

Comparative study of exponential smoothing models and Box-Jenkins ARIMA model of partitioned data of daily stock prices of the CRDB Bank in Tanzania *

K.K. Saxena^{1,†} and Juma Salehe Kamnge²

1,2. Department of Statistics, University of Dodoma,
Dodoma, Tanzania.

1. E-mail: haufiles@gmail.com

Abstract Data mining techniques and other analytical techniques have played a significant role in analysing data from different sources. Data from stock market consist of high volatility and hence it needs a special care to fit a model for forecasting future stock prices values. In the stock market, investors trade to get positive returns through buying at a lower price and selling at a higher price. However, not all investors get positive returns on their investment in stock market because of large amount of risk involved in the stock market due to vast fluctuation in the stock market prices. This study is conducted to compare and select the best model among autoregressive integrated moving average (ARIMA), single exponential smoothed model (SES), double exponential smoothed model (DES), and Damped trend linear exponential smoothed model. The modelling process was preceded by analysing the time series which revealed the presence of non-stationarity. The resultant models were found as ARIMA(1,1,2), Simple Exponential Smoothing (SES) and Double exponential smoothing (DES) whose parameters' estimates were also found to be statistically significant. Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to select the best model among all the three fitted models. The performance of the fitted model is analyzed and the market behavior for future forecast is studied. ARIMA(1,1,2) is selected as the best model for daily stock forecasting for the CRDB bank in Tanzania.

Key words Time series analysis, ARIMA model, Simple, Double and Damped trend linear exponential smoothing methods, ACF and PACF.

2020 Mathematics Subject Classification 62M10, 62M99, 62P99, 91B84.

1 Introduction

The stock market refers to the secondary market whereby investors buy or sell stocks or securities. It fails to address the dimensionality and expectancy of an investor. The general trend towards stock market prices among the investors is that it is highly risky for investment or not suitable for trade. To make a decision in the stock market, a proper decision is needed at appropriate time to sell or purchase a stock. Due to this reason, new investors, generally do not enter the stock market. An approach with

* Communicated, edited and typeset in Latex by *Lalit Mohan Upadhyaya* (Editor-in-Chief).
Received May 26, 2019 / Revised January 29, 2020 / Accepted February 17, 2020. Online First Published on June 30, 2020 at <https://www.bpasjournals.com/>.

[†]Corresponding author K.K. Saxena, E-mail: haufiles@gmail.com

adequate expertise is needed to help investors to ascertain patterns from the historic data that have feasible predictive ability in their investment decisions. The historical data has a significant role in helping the investors to get an overview about the future market behaviour. So, with this background, an attempt has been made in this paper to compare forecasting models: autoregressive integrated moving average (ARIMA), single exponential smoothing, double exponential smoothing, and Damped trend linear exponential smoothing and select the best model for forecasting daily stock prices of CRDB bank in Tanzania. The historical data of stock prices for CRDB bank are used in this study for 8 years (from January 2009 to June 2016) making a total of 2737 days to get the best forecasting model among ARIMA, SES, DES and DTLES models. It is felt that the results will contribute much to decision makers and investors since investors are highly motivated to estimate stock price trends in order for them to buy the stock and get high returns from stocks and avoid financial risks.

2 Review of literature

Exponential smoothing and ARIMA modelling are very important in economic forecasting which plays a very crucial role in decision making process in both public and private sector. These sectors base many of their decisions on future expected economic conditions or on the predictions of specific indicators of interest such as income growth, exchange rates, inflation, interest rates, unemployment and many others. But the realization of economic outcomes is a vast, dynamic and stochastic process which makes forecasting very difficult and forecast errors unavoidable. However, forecast accuracy and reliability can be improved by using appropriate models and methods such as ARIMA and VAR models.

Singh et al. [11] conducted the study on implementation of exponential smoothing for forecasting time series data. A geographical user interface was developed in Java to implement exponential smoothing technique to predict the tourist data of Himachal Pradesh based on historical data.

Latha and Nageswararao [5] conducted the study on forecasting of the time series of stock returns using ARIMA model using 30 top most companies listed as blue-chip companies. Pang et al. [10] studied the impact of market attributes on stock price prediction where traditional neural network algorithms were employed but they failed to predict the stock market exactly due to the fact of random selection of initial weights. The input was no longer a single index but multi stocked with high dimensional historical data. In their findings, they proposed the deep long-short memory neural network (LMND) to forecast the stock market. The findings showed that the accuracy of deep long short-term memory neural network model was 57.2% for the Shanghai A-shares composite index and 52.4% for individual stocks. Subashini and Karthikeyan [12] conducted the study on the time series forecast of stock price for different companies using ARIMA. Nandakumar and Lokeswari [8] conducted a study on the prediction of stock price market using the recurrent neural network that is known as long short memory by using the Artificial neural network and evaluated for the accuracy with various sizes of the data. Ogoteng [9] studied on the behaviour of stock return and the pricing volatility. This study was based on GARCH type models. The findings show that the volatility of stock returns depicted characteristics of volatility clustering, leptokurtic distribution and leverage effects overtime for all the Africa equity markets.

Saini et al. [6] conducted a study on time series forecasting of stock (ARIMA) model in forecasting the stock price using the daily stock data of State Bank of India for two years. This study showed that ARIMA (0, 1,1) was the best model found for forecasting the daily stock data of the State Bank of India. Abdalla [1] conducted a study on stock market return by exploring the relationship between macro variables such as exchange rate, money supply interest rate and stock market return. The findings of this study revealed that there was a need to focus on macro-economic variables in designing monetary and fiscal policy of the country since the capital market does respond quickly to the arrival of new information. Devi et al. [3] conducted a study on time series forecasting of the stock data for the past five years using ARIMA model with different parameters. Zhang [13] used autoregressive models to forecast regional Gross Domestic Product (GDP) in Sweden. The study applied Autoregressive Integrated Moving Average (ARIMA) model, the Vector Autoregressive (VAR) model and the First-order Autoregressive (AR(1)) model in modelling and forecasting the regional GDP. The findings showed that all the three models were valid in generating short-term forecasts. Furthermore, the analysis showed that the performance of AR(1) model was better compared to that of the ARIMA model, and the performance of VAR model in forecasting was the worst of all.

Green [4] studied on time series forecast using the Box-Jenkins methodology. In this study, the data of different companies for the stock price were obtained. By using Box-Jenkins methodology, different

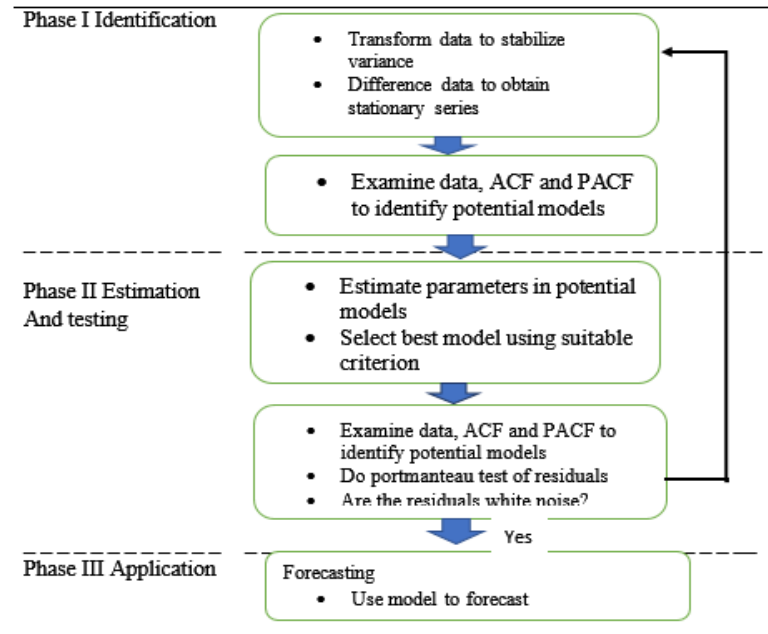
Summary of the Box Jenkins procedures

Fig. 3.1: The Box-Jenkins methodology and its steps (ARIMA).

ARIMA, AR and MA models were obtained. Marisa and Cancela [14] compared ARIMA and ANN for stock prices. The findings of this study showed that the predication accuracy between the classical ARIMA model and Artificial Neural Networks in the predictability of financial times series such as stock price index, is a complex task due to several reasons one of the major reasons being that financial times are usually very noisy. Miiler [7] conducted the study on seasonal exponential smoothing with damped trend. On his study, first he considered how the forecasting system supports overall production planning and it must be compatible with the cooperate forecast. The result from three and four parameters of winters whereby the fourth parameters was damped trend. The winter with damped trend provide most accurate forecast compared to winter without damped trend.

3 Material and methodology

This study was conducted in Dar es Salaam since the daily stock prices for the different companies are available in Dar es Salaam Stock Market Exchange. There is only one stock market located in Dar es Salaam. The population of the study consisted of six banks registered under the Dar Es Salaam stock exchange. These banks are registered at the Dar Es salaam Stock Exchange: NMB, CRDB, KCB, MKCB, MBP and DCB Bank in Tanzania. CRDB bank has the maximum 223 branches in the country and hence it was chosen purposely based on their coverage in the country. ("CRDB Bank Annual Report," [2]). This study used daily secondary data on stock prices for CRDB bank in Tanzania. Data were collected from Dar es Salaam Stock Market Exchange (DSE) on daily basis for 7 years (from January 2009 to June 2017) making a total of 2737 observations for CRDB bank.

3.1 The Box-Jenkins methodology and its steps (ARIMA)

The steps of this method are shown in Fig. 3.1 above.

3.2 Methodology for the simple exponential smoothing model

To get an exponentially smoothed time series, the value of weight associated with historical values ranging between 0 and 1, called exponential smoothing constant is computed as follows:

$$Y_t = \alpha X_t + (1 - \alpha) Y_{t-1} \text{ for } (t \geq 2) \quad (3.1)$$

where Y_t is exponentially smoothed forecasted value, X_t is the observed value of time series at time t , Y_{t-1} is exponentially smoothed forecasted value at time $t-1$, α is the smoothing constant where $0 \leq \alpha \leq 1$. Begin by setting:

$$\begin{aligned} Y_1 &= X_1, \\ Y_2 &= \alpha X_2 + (1 - \alpha) Y_1, \\ Y_3 &= \alpha X_3 + (1 - \alpha) Y_2, \\ &\dots \quad \dots \quad \dots \\ Y_t &= \alpha X_t + (1 - \alpha) Y_{t-1} \end{aligned} \quad (3.2)$$

3.3 Methodology for double exponential smoothing model

This method is generalization of simple exponential smoothing model. This method is also known as the Holt's method. This is applicable when the data has trend, to different smoothing constants for trend and slope in times series in this study two smoothing constants were used for both trend and slope in the time series - these are: α for the level and γ for the trend.

