1) X_t + \alpha_1 X_{t-1}$$
$$\hat{X}_t(3) = (\alpha_1 + 2) \hat{X}_t(2) - (1 + 2\alpha_1) \hat{X}_t(1) + \alpha_1 X_t$$
$$\hat{X}_t(k) = (\alpha_1 + 2) \hat{X}_t(k-1) - (1 + 2\alpha_1) \hat{X}_t(k-2) + \alpha_1 \hat{X}_t(k-3), \qquad k \geq 4$$

where $\hat{X}_t(k)$ is the k-step ahead forecast. That is the forecast of X_{t+k}.

TABLE 6. Forecasts

	Residuals	SNUMP	DNUMP	NUMP
October 2008	0.14626	0.4	0.4	6.6
November 2008	0.06016	-0.2	0.2	6.8
December 2008	0.16543	0.2	0.4	7.2
January 2009		0.33015	0.73015	7.9
February 2009		0.56511	1.29526	9.2
March 2009		0.75936	2.05462	11.3

Conclusion

We have successfully fitted an ARIMA(1, 2, 1) model to Nigerian monthly unemployment data. This means that the second differences SNUMP follow an ARMA(1,1) model. Its adequacy has been established and on its basis we have made forecasts.

REFERENCES

1. Akaike, H. (1970). *Statistical Predictor Identification.* Annals of the Institute of Statistical Mathematics, Volume 22: pp. 203 – 217.
2. Box, G. E. P. And Jenkins, G. M. (1976). Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco.
3. Etuk, E. H. (1987). *On the Selection of Autoregressive Moving Average Models.* An unpublished Ph. D. Thesis, Department of Statistics, University of Ibadan, Nigeria.
4. Etuk, E. H. (1988). *On Autoregressive Model Identification.* Journal of Official Statistics, Volume 4, No. 2; pp. 113 – 124.
5. Etuk, E. H. (1996). *An Autoregressive Integrated Moving Average(ARIMA) Simulation Model: A Case Study.* Discovery and Innovation, Volume 10, Nos 1 & 2:pp. 23 – 26.
6. Oyetunji, O. B. (1985). *Inverse Autocorrelations and Moving Average Time Series Modelling.* Journal of Official Statistics, Volume 1: pp. 315 – 322.
7. Priestley, M. B. (1981). Spectral Analysis and Time Sreies. Academic Press, London.
8. Schwarz, G. (1978). *Estimating the dimension of a model.* Annals of Statistics, Volume 6: pp. 461 – 464.