



Modeling and Forecasting of Total Supply Malaysia's Centrifugal Sugar Using ARIMA Model

Nur Amalina Shafie

Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia

Nurbaizura Borhan

School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Digital Innovation

Noor Amalina Nisa Ariffin

Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia

Khairil Anuar Md Isa

Department of Basic Sciences, Faculty of Health Sciences, Universiti Teknologi MARA Selangor Branch, Puncak Alam Campus, 42300 Selangor, Malaysia

Nor Azura Md Ghani

Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia

ABSTRACT

Centrifugal sugar is essential to Malaysia's food sector and culinary scene because it is high-quality, consistent, and convenient for a variety of food and beverage applications. Customers, chefs, and food manufacturers who are looking for dependable sweetening solutions for their regular cooking and dining experiences choose it because of its broad availability and adaptability. To support market stability, price control, production scheduling, trade policy development, risk management, policy formation, and consumer welfare, it is essential to forecast the centrifugal sugar supply in Malaysia. Autoregressive Integrated Moving Average (ARIMA) model is one of Box-Jenkins model which is common time series forecasting methodology that is widely utilised in various sectors including economics, finance, and business. Therefore, this research interest to model and forecast the total supply of centrifugal sugar in Malaysia using ARIMA model. This research used data from the IndexMundi website and analysed by using R software. This research found that the best model to forecast the total supply Malaysia's centrifugal sugar is ARIMA (1,1,2). The total supply of Malaysia's centrifugal sugar will increase in 6 years ahead.

Keywords: Centrifugal sugar, ARIMA, total supply, forecasting.

INTRODUCTION

Centrifugal sugar is included in the production and trading of sugar which is increased significantly worldwide; the top exporting countries are Brazil, Thailand, and India [1-2]. However, Malaysia and other ASEAN countries face challenges in the sugar industries despite their significant role in global trade [3]. Agidi and Igbeka [4] offers a particular illustration of a small-scale sugar centrifuge that would be able to supplement Malaysia's overall centrifugal sugar supply. Centrifugal sugar refers to sugar that has been extracted from a liquid by use of a centrifugal machine. It is the result of the machine's successful separation of sugar crystals through the application of centrifugal force. Using centrifugal force, sugar crystals are separated from the surrounding liquid, such as molasses or syrup. In the sugar processing sector, centrifugal sugar is essential, especially for small-scale operations. Although it can cause some crystal disintegration, it helps eliminate contaminants in sugar [5]. It has been discovered that applying centrifugal force to the solute extraction from sugar beetroot tissue greatly improves the extraction process [6].

Hoh [7] and Yuttitham et al. [8] mentioned that Malaysia's sugar production is primarily centrifugal sugar, with the country being only about 5 percent self-sufficient in domestic sugar production. This indicates that most of Malaysia's sugar production consists of centrifugal sugar, with a small percentage being domestically produced. Malaysia's cane production is predicted to reach 800 TMT in 2006 due to improved growth conditions and dry weather [9]. In 2006 and 2007, a 3% rise in domestic sugar consumption is anticipated. Tarimo and Takamura [10] mentioned that in terms of local sugar production, Malaysia is just approximately 5% self-sufficient.

Numerous studies have effectively employed the Auto-Regressive Integrated Moving Average (ARIMA) model to forecast sugar production. ARIMA models are widely used in time series analysis to understand and forecast data patterns. Integrated (I), Auto-Regressive (AR), and Moving Average (MA) are the three main components of the ARIMA model. In order to anticipate Malaysia's total supply of centrifugal sugar using ARIMA, the analysis of historical supply data will determine the essential ARIMA parameters (p , d , q) based on the autocorrelation and partial autocorrelation functions. These parameters reflect the number of moving average terms, differencing, and autoregressive terms, respectively. Once fitted to the data, the ARIMA model can be used to predict future levels of supply. Numerous studies have shown that ARIMA models perform well when used for time series analysis. It was discovered by Zhang [11] that hybrid ARIMA and neural network model increased predicting accuracy. Fattah et al. [12] forecasted demand in a food industry using ARIMA models with success; the chosen model was verified against historical data. Both Ariyo et al. [13] and Mondal et al. [14] discovered that ARIMA models could accurately forecast stock values; Mondal et al. [14] also pointed out that these models might be used for short-term forecasting. All these studies demonstrate how adaptable and trustworthy ARIMA models are for a range of forecasting applications.

