IJSGS

ISSN: 2488-9229
FEDERAL UNIVERSITY
GUSAU-NIGERIA

INTERNATIONAL JOURNAL OF SCIENCE FOR GLOBAL SUSTAINABILITY

Autoregressive Model for Cocoa Production in Nigeria

Emmanuel Oloruntoba Ajare[1] and Olayemi Joshua Ibidoja[2]

Department of Mathematical Sciences, Federal University Gusau, Gusau, Nigeria

Corresponding Author's E-mail: yemitty203@gmail.com Phone Number: +2347032825699.

Received: September, 2016 Revised and Accepted: November, 2016

Abstract

In this study, the trend and stationarity of cocoa production was examined to check whether it satisfied the statistical assumptions before the autoregressive model of order two after first difference was selected. The plot of the cocoa production was not stationary in mean, as the level changes over time. Phillips-Perron Unit Root Test was used to check the claim and equally showed that cocoa production was not stationary. There was very little evidence for non-zero autocorrelations in the forecasts errors at lags 1-20.The forecast errors seems to have roughly constant variance over time. The histogram of the time series showed that the forecast errors were roughly normally distributed and the mean seems to be close to zero. Therefore, it was plausible that the forecast errors were normally distributed with mean zero and constant variance. The R statistical package was used for all the analyses in this study.

Keywords: Trend; autoregressive model; first difference; cocoa production; stationarity.

1.0 INTRODUCTION

Cocoa is a bean that is in very high demand all over the world. It has several uses and benefits to the economy (www.nairaprojects.com/m/projects/030.html). Africa is the major producer of cocoa to the international markets which are in Europe and America. Ivory Coast, Ghana and Nigeria share the largest contribution to the world markets and with Ivory Coast by far the highest producer producing up to 39% of world output (UNCTAD,2004), the contributions of Nigeria and Ghana are19% and 6% respectively(UNCTAD,2004).Cocoa production in Nigeria is done mostly in South West of Nigeria. The bulk of the country was run with revenues from agricultural produce between (1960 -1970s). During this period, Nigeria belongs to category A. At that period, cocoa was sufficient to produce the bulk of the national need of foreign exchange while other agricultural produce complemented the efforts. The Northern parts were not left out because of revenue from groundnut. Southern parts were known for palm kernel and coal**.** If cocoa is well managed, it would be a support for the manufacturing sector by providing raw materials for industries and gainfully employed for the teeming population (www.nairaprojects.com/m/projects/030.html). Several authors have studied the trend of cocoa production in the past and the incomes generated from it, in Nigeria, cocoa has generated an internal rate of return as high as 42% (Abidogun,1982). The remarkable upsurge in West African production in which Nigeria participated fully can only be explained by the coming into bearing of your trees, an increasing proportion of which were of Amazon and selected high yielding Amelonado varieties and the effects of spraying against pests and diseases. It is important to study cocoa output in Nigeria(FAO, 1966). Thus research findings can be said to have made a significant impact on the Nigerian cocoa industry.This view was given credence and quantitative support by some empirical studies which found that cocoa research in Nigeria has generated an internal rate of return as high as 42%(Abidogun, 1982).

2.0 STATEMENT OF THE PROBLEM

Cocoa production has been reported to be on the decline in contrast to the history of cocoa production in West Africa. In view of this, this study examined the monthly production in tons of cocoa production in Nigeria and developed a parsimonious model for the data collected.

2.1 Objectives of the Study

1. To determine the stationarity of cocoa production, that is if the mean, variance and autocorrelation are constant over time.
2. To construct suitable parsimonious model for the cocoa production.
3. To forecast cocoa production for the next few years.
4. To study the trend and pattern of cocoa production.

3.0 METHODOLOGY

The study examined the methods used in analyzing the data. The data were entered into Microsoft Excel and imported into R GUI. All the analyses were carried out using R statistical package.

3.1Source of data

The data were secondary data extracted from the Central Bank of Nigeria Statistical Bulletin of 2015. The validity of the analysis depends on the accuracy of the data.

