



## Indian natural rubber price forecast—An Autoregressive Integrated Moving Average (ARIMA) approach

SHYJU MATHEW<sup>1</sup> and RAMASAMY MURUGESAN<sup>2</sup>

National Institute of Technology, Tiruchirappalli, Tamil Nadu 620 015, India

Received: 05 April 2019; Accepted: 18 July 2019

### ABSTRACT

The objective of this study was to forecast the price of natural rubber in India during April 2019 to March 2020 by employing autoregressive integrated moving average (ARIMA). The monthly pricing data for the period from April 2008 to March 2018 was used for the study. The analysis was carried out during the year 2018–19. RSS4 (Ribbed Smoked Sheets), latex (60% DRC (Dry Rubber Content)) and ISNR 20 (Indian Standard Natural Rubber) are the different types of Indian natural rubber that are competitive in international rubber market. The prices of these types of natural rubber were taken for modelling. AIC was used as a selection criterion for the best-fitted model. ARIMA(3,1,2) for RSS 4, ARIMA (3,1,2) for Latex 60% DRC, and ARIMA (4,1,3) for ISNR20 were the most suited models to forecast the price. The evaluation metrics were  $R^2$ , Adjusted  $R^2$ , Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). These were employed for validating the forecasting model. The price forecasting of natural rubber in India can be a better-suited tool for the policymakers to decide on their investment in natural rubber cultivation.

**Key words:** ARIMA, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Natural rubber, Rubber price forecast

Forecasting of price is a crucial part of the trading of a commodity. Before the economic reforms of 1991, the government controlled the price mechanism. At present, the price is determined by the forces of both the domestic and international markets. This prompts to accelerate price volatility making it crucial to examine the trends in price by using accurate statistical modelling tools which like this will help the policy makers to formulate policies (Yashavanth *et al.* 2017) and cultivators to strategize the production and marketing (Jha and Sinha 2013).

It is evident that agricultural economists have done various researches to forecast many particular agriculture commodity prices. However, there are not many studies which focus on predicting natural rubber prices. Romprasert (2009) conducted different forecasting models, which included regression analysis, exponential smoothing, Holt's linear exponential, and Box-Jenkins, to study the futures price of the Thailand natural rubber ribbed smoked sheets (RSS3). Khin (2011) used an autoregressive integrated moving average (ARIMA) and multivariate autoregressive moving average (MARMA) models to forecast the prices of

SMR20 over the period January 1990 to December 2008. Khin and Thambiah (2015) carried out a study to predict the natural rubber price using simultaneous supply-demand and price model equation and VECM model (vector error correction model).

Natural rubber is one of the major plantation crops in India. India holds the sixth position in the production of global natural rubber. During 2015–16, the country produced 562,000 tonnes of natural rubber. Among the state wise production, Kerala generated 78% of its total production. Among the various forms of rubber, solid block rubber contributes the significant portion (75.1%), and RSS Grades provide 23% of natural rubber. (The Statistics and Planning Department 2016).

The integration of international and domestic market brought instability in the price due to the advent of new economic policies in 1991 (Mohankumar 2008). This adversely affected the supply of natural rubber (Mohankumar and Chandy 2005). After the post-WTO period, the volatility in the price of natural rubber is rising every year (Raju 2016). Since the cultivators who invested in the initial period shifted to other crops in the midway of the livelihood, they incurred massive economic losses. So the forecasting of price helps the cultivators to decide on their production by the anticipated prices. It also helps the policy makers, industrialists and traders to take appropriate decisions in marketing and trade. Keeping all above points in view, the

<sup>1</sup>PhD Scholar (vettikunnelachen@gmail.com), <sup>2</sup>Professor (rmurugan@nitt.edu), Department of Humanities and Social Sciences, National Institute of Technology, Tiruchirappalli.

present study was conducted to forecast the price of natural rubber in India by employing autoregressive integrated moving average (ARIMA).

MATERIALS AND METHODS

Secondary data was used to forecast the price of Indian natural rubber. The data was collected from the publications of The Rubber Board, Ministry of Commerce and Industry, Government of India for the period from April 2008 to March 2018.

ARIMA is one of the widely used models for forecasting the time series variables. In this process, the time-dependent variables are assumed to be linear based on past values and random shocks. In general, in the notation ARIMA (p, d, q), ‘p’ denotes the orders of Auto Regression (AR), ‘d’ denotes the order of Integration (differencing) and ‘q’ denotes the order of Moving Average (MA).

