Forecasting of fuel oil supply using the transfer function approach (Case study: PT. Agrabudi Karyamarga Gas Station Division 64.706.07)

A U Labibah*, Y Sukmawaty, and D S Susanti
Statistics Study Program, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Achmad Yani Street Km. 36, Banjarbaru, 70714, South Kalimantan, Indonesia

*aleefalya@gmail.com

Abstract. In everyday life, fuel oil is quite important. The need for fuel is increasing every day, which means that the supply of fuel oil must keep up with the demand. As a reason, we require a way for predicting future fuel needs. The forecasting method is one that is frequently utilized. Forecasting is a method for predicting future conditions based on historical data. The transfer function approach is one way to forecast data with several variables in time series analysis. The objective of this research is to estimate the parameters of the transfer function model and use a transfer function approach to predict the movement of fuel, particularly pertalite. The parameter estimation results in this research are $\hat{\omega}_0 = 0.033; \hat{\omega}_1 = -0.0358; \hat{\delta}_1 = 0.0627; \hat{\delta}_2 = -0.9713; \hat{\delta}_3 = 1; \hat{\theta}_1 = -0.9141$, and the forecast value for the 214th period is 8762.61, based on the data used, namely for 213 periods starting from the 1st period until the 213th period.

1. Introduction
Every day, fuel oil plays a crucial role in people's lives. Fuel's existence can assist humans in reaching their everyday requirements. The demand for fuel fluctuates from time to time. The supply of fuel oil must be increased in tandem with the rise in fuel. To put it another way, the supply of fuel must be able to keep up with the people's ever-increasing demands. As a result, it is vital to perform research in order to forecast future fuel oil demand.

Forecasting, often known as time series analysis, is one way for predicting future fuel requirements. The basic concepts of time series analysis is that the factors that affect the pattern of the data set in the past, now tend to be much the same in the future. Thus doing the analysis of time series aims to identify those factors to assist researchers in making decisions. The Time series models that are commonly used is Autoregressive (AR), Moving Average (MA) and the combination of Autoregressive Integrated Moving Average (ARIMA) [1]. The model is used to forecast a single-variable time series (univariate) [2]. Meanwhile, because the ARIMA model cannot be used to solve data problems with more than one variable (multivariate), another model is required to solve data problems with more than one variable.

The transfer function approach is one of the multivariate approaches for evaluating time series data. The transfer function technique, according to [2], is a forecasting method that combines various properties of regression analysis models and ARIMA. As a result, the distribution pattern of fuel oil,
particularly pertalite, from day to day will be explored in this research with the objective of predicting the supply of fuel oil required in the future, and it will be tested using a transfer function approach.

2. Machine learning

2.1 Time series
A time series is a collection of observations made in the order of correlated observations. As a result, the time series' observational variable is associated with the variable at an earlier time [2]. In time series analysis, an ARIMA model predicts a dependent variable by regressing its own lagged (or previous) values only [3]. The number of days, weeks, months, and years between observations must be the same. Time-series analysis methods, including autoregressive, moving average, and autoregressive integrated moving average (ARIMA) have been widely used in the forecasting of time series [4].

2.2 Autoregressive concept
The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are the most significant tools in time series analysis for identifying a model from the data to be forecasted. The term "autocorrelation" refers to the correlation between observational data from a time series [2]. Then after, the covariance between $X_t$ and $X_{t−k}$ can be stated as follows:

$$y_k = Cov(X_t, X_{t−k}) = E[(X_t − μ)(X_{t−k} − μ)]$$  \hspace{1cm} (1)

and autocorrelation, between $X_t$ and $X_{t−k}$, namely:

$$ρ_k = \frac{Cov(X_t, X_{t−k})}{\sqrt{Var(X_t)Var(X_{t−k})}} = \frac{y_k}{\gamma_0}$$  \hspace{1cm} (2)

The partial regression coefficient $φ_{11}$ is defined as the first partial autocorrelation [5]:

$$X_t = φ_{11}X_{t−1} + a_t$$  \hspace{1cm} (3)

2.3 Stationary
The term "stationarity" refers to the absence of growth and decline in a data set. Because it can decrease the model, time series data stationarity is a crucial criterion [6]. If a time series model is stationary in the mean and variance, it is said to be stationary [7]. When the mean and variance of a time series are both constant, the data is said to be stationary. If there is no significant growth or decrease in the data, it is said to be stationary. The ACF and PACF plots also show that the system is stationary. The data is said to be stationary if the correlation coefficient value declines rapidly as the lag rises [8]. The identification of the ACF plot can be used to detect stationarity in the mean, whereas the Box-Cox test can be used to discover stationarity in variance [9].

