

Application of SARIMA model to forecasting the natural rubber price in the world market

Nhien T. Pham

Department of Economics, Nong Lam University, Ho Chi Minh City, Vietnam

ARTICLE INFO

Research Paper

Received: November 02, 2018

Revised: November 23, 2018

Accepted: December 09, 2018

Keywords

Natural rubber price
Rubber
Rubber market
Rubber price forecast
SARIMA

Corresponding author

Pham Thi Nhen

Email: ptnhien@hcmuaf.edu.vn

ABSTRACT

This study was conducted to develop a forecasting model to predict the price natural rubber in the world market by using the Seasonal Autoregressive Integrated Moving Average (SARIMA). The dataset for model development was collected from series data of average monthly closing average prices in the natural rubber - Ribbed Smoked Sheet No.3 (RSS3) on the Tokyo Commodity Exchange (TOCOM) for the period of January 2007 - September 2018. The RSS3 price on the TOCOM provided the reference price for natural rubber in the world market. It resulted SARIMA(2,1,2)(1,1,1)₁₂ model was selected as the best-fit model. The model achieved 0.000 for Probability value (*P-value*); 8.86 for Akaike Information Criterion (AIC) and 9.01 for Schwarz Information Criterion (SIC); 6.68% for Mean Absolute Percentage Error (MAPE) and 21.43 for Root Mean Square Error (RMSE). This model was used to forecast the world's natural rubber price during October 2018 - December 2020. This study may be helpful to the farmers, traders, and the governments of the world's important natural rubber producing countries to plan policies to reduce natural rubber production costs and stabilize the natural rubber price in the future, such as by setting suitable areas for natural rubber plantation in each country, and defining appropriate and sustainable alternative crop areas in each country.

Cited as: Pham, N. T. (2018). Application of SARIMA model to forecasting the natural rubber price in the world market. *The Journal of Agriculture and Development* 17(6), 1-7.

1. Introduction

Natural rubber (NR) is a major economic plant in the plantation sector in terms of providing income to the growers for a long-time and also it serves as raw materials for various industrial products. Over 20 million families are dependent on rubber cultivation for their livelihood in the world natural rubber market. Rubber has been one of the most essential economic plants in the world in past 10 years. According to the statistics of Association of Natural Rubber Producing Countries (ANRPC), the total global production of natural rubber increases from 10.06 million metric tons in 2007 to 13.54 million metric tons in 2017. In 2017, the largest rubber producing countries are Thailand, Indonesia, Malaysia,

Vietnam, India and China which collectively account for 90% of world production. Rubber is primarily exported to China, Germany, United States and Iran to be processed to produce vehicle tires, belts, shoe soles, medical gloves and parts for electronic equipment. Approximately 70% of primary processed natural rubber is used to produce automobile tires (ANRPC, 2017).

Furthermore, farmers have made a huge investment in the initial period (5 years), switching to other crops in the middle of the economic life of their plantation would involve huge losses. Therefore, price forecasting of natural rubber is significant to help the farmers to decide upon their production by the expected prices. This results in the requirement for statistical techniques to provide accurate and timely price forecast by tak-

ing into account the information to the farmers, traders and policymakers so that they may make production, marketing and policy decisions well in advance.

There are a few studies that used Autoregressive Integrated Moving Average (ARIMA) model to forecast natural rubber prices. By using monthly data, Rani & Krishnan (2018) showed that the model ARIMA(4,1,4) was found to be the best model to predict of prices of natural rubber in India in 2017. The forecast result was suitable for explaining the fluctuation of natural rubber prices in 2017. The other study was conducted to forecast the prices of natural rubber of Thailand. It was found that the ARIMA(1,0,1) model was the most suitable that could be explained for variations of the natural rubber price of Thailand. (Cherdchoongam & Rungreungnanun, 2016). The ARIMA forecasting model was developed to predict Malaysian natural rubber prices in 2014. An illustration for real-time forecasts for natural rubber prices in the Malaysians showed its ease of use. The forecast result of natural rubber prices was suitable (Khin & Thambiah, 2014).

