Forecasting Stock Market Series

with ARIMA Model

Fatai Adewole Adebayo[1], Ramysamy Sivasamy[2] and

Dahud Kehinde Shangodoyin[3]

Abstract

Forecasting financial time series such as stock market has drawn considerable attention among applied researchers because of the vital role which stock market play on the economy of any nation. To date, autoregressive integrated moving average (ARIMA) model has remained the mostly widely used time series model for forecasting stock market series the problem of selecting the best ARIMA model for stock market prediction has attracted a huge literature in empirical analysis in view of its implication for national economics. In this paper, we consider the problem of selecting best ARIMA models for stock market prediction for Botswana and Nigeria. Using the standard model selection criteria such as AIC, BIC, HQC, RMSE and MAE we evaluate the forecasting performance of various candidate ARIMA models with a view to determining the best ARIMA model for predicting stock market in each country under investigation. The outcome of the

---

[1] University of Botswana. E-mail: aadebayofatai@yahoo.com
[2] University of Botswana. E-mail: ramasamyr@mopipi.ub.bw
[3] University of Botswana. E-mail: shangodoyink@mopipi.ub.bw

empirical analysis indicated that ARIMA (3,1,1) and ARIMA (1,1,4) models are the best forecast models for Botswana and Nigeria stock market respectively.

Keywords: ARIMA; Information Criteria; Stock market; Forecasting; Unit root tests

1   Introduction

Stock market forecasters focus on developing a successful approach for forecasting or predicting index value of stock prices. The vital idea to successful stock market prediction is achieving best result and minimizes the inaccurate forecast of the stock price. Indisputably, forecasting stock indices is very difficult because the market indices are highly fluctuating as result of increase or decrease that characterize the stock price. Fluctuations are affecting the investor's belief. Determining more effective ways of stock market index prediction is important for stock market investors in order to make more informed and accurate investment decisions. Hsien-Lun et al. [5] (2010) used ARIMA model and vector ARMA model with fuzzy time series method for forecasting. Fuzzy time series method especially heuristic model performs better forecasting ability in short-term period prediction. The ARIMA model creates small forecasting errors in longer experiment time period. In this paper, the author investigated whether the length of the interval will influence the forecasting ability of the models or not. Contreras, et al. (2003) [3] used ARIMA models to predict next day electricity prices; they have found two ARIMA models to predict hourly price in the electricity market of Spain and California. The Spanish model needs 5 hours to predict future prices as proposed to the 2 hours needs by the California model. Kumar et al. (2004) [6] used ARIMA model to forecast daily maximum surface ozone concentration in

Brunei Darussalam. They have found that ARIMA (1, 0,1) was suitable for the surface 0_3 data collected at the airport in Brunei Darussalam. Tsitsika, et al. (2007) [9] used ARIMA model to forecast pelagic fish production. The final model selected was of the form ARIMA (1, 0, 1) and ARIMA (0, 1, 1) .Azad et al. (2011) [2] used ARIMA model in forecasting Exchange rates of Bangladesh. By using Box Jenkins methodology they tried to find out best model for forecasting. They have found that ERNN (exchange rate neural network) model shows better performance than ARIMA. Merh, (2011) [8] used ANN and ARIMA model in next day stock market forecasting. They used ANN (4-4-1) and ARIMA (1, 1,1) for forecasting the future index value of sensex (BSE 30). The forecasting accuracy obtained for ARIMA (1,1,1) is better than ANN(4-4-1). Liv et al (2011) [7] used ARIMA model in forecasting incidence of hemorrhagic fever with renal syndrome in china. The goodness of fit test of the optimum ARIMA (0,3,1) model showed non – significant autocorrelation in the residuals of the model. Similarly, Datta (2011) used ARIMA model in forecasting inflation in Bangladesh Economy. He showed that ARIMA (1,0,1) model fits the inflation data of Bangladesh satisfactorily. Al-Zeaud (2011) used ARIMA model in modeling and forecasting volatility. The result showed that best ARIMA models at 95% confidence interval for banks sector is ARIMA (2,0,2) model. Uko et al. (2012) examined the relative predictive power of ARIMA, VAR and ECM model in forecasting inflation in Nigeria. The result shows that ARIMA is good predictor of inflation in Nigeria and serves as a benchmark model in inflation forecasting.