2.4 ARIMA Box-Jenkins method
The approach method [7] is another name for the ARIMA Box-Jenkins method. This method implies that the series value is generated by an incomprehensible stochastic (random) process [10]. The Autoregressive (AR), Moving Average (MA), and Autoregressive Integrated Moving Average (ARIMA) models are the three types of Box-Jenkins models (ARIMA). ARIMA models are generally denoted ARIMA (p, d, q) where parameters p, d, and q are non-negative integers, p is the order of the Autoregressive model, d is the degree of differencing, and q is the order of the Moving-average model [11]. AR (p) is an autoregressive model with order p that satisfies the equation:

$$X_t = φ_1X_{t−1} + φ_2X_{t−2} + \cdots + φ_pX_{t−p} + a_t$$  \hspace{1cm} (4)

MA(q) denotes the MA model with order q, as follows:

$$X_t = a_t − θ_1a_{t−1} − θ_2a_{t−2} − \cdots − θ_qa_{t−q}$$  \hspace{1cm} (5)

ARIMA (p, d, q) is the time series model with the order p, d, and q [12]:

$$X_t = φ_1X_{t−1} + φ_2X_{t−2} + \cdots + φ_pX_{t−p} + a_t$$  \hspace{1cm} (6)

$$X_t = a_t − θ_1a_{t−1} − θ_2a_{t−2} − \cdots − θ_qa_{t−q}$$  \hspace{1cm} (7)

$$X_t = φ_1X_{t−1} + φ_2X_{t−2} + \cdots + φ_pX_{t−p} + a_t − θ_1a_{t−1} − θ_2a_{t−2} − \cdots − θ_qa_{t−q}$$  \hspace{1cm} (8)
\[ \phi_p(B)(1 - B)^d X_t = \theta_0 + \theta_q(B) a_t \]

(6)

with:
- \( \phi_t \) = autoregressive coefficient
- \( \theta_t \) = models parameter
- \( a_t \) = noise series
- \( B \) = operator backshift

2.5 Transfer function

According to [4], there is an output series \((Y_t)\) that is supposed to be affected by the input series \((X_t)\) and another series dubbed the "noise" \((n_t)\) disturbance series in the transfer function technique. It can obtain output variables based on input variables at different time periods [13]. The input series \((X_t)\) has an impact on the output series \((Y_t)\) via a transfer function that spreads out the impact of \((X_t)\) over time.

The transfer function's general model is as follows:

\[ Y_t = v(B)X_t + n_t \]

(7)

with:
- \( Y_t \) = output series (last stock)
- \( X_t \) = input series (first stock)
- \( n_t \) = noise series
- \( v(B) = (v_0 + v_1 B + \cdots + v_k B^k) \) where \( k \) is the transfer function's order

(1) Root mean square error (RMSE)

The Root Mean Square Error (RMSE) is another name for the estimated error rate, and the smaller (closer to 0) the RMSE number, the more accurate the estimation findings [4]. The following formula can be used to calculate the RMSE value [14]:

\[ RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{t=1}^{n}(X_t - \hat{X}_t)^2}{n}} \]

(8)

3. Method

3.1 Data source

Secondary data received from PT. Agrabudi Karyamarga Gas Station Division 64,706.07 will be used in this investigation. A petrol station with the code 64706.07 may be found on Jl. A Yani, Km 23.9 No. 1 A Platform Ulin, Banjarbaru City, South Kalimantan. From February 1, 2020, to August 31, 2020, 213 sales of fuel oil (BBM) data were acquired in the form of daily pertalite data.

3.2 Research variables

The input series is described in terms \(X_t\), while the output series is described in terms \(Y_t\). The variable \(X_t\), in this study represents the sale of fuel oil, specifically the first stock of pertalite on the first shift (at 07.00 WITA), and the variable \(Y_t\), represents the last stock of pertalite on the first shift (14.00 WITA).

3.3 Research procedure

The following procedures will be implemented in order to conduct this research:

1. Descriptive statistics are a type of statistical analysis that is used to describe something.
2. Establish, test, and assume stationarity.
3. Separate the data into two categories: training and testing.
4. Parameter estimation for the transfer function model.
5. Transfer function model diagnosis test
6. Forecasting with the transfer function model.
7. Look for the lower RMSE number to choose the optimal model.
8. Predicting the transfer function model’s future performance.
9. Draw conclusions based on the findings of previous research.

