DOI: 10.54083/ResBio/4.3.2022/156-160

Res. Bio., 2022, 4(3):156-160

Research Biotica

e-ISSN: 2582-6743

September, 2022

Article ID: RB137

Research Article

Time Series Analysis of Monthly Coffee (Robusta) Prices in India using Box-Jenkins Approach

Prema Borkar

Gokhale Institute of Politics and Economics (Deemed University), Pune, Maharashtra (411 004), India

Open Access

Corresponding Author

Prema Borkar

✉: drmoloysb@gmail.com

Conflict of interests: The author has declared that no conflict of interest exists.

How to cite this article?

Borkar, 2022. Time Series Analysis of Monthly Coffee (Robusta) Prices in India using Box-Jenkins Approach. *Research Biotica* 4(3): 156-160.

Abstract

Robusta coffee is a type of coffee made from the *Coffea canephora* plant's beans (seeds). It is the world's second most popular coffee, accounting for 43% of global coffee production with arabica constituting the remainder except for the 1.5% constituted by *Coffea liberica*. The purpose of this study is to focus on predicting monthly coffee prices in India by using the historic time series data. The objective of this paper is to fit an Autoregressive Integrated Moving Average model using Box-Jenkins approach. Numerous fields, including agricultural production, animal husbandry and dairy economics, stock price prediction, *etc.* depend heavily on forecasting. To choose the best model, Autoregressive (AR), Moving Average (MA), and Autoregressive Integrated Moving Average (ARIMA) processes was used to select the best model for monthly coffee prices in India. This paper discusses ARIMA (p, d, q) time series analysis and its components, ACF, PACF, Normalized BIC, Box-Ljung Q Statistics, and Residual analysis. According to the best fitted model *i.e.*, ARIMA (0,2,1) monthly coffee prices in India is expected to increase to INR 89.35 kg^{-1} in the month of November 2022. The outcomes are represented numerically and graphically.

Keywords: Autocorrelation function, Box-Jenkins Approach, Coffee Prices, Partial autocorrelation function, Residual Analysis

Introduction

Coffea canephora (syn. *Coffea robusta*, also known as robusta coffee) is a coffee species native to central and western Sub-Saharan Africa. It is a flowering plant in the Rubiaceae family. Though commonly referred to as *Coffea robusta*, the plant is scientifically known as *Coffea canephora*, which has two main varieties, *C. robusta* and *C. nganda* (Dagoon, 2005). Robusta coffee is a type of coffee made from the *Coffea canephora* plant's beans (seeds). It is the world's second most popular coffee, accounting for 43% of global coffee production with arabica constituting the remainder except for the 1.5% constituted by *Coffea liberica* (Anonymous, 2019). It is only second to arabica (from the *Coffea arabica* plant), which accounts for the remaining 57% (or more) of global coffee production. The chemical make-up of coffee beans from *C. robusta* and *C. arabica* differs in several ways (Urgert and Katan, 1996). As compared to *C. arabica* beans, *C. robusta* beans have a lower acidity, more bitterness, often a distinct woody and less fruity flavor. It is used primarily in instant coffee, espresso, and as a filler in ground coffee blends.

The robusta plant yields more than arabica, contains more caffeine (2.7% vs. arabica's 1.5%), and contains less sugar (3-7% vs. arabica's 6-9%) (Anonymous, 2016). Robusta requires far less herbicide and pesticide than arabica because it is less susceptible to pests and disease. Although it is also grown in India, Africa, and Brazil, where it is frequently called conilon, it is primarily grown in Vietnam, where French colonists introduced it in the late 19th century (Horowitz, 2004). With over 40% of global production, Vietnam, which primarily cultivates robusta, has recently surpassed all other countries as the world's top exporter of robusta coffee. It surpasses Brazil's 25% of global production, Indonesia's 13%, India's 5%, and Uganda's 5%. Brazil is still the biggest coffee producer in

Article History

RECEIVED on 29th July 2022 RECEIVED in revised form 19th Septemer 2022 ACCEPTED in final form 21st September 2022

the world, producing one-third of the world's coffee, though 69% of that is *C. arabica* (Anonymous, 2022).

