Research Article

Box-Jenkins Modelling of Nigerian Stock Prices Data

Ette Harrison Etuk*, Bartholomew Uchendu, Ephraim Okon Udo

Department of Mathematics/Computer Science, Rivers State University of Science and Technology, Nigeria

*Corresponding Author’s Email: ettetuk@yahoo.com

ABSTRACT

Nigerian stock prices data is modelled by Box-Jenkins approach and the use of automatic model selection criteria: Akaike Information criterion (AIC), Schwarz Information Criterion (SIC), $R^2$. It is inferred that the most adequate model is autoregressive integrated moving average of orders 2, 1 and 3 (ARIMA (2,1,3)). Forecasts are obtained on the basis of the model.

Key Words: Stock prices, 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:

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

Or

$$\alpha(B)X_t = \beta(B)\varepsilon_t$$ (2)

where $\{\varepsilon_t\}$ is a sequence of random variables with zero mean and constant variance, called a white noise process, and the $\alpha_i$’s and $\beta_i$’s constants; $\alpha(B) = 1 + \alpha_1 B + \alpha_2 B^2 + ... + \alpha_p B^p$ and $\beta(B) = 1 + \beta_1 B + \beta_2 B^2 + ... + \beta_q B^q$ 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 series is unique (Priestley, 1981). Moreover it allows for meaningful association of current events with the past history of the series (Box and Jenkins, 1976).

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

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} = \varepsilon_t$$

Then the sequence of the last coefficients $\alpha_i$ 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 behavior as the ACF of an MA model.
5. An AR model is always invertible but is stationary if \( \alpha(B) = 0 \) has zeros outside the unit circle.
6. An MA model is always stationary but is invertible if \( \beta(B) = 0 \) 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 \( \alpha(B) = 0 \) and \( \beta(B) = 0 \) should have roots outside the unit circle respectively.

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

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

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 Nigerian stock prices.

MATERIALS AND METHODS

The data for this work are monthly stock prices data from January 1987 to December 2006 obtained from Nigerian Stock Exchange Office, Port Harcourt, Nigeria.

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 \( 0 \leq a \leq p \) and \( 0 \leq b \leq q \) for an optimum one. To do this we shall use the automatic model determination criteria AIC and SIC (e.g. Akaike(1970), Etuk (1987, 1988), and Schwarz(1978)) defined by:

\[
AIC(p + d + q) = n \ln \hat{\sigma}^2_{p+d+q} + 2(p + d + q)
\]

\[
SIC(p + d + q) = n \ln \hat{\sigma}^2_{p+d+q} + (p + d + q) \ln(n) / n
\]

where \( \hat{\sigma}^2 \) 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, \( \alpha_i \)'s and \( \beta_j \)'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 Box and Jenkins (1976), Oyetunji (1985)). There are attempts to adopt linear methods to estimate ARMA models (See for example, Etuk(1987, 1988, 1996)).
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 NSP in Figure 1, the correlogram of Figure 2 and the ADF test of Table 1 clearly depict non-stationarity. Differencing the series once yields a stationary process, DNSP; the time plot is in Figure 3, the correlogram in Figure 4 and the ADF test in Table 2. We note that in this table the dependent variable is the second difference SNNSP of the original series. From fig. 4, the ACF cuts off at lag 5 and PACF at lag 4. Exploring the range of models \( \text{ARMA}(p,q): 0 \leq p \leq 4, 0 \leq q \leq 5 \) which are stationary and invertible with positive \( R^2 \) for the optimal on the basis of AIC and SIC yields an ARMA(2, 3). The model estimation is summarized in Table 3.

![FIG. 1: STOCK PRICES PLOT](image1)

![FIGURE 2: CORRELOGRAM OF NSP](image2)
TABLE 1: ADF TEST FOR NSP

| Variable   | Coefficient | Std. Error | t-Statistic | Prob. |
|------------|-------------|------------|-------------|-------|
| X(-1)      | -0.005154   | 0.019379   | -0.265938   | 0.7906|
| D(X(-1))   | -0.016566   | 0.076634   | -0.231399   | 0.8172|
| D(X(-2))   | -0.051026   | 0.074853   | -0.681676   | 0.4961|
| D(X(-3))   | 0.049534    | 0.074899   | 0.541182    | 0.5849|
| D(X(-4))   | 0.129850    | 0.074983   | 1.732766    | 0.0845|
| C          | 0.131795    | 0.151346   | 0.870810    | 0.3648|

