



MODELLING ANNUAL COCOA PRODUCTION USING ARIMA TIME SERIES MODEL

Oni O.V.¹, Oni O.A.², Akanle Y.O.¹, Ogunleye T.B.³

¹Department of Statistics, Federal College of Animal Health and Production Technology, Ibadan.

²Department of Computer Studies, The Polytechnic, Ibadan

³Department of Agribusiness Management, Federal College of Animal Health and Production Technology, Ibadan.

Cite this article:

Oni O.V., Oni O.A., Akanle Y.O., Ogunleye T.B. (2021), Modelling Annual Cocoa Production Using ARIMA Time Series Model. African Journal of Mathematics and Statistics Studies 4(3), 135-144. DOI: 10.52589/AJMSS-OLPATDNK.

Manuscript History

Received: 22 July 2021

Accepted: 12 Aug 2021

Published: 25 Nov 2021

Copyright © 2020 The Author(s).

This is an Open Access article distributed under the terms of Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), which permits anyone to share, use, reproduce and redistribute in any medium, provided the original author and source are credited.

ABSTRACT: *Cocoa is the most valuable tropical agricultural commodity, comes next to oil; a major target in Nigeria's export diversification strategies. Cocoa production forecasting is important to the Nigerian agricultural transformation agenda. This study attempts to forecast Nigerian cocoa production between 2019 and 2025 using the ARIMA model. The automated analytical procedure implemented in the R software package indicated that ARIMA (0, 1, 1) is the combination with the least Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) and hence, the most appropriate for forecasting. The results revealed that cocoa production would fall by more than 20% in 2025 in comparison with 2018. Thus, to guard against the fall, cocoa farmers in the country should be incentivized through adequate financial and technical assistance.*

KEYWORDS: Annual cocoa production, AR, MA, ARIMA, AIC.



INTRODUCTION

In the international commodity market, cocoa is one of the world's smallest tropical agricultural commodities, its global implications on food, candy producers and the retail industry generate an annual consumption of 45 million tons of cocoa beans. An important source of income and employment for the rural populations of developing countries, particularly for the five to six million small-scale farmers who grow over 90 percent of the world's cocoa.¹

A further 14 million rural workers directly depend on cocoa for their livelihoods^{2,3} and close to 50 million people in total are dependent on cocoa for their livelihoods,⁴ on average, four million tons of cocoa beans are produced each year. Production is dominated by Côte d'Ivoire (1.5 million tons), Ghana (842,000 tons) and Indonesia (447,000 tons), which between them account for around 68 percent of the world's output.⁵ Exports of all cocoa products – cocoa beans, butter, paste and powder – were worth \$20.3bn in 2014, more than double the value in 2005 (\$8.4bn).⁶

Cocoa is a vital source of foreign exchange for governments of developing countries to invest in economic and social development. Cocoa is Côte d'Ivoire's largest export, worth \$2bn in 2013, accounting for 17 percent of all exports. Cocoa is Ghana's third most important export, worth \$1.4bn in 2013 and accounting for 11 percent of all exports,⁶ while the cocoa sector provides a livelihood for around six million people, 30 percent of the population.³

Global demand for cocoa is expected to exceed 4.5 million tons by 2020.⁷ While some predictions foresee production stagnating or falling, leading to a global shortage of cocoa, the International Cocoa Organization (ICCO) forecast is that production will broadly keep pace with demand.⁸

In the light of the falling prices of crude oil globally, it is becoming obvious that Nigeria risks dwindling revenue generating power as well as a low foreign reserve if nothing is done to revive the non-oil sector of the economy especially the agricultural export crops of which cocoa is a factor. Therefore, there is the need to forecast the yield of cocoa in Nigeria into the future. This will assist policy makers in the economy to come up with necessary policy decisions that will preserve the cocoa export sector as well as help prevent a near total loss of the nation's income generating capacity from the non-oil sector.