India is one of the top ten coffee-producing countries, accounting for about 3% of global output in 2020. Due to its high quality, Indian coffee is regarded as one of the best in the world and commands a high premium in international markets. Because of its mild aromatic flavor, Arabica coffee has a higher market value than Robusta coffee. Because of its strong flavor, Robusta coffee is primarily used in the preparation of various blends. Robusta is the most commonly produced coffee, accounting for 72% of total production. More than 2 million people in India are directly employed by the industry (Anonymous, 2022). Because coffee is primarily an export commodity for India, domestic demand and consumption have little impact on coffee prices.

The southern part of India produces the majority of coffee. Karnataka is India's largest producer, accounting for roughly 70% of total coffee production. Kerala is the second-largest producer of coffee, but it is far behind, accounting for only about 23% of total output. Tamil Nadu is India's third-largest producer, producing 6% of the country's coffee (Anonymous, 2022). The production in Orissa and the Northeastern areas is lower.

According to the Food and Agriculture Organization, India is the eighth largest volume exporter of coffee (Anonymous, 2022). Seasonality is evident in Indian coffee exports, with exports peaking from March to June. Over 70% of the country's output is exported. Coffee exports were valued at US$ 114.7 million in March 2022, a 22% increase from February 2022. This rapid increase in coffee exports has increased earnings for coffee growers in key states such as Karnataka, Kerala, and Tamil Nadu. The purpose of this study is to focus on predicting monthly coffee prices in India by using the historic time series data. The objective of this paper is to fit an Autoregressive Integrated Moving Average model using Box-Jenkins approach (Box and Jenkins, 1976).

Materials and Methods

The study used secondary sources of information. The study's time series data on robusta coffee price month^{-1} was gathered from the International Coffee Organization. The data was collected for a period of five years between June 2017 and June 2022.

Autoregressive Integrated Moving Average

The forecasting algorithm known as ARIMA, or "Autoregressive Integrated Moving Average," is founded on the notion that past values of a time series can be used to predict future values on their own. A time series is "explained" by an ARIMA model based on its own past values, *i.e.*, its own lags and lagged forecast errors, which allows for the prediction of future values. ARIMA models can be used to model any "non-seasonal" time series that has patterns and is not random white noise. An ARIMA model is defined by three terms: p, d, and q, where p is the order of the AR term. The order of the MA term is denoted by q. The number of differencing required to make the time series stationary is denoted by d. Making the time series stationary is the first step in developing an ARIMA model. Because the term "Auto Regressive" in ARIMA refers to a linear regression model that employs its own lags as predictors. Linear regression model performs best when the predictors are uncorrelated and independent of one another. The most common method is to differentiate it. To put it another way, subtract the previous value from the current value. Depending on the complexity of the series, more than one differencing may be required at times. As a result, the value of d is the smallest number of differencing required to make the series stationary, and d=0 if the time series is already stationary. The right order of differencing is the smallest amount of differencing required to obtain a near-stationary series that roams around a defined mean and the ACF plot quickly reaches zero.

If the autocorrelations are positive for a large number of lags (10 or more), the series requires additional differencing. If, on the other hand, the lag 1 autocorrelation is too negative, the series is most likely over differenced. If you can't decide between two differentiating orders, choose the one with the lowest standard deviation in the differenced series. Differencing is only required if the series is non-stationary. Otherwise, no differencing is required, so d=0. The another method to check differencing is through Augmented Dicky Fuller (ADF) Test. Assis *et al.* (2010) forecasted cocoa bean prices in Malaysia along with other competing models. Nochai and Nochai (2006) also forecasted palm oil prices in Thailand.

Augmented Dicky Fuller (ADF) Test

The ADF test's null hypothesis is that the time series is non-stationary. So, if the p-value of the test is less than the significance level (0.05), the null hypothesis is rejected and the time series is inferred to be stationary. If p value is more than 0.05, we proceed with determining the order of differencing.