The ARIMA forecasting model would not be really effective for seasonal time data. To develop a forecasting model consistent with the data series available seasonality, the ARIMA model is expanded by adding autoregressive and moving averages to the number of seasons as known as Seasonal Autoregressive Integrated Moving Average (SARIMA). By using the SARIMA(1,1,1)(1,0,0)₁₂ model, the study demonstrated the volatility of seasonal potato prices in Delhi and found that the model had the most reasonable forecast results. The forecast result of potato prices was suitable (Chandran & Pandey, 2007). Similarly, Adanacioglu & Yercan conducted a forecast of tomato prices in the Antalya city in Turkey. Their study showed that the SARIMA(1,0,0)(1,1,1)₁₂ model was the most appropriate for the monthly wholesale price series. The forecast result of tomato prices was roughly equal to the real ones (Adanacioglu & Yercan, 2012).

There is hardly any literature available on forecasting the prices of natural rubber by using Seasonal Autoregressive Integrated Moving Average (SARIMA) model. This study was designed to develop a Seasonal ARIMA temporal model using long-term historical prices in the world's natural

rubber market. This model is selected because of the capability to correct the local trend in data, where the pattern in the previous period can be used to forecast the future. Thus, this model also supports the modeling of one perspective as a function of time. Due to the seasonal trend of time series used, the SARIMA is selected for the model development.

2. Material and Methods

The dataset for model development was collected from time series data of Ribbed Smoked Sheet No.3 (RSS3) prices on the Tokyo Commodity Exchange (TOCOM) for the period of January 2007 - September 2018 because it is Japan's largest commodity exchange and one of the largest markets in the world for the buying and selling of natural rubber. TOCOM has offered RSS3 rubber contracts for almost 66 years and it has the largest trading volumes for rubber futures. The RSS3 price on the TOCOM provides the reference price for natural rubber on the world market. The original dataset is plotted as presented in Figure 1.

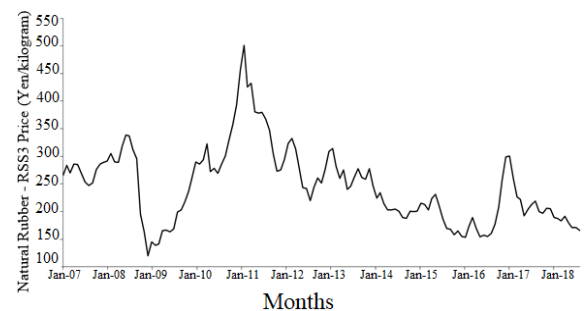


Figure 1. Natural rubber - RSS3 prices from January 2007 to September 2018 on TOCOM (TOCOM, 2018).

Since time series plot of the historical data exhibited the seasonal variations which present similar trend every year, then SARIMA was chosen as the appropriate approach to develop a model prediction.

One of time series models which is popular and mostly used is Box - Jenkins ARIMA model. Seasonal ARIMA is based on the theory of ARIMA. ARIMA models employ a combination of linear operators for the representation of a time series. The general class of ARIMA(p,d,q) comes from three parts: d is the level of differencing, p is the

autoregressive order, and q is the moving average order.

We have a time series, $Z_t, t = 1, 2, \dots, n$ (here n is 129), the first, an autoregressive-moving average ARMA(p,q) model has the form:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} \\ = C + U_t - \theta_1 U_{t-1} - \theta_2 U_{t-2} - \dots - \theta_q U_{t-q}$$

Or

$$Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \dots - \phi_p Z_{t-p} = \\ C + U_t - \theta_1 U_{t-1} - \theta_2 U_{t-2} - \dots - \theta_q U_{t-q} \quad (1)$$

Where:

The constant is notated by C, while ϕ is an autoregressive operator, U is a random shock corresponding to time period t, and θ is a moving average operator.

