International Journal of Tropical Agriculture
Volume 40, Numbers (3-4) • 2022 : pp. 241-246
ISSN: 0254-8755

Non-Seasonal Cocoa Monthly Price Forecasting using Univariate Time Series Approach

Prema Borkar
Assistant Professor, Gokhale Institute of Politics and Economics (Deemed University), Pune – 411004
E-mail: premaborkar@rediffmail.com

Abstract: Cocoa is a major plantation crop grown for chocolate production worldwide. Cocoa is grown in Kerala, Karnataka, Andhra Pradesh, and Tamil Nadu in India. Andhra Pradesh leads the way in terms of area and production. There are currently ten multinational firms active in the cocoa sector in India, exporting items such as beans, chocolates, cocoa butter, cocoa powder, and cocoa-based products to other countries. A univariate time series approach was used in this study to forecast non-seasonal cocoa monthly prices in India. Monthly cocoa prices in India from September 2017 to September 2022 were used in the study. The ARIMA model is an extrapolation method that only requires historical time series data on the variable under consideration. The Box-Jenkins model provides a validated method for identifying and filtering the most appropriate variations for the series under consideration.

The autocorrelation and partial autocorrelation functions were used to estimate the model. Model parameters were estimated using the R programming language. The performance of the fitted model was evaluated using various goodness of fit measures such as AIC, BIC, and MAPE. According to empirical findings, the ARIMA (0,1,0) model performed best in predicting monthly cocoa prices in India. Projections for the period from October 2022 to March 2023 are calculated using the chosen model. Based on the fitted model, monthly cocoa prices are found to be constant for the next six months.The outcomes of the monthly cocoa prices are represented graphically and numerically. Forecasting is a critical phase in the process of advising policymakers. In an ideal situation, decision-makers would make judgments based on precise projections to tighten restrictions and cause unexpected outcomes. The most accurate forecasting models are also taken into consideration by marketing strategists when developing, planning, and executing marketing plans. These projections are especially helpful for changing their marketing and policy plans to account for upcoming changes.

Keywords: ACF - Autocorrelation Function, AIC – Akaike Information Criterion, ARIMA – Autoregressive Integrated Moving Average, MAPE - Mean Absolute Percent Error, PACF - Partial Autocorrelation Function, Residual Analysis.

INTRODUCTION

Cocoa is a major plantation crop produced for chocolate production around the world. The cocoa tree is a tiny (4 to 8 m tall) evergreen tree. Cocoa is the dried and fully fermented fatty seed of the cacao tree that is used to make chocolate. “Cocoa” can also refer to the drink known as hot chocolate; cocoa powder, the dry powder created by grinding cocoa seeds and separating the cocoa butter from the dark, bitter cocoa solids; or the combination of cocoa powder and cocoa butter. Because the tree needs between 40 and 50 percent of shade, cocoa is mostly produced in India’s Andhra, Tamil Nadu, Karnataka, and Kerala as an intercrop. The majority of cocoa is grown in coconut groves, followed by around a fifth in areca nut groves, and the last third in oil palm and rubber plantations.

Cocoa is grown on an area of 1,03,376 ha in the Indian states of Kerala, Karnataka,

Article History: Received: 11 June 2022; Revised: 27 July 2022; Accepted: 22 August 2022

Andhra Pradesh, and Tamil Nadu, with a total production of 27,072 MT. Andhra Pradesh leads the way in terms of area with 39,714 hectares and production with 10,903 MT. Andhra Pradesh has the highest production at 950 kg/ha. The average cocoa productivity in India is 669 kg/ha. Numerous chocolate-making enterprises are gradually promoting the region under cultivation as contract farming.

Cocoa is primarily an export commodity. There are currently ten multinational firms in India active in the cocoa sector, exporting items such as beans, chocolates, cocoa butter, cocoa powder, and cocoa-based products to other nations. Exports of cocoa beans and products earn India Rs. 1108 crores in foreign exchange. The demand of the chocolate industry and confectionaries in India is 50,000 MT of dry beans per annum. The current domestic production of cocoa beans is not sufficient to meet the demand of the industry. Hence India is importing a lion's share of its requirement from other cocoa-growing countries worth Rs.2021crores.