Forecasting equation

$$Y_{t+p} = L_t + pT_t \quad (3.3)$$

Level equation

$$L_t = \alpha X_t + (1 - \alpha) (L_{t-1} + T_{t-1}), \quad 0 < \alpha < 1 \quad (3.4)$$

Trend equation

$$T_t = \gamma (L_t - L_{t-1}) + (1 - \gamma) T_{t-1}, \quad 0 < \gamma < 1 \quad (3.5)$$

where Y_{t+p} is forecast for p periods into the future, L is estimate level at time t , T_t is slope (trend) estimate at time t , p are periods to be forecast into future, α is smoothing constant for level ($0 \leq \alpha \leq 1$), and γ is the smoothing constant for trend ($0 \leq \gamma \leq 1$).

3.4 Methodology for damped-trend linear exponential smoothing

Holt's linear method of forecasting usually displays a constant trend may be (decreasing or increasing) indefinitely into future, this is for reason of over forecasting so this method provides good forecast for a short period of time. Damped trend exponential is the extension of Holt's linear exponential smoothing model where in the damped trend linear exponential smoothing their is introduction of damping parameter $0 < \varphi < 1$.

$$y_{t+h|t} = L_t + (\varphi + \varphi^2 + \dots + \varphi^h) T_{t-1} \quad (3.6)$$

$$L_t = \alpha y_t + (1 - \alpha) (L_{t-1} + \varphi T_{t-1}) \quad (3.7)$$

$$T_t = \gamma (L_t - L_{t-1}) + (1 - \gamma) \varphi T_{t-1} \quad (3.8)$$

where, $y_{t+h|t}$ is the forecast and also if $\varphi = 1$ the model is identical to Holt's linear exponential smoothing model.

3.5 Selection of the best fitted model

Model selection is done by using AIC and BIC. The one having the lowest AIC and BIC is considered as the best model.

3.6 Forecasting model evaluation

After diagnostic checking, the researcher is interested to check the accuracy of the model in forecasting the future values which is normally done as the part of model validation. Then, this is achieved by analyzing the forecast errors which are obtained from one-step-ahead forecasting.

The following are the standard forecasting accuracy for the forecasting models:

The **mean absolute deviation** (MAD) or **mean absolute error** (MAE)

$$MAE = \frac{1}{T} \sum_{t=1}^T |e_t(1)| \quad (3.9)$$

The **mean squared error** (MSE)

$$MSE = \frac{1}{T} \sum_{t=1}^T [e_t(1)]^2 \quad (3.10)$$

The **average error or mean error** (ME), which is given by

$$ME = \frac{1}{T} \sum_{t=1}^T e_t(1) \quad (3.11)$$

The measures of **relative forecast error used in this study** are Mean Percent **forecast error** (MPE) and the **mean absolute percent forecast error** (MAPE), which are given as:

$$MPE = \frac{1}{T} \sum_{t=1}^T \left[\left(\frac{e_t(1)}{Y_t} \right) \times 100 \right] \quad (3.12)$$

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left[\left(\frac{|e_t(1)|}{Y_t} \right) \times 100 \right] \quad (3.13)$$

3.7 Reliability and validity

Validity and Reliability are significant concepts in research as they are applied for enhancing the accuracy of the evaluation and assessment of a research work. Reliability is concerned with the accuracy of the research process or procedure, whereas, validity is concerned with the success of the study to attain what the researcher is considering or looking for. The reliability of any study can be assessed by repetition of the study technique or procedures but its validity can be assessed by comparing its performance with some standard measures or criteria.

3.7.1 Reliability

Reliability refers to the repeatability, stability and consistency of findings when the research is carried. The outcomes of any research are considered reliable if consistent outcomes or results have been obtained in identical situations but in different circumstances. In this work, reliability was realised from data that are usually published in DSE website for the sake of customers or stock holders to foresee the future trend of their stock price so as to sell, hold or buy the stock price. These data are clean because the updating of these data is done daily for the sake of the customers to decide to buy or sell the stock price.

3.7.2 Validity

Validity is concerned with whether the study investigates the right concept or not. Under validity the research is required to show the details that whether the findings are valid or not. Validity in this research is guaranteed by the Box-Jenkins Procedure and exponential smoothing method which were applied in fitting the accurate models.

The validity of outcomes or results of the best fitted model from those fitted models ARIMA, SES, DESES and DTL were verified by the diagnostic examining of the residuals of the fit which revealed that the residuals are random. Also the validity, ensured by model validation, was also done to check how forecasted value fits on the historical data. It was found that the forecasted data match the observed data, this ensure that the model fit well the historical data.

4 Results and discussion

4.1 Model identification process for portioned time series

The time series plots for daily stock prices for CRDB (Y_t) from January 2009 to July 2016 are shown in Fig. 4.1, Fig. 4.2 and Fig. 4.3 show the first part or portion (x_t) and the last portion of the time series data for CRDB (y_t), that is, the first portion is from 2009 up to 2014 and the last part is from 2015 up to 2016. These figures were portioned because of high fluctuation that occurred and which cannot be captured.

From the time series plots for CRDB data in Fig. 4.1, Fig. 4.2 and Fig. 4.3 some patterns such as trends, seasonal variations and cyclical fluctuations are recognized by visual inspection.

Fig. 4.1, Fig. 4.2 and Fig. 4.3 show that the data were nonstationary. In Fig. 4.1 the whole stock price data for CRDB from 2009 to 2016 show that there was upward increase in price of stock as time went on, that is, from 2009 up to 2015 then it started decreasing showing that the data were not stationary. For the first portion of CRDB stock price in Fig. 4.2 there is upward increase in price as time goes on and also for the last portion in Fig. 4.3 there is a downward increase as time goes on. So these figures depict the non-stationary nature of the time series data. Therefore, since these data

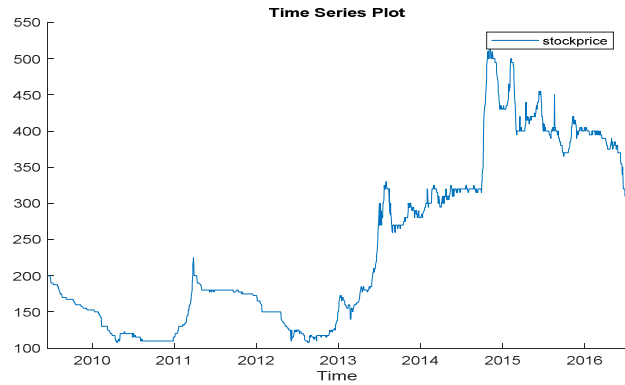


Fig. 4.1: Time series plot for the whole data of the CRDB stock price.

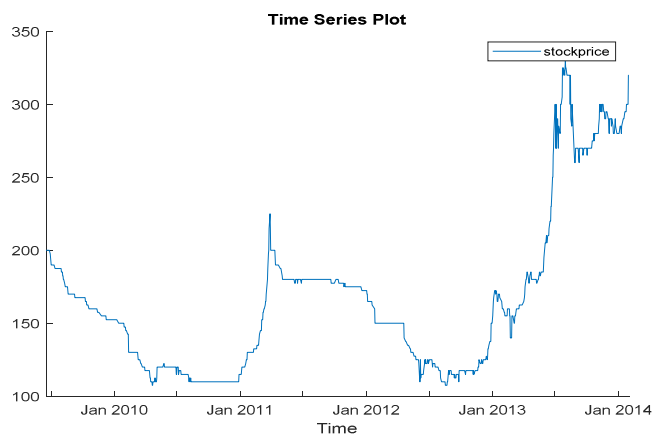


Fig. 4.2: The time series plot for the first portion of the stock price for the CRDB Bank.

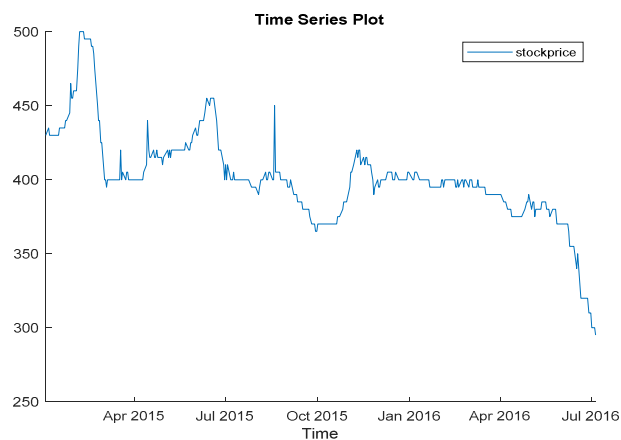


Fig. 4.3: The time series plot for the last portion of the stock price for the CRDB Bank.

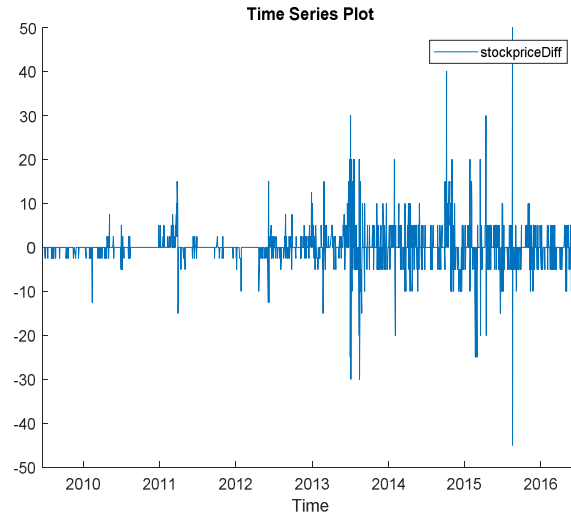


Fig. 4.4: The time series plot for the difference of the whole data of the CRDB Bank.

Table 4.1: Augmented Dickey-Fuller test for the whole stock price data for CRDB time series (W_t).

S/N	Significance Level	Lags	p -Value	Test Statistic	Critical Value
1	0.05	0	0.001	-46.0562	-1.9416
2	0.05	5	0.001	-11.8297	-1.9416
3	0.05	10	0.001	-9.4465	-1.9416
4	0.05	15	0.001	-8.5827	-1.9416
5	0.05	20	0.001	-7.9967	-1.9416

are nonstationary, the transformation is taken by differencing at order one ($d = 1$) in order to achieve stationarity of data for both the whole data and the second portion data. But for the first portion, the transformation was taken by differencing at order two ($d = 2$).

