Forecasting of primary energy consumption data in the United States: A comparison between ARIMA and Holter-Winters models

A. Rahman, and A. S. Ahmar

Forecasting of Primary Energy Consumption Data in the United States: a comparison between ARIMA and Holter-Winters Models

A Rahman[1,a)] and A S Ahmar[2,b)]

[1]*Departement of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Daeng Tata Kampus UNM Parangtambung, 90223 Tamalate, Makassar, Indonesia*
[1]*Departement of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Daeng Tata Kampus UNM Parangtambung, 90223 Tamalate, Makassar, Indonesia*

Corresponding author: [a)]abdul.rahman@unm.ac.id
[b)]ansarisaleh@unm.ac.id

Abstract. This research has a purpose to compare ARIMA Model and Holt-Winters Model based on MAE, RSS, MSE, and RMS criteria in predicting Primary Energy Consumption Total data in the US. The data from this research ranges from January 1973 to December 2016. This data will be processed by using R Software. Based on the results of data analysis that has been done, it is found that the model of Holt-Winters Additive type (MSE: 258350.1) is the most appropriate model in predicting Primary Energy Consumption Total data in the US. This model is more appropriate when compared with Holt-Winters Multiplicative type (MSE: 262260,4) and ARIMA Seasonal model (MSE: 723502,2).

INTRODUCTION

Forecasting is a very important matter in business or economy. Forecasting can help a person to determine the trend of a data and can provide an information about how fluctuations in a data so that users, in this case, policy makers can use it as one of the tools for decision making. Forecasting is an effort that is used to know what will happen in the future. Random data fluctuation requires a policy maker to think about how is the strategy that should be done in the future in order to give maximum results and in line with expectations.

Data changes from time to time can be predicted or predictable. Forecasting is an activity to predict what will happen in the future. Its data is usually associated with time, so that it can be stated that the data is the time series data. To predict a time-series data, the time series forecasting method is used i.e. ARIMA Box-Jenkins method, Holt-Winters method, and linear regression model, either without season or with season. There are two types of time series forecasting that are often used in data forecasting i.e. ARIMA [1] [2] and Holt-Winters [3] [4]. The comparison of forecasting between ARIMA and Holt-Winters has been studied by several researchers including: Da Veiga, et. al. 2014 discussed the comparison of performance between Holt-Winters model and ARIMA in predicting the group of degradable dairy products, the result of this study is that the Holt-Winters method is better than ARIMA model in terms of performance [5]. The next research is a study conducted by Omane-Adjepong, Oduro, Oduro in 2013. This study discussed the best approach between seasonal ARIMA model and Holt-Winters in forecasting short-term inflation in Ghana and the results of this research indicate that the seasonal ARIMA model is the most appropriate model in predicting short-term inflation in Ghana [6]. And the third research was conducted by Mini, Kuriakose and Sathianandan in December 2015, this research discusses about Catch Modeling per unit effort (CPUE) series for the fishery along the northeast coast of India by comparing 3 methods i.e. Holt-Winters, ARIMA and NNAR method. Based on the comparison of these 3 methods, Holt-Winters method is the method that gives the

best accuracy level compared to ARIMA and NNAR models. Based on a review of some research on forecasting method between Holt-Winters and ARIMA Model, this research will discuss about the suitable model (Model ARIMA or Holt-Winters) in predicting data of Primary Energy Consumption Data of U.S.

ARIMA Model

If the observations in the time series can be predicted with certainty, and does not require further investigation, it is called *deterministic* time series and if observations can only show the structure of the probabilistic of the future condition of a time series, it is called *stochastic* time series [7].

In forming analysis model of a time series by assuming that the data is in a stationary state. The time series is said to be stationary if there is no change in tendency in the mean and changes in variance. The stationary time series of relative neither extreme increase or decline in value of data or data fluctuations is at about a constant average value.

The stationary can be seen by using the time series diagram i.e. a scatter diagram between the Z changing values and T time. If the time series diagram fluctuates around a line that parallels to the time axis (t) then the stationer is said to be in the average. If the condition of stationary is in the unmet average, it needs a differentiating or *differencing* process [8].

The process of *differencing* in the first order is the gap between the *t* data and *t – 1* data, $\Delta Z_t = Z_t - Z_{t-1}$. The form for the second-order *differencing* is,

$$\Delta^2 Z_t = \Delta Z_t - \Delta Z_{t-1}$$
$$= (Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2})$$
$$= Z_t - 2Z_{t-1} + Z_{t-2}$$

If the stationary condition in variance is not met, *Box* and *Cox* [9] introduced the *rank transformation,* namely $Z_t^{(\lambda)} = \dfrac{Z_t^{(\lambda)} - 1}{\lambda}$, ($\lambda$ is a parameter), which known as *the Box-Cox transformation.* Here are some provisions to stabilize the variance:

1. Transformation may be done only for the positive Z_t series.
2. Transformation is done before doing differencing and modeling of time series.
3. Value is selected based on the sum of squares error (SSE) of the transformed series. The smallest SSE values provide the most constant variance results.
4. Transformation does not only stabilize variance, but it can also normalize distribution.