The researcher has conducted numerous studies to fit an ARIMA model in the agriculture industry across the globe for various agricultural crops. Different agricultural sectors employ the ARIMA model to forecast agricultural production. The pertinent work for forecasting utilizing the ARMA model developed by Box-Jenkins in 1970, from which we can learn about forecasting methods for various agricultural productions. The development of this literature gained additional aspects because of forecasts by Goodwin and Ker (1998)[3]. To effectively describe agricultural yield series, they developed an ARIMA (0, 1, 2) univariate filtering model.Falak and Eatzaz (2008)[2] examined the potential for Pakistan to produce wheat in the future using the ARIMA model. With monthly data from January 1998 to December 2000, Hossian et al. (2006) [4] predicted three various types of pulse prices in Bangladesh, including motor, mash, and mung. Rahman (2010)[6] adapted an ARIMA model to predict the production of Boro rice in Bangladesh. The study by M. A. Awal and M. A. B. Siddique [5] looked at growth patterns and the best ARIMA model for accurately forecasting Aus, Aman, and Boro rice production in Bangladesh.The main objective of this study is to develop an ARIMA model for forecasting monthly cocoa prices in India.

MATERIAL AND METHODS

The present study made use of time series data. The information was gathered from the International Cocoa Organization. The cocoa prices information ranged from September 2017 to September 2022. The data were subjected to the linear time series model of Box and Jenkin (1970)[1]. Common names for this approach include the Autoregressive Integrated Moving Average Model (ARIMA Model).

One of the time series models that is often utilized is the ARIMA model. The ARIMA (p, d, q) model combines the Autoregressive (AR) model, which illustrates the relationship between the present (Y) and historical (Y_{t-k}) values, with a random value and moving average (MA) model that shows that there is a relationship between a present value (Y_t) and residuals in the past(Z_{t-k} k = 1,2...) with a non-stationary data pattern and d differencing order.

The form of ARIMA (p,d,q) is:

$$\Phi_p(B)(1 - B)^d Y_t = \Theta_q (B)\, a_t$$

Where, p is AR model order, q is the MA model order, d is differencing order and

$$\Phi_p (B) = (1 - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_p B^p)$$

ARIMA model types are listed using normal ARIMA notation (p,d,q). Autoregressive (p) terms mean how many autoregressive orders are included in the model. The prior values from the series that are used to predict the current values are defined by autoregressive orders. Difference (d) specifies the sequence in which differencing is done to the series before model estimation. Since ARIMA modeling implies stationarity and trends are often nonstationary, differencing is required when trends are present in order to mitigate their effects. First-order differencing accounts for linear trends, second-order differencing accounts for quadratic trends, and so on. The order of differencing corresponds to the degree of the series trend. Moving Average (q) terms indicate the model's number of moving average orders. Moving average orders indicate how previous values' departures from the series mean are utilized to predict current values.

Model identification

We looked at the ACF graph to see whether the series is stationary. The time series value should be regarded as stationary if an ACF graph breaks off or dies out pretty rapidly. The Autoregressive Integrated Moving Average (ARIMA) model is used to analyze non-seasonal series (p,d,q). Here, the letters p stands for the order of the autoregressive part, d for the average amount of the difference, and q for the order of the moving average part. The ARIMA models become the ARMA models if the original series is stationary and d = 0.

The difference linear operator is denoted by:

$$\Delta Y_t = Y_t - Y_{t-1} = Y_t - BY_t = (1 - B)Y_t$$

The stationary series

$$W_t = \Delta^d Y_t = (1 - B)^d Y_t = \mu + \Theta_q(B)\varepsilon_t$$

Model estimation and Diagnostics

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

Estimate the parameters for the given preliminary model. The generated model must be evaluated for suitability when determining if the residuals from an ARIMA model have a normal and random distribution. Ljung-Box Q statistics offer a general assessment of the model's suitability. The following equation contains the test statistics Q:

where *n* is the sample size, $\rho_k$isthe sample autocorrelation at lag *k*, and *h* is the number of lags being tested.A common technique in autoregressive integrated moving average (ARIMA) modeling is the Ljung-Box test. It should be noted that it is applied to the residuals of a fitted ARIMA model rather than the original series, and in such applications, the premise that the ARIMA model's residuals lack autocorrelation is evaluated. The degrees of freedom must be changed to reflect parameter estimation when testing the residuals of an estimated ARIMA model.

