Auto-Regressive Integrated Moving Average (ARIMA) Forecasting for Monthly Household LPG Demand in Indonesia

Aan Nurochman and Armand Omar Moeis

March 1, 2021

Auto-Regressive Integrated Moving Average (ARIMA) Forecasting for Monthly Household LPG Demand in Indonesia

Aan Nurochman*

Energy Systems Engineering Study Program, University of Indonesia, aan.nurochman@ui.ac.id

Armand Omar Moeis

Department of Industrial Engineering, University of Indonesia, armand.omar@ui.ac.id

Forecasting demand for household LPG in Indonesia, which uses this energy source as the primary commodity in people's cooking activities, is an essential policy tool used by decision-makers to make plans and decisions. Besides that, most of the initial studies related to the supply chain always put forward the demand side in determining the next policy planning steps. Based on this, a method is needed to obtain the estimated value of future LPG needs. The decision-makers that decide at this time in terms of investment, subsidy planning, and supply planning and control systems in their implementation become more appropriate and effective. Based on some studies and best practices, time series forecasting can provide better results found on these problems. This study uses the Autoregressive Integrated Moving Average (ARIMA) method to estimate household LPG demand in Indonesia. ARIMA forecasting of the total household LPG demand will provide a clearer picture of taking strategic steps later.

CCS CONCEPTS

Computing methodologies~Modeling and simulation~Model development and analysis~Uncertainty quantification

Additional Keywords and Phrases: Household LPG Demand, ARIMA Forecasting, Indonesia

ACM Reference Format:

Aan Nurochman and Armand Omar Moeis, 2021, Auto-Regressive Integrated Moving Average (ARIMA) Forecasting for Monthly Household LPG Demand in Indonesia.

1 Introduction

Energy demand forecasting is one of the most important policy tools used by decision-makers all over the world [1]. Energy demand is also a critical factor in planning the development of supporting infrastructure and strategies in the procurement process for meeting this energy. One type of fuel currently needed in Indonesia is LPG (Liquified petroleum gas), especially household needs.

Based on the data, the development of LPG use for households in Indonesia had experienced very significant growth. Before conversion in 2006, the consumption was only 1.1 million tons to 7.8 million tons in 2019 [2]. This development should be a serious concern considering that the increase in LPG consumption must be balanced with the development of infrastructure and the proper LPG supply process to positively impact PT Pertamina (Persero) as the executor of the assignment in its supply and distribution. From this, we can see the role of accuracy in demand in the planning process, especially the procurement of LPG in Indonesia, as shown below.

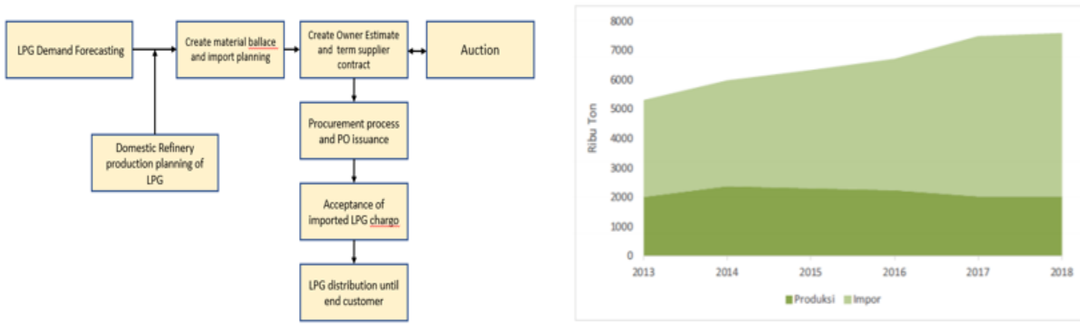


Figure 1: **Flow LPG Procurement and distribution planning (left) and Production LPG versus Import (right)**

Figure 1 above shows that in the chain of LPG procurement and distribution plans in Indonesia, the position of demand is in the first chain and becomes the basis for subsequent planning. The figure also shows that demand forecasting does not affect the amount of production planning at domestic LPG refineries in Indonesia's case. This happens because all refineries' total LPG production capacity in Indonesia is far from the total consumption [3, 4], as shown in the following graph.