By visual inspection, the differenced time series in Fig. 4.4, Fig. 4.5 and Fig. 4.6 appear to be stationary both in the mean and the variance. However, the stationarity of a time series is confirmed by a formal test usually a unit root test. Therefore, we confirm the stationarity of the first differenced whole data of CRDB (W_t), the second differenced data of CRDB first portion (Z_t) and the first differenced last portion of CRDB stock prices (w_t) by conducting the unit root test and in this case, the Augmented Dickey-Fuller (ADF) test was applied.

Testing for stationarity: Augmented Dickey-Fuller (ADF) Test

$$y_t = c + \delta t + \phi y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_p \Delta y_{t-p} + \varepsilon_t$$

$$H_0 : \phi = 1$$

$$H_a : \phi \neq 1$$

This test was carried out using MATLAB and the results are presented in the Tables shown. This has been done at 5% level of significance but by a different number of lags.

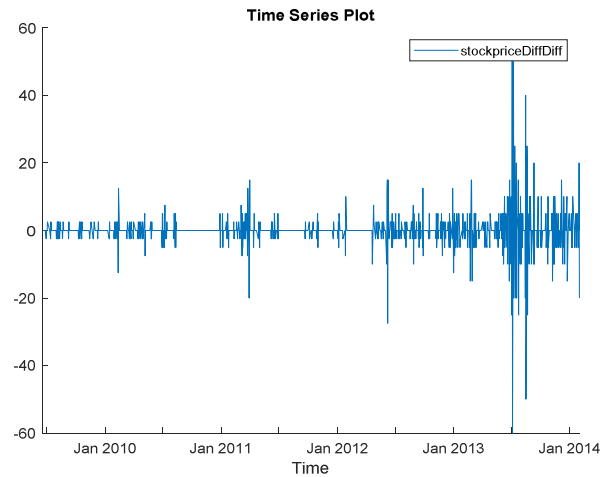


Fig. 4.5: The time series plot of the second-order difference stock price for the first portion of the CRDB Bank.

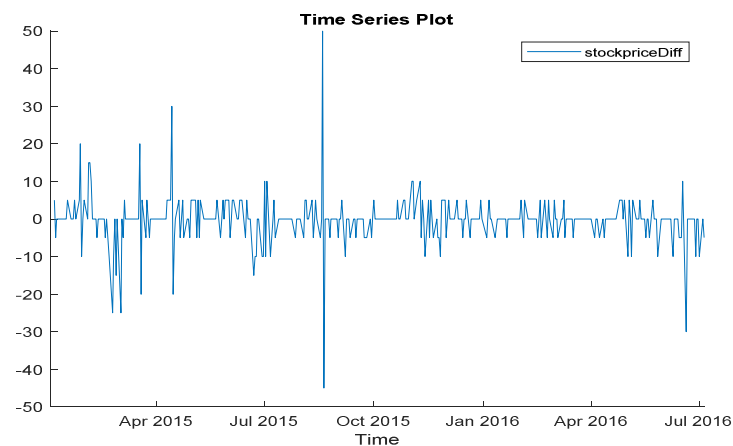


Fig. 4.6: The time series plot of the first-order difference stock price for the last portion of the CRDB Bank.

Table 4.2: Augmented Dickey-Fuller test for the stock price for CRDB time series data for the first portion (Z_t).

No	Lags	p -Value	Test Statistic	Critical Value	Significance Level	Model
1	0	0.001	-66.9683	-1.9416	0.05	AR
2	5	0.001	-23.5487	-1.9416	0.05	AR
3	10	0.001	-14.3503	-1.9416	0.05	AR
4	15	0.001	-13.3911	-1.9416	0.05	AR
5	20	0.001	-10.7767	-1.9416	0.05	AR

Table 4.3: Augmented Dickey-Fuller test for the stock price for CRDB time series data for the last portion (w_t .)

No	Lags	p -Value	Test Statistic	Critical Value	Significance Level	Model
1	0	0.001	-23.7415	-1.9413	0.05	AR
2	5	0.001	-5.7678	-1.9413	0.05	AR
3	10	0.001	-4.8176	-1.9413	0.05	AR
4	15	0.001	-4.9094	-1.9413	0.05	AR
5	20	0.001	-4.3582	-1.9413	0.05	AR

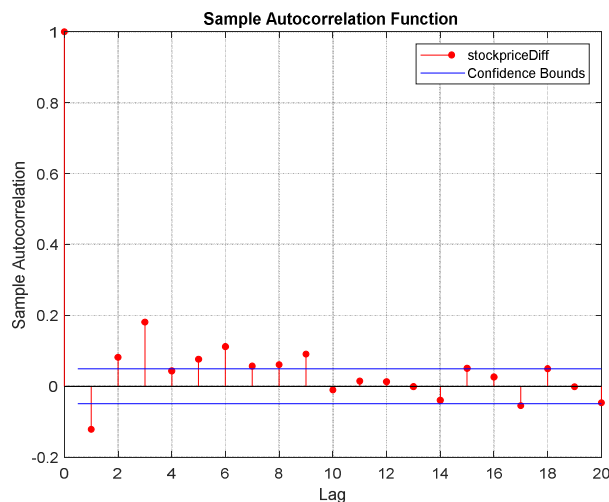


Fig. 4.7: Sample autocorrelation function for the first order difference of the whole data base of CRDB stock price.

It is revealed from Table 4.1 that the ADF test statistic is -46.0562 for lag order 0, -11.8297 for lag order 5 and a p -value of less than 0.01 indicating that the series of the first order difference of whole data of stock price of CRDB (W_t) is stationary.

In Table 4.2, the ADF test statistic is -23.5487 for lag order 5, -10.7767 for lag order 20 and a p -value of less than 0.001. This indicates that the series' first order difference of stock price for CRDB first portion is stationary (Z_t).

Also it is revealed by Table 4.3, the ADF test statistic is -5.7678 for lag order 5, -4.3582 for lag order 20 and a p -value less than 0.001. This indicates that the series' first order difference of stock price for CRDB for last portion is stationary (w_t).

Since the first difference of time series data for the stock prices of CRDB for the whole data and last portion are stationary, this indicates that the whole data and the last portion stock price for CRDB integrated in the first order, while, the first portion is stationary at second order difference which shows that the time series data for the stock prices of CRDB for the first portion integrated in the second order.

4.2 Model building process

After confirmation of stationarity in these series W_t , Z_t and w_t , we suggest some ARIMA models by looking at the ACF and PACF plots.

The model building process is done by observing the ACF and PACF of the stationary data to determine

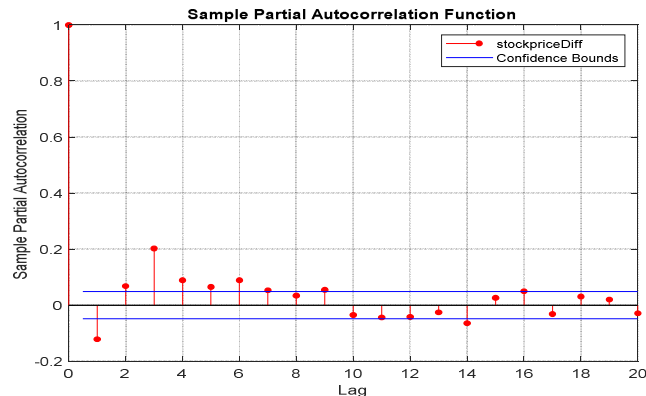


Fig. 4.8: Sample partial autocorrelation function for the first order difference of whole data base of CRDB stock prices.

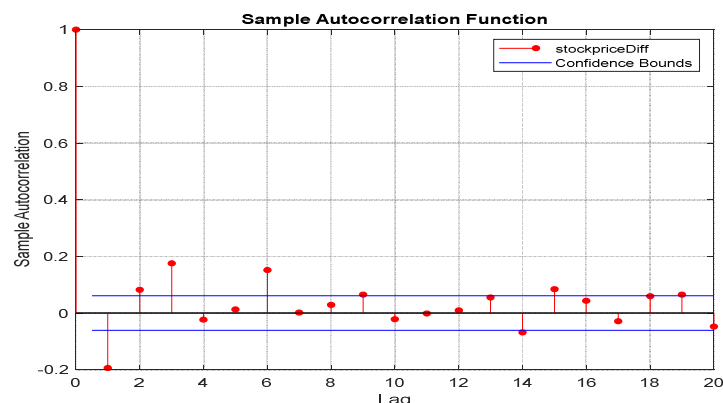


Fig. 4.9: Sample autocorrelation function for the first portion of CRDB stock prices.

the order of AR and MA for both seasonal and non-seasonal observation component. As it is revealed in the Fig. 4.7, the ACF of the whole data of CRDB stock price, for the first difference shows that the first difference is stationary since there is a quick decay of lags. Again, Fig. 4.7 represents the ACF for lags 1 to 20 of the first difference time series of the whole data for stock price of CRDB bank. The above ACF infers that the first nine lags exceed the significance limits except the fourth lag that does not exceed the significance limit but touches the significance limit and the autocorrelation tails off to zero at lag 9 but the rest of the coefficients from lag 9 to lag 20 are within the limit. The Fig. 4.8 represents the partial autocorrelation function (PACF) of first series of the whole data of CRDB bank. It also infers that the PACF from lag 1 up to lag 7 exceeds the significance limit except at lag 2 where it touches the significance limit. This infers that after lag 7, the partial autocorrelation tail off to zero. The above correlogram in Fig. 4.9 infers that the autocorrelation at lag 1 exceeds the significance limits. At lag 2, it exceeds the significance limits and at lag 3, it exceeds the significance limits and the rest all are within the significance limits except at lag 6. At lag 14 and at lag 15, it exceeds significance limits and tails off to zero after lag 3 and at lag 6, the autocorrelation is assumed to be the error that happened by chance alone.