The general model for ARMA (p, q) is shown in equation 1

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \tag{1}$$

Where ‘μ’ represents the drift, and ‘ε<sub>t</sub>’ represents the error term. Before modelling for ARIMA, checking and transforming the data into stationarity is an important process. For testing the stationarity, one of the simplest techniques is visualisation, i.e., plotting the data against time, based on the visual inspection (nature of line) one can identify the stationarity. The most common technique is correlogram which consists of Autocorrelation (AC), Partial Autocorrelation (PAC) and Ljung-Box or Q-stat is used. Similarly, the most popular model-based approach is unit root test (Dickey-Fuller Test and Augmented Dickey-Fuller Test) which are also used for the stationarity check. In general, the Dickey-Fuller Test is based on three models. The first model is a random walk without drift and trend. The second model is a random walk with drift, and the final model is a random walk with drift and trend.

The most suitable ARIMA model was selected using the smallest Akaike Information Criterion (AIC) or Schwarz Criterion (SC) value (Makridakis *et al.* 2003) or Hannan-Quinn Criterion values (HQC). The AIC, SC and HQC formula is given below:

$$AIC = -2l / n + 2k / n \tag{2}$$

$$SC = -2l / n + (k \log n) / n \tag{3}$$

$$HQC = -2 \left( \frac{l}{n} \right) + 2k \log(\log(n)) / n \tag{4}$$

The performance measures used to validate models were R<sup>2</sup>, Adjusted R<sup>2</sup>, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). R<sup>2</sup> is otherwise known as the coefficient of determination which lies between 0 to 1, where 0 denotes that model didn’t capture any variance of the target variable. Similarly 1 denotes that the model captures all the variance of the target variable. R<sup>2</sup> and Adj R<sup>2</sup> higher values show

a better model.

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y}_t)^2} \tag{5}$$

$$Adj R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - P - 1} \tag{6}$$

where ‘P’ is the number of predictors.

MAE, RMSE and MAPE are the minimising objective function. The formula for calculating this evaluation metrics are as follows:

$$RMSE: \sqrt{\sum_{t=1}^n (\hat{y}_t - y_t) / n} \tag{7}$$

$$MAE: \sum_{t=1}^n |\hat{y}_t - y_t| / n \tag{8}$$

$$MAPE: 100 * \left( \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right| / n \right) \tag{9}$$

Where ‘y<sub>t</sub>’ is the actual stationary data,  $\hat{y}_t$  is the predicted value, t = {1, 2, ..., n}, n is the total number of observations.

RESULTS AND DISCUSSION

The most common methods were used for the stationarity check. They were ACF, PACF and LB stat. ACF and PACF are computed for 36 lags. The significance was tested through LB-stat. In general, a stationary data has no trend; its variance is constant (i.e., the variations around its mean). Before entering into the statistical test, the common graphical method was applied for visual interpretation of the data. The graphical representation of zero difference and First difference data is shown in the below Fig 1. The x-axis consists of the time (in monthly), and y-axis comprises of the price (₹).

This graphical representation clearly shows the zero difference data, the variance is not constant (i.e., the fluctuation of data is not in constant form) but in the first difference data, the variance of the data is constant. The comparison of LB-Stat value for zero difference and first difference is listed in Table 1.

To test the significance, the null hypothesis (H<sub>0</sub>) is framed as autocorrelations up to k lags are equal to zero (i.e., the values are random and independent up to a certain number of lags) and the alternate hypothesis (H<sub>1</sub>) is the autocorrelations upto lag k not equal to zero (i.e., the data values are not random, and the values are dependent up to a certain number of lags). Table 1 shows that for the zero difference data, the LB-stat values are higher than the

Table 1 Comparison of LB-stat Values

| Segments        | Zero Difference | First Difference |
|-----------------|-----------------|------------------|
| ISNR 20         | 1024.47         | 27.57            |
| Latex (60% DRC) | 668.33          | 25.714           |
| RSS 4           | 985.29          | 31.113           |



Fig 1 Graphical representation of price of zero difference and first difference.

table values. This results in the acceptance of the alternate hypothesis ( $H_1$ ). Similarly, for the first difference data, the values are lesser than the table values. This results in the acceptance of the null hypothesis. The data is random and independent up to 36 lags.

Another popular method for finding stationary is Unit root test (i.e. Augmented Dickey-Fuller Test). The comparison of  $\tau$ -values for zero difference and first difference is listed in Table 2.

Table 2 Comparison of unit root test

|                | Zero difference |        | First difference |      |
|----------------|-----------------|--------|------------------|------|
|                | t-Statistic     | Prob   | t-Statistic      | Prob |
| ISNR 20        | -1.5859         | 0.4862 | <b>-6.2629</b>   | 0    |
| Latex(60% drc) | -2.5957         | 0.097  | <b>-7.7721</b>   | 0    |
| RSS 4          | -1.7569         | 0.3999 | <b>-6.5197</b>   | 0    |

For 1% level, the table value is -3.493, 5% level, the table value is -2.889 and 10% level, the table value is -2.58. Here, the calculated values are greater than the table value. This shows that the acceptance of an alternate hypothesis, i.e., all the data are stationary in the first difference. In the first difference, the  $\tau$ -values values are significant compared to zero difference. This shows that the first difference data became stationary.

