__Original Research Article__

Linear Time Series Model Selection: The Out-of-Sample Approach

ABSTRACT

Background: In linear time series, the in-sample model selection and the out-of-sample model selection are the two common approaches to model selection. However, empirical evidence based on out-of-sample forecast performance is generally considered more trustworthy than evidence based on in-sample performance, which is deficient in providing information about future observations.
Aim: The aim of this study is to apply the out-of-sample approach in model selection on Nigeria exchange rate with a view to identifying and selecting the best model for predicting Nigeria exchange rate.
Material and Methods: Data from the Nigeria exchange rate of naira to pound and naira to euro from January 2002 to December 2018, comprising of 204 data points were considered. The Box-Jenkins ARIMA iterative procedure was used in model building while the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and Theil's U coefficient were the measures of accuracy adopted in selecting the best out-of sample model.
Results: Our result reviewed that, based on in-sample model selection, ARIMA (0, 1, 1) and ARIMA (1, 1, 0) were the appropriate models with minimum information criteria. However, on the basics of out-sample forecast performance evaluation, ARIMA (1, 1, 0) and ARIMA (1, 1, 2) were found to be appropriate out-of-sample models with minimum forecast evaluation criteria. In all, our results revealed that, the out-of sample models performed better than their in-sample counterparts in their ability to forecast future values.
Conclusion: So far, this study showed that out-of-sample is a better model selection criterion than the in-sample counterpart as evident in its ability to predict future values which is the very essence for modelling in time series.

Keywords; in-sample, Out-of-sample, model selection, ARIMA model, time series

1.0 INTRODUCTION

Model selection is the task of selecting a statistical model from a set of candidate models [1]. The objective of model selection is to discover a model that optimises a process because exact model that describe a particular series is not known. Also, model selection is useful in comparing competing models for the purpose of selecting the best model that describe the series [2]. One sensitive challenge of model selection is that several competing models may fit a particular series appropriately [3]. According to [4] inappropriate model selection results in a choice of a poor model with far reaching consequences. [5] Opined that modelling is approximation of reality, thus model selection is to reject a model far from reality and select that which is close to reality.

In linear time series several competing models may adequately fit a particular series and basically there are two approaches to model selection namely; the in-sample model selection and the out-of-sample model selection. The in-sample model selection criteria include AIC [6], BIC [7] and Hanna and Quinn information criteria [8]. Model selection based on in-sample criteria such as Akaike information criteria, Schwarz information criteria and Hannan Quinn information criteria may not provide more genuine forecasts because it is the same data used in model estimation that is also used in forecast evaluation [9]. Also, a model selected on the basis of in-sample criteria does not give information about the future observations. On the other hand, out-of-sample model selection procedure is applied to achieve best predictive performance. The out-of-sample model selection procedure is accomplish by withholding some of the sample data from the model identification and estimation process, then use the model to make predictions for the hold-out data in order to see how accurate they are and to determine whether the statistics of their errors are similar to those that the model made within the sample of data that was fitted. The data which are not held out are used to estimate the parameters of the model. The model is then tested on data in the validation period, and forecasts are generated beyond the end of the estimation and validation periods [10]. The out-of-sample forecasting is advantageous over the in-sample forecasting in that, the model selection is based on how best the forecasts perform and able to provide information about future observations.

Thus, the aim of this work is to apply the out-of-sample in model selection on Nigeria exchange rate to improve on the work of several authors such as [11, 12, 13, 14, and 15] who applied in-sample model selection on Nigeria exchange rate.

The remaining part of this work is organized as follows; materials and methods are presented in section 2, results and discussion treated in section 3 while conclusion of the study is handled in section 4.

2.0 MATERIALS AND METHODS

2. 1. An Autoregressive Process AR (p)

Mathematically the AR (p) model can be expressed as [16]

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t. \quad (1)$$

Here $X_t$and $\varepsilon_t$are respectively the actual value and random error (or random shock) at time t,φ_i(i =1, 2, p) = are model parameters and c is a constant. The integer constant p is known as the order of the model. Sometimes the constant term is omitted for simplicity.

The model in back shift operator is specified as:

$$(1 - \varphi_1 B + \varphi_2 B^2 - \cdots - \varphi_p B^p)X_t = \varepsilon_t, \quad (2)$$

where the lag backshift operator B is defined as $BX_t = X_{t-p}$, p = 0, 1, 2,…

More precisely we express the model as: $\varphi(B)X_t = \varepsilon_t$

The autoregressive operator $\varphi(B)$ is define as $\varphi(B) = 1 - \varphi_1 B + \varphi_2 B^2 - \cdots - \varphi_p B^p$

2.2 Moving Average Model MA (q)

The notation MA (q) is the moving average model of order q [16]

$$X_t = \mu + \varepsilon_t + \sum_{i=j}^{q} \theta_i \varepsilon_{t-j} = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \quad (3)$$

where θ_1,…, θ_q are MA (q) parameters to be estimated, μ is the mean of X_t and ε_t is the error term.