On the other hand, in planning the supply and distribution chain in fulfilling LPG needs, this demand position is critical because Indonesia's geography is very demanding for challenges as an archipelago consisting of approximately 17,508 islands. This condition causes Indonesia to have a coastline of about 81,000 km or become a country with the second-longest coastline after Canada [5]. Thus, the accuracy of demand as an input in the supply and distribution chain will affect the process's effectiveness and efficiency. On the supply side, the following is an overview of current LPG fulfillment in Indonesia.

1.1 LPG Supply Chain in Indonesia

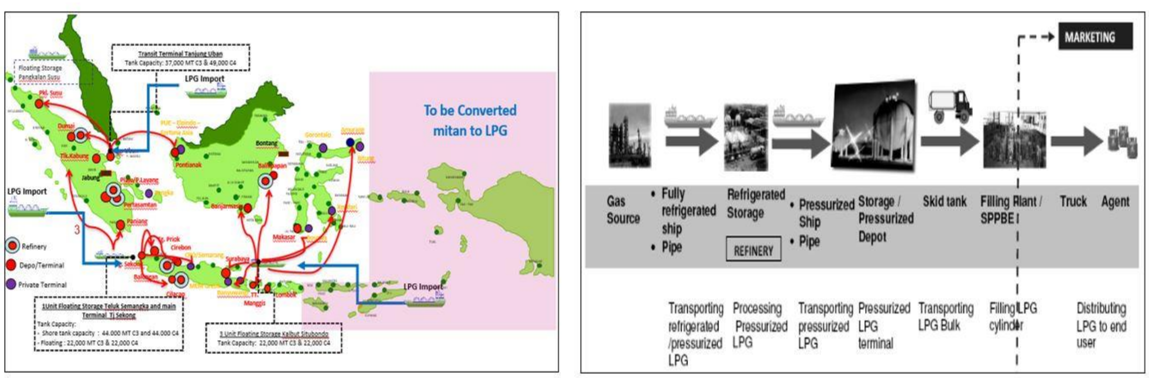


Figure 2: **LPG Supply and distribution pattern in Indonesia**

Figure 2 above shows the complexity in the context of LPG supply in Indonesia because the LPG terminal infrastructure has not been able to keep up with the current demand growth rate. As the demand for LPG increases, the existing LPG terminals' capacities must be added and improved [2, 6]. However, to keep the operation running well. The supply chain still requires floating storage to support its operations. Besides that, the land distribution is no less complex. The following is an overview of the LPG supply chain in Indonesia [7]

From the figure above, it can be seen that the distribution chain is also quite long. It starts from main LPG terminals to agents and LPG bases with more than 30 thousand distribution channels. Based on the above, in terms of the complexity of the LPG distribution chain in Indonesia, the accuracy of demand forecasting has a direct impact on the following:

- The number of subsidies that the government will provide, precisely the case of LPG PSO (Public Service Obligation), where the more significant the demand for the value of subsidies issued by the government, the greater [2]
- How much infrastructure capacity in the distribution chain must be prepared by PT Pertamina (Persero) as the implementer of supply and distribution activities.
- How much LPG supply is needed through imports, where the value continues to increase so that the procurement strategy must be right.