RESULTS AND DISCUSSION

For modeling purposes, monthly cocoa price data from the previous 60 months was employed from September 2017 to September 2022. Figure 1 displays the monthly trend for cocoa prices in India.Box-Jenkins ARIMA methodology partially solved the challenge of determining optimal values for p, d, and q by following the previously outlined procedures. The stationarity test was the first stage in fitting the ARIMA model. At the 5%, significant levels, the computed ADF test statistic 'tau' (-0.00639) was determined to be greater than critical values (-2.1472), resulting in the acceptance of the null hypothesis (H_0: Data set is nonstationary). As a result, the production data series was nonstationary. After accounting for the initial difference, the ADF test statistic 'tau' (-5.2874) was found to be less than crucial values (-2.1472) at the 5% significant level, indicating that the series was stationary. As a result, the value of 'd' was believed to be 1.

Figure 1: Monthly Trend of Cocoa Prices in India from Sep 2017 to Sep 2022

The basic tools for determining stationarity in time series are the autocorrelation and partial autocorrelation functions. Figures 2 and 3 display the ACF and PACF functions for the initial dataset's monthly cocoa price data.We derive a correlation of forecast errors for lags 1 to 12. The correlogram demonstrates that the autocorrelations for lags 1, 2, 3, and 4 are over the significance

boundaries and that they start to decline after lag 4. Positive autocorrelations exist for lags 1, 2, 3, and 4; their magnitudes get smaller as the lag increases. The autocorrelation for lags also exceeds the significance constraints, although this is most likely due to chance, as they only exceed the significance limitations.From the partial autocorrelogram, we see that none of the partial autocorrelations is positive. The partial autocorrelation tails off to zero after lag 1.

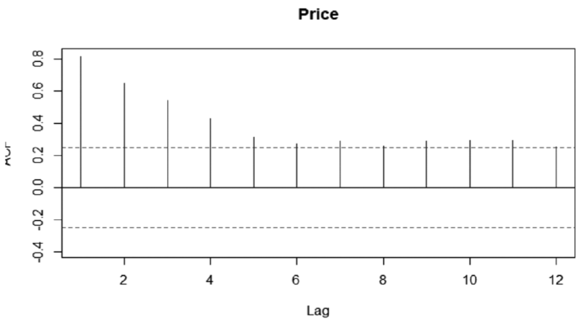

Figure 2: Autocorrelation function for Monthly Cocoa Prices

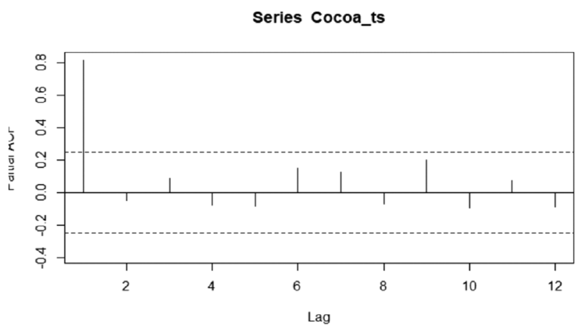

Figure 3: Partial Autocorrelation function for Monthly Cocoa Prices

After taking the first difference, the data of monthly cocoa prices becomes stationary. Figure 4 depicts the monthly trend of the first differenced cocoa prices. Figures 5 and 6 display the first differenced ACF and PACF plots for the monthly cocoa prices data. From figure 5, we see that none of the autocorrelations at lag exceeds the significance bounds. From the partial autocorrelogram i.e., figure 6, we see that the partial autocorrelation at lag 4 is positive and exceeds the significance bounds, while others are negative and do not exceed the significance bounds.

