Data Science and Engineering (DSE) Record, Volume 4, issue 1.

Price Forecasting of Ribbed Smoked Sheet in Thailand

Witidtayapond Promana [1] and Sakgasit Ramingwong [2]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand
witidtayapond_p@cmu.ac.th

Abstract. This independent study attempts to find the most appropriate model for forecasting price of Ribbed Smoked Sheet No.3 in Thailand. Daily price from Rubber Authority of Thailand since 2011 to 2021 are used as raw data. A total of 2,618 values of daily rubber price are divided into training and test sets. The training set involves 2,376 values from 2011 to 2020. It is used for constructing four forecasting models i.e. moving average, Holt's method, Box-Jenkins method and Neural network. The test set, including 242 values from 2021, is used for comparing accuracy of the forecast via criteria of the lowest. The finding indicates that the Neural network by nonlinear autoregressive neural network (NNAR) is the most suitable for forecasting price of Ribbed Smoked Sheet No.3. This method has the least Mean absolute error (MAE) of 4.5352, Root mean square error (RMSE) of 5.7807 and Mean absolute percentage error (MAPE) of 7.3309. Respectively, Box-Jenkins method, Moving average and Holt's Method are found to provide less accurate result.

Keywords: Price forecasting, Ribbed smoke sheet, NNAR model

1 Introduction

Para rubber production is an important economic and strategic industry in Thailand. Thailand has been accounting 87% of para rubber product in the world, i.e. approximately 14.56 million tons per year. This generates 500 billion bath on annual revenue for the country. The total rubber plantation area in Thailand involves 23 million rai (36,800 square meters). This industry provides income for more than 1.6 million households of rubber farmers. Almost all of para rubber production are exported. Only 15% of the production are used as raw materials in domestic factories. At the same time, the price of rubber traded in the world market is still determined by the futures market. More than 90% of which is driven by speculation. In fact, the price of para rubber is quite volatile. Additionally, the price of para rubber can be affected by the price of synthetic rubber, which is a substitute for natural rubber. This usually results in the decline of natural rubber price [1]. The highest value export of para rubber in Thailand is a Ribbed Smoked Sheet No.3 or RSS3 grade. Ribbed smoked sheet No.3 is a grade of ribbed smoke sheet from the international Standard of Quality and Packing for Natural rubber grades (The Green Book). RSS3 is commonly used in rubber industry [2]. A price of ribbed smoked sheet No.3 had consistently increased until 2012. After that, it began to decline. Due to the large amount of product released to market and the sluggish of global economy. Furthermore, the demand of raw material have been continuously decreasing. This results in the even more declining price of rubber product [3]. A precise forecasting model for predicting the price of Ribbed smoked sheet No.3 would help the rubber farmers to strategize their plantation and production. This study hence aims to develop an appropriate model for such forcasting based on actual historical prices. It can be used as a guidance for analysis and production planning for aligning with the trends and directions of future rubber price.

2 Literature Review

2.1 Time series analysis

Time series analysis is the prediction of the future values of the selected dependent variable. It aims to investigate various patterns of relationships. Time series analysis consists of decomposing components, time series data, and analysis of the relationship patterns of that component, in order to forecast the future value of that time series data. This research utilizes 4 methods of time series analysis.

2.2 Moving Average

The moving average is a widely used tool for analyzing time series data. In this research, the moving average is calculated based on the current day's price, the previous day's total, and the average number of days. Unlike other technical tools, averaging essentially spreads out anomalies in the data and provides unambiguous information. There are three commonly used moving averages.

1) **Simple Moving Average (SMA):** It is the most widely used method. By using a weighting method, all values that are calculated are equally important to the price. and take the data over a period to find the mean.

2) **Weight Moving Average (WMA):** It is an analysis that gives the most im portance to the last day of calculation. The next day will be reduced in importance. The signaling of this weighted average is more sensitive than that of a simple moving average. therefore, it is usually ahead of the simple moving average. However, this weighted average line It describes only the changes that occur over the period under consideration, like the simple moving average method.