3.2Autocorrelation

The initial correlation of the observations in a time series usually expressed as a function of the time lag between observations. The autocorrelation at lag k, $\gamma(k)$, is defined mathematically as;

$$\gamma(k) = \frac{E(X_t - \mu)(X_{t+k} - \mu)}{E(X_t - \mu)^2}$$

Where $X_t, t = 0, \pm1, \pm2, \pm3, \ldots$ represent the values of the series and μ is the mean of the series. E denotes the expected value, the corresponding sample statistic is calculated as follows

$$\hat{\gamma}(k) = \frac{\sum_{i=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{i=1}^{n}(x_t - \bar{x})^2}.$$

Where $\bar{x}$ is the mean of the series of observed values $x_1, x_2, \ldots, x_n$. A plot of the sample values of the autocorrelation against the lag is known as the autocorrelation function or correlogram and is a basic tool in the analysis of time series particularly for indicating possibly suitable models for the series. The term in the numerator of $\gamma(k)$ is the autocovariance. A plot of the autocovariance against lag is called the autocovariance function.

3.3 Partial Autocorrelation

This measures the correlation between the observations a particular number of time units apart in a time series, after controlling for the effects of observations at intermediate time point.

3.4 Autoregressive Model

A model used primarily in the analysis of time series in which the observations z_t at time t function of previous values of the series.

$z_t = \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \cdots + \Phi_p z_{t-p} + a_t$ where a_t is the random disturbance and $\Phi_1, \Phi_2, \ldots, \Phi_p$ are finite set of weight parameters. The process above is called autoregressive process of order p. The autoregressive process of first order $(p = 1)$ and second order $(p = 2)$ are

$z_t = \Phi_1 z_{t-1} + a_t$ and $z_t = \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + a_t$ respectively.

3.5 Moving Average Process

$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$ where $a_t, a_{t-1}, \ldots, a_{t-q}$ are white noise sequence, $\theta_1, \theta_2, \ldots, \theta_q$ are finite set of weight parameters. The process above is called moving average of order q. The moving average process of first order $(q = 1)$ and second order $(q = 2)$ are

$z_t = a_t - \theta_1 a_{t-1}$ and $z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ respectively.

3.6 Autoregressive Moving Average Model

A model for a time series that combines both an autoregressive model and a moving-average model. The general model of order p, q usually denoted by $ARMA(p, q)$ is

$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$, where $\Phi_1, \Phi_2, \ldots, \Phi_p$ and $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the model and $a_t, a_{t-1}, \ldots, a_{t-q}$ are a white noise sequence. In some cases, such models are applied to the time series observations after differencing to achieve stationarity, in which case they are known as autoregressive integrated moving-average models.

4.0 RESULTS AND DISCUSSION

The time plot in Figure 1 showed that the cocoa production in tons is not stationary in mean, as the level changes over time.

4.1 Test for Stationarity

H_o : The cocoa productions contain unit roots.

H_a : The indices are not stationary.

Phillips-Perron Unit Root Test

Dickey-Fuller = -5.9953, Truncation lag parameter = 3, p-value = 0.01

The P-value is less than level of significance $\alpha = 0.05$, the null hypothesis is rejected.

Which means that the cocoa production (in tons) are not stationary at 0.05 level of significance.

ARIMA models are defined for stationary series. Therefore, we need to difference the non-stationary time series to make it stationary.For the ARIMA (p, d,q) model, d is the order of differencing used.The correlogramin figure 4 showed that the partial autocorrelations at lags 1,2 and 15 clearly exceed the significance bounds. Lags 1 and 15 are positive. The partial correlogram shows that the partial autocorrelations at lags 1 and 2 exceed the significance bounds and are negative. The partial autocorrelations tail off to zero after lag 3.The time series does appear to be stationary in mean and variance, after the first difference, therefore ARIMA (p,1,q) is appropriate for the Cocoa Production.By taking the time series of the first differences, the trend component of the time series of the Cocoa Production has been removed and are left with irregular component.Because the correlogram is zero after lag 7 and the partial correlogram tails off to zero after lag 3,this means that different ARMA (autoregressive moving average) models are possible for the time series of first differences.

4.2 Appropriate ARIMA Model

The auto.arima in R package was used to select the appropriate ARIMA model.

Series: cocoatonnes2diff1
ARIMA(2,1,0) with zero mean

Coefficients:

	ar1	ar2
	-0.8787	-0.3965
s.e.	0.1330	0.1313

sigma^2 estimated as 4593: log likelihood=-265.27
AIC=536.55 AICc=537.11 BIC=542.1
An ARIMA(2,1,0) that is with p=2,d=1 and q=0,where d is the order of the differencing used, is selected to be the best candidate model for the time series of first differences of the cocoa production.