If we introduce the backshift operator, B, where:

$$BZ_t = Z_{t-1}; B^2 Z_t = B(BZ_t) = Z_{t-2}$$

And so on (1) can be rewritten as:

$$Z_t - \phi_1 BZ_t - \phi_2 B^2 Z_t - \dots - \phi_p B^p Z_t = \\ C + U_t - \theta_1 B U_t - \theta_2 B^2 U_t - \dots - \theta_q B^q U_t \quad (2)$$

Or

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Z_t = \\ C + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) U_t \quad (3)$$

Because the series is not stationary (i.e., has no fixed mean level), then the autoregressive portion of the ARMA(p,q) model must include a stationary inducing operator. For a non-seasonal series, this is most frequently accomplished through a differencing operator (or product of differencing operators) of the form (1-B). That is, instead of modeling the nonstationary series Z_t , we model the series

$$(1 - B)Z_t = Z_t - Z_{t-1}$$

Physically this corresponds to modeling the change in the series rather than the series itself. Usually only a single differencing operator is required. On rare occasions in the modeling of non-seasonal series, the operator may need to be repeated, say d times. The model we then consider

is an autoregressive-integrated moving average or ARIMA(p,d,q) model of the form

$$(1 - \phi_1 B - \phi_1 B^2 - \dots - \phi_1 B^p)(1 - B)^d Z_t = \\ C + (1 - \theta_1 B - \theta_1 B^2 - \dots - \theta_1 B^q) U_t \quad (4)$$

Some situations were encountered in which a time series exhibits some periodic or seasonal pattern. For example, data recorded monthly may exhibit “similar” behavior from year to year; that is, a seasonality of period 12. Data recorded quarterly may have 4 as its seasonality, and data recorded hourly may have 24 as its periodicity. In such situations, seasonal ARIMA models need to be employed to account for any seasonal pattern present in the series.

Multiplicative seasonal ARIMA models are often described as SARIMA(p,d,q)(P,D,Q)_s models, where s is the seasonality, and P, D and Q refer to the orders of the seasonal AR, seasonal differencing and seasonal MA parts of the model, and s is the length of the seasonal period (s = 12).

This multiplicative seasonal model can be expressed as:

$$(1 - \phi_1 B - \phi_1 B^2 - \dots - \phi_1 B^p)(1 - \phi_1 B^s - \\ \phi_1 B^{2s} - \dots - \phi_1 B^{ps})(1 - B)^d (1 - B^s)^D Z_t \\ = C + (1 - \theta_1 B - \theta_1 B^2 - \dots - \theta_1 B^q)(1 - \\ \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_1 B^{qs}) U_t \quad (5)$$

Where

ϕ is a seasonal autoregressive operator and Θ is a seasonal moving average operator.

A SARIMA (p,d,q)(P,D,Q)₁₂ model was constructed using monthly natural rubber price data from January 2007 to September 2018 and a forecast of natural rubber prices from October 2018 to September 2020, following the four steps below:

- Step 1: Identification of model
This step focus on selection of the order of regular differencing (d), seasonal differencing (D), the non-seasonal order of Autoregressive (p), the seasonal order of Autoregressive (P), the non-seasonal order of Moving Average (q) and the seasonal order of Moving Average (Q). The number of order can be identified by observing the autocorrelations function (ACF) and partial autocorrelations function (PACF).

- Step 2: Estimation of parameters
The historical data is used to estimate the parameters of the tentatively model in Step 1.
- Step 3: Diagnostic checking
Diagnostic test is used to check the adequacy of the tentatively model.
- Step 4: Forecasting
The final model in Step 3 is used to forecast the forecast values (Box et al., 2008).

3. Results and Discussion

As it was earlier stated that development of SARIMA model for any variable involves four steps namely identification, estimation of parameters, diagnostic checking and forecasting. Each of these steps is explained for natural rubber prices.

3.1. SARIMA model identification

SARIMA model is estimated only after transforming the variable under forecasting into a stationary series. Stationary series is the one whose values vary over time only around a constant mean and a constant variance. A popular formal method of determining stationarity is the Augmented Dickey Fuller (ADF) test. The estimates of necessary parameters and related statistics for the time series of natural rubber prices without differencing and after first differencing are presented in Table 1.

The analysis exposed that the hypothesis of random walk that underlying process of generating the time series is nonstationary cannot be rejected without differencing, as the ADF test statistics is less than the critical value at 1% and 5% level. The ADF tests for the differenced time series of natural rubber prices revealed that the series were stationary after first difference.