3) **Exponential Moving Average (EMA):** This method is a weighted average method that focuses on one value that influences price changes. and weighted to make the final value more important. This method does not consider the time of analysis.

Every price is affected by the EMA, although the most recent one is the most important. This method solves the shortcomings of the simple moving average method. The weighted average method gives the most importance to the last day and takes all values to the average without discarding the past data. This will cause every value to reflect the price change. While the other averages focus on periods, weighted averages focus on a value known as the Smoothing Factor (SF), or Smoothing Constant.

2.3 Holt's exponential smoothing method

Smoothing method with Holt's Exponential Curve is suitable for a time series that tends to be linear and does not contain seasonal components. There are two smoothing constants, namely the level smoothing constant (level : α) and the smoothing constant of Slope (trend : γ).

2.4 Box and Jenkins Method

Box and Jenkins method is a time consuming and complicated calculation which can be used with all moving data. It produces relatively high accuracy. At least 30 values are needed to determine the forecast model by checking the correlation function and some correlation of stationary. Mean and variance are constant over time. If the time series is moving unsteadily, for a non-stationary time series, it must be converted to a fixed time series before modeling. Box– Jenkins method is suitable for time series with trend components. and seasonal variations. There is a generalized model, e.g. SARIMA(p, d, q)(P, D, Q,). For time series which contains only trend components, the forecast model can be reduced to ARIMA(p, d, q).

2.5 Artificial Neural Network

Artificial Neural Network (ANN) is a simulation of the neural network in the human brain. It is capable of learning, remembering patterns and analogy knowledge, similar to the ability of the human brain. The initial idea for this technique came from the study of the bioelectrical network in the brain, which consists of neurons and synapses. The structure of a neural network consists of interconnected layers of neutrons. The first layer is the input and the last layer is the output layer. Between the first layer and the last layer there can be any number of hidden layers, or none at all [5]. The link of the artificial neural network can be divided into 2 types. Firstny, Feed forward network is the input signal sent to the next layer without returning it. Secondly, Feedback network (Recurrent network) is a processing that has the result back to the previous layer, causing a loop (Loop) and may cause transmission in the same layer [6].

Nonlinear autoregressive neural network (NNAR) is a type of neural network that has been developed for use in forecasting time series with one variable as follows: With NNAR, recurrent feedbacks are sent from the output layer to the input layer with a delay. The predicted output value can be estimated using only the data back in the past. This historical value with this delay is predicted as the time lag in the NAR model. The dependent variable output at time t+k, yt+k, is calculated from the historical output yt, . yt-1,…, yt-d where d is the number of historical data[4].

3 Data and Methodology

3.1 Data

Daily price of rubber from the Rubber Authority of Thailand is used in this independent study. This involves data from January 2011 to December 2021. Then it is divided into train and test sets as follows:

1) **Train set**: Data from January 2011 – December 2020, 2,376 values for build-model in 4 types of models including Moving average, Holt's exponential smoothing, Box – Jenkins and Artificial neural network.

2) **Test set**: Data from January 2021 – December 2021, 242 values for measuring accuracy of model.

3.2 Methodology

1) **Moving average:** This research calculates three types of movine averages as follows:
   - Simple moving average (SMA): calculated average data within the selected time period
   - Weighted moving average (WMA): calculated average data with weighted average points
   - Exponential moving average (EMA): calculated average data with specific attention on more recent data.
2) **Holt's exponential smoothing**: This method is appropriate for time series

data with trends, that have 2 constants (α for level and γ for trend).

3) **Box – Jenkins**: This method involves identifying if the time series data is

stationary, determining an appropriate autoregressive integrated moving average (ARIMA) model based on the ACF and PACF plots, and check suitability using Ljung-Box Chi-square statistics.