The ARIMA model would specifically be used for the sugar supply in Malaysia to examine past supply data, spot trends, seasonality, and other supply-influencing factors, and then project future supply levels based on these patterns. Putra et al. [15] created a forecasting system for

Indonesian sugar production, Wardah et al. [16] combined demand forecasting and inventory control for coconut sugar. The ARIMA model was employed by Devaki and Mohideen [17] and Muhammad et al. [18] to predict the production of sugarcane in Tamil Nadu and Pakistan, respectively. By using the ARIMA model to anticipate maize production in India, Sharma et al. [19] demonstrated how adaptable this forecasting method is for a variety of crops and nations. Tripathi et al. [20] used the ARIMA model as well to estimate Odisha's rice production and compared the results with those for the entire country of India. Kumar et al. [21] used Box-Jenkin's ARIMA model to forecast sugarcane production in India. Saleem et al. [22] also used the ARIMA model to forecast the production of centrifugal sugar in Pakistan. This research focuses on modelling and forecasting the total supply of Malaysia's centrifugal sugar by using the ARIMA model.

METHODOLOGY

This research used annual data total supply of centrifugal sugar in Malaysia from 1972 to 2023 from the IndexMundi website. The estimating and evaluation portions of the data were separated into two groups. In the estimating stage, centrifugal sugar measurements from 1972 to 2002 were used as training data. Meanwhile, the evaluation portion employed centrifugal sugar data from 2003 to 2023. This research used descriptive analysis and Autoregressive Integrated Moving Average (ARIMA) model. Descriptive statistics are used to get the trend of the raw data. While the ARIMA model was focused on time series prediction.

ARIMA Model

The ARIMA model is one type of Box-Jenkins model that is used to predict future time series data based on historical time series data. The ARIMA model is a generalization of the ARMA model. The ARMA model is used for stationary data. A stationary zero mean ARMA (p, q) model for a time series $\{X_t\}$ is defined as Equation 1.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (1)$$

where

- X_t is the value of time series at time t.
- $\phi_1, \phi_2, \dots, \phi_p$ are the parameters of the autoregressive part of the model.
- $\theta_1, \theta_2, \dots, \theta_q$ are the parameters of the moving average part of the model. $\{\varepsilon_t\}$ is the white noise error term at time t, with zero mean and constant variance.

When a time series is not stationary, then differencing operations are applied with the appropriate lag to achieve stationarity then is called the ARIMA model. The ARMA model has two order parameters which are p and q. While, the ARIMA model has three order parameters which are p, d, and q. p refers to the order of the autoregressive process, d refers to the order of integration or differencing and q refers to the order of the moving average method used in the model.

The raw data is separated into two parts which are training data and testing data. The training data is used to fit the model while the testing data is used to evaluate the model's performance. The training data is identified whether is stationary or not by using a unit root test. The data is considered stationary when its mean, variance, and autocovariance are all-time invariant. This

research used the Augmented Dickey-Fuller (ADF) test as Equation 2 to determine whether the data is stationary or not by following the hypothesis below:

- H_0 : The series has a unit root
- H_1 : The series is stationary

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \dots + \delta_p \Delta y_{t-p} + \epsilon_t \quad (2)$$

where

- Δ is the first difference operator.
- y_t is the time series.
- t is the time trend.
- ϵ_t is the error term.

The data is stationary if the p-value is less than $\alpha=0.05$, the null hypothesis is rejected. This means that the unit root is absent from the series. If the series is not stationary, it should be adjusted to become so, and differentiation is a typical way to do this. A data is differencing by subtracting the current and previous values d times. When the stationarity assumption is violated, differencing is frequently utilized to stabilize the series. Autocorrelation Function (ACF) correlogram also can help to identify stationary data. The data is stationary if the data fluctuates randomly around some fixed values.

The ACF correlogram is used to determine the q order by counting the number of significant lags. The Partial Autocorrelation Function (PACF) correlogram is used to determine the p order by counting the number of significant lags. After that, the residuals of suggestion models are checked to either fulfill the assumption of white noise or not by using the Ljung Box test as Equation 3 by following the hypothesis below:

- H_0 : The residuals have white noise (there is no autocorrelation)
- H_1 : The residuals have no white noise (there is significant autocorrelation)

$$Q = n(n + 2) \sum_{k=1}^m \frac{\hat{r}_k^2}{n-k} \quad (3)$$

where

- n is the number of observations
- m is the number of lags being tested
- $\hat{r}_k$ is the sample autocorrelation at lag k .