Based on the total realization of LPG consumption for households, there is a trend of increasing consumption every year, as in the figure below [8]:

Figure 3: Yearly LPG Consumption (Million Tons)

The following factors influenced the increase:

- There is a population growth rate: It is known from BPS 2020 data that the total population of Indonesia based on the final population census report as of September 2020 is 268.1 million people with a population growth rate of 1.15% in 2019 [8]
- Economic growth rate: On this economic growth side, starting from 2015 - 2019, it tends to be stable at 5% [8] .so that the effect on demand growth is not significant.
- The rate of development number of UMKM: Even though economic growth is stagnant at 5%, if the development of UMKM is massive enough, this can also affect the total LPG Demand.
- Addition of conversion areas: Currently, several regions have not converted kerosene to LPG. The addition of conversion areas will increase the overall demand.

For the planning process both in terms of LPG procurement and infrastructure development to be effective, a practical time series forecasting method is needed. In this paper, LPG forecasting demand will be carried out using ARIMA (Auto Regression Integrated Moving Average) modeling as a scientific approach to existing data realization.

1.2 Why ARIMA?

Auto Regression integrated moving average (ARIMA) is a forecasting method with excellent accuracy for short-term forecasting and for non-stationary time series data when it is linear [9]. ARIMA is widely used internationally for forecasting problems in various fields such as energy demand, inflation, production, and others.

Tabel 1 : The Country that uses the ARIMA method for forecasting

Country	ARIMA Forecasting topic
Turkey[1]	Forecasting energy Primer
Turkey[10]	Forecasting Irish inflation using ARIMA models
Turkey[11]	Natural Gas Energy Demand
Saudi Arabia[12]	Forecasting Monthly electric energy consumption
Spanish[13]	Next-Day Electricity price
Marocco[14]	Modeling and forecasting energy demand
China[15]	Forecasting of China's coal consumption

2 Literature Review

2.1 Forecasting Demand

Forecasting is an activity that tries to predict future conditions by using past mass data from a variable. Another definition of forecasting is the art of determining meaningful information about the future [16]. Forecasting is a vital part of an organization's statistics and operations that play an essential role in decision-making. Forecasting functions to predict what will happen in the future based on past data.

In Indonesia, post-conversion use of LPG is undergoing sudden and significant changes affecting many sectors where all developments require an accurate and practical reading into the future. Here the forecast is fundamental because it is a sign of the continuity of energy management, especially LPG for households in Indonesia. Divination is the science that estimates the future rates of several variables. The variable most often in demand, but it can also be something else, such as supply or price. Forecasting is the operation of making assumptions about the future value of the studied variable. One of them is by using the time series forecasting method.

A time series is nothing but observations according to the chronological order of time.[17] Time-series forecasting models use mathematical techniques that are based on historical data to forecast demand. It is founded on the hypothesis that the future is an expansion of the past; that's why we can use historical data to predict future demand.

By time series analysis, the forecasting accuracies depend on the characteristics of the time series of demand. If the transition curves show stability and periodicity, we will reach high forecasting accuracy, whereas we can't expect high accuracy if the curves contain highly irregular patterns.

2.2 Auto-Regressive Moving Average

We can work with the traditional statistical models to model time series, including moving average, exponential smoothing, and ARIMA. These models are linear since the future values are cramped to be linear functions of past data.

During the past few decades, researchers have focused much on linear models since they had proved simplicity in comprehension and application. Time-series forecasting models are mostly used to predict demand. Under an autoregressive moving average hypothesis, Kurawarwala and Matsuo [18] calculated the seasonal variation of demand using historical data and validated the models by examining the forecast performance. Miller and Williams [19] mixed seasonal factors in their model to improve forecasting accuracy. The seasonal factors are calculated from the multiplicative model. Hyndman widened Miller and Williams [19].

Work by applying different relationships between trend and seasonality under the seasonal ARIMA hypothesis. The classical ARIMA approach becomes prohibitive. In many cases, it is impossible to determine a model when the seasonal adjustment order is high, or its diagnostics fail to indicate that time series is stationary after seasonal adjustment. In such cases, the static parameters of the classical ARIMA model are considered the principal constraint to forecasting high variable seasonal demand. Another limitation of the classical ARIMA approach is that it requires many observations to determine the best fit model for a data series. An ARIMA model is labeled as an ARIMA model (p, d, q), wherein:

- p is the number of autoregressive terms;
- d is the number of differences; and
- q is the number of moving averages.