Autoregressive Integrate Moving Average (ARIMA) models have been studied by George Box and Gwilym Jenkins in 1976 and their names are then often synonymous with ARIMA process that is applied to the analysis of *time series.* In general ARIMA model was rendered with the notation ARIMA (p, d, q), where p denote the order of the process *autoregresive (AR),* d express distinction *(differencing),* and q express order of the *moving average* (MA). The Autoregressive Model (AR) was first introduced by Yule in 1926 [10] and later developed by Walker in 1931 [11], while the Moving Average (MA) model was first used by Slutzky in 1937 [12]. And in 1938, then Wold produced the theoretical underpinnings of the ARMA combination process [13]. And these combinations are then used frequently.

Box and Jenkins have effectively reached agreement on relevant information needed to understand and use ARIMA models for a time series of one variable [14].

A process (Z_t) is said to follow a mixed-auto-regressive-moving average or ARMA (p, q) model (Eq. 1) if it meets:

$$\phi_p(B)Z_t = \theta_q(B)a_t \tag{1}$$

where $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)$ (for AR(p))

and $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$ (for MA(q))

If there is a differencing then the model becomes ARIMA as show in Eq. 2 [14]:

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)a_t \qquad (2)$$

where $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)$ (for AR(p)), $(1 - B)^d$ (for non-seasonal differencing) and $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$ (for MA(q)) [14].

In performing Time Series data forecasting methods ARIMA (p, d, q) there are so-called steps or stages. The stages in doing the forecasting is [14].

1. Model Identification
    Model identification is done to see the meaning of autocorrelation and data stationary, so whether or not necessary transformation or process differencing is done. From this stage, it is obtained a temporary model that will be done its model testing process whether it is appropriate or not on the data.

2. Model Assessment and Testing
    After the process of identifying the model, the next step is the assessment and testing model. At this stage, it is divided into two parts: parameter assessment and diagnostic examination model.
    a. Parameter Assessment
        Having obtained one or more interim models then the next step is to look for estimates for the parameters in the model.
    b. Model Diagnostics
        Diagnostic checking is passed to check whether the estimated models fit enough or adequately with existing data. *Diagnostic checking* based on residual analysis. The basic assumption of the ARIMA model is that residuals are normal independent distributed random variables with zero mean constant variance.

Holt-Winters

Holt-Winters Exponential Model (HW) is an extension of Holt's linear model. The Holt-Winters Exponential Model was proposed in 1960 by Holt and Winters, who later named this method based on its discoverers Holt and Winters. This model is an extension of the Holt linear equation model which directly captures the seasonality because the Holt linear model can only predict linear trend data. The Holt-Winters model is often used in a time series that shows seasonal patterns of increasing or decreasing seasonality. Model HW basically has three smoothing equations are *level, trend,* and *seasonal.* Each equation of HW is designed in order to deal with the effect of *the level, trend,* or *seasonal* in a time series.

This HW method can also be used to predict time series in the short, medium and long term. This HW forecasting method is different from other forecasting methods in the sense that this model is independent of the match model of any statistical modeling technique but it uses an iterative approach in generating approximate values. In general, there are two versions of the Holt-Winters smoothing model that are additive and multiplicative models. The HW model for seasonal multiplication does not apply if data from a time series contains a null value or a negative value. The HW model equation is presented as follows;

METHOD

The data to be forecast is data on the Primary Energy Consumption Total. This data is obtained from the US Energy Information Administration (EIA) website (https://www.eia.gov). Primary Energy Consumption Total data is the total consumption data of primary energy consumed by the residential sector, primary energy consumed by the commercial sector, primary energy consumed by the industrial sector, primary energy consumed by the transportation sector, primary energy consumed by the electric power sector, and energy consumption balancing item. Data from this research from January 1973 to December 2016. This data will be processed using R Software.

To see the comparison of the accuracy level of ARIMA Model and Holt-Winters Model, four criteria were used e.g. mean absolute error (MAE), residual sum of squares (RSS), mean squared error (MSE), and root mean square error (RMSE). In general, to obtain the best model will be compared the value of the four criteria. The model is said to be good if the value of MAE, RSS, MSE, and RMSE has the smallest value.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where:

y_i = actual data;

$\hat{y}_i$ = predicted data.