The partial correlogram in Fig. 4.10 shows that the partial auto-correlation at lag 1 exceeds the significance limits. At lag 2 it does not exceed the significance limits and at lag 3 it exceeds the

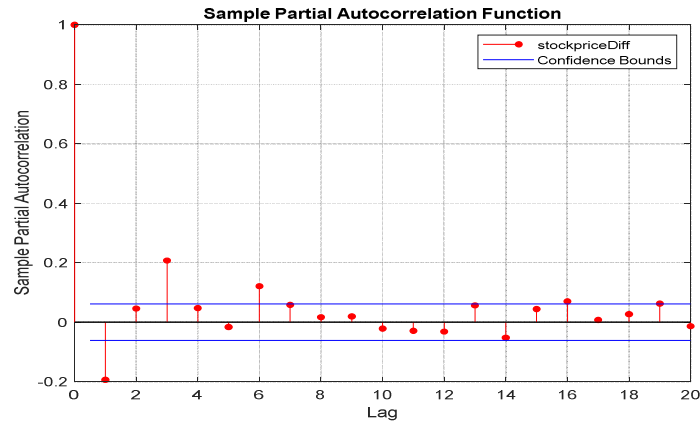


Fig. 4.10: Sample autocorrelation function for the last portion of CRDB stock price.

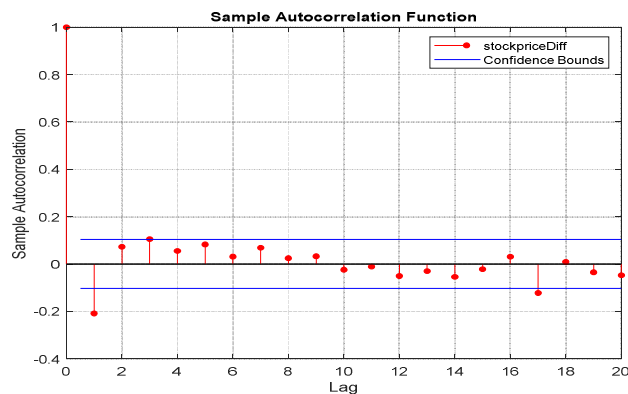


Fig. 4.11: Sample partial autocorrelation function for the last portion data of CRDB stock price.

significance limits. The rest lags are all within the significance limit except at lag 6 where it exceeds the significance limit and at lag 16 it touches the significance limit. With these partial autocorrelation at lag 6 and at lag 16, we infer that the partial autocorrelation tail off to zero after lag 3.

The correlogram in Fig. 4.11 infers that the autocorrelation at lag 1 exceeds the significance limit and tails off to zero after lag 1. Although at lag 17 it exceeds the significance limits, the rest of the coefficients between lag 1 to lag 20 are within the significance limits. Therefore, we assume that the autocorrelation at lag 17 is an error that happened by chance.

The partial correlogram in Fig. 4.12 infers that the partial autocorrelation at lag 1 exceeds the significance limits and tail off to zero after lag 1. Although at lag 5, at lag 12 and at lag 14, it exceeds the significance limits. We assume these errors to have happened by chance.

4.3 Model suggestion

After studying the ACF and PACF plots, the following models are suggested:

- I. ARIMA(1,1,1) Model
- II. ARIMA(2,1,1) Model
- III. ARIMA(2,1,2) Model
- IV. ARIMA(3,1,1) Model
- V. ARIMA(3,1,2) Model

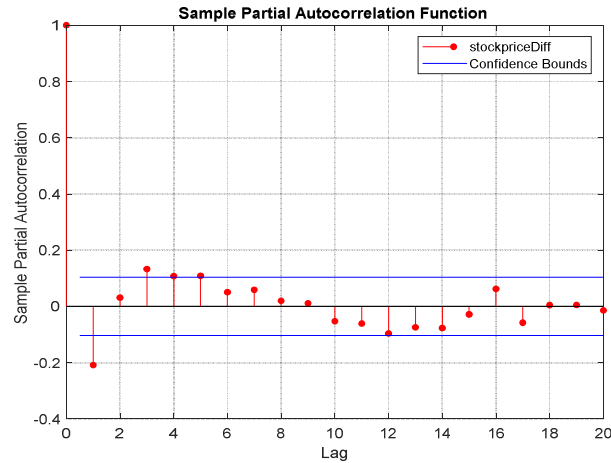


Fig. 4.12: Sample partial autocorrelation function for the last portion data of CRDB stock price.

Table 4.4: Model selection for whole data base of stock prices for CRDB.

Model	AIC	BIC
ARIMA(1,1,1) Model	9911.1136	9927.3645
ARIMA(2,1,1) Model	9830.1550	9851.8205
ARIMA(2,1,2) Model	9810.4372	9837.5191
ARIMA(3,1,1) Model	9805.3408	9832.4197
ARIMA(3,1,2) Model	9806.0163	9838.5109

Suggested models for the first portion of CRDB stock price data X_t

- I. ARIMA (1, 2,1) Model
- II. ARIMA (2, 2,1) Model
- III. ARIMA (2, 2,2) Model
- IV. ARIMA (2, 2,3) Model

Suggested models for the last portion of CRDB stock price data X_t

- I. ARIMA (2, 1,1) Model
- II. ARIMA (1, 1,2) Model
- III. ARIMA (2, 1,2) Model

To select as the best suitable model for forecasting out of the above five models, we select the one with the lowest Bayesian information criterion and Akaike information criterion values. Table 4.4 shows the summaries of output of each ARIMA model in our time series for the whole data of CRDB and it shows clearly that the lowest AIC and BIC are for the ARIMA (3, 1,1). Hence, this model can be the best model for forecasting the time series data for whole data base of stock prices for the CRDB bank. This is the best predictive model for making forecasting.

To select the best suitable model for forecasting out of the thirteen above, we select the one with the lowest Bayesian information criterion and Akaike information criterion values. Table 4.5 shows the summaries output of each of the ARIMA models in our time series for the first portion data of CRDB and the lowest AIC and BIC are for the ARIMA(2,2,3). Hence, this model can be the best model for forecasting the time series data for the first portion data of stock prices.

Table 4.5: Model selection for the first portion of stock prices for CRDB.

Model	AIC	BIC
I. ARIMA(1,2,1) Model	5742.8628	5757.7722
II. ARIMA(2,2,1) Model	5743.0122	5762.8876
III. ARIMA(2,2,2) Model	5738.4082	5763.2524
IV. ARIMA(2,2,3) Model	5713.5817	5743.3948

Table 4.6: Model selection for the last portion of stock prices for CRDB.

Model	AIC	BIC
I. ARIMA(2,1,1) Model	2441.3041	2456.9797
II. ARIMA(1,1,2) Model	2434.6641	2450.3504
III. ARIMA(2,1,2) Model	2435.1570	2454.7514

Table 4.6 clearly shows that ARIMA(1,1,2) is the best model for the last portion of stock prices compared to other selected models since it has the lowest BIC and AIC compared to the other selected models. Hence this model can be the best predictive model for making predictions for future values of our times series data.

4.3.1 Model estimation

Table 4.7 shows that all the parameters are statistically significant since the p -value for all parameters are less than 0.05. The following is an equation of autoregressive integrated moving average model of time series data for the whole stock price of CRDB bank:

$$\begin{aligned}
 (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3)(1 - L)y_t &= (1 + \theta_1 L)\varepsilon_t \\
 y_t [(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3) + (-L + \phi_1 L^2 + \phi_2 L^3 + \phi_3 L^4)] &= (1 + \theta_1 L)\varepsilon_t \\
 y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3} + y_{t-1} + \phi_1 y_{t-2} + \phi_2 y_{t-3} + \phi_3 y_{t-4} &= \varepsilon_t + \theta_1 \varepsilon_{t-1} \\
 y_t - (\phi_1 + 1)y_{t-1} - (\phi_2 - \phi_1)y_{t-2} - (\phi_3 - \phi_2)y_{t-3} + \phi_3 y_{t-4} &= \varepsilon_t + \theta_1 \varepsilon_{t-1} \\
 y_t - (\phi_1 + 1)y_{t-1} - (\phi_2 - \phi_1)y_{t-2} - (\phi_3 - \phi_2)y_{t-3} + \phi_3 y_{t-4} &= \varepsilon_t + \theta_1 \varepsilon_{t-1} \\
 y_t = 1.48490y_{t-1} - 0.32592y_{t-2} - 0.0093y_{t-3} - 0.14967y_{t-4} - 0.65008\varepsilon_{t-1}
 \end{aligned}$$

Table 4.8 shows that all the parameters are statistically significant since the P -value for all the parameters are less than 0.05. The following is an equation of autoregressive integrated moving average model of time series data for the first portion stock price of the CRDB bank:

$$(1 - \phi_1 L - \phi_2 L^2)(1 - L)^2 y_t = (1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3)\varepsilon_t$$

Table 4.7: Estimation results for the whole database of CRDB Bank.

Parameter	Value	Standard Error	T-Statistic	p-Value
AR{1}	0.48490	0.037279	13.0075	3.8411e-12
AR{2}	0.15897	0.012595	12.6216	1.393e-07
AR{3}	0.14967	0.018423	8.12400	4.5921e-112
MA{1}	-0.65008	0.037507	-17.3323	6.5758e-18
Variance	20.94040	0.219820	95.26270	0

Table 4.8: Estimation results for the first portion of the data of CRDB Bank.

Parameter	Value	Standard Error	T-Statistic
AR{1}	-0.6922	0.1007	-6.8739
AR{2}	-0.6709	0.0939	-7.1448
MA{1}	-0.495	0.115	-4.3043
MA{2}	-0.1441	0.089	-1.6191
MA{3}	-0.5598	0.0999	-5.6036

Table 4.9: Estimation results for the last portion data of the CRDB Bank.

Parameter	Value	Standard Error	T-Statistic	p-Value
Constant	0	0		
AR{1}	0.81628	0.083613	9.7626	1.6294e-22
MA{1}	-1.0735	0.087864	-12.2179	2.4943e-34
MA{2}	0.32831	0.028256	11.619	3.301e-31
Variance	37.8347	0.97918	38.639	0

$$\begin{aligned}
& y_t [(1 - \phi_1 L - \phi_2 L^2) (1 - 2L + L^2)] = (1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3) \varepsilon_t \\
& y_t [1 (1 - \phi_1 L - \phi_2 L^2) - 2L (1 - \phi_1 L - \phi_2 L^2) + L^2 (1 - \phi_1 L - \phi_2 L^2)] \\
& \quad = (1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3) \varepsilon_t \\
& y_t - (\phi_1 + 2) y_{t-1} - (\phi_2 - 2\phi_1 - 1) y_{t-2} - (\phi_1 - 2\phi_2) y_{t-3} - \phi_2 y_{t-4} \\
& \quad = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \varepsilon_t
\end{aligned}$$

$$y_t = 1.3078y_{t-1} - 0.2865y_{t-2} + 0.7135y_{t-3} - 0.6709y_{t-4} - 0.495\varepsilon_{t-1} - 0.1441\varepsilon_{t-2} - 0.5598\varepsilon_{t-3}$$

Table 4.9 shows that the parameter estimates are all statistically significant since all p -values are less than 0.05. The following is an equation of autoregressive integrated moving average model of time series data for the last portion stock price of CRDB bank:

$$\begin{aligned}
& (1 - \phi_1 L - \phi_2 L^2) (1 - L) y_t = (1 + \theta_1 L) \varepsilon_t \\
& y_t (1 - \phi_1 L - \phi_2 L^2 - L + \phi_1 L^2 + \phi_2 L^3) = \varepsilon_t + \theta_1 \varepsilon_{t-1} \\
& y_t = 1.4849y_{t-1} + (\phi_2 - \phi_1)y_{t-2} - \phi_2 y_{t-3} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \\
& y_t = 1.81628y_{t-1} - 1.88978y_{t-2} - 1.0735y_{t-3} + 0.32831\varepsilon_{t-1}
\end{aligned}$$

4.3.2 Diagnostic checking

Model validation for the whole of the CRDB data

Furthermore, the plots of residuals versus fitted value and versus observation order indicate a stable variance in the two series. In general, the residual plots indicate that the fits are certainly acceptable as indicated in Fig. 4.13. This indicates that there is high fluctuation around zero mean.