4. Result and discussion

4.1 Transfer function

The following formula can be used to compute the transfer function:

\[ Y_t = v(B)X_t + n_t \]

with \( v(B) = \frac{\omega(B)}{\delta(B)} \) dan \( n_t = \frac{\theta(B)}{\phi(B)} \alpha_t \)

As a result, the transfer function model looks like this:

\[ y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} \alpha_t \]  

(9)

Application of the transfer function in fuel supply: a case study

4.1.1. Descriptive data.

Table 1. Descriptive statistics

|          | Mean  | Variance     | St. Dev | Min. | Max.  |
|----------|-------|--------------|---------|------|-------|
| First stock | 10958.88 | 11757001.21 | 3428.84 | 1678 | 19834 |
| Last     | 10907.69 | 11930809.85 | 3454.10 | 1678 | 19834 |

The mean of first stock of pertalite is 10958.88 litres (L), with a minimum value 1678 L and a maximum value 19834 L, as shown in table 1. The last stock of pertalite has a mean value 10907.69 L, with a minimum value of 1678 L and a maximum value of 19834 L.

4.1.2. Data stationarity test in variance.

According to the plot in figure 1, the rounded value is already worth 1.00, indicating that the data is stationary in variance.

4.1.3. Data stationarity test in mean.
After differencing 1 time, the ACF plot was on mean stationary, as seen in figure 2(a). The PACF plot was then used to perform stationary detection with data that had been differencing 1 time, as seen in figure 2(b).

4.1.4. Model identification.

First, the data will be separated into two groups: training data (200 data in this case) and testing data (13 data in this case). The PACF plot is cut off after lag 1 as can be seen in Figure 2(b), hence the model is $p = 1$. The ARIMA model $(1,1,1)$ is obtained in Figure 2 by differencing 1 time, $d = 1$, and having a cut off at lag 1 ($q = 1$). The values $\theta_x = 0.3090$ and $\phi_x = 0.9503$ were calculated using the ARIMA model $(1,1,1)$.

4.1.5. Transfer function model identification.

a. Prewhitening input series ($X_t$)

This is implemented by reassembling the tribes and modifying the ARIMA equations [15]. The following results are achieved using ARIMA $(1,1,1)$:

\[
(1 - \phi_x B)x_t = (1 - \theta_x B)\alpha_t \\
\]

\[
(1 - 0.9503 B)x_t = (1 - 0.3090 B)\alpha_t \\
\]

to transform $x_t$ series to (white noise) $\alpha_t$ using the equation:

\[ \alpha_t = x_t - 0.9503 x_{t-1} - 0.3090 \alpha_{t-1} \] (10)

Then, using the input series that has been differencing, set $\alpha_1 = 0$, and you will get:

\[ \alpha_2 = x_2 - 0.9503 x_{2-1} - 0.3090 \alpha_{2-1} = -3346 - 0.9503(0) - 0.3090(0) = -3346 \]

Use the formula in equation (10) to calculate $\alpha_3, ..., \alpha_{200}$.

b. Prewhitening output series ($Y_t$)

The following results are achieved using ARIMA $(1,1,1)$:

\[ \beta_t = y_t - 0.9503 y_{t-1} - 0.3090 \beta_{t-1} \] (11)

Then, using the output series that has been differencing, set $\beta_1 = 0$, and use equation (11) to calculate the values of $\beta_2, \beta_6, ..., \beta_{200}$.

\[ \beta_2 = y_2 - 0.9503 y_{2-1} - 0.3090 \beta_{2-1} = -79 - 0.9503(0) - 0.3090(0) = -79 \]
c. Calculating the cross correlation of $\alpha_t$ dan $\beta_t$ series

The cross-correlation value can be calculated by following the value of $k = 0, 1, 2, 3, \ldots, 24$ for $k = 0$, and then calculating the cross-correlation value as follows:

$$ r_{\alpha\beta}(k) = \frac{c_{\alpha\beta}(k)}{\sigma_{\alpha}(0)\sigma_{\beta}(0)} $$

$$ r_{\alpha\beta}(0) = \frac{c_{\alpha\beta}(0)}{\sigma_{\alpha}(0)\sigma_{\beta}} $$

$$ = \frac{-35659427.36}{(7083.4112)(7074.3363)} $$

$$ = -0.711616828 $$

d. Calculating the impulse response weight appraisal value

For $k = 0$, the impulse response weight is calculated as follows:

$$ v_k = r_{\alpha\beta}(k) \frac{S_{\beta}}{S_{\alpha}} $$

$$ v_0 = r_{\alpha\beta}(0) \frac{S_{\beta}}{S_{\alpha}} $$

$$ = -0.711617 \frac{7074.3363}{7083.4112} $$

$$ = -0.710705 $$

e. Identifying $(r, s, b)$ values for the transfer function model

The transfer function model's most important parameters are $(r, s, b)$, where $r$ represents the degree of function $\delta(B)$, $s$ represents the degree of function $\omega(B)$, and $b$ represents the absolute magnitude of the delay before the output series is influenced input series. As a result of data processing, the value $(r, s, b) = (2, 2, 5)$ is calculated using the equation:

$$ y_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-5} + n_t $$

f. First observation of noise series

The noise series value is calculated as follows:

$$ n_t = y_t - v_0 x_t - v_1 x_{t-1} - v_2 x_{t-2} - v_3 x_{t-3} - v_4 x_{t-4} $$

$$ n_{25} = y_{25} - (-0.710707)x_{25} - (0.3187)x_{24} - (-0.0763)x_{23} $$

$$ = (1118) - (-0.7107)(-1618) - (-0.2309)(-1805) - (-0.0763)(0) $$

$$ = 403.57 $$

Use the same equation to calculate the following value.

g. Noise series ARIMA model identification

Based on the ARIMA auto test, the best ARIMA for noise series is ARIMA $(0, 1, 1)$, which has the lowest RMSE value of 2984.909. The model is then created as follows:

$$(1 - B)n_t = (1 - \theta_1 B)\alpha_t$$

so $n_t = \frac{(1-\theta_1 B)}{(1-B)} \alpha_t$

h. Estimation of model parameters

The following is a model of the transfer function that has been determined tentatively:

$$ y_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-5} + \frac{(1-\theta_1 B)}{(1-B)} \alpha_t $$

The parameters $\omega_0, \omega_1, \omega_2, \delta_1, \delta_2, \theta_1$ will be evaluated. The following are the results $\hat{\omega}_0 = 0.033$; $\hat{\omega}_1 = -0.0358$; $\hat{\omega}_2 = 0.0627$; $\hat{\delta}_1 = -0.9713$; $\hat{\delta}_2 = 1$; $\hat{\theta}_1 = -0.9141$.

As a result, when the parameter values are placed into the transfer function model, the result will be as follows
Calculations can be made to forecast the 201st forecast: 

\[ y_t = \frac{(0.0330 - (0.0358)B - 0.0627B^2)}{(1 - (-0.9713)B - 1B^2)} x_{t-5} + \frac{(1 - (-0.9141)B))}{(1 - B)} a_t \]

\( y_{t} \) is the forecast for the 201st period, \( y_{t-1} \) is the value of the 200th period, and \( a_t \) is the value of the 201st period.

**i. Calculating \( a_t \) for transfer function model**

Use the following equations to calculate the value of \( a_t \):

\[
\begin{align*}
    n_t &= (1 - \theta_1 B) a_t \\
    n_t - B n_t &= a_t - \theta_1 B a_t \\
    n_t - n_{t-1} &= a_t - \theta_1 a_{t-1} \\
    a_t &= n_t - n_{t-1} + \theta_1 a_{t-1} 
\end{align*}
\]

Kemudian asumsikan \( (a_1, \ldots, a_{24}) = 0 \) untuk mencari nilai \( a_{25}, \ldots, a_{200} \).

\[
\begin{align*}
    a_{25} &= n_{25} - n_{24} + \theta_1 a_{24} \\
    &= 403.57 - 0 + (-0.9141)(0) \\
    &= 403.57
\end{align*}
\]

**j. Using the transfer function model for forecasting**

The transfer function model in the simplified equation below will be used to produce forecasting:

\[
y_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-5} + \frac{(1 - \theta_1 B)}{(1 - B)} a_t
\]

becomes:

\[
\hat{y}_t = -k_1 y_{t-1} - k_2 y_{t-2} - k_3 y_{t-3} + l_0 x_{t-b} + l_1 x_{t-b-1} + l_2 x_{t-b-2} + \ldots + l_3 x_{t-b-3} + m_0 a_t + m_1 a_{t-1} + m_2 a_{t-2} + m_3 a_{t-3}
\]

where the values of the above-mentioned equations \( k, l, \) and \( m \) are:

\[
\begin{align*}
    k_1 &= 1 - \delta_1 = -0.0287 \\
    k_2 &= \delta_1 - \delta_2 = -1.9713 \\
    k_3 &= \delta_2 = 1 \\
    l_0 &= \omega_0 = 0.0330 \\
    l_1 &= -\omega_0 - \omega_1 = 0.0028 \\
    l_2 &= \omega_1 - \omega_2 = -0.0985 \\
    l_3 &= \omega_2 = 0.0627 \\
    m_0 &= 1 \\
    m_1 &= -\delta_1 - \theta_1 = -0.0572 \\
    m_2 &= -\delta_2 + \theta_1 = -1.9141 \\
    m_3 &= \theta_1 \delta_2 = -0.9141
\end{align*}
\]