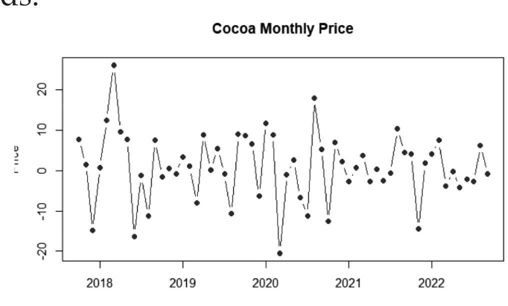

Figure 4: Monthly Trend of first differenced Cocoa Prices

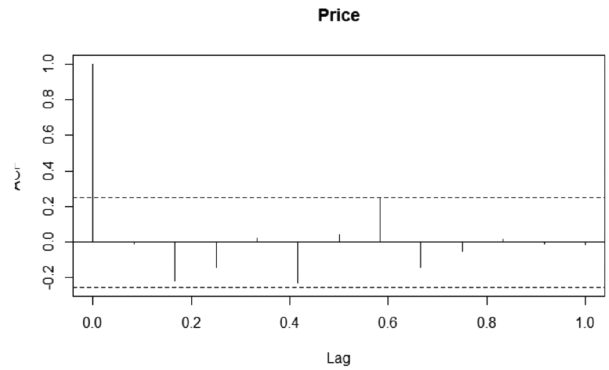

Figure 5: First Differenced Autocorrelation function for Monthly Cocoa Prices

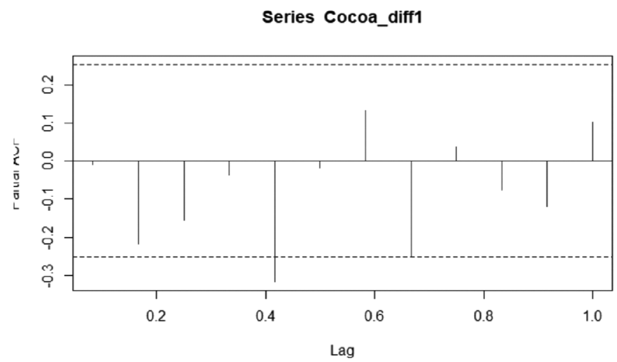

Figure 6: First Differenced Partial Autocorrelation function for Monthly Cocoa Prices

Model Building

After deciding the the 'd' in the model, the next step is to estimate the 'p' and 'q' in the model. This issue was partially solved by examining the series' ACF and PACF. The coefficients of the chosen model were estimated based on the minimum AIC, BIC, and RMSE criteria. Finally, the ARIMA (0, 1, 0) model was chosen. Table 1 displays the model fit statistics for the first differenced data set. In statistics, the mean absolute percentage error (MAPE), also referred to as the mean absolute percentage deviation (MAPD), is a metric for forecasting model accuracy. The model's mean absolute percentage error (MAPE) is 3.71 percent. The AIC and BIC values were estimated as 426.89 and 428.95, respectively.

A measure of how far a dependent series deviates from its projected level. Root Mean Square Error (RMSE), sometimes known as the square root of mean square error, is a measure of how far a dependent series deviates from its projected level, stated in the same units as the dependent series.The RMSE of the fitted model is 8.2778.

Table 1: Model fit statistics for the Monthly Cocoa Prices

Model Fit Statistics	*Values*
AIC (Akaike Information Criterion)	426.89
BIC (Bayesian Information Criterion)	428.95
ME (Margin of Error)	0.9013
RMSE (Root Mean Square Error)	8.2778
MAE (Mean Absolute Error)	6.1692
MPE (Mean Percent Error)	0.4530
MAPE (Mean Absolute Percentage Error)	3.7141
MASE (Mean Absolute Scaled Error)	0.4915
ACF1	-0.0108

Model Verification

We created an ACF and PACF plot as well as a forecast error histogram to determine whether the forecast errors are normally distributed with a mean zero and constant variance. Figure 7 displays the autocorrelation and partial autocorrelation residual function. The residual ACF and PACF did not exhibit any noteworthy value. Insignificant values were returned by the box-Ljung statistics (8.1546), which was in line with the idea that randomness governs residuals. The variance of the forecast errors appears to be approximately consistent throughout time, according to the forecast error time plot. The time series' histogram (Figure 8) reveals that the forecast errors are nearly normally distributed and that the mean appears to be very low. Consequently, it is conceivable that the forecast errors have a normal distribution with a mean zero and a constant variance.