Validation is usually done to assess the precision of the fit in estimating the observed values. In the context of this study, this was done for stock prices of the whole database of CRDB time series and the results are displayed in Fig. 4.14. It is seen that the fitted stock prices series track well the series of their observed data. Also, it looks like the fitted values smooth out the highs and lows in the observed values.

In the Fig. 4.15 the plot of residual histogram suggests that the residuals are approximately normally distributed. However, there is some indication of an excess of large residuals. Also the Fig. 4.16 of the residual QQ plot suggests that the data does not fit well with the normal distribution.

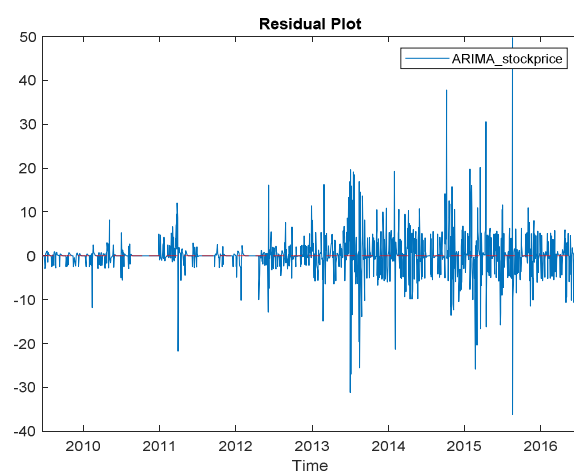


Fig. 4.13: The residual plot.

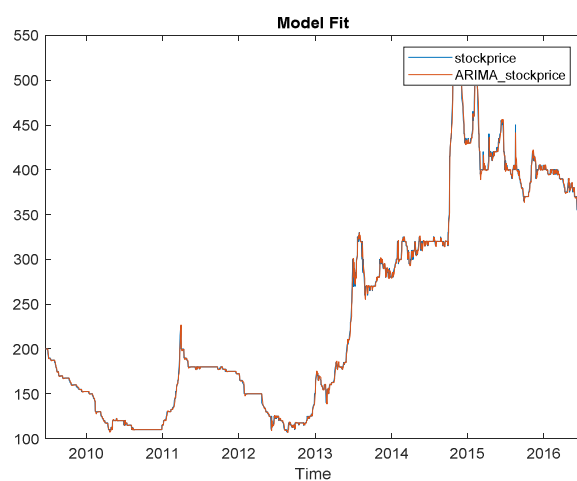


Fig. 4.14: The model fit for the whole database of CRDB.

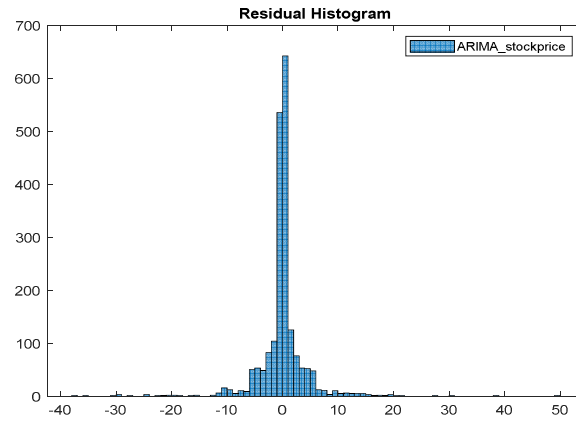


Fig. 4.15: The histogram of residuals.

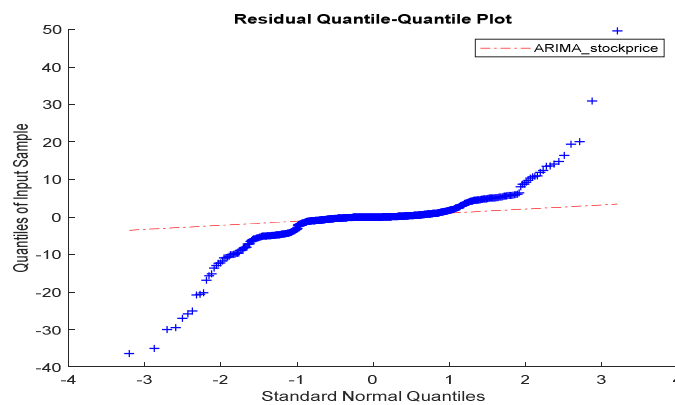


Fig. 4.16: The QQ plot of residuals.

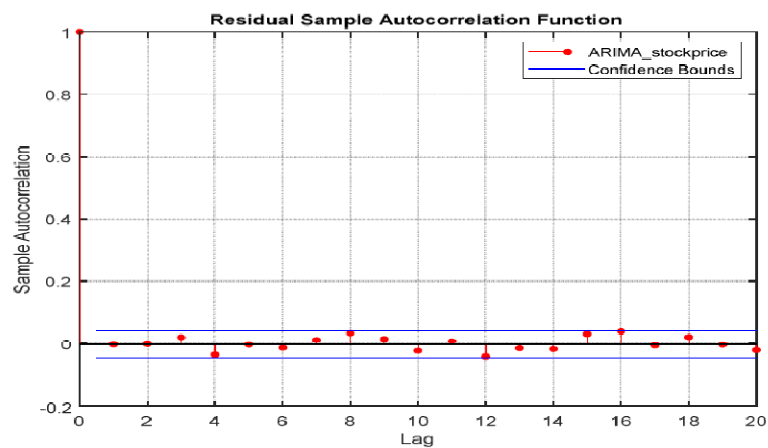


Fig. 4.17: The ACF of residuals.

Table 4.10: Box-Ljung test.

Test	X-squared	Deg. of free- dom	p-value
Box-Ljung test	20.095	20	0.452

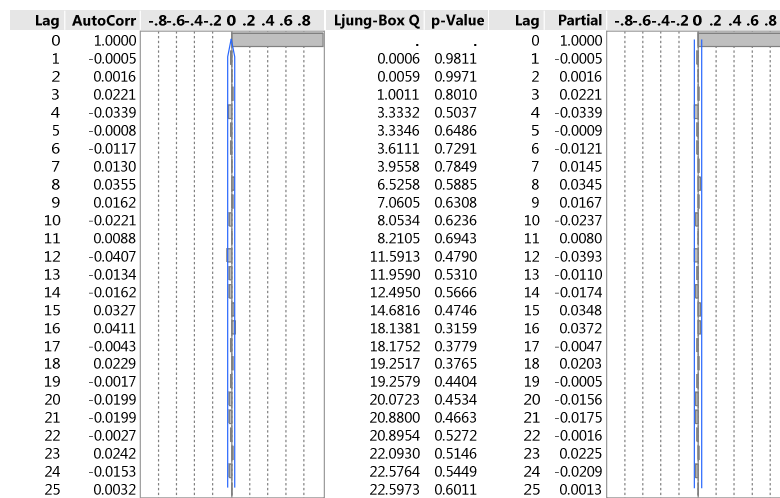


Fig. 4.18: The plots of residuals.

It clearly observed from the ACF plot Fig. 4.17 that none of the autocorrelation between lag 1 and lag 20 exceeds the significance limits, i.e., all the ACF values are well within the significance bounds. This infers that the errors are purely random. The Ljung-Box test below has been done to test whether the errors are correlated or otherwise.

Ljung-Box Q-test

$$H_0 : \rho_1 = \rho_2 = \dots \rho_m = 0$$

$$H_a : \rho_j \neq 0, j \in \{1, \dots, m\}$$

The values in the Table 4.10 and Fig. 4.18 suggest us to accept the null hypothesis that all of the autocorrelation functions in lag 1 to lag 20 are zero. In other words, we can conclude that there is no evidence for non-zero autocorrelations in forecast errors at lags 1 to 20 in our fitted model.

Diagnostic checking of the first portion of the stock prices for CRDB.

In Fig. 4.19 the plot of residual histogram suggests that the residuals are approximately normally distributed. However, there is some indication of an excess of large residuals. Also, as indicated in the Fig. 4.20 the residual QQ plot suggests that the data does not fit well into the normal distribution.

To further investigate whether there are any correlations between the successive forecast errors, the plot of autocorrelation function (ACF) and partial autocorrelation function (PACF) of forecast errors are shown in the Fig. 4.21 and Fig. 4.22 represent the ACF and the PACF of forecast errors respectively. It is clearly observed from the ACF plot in Fig. 4.21 that none of autocorrelation from lag 1 up to lag 20 exceed the significance limits, i.e., all the ACF values are well within the significance bounds.

Similarly, the PACF plot of Fig. 4.22 shows that all partial autocorrelation functions (PACF) are within the bound. This means that ACF and PACF concludes that there is no non-zero autocorrelation in forecast residuals at lag 1 to lag 20 in the fitted model ARIMA(2,2,3). This suggests that the errors are purely random.

The Table 4.11 shows that the errors are purely random since the p -value is greater than 0.05.