R-squared: 0.019557  Mean dependent var: 0.104609
Adjusted R-squared: -0.002872  S.D. dependent var: 1.310637
S.E. of regression: 1.312518  Akaike info criterion: 3.406972
Sum squared resid: 394.4996  Schwarz criterion: 3.493602
Log likelihood: -394.3192  F-statistic: 0.865994
Durbin-Watson stat: 1.897009  Prob(F-statistic): 0.504663

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(X)
Method: Least Squares
Date: 01/08/12  Time: 20:32
Sample(adjusted): 1987:06 2006:12
Included observations: 235 after adjusting endpoints

---

![FIG. 3: FIRST DIFFERENCES](image-url)
TABLE 2: ADF TEST ON DNSP

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|----|-----|--------|------|
| 1               | -0.670              | 64.302 | 0.000 |
| 2               | 0.217              | 60.069 | 0.000 |
| 3               | 0.843              | 60.292 | 0.000 |
| 4               | -0.222         | 66.411 | 0.000 |
| 5               | 0.252              | 74.339 | 0.000 |
| 6               | -0.156         | 77.407 | 0.000 |
| 7               | 0.047             | 77.685 | 0.000 |
| 8               | 0.034             | 77.835 | 0.000 |
| 9               | 0.094              | 83.955 | 0.000 |
| 10              | -0.098            | 79.099 | 0.000 |
| 11              | 0.124             | 81.123 | 0.000 |
| 12              | -0.111           | 82.769 | 0.000 |
| 13              | 0.094             | 83.955 | 0.000 |
| 14              | -0.038           | 84.135 | 0.000 |
| 15              | 0.066             | 84.375 | 0.000 |
| 16              | 0.124             | 86.859 | 0.000 |
| 17              | -0.190           | 86.618 | 0.000 |
| 18              | 0.090             | 89.022 | 0.000 |
| 19              | -0.071           | 90.633 | 0.000 |
| 20              | 0.023             | 90.612 | 0.000 |
| 21              | 0.034             | 90.769 | 0.000 |
| 22              | -0.083           | 91.802 | 0.000 |
| 23              | 0.147             | 95.030 | 0.000 |
| 24              | -0.159           | 96.669 | 0.000 |
| 25              | 0.047             | 99.014 | 0.000 |
| 26              | -0.026           | 99.116 | 0.000 |
| 27              | 0.035             | 99.902 | 0.000 |
| 28              | 0.029             | 99.241 | 0.000 |
| 29              | -0.015           | 99.477 | 0.000 |
| 30              | 0.039             | 99.716 | 0.000 |
| 31              | -0.049           | 100.11  | 0.000 |
| 32              | 0.041             | 100.38  | 0.000 |
| 33              | -0.198           | 102.35  | 0.000 |

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(X(2))
Method: Least Squares
Date: 01/09/12   Time: 08:27
Sample (adjusted): 1987:07 2006:12
Included observations: 234 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|-------|
| D(X(-1))| -1.02694    | 0.156939   | -6.543283   | 0.000 |
| D(X(-1.2))| 0.067282  | 0.140469   | 0.051843    | 0.9587 |
| D(X(-2.2))| -0.035774 | 0.127947   | -0.279604   | 0.7800 |
| D(X(-3.2))| -0.014370 | 0.102848   | -0.139715   | 0.8890 |
| D(X(-4.2))| 0.118165  | 0.077514   | 1.523660    | 0.1290 |
| C        | 0.094750    | 0.085966   | 1.024275    | 0.2714 |

R-squared   | 0.508131   | Mean dependent var | 0.015296 |
Adjusted R-squared | 0.497344  | S.D. dependent var | 1.835381 |
S.E. of regression | 1.299577  | Akaike info criterion | 3.387876 |
Sum squared resid | 385.3062  | Schwarz criterion | 3.476474 |
Log likelihood | -390.3815 | F-statistic | 47.10756 |
Durbin-Watson stat | 1.940182 | Prob(F-statistic) | 0.000000 |
TABLE 3: MODEL ESTIMATION

| Variable  | Coefficient | Std. Error | t-Statistic | Prob. |
|-----------|-------------|------------|-------------|-------|
| AR(1)     | -0.061736   | 0.019321   | -3.195196   | 0.0016|
| AR(2)     | -0.591882   | 0.020070   | -49.42218   | 0.0000|
| MA(1)     | 0.046505    | 0.070918   | 0.655756    | 0.5126|
| MA(2)     | 0.970226    | 0.014925   | 65.00586    | 0.0000|
| MA(3)     | 0.035801    | 0.067853   | 0.527630    | 0.5983|