Following the introduction, the paper is organized as follows. Section two discusses the methodology to be used to analysis the data. Section three examines ARIMA model selection. Section four describes the data use, econometric methodology and the model with empirical investigation with result reported. Section five ends with conclusion.

2   Preliminary

Unit Root Test

The popular way of detecting a unit root is to examine a series, mean and covariance, if the mean is increasing over period of time. Another way would be to perform the t-test and F test to find out whether the autocorrelation function (ACFs) for the variable tend to zero as the length of the lag increases. If the ACFs tend to zero quickly then the variable is stationary; otherwise the variable is non-stationary. Another known method for the testing of non-stationary status is the Dickey-Fuller (DF) test developed by Dickey and Fuller (1979). The test examines the hypothesis that the variable being tested contains a unit root and therefore should be expressed in first difference form.

In order to run a DF test, the following equations are estimated.

$$\Delta Y_t = \delta Y_{t-1} + U_t \text{ No drift, no intercept} \tag{1}$$

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + U_t \text{ Intercept, no drift term} \tag{2}$$

$$\Delta Y_t = \beta_0 + \beta_1 t + \delta Y_{t-1} + U_t \text{ With intercept and trend} \tag{3}$$

However, the Dickey-Fuller test is only valid if the there is no autocorrelation of U_t, the white noise. When U_t is auto-correlated then chance of rejecting a correct null hypothesis is high, hence augmenting the test using lags of the dependent variable would be necessary thus the use of the Augmented Dickey-Fuller (ADF) test Said and Dickey (1984) . Just like the DF and ADF test can be specified with no drift and no trend; with trend and no drift; lastly with both trend and drift as follows.

$$\Delta Y_t = \delta Y_{t-1} + \sum \alpha_i \Delta Y_{t-1} + U_t \text{ No drift, no intercept} \tag{4}$$

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \sum \alpha_i \Delta Y_{t-1} + U_t \text{ Intercept, no drift term} \tag{5}$$

$$\Delta Y_t = \beta_0 + \beta_1 t + \delta Y_{t-1} + \sum \alpha_i \Delta Y_{t-1} + U_t \text{ With intercept and trend} \tag{6}$$

Furthermore, Phillips-perron unit root tests are used to reinforce the ADF. One advantage that the Phillips-Perron (PP) test has over the ADF test is that it is robust with respect to unspecified autocorrelation and heteroscedasticity in the disturbance process of the test regression, Brooks (2000). Therefore the PP test works well with financial time series.

The two tests specify the Null hypothesis (H_0) as that the time series has unit root, thus the time series is non-stationary against the Alternative Hypothesis (H_1) that the time series has no unit root, thus a stationary time series:

H_0: Time series has a unit root ($\delta = 1$)

H_1: Time series has no unit root ($\delta \neq 1$)

The decision rule of whether to reject or not reject the null hypothesis is as follows;

Reject the Null hypothesis if the absolute ADF/PP test statistics $<$ critical value

Do not reject the Null hypothesis if the absolute ADF/PP test statistics $>$ critical value.

3   ARIMA Model Selection

This procedure is based on fitting an autoregressive integrated moving average (ARIMA) model to the given set of data and then taking conditional expectations. The Box-Jenkins methodology refers to the set of procedures for model identification, estimation, and diagnostic checking ARIMA model with time series data. Forecasts follow directly from the form of the fitted model.

The first step in the Box-Jenkins procedure is to difference the data until it is stationary.

A pth-order autoregressive model: AR (p), which has general form

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$$

where

Y_t = Response (dependent) variable at time t.

$Y_{t-1}, Y_{t-2}, ... Y_{t-p}$ = Response variable at time lags t-1, t-2, ..., t-p, respectively.

$\phi_0, \phi_1, \phi_2, ..., \phi_p$ = Coefficients to be estimated

ε_t = Error term at time t.

A qth-order moving average model: MA (q), which has the general form

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q}$$

where

Y_t = Response (dependent) variable at time t

μ = constant mean of the process

$\theta_1, \theta_2, ... \theta_q$ = coefficient to be estimated

ε_t = Error term at time t

$\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-q}$ = Errors in previous time periods that are incorporated in the response Y_t

Autoregressive Moving Average Model: ARMA (p,q), which has the general form

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q}$$

We can use the graph of the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) to determine the model which processes can be summarized as follows:

Model	ACF	PACF
AR(p)	Dies down	Cut off after lag q
MA (q)	Cut off after lag p	Dies down
ARMA (p,q)	Dies down	Dies down

4   Main Results

For empirical illustration we use the quarterly data on stock market data that were collected from Botswana and Nigeria during 2002 to 2012. We carried out unit root test to determine the order of integration of series. We present the empirical result of our data analysis in Table 1 and 2 below.