Determining stability whether a time series is constant with correlation function are autocorrelation function (ACF) and partial autocorrelation function (PACF). If it is found that the mean of the time series is not constant, the time series must be converted by variance or seasonal variance. If a time series has an unstable variance Time series should be converted with square root natural logarithm.

Determine possible predictive models from ACF and PACF curves of fixed time series.Checking the suitability of the forecasting model by checking the independent forecasting error values from the ACF and PACF curves and the Ljung Box Chi-square test statistics. multiple suitable The smallest Akaike's information criterion (AIC) can be used. and forecast using the most appropriate model.

4) Artificial Neural Network: In this study, a model selection method was used. nonlinear autoregressive neural network (NNAR) by experimentally constructing a feasible and suitable model as follows

1. The number of input data 1 node (using the time series at time t-1 gives the result is the forecast value at time t
2. 1 hidden layer and the number of nodes in the hidden layer is 1 – 10 nodes.
3. Number of time lags 1-9
4. The performance measurement function to select suitable models is model have the lowest MAPE.

After obtaining the most suitable model of the neural network method, Therefore, it is used for forecasting to compare the efficiency with other methods.

5) Accuracy: Evaluation of the performance of predictive models of various models. Criteria for evaluating forecast validity are used to measure performance. There are 3 criteria used for evaluation in this research as follows:

1. Mean Absolute Error: MAE

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|A_i - F_i|$$

A_i is The actual value of the price of para rubber from the test set.
F_i is the predicted value obtained from train set.
n is the total number of data.

2. Root Mean Squared Error:RMSE

$$\boldsymbol{RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2}}$$

A_i is The actual value of the price of para rubber from the test set.
F_i is the predicted value obtained from train set.
n is the total number of data.

a measure of the accuracy of forecasting, it is the square root of the mean squared error (MSE). The MSE value is the positive value of the error before being squared.

3. Mean Absolute Percentage Error: MAPE

$$\boldsymbol{MAPE = \frac{1}{n}\sum_{i=1}^{n} |\frac{A_i - F_i}{A_i}| \times 100}$$

A_i is The actual value of the price of para rubber from the test set.
F_i is the predicted value obtained from train set.
n is the total number of data.

Determining the accuracy of forecasting is comparing the data between the forecasted value and the actual value obtained in order to consider how close the forecasted value is to the actual value, regardless of the mark high accuracy. Calculation of Mean Absolute Percentage Error is shown in the above equation.

4 Results

Price movement of rubber smoked sheet considering the movement characteristics of the time series, the daily price of RSS3 rubber from 2011 – 2021 are analyzed based on 2,618 values. During 2011-2015, the price of RSS3 rubber has been continuously decreaseding. Then, in 2016, the price increased. Unfortunately, it began to decline again after the first quarter of 2017. It can be seen that the price of rubber is constantly changing. This can be affected by various factors. In this independent study, the data were divided into 2 sets, a learning set (green) and a test set (blue) to evaluate the efficiency of forecasting (see Figure 1).

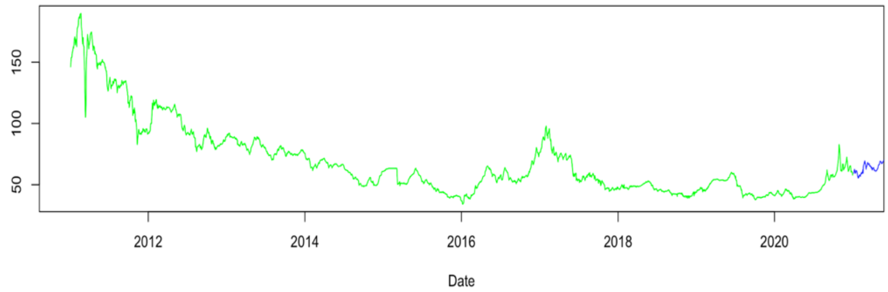


Fig. 1. The historical price movement of RSS 3.