The model in backshift operator is given as

$$X_t = \theta(B)\varepsilon_t, \quad (4)$$

where,

(i) $B^q \varepsilon_t = \varepsilon_{t-q}$ is a backward shift operator.

(ii) $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$.

(iii) $\theta_1, \theta_2, \ldots, \theta_q$ is a finite set of weighted parameters.

(iv) ε_t is a white noise process with mean zero, and constant variance σ^2.

2.3 Autoregressive moving average

Autoregressive (AR) and moving average (MA) models can be effectively combined together to form a general and useful class of time series models known as the ARMA models. Mathematically an ARMA (p, q) model is represented as [16]

$$X_t = c + \varepsilon_{t.} \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_j \varepsilon_{t-j}, \quad (5)$$

where, the model order p, q refer to p autoregressive and q moving average terms.

The model in backshift operator is given as

$$\varphi(B)X_t = \theta(B)\varepsilon_t, \quad (6)$$

$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 \ldots - \ldots \varphi_p B^p$is the autoregressive coefficient polynomial

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 \ldots - \ldots \theta_p B^p$is the moving average coefficient polynomial

2.4 Autoregressive Integrated Moving Average.

Autoregressive integrated moving average (ARIMA) models are specific subset of univariate modelling, in which a time series is expressed in terms of past values of itself (the autoregressive component) plus current and lagged values of a “white noise” error term (the moving average component). ARIMA models are univariate models that consist of an autoregressive polynomial, an order of integration (d), and a moving average polynomial.

A process (X_t) is said to be an autoregressive integrated moving average process, denoted by ARIMA (p, d, q) if it can be written as:

$$\varphi(B)\nabla^d X_t = \theta(B)\varepsilon_t, \quad (7)$$

where ∇^d=$(1 - B)^d$ (1−B) with $\nabla^d X_t$and d^{th} consecutive differencing (Vandale, 1983)

if the expectation of $\nabla^d X_t$) = μ, we write the model as

$$\varphi(B)\nabla^d X_t = \alpha + \theta(B)\varepsilon_t, \quad (8)$$

where: α is a parameter related to the mean of the process (X_t), by α = μ ($\varphi_1 \dots \varphi_p$) and this process is called a white noise process, that is, a sequence of uncorrelated random variables from a fixed distribution with constant mean, μ ,usually assumed to be "zero" and constant variance. If d=0, it is called ARMA (p, q) model while when d=0 and q=0, it is referred to as autoregressive of order *p* model and denoted by AR *(p)*. When p=0 and d=0, it is called Moving Average of order *q* model, and is denoted by MA (q).
There are three steps we will take to achieve our aims, and these are listed as (1) model identification (2), model estimation and (3) model diagnostic and forecasting accuracy.

2.5 Box-Jenkins Modelling Approach

The Box-Jenkins model uses iterative three-stage modelling approach which is:

1 Model identification
2 Model estimation
3 Model checking

2.5.1 Model identification

The first step in developing a Box–Jenkins model is to determine if the time series is stationary. Stationarity, which in time series is a stochastic process whose unconditional joint distribution does not change when shifted by time, consequently, parameters such as mean and variance do not change over time can be assessed from time plot of the series. It can also be detected from an autocorrelation plot. Specifically, non-stationarity is often indicated by an autocorrelation plot with very slow decay. Finally, unit root tests provide a formal approach to determining the degree of differencing such as Augmented Dickey Fuller test (ADF)[17], Rothenberg and Stock (2001) introduced Efficient unit root (ERS) test statistic to test whether the series contains unit root or not [18] and Phillips-Perron Unit Root Tests[19]. But for the purpose of the work we shall consider the ADF test. Once stationarity has been addressed, the next step is to identify the order (i.e. the *p* and *q*) of the autoregressive and moving average terms. These are determined by examining the values of the autocorrelations and the partial autocorrelations with their corresponding plots (12).

2.5.1.1 Unit Root Test.

A test of stationarity that has become widely popular over the past years is the unit root test. Consider the following random walk model without drift and trend

$X_t = \rho X_{t-1} + \varepsilon_t$, -1≤ $\rho \leq 1$ (9)

Where,X_t is the actual series, $X_{t-1}$is the immediate previous observation, ε_t is a white noise error term and ρ is a parameter to be estimated.

$X_t - X_{t-1} = \rho X_{t-1} - X_{t-1} + \varepsilon_t;\ -1 \leq \rho \leq 1$ (10)

By subtracting X_{t-1} from both sides

$\Delta X_t = \sigma X_{t-1} + \varepsilon_t$, where $\sigma = (\rho - 1)$ and Δ is the first difference operator.

We test the hypothesis: $H_0: \sigma = 0$vs $H_1: \sigma < 0$, if the null hypothesis is accepted, $\sigma = 0$ and$\rho = 1$, that is we will have a unit root, meaning that the time series under consideration is not stationary.

For the purpose of this work we shall consider the Augmented Dickey Fuller test to test staionarity of the data because the ADF test is popular. Also, ADF test ensures that the null hypothesis is accepted unless there is strong evidence against it to reject in favour of the alternative stationarity hypothesis.