Botswana Stock Market Series Result.

Table 1a show Statistical Result for ARIMA model

ARIMA Model (p,d,q)	ϕ_p	t-test	P-value	θ_q	t-test	P-value
(1,1,1)	-0.625031	-1.735227	0.1032	0.277418	0.601547	0.5565
(1,1,2)	-0.682142	-3.162906	0.0064	-0.930660	-13.29535	0.0000
(1,1,3)	-0.311211	-1.219017	0.2417	-0.994112	-21.16636	0.0000
(1,1,4)	-0.449009	-1.904846	0.0762	-0.273859	-0.888345	0.3884
(1,1,5)	-0.503587	-2.827363	0.0127	-0.903058	-15.28625	0.0000
(2,1,1)	0.034501	0.089367	0.9301	-0.999667	-4.038351	0.0012
(2,1,2)	0.175904	0.347203	0.7336	-0.158414	-0.262733	0.7966
(2,1,3)	0.023266	0.086253	0.9325	-0.982327	-17.75056	0.0000
s(2,1,4)	0.278042	1.094467	0.2922	0.911112	21.48394	0.0000
(2,1,5)	0.332179	1.361514	0.1949	-0.893050	-13.47365	0.0000

(3,1,1)	-0.077446	-2.129337	0.0000	-0.050985	-3.666329	0.0000
(3,1,2)	-0.005831	-0.018286	0.9857	0.020734	0.055637	0.9565
(3,1,3)	-0.036200	-0.148689	0.8841	-0.999960	-8.402277	0.0000
(3,1,4)	-0.544494	-1.994568	0.0675	0.923013	24.38043	0.0000
(3,1,5)	-0.539615	-2.109261	0.0549	-0.937170	-16.05917	0.0000

The summary of the various statistics obtained from fitting the model are tabulated in Table 2.

Table 1b summary of Portmanteau Test of stock market in Botswana

Variable	(1,1,2)	(1,1,3)	(1,1,5)	(2,1,3)	(2,1,4)	(2,1,5)	(3,1,1)	(3,1,3)	(3,1,4)	(3,1,5)
AIC	2.671	2.432	2.287	2.547	2.551	2.449	1.565	2.329	2.318	2.314
SIC	2.969	2.730	2.586	2.846	2.850	2.748	1.864	2.627	2.616	2.612
HQC	2.735	2.497	2.352	2.605	2.610	2.508	1.616	2.379	2.368	2.364
RMSE	0.892	0.976	0.919	0.961	0.914	0.928	0.754	0.976	0.923	0.905
MAE	0.723	0.748	0.686	0.759	0.641	0.727	0.618	0.778	0.793	0.718

In selecting the best ARIMA model of stock market series in Botswana we subject all the ARIMA model to Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (SBIC). The above result of table 1b show that ARIMA (3,1,1) is preferred to others since it has the lowest value of AIC as well, as compared with (p, d, q) $\neq$ (3,1,1)

FORECAST EVALUATION FOR ARIMA (3,1,1) MODEL

In the next step, the forecast of Botswana stock market series using ARIMA (3,1,1) model was conducted. The duration of forecasts is from 2008Q1 to 2013Q3. Then those forecasted value are plotted in Figure 1.
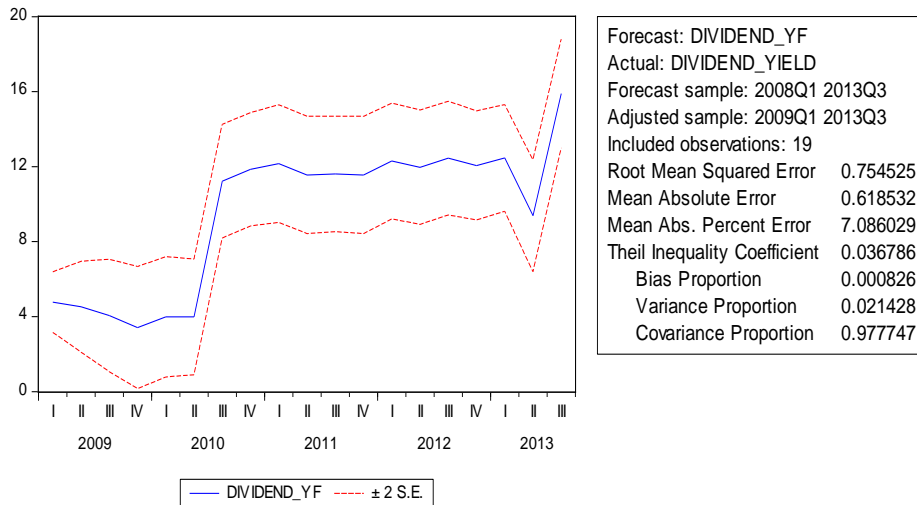
Figure 1: Forecast of Stock Market series by ARIMA (3.1.1) model.