The paper therefore, is forecasting the annual cocoa production in Nigeria using Auto-Regressive Integrated Moving Average also called Box-Jenkins Methodology of forecasting.

METHODOLOGY

The annual cocoa production data for Nigeria for the period of 1961 to 2018 was obtained from the Food and Agriculture Organization of the United Nations statistical database.⁹ Model identification, estimation and forecast will be carried out using the appropriate model. Both inferential and descriptive statistics were used to analyze this study, while the descriptive statistics involved the use of graphs to examine the cocoa production trend, as well as some ARIMA properties of the data, the ARIMA model was used to forecast cocoa production in Nigeria.

ARIMA Modelling

ARIMA modelling also known as Box-Jenkins Methodology has evolved from the combination of AR (Autoregressive) and MA (Moving Average), the ARMA models.^{10,11,12}

Let z_t be a discrete time series variable which takes different variables over a period of time. the corresponding AR (p) model of z_t series, which is the generalizations of the autoregressive model, is expressed as;

$$AR(p): z_t = \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \dots + \Phi_p z_{t-p} + a_t \quad \dots .1$$

where a_t is the random disturbance and $\Phi_1, \Phi_2, \dots, \Phi_p$ are finite set of weight parameters. The process above is called autoregressive process of order p .

Similarly, the MA(q) model which is the generalization of the moving average model is specified as:

$$MA(q): z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad \dots .2$$

where $a_t, a_{t-1}, \dots, a_{t-q}$ are white noise sequence, $\theta_1, \theta_2, \dots, \theta_q$ are finite set of weight parameters. The process above is called moving average of order q .

When (z_t) in the data is replaced with $\Delta z_t = z_t - z_{t-1}$, then the ARMA models become the **ARIMA** (p, d, q) models, where p is order of autocorrelation (indicates weighted moving average over past observations), d is order of integration (differencing) and q is order of moving averaging. By combining the models in (1) and (2), this resulting model is referred to as ARIMA model, which have the general form of:

$$\begin{aligned} ARIMA(p, q): z_t \\ = \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \dots + \Phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \\ - q \dots .3 \end{aligned}$$

where $\Phi_1, \Phi_2, \dots, \Phi_p$ and $\theta_1, \theta_2, \dots, \theta_q$ are the parameters of the model and $a_t, a_{t-1}, \dots, a_{t-q}$ 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.

Test of Stationarity

Time series is said to be stationary if both the mean and the variance are constant over time. A time plot of the data can suggest whether the time series needs any differencing before performing formal tests. Also, the stationarity and non-stationarity properties are checked by applying Augmented Dickey–Fuller test (ADF) for a unit root in a time series sample.

Model Selection Criteria

When fitting models, there is the tendency of two or more models competing and for that reason it is appropriate to use good model selection criteria to select the most adequate model. In this study, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were the measures of goodness of fit that were employed to select the most adequate model. For a given data set, several competing models may be ranked according to their AIC, or BIC

values with the one having the lowest information criterion value being the best. In general, the AIC, and BIC are given by;

$$AIC = 2k + n \log (RSS/n)$$

$$BIC = \log (\delta_e^2) + \frac{k}{n} \log (n)$$

where k is the number of parameters in the statistical model,

RSS is the residual sum of squares of the estimated model,

n is the number of observations in the data, δ_e^2 is the error variance.

Model identification

The ARIMA model needs to be stationary, that is should have a constant mean, variance and autocorrelation through time. Therefore, usually the series' first need to be different until it is stationary. However, the first step in the identification of an ARIMA process is to examine the time plot if there is no seasonal variation. The three commonly used tools graphical methods in identification in time series are: - the time plot of the series, the time plot of autocorrelation at various lags (ACF) and the time plot of the partial autocorrelation function (PACF).