If the result has a p-value more than $\alpha = 0.05$, so the residuals have white noise since it failed to reject the null hypothesis. The models that have white noise are chosen. After that, the performance of forecasting models can be evaluated using forecasting accuracy measures such as Mean Square Error (MSE) and Root Mean Square Error (RMSE). The formula for the measures is shown in Equations 4 and 5. The best forecasting model is found at the smallest MSE and RMSE.

$$MSE = \sum_{t=1}^N \frac{(\sigma_t^2 - \hat{\sigma}_t^2)^2}{N} \quad (4)$$

$$RMSE = \sqrt{\sum_{t=1}^N \frac{(\sigma_t^2 - \hat{\sigma}_t^2)^2}{N}} \quad (5)$$

where

- N is the number of total observations t is the t -th observation in out-of-sample
- σ_t^2 and $\hat{\sigma}_t^2$ correspond to the actual and predicted condition variance at time t , respectively.

RESULTS AND DISCUSSION

Raw Data

Based on Figure 1, the trend of centrifugal sugar was on the rise. It means the total supply of centrifugal sugar in Malaysia from 1972 to 2023 is increasing.

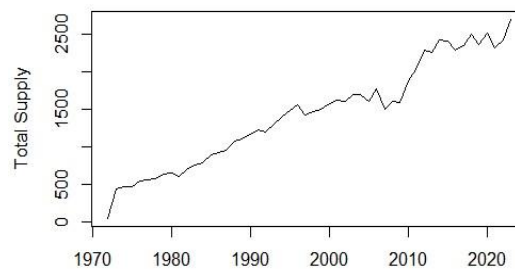


Figure 1: Trend analysis for total supply Malaysia's centrifugal sugar

Test of Stationary

As mentioned above the data were separated into two groups which are from 1972 to 2002 were used as training data and data from 2003 to 2023 as testing data. For the test of stationary, training data were used. The ACF of the data, as seen in Figure 2, displayed a declining pattern, supporting the idea that the data was not stationary. Additionally, the ADF test in Table 1 also verifies that the data is not stationarity since the p-values above $\alpha = 0.05$ (0.4559). As a result, there is insufficient proof that the data has a unit root at the 5% level of significance. Hence, the data is not stationary.

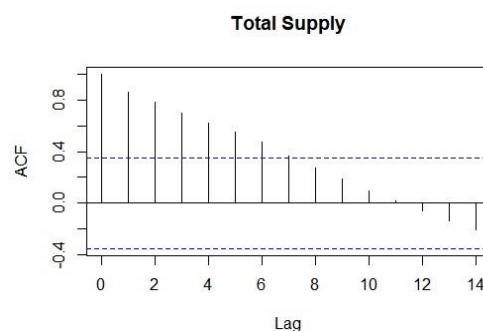


Figure 2: ACF correlogram of original data

Table 1: Unit root test of original data

Test	p-value
Augmented Dickey-Fuller (ADF)	0.4559

The differencing technique was used to fulfill the stationary analysis assumptions. ACF of first-order differencing is displayed in Figure 3. The initial differencing between data is stationary since the plot does not exhibit a declining pattern. The null hypothesis was rejected when the results of the ADF test in Table 2 showed p-values of 0.04526, which are less than $\alpha = 0.05$. Consequently, there is sufficient evidence to conclude that the data lacks a unit root at the 5% significance level. The initial set of differencing data is therefore stationary.

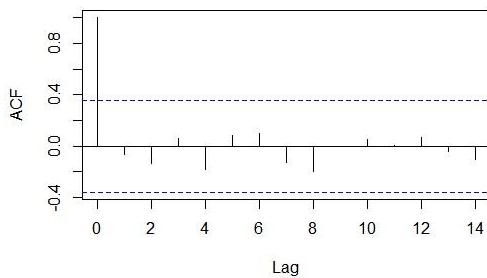


Figure 3: ACF correlogram of first differencing

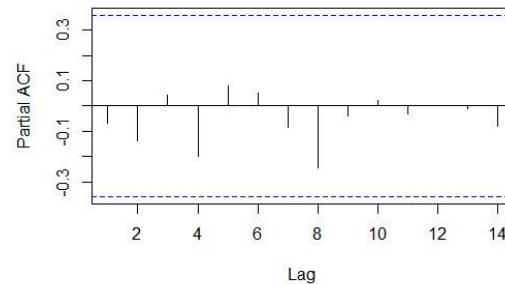


Figure 4: PACF correlogram of first differencing

Table 2: Unit root test after differencing

Test	p-value
Augmented Dickey-Fuller (ADF)	0.04526

Modelling

Figures 3 and 4's ACF and PACF can be used to create an ARIMA model. The ACF indicates testing MA(1) since it displays one notable spike at lag 0. In the meantime, PACF displays a single notable peak at lag 8. It therefore recommends testing the AR (1) model. For this investigation, a total of six reasonable ARIMA models were examined.