The order of the 'Auto Regressive' (AR) term is 'p'. It denotes the number of Y lags to be used as predictors. 'q' represents the order of the 'Moving Average' (MA) term. It is the number of lag forecast errors that should be included in the ARIMA Model.

When Y_t depends only on its own lags, the model is said to be pure auto-regressive (AR only). In other words, Y_t is a function of Y_t's "lags."

$$Y_t=\alpha+\beta_1Y_{t-1}+\beta_2Y_{t-2}+..........+\beta_pY_{t-p}+\varepsilon$$

Where, Y_{t-1} is the lag 1 of the series, β_1 is the coefficient of lag 1 that the model estimate and α is the intercept term estimated by the model.

Likewise, a pure moving average model is one where Y_t depends only on the lagged forecast errors.

$$Y=\alpha+e_t+\phi_1e_{t-1}+\phi_2e_{t-2}+..........+\phi_qe_{t-q}$$

Where, the error terms are the errors of the autoregressive models of the respective lags. The errors e_t and e_{t-1} are the errors from the following equations:

$$Y_t=\beta_1Y_{t-1}+\beta_2Y_{t-2}+..........+\beta_0Y_0+\varepsilon_t$$

$Y_{t-1}=\beta_1Y_{t-2}+\beta_2Y_{t-3}+..........+\beta_0Y_0+\varepsilon_{t-1}$

An ARIMA model is one in which the AR and MA terms are combined, and the time series has been differentiated at least once to make it stationary. Thus, the equation is:

$Y_t=\alpha+\beta_1Y_{t-1}+\beta_2Y_{t-2}+..........+\beta_pY_{t-p}\varepsilon_t+\phi_1e_{t-1}+\phi_2e_{t-2}+.........+\phi_qe_{t-q}$

For the present study, the ARIMA model is divided into four stages.

Identification Stage

A graphical plot of the data is a good starting point for time series analysis. It is useful to recognise the existence of trends. It was determined whether the time series data were stationary or not by performing the stationarity check. A non-stationary time series can frequently be made stationary by taking the first differences in the series, which results in the creation of a new time series with successive differences (Y_t - Y_{t-1}). First differences can be produced if the series is not stationary after the first differences. Second-order differencing refers to this. It is possible to distinguish between second-order differences (Y_t- Y_{t-2}). After performing data differencing, we obtain the value of "d." The next step is to calculate the value of p and q of the model.

Before estimating the model's parameters p and q, the data is examined to determine which model best explains the data. This is accomplished by investigating the sample ACF (Autocorrelation function) and PACF (Partial autocorrelation function) of the differenced series Y_{t-1}. ACF and PACF are both used to identify appropriate models. They are discovered by searching for significant spikes in the ACF and PACF functions. One or more models that appear to provide statistically adequate representations of the available data are tentatively chosen during the identification stage. The model's parameters are then precisely estimated using least squares.

Estimating the Parameters

Obtaining least square estimates of the parameters, such as R^2, Root mean square error (RMSE), Mean absolute percentage error (MAPE), Mean absolute error (MAE), and normalized Bayesian Information Criterion (BIC), is the next step after making a tentative determination of the most appropriate model. These estimates are used to evaluate the model's accuracy.

Diagnostic Checking

After estimating the parameters of a tentatively identified ARIMA model, diagnostic checking is required to ensure that the model is adequate. To do so, we must examine the ACF and PACF of residuals, which may indicate the model's adequacy or inadequacy. If it has random residuals, it means that the model that was tentatively identified is adequate. When an inadequacy is detected, the checks should indicate how the model should be modified, followed by more fitting and checking. When all of their ACF were within the limits, the residuals of ACF and PACF were considered random.

Ljung-Box Statistic: The Ljung-Box test is used to examine residual autocorrelations. The residuals shouldn't be correlated, or if they are, the correlation should be minimal. Ljung-Box statistics are used in this instance to test the null hypothesis.

Forecasting

After determining that the fitted model is adequate, it can be used to forecast future values. This was accomplished with the R-Programming Software.