Based on the AIC value, the best-fitted model is chosen for the forecasting process. For ISNR 20 the best-fitted model is selected as ARIMA(4,1,3). For Latex (60% DRC) the ARIMA(3,1,2) is selected as the best-fitted model. Similarly, for RSS4, the ARIMA(3,1,2) is selected as the most suited model.

Table 3 is the comparison of the performance measure such as  $R^2$ ,  $AdjR^2$ , AIC, SC, HQC, RMSE, MAE and MAPE. AIC, SC and HQC are information criteria used to select the appropriate model. The  $R^2$  and  $Adj R^2$  value

Table 3 Comparison of performance measure

|                 | AIC    | (p,d,q) | $R^2$ | $Adj R^2$ | SC       | HQC      | RMSE    | MAE      | MAPE |
|-----------------|--------|---------|-------|-----------|----------|----------|---------|----------|------|
| ISNR 20         | 16.699 | (4,1,3) | 0.948 | 0.9352    | 16.87335 | 16.77    | 956.18  | 733.22   | 5.69 |
| Latex (60% DRC) | 17.527 | (3,1,2) | 0.886 | 0.877     | 17.75056 | 17.61768 | 1147.39 | 1145.051 | 6.76 |
| RSS 4           | 16.673 | (3,1,2) | 0.948 | 0.944     | 16.89727 | 16.76439 | 906.152 | 667.83   | 4.74 |

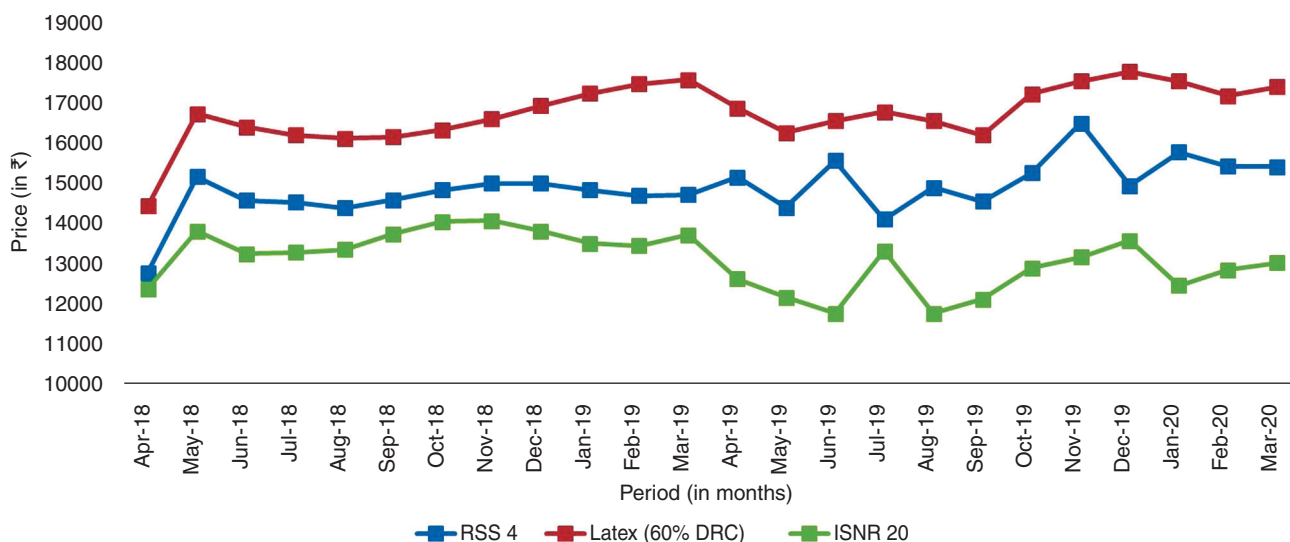


Fig 2 Comparison of forecasting price of RSS 4, Latex (60% DRC) and ISNR 20.

for the best-fitted model is higher. Similarly, the RMSE, MAE and MAPE are minimum for the most suited model.

Fig 2 is the graphical representation of forecasting the price of Indian natural rubber from April 2018 to March 2020. The above forecasting may help the farmers in the investment decisions of maintaining rubber trees, scheduling the days for tapping and even shifting the pattern of cropping. This may also help the industrialists to take appropriate policy measures.