RESULT AND DISCUSSION

Exploratory Data Analysis

This section briefly examines the descriptive statistics of cocoa bean production, the time plot and the correlogram of annual production of cocoa beans in Nigeria. Figure 1 below showed an irregular movement which depicts a non-seasonal movement of cocoa production. Also, the observed time plot showed that the series has no constant mean and variance with all evidence of stationarity. The ACF and PACF plots indicate that the cocoa bean production data is stationary at level since there are enough spikes in the plots outside the significant zone, so we can conclude that the residuals are not random.

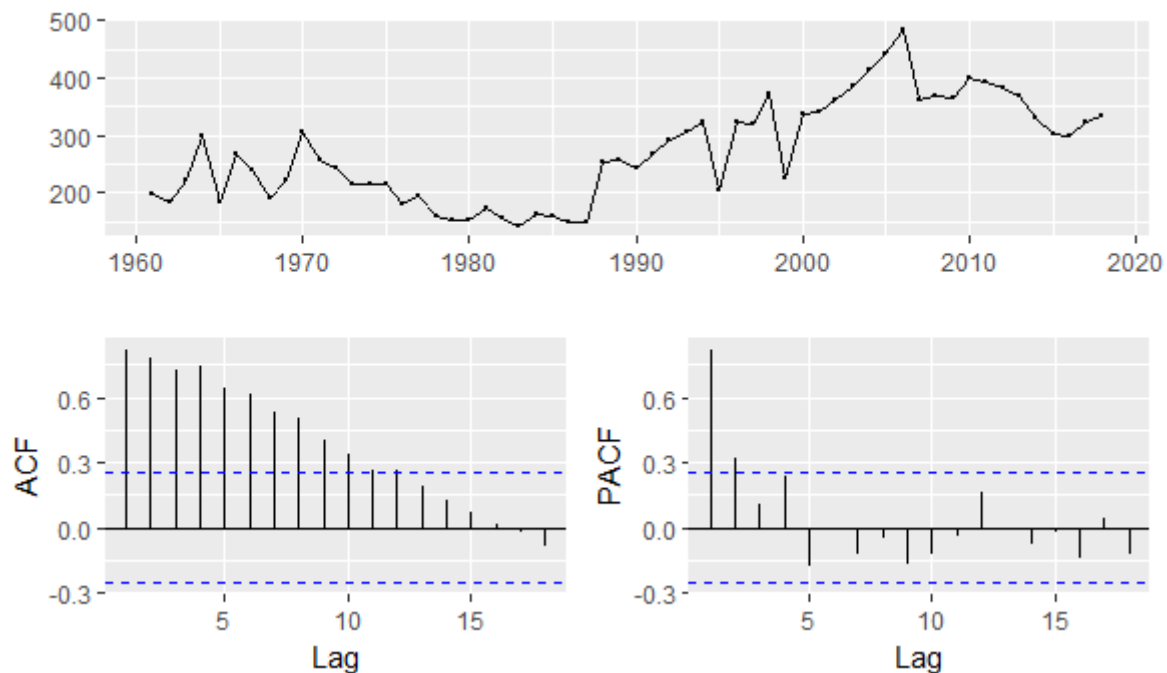


Figure 1: Time Plot and Correlogram of the Annual Cocoa Production from 1961-2018

Test for Stationarity Cocoa Production

We make use of the R statistical package to examine the correlogram and Dickey-Fuller test to examine the time series property of cocoa beans production in Nigeria.

Table 1: Augmented Dickey-Fuller Test

Dickey-Fuller	Lag order	p-value	Alternative Hypothesis
-1.7228	3	0.6866	Stationary

The result of the ADF test in table 1 above, indicates that the observed p-value is much higher than the critical value, thus it is clear that the series suffers from non-stationary at level.

Time Series Plot After First Difference

It can be observed in figure 2 that the time plot of the data set at first difference is stationary, at both mean and variance. The autocorrelation function ACF and PACF clearly shows that the data is stationary since most of the spikes lie within the control limit in the circulation of the lags of the years.