The ARIMA method has been studied in depth by George Box and Gwilyn-Jenkins (1976), so that this method is called the Box Jenkins method. The ARIMA method consists of several parts, including:

- The Auto-Regressive (AR) method was first introduced by Yule Walker (1927)
- The Moving Average (MA) method was first used by Slutsky (1937)
- Wold (1938) produced the theoretical underpinnings of the ARMA combination process
- ARIMA refines processes that include periodic seasonal series and simple development, which provides for non-stationary functions.

According to Makridakis, the ARIMA method is often written in the backshift operator. Notation B can be more than one [20]. Generally defined as follows:

$$B^k Y_t = Y_{t-k} \qquad (1)$$

The backshift operator B can be extended by definition to different (1 - B). If Y_t is multiplied by (1 - B), the following equation will be obtained:

$$(1 - B)Y_t = Y_t - BY_t = Y_t - (1 - B)Y_t = Y_t - BY_t = Y_t - Y_{t-1}) \qquad (2)$$

Keep in mind that B is not a number, so (1 - B) is also not a specific number but an operator.

2.3 Stationary of data

In using the ARIMA method, the data used must be stationary in terms of both the variety and the mean side. If the data used for forecasting is stationary from the start, it will be easier. It can be directly carried out by the autoregressive or moving average process, but if the data used is not stationary. It must be made stationary first because it is a condition that this ARIMA model can be used.

2.3.1Stationary in variants

One of the efforts to make data into a questionnaire in variety is to use the Box-cox transformation. The Box-Cox transformation estimates the parameter λ which is assigned to the Y variable to get the minimum rounded value. If the minimum round value is obtained with λ = 1, then the data does not need to be transformed because there has been freedom between the various responses and the average response. With the Box-Cox transformation, the normalcy of the spread, the consistency of the error range, and the linearity of the model structure simultaneously can be achieved [21].

2.3.2Stationary in mean

Besides being stationary in variety, the data used must also be stationary in its mean. After the Box-cox transformation is carried out, the unit root test is carried out with an Augmented decay fuller (ADF) method. Suppose the observed data is not stationary at the degree level. A transformation must be carried out using the difference method where the amount of different states the number of orders in the ARIMA model will be used later. The following is the process difference equation:

$$D.Y_t = Y_t - Y_{t-1} = Y_t - L.Y_t = (1 - L)\, Y_t \qquad (4)$$

If the differentiation process has not produced a stationary series, the next level of differentiation (d times) is carried out, namely

$$D2.Y_t = D.Y_t - DY_{t-1} \qquad (5)$$

2.4 ARIMA Equation

The main requirement of the ARIMA model is when AR (p), MA (q), and ARMA (p, q) use stationary time series data so that the ARIMA model can be denoted by [22]:

$$Y_t = ARIMA\,(p, d, q) \qquad (6)$$

If the data used has a seasonal nature like monthly, then a sessional difference is made as in the equation below:

$$S12.Y_t = Y_t - DY_{t-12} = SARIMA\ (p, d, q)(P, D, Q)12 \quad (7)$$

3 Methodology and Discussion

3.1 Methodology

In modeling to make forecasting of the demand or need for LPG in the future, there are several steps in implementing ARIMA, namely:

Figure 4: Step of ARIMA Methode

3.1.1 Data Collection and examination

We are collecting and identifying data that will be used for forecasting.

3.1.2 Determine Stationarity of time series

ARIMA model can only be used for stationary time series. Therefore, the first thing to do is to investigate whether the time series data is stationary or not. If the time series data are not stationary, you have to check with differencing. Some of the data will be stationary.