The ACF and PACF values of the series from which the seasonal differences are taken are presented in Figure 2. The seasonal spikes at ACF and PACF after 1 lag are observed as being cut off after taking the seasonal difference of the series. This suggests that the seasonal order of Autoregressive equals 1 ($P = 1$), and the seasonal order of Moving Average equals 1 ($Q = 1$). In addition, the discontinuation of ACF and PACF values after 2 lags indicates that the non-seasonal

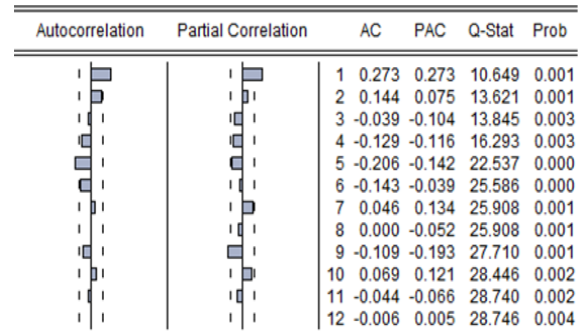


Figure 2. ACF and PACF of natural rubber - RSS3 price at the first difference.

order of Autoregressive and the non-seasonal order of Moving Average can not be greater than 2. Therefore, this study choose the non-seasonal order of Autoregressive equals 2 ($p = 2$) and the non-seasonal order of Moving Average equals 2 ($q = 2$).

So, the tentative specifications were SARIMA(1,1,1)(1,1,1)₁₂, SARIMA(1,1,2)(1,1,1)₁₂, SARIMA(2,1,1)(1,1,1)₁₂, and SARIMA (2,1,2)(1,1,1)₁₂.

In addition, the goodness-of-fit statistics employed were the Adjusted R-squared (R_a^2), Akaike information criterion (AIC), Schwarz information criterion (SIC) and Standard error (SE). R_a^2 must be larger as better, while AIC, SIC and SE must be lower as better.

Criteria were given in Table 2 suggest that SARIMA(2,1,2)(1,1,1)₁₂ model as the most suitable model for forecasting. This model was selected with the lowest AIC, SIC and SE values and the largest R_a^2 value.

3.2. Estimate the parameters of the tentatively model

Parameters of the model can be estimated by using the result of regression SARIMA(2,1,2)(1,1,1)₁₂ model.

The Table 3 shows that the *P-values* associated with the variable coefficients are 0.000. These indicate that all the coefficients of variables in this equation is statistically significant at the level of 1% significance. There are three negative parameters such as AR(2), MA(1) and SAR(1). The positive parameters are AR(1), MA(2) and SMA(1). The particular parameters of the model were gen-

Table 1. ADF tests of natural rubber prices at levels

ADF test statistics	Critical value at 1%	Critical value at 5%	Probability	Level of Integration
-2.835	-3.478	-2.883	0.056	I(0)
-8.8675	-3.477	-2.882	0.000	I(1)

Table 2. Three model selection criteria

Models	R_a^2 (%)	AIC	SIC	SE
SARIMA(1,1,1)(1,1,1) ₁₂	30,31	8.89	9.01	20.19
SARIMA(1,1,2)(1,1,1) ₁₂	30,62	8.90	9.03	20.23
SARIMA(2,1,1)(1,1,1) ₁₂	30,65	8.91	9.03	20.31
SARIMA(2,1,2)(1,1,1) ₁₂	34,95	8.85	9.01	19.75

erated as the following results: AR(1) = 0.969; AR(2) = -0.831; MA(1) = -0.764; MA(2) = 0.834; SAR(1) = -0.591; SMA(1) = 0.946 and Constant = -1.253.

3.3. Diagnostic checking

ACF and PACF of residuals were given in Figure 3 indicated that the autocorrelation values less than to |0.2|. It implies that the residuals of the respective time series are white noise. (i.e. there is no autocorrelation).