Call:
arima(x = cocoatonnes2, order = c(2, 1, 0))

Coefficients:

	ar1	ar2
	-0.8787	-0.3965
s.e.	0.1330	0.1313

sigma^2 estimated as 4593: log likelihood = -265.27, aic = 536.5

4.3 Forecasts Between January 2016 to December 2017
The forecasts for 2 years are shown below
>forecast.Arima (cocoaarima, h=24)

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2016	49.53648	-37.32099	136.3939	-83.30055	182.3735
Feb 2016	41.15600	-46.33862	128.6506	-92.65547	174.9675
Mar 2016	42.20096	-55.35926	139.7612	-107.00452	191.4064
Apr 2016	44.60554	-62.73120	151.9423	-119.55183	208.7629
May 2016	42.07843	-69.48563	153.6425	-128.54408	212.7010
Jun 2016	43.34550	-75.99814	162.6891	-139.17485	225.8659
Jul 2016	43.23415	-81.87924	168.3475	-148.11027	234.5786
Aug 2016	42.82962	-87.65556	173.3148	-156.73024	242.3895
Sep 2016	43.22921	-93.07730	179.5357	-165.23361	251.6920
Oct 2016	43.03850	-98.34269	184.4197	-173.18537	259.2624
Nov 2016	43.04763	-103.41599	189.5113	-180.94915	267.0444
Dec 2016	43.11522	-108.29016	194.5206	-188.43932	274.6698
Jan 2017	43.05221	-113.04451	199.1489	-195.67712	281.7815
Feb 2017	43.08078	-117.63952	203.8011	-202.71971	288.8813
Mar 2017	43.08066	-122.10993	208.2713	-209.55655	295.7179
Apr 2017	43.06944	-126.46807	212.6069	-216.21581	302.3547
May 2017	43.07934	-130.71016	216.8688	-222.70876	308.8674
Jun 2017	43.07509	-134.85510	221.0053	-229.04565	315.1958
Jul 2017	43.07490	-138.90492	225.0547	-235.23922	321.3890
Aug 2017	43.07675	-142.86552	229.0190	-241.29741	327.4509
Sep 2017	43.07520	-146.74478	232.8952	-247.22941	333.3798
Oct 2017	43.07583	-150.54560	236.6973	-253.04259	339.1942
Nov 2017	43.07589	-154.27336	240.4251	-258.74374	344.8955

Dec 2017	43.07559	-157.93219	244.0834	-264.33927	350.4904

4.4 Test for Non-Autocorrelation

Box.test(cocoa$residuals, lag=20, type="Ljung-Box")

Box-Ljung test

data: cocoa$residuals

X-squared = 12.711, df = 20, p-value = 0.8894

Since the correlogram shows that none of the sample autocorrelations for lags 1-20 exceed the significance bounds and the p-value =0.8894, it can be concluded that there is very little evidence for non-zero autocorrelations in the forecasts errors at lags 1-20.

4.5 Forecast Errors

To be certain that the predictive model cannot be improved upon, we checked whether the forecast errors are normally distributed with mean zero and constant variance in Figure 6 and Figure 7, we can make a time plot of the in-sample forecast errors and histogram (with overlaid normal curve) of the forecast errors.The plot shows that the in-sample forecast errors seem to have roughly constant variance over time.The histogram of the time series shows that the forecast errors are roughly normally distributed and the mean seems to be close to zero. Therefore, it is plausible that the forecast errors are normally distributed with mean zero and constant variance.

4.6 Adequacy of the Model

```
> model1=arima(cocoatonnes2,order=c(2,1,0))
> model2=arima(cocoatonnes2,order=c(2,1,1))
> model3=arima(cocoatonnes2,order=c(2,1,2))
> model4=arima(cocoatonnes2,order=c(2,1,3))
> model5=arima(cocoatonnes2,order=c(3,1,1))
> model6=arima(cocoatonnes2,order=c(3,1,2))
>AIC(model1,model2,model3,model4,model5,model6)
df     AIC
model1  3   536.5488
model2  4   537.7442
model3  5   537.9398
model4  6   537.8590
model5  5   539.4240
model6  6   539.6131
```

Model1 has the lowest AIC (Akaike's Information Criterion), therefore the model is adequate and will have the best fit.