RESULT AND DISCUSSION

Data modeling of a time series data is usually influenced by previous data. Time series data are analyzed to understand the past and predict the future. In order for a data modeling to produce good results then the data should be divided into 2 groups i.e. data groups used to understand the past (data train) and data groups used for validation of modelers from the data train results in order to obtain predictions of the future that is more Directed (data cross validation). In modeling the Primary Energy Consumption Total data, 70% of the data will be used as data train and 30% of the remaining data will be used as cross validation data.

Before doing the modeling process, the data is first decomposed in order to see whether the data contained the trend and seasonal. This decomposition process is needed to assist the researcher in determining a suitable forecasting analysis method. The results of data decomposition are as shown in Fig. 1 and 2.

From the conducted analysis, it is found ARIMA $(0,0,2)$ $(1,0,1)^{12}$ and Holt-Winters model with $\alpha = \beta = \gamma = 1$. The accuracy of the comparison table between the ARIMA model and Holt-Winters Model is shown in Table 1.

TABLE 1. The accuracy of the comparison between the ARIMA and Holt-Winters Model

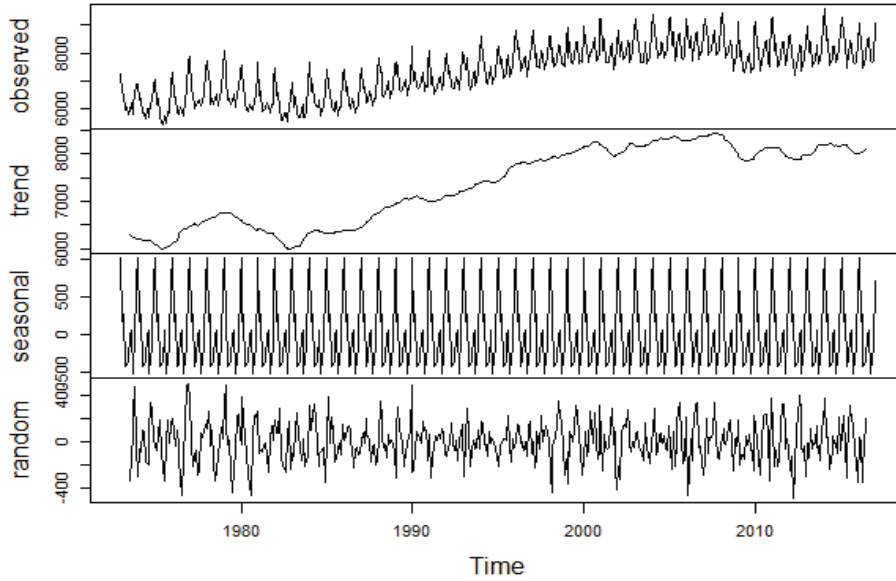| Type	MAE	RSS	MSE	RMSE
HW Additive	420.9939	41077663	258350.1	508.2815
HW Multiplicative	425.2001	41699399	262260.4	512.1136
Seasonal ARIMA	700.2497	115036847	723502.2	850.5893

FIGURE 1. Data decomposition of Primary Energy Consumption Total.
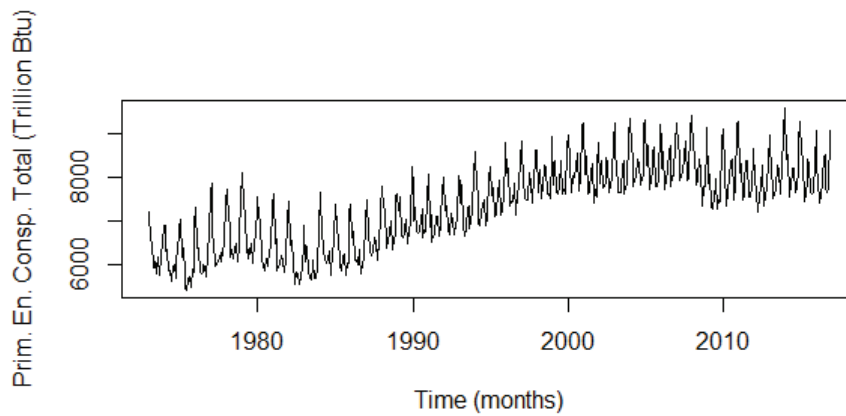

FIGURE 2. Plotting of Primary Energy Consumption Total.