2.5.2 Model Estimation

After an optimal model has been identified, the model estimation methods make it possible to estimate simultaneously all the parameters of the process, the order of integration coefficient and parameters of an ARMA structure. There are many methods of estimating parameters of linear time series models but for the purpose of this study we shall consider the maximum likelihood method. It is often convenient to work with the log-likelihood function, which contains an *arbitrary additive constant*. One reason that the likelihood function is of fundamental importance in estimation theory is because of the *likelihood principle*. This principle states that (given that the assumed model is correct) all that the *data* have to tell us about the parameters is contained in the likelihood function, all other aspects of the data being irrelevant. Also from a Bayesian point of view, the likelihood function is equally important since it is the component in the posterior distribution of the parameters that comes from the data [20].

Only two methods of maximum likelihood estimation considered in time series analysis are discussed.

Given: $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \dots, \varphi_p)$, $\mu = E(X_t)$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_q)$ and $\sigma_e^2 = E(e_t^2)$ from observations of the causal $ARMA(p, q)$ process defined by

$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + a_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$. (11)

(i) Conditional Maximum Likelihood Estimation

For an $ARMA(p,q)$ model, the joint probability density function of [21]
$e = \{e_1, e_2, \ldots, e_n\}'$is given by

$$P(\boldsymbol{a}|\boldsymbol{\varphi}, \mu, \theta, \sigma_e^2) = (2\pi\sigma_e^2)^{\frac{2}{n}} exp\left\{-\frac{1}{2\sigma_e^2}\sum_{i=1}^{n} e_t^2\right\}. \quad (12)$$

rewriting (11) as

$$e_t = \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \cdots - \varphi_p X_{t-p}. \quad (13)$$

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ and assume initial conditions $\boldsymbol{X}_* = \left(X_{1-p}, \ldots, X_{-2}, X_{-1}, X_0\right)'$ and $\boldsymbol{e}_* = \left(e_{1-p}, \ldots, e_{-2}, e_{-1}, e_0\right)'$ are known. The conditional log-likelihood function is

$$lnL_*(\boldsymbol{\varphi}, \mu, \theta, \sigma_e^2) = -\frac{2}{n}(2\pi\sigma_e^2) - \frac{S_*(\boldsymbol{\varphi}, \mu, \theta)}{2\sigma_e^2} \quad (14)$$

where$S_*(\boldsymbol{\varphi}, \mu, \theta) = \sum_{i=1}^{n} e_t^2\,(\boldsymbol{\varphi}, \mu, \theta|\boldsymbol{X}_*, \boldsymbol{a}_*, \boldsymbol{X})$ (15)

is the conditional sum of squares function. The quantity of $\hat{\boldsymbol{\varphi}}, \hat{\mu}, \text{and}\hat{\theta}$ which maximize (15) are called the conditional estimators. Because $L_*(\boldsymbol{\varphi}, \mu, \theta, \sigma_e^2)$ involves the data only through $S_*(\boldsymbol{\varphi}, \mu, \theta)$, these estimators are the same as the conditional least squares estimators obtained from minimizing the conditional sum of squares function $S_*(\boldsymbol{\varphi}, \mu, \theta)$. The estimator $\hat{\sigma}_e^2$ of σ_e^2 is obtained from

$$\hat{\sigma}_e^2 = \frac{S_*(\boldsymbol{\varphi}, \mu, \theta)}{n-p-q-1}, \quad (16)$$

where $n - p - q - 1$ (degrees of freedom) equals the number of terms used in the sum of $S_*(\boldsymbol{\varphi}, \mu, \theta)$ minus the number of parameters' estimator.

(ii) Unconditional Maximum Likelihood Estimation

Box, Jenkins and Reinsel (2008) suggest the following unconditional log-likelihood function;

$$lnL_*(\boldsymbol{\varphi}, \mu, \theta, \sigma_e^2) = -\frac{2}{n}(2\pi\sigma_e^2) - \frac{S_*(\boldsymbol{\varphi}, \mu, \theta)}{2\sigma_e^2}, \quad (17)$$

where

$$S(\boldsymbol{\varphi}, \mu, \theta) = [E(e_t|\boldsymbol{\varphi}, \mu, \theta, \boldsymbol{X})]^2 \quad (18)$$

is the conditional sum of squares function.
where $E(e_t|\boldsymbol{\varphi}, \mu, \theta, \boldsymbol{X})$ is the conditional expectation of e_t given $\boldsymbol{\varphi}, \mu, \theta, \text{and}\boldsymbol{X}$.
The quantities, $\hat{\boldsymbol{\varphi}}, \hat{\mu}, \text{and}\hat{\theta}$ that minimize function (18) are called unconditional maximum likelihood estimators and are equivalent to the unconditional least squares estimators obtained by minimizing (19). In practice, the summation in (14) is approximated by a finite form

$$S(\boldsymbol{\varphi}, \mu, \theta) = \sum_{i=-M}^{n}[E(e_t|\boldsymbol{\varphi}, \mu, \theta, \boldsymbol{X})]^2, \quad (19)$$

where M is sufficiently large integer.
The estimator $\hat{\sigma}_e^2$ of σ_e^2 is obtained from

$$\hat{\sigma}_e^2 = \frac{S_*(\boldsymbol{\varphi}, \mu, \theta)}{n}$$

2.5.3 Model Verification

The last step in Box-Jenkins iterative procedure is model verification or model diagnosis. The conformity of white noise residual of the model fit will be judged by plotting the ACF and PACF of the residual to see whether it does not have any pattern or we perform Ljung-Box Test on the residual.