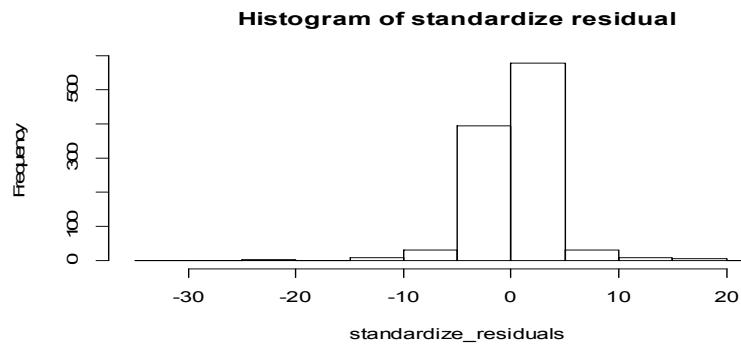


Fig. 4.19: The histogram of residuals.

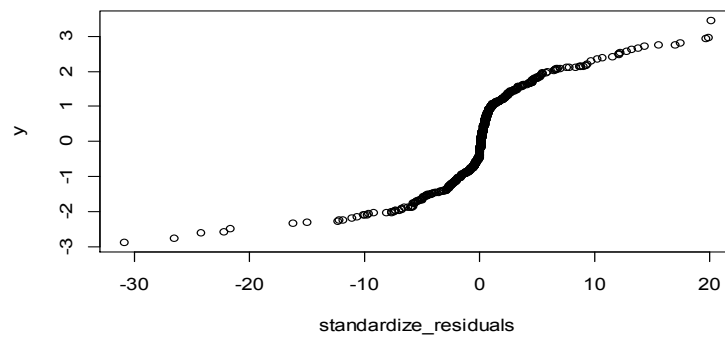


Fig. 4.20: The QQ plot of residuals.

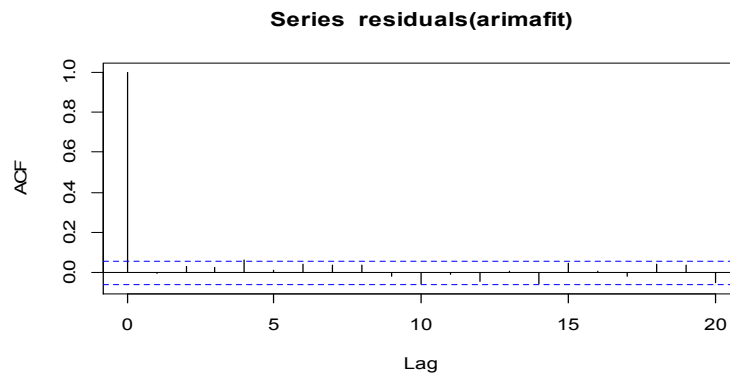


Fig. 4.21: The ACF of residuals.

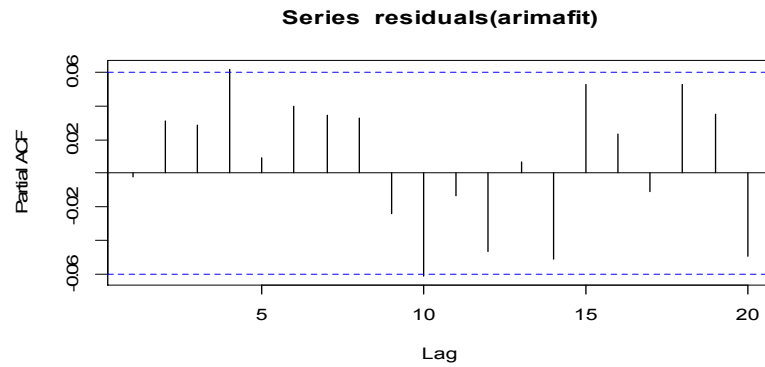


Fig. 4.22: The PACF of residuals.

Table 4.11: Box-Ljung test.

Test	X -squared	Deg. of free- dom	p -value
Box-Ljung test	27.368	19	0.09638

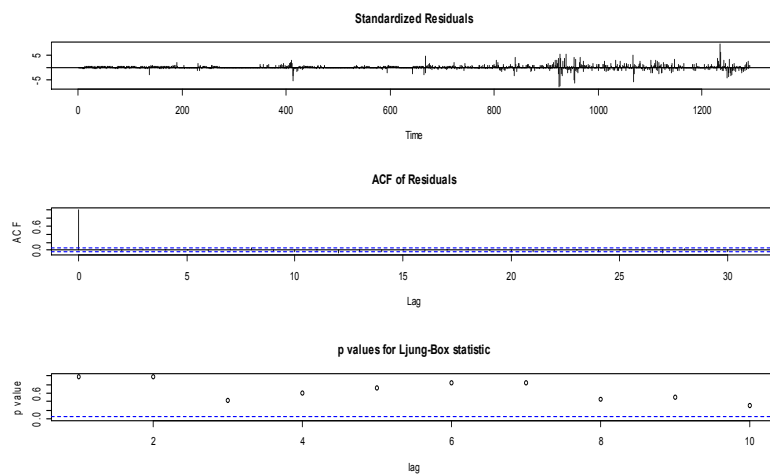


Fig. 4.23: The plots of residuals.

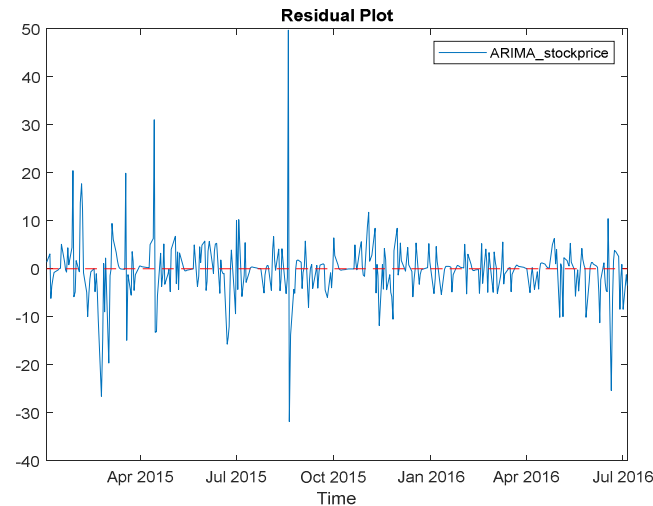


Fig. 4.24: The residual plot.

The Fig. 4.23 shows the plot of p -values for Ljung statistic and the ACF and the PACF of residuals which depicts that all the values are statistically significant as all the p -values are above 0.05. This clearly suggests us to accept the null hypothesis that all the autocorrelation functions in lag 1 to lag 20 are purely random. In other words, there is no evidence for non-zero autocorrelations in the forecast errors at lag 1 to lag 20 in our fitted model.

4.3.3 Model validation for the first portion data of CRDB

Furthermore, the plots of residuals versus fitted value and versus observation order indicate a stable variance in the two series. In general, the residual plots indicate that the fits are certainly acceptable. Fig. 4.24 indicates that there is high fluctuation around the zero mean.

Validation is usually done to assess the precision of the fit in estimating the observed values. This is done for stock prices of the whole data of CRDB time series and the results are displayed in Fig. 4.25. It is seen that the fitted stock prices series track well the series of their observed data. Also, it looks like the fitted values smooth out the highs and lows in the observed values.

As is revealed in Fig. 4.26 the plot of residual histogram suggests that the residuals are approximately normally distributed. However, there is some indication of an excess of large residuals. Also from the Fig. 4.27 of the residual QQ plot it is seen that some data did not fit well in the line. This suggests that the data does not fit well into the normal distribution.

It is clearly observed in the ACF plot of Fig. 4.28 above that none of the autocorrelation between lag 1 up to lag 20 exceed the significance limits, i.e., all the ACF values are well within the significance bounds. This infers that the errors are purely random. Also the Ljung-Box test below has been done to test if the error are correlated:

Ljung-Box Q-Test

$$H_0 : \rho_1 = \rho_2 = \dots \rho_m = 0$$

$$H_a : \rho_j \neq 0, j \in \{1, \dots, m\}$$

The statistics in Table 4.12 and Fig. 4.29 show that the large p -value in both the tests suggests us to accept the null hypothesis that all of the autocorrelation functions in lag 1 to lag 20 are zero. In other words, we can conclude that there is no evidence for non-zero autocorrelations in forecast errors at lags 1 to 20 in our fitted model.

4.4 Exponential smoothing model for the whole data

4.4.1 Simple exponential smoothing

The Table 4.13 shows that the level smoothing weight is statistically significant since the p -value is less than 0.05.

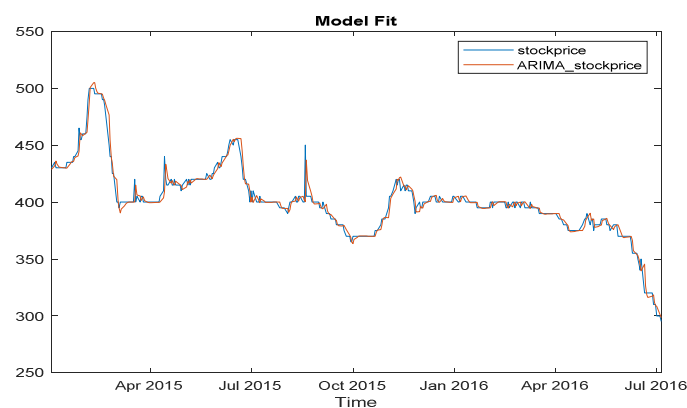


Fig. 4.25: The model fit for the whole data of CRDB.

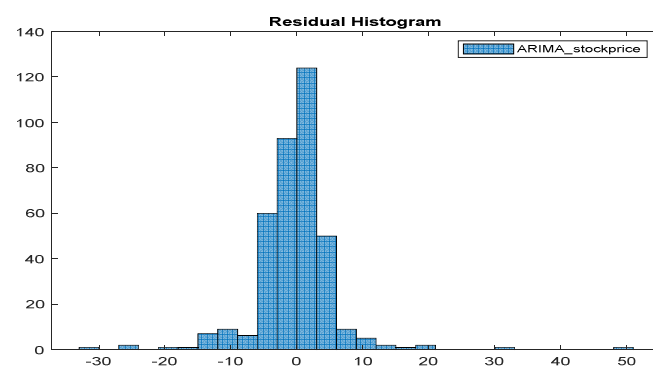


Fig. 4.26: The histogram of residuals.