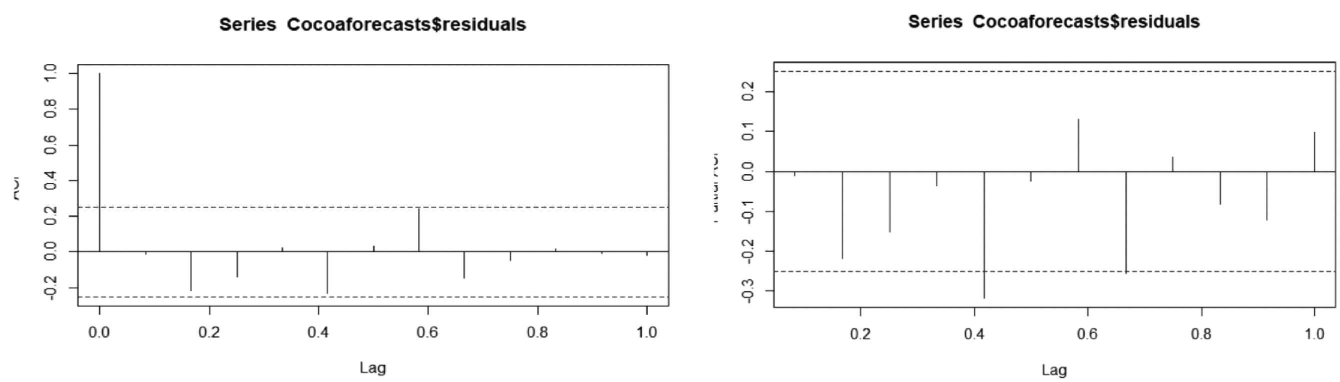

Figure 7: Autocorrelation and Partial autocorrelation function of residuals

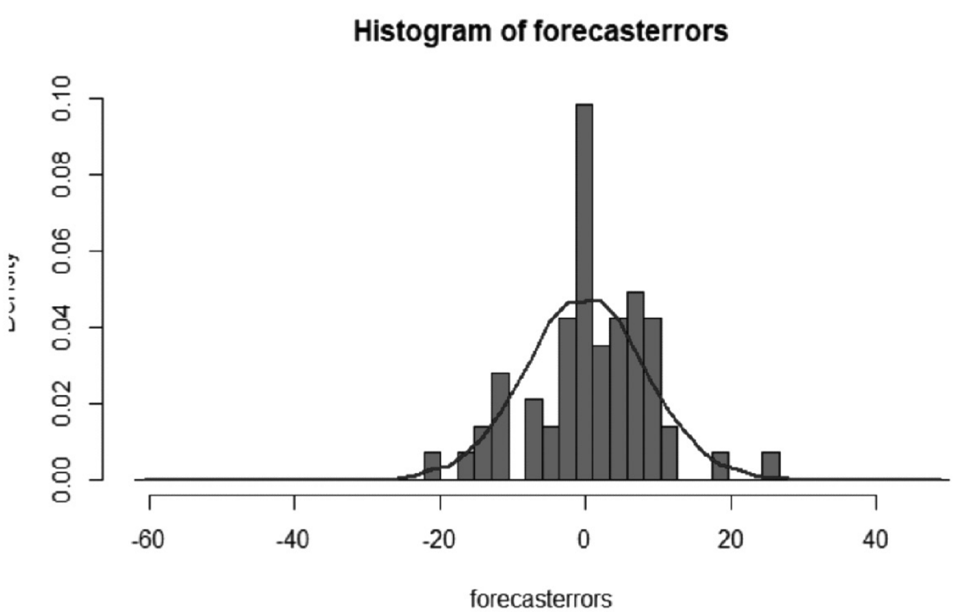

Figure 8: Histogram of forecast errors for Monthly Cocoa Prices

The ARIMA (0, 1, 0) does appear to provide a sufficient predictive model for the monthly cocoa prices because the forecast errors do not appear to be associated with one another and appear to be normally distributed with mean zero and constant variance.