3.1.3 Model Identification and Estimation

In addition to determining the d value, at this stage, the residual lag value (q) and the dependent lag (p) value are also selected to be used in the model. The main tools used to identify q and p are ACF and PACF (Partial Auto Correlation Function/Partial Autocorrelation Coefficient). The correlogram plots the ACF and PACF values against lag. The partial autocorrelation coefficient measures the closeness of the relationship between Y_t and Y_{t-k} while the effect of time lab 1, 2, 3,… k-1 is considered constant.

Several alternative ARIMA models will be obtained from the results of stationarity identification and ACF and PACF identification. The next step is to estimate the autoregressive and moving average parameters included in the model.

3.1.4 Diagnostic Checking

After estimating and obtaining parameter estimators so that the temporary model can be used for forecasting, it is necessary to conduct a feasibility test on the model. This stage is called diagnostic checking, where at this stage, it is tested whether the model specifications are correct or not.

3.1.5 Forecasting and forecasting evaluation

After the best model is obtained, then forecasting can be done. In many cases, forecasting using this method is more reliable than forecasting using a traditional econometric model.

3.2 Discussion

Based on realization data of LPG consumption for households in Indonesia from January 2016 to December 2020, there is a trend of data that continues to increase but with varying rates of increase as in the figure below:

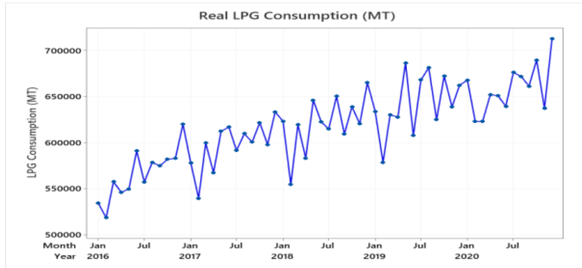


Figure 4: **Plot data Household LPG Demand from Jan 2016 - Dec 2020**

The ARIMA process is then applied to obtain a forecast for the next 24 months from the graph above. As follows:

3.2.1Determine Stationary of time series

Model identification is applied to determine whether the data to be used in the LPG forecasting demand is stationary or not. From Figure 7 above, it can be seen that the data is not stationary either in a variety or on average. In order for the ARIMA method to be used, the data used must be stationary data. Therefore, from the data above, a box-Cox transformation is carried out so that the following data are obtained:

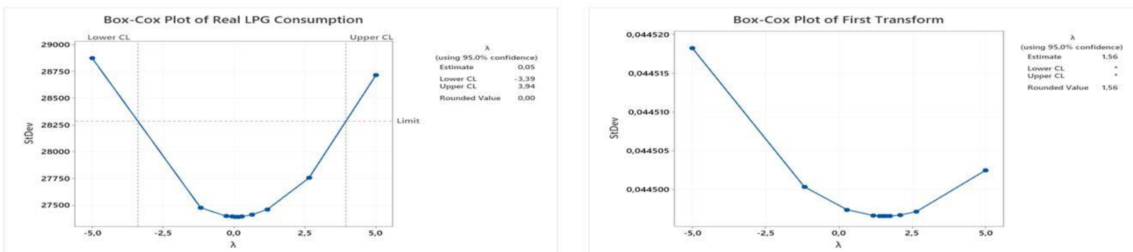


Figure 5: **Box-Cox Plot for data stationer in Variants**

The data above shows that the first stationarity test uses the Box-cox transformation method so that the data used in forecasting is stationary in variants. In Figure 5 left, the rounded value is still zero. This shows that the data is still not stationary to the variants. A second Box-Cox transformation is required. The results can be seen in Figure 7 (right). From this transformation, it can be seen that the rounded value is greater than or equal to one, which indicates that the data is stationary concerning the variant. After that, check the ACF with the results as shown below:

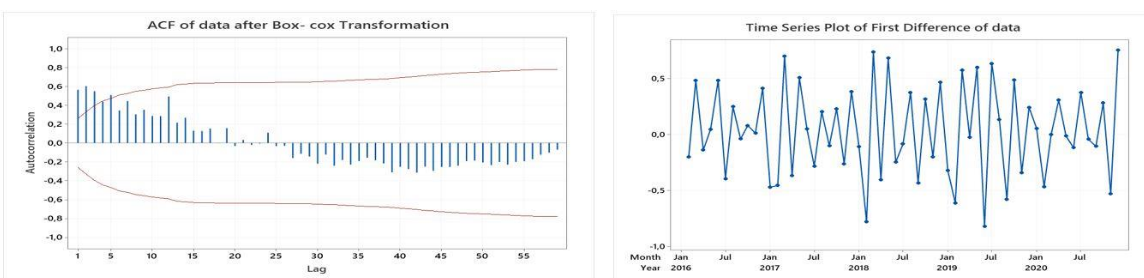


Figure 6: **ACF data (left) and Time-series plot of first difference data**

Figure 6 (left) shows that the first three lags come out of the confidence interval, so the data is not stationary to the average. A first difference is made to obtain stationary data both in terms of the variety and the average. Then the demand data plot is obtained, as shown in Figure 7 (right). For the next stationary test using the unit root test with the Fuller Augmented decay method, the results are shown below. The figure shows that at α = 5% is -2.9237 greater than the statistical value t = -7.0227, and the probability is smaller than 0.05, so the data can be ascertained to be stationary.

Exogenous: Constant
Lag Length: 10 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.022708	0.0000
Test critical values:	1% level	-3.574446	
	5% level	-2.923780	
	10% level	-2.599925	

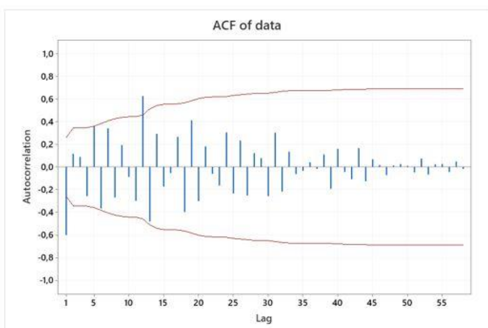
*MacKinnon (1996) one-sided p-values.

***Figure 7:** of augmented Dickey fuller unit rot test on D(Y)*

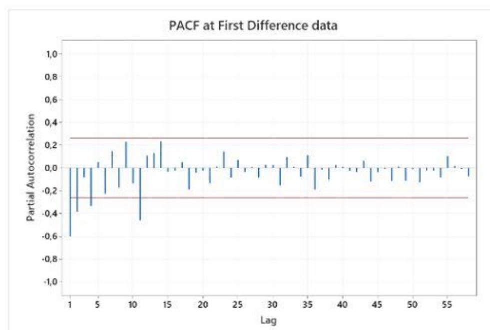
From this test, we can write an overview of the ARIMA model in ARIMA (p, d, q), where the stationarity test level represents the value of d. So the ARIMA model for household LPG Forecasting demand is ARIMA (p, 1, q).

3.2.2Model Identification and estimation

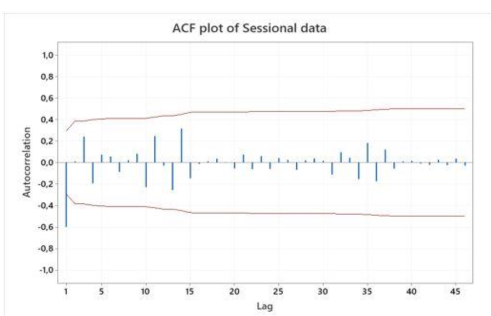
PACF and ACF identification will be carried out at this stage, as shown in Figure 10. In Figure 10 (a) it can be seen that the ACF correlogram shows the first lag out of the confidence interval and repeated at the 12th lag so that it can be said that the observational data are sessional. The professional model analysis with the ACF display is then carried out, as shown in Figure 10 (c). and analysis of the PACF correlogram plot on the Sessional model is also performed.