R-squared: 0.040931
Adjusted R-squared: 0.024396
S.E. of regression: 1.094975
S.E. of log residual: 0.019321
Akaike info criterion: 3.366589
Schwarz criterion: 3.439764
Prob(F-statistic): 0.045093

The chosen model as summarized in Table 3 is ARIMA(2, 1, 3) and is given by

\[ DNSP_t = -0.061736DNSP_{t-1} - 0.591882DNSP_{t-2} + 0.046505e_{t-1} + 0.970226e_{t-2} + 0.035801e_{t-3} + e_t \]

Clearly non-linear techniques used by Eviews 3.1 involved an iterative process that converged after thirty one iterations. We observe that only the first and third MA coefficients are not significant, each being less than twice its standard error. The roots of \( \alpha(B) = 0 \) and \( \beta(B) = 0 \) all lie 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 0.1) mean.

Forecasting: An ARIMA(2, 1, 3) is of the form
\[ \nabla X_t = \alpha_1 \nabla X_{t-1} + \alpha_2 \nabla X_{t-2} + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \beta_3 \epsilon_{t-3} + \epsilon_t \]

\[ X_t = (1 + \alpha_1) X_{t-1} + (\alpha_2 - \alpha_1) X_{t-2} - \alpha_2 X_{t-3} + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \beta_3 \epsilon_{t-3} + \epsilon_t \]

At time \( t+k \),

\[ X_{t+k} = (1 + \alpha_1) X_{t+k-1} + (\alpha_2 - \alpha_1) X_{t+k-2} - \alpha_2 X_{t+k-3} + \beta_1 \epsilon_{t+k-1} + \beta_2 \epsilon_{t+k-2} + \beta_3 \epsilon_{t+k-3} + \epsilon_{t+k} \]

Taking conditional expectations at time \( t \),

\[ \hat{X}_t(1) = (1 + \alpha_1) X_t + (\alpha_2 - \alpha_1) X_{t-1} - \alpha_2 X_{t-2} + \beta_1 \epsilon_t + \beta_2 \epsilon_{t-1} + \beta_3 \epsilon_{t-2} \]

\[ \hat{X}_t(2) = (1 + \alpha_1) \hat{X}_t(1) + (\alpha_2 - \alpha_1) X_t - \alpha_2 X_{t-1} + \beta_1 \epsilon_t + \beta_3 \epsilon_{t-1} \]

\[ \hat{X}_t(3) = (1 + \alpha_1) \hat{X}_t(2) + (\alpha_2 - \alpha_1) \hat{X}_t(1) - \alpha_2 X_t + \beta_3 \epsilon_t \]

\[ \hat{X}_t(k) = (1 + \alpha_1) \hat{X}_t(k-1) + (\alpha_2 - \alpha_1) \hat{X}_t(k-2) - \alpha_2 \hat{X}_t(k-3), \ k \geq 4 \]

where \( \hat{X}_t(k) \) is the \( k \)-point ahead forecast. That is the forecast of \( X_{t+k} \) given the series up to \( X_t \).

Table 6. Forecasts

| TIME            | RESIDUALS | DNSP  | NSP  |
|-----------------|-----------|-------|------|
| October 2006    | 3.16106   | 2.60  | 26.25|
| November 2006   | -5.17258  | -5.26 | 20.99|
| December 2006   | 5.35089   | 6.01  | 27.00|
| January 2007    |           |       | 27.19|
| February 2007   |           |       | 26.22|
| March 2007      |           |       | 26.28|
| April 2007      |           |       | 27.24|

CONCLUSION:

We have fitted an adequate ARIMA (2,1,3) model to Nigerian Stock Prices. That means that the first differences DNSP follow an ARMA (2,3) model. On the basis of the model, we have made some forecasts.

REFERENCES

Akaike, H. (1970). Statistical Predictor Identification. Annals of the Institute of Statistical Mathematics, Volume 22: pp. 203 – 217.

Box, G. E. P. And Jenkins, G. M. (1976). Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco.

Etuk, E. H. (1987). On the Selection of Autoregressive Moving Average Models. An unpublished Ph. D. Thesis, Department of Statistics, University of Ibadan, Nigeria.

Etuk, E. H. (1988). On Autoregressive Model Identification. Journal of Official Statistics, Volume 4, No. 2; pp. 113 – 124.

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.

Oyetunji, O. B. (1985). Inverse Autocorrelations and Moving Average Time Series Modelling. Journal of Official Statistics, Volume 1: pp. 315 – 322.

Priestley, M. B. (1981). Spectral Analysis and Time Series. Academic Press, London.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, Volume 6: pp. 461 – 464.