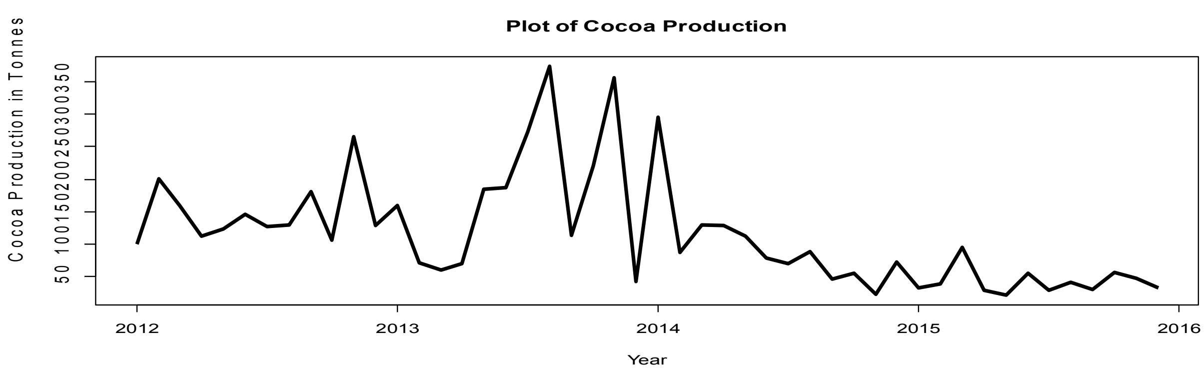

Figure 1: Plot of Cocoa Production.

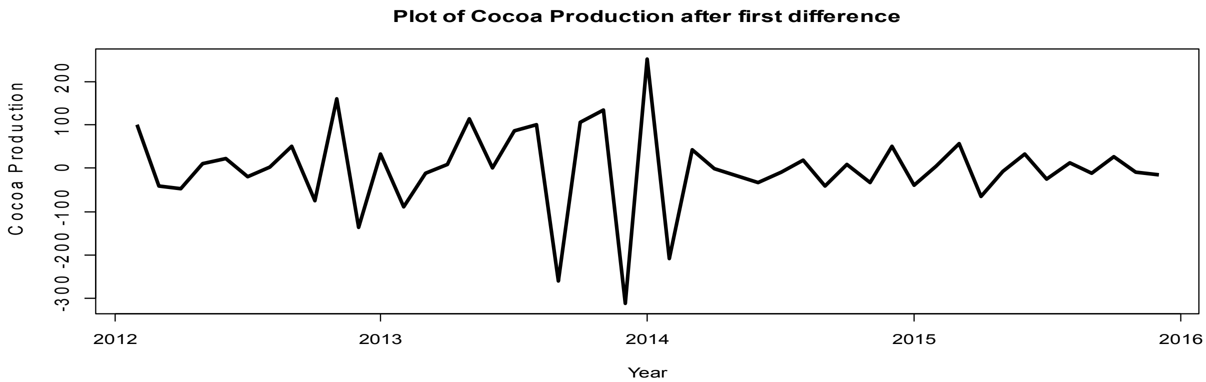

Figure 2: Plot of Cocoa Production after first difference.

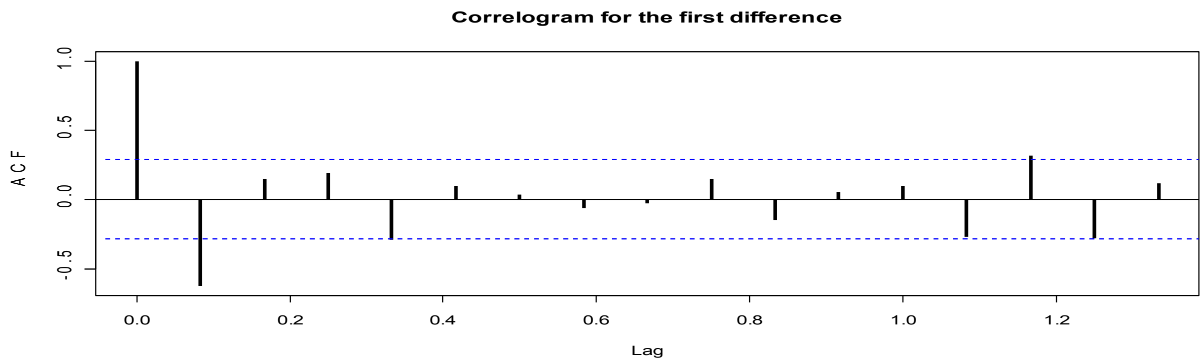

Figure 3: Correlogram for the first difference (ACF)