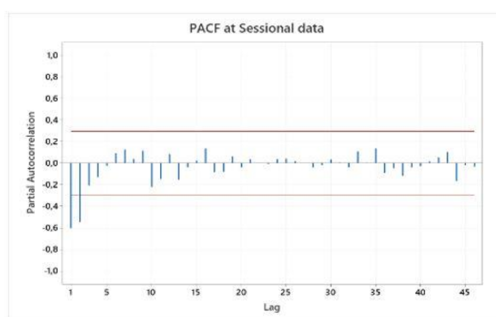
(a)



(b)



(c)



(d)

***Figure 8:** Correlogram of normal and sessional ACF, PACF*

From the picture above, it can be seen that the data model being observed contains seasonal elements so that the general form of the ARIMA model to be used is the professional model with the notation SARIMA (p, d, q) (P, D, Q) 12 with some alternative SARIMA (2,1,1) (1,1,0)12, and SARIMA (2,1,1) (1,1,1)12.

3.2.3Diagnostic Checking

From the two selected models, the final parameter comparisons were carried out where from the figure below, the P-value was obtained less than 0.05 for both models. From the two last parameters obtained, it is found that the most appropriate model for forecasting household LPG demand in Indonesia.

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0,655	0,146	-4,50	0,000
AR 2	-0,455	0,155	-2,93	0,005
MA 1	0,9360	0,0886	10,56	0,000
SMA 12	0,707	0,191	3,69	0,001
Constant	-0,011535	0,000829	-13,91	0,000

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0,744	0,161	-4,62	0,000
AR 2	-0,443	0,155	-2,87	0,006
SAR 12	-0,9789	0,0427	-22,90	0,000
MA 1	0,764	0,157	4,88	0,000
SMA 12	-0,775	0,203	-3,81	0,000
Constant	-0,02199	0,00910	-2,42	0,020

Figure 9:* *Diagnostic check final parameter

From the Comparison table, we obtain that the best forecasting models of SARIMA (2,1,1)(1,1,0)12

3.2.4 Forecasting and Forecasting evaluation

From these criteria, a suitable model was obtained, and forecasting was carried out for data until 2022

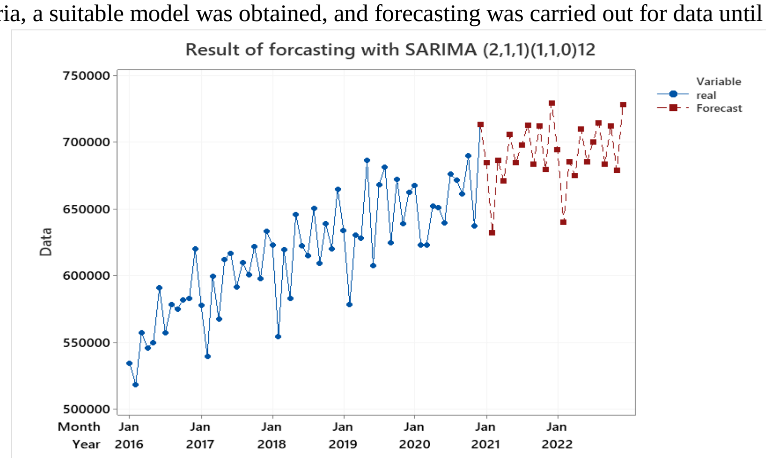


Figure 13: Result of forecasting with ARIMA Models

4 Conclusion

With the implementation of household LPG forecasting demand, it is hoped that it can provide a future picture of the opportunities and challenges in LPG supply in stock planning and infrastructure preparation to support the LPG supply and distribution activities to the community. Based on the result of forecasting, we can see that the trend of consumption LPG in Indonesia increases from year to year and will grow with population growth, economic growth, and new conversion areas. This demand forecast can also be used to plan policies for developing alternative energy sources for households to reduce dependence on imports.