1) Moving average method

Table 1. the error of forecasting by the method. Different types of moving averages

	MAE	RMSE	MAPE
SMA	9.349062	10.64503	16.25084
EMA	9.08153	10.32179	15.73664
WMA	6.4838	7.9440	11.2583

Table 1 shows the error of the forecasting. It was found that the Weight moving average (WMA) method is the most accurate method. The error of forecasting MAE, RMSE and MAPE are respectively 6.4838, 7.9440, 11.2583. EMA method and the SMA method result in less accurate results.

2) Stationary test

Considering the price movement of RSS3 rubber, it can be seen that the data is in the form of daily data. It moves in an upward and downward direction at certain times due to the influence of the trend. But there is no obvious repetitive movement. Since it is a daily data, there is no seasonal influence. The Holt exponential smoothing method was therefore used in this forecasting, that values recent data and trends over time.

Stationary test by Augmented Dickey-Fuller test can be considered from Augmented Dickey-Fuller test statistic is equal to -3.4146 which is smaller than the critical value at statistical significance level 0.05. This rejects the main hypothesis H0 : $\alpha = 0$ and Accept H1 :$\alpha < 0$, indicating that the variable has stationary properties and was tested at level with p-value equal 0.05089, which is greater than the critical value.

```
        Augmented Dickey-Fuller Test

data:  train_data
Dickey-Fuller = -3.4146, Lag order = 13, p-value = 0.05089
alternative hypothesis: stationary
```

Fig 2. stationary test by Augmented Dickey-Fuller

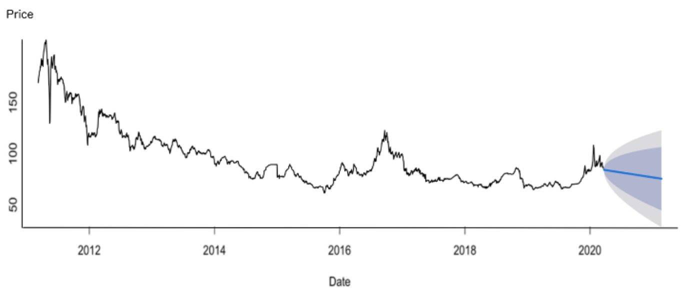


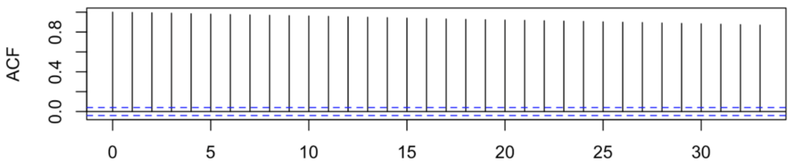
Fig 3. The results from Holt's exponential smoothing method

MAE	RMSE	MAPE
29.31461	32.62875	50.43724

Table 2. the error of forecasting by Holt's exponential smoothing method

The result of the Holt's exponential smoothing method is shown in Figure 3. The error of forecasting MAE, RMSE and MAPE is 29.31461, 32.62875, 50.43724 respectively according to Table 2.

3) **Box-Jenkins**: Initially characterize time series with self-correlation and partial self-correlation.



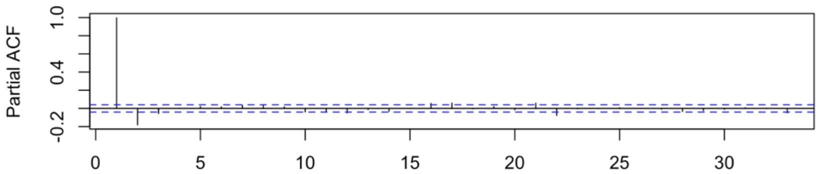
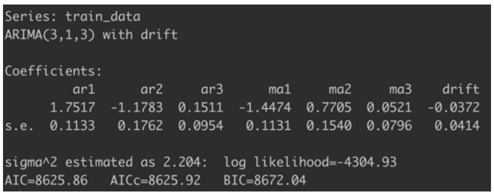