In Figure 1 the middle line represents the forecast value of stock market. Meanwhile, the line which are above or below the forecast quarterly stock market series show the forecast with ±2 of standard errors. Some forecast evaluation measures such as root mean square error (RMSE), mean absolute error (MAE) and Theil Inequality coefficient value were used to select ARIMA (3,1,1) model. Thus, smaller RMSE, MAE and Theil Inequality coefficient are preferred.

Nigeria stock market Result

Table 2a show Statistical Result for ARIMA Models

ARIMA Model (p,d,q)	ϕ_p	t-test	p-value	θ_q	t-test	p-value
(1,1,1)	0.425419	1.180375	0.2607	-0.999371	-3.996200	0.0018
(1,1,2)	-0.133658	-0.437855	0.6693	-0.09524	-0.442656	0.6659
(1,1,3)	-0.151553	-0.497666	0.6277	-0.255456	-0.973062	0.4544
(1,1,4)	-0.099907	-0.329803	0.0000	-0.817183	-3.455509	0.0000
(1,1,5)	-0.120289	-0.380096	0.7105	-0.187350	-0.266598	0.7943
(2,1,1)	-0.147928	-0.426277	0.6781	-0.220886	-0.6336058	0.5377
(2,1,2)	-0.060734	-0.181037	0.8596	-0.030115	-0.714277	0.4899
(2,1,3)	-0.193540	-0.526842	0.6088	-0.306873	-0.828664	0.4249
(2,1,4)	-0.076540	-0.235897	0.8178	-0.823817	-3.170167	0.0089
(2,1,5)	-0.099084	-0.270503	0.7918	-0.167787	-0.202762	0.8430
(3,1,1)	-0.257760	-0.742137	0.4751	-0.274340	-0.755570	0.4674
(3,1,2)	-0.252999	-0.773623	0.4571	-0.162377	-343.4140	0.0000
(3,1,3)	0.175773	0.128446	0.9003	-0.426739	-0.297263	0.7723
(3,1,4)	-0.303875	-0.931136	0.3737	-0.870743	-3.336504	0.0075
(3,1,5)	-0.237740	-0.735242	0.4791	-0.338568	-0.487362	0.6365

In selecting the best ARIMA model of stock market series we subject all the ARIMA model to Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (BIC). The result above show that ARIMA (1,1,4) is preferred to others since it has the lowest values of AIC and BIC.

The summary of the various statistics obtained from fitting the model are tabulated in Table 2b.

ARIMA Model	(1,1,1)	(1,1,4)	(2,1,4)	(3,1,2)	(3,1,4)
AIC	2.357300	2.178277	2.281008	2.770460	2.735875
SBIC	2.654091	2.475088	2.575083	3.060181	2.608197
HQC	2.398223	2.219200	2.310240	2.785296	2.333312
RME	0.746670	0.635303	0.726514	0.703340	0.739170
MAE	0.548636	0.487253	0.543786	0.581307	0.636960