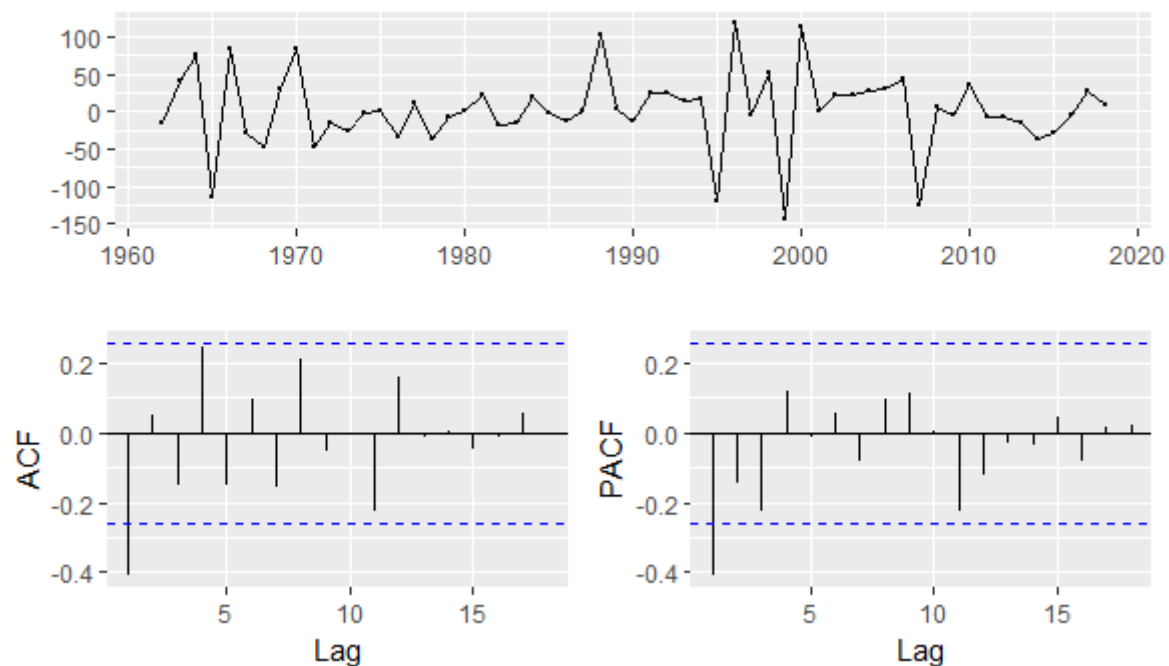


Figure 2: Time Plot of the First Difference of the Annual Cocoa Production

The result of table 2, shows that the test statistic for the ADF test is less than asymptotic critical values at all levels of significance, we therefore reject the null hypothesis and conclude that the series is stationary.

Table 2: Augmented Dickey-Fuller Test

Dickey-Fuller	Lag order	p-value	Alternative Hypothesis
- 3.9114	3	0.01979	Stationary

Model identification

Since we achieved stationarity at first difference of the data, therefore the order of the Integration is 1, and we proposed (p and q) to be 0, 1 and 2 in which 13 ARIMA models are generated and we select the best among them by considering Akaike Information Criterion (AIC).

Table 3: Results of ARIMA Model Identification for Cocoa Production

MODEL	AIC
ARIMA(2,1,2) with drift	610.1448
ARIMA(0,1,0) with drift	614.0419
ARIMA(1,1,0) with drift	605.8301
ARIMA(0,1,1) with drift	603.7172
ARIMA(0,1,0)	612.0172
ARIMA(1,1,1) with drift	605.9873
ARIMA(0,1,2) with drift	605.9852
ARIMA(1,1,2) with drift	Inf
ARIMA(0,1,1)	601.9641*
ARIMA(1,1,1)	604.1377
ARIMA(0,1,2)	604.1342
ARIMA(1,1,0)	603.9068
ARIMA(1,1,2)	Inf

From the above table we found that the ARIMA (0,1,1) model has the least values of all the criteria, therefore we will make use of this model for further analysis.