From the table above, it can be seen that the MAE, RSS, MSE, and RMSE values of the HW model are smaller than the ARIMA model in other words that the HW model is more precise in predicting the Primary Energy Consumption Total data in the US. Of the two appropriate HW models then the most appropriate one is HW additive model. In addition to the table, we can also see the comparison of the following Fig. 3 and 4.
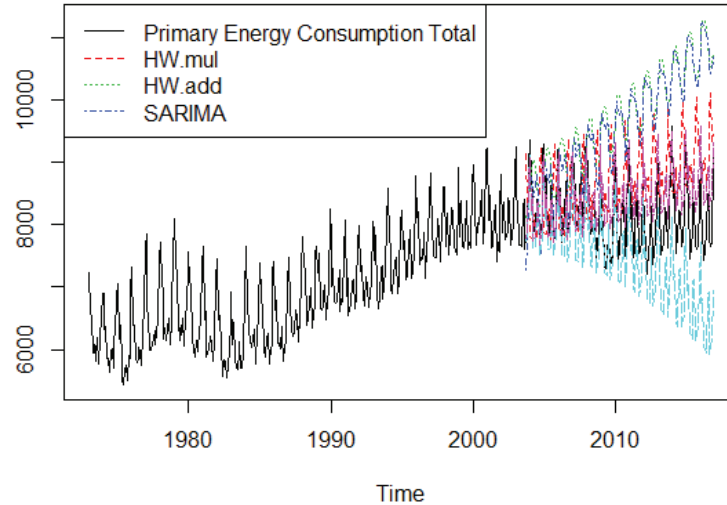
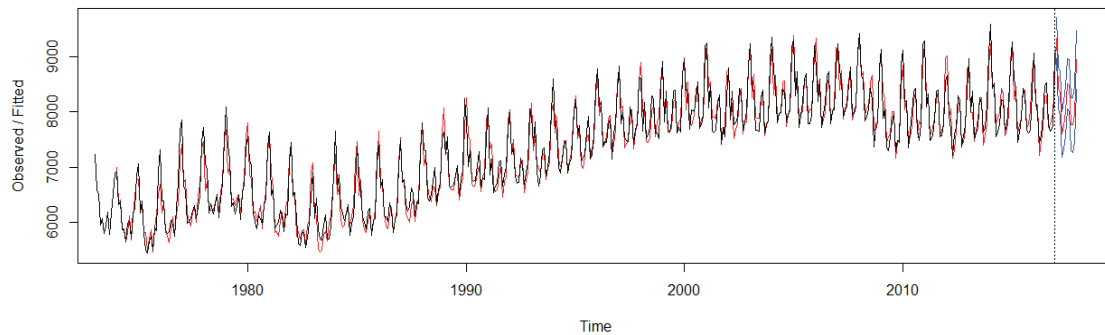**FIGURE 3.** Comparison of result predicting

FIGURE 4. Predicting of Holt-Winters

CONCLUSIONS

This study aims to compare ARIMA Model and Holt-Winters Model based on MAE, RSS, MSE, and RMS criteria in predicting Primary Energy Consumption Total data in the US. From the results of data analysis that has been done, it is found that the model of Holt-Winters Additive type (MSE: 258350.1) is the most appropriate model in predicting Primary Energy Consumption Total data in the US. This model is more appropriate when compared with Holt-Winters Multiplicative type (MSE: 262260,4) and ARIMA Seasonal model (MSE: 723502,2).

ACKNOWLEDGMENTS

REFERENCES

1. B. Guha and G. Bandyopadhyay, "Gold Price Forecasting Using ARIMA Model," J. Adv. Manag. Sci. **4**, 2, pp. 117–121 (2016).
2. G. P. Zhang, Neurocomputing **50**, 159–175 (2003).
3. P. Deossa, A. Márquez, and J. Espinosa, "Integration of economic MPC, energy load and price estimation with Holt Winters models," (Transmission & Distribution Conference and Exposition-Latin America, IEEE 2014), pp. 1–5.
4. D. S. Dhakre, K. A. Sarkar, and S. Manna, Int. J. Bio-res. Env. Agril. Sci, 2016.
5. C. P. Da Veiga, C. R. P. Da Veiga, A. Catapan, U. Tortato, and W. V Da Silva, WSEAS Trans. Bus. Econ. **11**, 608–614 (2014).

6.   M. Omane-Adjepong, F. T. Oduro, and S. D. Oduro, Int. J. Business, Humanit. Technol. **3**, 69–79 (2013).
7.   A. S. Ahmar, S. Guritno, and Abdurakhman, "Pendeteksian dan Pengoreksian Data Yang Mengandung Additive Outlier (Ao) Pada Model Arima(p,d,q)" (Universitas Gadjah Mada, 2013).
8.   W. W. S. Wei, "Time Series Analysis : Univariate and Multivariate Methods, 2nd ed." (New York, Pearson Addison Wesley, 2006).
9.   G. E. P. Box and D. R. Cox, J. R. Stat. Soc. Ser. B, 211–252 (1964).
10. G. U. Yule, J. R. Stat. Soc., 89(1), 1–63 (1926).
11. G. Walker, Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character, 131, 518–532 (1931).
12. E. Slutzky, Econom. J. Econom. Soc., 105–146 (1937).
13. H. Wold, "A Study in the Analisys of Stationery: Time Series" (Almqvist & Wiksells Boktryckeri, 1938).
14. S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, "Forecasting methods and applications" (John Wiley & Sons, 2008).