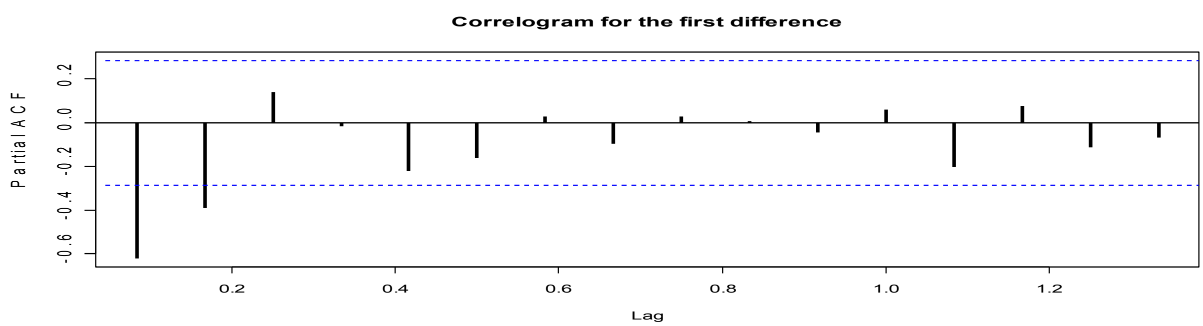

Figure 4: Correlogram for the first difference (PACF)

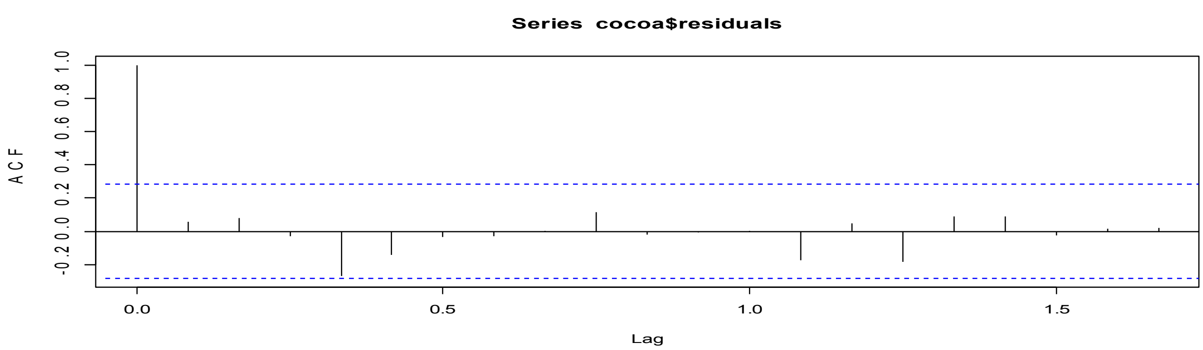

Figure 5: Cocoa Residuals (ACF).

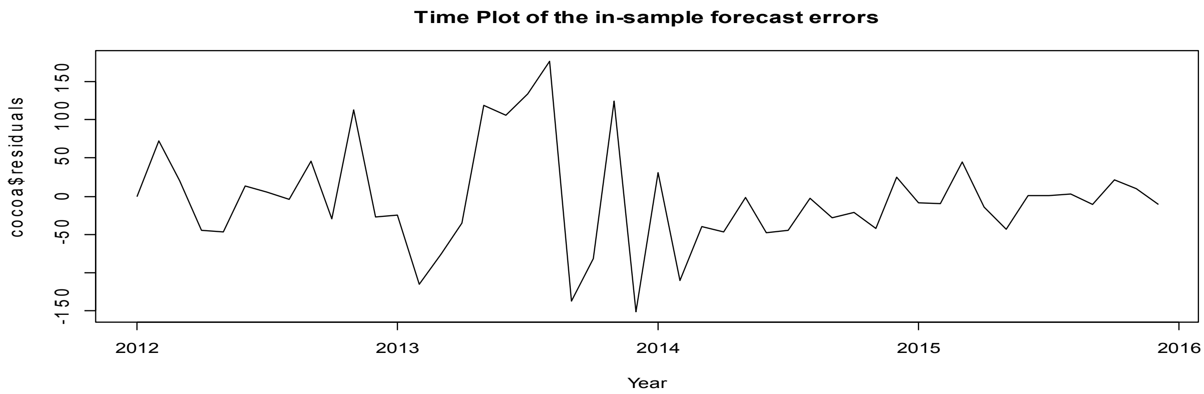

Figure 6: Time Plot of the Forecast Errors.