Fig 4. the self-correlation and some self-correlation

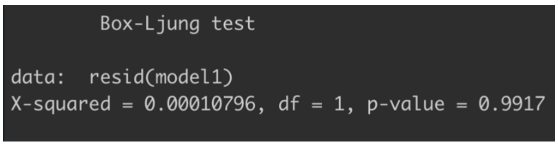
From the ACF and PACF graphs in Fig 4, the time series is stable. The time-series data and PACF curves are rapidly decreasing. which the time series can be used to define the model. The model ARIMA(p,d,q) was obtained from the analysis results in Fig. 5. The model used in the study was ARIMA(3,1,3) as the best model because it had the lowest AIC value equal 8,625.92.


```
Series: train_data
ARIMA(3,1,3) with drift

Coefficients:
         ar1      ar2     ar3      ma1     ma2     ma3    drift
      1.7517  -1.1783  0.1511  -1.4474  0.7705  0.0521  -0.0372
s.e.  0.1133   0.1762  0.0954   0.1131  0.1540  0.0796   0.0414

sigma^2 estimated as 2.204:  log likelihood=-4304.93
AIC=8625.86   AICc=8625.92   BIC=8672.04
```

Fig 5. the values of the most appropriate forecasting model


```
        Box-Ljung test

data:  resid(model1)
X-squared = 0.00010796, df = 1, p-value = 0.9917
```

Fig 6. the values Ljung-Box test

The independence of each other was examined by considering the Ljung-Box value. It is found that the p-value is 0.9917, which is greater than the significance level of 0.05. It can be concluded that a suitable forecasting model is ARIMA(3,1,3) for the time series data, ARIMA(3,1,3) is the model value ARIMA(p,d,q) include I. model Auto Regressive (AR(p)) 3rd order II. Moving Average (MA(q)) model 3rd order III. Integrated (I(d)) process at 1st difference.

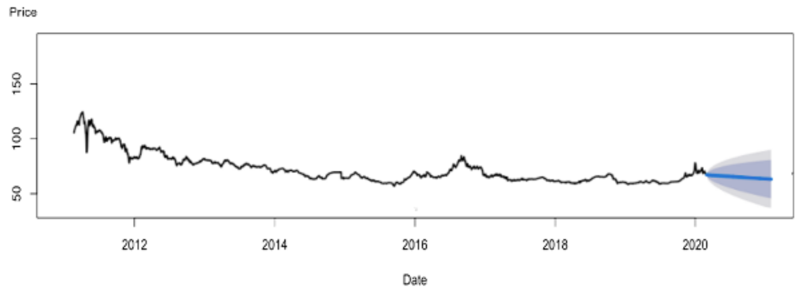


Fig. 7 shows the result from ARIMA (3,1,3) model

Table 3. shows the error of forecasting by Box – Jenkins method

MAE	RMSE	MAPE
6.12577	7.41726	9.94420

Forecasting by Box-Jenkins method using ARIMA(3,1,3), the forecasting results are as shown in Figure 7. The errors of forecasting MAE, RMSE and MAPE are 6.12577, 7.41726, 9.94420 respectively according to Table 3.

4) Artificial Neural Network

Artificial neural network method utilizes nonlinear autoregressive neural network (NNAR) from train set data by finding the number of nodes in the hidden layer and the appropriate amount of time lag. MAPE was used as a criterion to evaluate the model. and repeated for 10 times. The values of the suitable model are shown in Table 4.