Table 3 shows that all suggested models have p-values more than $\alpha = 0.05$ except ARIMA (2,1,1). It means all the suggested models except ARIMA (2,1,1) are white noise since all the suggested models failed to reject the null hypothesis.

Table 3: Portmanteau test

Model	Ljung-Box Test (p-value)
ARIMA(0,1,1)	0.3214
ARIMA(1,1,0)	0.3875
ARIMA(1,1,1)	0.149
ARIMA(1,1,2)	0.3314
ARIMA(2,1,1)	0.03525
ARIMA(2,1,2)	0.0758

Forecast

The values for the forecasting accuracy metrics are displayed in Table 4. The lowest values of MSE and RMSE values is ARIMA (1,1,2) when compared to other models. As a result, this model might be regarded as the series' best forecasting model.

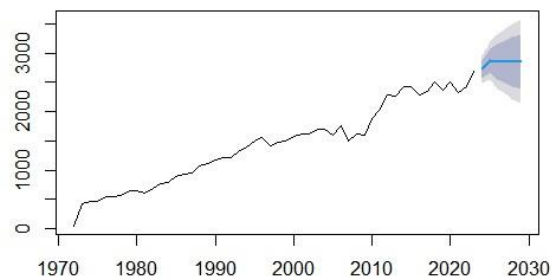
Table 4: Forecast accuracy

Model	MSE	RMSE
<i>ARIMA(0,1,1)</i>	1931872.7631	1389.9183
<i>ARIMA(1,1,0)</i>	2024512.3923	1422.8536
<i>ARIMA(1,1,1)</i>	1931820.0783	1389.8993
<i>ARIMA(1,1,2)</i>	1923849.0835	1387.0288
<i>ARIMA(2,1,2)</i>	1925908.6152	1387.7711

After selecting the ARIMA (1,1,2) model, the model was used to forecast 6 years ahead. Table 5 shows the forecast value and Figure 5 shows the graph of forecast value.

Table 5: Forecast value

Year	Forecast value (1000 MT)
2024	2725.946
2025	2874.475
2026	2857.472
2027	2859.418
2028	2859.195
2029	2859.221

**Figure 5: Forecasting plot**

CONCLUSIONS

Centrifugal sugar is essential to Malaysia's food sector. Centrifugal sugar is sugar that has been extracted from liquid using a centrifugal machine. While non-centrifugal cane sugar (NCS) is a technical term for conventional raw sugar made by evaporating water from sugarcane juice. The sugar industry in Malaysia is facing several challenges, including those related to market dynamics and technology utilization. Encompassing economic, social, and environmental aspects, centrifugal sugar forecasting is very crucial to be taken into account. The accurate forecasting techniques can help to ensure that there will be a sufficient supply to meet the demand of the population. In addition, forecasting supply accurately will help in stabilizing food prices by preventing overproduction or underproduction that somehow will lead to price volatility. Therefore, this research has been devoted to predicting the total supply of centrifugal sugar prediction in Malaysia using the ARIMA model approach. Using the ARIMA model approach will involve several steps. Starting from checking stationary to model diagnostics and then forecasting process. It is indeed a powerful forecasting tool and can be effectively applied for centrifugal sugar supply data in predicting future trends which helps in efficient future planning and decision making. This prediction of centrifugal sugar plays a critical role in enhancing economic stability, public health, environmental sustainability, and overall society

well-being. By leveraging accurate forecasts, it can help stakeholders to make informed decisions that not only benefit the food sector but also the broader economy and society. This research found that the best model to forecast centrifugal sugar is ARIMA (1,1,2) since it has the lowest accuracy error. This research also found that the values of six years ahead have gradually increased pattern.

ACKNOWLEDGEMENT

We would like to thank Universiti Teknologi MARA for their financial support through (600-RMC/GPM LPHD 5/3 (136/2021)).