Results and Discussion

After determining that the variable under forecasting was a stationary series, the ARIMA model was developed. Non-stationary data were used as shown in figure 1. Non-stationarity in mean was corrected once more by first order differencing the data. Stationarity could now be tested on the newly constructed variable Y_{t-1}. After first differencing, the series was found to be nonstationary. Then second-order differencing was carried out Y_{t-2} and then the series was found to be stationary. Figure 2 shows the second order differenced time plot of Coffee (Robusta) prices in India. Y_{t-2} was stationary in mean. The next step was to determine the p and q values. Figure 3 plots shows the autocorrelation and partial autocorrelation function of original time series data. The autocorrelation and partial autocorrelation coefficients (ACF and PACF) of various orders of Y_{t-2} were computed for this purpose and are shown in table 1 and figure 4.

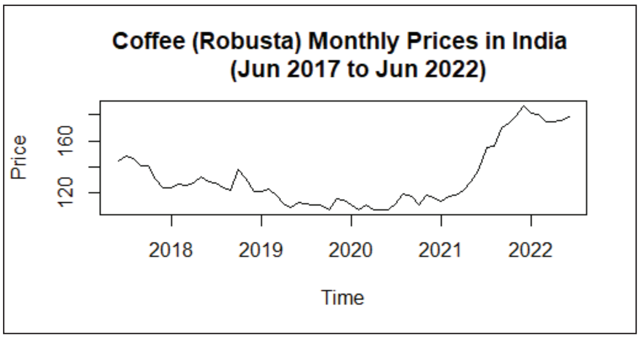

Figure 1: Time plot of Coffee (Robusta) prices in India

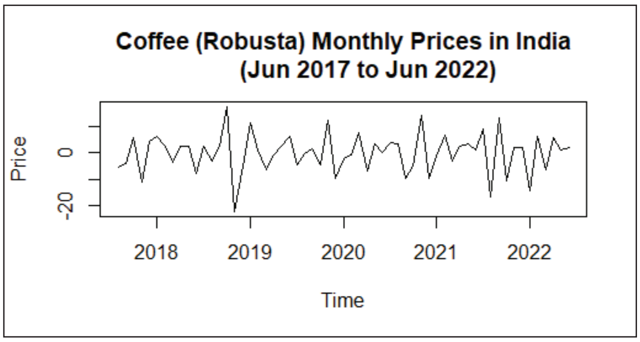

Figure 2: Second Differenced Time plot of Coffee (Robusta) prices in India

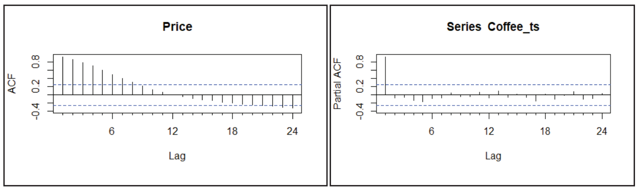

Figure 3: ACF and PACF of original data

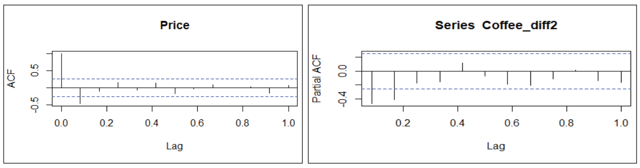

Figure 4: ACF and PACF of second order differenced data

It can be theoretically and visually verified using the Dickey-Fuller unit root test. The stationarity requirement is satisfied at the second order difference with the P_r (|t|>-6.2228) < 0.01, indicating that there is no unit root at the second order difference of coffee (robusta) monthly prices in India at the 5% significance level, according to the results of the Dickey-Fuller unit root test.

The preliminary ARIMA models are discussed with values differenced once (d=2), and the model with the lowest normalized AIC is chosen. Table 1 lists the various ARIMA models and their corresponding normalized AIC values. The chosen ARIMA's normalized AIC value was 379.1261.