The test hypothesis:
H_0: There is no serial correlation
H_1: There is serial correlation
The test statistics of the Ljung-Box (LB)

$$LB = n(n+2)\sum_{k=1}^{m}\frac{\rho_k^2}{n-k}. \quad (20)$$

where n is the sample size, $m = lag\ length$ and p is the sample autocorrelation coefficient [Ljung and Box, 1978], (K = 1, 2, ….) LB is asymptotically a Chi-squared random variable with m-p-q degrees of freedom [22 and23] The decision: if LB is less than critical value of x^2, then we do not reject the null hypothesis. This means that a small value of Ljung-Box statistic will be in support of no serial correlation or i.e. the errors are normally distributed.

2.6 Information Criteria

There are several information criteria available to determine the order, p, of an AR process and the order, q, of MA (q) process; all of them are likelihood based. For this work, we shall consider Akaike information criterion (AIC) and Bayesian information criterion (BIC).

The idea of AIC [6] is to select the model that minimizes the negative likelihood penalised by the number of parameters (17). The AIC is specify in equation (12) below

$$AIC = -2logP(L) + 2P, \quad (21)$$

where *L* refers to the likelihood under the fitted model and *p* is the number of parameters in the model. Specifically, AIC is aimed at finding the best approximating model to the unknown true data generating process and its applications [6]

Unlike Akaike Information Criteria, BIC is derived within a Bayesian framework as an estimate of the Bayes factor for two competing models [7]. BIC is defined as

$$BIC = -2logP(L) + Plog(n), \quad (23)$$

Superficially, BIC differs from AIC only in the second term which now depends on sample size n. Models that minimize the Bayesian Information Criteria are selected. From a Bayesian perspective, BIC is designed to find the most probable model given the data (24).

2.7 Forecast Performance Measures

The methods of forecast performance based on forecast error include Mean Squared Error (MSE) [25], Root Mean Squared Error (RMSE)[25] , Mean Absolute Error (MAE)[25] and Theil's U-statistics[25] These criteria measure forecast performance.

Now we shall discuss the commonly used forecast performance measures. Here X_t is the actual value, $\hat{X}_t$ is the forecasted value and n is the size of the test set.

The mathematical definitions of performance measures are given below

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(\hat{X}_t - X_t)^2. \quad (24)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{X}_t - X_t)^2}. \quad (25)$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|\hat{X}_t - X_t|. \quad (26)$$

$$\text{Theil U- Statistic} = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{X}_t - X_t)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{X}_t)^2}\sqrt{\frac{1}{n}\sum_{t=1}^{n}{X_t}^2}}. \quad (27)$$

3.0 RESULT AND DISCUSSION

In this section, we shall use the monthly official exchange rate in Nigeria to identify and estimate ARIMA model that adequately represents the series and use some diagnostic tests to evaluate the model. The data set is from Nigeria official exchange rate for the Naira to British pound and Euro from January 2002 to December 2018. Gretl and E-views are statistical software's used for data analysis.

3.1 Stationary Test

Figures 1-2 showed that there is an upward trend in the series and the series tend to be moving with time which signifies that the series are not stationary.