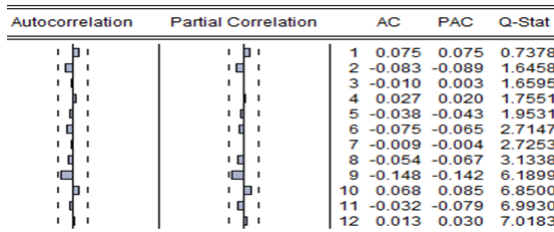


Figure 3. ACF and PACF of residuals of the selected SARIMA(1,2,1)(1,1,1)₁₂ model.

In Table 4, the Breusch-Godfrey Prob. Chi-square statistic is 0.525. This value is greater than 1% presented for the best selected model. It also implies that the residuals of the respective time series are white noise, implying that the model fitness is acceptable.

Moreover, Durbin-Watson statistic of the SARIMA(2,1,2)(1,1,1) equals 1.85 also showed no autocorrelation.

Otherwise, in Box-Jenkins models, the random error component plays a dominant role in determining the structure of the model. The result of heteroskedasticity test (Autoregressive Conditional Heteroscedasticity - ARCH test) with

Prob. Chi-Square statistic is 0.374. This value is greater than 1%, demonstrated that there is no heteroscedasticity. This result was showed in Table 5.

Based on the results of the analysis, the SARIMA(2,1,2)(1,1,1)₁₂ model was selected as the best-fit model for forecasting, since it provides a reasonable fit to the highly seasonal time series data.

3.4. Forecasting

The final SARIMA(2,1,2)(1,1,1)₁₂ model was used to forecast the prices of world natural rubber during October 2018 - December 2020. This model was the most suitable for explaining the fluctuation of natural rubber prices on the world market prices. The result of forecasting was presented in Table 6.

The result of forecast shows that the world natural rubber prices tend to increase slightly from 173.63 yen per kilogram in October 2018 to 219.31 yen per kilogram in July 2019, and have decreasing trend to 167.45 yen per kilogram in May 2020. After that there is a slight increase and reaches 171 yen per kilogram in December 2020. Therefore, the world natural rubber price remain stable about 170 yen per kilogram from October 2018 to December 2020.

3.5. Error measures

The accuracy of the forecasting can be evaluated using error measures. It is achieved by comparing the original data and the forecast values. In this paper, Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) were used as the error measures.

Table 3. Estimated regression results of SARIMA(2,1,2)(1,1,1)₁₂ model

Variables	Coefficients	Standard Error	t Stat	P-value
C	-1.253	2.461	-0.509	0.611
AR(1)	0.969***	0.089	10.808	0.000
AR(2)	-0.831***	0.082	-10.092	0.000
SAR(1)	-0.591***	0.064	-9.117	0.000
MA(1)	-0.764***	0.087	-8.782	0.000
MA(2)	0.834***	0.078	10.665	0.000
SMA(1)	0.946***	0.013	71.209	0.000

***Corresponds to the significance level statistics, 1%.

Table 4. Breusch-Godfrey Serial Correlation LM Test

F-statistic	0.868362	Prob. F(12,109)	0.581
Obs*R-squared	11.04076	Prob. Chi-Square(12)	0.525

The result showed MAPE and RMSE values for the selected model were 6.68% and 21.43 respectively. Thus, the empirical result indicated that the model was able to accurately represent the natural rubber prices historical dataset.

The trend of the predicted natural rubber prices on the world market during October 2018 - December 2020 was showed in Figure 4.

**Figure 4.** The forecasting results of natural rubber prices in the world market during October 2018 - December 2020 (Yen/kg).

This information is necessary and useful for the natural rubber producers and consumers as well as traders and planners for new investment decisions in the natural rubber world market. Especially, the forecasting result may be useful for the farmers because globalization and market integration, there is a huge price fluctuation where farmers can not decide upon their farming practices. Time series forecasts are generated by models based on changes over time in previously observed values or historical datasets. The SARIMA

forecasting model can serve as a useful tool for the farmers to decide upon their production. However, it should be up-dated from time to time with incorporation of current data.