So, using equation (20), calculate the value of the 201st forecast:

\[
\hat{y}_{201} = -k_1 y_{200} - k_2 y_{199} - k_3 y_{198} + l_0 x_{196} + l_1 x_{195} + l_2 x_{194} + l_3 x_{193} + m_0 a_{201} + m_1 a_{200} + m_2 a_{199} + m_3 a_{198}
\]

\[
= (-0.0287)(1527) - (-1.9713)(2865) - \ldots + (-1.9141)(3978.44) + (-0.9141)(2164.14) \\
= -1340.41
\]

Calculations can be made to forecast the 201st data as follows:

\[
\begin{align*}
    \text{last stock value for the 201st period} &= \text{(ending stock value for the 200th period)} + \hat{y}_{201} \\
    \hat{y}_{201} &= 12876 + (-1340.41) \\
    &= 11535.59
\end{align*}
\]

Afterwards, Table 2 below shows the next forecasting value.
Table 2. 201st to 213th Forecast Values

|   | $\hat{Y}_t$ | $Y_t$ | Error       | RMSE  |
|---|-------------|-------|-------------|-------|
| 201| 11535.59    | 8613  | -2922.59    |       |
| 202| 10195.18    | 11548 | 1352.82     |       |
| 203| 11245.95    | 12968 | 1722.05     |       |
| 204| 5041.43     | 4978  | -63.43      |       |
| 205| 6092.20     | 7898  | 1805.8      |       |
| 206| 7432.61     | 8702  | 1269.39     |       |
| 207| 8483.38     | 9837  | 1353.62     |       |
| 208| 12275.38    | 11215 | -1060.38    |       |
| 209| 14918.38    | 13290 | -1628.38    | 1397.75|
| 210| 8713.86     | 8259  | -454.86     |       |
| 211| 8659.36     | 7950  | 709.56      |       |
| 212| 9132.18     | 9138  | -5.82       |       |
| 213| 4722.61     | 3729  | 930.61      |       |

As a result, the forecasted value for the 214th data is 8762.61.

5. Conclusion

It can be concluded, based on the results of the research in the discussion, that:

(1) In this case, the transfer function model is:

$$y_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-5} + \frac{(1 - \theta_1 B)}{(1 - B)} d_t$$

with the following procedure for obtaining the estimated value of the transfer function model parameters:

$$y_t = \frac{(0.0330 - (-0.0358)B - 0.0627B^2)}{(1 - (-0.9713B) - 1B^2)} x_{t-5} + \frac{(1 - (-0.9141B))}{(1 - B)} a_t.$$  

(2) The forecasted value of fuel oil supply for the 214th period is 8762.61 L.

References

[1] Lay P A I A E G 2019 Int. J. Innov. Creat. Chang. 5 147–61
[2] Wei W S 1991 Technometrics 33 108
[3] Ho W T and Yu F W 2021 P. J. Build. Eng. 33
[4] Zhao L, Mbachu J and Liu Z 2020 Math. Probl. Eng. 2020
[5] Bovas Abraham J L 2005 Statistical Methods For Forecasting (Canada: John Wiley & Sons Inc Publication)
[6] Pascarianti N 2011 Peramalan Harga Premium Non Subsidi degan Model Fungsi Transfer
[7] Makridakis S, Wheelwright SC H R 1997 Forecasting: Methods and Applications Third Ed. 1–632
[8] Kutner M H 2005 Applied Linear Statistical Models (America: McGraw-Hill)
[9] Box G E P and Cox D R 1964 J. R. Stat. Soc. Ser. B 26 211–43
[10] Safriani N, El Husna N and Rizkya, R 2015 J. Agroindustri 5 85-94.
[11] Zhu J 2015 Appl. Comput. Math. 4 456
[12] Yuan S and Shen Y 2020 J. Phys. Conf. Ser. 1650 032095
[13] Cook T 1979 Ann. Reg. Sci. 13 81–92
[14] Wang W and Lu Y 2018 IOP Conf. Ser. Mater. Sci. Eng. 324
[15] Listya D and Sutijo B 2013 J. Sains Dan Seni Pomits 2 D-271-D-274