Table 1: AIC values of ARIMA (p,d,q)

ARIMA models	AIC values
(0,2,0)	408.3242
(1,2,1)	381.2419
(0,2,2)	381.2398
(1,2,0)	395.9318
(1,2,2)	382.7503
(0,2,1)	379.1261

Model Estimation

Table 2 determines the estimated ARIMA model and their value of significance. Table 3 determines the model fit statistics of the fitted ARIMA model (0,2,1). The mean absolute percent error (MAPE) of the fitted model is estimated as 3.1990 which indicates the model is good fit. The fitted models accurately forecast 96.80% of monthly coffee (robusta) prices in India, according to the mean absolute percentage error (MAPE).

Table 2: Estimated ARIMA model

	Estimate	SE	t value	Significance
MA1	-0.8958	0.0723	-12.387	< 2.2e^{-16***}

Table 3: Fitted ARIMA model fit statistics

Fit Statistic	Mean	Fit Statistic	Mean
ME	0.4378	MPE	0.3513
RMSE	5.6286	MAPE	3.1990
MAE	4.2169	MASE	0.1951

Diagnostic Checking

Again, we should look into whether the forecast errors appear to be correlated, as well as whether they are normally distributed with a mean of zero and a constant variance. We can use a correlogram and the Ljung-Box test to look for correlations between successive forecast errors. The ACF and PACF of residuals of various orders were estimated. Figure 5 depicts the ACF and PACF of the residuals. Various autocorrelations up to lag 12 were computed for this purpose, and their significance was tested using the Box-Ljung statistic. Given that none of the sample autocorrelations for lags 1-12 exceed the significance bounds, we can conclude that there is very little evidence for non-zero autocorrelations in forecast errors at lags 1-12. This demonstrated that the chosen ARIMA model was suitable for forecasting monthly coffee (robusta) prices in India. Furthermore, the p-value for the Ljung-Box test is 0.4596, indicating that there is little evidence for non-zero autocorrelations in the forecast errors for lags 1-12.

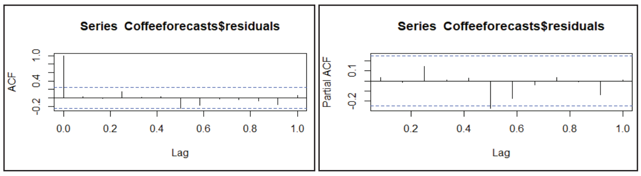

Figure 5: ACF and PACF of the residuals

To check whether the forecast errors are normally distributed with mean zero and constant variance, we make a time plot of the forecast errors, and a histogram. Figure 6 shows the histogram plot of forecasted errors. The variance of the forecast errors appears to be roughly constant over time, as shown by the time plot of the forecast errors. The time series' histogram reveals that the forecast errors are roughly normally distributed and that the mean appears to be very low. It is conceivable that the forecast errors have normal distributions with a mean of zero and constant variance. Because successive forecast errors do not appear to be correlated, and forecast errors appear to be normally distributed with mean zero and constant variance, the ARIMA (0,2,1) model appears to be an adequate predictive model for monthly coffee (robusta) prices in India.

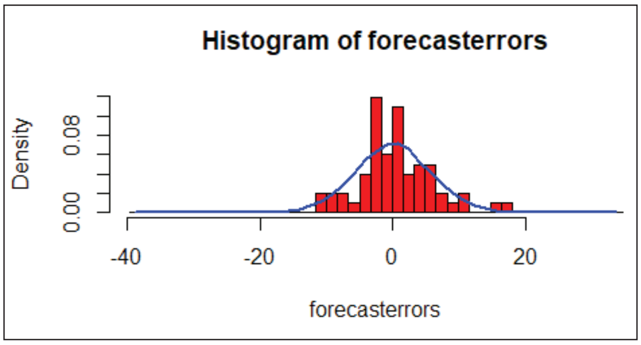

Figure 6: Histogram of Forecast Errors

Forecasting

Based on the model fitted, forecasted prices of monthly coffee (robusta) for the leading 5 months were estimated from July 2022 to November 2022. Figure 7 shows the actual and forecasted value of prices of monthly coffee (robusta) (with 95% confidence limit) in India. It is clear from the forecasted series that the monthly coffee (robusta) prices show an increasing trend.