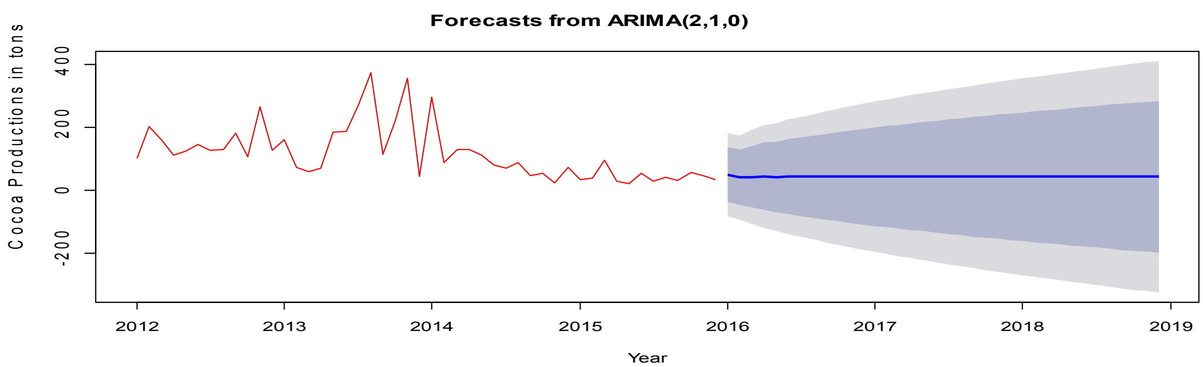

Figure 7: Plot of Forecasts from ARIMA (2,1,0).

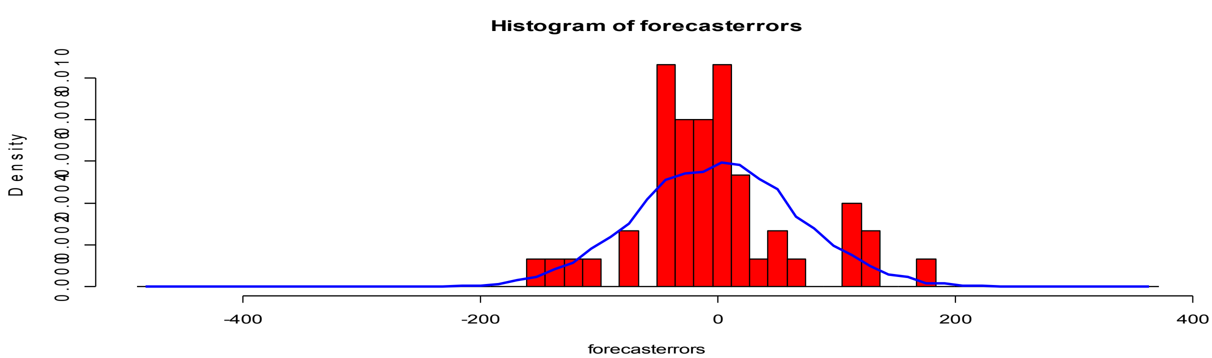

Figure 8: Plot of Histogram of Forecast Errors.

5.0 CONCLUSIONS

The forecasts give the predictions for cocoa production for the next two years with 80% and 95% intervals for the predictions. The autoregressive model after first difference model for the cocoa production gave the forecasted cocoa production for December, 2017 to be 43 tons. An ARIMA (2,1,0) that is with p=2,d=1 and q=0,where d is the order of the differencing used, is selected to be the best candidate model for the time series of first differences of the cocoa production.

6.0 REFERENCES

Bloomfield, P. (2000). Fourier Analysis of Time Series: An Introduction, (2nd edition). Wiley,

Box, G.E.P., Jenkins, G.M. &Reinsel, C.G. (1994). Time Series Analysis, Forecasting and

Control.(3rd Edition). Prentice Hall, Englewood Cliffs, New Jersey.
Crawley, M.J. (2007). The R Book. John Wiley & Sons Ltd, England.
Durbin, J. & Koopman, S.J. (2001). Time Series Analysis by State SpaceMethods. Oxford University Press, Oxford.
Fuller, W.A. (1995). Introduction to Statistical Time Series, (2nd edition).Wiley,New York.
Percival, D.B. and Walden, A.T. (2000).Wavelet Methods for Time SeriesAnalysis. Cambridge University Press, Cambridge.

New York.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL https://www.R-project.org/.
Robert, H.S. and David, S.S. (2005). Time Series Analysis and its applications with R examples.(2nd Edition). Springer, USA.