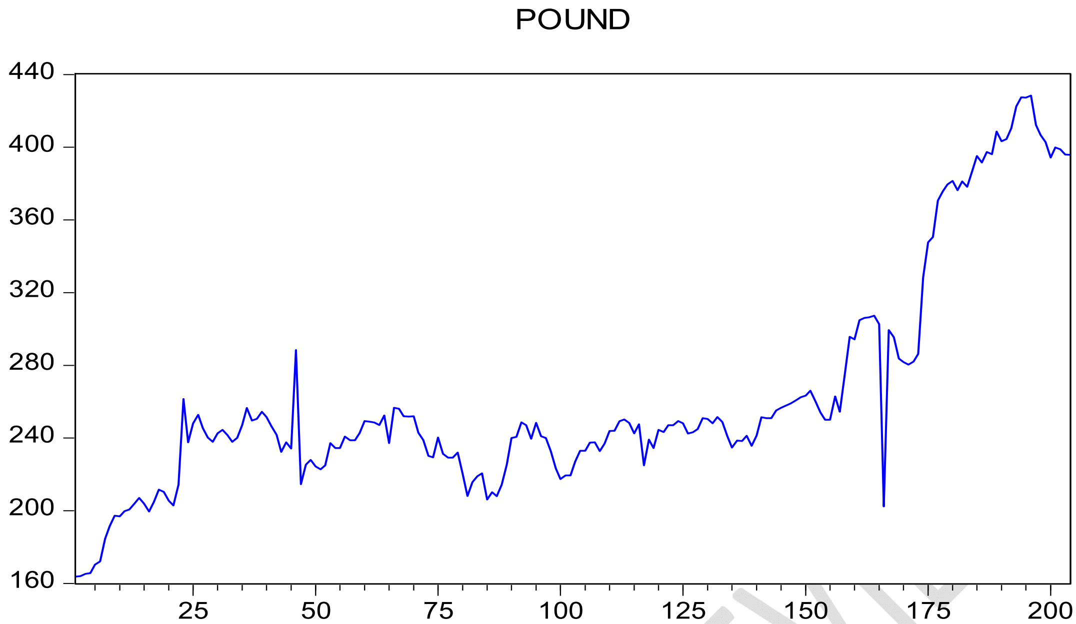

Figure 1: Time plot of Naira to Pound Exchange Rate

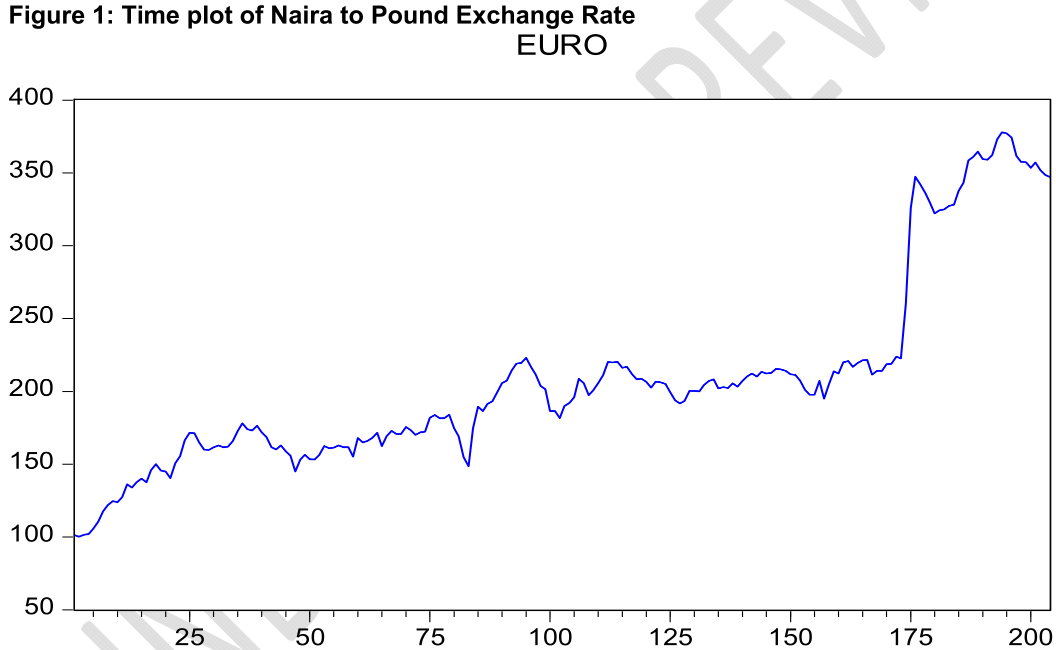

Figure 2: Time plot of Naira to Euro Exchange Rate

Evidence from Figures 3-4 suggest that the two series are stationary at first difference because the series is found to fluctuate around a common mean.

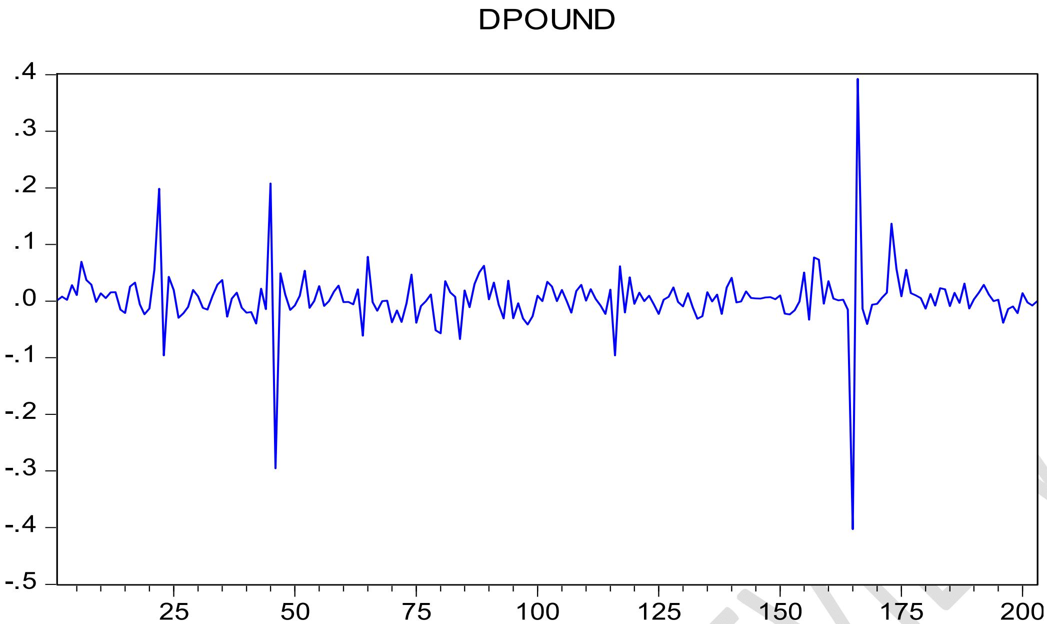

Figure 3: time plot of Naira to pound exchange rate at first difference.