Parameter Estimates of The Model

Table 4: Results of ARIMA (0, 1, 1) Model Estimation for the Annual Cocoa Production

Parameter	Coefficient Estimate	Standard Error
ma1	-0.4705	0.1091
sigma ²	2126	
log likelihood=-298.87		
AIC=601.74 AICc=601.96 BIC=605.83		

Model Diagnostic Checking

Before we use the model, we are to look at some tests/plots to check whether these volatility models have adequately captured all of the persistence in the variance of returns. And if as much as the model is adequate, then standard residuals are uncorrelated as specified in the results below.

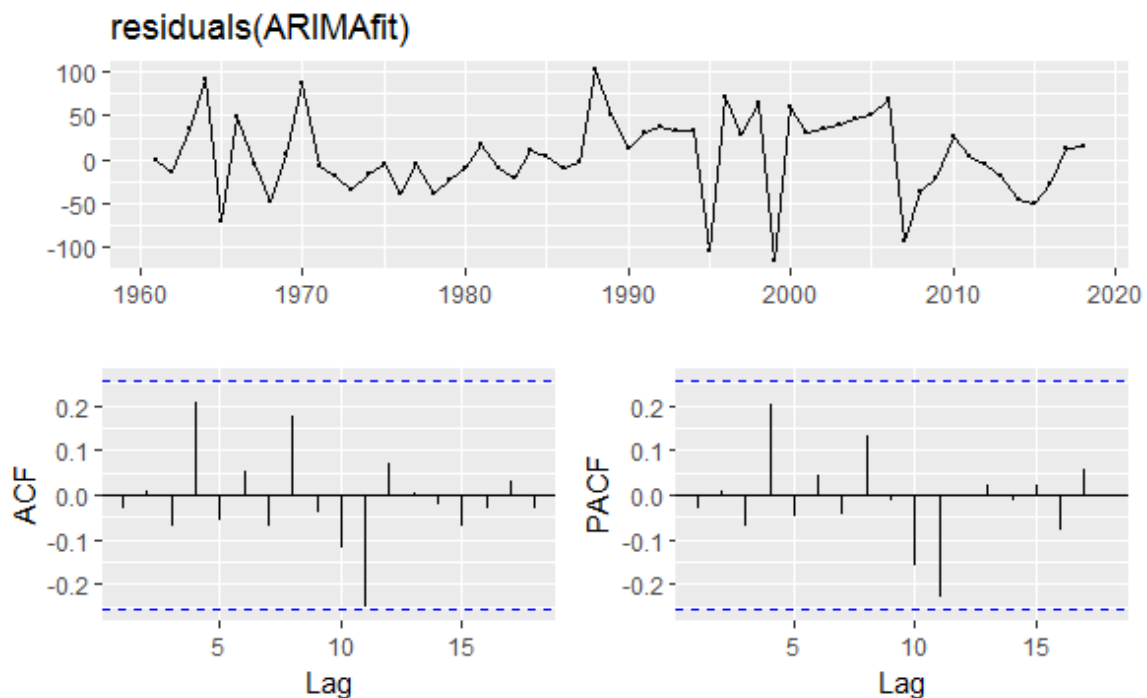


Figure 3: The ACF and PACF of ARIMA (0, 1, 1) Model

It could be observed from figure 3 that the residual ACF and PACF of ARIMA (0, 1, 1) model lie within the 90% confidence interval. The model has passed the standard test criteria of being white noise.

Forecast

The ability of a well fitted ARIMA model is in its ability to forecast the time series data that is tested. This further testifies the validity of the model.