Forecasting the Monthly Cocoa Prices

Table 2 shows a sample of monthly cocoa prices forecasted from October 2022 to March 2023. The forecasted values are the monthly point predictions from October 2022 to March 2023. It indicates that monthly cocoa market prices will remain stable for the next six months. Figure 9 also indicates that the monthly prices of cocoa show a constant trend.The upper and lower confidence limits indicate the researcher's expectation that research cocoa prices will not fall below the lower confidence limit or rise above the upper confidence limit in the next six months.

Table 2: Forecasted monthly Cocoa Prices using ARIMA (0, 1, 0)

Month	*Forecasted Prices (Rs/Kg)*	*Lower Confidence Limit*	*Upper Confidence Limit*
October 2022	183.77	167.41	200.13
November 2022	183.77	160.64	206.91
December 2022	183.77	155.44	212.10
January 2023	183.77	151.05	216.48
February 2023	183.77	147.19	220.35
March 2023	183.77	143.69	223.84

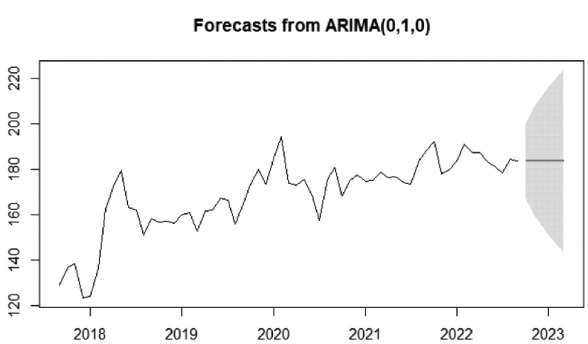

Figure 9: Observed and Forecasted monthly prices of Cocoa

CONCLUSION

A time series model looks for patterns in a variable's past movement and uses that information to forecast future values. This analysis attempted to fit the best model to forecast cocoa monthly prices. Model selection criteria such as AIC, BIC, and MAPE were used to select the best model for forecasting. ARIMA (0,1,0) is the best-chosen Box-Jenkins ARIMA model for forecasting monthly cocoa prices. It suggests that cocoa prices will remain stable for the next six months. These models were ultimately selected to forecast because they can adequately describe real-world scenarios. In the entire process of advising policymakers, forecasting is essential. When changes are assessed from two angles—current events and what is expected to happen in the future—perfect decision-making results. To tighten regulations and produce final results that deviate from expected outcomes, authorities should ideally base their actions on reliable projections.

References

Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis, Forecasting and Control. San Francisco, Holden- Day, California, USA.

Falak, S. and Eatzaz, A. (2008) Forecasting Wheat Production of Pakistan. The Lahore Journal of Economics, 13(1): 57-85.

Googwin, B. K. and Ker, A. P. (1998) Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group Risk Crop Insurance Contracts. Amer. J. Agr. Econ. 80: 139-153.

Hossain, M. Z., Samad, Q. A. and Ali, M. Z. (2006). ARIMA model and forecasting with three types of pulse prices in Bangladesh: a case study. International Journal of Social Economics 33 (4):344-53.

M. A. Awal and M.A.B. Siddique (2011). Rice Production in Bangladesh Employing By ARIMA Model, Bangladesh J. Agril. Res. 36(1): 51-62.

N. M. F. Rahman (2010). Forecasting of boro rice production in Bangladesh: An ARIMA approach. Journal of Bangladesh Agricultural University. 8(1): 103–112.