As additional information, the results of this forecasting are not the only source in determining the amount of LPG demand for households in the future. It can become a basis for seeing future phenomena while still paying attention to the development of existing substituted energy sources so that in number, they will complement each other.

REFERENCES

[1] Ediger, V. Ş. and Akar, S. J. E. p. ARIMA forecasting of primary energy demand by fuel in Turkey, 35, 3 (2007), 1701-1708.

[2] Budya, H. and Arofat, M. Y. J. E. P. Providing cleaner energy access in Indonesia through the megaproject of kerosene conversion to LPG, 39, 12 (2011), 7575-7586.

[3] ESDM, K. Hand Book of Energy and economic Statistic in Indonesia, IS SN 2528-3464 (2019).

[4] DEN Neraca Energy Nasional 2019, Dewan Energ Nasional (2019).

[5] Shalih, O. Tipologi Pulau di Indonesia (

[6] Thoday, K., Benjamin, P., Gan, M. and Puzzolo, E. J. E. f. S. D. The Mega Conversion Program from kerosene to LPG in Indonesia: Lessons learned and recommendations for future clean cooking energy expansion, 46 (2018), 71-81.

[7] Masudin, I. J. A. I. J. o. S. and Technology Facility location modeling in multi-echelon distribution system: A case study of Indonesian Liquefied petroleum gas supply chain, 2, 1 (2013), 37-43.

[8] Institute, P. E. Pertamina Energy Outlook, PT Pertamina (Persero) (2020).

[9] Wiyanti, D., Pulungan, R. J. I. J. o. M. and Sciences, N. Peramalan Deret Waktu Menggunakan Model Fungsi Basis Radial (RBF) dan Auto Regressive Integrated Moving Average (ARIMA), 35, 2 (2012).

[10] Meyler, A., Kenny, G. and Quinn, T. Forecasting Irish inflation using ARIMA models (1998).

[11] Erdogdu, E. J. A. E. Natural gas demand in Turkey, 87, 1 (2010), 211-219.

[12] Abdel-Aal, R. and Al-Garni, A. Z. J. E. Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis, 22, 11 (1997), 1059-1069.

[13] Contreras, J., Espinola, R., Nogales, F. J. and Conejo, A. J. J. I. t. o. p. s. ARIMA models to predict next-day electricity prices, 18, 3 (2003), 1014-1020.

[14] El Kafazi, I., Bannari, R. and Abouabdellah, A. *Modeling and forecasting energy demand*. IEEE, City, 2016.

[15] Jiang, S., Yang, C., Guo, J., Ding, Z. J. E. S., Part B: Economics, Planning, and Policy ARIMA forecasting of China's coal consumption, price and investment by 2030, 13, 3 (2018), 190-195.

[16] Hartati, H. J. J. M. S. d. T. Penggunaan Metode ARIMA Dalam Meramal Pergerakan Inflasi, 18, 1 (2017), 1-10.

[17] Bozarth, C. C., Handfield, R. B. and Weiss, H. J. *Introduction to operations and supply chain management*. Pearson Prentice Hall Upper Saddle River, NJ, 2008.

[18] Kurawarwala, A. A., Matsuo, H. J. T. F. and Change, S. Product growth models for medium-term forecasting of short life cycle products, 57, 3 (1998), 169-196.

[19] Miller, D. M. and Williams, D. J. I. J. o. F. Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy, 19, 4 (2003), 669-684.

[20] Makridakis, S., Wheelwright, S. C. and McGee, V. E. J. T. B. A. P. Metode aplikasi dan peramalan (2002).

[21] Dwi Ispriyanti, D. I. J. J. M. Pemodelan Statistika dengan Transformasi Box Cox, 7, 3 (2004), 8-17.

[22] Juanda, B. and Junaidi, J. *Ekonometrika deret waktu: teori dan aplikasi*. IPB Press, City, 2012.