References

- [1] Mariotti, M., & Lucisano, M. (2014). Sugar and sweeteners. *Bakery products science and technology*, 199-221.
- [2] Godshall, M. A. (2007). Sugar and other sweeteners. In Kent and Riegel's *Handbook of Industrial Chemistry and Biotechnology* (pp. 1657-1693). Boston, MA: Springer US.
- [3] Solomon, S., Swapna, M., Xuan, V. T., & Mon, Y. Y. (2016). Development of sugar industry in ASEAN countries. *Sugar Tech*, 18, 559-575.
- [4] Agidi, G., & Igbeka, J. C. (2003). Development and performance of a sugar centrifuge. *Sugar tech*, 5, 131-136.
- [5] Jullienne, L. M. S. A. (1983). Washing sugar in batch A-centrifugals. *Proceedings of The South African Sugar Technologists' Association-June*, 43.
- [6] EL-BELGHITI, K. A. M. A. L., Rabhi, Z., & Vorobiev, E. (2005). Effect of centrifugal force on the aqueous extraction of solute from sugar beet tissue pretreated by a pulsed electric field. *Journal of Food Process Engineering*, 28(4), 346-358.
- [7] Hoh, R. (2006). *Malaysia Sugar Annual 2006*. USDA Foreign Agricultural Services Global Agriculture Information Network (GAIN) Report.
- [8] Yuttitham, M., Gheewala, S. H., & Chidthaisong, A. (2011). Carbon footprint of sugar produced from sugarcane in eastern Thailand. *Journal of Cleaner Production*, 19(17-18), 2119-2127.
- [9] Li, Y. R., & Yang, L. T. (2015). Sugarcane agriculture and sugar industry in China. *Sugar Tech*, 17(1), 1-8.
- [10] Tarimo, A. J., & Takamura, Y. T. (1998). Sugarcane production, processing and marketing in Tanzania. *African Study Monographs*, 19(1), 1-12.
- [11] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
- [12] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, 10, 1847979018808673.
- [13] Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014, March). Stock price prediction using the ARIMA model. In *2014 UKSim-AMSS 16th international conference on computer modelling and simulation* (pp. 106-112). IEEE.
- [14] Mondal, P., Shit, L., & Goswami, S. (2014). Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. *International Journal of Computer Science, Engineering and Applications*, 4(2), 13.

-
- [15] Putra, J. A., Basbeth, F., & Bukhori, S. (2019, September). Sugar production forecasting system in PTPN XI semboro jember using autoregressive integrated moving average (ARIMA) method. In 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) (pp. 448-453). IEEE.
- [16] Wardah, S., Nurhasanah, N., & Sudarwati, W. (2023). Integration models of demand forecasting and inventory control for coconut sugar using the ARIMA and EOQ modification methods. *Jurnal Sistem dan Manajemen Industri*, 7(2), 127-138.
- [17] Devaki, C., & Mohideen, A. K. (2022). A sugarcane production in tamilnadu using a comparison of arima, state space and linear mixed models. *International Journal of Mechanical Engineering*, 970-974.
- [18] Muhammad, F., Javed, M. S., & Bashir, M. (1992). Forecasting sugarcane production in Pakistan using ARIMA Models. *Pak. J. Agric. Sci*, 9(1), 31-36.
- [19] Sharma, Pawan Kumar, Sudhakar Dwivedi, Lyaqat Ali, and R. K. Arora. "Forecasting maize production in India using ARIMA model." *Agro-Economist* 5, no. 1 (2018): 1-6.
- [20] Tripathi, Rahul, A. K. Nayak, R. Raja, Mohammad Shahid, Anjani Kumar, Sangita Mohanty, B. B. Panda, B. Lal, and Priyanka Gautam. "Forecasting rice productivity and production of Odisha, India, using autoregressive integrated moving average models." *Advances in Agriculture* 2014 (2014).
- [21] Kumar, Manoj, and Madhu Anand. "An application of time series ARIMA forecasting model for predicting sugarcane production in India." *Studies in Business and Economics* 9, no. 1 (2014): 81-94.
- [22] Saleem, Muhammad Kashif, Parwarasha Nazir, and Fahim Nazeer. "Forecasting the Production of Centrifugal Sugar in Pakistan using ARIMA Modelling." *Journal of Statistics* 27 (2023): 96.
- [23] IndexMundi (2024).
<https://www.indexmundi.com/agriculture/?country=my&commodity=centrifugalsugar&graph=total-supply>