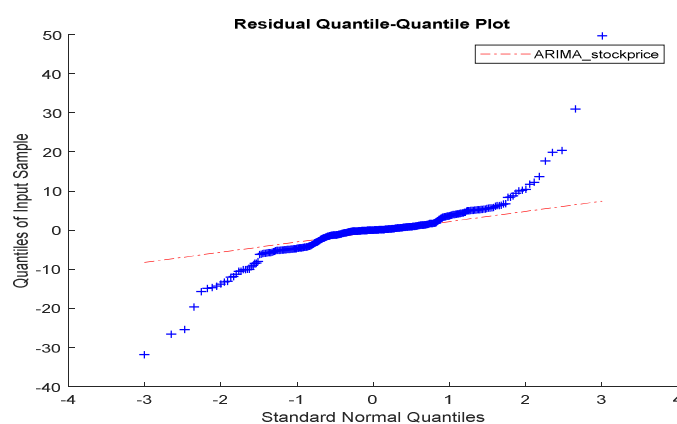


Fig. 4.27: The QQ plot of residuals.

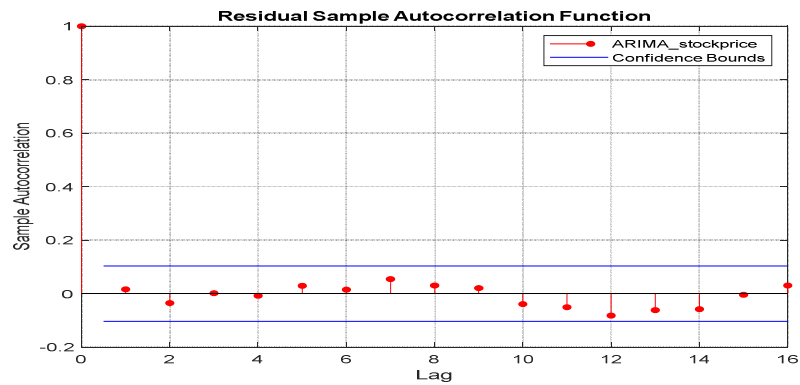


Fig. 4.28: The ACF of residuals.

Table 4.12: Box-Ljung test.

Test	X-squared	Deg. freedom	p-value	Critical Value
Box-Ljung test	20.095	20	0.452	31.4104

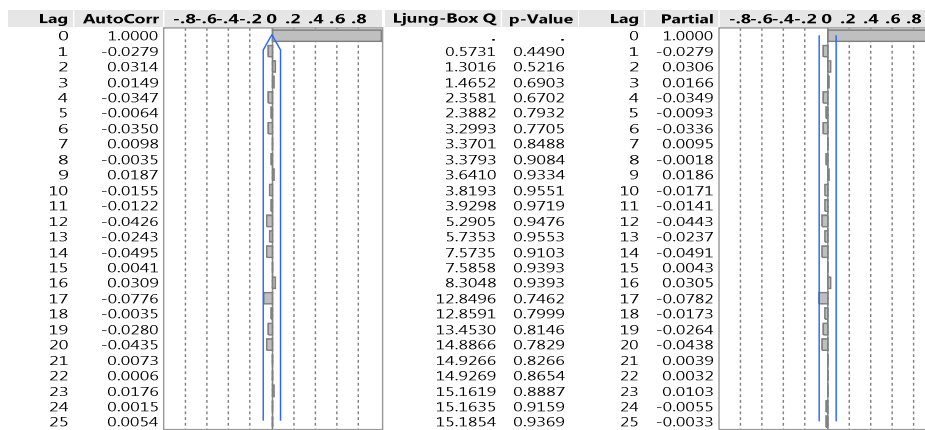


Fig. 4.29: The residual plot.

Table 4.13: Simple exponential smoothing.

Term	Estimate	Std Error	t Ratio	Prob> t
Level Smoothing Weight	0.899561	0.021525	41.79	<.0001

Table 4.14: Double (brown) exponential smoothing.

Term	Estimate	Std Error	t Ratio	Prob> t
Level Smoothing Weight	0.419613	0.010979	38.22	<.0001

Table 4.15: Damped-trend linear exponential smoothing.

Term	Estimate	Std Error	t Ratio	Prob > t
Level Smoothing Weight	1	0.023017	43.45	<.0001
Trend Smoothing Weight	0	2.24E-07	0	1
Damping Smoothing Weight	1	0.000857	1166.8	<.0001

4.4.2 Double (brown) exponential smoothing

The Table 4.14 shows that the level smoothing weight is statistically significant since the p -value is less than 0.05.

4.4.3 Damped-trend linear exponential smoothing

The Table 4.15 shows that the all the smoothing weights are statistically significant ($p < 0.0001$).

4.4.4 Exponential smoothing model for the last portion data

The Table 4.16 shows that the level smoothing weight is statistically significant since the p -value is less than 0.05.

4.4.5 Double (brown) exponential smoothing

The Table 4.17 shows that the level smoothing weight is statistically significant since the p -value is less than 0.05.

4.4.6 Damped-trend linear exponential smoothing

The Table 4.18 shows that the all smoothing weights are statistically significant ($p < 0.0001$).

4.5 Selection of the best fitted model of CRDB by using the model selection criterion

The Table 4.19 shows AIC and BIC of the whole data for stock price for CRDB and the last portion data for CRDB stock price. From the, table it is clearly observed that the last portion data have the smallest AIC and BIC for all models selected compared to the whole data. This is because of the sudden change observed from the whole data of stock price of CRDB that cannot be captured well. As Table 4.19 shows that ARIMA(1,1,2) model is the best model compared to the other selected models since it contains smallest AIC and BIC as compared to the other selected models for both the whole

Table 4.16: Simple exponential smoothing.

Term	Estimate	Std Error	t Ratio	Prob> t
Level Smoothing Weight	0.399958	0.021573	18.54	<.0001

Table 4.17: Double (brown) exponential smoothing.

Term	Estimate	Std Er- ror	t Ra- tio	Prob> t
Level Smoothing Weight	0.40827605	0.0170578	23.93	<.0001

Table 4.18: Damped-trend linear exponential smoothing.

Term	Estimate	Std Er- ror	t Ra- tio	Prob> t
Level Smoothing Weight	0.724627	0.04952	14.6	<.0001
Trend Smoothing Weight	0.15274	0.057586	2.65	0.0082
Damping Smoothing Weight	0.821431	0.059341	13.8	<.0001

Table 4.19: The model criteria selection for the different models.

	For the whole data			For the last por- tion data	
Statistics	AIC	BIC	Statistics	AIC	BIC
DTLES	9932.752	9949.005	DTL	2460.871	2472.645
ARIMA(3,1,1)	9798.715	9820.385	ARIMA(1,1,2)	2426.682	2438.455
DES	9886.371	9891.788	DES	2440.380	2444.302
SES	9908.009	9913.427	SES	2443.103	2447.028

Table 4.20: One-step-ahead forecast of stock prices for the CRDB Bank by the last portion data of stock prices of the CRDB (Continued in the Table 4.21).

Time period	Observed value	Forecast	Forecast error	Absolute error	Squared error	Absolute (%) error
8-Jul-16	300	294.268915	5.73108518	5.73108518	32.84533734	1.910361727
11-Jul-16	295	292.510796	2.48920443	2.48920443	6.196138694	0.843798112
12-Jul-16	290	291.059507	-1.05950745	1.05950745	1.122556037	0.365347397
13-Jul-16	285	289.861502	-4.86150159	4.86150159	23.63419771	1.705790032
14-Jul-16	290	288.872575	1.12742545	1.12742545	1.271088145	0.388767397
15-Jul-16	290	288.056237	1.94376260	1.9437626	3.778213045	0.670262966
18-Jul-16	290	287.382369	2.61763065	2.61763065	6.85199022	0.902631259
19-Jul-16	290	286.826106	3.17389365	3.17389365	10.0736009	1.094446086
20-Jul-16	290	286.366924	3.63307631	3.63307631	13.19924347	1.252784934
21-Jul-16	290	285.987879	4.01212133	4.01212133	16.09711757	1.383490114
22-Jul-16	300	285.674985	14.32501451	14.32501451	205.2060407	4.775004837
25-Jul-16	300	285.416699	14.58330083	14.58330083	212.6726631	4.861100277

data of stock price for CRDB and the last portion data of stock price for CRDB.

4.6 Forecasting with the best fitted model for the CRDB stock prices

4.6.1 One-step-ahead forecast of stock prices for the CRDB Bank

We have explained that after diagnostic checking, the fit can be used to forecast the future values of the variable of interest if it is adequate. However, before the fit is used to forecast, we need to assess the accuracy of its forecasts.

This is done by the one-step-ahead forecasting. Therefore, the fitted models ARIMA (1, 1,2) is used to predict the stock price of CRDB bank for the period of 23 days. That is from 8th July 2016 to 10th August 2016.

4.6.2 Evaluation of forecast accuracy

To assess the model forecasting capability, the standard measures of forecast accuracy are obtained as shown in the Table 4.20 and Table 4.21. The values of these measures are obtained using the formulae given from (3.9) to (3.12) and the forecast errors are presented in Table 4.22.

4.6.3 Analysis of forecast errors

The analysis of forecast errors is really important because this view is used in assessing the accuracy of future forecasts of the fitted model. To evaluate the model forecasting capability, we consider the

Table 4.21: One-step-ahead forecast of stock prices for the CRDB Bank by the last portion data of stock prices of the CRDB (Continued from the Table 4.20).

Time period	Observed value	Forecast	Forecast error	Absolute error	Squared error	Absolute (%) error
26-Jul-16	295	285.20349	9.79651040	9.796510400	95.97161602	3.320850983
27-Jul-16	295	285.02749	9.97251014	9.972510140	99.45095849	3.380511912
28-Jul-16	295	284.882206	10.11779398	10.11779398	102.369755	3.429760671
29-Jul-16	295	284.762277	10.23772255	10.23772255	104.810963	3.470414424
1-Aug-16	295	284.663279	10.33672089	10.33672089	106.8477988	3.503973183
2-Aug-16	295	284.581558	10.41844181	10.418441810	108.5439297	3.53167519
3-Aug-16	290	284.514099	5.485900600	5.485900600	30.09510539	1.891689862
4-Aug-16	290	284.458414	5.541586330	5.541586330	30.70917905	1.910891838
5-Aug-16	295	284.412446	10.58755365	10.58755365	112.0962923	3.589001237
9-Aug-16	295	284.374501	10.62549864	10.62549864	112.9012213	3.601863946
10-Aug-16	295	284.343179	10.65682138	10.65682138	113.5678419	3.612481824
Total			151.4925663	163.3345844	1550.312848	55.3969002

Table 4.22: Some measures of forecast accuracy.