*Policy implications:* Even though a newcomer in the plantation sector of India, natural rubber, particularly in the post-independent epoch, developed as a vital commodity with grave ramifications extending from income security to more than a million population to national security. In recent times, among the states, the major portion of natural rubber is produced by Kerala. Now the cropping of natural rubber is gradually spreading to other regions, especially in the North-East. Anyway, the rubber plantation economy of India is currently standing up to a relevant challenge i.e., the price instability, particularly in the neo-liberal scenario. Hence, a dynamic national policy on natural rubber is a critical need for the present situation. The driving force, behind the policy ought to be to strengthen the natural rubber in production, processing and marketing. In this process, the role of Rubber Board of India is very significant. As of now, an arrangement of organisations under Rubber Board is comprised for the advancement and sustenance of natural rubber economy in India. It includes co-operative institutions, Rubber Producer’s Societies (RPSs), Private Limited Companies and Self Help Groups (SHGs). Most of these organisations are coordinating with other organisations like Agricultural Technology Management Agency (ATMA). Based on evaluating the performance, studies are conducted which reveals some as well functioning; others in a moderate way and rest of them are ill-functioning.

Thus, a policy should go for fortifying the institutional framework through sufficient subsidising, programming,

observing and assessing. Among the adversely affecting factors in the Indian natural rubber economy, the most relevant one is price instability. Natural rubber in the domestic market attained ₹ 238 per kg in April 2011 endured a drastic decrease to ₹ 118 per kg in November 2014 for RSS 4 graded ribbed smoked sheets. This tremendous price fall is due to massive import of natural rubber. The huge import of natural rubber may be the reflections of neo-liberal trade policy adopted by India, i.e. the removal of quantitative restrictions for the import of 715 items including natural rubber to compromise with the policy of WTO which came into force from April 2001. Therefore a foreign trade policy should be adopted in this dimension too. An internal looking is vital, whether these policies really harm or sustain the rubber economy of India. The occurrence of price instability also can be regulated to some degree by effectively checking the price stabilisation fund scheme. To protect the Indian rubber economy from outside stuns, WTO forum must be legitimately used. For example, the present price fall of Indian natural rubber is because of dumping rehearsed by countries like Sri Lanka, Vietnam and Indonesia. This could be stood up to through the mediation of WTO.

Forecasting the price of Indian natural rubber is important for the industrialists for decision or policy-making process and farmers for cultivation process. In this paper, the ARIMA (p,d,q) is used for forecasting the price of Indian natural rubber. For constructing the ARIMA model, the first process is to check whether the data is stationary or not. For this testing process, the Autocorrelation, Partial Autocorrelation, LB-stat and unit root test was used. Three segments were considered for this study. The first segment was RSS4, second was Latex (60% DRC), and the third segment was ISNR 20. For all the three segments, first differenced data became stationary. By using this first difference data, the ARIMA model was constructed. The best model was chosen based on the information criteria. The most suited ARIMA models for RSS4, Latex (60%

DRC) and ISNR 20 are ARIMA(4,1,3), ARIMA(3,1,2), and ARIMA(3,1,2).

#### REFERENCES

- Jha G K and Sinha K. 2013. Agricultural price forecasting using neural network model: An innovative information delivery system. *Agricultural Economics Research Review* **26**(2): 229–39.
- Khin AA. 2011. *Econometrics forecasting models for short term natural rubber prices*. Lambert Academic Publishing. Saarbrücken, Germany.
- Khin AA and Thambiah S. 2015. Natural rubber prices forecasting using simultaneous supply-demand and price system equation and VECM model: Between theory and reality. (In) *Proceedings of 2nd International Conference on Agriculture and Forestry, ICOAF-2015*, Colombo, Sri Lanka, June 10-12.
- Makridakis S, Wheelright S C and Hyndman R J. 2003. *Forecasting: Methods and Applications*. Wiley. New Delhi, India.
- Mohanakumar S and Binny C. 2005. Impact of uncertainty on investment and employment in rubber smallholdings sector: An analysis of trends in the post-reform phase. *Economic and Political weekly* **40**(46): 4850–6.
- Mohanakumar S. 2008. Sources of instability in natural rubber price: An analysis in the post-reform phase, *Journal of Plantation Crops* **36**(3): 517–23.
- Raju K V. 2016. Instability in Natural Rubber Prices in India: An Empirical Analysis, *IOSR Journal of Economics and Finance* **7**(3): 24–8.
- Romprasert S. 2009. Forecasting model of RSS3 price in futures market. *Kasetsart University Journal of Economics* **16**(1): 54–74.
- The Statistics and Planning Department. 2017. *Indian Rubber Statistics 2017*, Vol. 38. The Rubber Board. Kottayam, India.
- Yashavanth B S, Singh K N, Paul A K and Paul R K. 2017. Forecasting prices of coffee seeds using vector autoregressive time series model. *Indian Journal of Agricultural Sciences* **87**(6): 754–8.