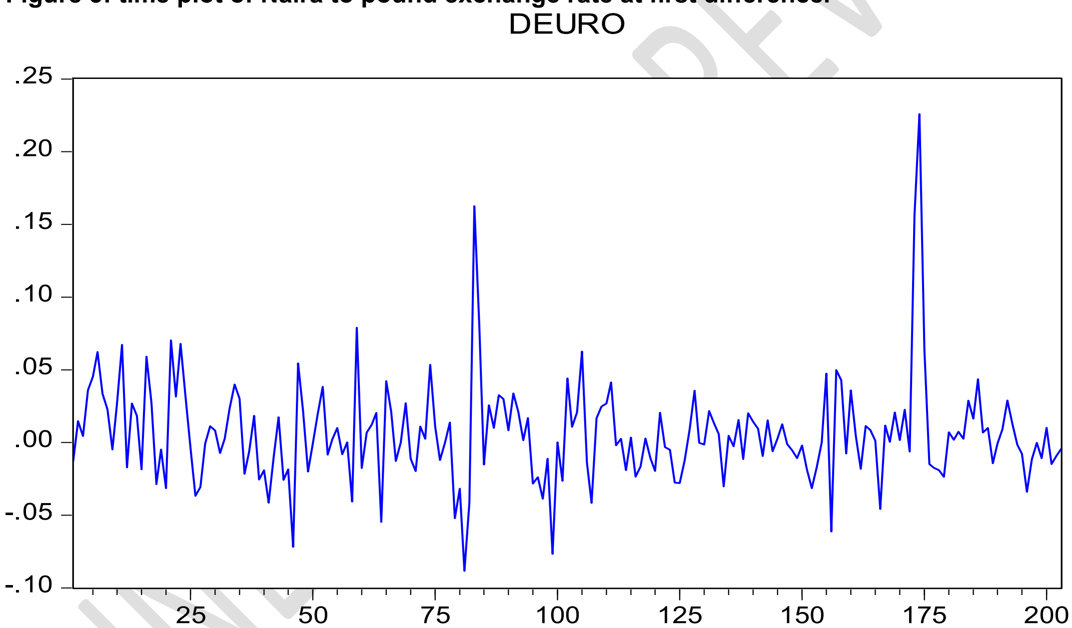

Figure 4: Time plot of Naira to Euro exchange rate at first difference.

To further test for stationarity of the data, we applied Augmented Dickey Fuller test as shown in Tables 1-2. At level form, the series were not stationary because at each assumption; intercept, intercept and trend, and no drift, each ADF test statistic were less than the corresponding value of level of significance. But at first difference the series became stationary given that the ADF statistics at various assumptions were greater than the corresponding level of significance.

Table 1: ADF test of Naira to Pound Exchange Rate

ADF Test					
Level			**First Difference**		
Intercept	**t-statistic**	**Prob.**	**Intercept**	**t-statistic**	**Prob.**
1% = -3.462737	-0.690578	0.8455	1% = -3.462737	-20.90263	0.0000
5% = -2.875680			5% = -2.875680		
10%= -2.574385			10%= -2.574385		
Intercept and trend	**t-statistic**	**Prob.**	**Intercept and trend**	**t-statistic**	**Prob.**
1% = -4.004132	-1.627799	0.7787	1% = -4.004132	-20.85049	0.0000
5% = -3.432226			5% = -3.432226		
10%= -3.139858			10%= -3.139858		
No Drift	**t-statistic**	**Prob.**	**No Drift**	**t-statistic**	**Prob.**
1% = -2.576460	1.446032 0.9632	0.9632	1% = -2.576460	-20.75971	0.0000
5% = -1.942407			5% = -1.942407		
10%= -1.615654			10%= -1.615654		

Table 2: ADF test of Naira to Euro Exchange Rate

ADF Test					
Level			**First Difference**		
Intercept	**t-statistic**	**Prob.**	**Intercept**	**t-statistic**	**Prob.**
1% = -3.462737	-0.915701	0.7818	1% = -3.462737	-10.75595	0.0000
5% = -2.875680			5% = -2.875680		
10%= -2.574385			10%= -2.574385		
Intercept and trend	**t-statistic**	**Prob.**	**Intercept and trend**	**t-statistic**	**Prob.**
1% = -4.004132	-2.275960	0.4447	1% = -4.004132	-10.73855	0.0000
5% = -3.432226			5% = -3.432226		
10%= -3.139858			10%= -3.139858		
No Drift	**t-statistic**	**Prob.**	**No Drift**	**t-statistic**	**Prob.**
1% = -2.576460	1.141854	0.9344	1% = -2.576460	-10.53163	0.0000
5% = -1.942407			5% = -1.942407		
10%= -1.615654			10%= -1.615654		

3.2 In-Sample Model Selection

Three models, ARIMA (1, 1, 0), ARIMA (0, 1, 1) and ARIMA (0, 1, 2) were considered and fitted tentatively for Naira exchange rate. The models were estimated and the parameters were found to be significant at5% level of significance as shown in Table 3 and adequate [see Table 4]. Based on minimum in-sample information criteria, ARIMA(0, 1, 1) was selected

Table 3: ARIMA models for Naira to Pound Exchange Rate.