Based on the fitted model, forecasting prices for monthly

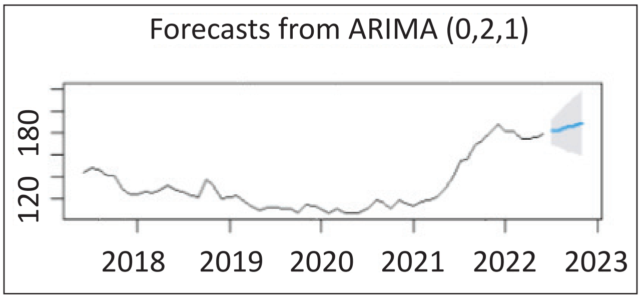

Figure 7: Observed and forecasted prices of monthly coffee (robusta)

coffee (robusta) with upper limit and lower limit are shown in table 4. It has been found that the monthly prices of coffee (robusta) will be around Rs. 189.35 kg^{-1} with lower limit of Rs. 158.55 kg^{-1} and upper limit of Rs. 220.14 kg^{-1} in the month of November 2022.

Table 4: Forecasted prices for monthly coffee (robusta) (INR kg^{-1})

Month	Forecasted Prices	Lower Limit	Upper Limit
July 2022	180.91	169.59	192.22
August 2022	183.02	166.16	199.87
September 2022	185.13	163.43	206.83
October 2022	187.24	160.94	213.53
November 2022	189.35	158.55	220.14

Conclusion

It has been found that there is an increasing trend in the monthly coffee (robusta) prices in India. ARIMA (0,2,1) model quite satisfactorily captured the variation present in the data set. From the forecast available from the fitted ARIMA model, it can be found that forecasted monthly prices of coffee (robusta) will increase form Rs. 180.91 kg^{-1} in July 2022 to Rs. 189.35 kg^{-1} in November 2022. That is, using time series data from June 2017 to June 2022 of monthly coffee (robusta) prices, this study provides an evidence on future prices in India, which can be considered for future policy making and formulating strategies for augmenting and sustaining coffee (robusta) prices. The model demonstrated a good performance in terms of explained variability and predicting power. The findings of the present study provided direct support for the potential use of accurate forecasts in decision making of monthly coffee (robusta) prices in India.

References

Anonymous, 2016. Understanding the Difference: Arabica vs. Robusta. The Coffee Barrister. Available at: https://www.coffeeb.net/arabica-vs-robusta. Retrieved on: 2nd August 2016.

Anonymous, 2019. Coffee: World Markets and Trade (PDF). United States Department of Agriculture - Foreign Agricultural Service. December 2019. Available at: https://downloads.usda.library.cornell.edu/usdaesmis/files/m900nt40f/6m3129089/r494w654j/coffee.pdf. Retrieved on: 8th May, 2020.

Anonymous, 2022. Coffee Industry and Exports. Available at: https://www.ibef.org/exports/coffee-industry-in-india. Retrieved on: 10th July, 2022.

Assis, K., Amran, A., Remali, Y., Affendy, H., 2010. A comparison of univariate time series methods for forecasting cocoa bean prices. *Trends Agric. Econ* 3(4), 207-215. DOI: 10.3923/tae.2010.207.215.

Box, G.E.P., Jenkins, G.M., 1976. *Time Series Analysis: Forecasting and Control*, Revised Ed., Holden-Day, San Francisco, CA. ISBN: 978-0-8162-1104-3. p. 567.

Dagoon, J., 2005. Agriculture & Fishery Technology IV. Rex Bookstore, Inc., Manila, Phillippines, ISBN: 9789712342233. p. 58.

Horowitz, A.R., 2004. Insect pest management: field and protected crops. Springer. ISBN: 978-3-540-20755-9, p. 41.

Nochai, R., Nochai, T., 2006. ARIMA model for forecasting Oil Palm Price. Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications. Universiti Sains Malaysia, Penang, June 13-15, pp. 1-7.

Urgert, R., Katan, M.B., 1996. The cholesterol raising factor from coffee beans. *Journal of the Royal Society of Medicine* 89(11), 618-623.