Statistics/Variable	Stock price of NMB
Mean error (ME)	6.059702651
Mean absolute deviation (MAD)	6.533383374
Mean squared error (MSE)	62.01251392
Mean absolute percentage error (MAPE)	2.215876008

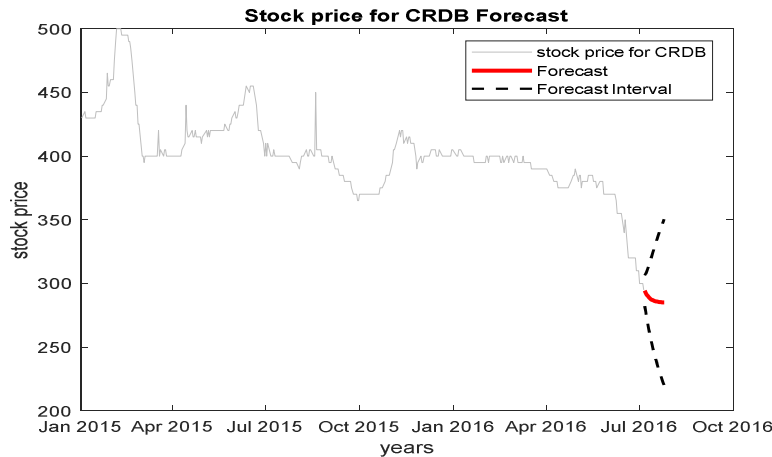


Fig. 4.30: The forecasting for the CRDB stock prices.

standard measures of forecast accuracy in Table 4.22. These measures of forecast accuracy are taken into consideration as part of validation of the fitted model. The average forecast error values for the CRDB stock series are close to zero, indicating that the forecasts produced by the fits ARIMA (1,1,2) for stock prices of CRDB are unbiased.

Also, it is noted that the fit for the stock price of CRDB with ME value 6.059702651 produces more bias. The positive sign indicates that the model is forecasting too low on average since we have almost all errors for the twenty three time periods in the stock price of CRDB (see, Owusu [15]).

Both MAD and MSE indicate that there is low variability in the forecast errors with the fitted model for the CRDB stock price. The relative (or percent) forecast error (MAPE) value for CRDB stock price is 2.215876008%. This implies that forecasting by ARIMA model is more accurate.

4.7 General observation on the forecasts

Fig. 4.25 shows that the agreement between the observed and the predicted values is good since the predicted values and the observed values are unbiased, this shows that the best fitted model is good in generating point forecasts as presented in the Fig. 4.30. The model captures well the trends of the data and it predicts well the patterns of the series. In this study the point forecasts are used even though the interval and the density forecasts contain more comprehensive and important information. They are not often applied by the business planners in forecasting. The point forecast is the most usually practiced type of forecast by the policymakers and the business managers because it is simple to calculate and understand (see, Awokuse and Ilvento, [16]).

4.8 Out-of-Sample Forecast of Daily Stock Price of CRDB bank in Tanzania

After evaluation forecast accuracy, the ARIMA (1,1,2) model is used to generate the out-of-sample forecasts of daily stock prices of the CRDB Bank for the period starting from 10th Aug. 2016 up to 25th Aug. 2016. The forecasts along with their corresponding standard error of predicted stock price of the CRDB Bank and also along with the upper and the lower confidence limit at 95 per cent confidence level are presented in Table 4.23.

Table 4.23: Out of sample forecast of the stock price for the CRDB Bank in Tanzania.

Date	Predicted stock price	Standard error of predicted	Upper control limit	Lower control limit
11-Aug-16	284.3173224	37.940519084	360.58567509	208.00628223
12-Aug-16	284.29597866	38.924029743	362.45195065	206.10476913
15-Aug-16	284.27835989	39.885217983	364.2800003	204.24763167
16-Aug-16	284.26381598	40.825334009	366.07156028	202.43206036
17-Aug-16	284.25181032	41.745537472	367.82829527	200.65550454
18-Aug-16	284.24189991	42.64690373	369.5517941	198.91564407
19-Aug-16	284.23371908	43.530429991	371.2435678	197.21036419
22-Aug-16	284.226966	44.397041217	372.9050495	195.53773345
23-Aug-16	284.22139148	45.247595733	374.53759553	193.89598415
24-Aug-16	284.21678984	46.082890503	376.14248748	192.28349508
25-Aug-16	284.21299128	46.90366605	360.58567509	208.00628223

Acknowledgments The authors express their most sincere thanks to the Editor-in-Chief of this Journal for suggesting many corrections and for pointing out many errors in the original version of this manuscript which have led to a much improved final version of this paper.

References

- [1] Abdalla, Kassim Ali (2014). An empirical analysis exploring relationship between macroeconomic variables and stock market return: A study of Dar es Salaam stock exchange (DSE), A Dissertation submitted in partial fulfillment of the requirements for the award of the Degree of Master of Science in Accountancy and Finance of Mzumbe University. <http://hdl.handle.net/11192/887>
- [2] CRDB Bank Annual Report. (2018). <https://crdbbank.co.tz/wp-content/uploads/2019/05>
- [3] Devi, B.U., Sundar, D. and Alli, P. (2013). An effective time series analysis for stock trend prediction using ARIMA model for Nifty Midcap-50, *International Journal of Data Mining & Knowledge Management Process*, 3(1), 65–78. <https://doi.org/10.5121/ijdkp.2013.3106>
- [4] Green, S. (2011). Time series analysis of stock prices using the Box-Jenkins Approach, *Electronic Theses & Dissertations*, 156. Retrieved from <files/206/Green-2011-TimeSeriesAnalysisofStockPricesUsingtheBox-JenkinsApproach.pdf>
- [5] Latha, C.M. and Nageswararao, K.S. (2018). Forecasting time series stock returns using arima, *Evidence from S & P BSE Sensex*, 118(24), 1–21.
- [6] Saini, S., Singh, N.P. and Laxmi, Ratna Raj (2016). Application of ARIMA models in forecasting stock prices, *International Journal of Mathematics and Computer Applications Research (IJM-CAR)*, 6(6), 1–10.
- [7] Miiller, T. (1993). Seasonal exponential trends - an application smoothing planning with damped for production, *International Journal of Forecasting*, 9, 509–515.
- [8] Nandakumar, R.R.U.K. and Lokeswari, Y.V. (2018). Stock price prediction using long short term memory, *International Research Journal of Engineering and Technology*, 5(3), 3342–3348. Retrieved from www.irjet.net
- [9] Ogoteng, O. (2017). Stock returns behaviour and the pricing of volatility in Africa's equity markets. Dissertation Masters in Management of Finance and Investment, WITS Business School, University of the Witwatersrand, Johannesburg, South Africa. <https://pdfs.semanticscholar.org/111e/a3aa2697e13875f0dab4ec651e5b5768ee18.pdf>
- [10] Pang, X., Zhou, Y., Wang, P., Lin, W. and Chang, V. (2018). Stock market prediction based on deep long short term memory neural network, in *Proceedings of the 3rd International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS 2018)*, 102–108 (ISBN: 978-989-758-297-4). DOI: 10.5220/0006749901020108.
- [11] Singh, Kuljeet, Shastri, Sourabh, Bhadwal, Arun Singh, Kour, Paramjit, Kumari, Monika, Sharma, Anand and Mansotra, Vibhakar (2019). Implementation of exponential smoothing for forecasting time series data, *International Journal of Scientific Research in Computer Science Applications and Management Studies*, 8(1). https://www.researchgate.net/publication/330970319_Implementation_of_Exponential_Smoothing_for_Forecasting_Time_Series_Data
- [12] Subashini, A. and Karthikeyan, M. (2018). Forecasting on stock market time series data using data mining techniques, *International Journal of Engineering Science Invention (IJESI)*, 6–13. Retrieved from www.ijesi.org, <http://www.ijesi.org/papers/NCIOT-2018/Volume-4/2.%2006-13.pdf>
- [13] Zhang, H. (2013). Modeling and forecasting regional GDP in Sweden using autoregressive models, Master's Thesis in Microdata analysis, Business Intelligence Program, School for Technology and Business Studies, Dalarna University, Sweden.
- [14] Marisa, Angila and Cancela, Roldao (2008). Comparative study of artificial neural network and Box- Jenkins ARIMA for stock price indexes, Dissertation for the degree of Master in Data Analysis Prospecting, ISCTE Business School. <https://repositorio.iscte-iul.pt/bitstream/10071/1472/1>
- [15] Owusu, Frank Kofi (2010). Time series ARIMA modelling of inflation in Ghana, A Thesis submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Tech-

nology in partial fulfillment of the requirements for the degree of Master of Science. <http://hdl.handle.net/123456789/866>

- [16] Awokuse, T. and Ilvento, T. (2012). Using Statistical Data to Make Decisions Module 6: Introduction to Time Series Forecasting, University of Delaware, College of Agriculture and Natural Resources, Food and Resource Economics. <https://docplayer.net/7292575-Module-6-introduction-to-time-series-forecasting.html>

Appendix

Appendix-1: Code used for forecasting stock price for CRDB data

```
[stockpriceF,stockpriceMSE] = forecast(ARIMA_stockprice,20,'Y0',stockprice);
UB = stockpriceF + 1.96*sqrt(stockpriceMSE);
LB = stockpriceF - 1.96*sqrt(stockpriceMSE);
dateF = date(end) + days(1:20);
figure
h4 = plot(date,stockprice,'Color',[.75,.75,.75]);
hold on
h5 = plot(dateF,stockpriceF,'r','LineWidth',2);
h6 = plot(dateF,UB,'k-','LineWidth',1.5);
plot(dateF,LB,'-','LineWidth',1.5);
legend([h4,h5,h6],'stock price for CRDB','Forecast',...
'Forecast Interval','Location','Northwest')
title('Stock price for CRDB Forecast')
xlabel('years')
ylabel('stock price')
hold off
```

Appendix 2: Code for R Software

```
code for software
#loading data into R software
library(readxl)
p3 <- read_excel("C:/Users/hp/Desktop/gebo/juma/p3.xlsx")
view(p3)
#for performing ARIMA and AUTOARIMA
attach(p3)
library(readxl)
library(forecast)
# Defining variables
Y <-closing_price
d.Y<- diff(Y)
dd.Y<-diff(d.Y)
t <-date
arimafit<- auto.arima(Y)
fcast<- forecast(arimafit)
plot(fcast)
standardize_residuals<-residuals(arimafit)
hist(standardize_residuals,plot = TRUE,main = "Histogram of standardize residual")
y=rnorm(1885)
qqplot(standardize_residuals,y,plot = TRUE)
acf(residuals(arimafit),plot = TRUE,lag.max = 20)
Box.test(residuals(arimafit),lag = 20,type = "Ljung-Box",fitdf =1)
tsdiag(arimafit)
```