Model	Parameter	Estimate	S.e	z-ratio	P-value	In-sample criteria	
						AIC	BIC
ARIMA(1,1,0)	φ_1	−0.354587	0.0674131	−5.260	<0.0001	1544.925	1551.430
ARIMA(0,1,1)	$\boldsymbol{\theta_1}$	**−0.364266**	**0.0618503**	**−5.889**	**<0.0001**	**1543.637**	**1550.142**
ARIMA(0,1,2)	θ_1	−0.400325	0.0742948	−5.388	<0.0001	1544.543	1554.300
	θ_2	0.0787839	0.0750629	1.050	0.0046		

Table 4: Ljung-Box Diagnostic check for Naira to Pound Exchange Rate models

Model	**Test Statistic**	**p-value**
ARIMA (0,1,1)	3.48400	0.9827
ARIMA (1,1,0)	6.02421	0.8717
ARIMA (2,1,0)	2.88988	0.9839

Also, for Naira to Euro exchange rate, three models, ARIMA (1, 1, 0), ARIMA (0, 1, 1) and ARIMA (1, 1, 2) were fitted tentatively. From Table 5, ARIMA (1, 1, 0) is the optimal model based on smallest in-sample selection criteria. Also diagnostic test of the model from Table 6 suggested that the residuals of the model are white at 5% level of significance.

Table 5:ARIMA models for Naira to Euro Exchange Rate.

Model	Parameter	Estimate	S.e	z-ratio	P-value	In-sample criteria	
						AIC	BIC
ARIMA(1,1,0)	$\boldsymbol{\varphi_1}$	**0.391906**	**0.0605581**	**6.472**	**<0.0001**	**1303.534**	**1310.039**
ARIMA(0,1,1)	θ_1	0.388483	0.0664185	5.849	<0.0001	1304.055	1310.560
ARIMA(1,1,2)	φ_1	−0.661930	1.09454	−0.604	<0.0001	1305.927	1318.936
	θ_1	0.783590	1.06604	1.016	<0.0001		
	θ_2	0.328470	0.361495	0.9086	<0.0001		

Table 6: Ljung-Box Diagnostic check for Naira to Euro Exchange Rate models

Model	**Test Statistic**	**p-value**
ARIMA (1,1,0)	7.28457	0.7756
ARIMA (0,1,1)	6.39473	0.8458
ARIMA (1,1,2)	4.48403	0.8768

3.3 Out-of-Sample Forecast Model Selection

We applied forecast performance evaluation criteria, MSE, RMSE, MAE and Theil's U coefficient for each of the models fitted for the exchange rate series. Result from Tables 7- 8 revealed that ARIMA (1, 1, 0) and ARIMA (1, 1, 2) models have smallest out-of-sample forecast performance evaluation criteria for naira to pound and naira to euro exchange rates, respectively.

Table 7: Out-of-Sample Forecast Performance Evaluation Criteria for Naira to pound Exchange rate

Model	MSE	RMSE	MAE	Theil's U coefficient
ARIMA(1, 1, 0)	**0.0030**	**0.0545**	**0.02563**	**0.8992**
ARIMA(0, 1, 1)	0.0034	0.0585	0.02863	0.9692
ARIMA(0, 1, 2)	0.0033	0.0583	0.02848	0.9992

Table 8: Out-of-Sample Forecast Performance Evaluation Criteria for Naira to Euro Exchange Rate

Model	MSE	RMSE	MAE	Theil's U coefficient
ARIMA(1, 1, 0)	0.0014	0.0377	0.0244	0.9928
ARIMA(0, 1, 1)	0.0013	0.0366	0.0243	0.9971
ARIMA(1, 1, 2)	**0.0012**	**0.0356**	**0.0234**	**0.7891**

The purpose of this work is to apply out-of-sample technique in model selection. Data from Nigeria exchange rate from January 2002 to December 2018 comprising of 204 observations were considered. The first 192 observations were used for model estimation while the remaining 12 observations were used for forecast performance evaluation. Our results revealed that ARIMA (0, 1, 1) and ARIMA (1, 1, 0) were the best in-sample models for naira to pound and naira to euro exchange rates, respectively. However, on the basis of out-of-sample forecast performance evaluation ARIMA (1,1,0) and ARIMA (1, 1, 2) were appropriate models for naira to pound and naira to euro exchange rates, respectively. This suffices from the fact that a best model in the in-sample fit may not necessary give genuine forecast. Therefore, we can say that ARIMA (1, 1, 0) and ARIMA (1, 1, 2) are the appropriate models for forecasting naira to pound and naira to euro exchange rate series, respectively.