Table 4. the values of NNAR model that are suitable for forecasting

Hidden node	Time lag	MAPE
10	9	1.109419

```
Series: train_data
Model:  NNAR(9,10)
Call:   nnetar(y = train_data, p = 9, size = 10)

Average of 20 networks, each of which is
a 9-10-1 network with 111 weights
```

Fig 8. Forecasting model for Artificial Neural Network Method

Forecasting with NNAR model of 1 layer with 9 nodes in the hidden layer and using historical data for 9 time periods (9-Time lags) with a weight of 111 to test the forecasting efficiency. The results are as shown in the picture. at 4.8, the error of forecasting MAE, RMSE and MAPE is 4.535161, 5.780739, 7.330895 respec-tively according to Table 4.

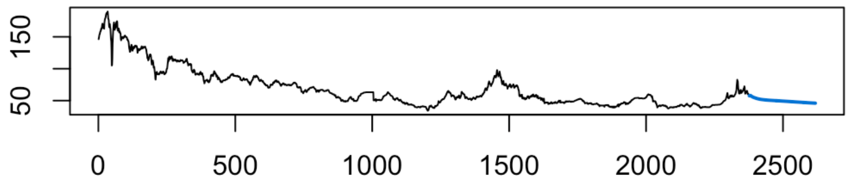


Fig 9. Forecasting from nonlinear autoregressive neural network model

Table 5. the error of forecasting from Artificial Neural Network

MAE	RMSE	MAPE
4.535161	5.780739	7.330895

Table 6. Accuracy of forecasting model

Method	**MAE**	**RMSE**	**MAPE**
Moving Average)WMA)	6.4838	7.9440	11.2583
Holt’s exponential smoothing	29.3146	32.6288	50.4372
Box – Jenkins	6.1258	7.4173	9.9442
Artificial neural network	4.5352	5.7807	7.3309

Table 6 compares the errors of various forecasting models. It indicates that Holt's exponential smoothing method has the largest error, followed by moving average method, Box-Jenkins method with ARIMA(3,1,3) and Artificial neural network method, respectively. The method with the least error is Artificial neural network method with MAE value of 4.5352, RMSE value of 5.7807, and MAPE value of 7.3309. Box–Jenkins method, artificial neural network method, and moving average method have similar tolerances because Box–Jenkins method and artificial neural network method are used to test data to find a model that is suitable for the data first. They are used to make the forecast model suitable for making the forecast value close to the actual data. The forecast obtained is closer to the true value.

5 Conclusion

This independent study attempts to find the most appropriate model to forecast the price of rubber smoked sheet in Thailand based on different forecasting models. The weight moving average (WMA) is found to be the most suitable method to identify the errors. Artificial Neural network is found to be is the most suitable method to be used as a model for forecasting price of ribbed smoke sheet.

6 Recommendation

1) To increase the efficiency of forecasting model, additional historical price and other factors that may affect price of ribbed smoke sheet should be included. This includes world oil prices, exchange rate, etc.

2) Increasing the number of hidden layers may increase the effectiveness of the forecasting model.

References

1. Dawan Buangam and Sontila B., Rubber Product in Yala, Information segment, Department of Agricultural Extension, Yala, 2019.
2. The office of Industrial Economics and Plastics Institute of Thailand, Natural RubberExport Statistics,[Online].Avaliable:http://rubber.oie.go.th/ImExThaiByProduct.aspx?pt=im[Access September 2022]
3. Punnawit Lertsiwanon, Factor effects Ribbed smoke sheet No.3 export FOB price of Thailand, Independent study, Master of Economics, Faculty of Economics, Thammasat University, 96 pages, 2017.
4. Analaya Nansailao, Forecasting model export sugar value of Thailand use nonlinear autoregressive neural network, Khonkaen Agriculture Journal 49(1): 179-191, 2021.
5. Nerawan Kaiwan, Scrap forecasting using neural networks Case study of workpiece welding process, Kasetsart University, Bangkok, 2017.
6. Kanyakorn Saejeen, Forecasting of cassava root yield by artificial neural network using cassava morphology, Kasetsart University, Bangkok. 2018.