Forecast Evaluation for ARIMA (1,1,4) Model

The forecast of Nigeria stock market series using ARIMA (1,1,4) model was conducted. The duration of forecasts is from 2008Q1 to 2012Q4. The forecast are plotted in Figure 2.
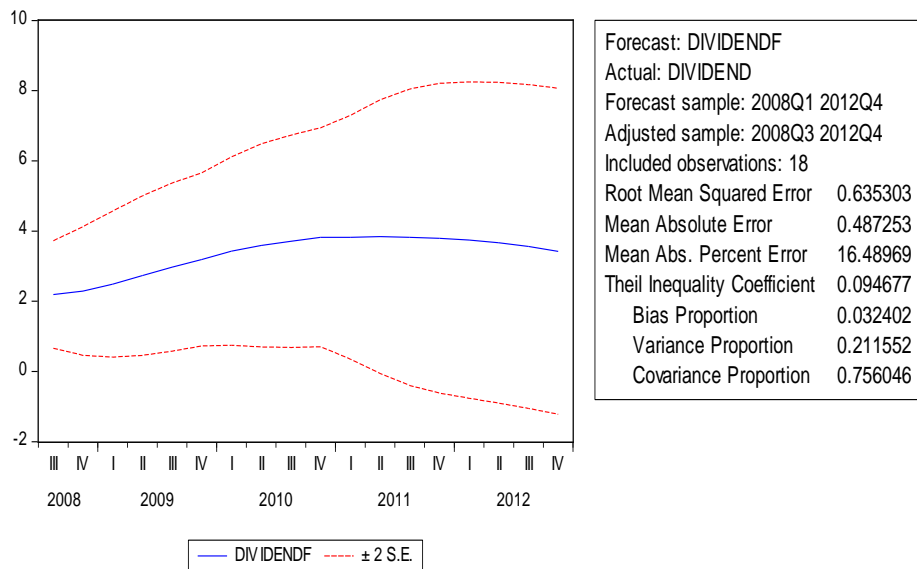

Figure 2:   Forecast of Stock Market series by ARIMA (1.1,4) model.

In Figure 1 the middle line represents the forecast value of stock market. Meanwhile, the line which are above or below the forecasted quarterly stock market series show the forecast with ±2 of standard errors. Some forecast evaluation measure such as root mean square error (RMSE), mean absolute error (MAE) and Theil Inequality coefficient value were used to select ARIMA (1,1,4) model. Thus, smaller RMSE, MAE and Theil Inequality coefficient are preferred.

5 Conclusion

This study investigates the choice of the best ARIMA model in forecasting the stock market series in Botswana and Nigeria. Having a reliable statistical forecast model for predicting stock market is a very crucial issue among investors, portfolio managers and developing economies such as Botswana and Nigeria as it helps the investors as well as other stakeholders that are connected to the business of stock market to make informed decision about investment in stock market. The empirical analysis indicated that the ARIMA (3,1,1) and (1,1,4) models are the best models for forecasting stock market series in Botswana and Nigeria.

References

[1] H.A. Al-Zeaud, Modeling & Forecasting Volatility using ARIMA model, European Journal of Economics, *Finance and Administrative Science*, **35**, (2011), 109-125.985.

[2] A.K. Azad and M. Mahsin, Forecasting Exchange Rates of Bangladesh using ANN and ARIMA models: A comparative study, *International Journal of Advanced Engineering Science & Technologies*, **10**(1), (2011), 031-036.

[3]   J. Contreras, R. Espinola, F.J. Nogales and A.J. Conejo, ARIMA models to predict Next Day Electricity Prices, IFEE *Transactions on power system*, **18**(3), (2003), 1014-1020.

[4]   K. Datta, ARIMA Forecasting of Inflation in the Bangladesh Economy, *The IUP Journal of Bank Management*, **X**(4), (2011), 7-15.

[5]   Hsien-Lun Wong, Yi-HsienTu and Chi-Chen Wang, Application of fuzzy time series models for forecasting the amount of Taiwan export, *Experts Systems with Applications*, (2010), 1456-1470.

[6]   K. Kumar, A.K. Yadav, M.P. Singh, H. Hassan and V. K. Jain, Forecasting Daily Maximum Surface Ozone, (2004).

[7] Q. Liv, X. Liu, B. Jiang and W. Yang, Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model, *Biomed Central*, (2011), 1-7.

[8]   N. Merh, V. P. Saxena and K. R. Pardasani, Next Day Stock market Forecasting: An Application of ANN & ARIMA, *The IUP Journal of Applied Science*, **17**(1), (2011), 70-84.

[9]   E.V. Tsitsika, C. D. Maravelias and J. Haralatous, Modeling & forecasting pelagic fish production using univariate and multivariate ARIMA models, *Fisheries Science*, **73**. (2007), 979-988.

[10] A.K. Uko and E. Nkoro, Inflation Forecasts with ARIMA, Vector Autoregressive and   Error Correction Models in Nigeria, *European Journal of Economics, Finance & Administrative Science,* **50**, (2012), 71-87.