4.0 CONCLUSION

Results from our work showed that the out-of-sample models performed better than their in-sample counterparts. The major strength of this works is that the models selected for the exchange rate is actually based on their ability to predict future values which is the essence of modelling in time series. One conspicuous weakness of this study is that it capitalizes on fitting the best model for forecasting which could be impaired as a result of over-fitting. However, this work could be improved by considering larger length of sample size for forecast evaluation and smaller sample size for model formulation to guarantee accurate forecasting of future observations.

References

[1] Burnham K P, Anderson DR. Model Selection and Multimodel Inference A Practical Information-Theoretic Approach 2002;Second edition. (Springer-Verlag New York, Inc, 175 Fifth Avenue, New York, NY 10010, USA)

[2] Ongbali SO et.al. Model Selection Process in Time Series Analysis of Production System with Random Output.*IOP Conf. Series: Materials Science and Engineering* 2018;**413,** 012057. doi:10.1088/1757-899X/413/1/012057

[3] Igboanugo AC, Ongbali SO. A Factorial Study of Bottleneck Problems in Multi- Stage Production System: Results from Textile Industry .*University of Benin Journal of Science & Technology.* 2012; (1)

[4] Shibata R. Statistical Aspects of Model Selection. Working Paper [C]. International Institute for Applied Systems Analysis. 1989; A-2361 Laxenburg, Austria.http://pure.iiasa.ac.at/3267/

[5] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining .Inference and Prediction. Springer Series in Statistics. Springer- Verlag, New York, NY, 2nd edition.

[6] Akaike HA. New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control.*1973;19(6):716-723. https://doi.org/10.1109/TAC.1974.1100705

[7] Schwarz. G. Estimating the Dimension of a Model. *Annals of Statistics* 1978; *6(2): 461-464.*https://www.jstor.org/stable/2958889

[8] Hannan E, Quinn B. The Determination of the Order of an Autoregression.*Journal of RoyalStatisticalSociety,SeriesB.*1979;41,190-195.https://www.jstor.org/stable/2985032

[9] Moffat I U, Akpan EA. Time Series Forecasting: A Tool for Out- Sample Model and evaluation. *American Journal of Scientific and Industrial Research* 2014; **5.**(6), 185-194.doi:10.5251/ajsir.2014.5.6.185.194

[10] Chatfield C. Time Series Forecasting (5th ed.).New York: Chapman and Hall CRC, Boca Raton-London-New York-Washington. 2000

[11] Ette HE. Forecasting Nigeria Naira – US dollar Exchange Rate by a Seasonal ARIMA Model. *America Journal of Scientific Research.* 2012;59.71-78

[12] Muhammed M B, Abdulmuahymin SA. Modelling the Exchange Ability of Nigerian currency (Naira) with respect to US Dollar. *International Journal of Scientific and Engineering research.*2016;**7**(7), 86-104

[13] Nwankwo SC. Autoregressive Integrated Moving Average (ARIMA) Model for Exchange Rate (Naira to dollar) *Academic Journal of Interdisciplinary Studies* 2014;**3.** (4), 182-195.Doi:10.5901/ajis.2014.v3n4p429

[14] Oladejo, M .O, Abdullahi. A Suitable Model for the Forecast of Exchange Rate in Nigeria (Naira versus US Dollar).*International Journal of Science and Research* 2015;**4.**(5), 2669-2676

[15] Osbuohien IO, Edokpa IW. Forecasting Exchange Rate Between the Nigeria Naira and the US Dollar using ARIMA model. *International Journal of Engineering Science I Invention* 2013;**2.**(4), 16-22

[16] Hipel KW, McLeod AI.**Time Series Modelling of Water Resources and Environmental Systems** 1994; Amsterdam, Elsevier.

[17] Dickey D, Fuller WA. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association.*1979; 74, 427- 431.

[18] Ellliot G, Rothenberg TJ, Stock JH. Efficient Tests for an Autoregressive Unit Root. *Econometrica.*1996; 64(4): 813-836.

[19] Phillips PCB, Perron P. Testing for Unit Roots in Time Series Regression.Biometrika.1988; 75,335-346.

[20] Fisher R A, Statistical Methods and Scientific Inference, Oliver & Boyd, Edinburgh, 1956.

[21] Box GEP, Tiao G C. Intervention Analysis with Applications to Economic and Environmental Problems. *Journal of American Statistical Association*, 1975; 7, 70- 79.

[22] Tsay RS. Analysis of Financial Time Series. 3rd ed. New York: John Wiley & Sons Inc. 2010; 97-140.

[23] Akpan EA, Lasisi KE, Adamu A. Identification of Heteroscedasticity in the Presence of Outliers in Discrete-Time Series. *Asian Research Journal of Mathematics.* 2018; 10(1): 1-20.

[24] Henry de-GraftAcquah. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship.*Journal of Development and Agricultural Economics*2010; 2(1). 001-006, http://www.academicjournals.org/JDAE

[25] Hamzacebi C. Improving artificial neural networks' performance in seasonal time series forecasting, *Information Sciences* 2008;178, 4550-4559.