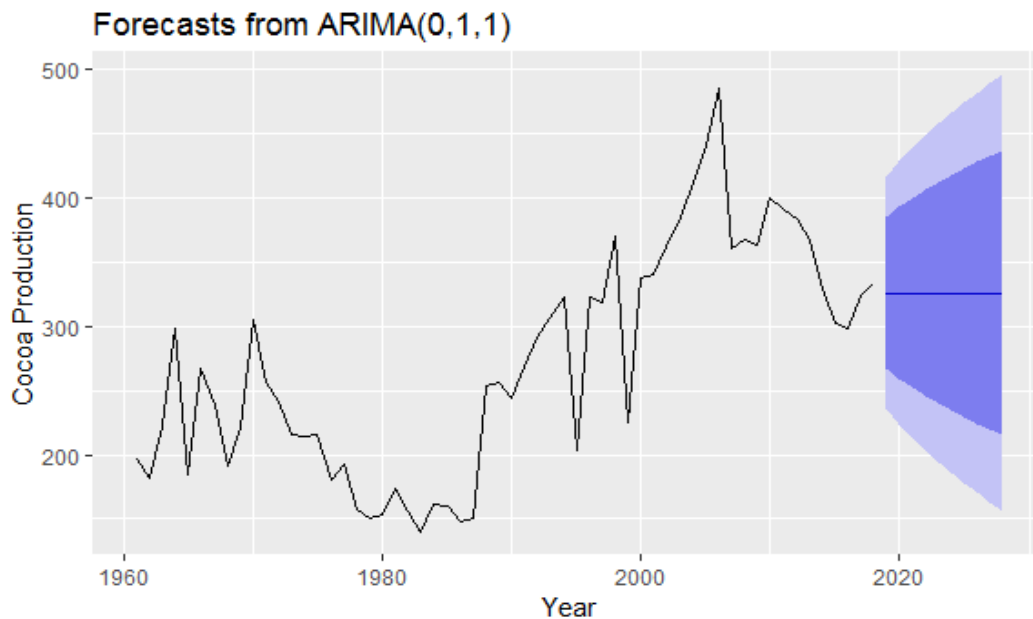


Figure 4: Plot of Forecast from ARIMA (0, 1, 1) Model

CONCLUSION AND RECOMMENDATION

The use of autoregressive integrated moving average modeling strategy is now among the most popular ways of analysis of time series data. We employed this method to model the annual cocoa production and forecast for 2019 to 2025. An ARIMA (0,1,1) that is with $p=0$, $d=1$ and $q=1$, is selected to be the best candidate model for the time series of first differences of the cocoa production considering Akaike Information Criterion (AIC).

The forecasts give the predictions for cocoa production for the next seven (7) years with 80% and 95% intervals for the predictions. The results revealed that cocoa production would fall by more than 20% in 2025 in comparison with 2018.

In order to guard against the fall in cocoa production in Nigeria, cocoa farmers in the country should be incentivized through adequate financial and technical assistance. The cocoa rebirth programme initiated by the government should be resuscitated, and likewise, partnership with research institutions should be encouraged in developing new high yielding and disease resistant cocoa varieties.

REFERENCES

- [1]. Cocoa market update, World Cocoa Foundation, May 2010
- [2]. Cocoa Barometer 2012, tropical commodity coalition
- [3]. Cocoa barometer 2015, fountain A.C and hutz-Adams. F (2015)
- [4]. Cocoa market update, world cocoa foundation, may 2010



-
- [5]. Average annual production 2009/10 to 2013/14, ICCO, www.icco.org
 - [6]. UN International Trade Statistics, 2014 Yearbook, vol. 11
 - [7]. Chocolate crisis: Cocoa demand will be unsustainable by 2020, Benzeinga.com, 10 October (2013)
 - [8]. Cocoa supply and demand: what to expect in the coming year? ICCO London, 22 September (2015).
 - [9]. FAO, FAOSTAT database, data by volume for 2018
 - [10]. Box, George E.P, and Jenkins, G.M. (1976): Time series analysis: Forecasting and control, Revised edition, Oakland, CA: Holden-Day.
 - [11]. Padhan, P.C (2012). Application of ARIMA model for forecasting agricultural productivity in India. Journal of agriculture & social sciences,8;50-56
 - [12]. Brockwell, P. J. & Davis, R. A. (2006), Introduction to time series and forecasting, Springer Science & Business Media.