Beside the prices of natural rubber in the past periods, there are other factors having potential influence on natural rubber prices on the world market. The factors comprise of world economic growth, crude oil price, exchange rate (yen/\$), world natural rubber consumption and world synthetic rubber consumption. . . Oil price change by all means affects nearly everything in the world. Its byproduct is also used as one of the main content to produce synthetic rubber. As synthetic rubber is a perfect substitute goods of natural rubber so when oil price increases, synthetic rubber price increases, and then natural rubber price also increases. Conversely, when oil price decreases lead to natural rubber price also reduces. In addition, the devalued yen made rubber futures at TOCOM economically more attractive to the overseas investors so natural rubber price also increases and on the contrary. Therefore, to make the best decision, decision makers should be combine this forecasting result with these factors.

4. Conclusion

In this paper, an efficient technique was presented to accurately predict time series data of world natural rubber prices. Multiplicative seasonal SARIMA models provide an economical way to model time series whose have seasonal tendencies. Based on the results of the analy-

Table 5. Heteroskedasticity test - ARCH

F-statistic	1.077348	Prob. F(12,103)	0.387
Obs*R-squared	12.93619	Prob. Chi-Square(12)	0.374

Table 6. The forecasting results of natural rubber prices in the world market (Yen/kg)

Months	Prediction		
	2018	2019	2020
January	-	195.188	187.980
February	-	200.199	189.484
March	-	197.971	187.716
April	-	193.310	177.972
May	-	193.364	167.449
June	-	205.852	168.569
July	-	219.311	170.756
August	-	213.621	172.769
September	-	206.939	175.484
October	173.628	200.890	176.425
November	181.885	193.306	174.178
December	187.354	188.101	171.096

sis, SARIMA(2,1,2)(1,1,1)₁₂ model was the most suitable model that can be able to represent the historical data. The model attained 0.000 for P value; 8.86 for Akaike Information Criterion; 9.01 for Schwarz Information Criterion; 6.68% for Mean Absolute Percentage Error and 21.43 for Root Mean Square Error. This model can be used to forecast the prices of natural rubber on the world market. The world natural rubber prices have a light increasing trend during the October 2018 to July 2019 and reach 219.31 yen/kg and have decreasing trend to 171.09 yen/kg after that. Forecasting the future prices of natural rubber through the most accurate time series model can help the governments of the world's important natural rubber producing countries as well as consumers and traders to perform better strategic planning and also to help them in maximizing revenue and minimizing the loss.

Acknowledgments

The author would like to express their sincere gratitude to the Tokyo Commodity Exchange for providing time series data of natural rubber prices.

Conflict of interest statement

The author declares that there is no conflict of interest.

References

- Adanacioglu, H., & Yercan, M. (2012). An analysis of tomato prices at wholesale level in Turkey: an application of SARIMA model. *Custos e@ gronegocio on line* 8(4), 52-75.
- ANRPC (The Association of Natural Rubber Producing Countries). (2017). *A monthly bulletin of market trends and statistics published by The Association of Natural Rubber Producing Countries for the period of 2007 - 2017*. Retrieved from July 1, 2018, <http://www.anrpc.org>.
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis: forecasting and control* (4th ed.). New Jersey, USA: John Wiley & Sons.
- Chandran, K. P., & Pandey, N. K. (2007). Potato price forecasting using seasonal ARIMA approach. *Potato Journal* 34(1-2), 137-138.
- Cherdchoongam, S., & Rungreunganun, V. (2016). Forecasting the price of natural rubber in Thailand using the ARIMA model. *King Mongkut's University of Technology North Bangkok Int Journal of Applied Science Technology* 9(4), 271-277.
- Khin, A. A., & Thambiah, S. (2014). Forecasting analysis of price behavior: A case of Malaysian natural rubber market. *American-Eurasian Journal of Agricultural and Environmental Science* 14(11), 1187-1195.
- Rani, V. J., & Krishnan, S. (2018). Forecasting the prices of Indian natural rubber using ARIMA model, *International Journal of Pure and Applied Bioscience* 6(2), 217-221.
- TOCOM (Tokyo Commodity Exchange). (2018). *A monthly series data of Ribbed Smoked Sheet No.3 (RSS3) prices on the Tokyo Commodity Exchange (TOCOM) for the period of January 2007 - September 2018*. Retrieved from October 12, 2018, <https://www.tocom.or.jp/historical/download.html>.