Abstract

Nowadays, there is an increasing demand for electricity however overproduction of electricity lead to wastage. Therefore, electricity load forecasting plays a crucial role in operation, planning and maintenance of power system. There were many ways that have been employed towards electricity load forecasting. The present study was designed to study the effect of deseasonalizing the electricity load data in forecast performance and to compare the methods of Exponential Smoothing and Box-Jenkins in electricity load forecasting. The daily seasonality in electricity load data was removed and the forecast methods were employed on both the seasonal data and non-seasonal data. Holt Winters method and Seasonal-Autoregressive Integrated Moving Average (SARIMA) methods were used on the seasonal data. Meanwhile, Simple and Double Exponential Smoothing methods as well as Autoregressive Integrated Moving Average (ARIMA) methods were used on the non-seasonal data. Previous studies employed similar approach in electricity load forecasting for neural network method. This paper focused on the traditional time series forecasting method. The forecast accuracy measures used for this research were mean absolute error (MAE) and mean absolute percentage error (MAPE). The results revealed that both Exponential Smoothing method and Box-Jenkins method produced better forecast for deseasonalised data. Besides, the study proved that Box-Jenkins method was better in forecasting electricity load data for both seasonal and non-seasonal data.

Keywords: Box-Jenkins; Deseasonalisation; Exponential Smoothing; Forecast Accuracy

1. Introduction

Electricity supply systems regard load demand as a sensitive factor as the demand has to be balanced with the supply as to avoid overproduction of load or interruption in electricity load supply. Thus, it is important to properly plan and manage the electricity load production. An accurate forecast can help to minimize the operating cost and allow good management of electricity load [1]. Earlier, electricity load forecasting was done by using simple methods as the demands for electricity load were very predictable [2]. As the electricity load data became more complex, the more critical forecasting techniques were used to predict the load. Forecasting of electricity load data these days have become a challenge due to the complexity of the data. The electricity demand also has been an important part of energy equation. Therefore, having a good understanding about the electricity demand may help to produce better prediction on the data. The information should be gathered from every related perspective in order to employ better statistical techniques on the data.

The seasonal components present in the data served no function in analyzing the data but make it harder to forecast them. Seasonality is the occurrences of variance at a fixed point of time such as daily, weekly, monthly and yearly. Many factors had contributed to seasonality in data such as holidays, weather and vacation [3]. There were various studies conducted in modeling the seasonality that exist in electricity load forecasting. Seasonal adjustments were made before employing forecasting method on a time series data. This approach had been very useful in predicting the future of a data. Eliminating the seasonality from data is known of deseasonalisation of data. Deseasonalization improved the forecasting accuracy of traditional statistical methods applied to the M-competition [4]. Similarly, neural network method also showed good result when applied to the deseasonalised data [5]. Therefore, deseasonalisation was regarded as a useful method in forecasting time series data.

The aim of this research is to examine the impact of deseasonalisation data on electricity load forecasting using traditional methods that is exponential smoothing method and Box-Jenkins method. These two methods have been used widely in time series forecasting in many areas of study. As example Box Jenkins approach were applied in forecasting of gold price [6], import of palm oil products [7], tourism demand forecasting [8], [9] and [10].

2. Materials and Methods

In order to verify the effect of data deseasonalisation on the forecast performance, two well-known forecasting approaches were used that are Box-Jenkins and exponential smoothing.

2.1 Data

There were two datasets used in this research which were the electricity load data from three stations in Johor (Central City Johor Bahru and Pasir Gudang). These datasets were obtained from Tenaga Nasional Berhad and was recorded in every half hour. The
average electricity load of half hourly data for each day was computed in order to convert the data into daily scale. The datasets were split into in-sample and out-of-sample period for the purpose of estimation and forecast evaluation respectively. The out sample period taken for this research were the last seven days of each dataset while the remaining days were used as the in-sample dataset.

2.2 Desseasonalisation of dataset

The daily seasonality of each dataset was eliminated by unstacking into seven columns of same scale. Then, the datasets that were seasonal and non-seasonal were forecasted separately by using exponential smoothing method as well as Box-Jenkins method.

2.3 Exponential Smoothing Method

The dataset that do not exhibit any trend or seasonality was forecasted by using simple exponential smoothing method [6]. The forecast values were computed by using formula in (1) and smoothing equation in (2);

\[
\hat{y}_t = \ell_{t-1} + b_{t-1} \\
\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}
\]

Where \(\alpha\) is smoothing constant between 0 < \(\alpha\) < 1, \(y_t\) is the actual value in period \(T\) and \(\ell_t\) is the level of series at time \(T\) and \(\hat{y}_t\) is the forecast value computed at time \(t\).

The dataset that exhibit trend but did not exhibit seasonality were forecasted by using Double Exponential Smoothing method. The forecast equation of (3) and smoothing equations of (4) and (5) were used to compute the forecast values.

\[
\hat{y}_t = \ell_{t-1} + b_{t-1} \\
\ell_t = y_t + \delta (\ell_{t-1} - \ell_{t-1}) + b_{t-1} \\
b_t = \beta (\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}
\]

Where \(b_t\) is the trend at time \(t\), \(\alpha\) is the smoothing parameter for the level and \(\beta\) is the smoothing parameter for the trend.

Lastly, the dataset that exhibit both trend and seasonal pattern were forecasted by using Holt-Winters method. The forecast equation of (6) and smoothing equations of (7), (8) and (9) were used to compute the level, trend and seasonal respectively.

\[
\hat{y}_{t+h} = \ell_t + h b_t + s_{t-h} \\
\ell_t = y_t + \delta (\ell_{t-1} - \ell_{t-1}) + b_{t-1} \\
b_t = \beta (\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\
s_t = \delta (y_t - \ell_t) + (1-\delta)(s_{t-1})
\]

Where \(s_t\) is seasonal component and \(\delta\) is the smoothing parameter for the seasonal component.

2.4 Box-Jenkins Method

There were three steps in Box-Jenkins approach which consist of identification, estimation and diagnosis check.

The general ARIMA model which allocates seasonality can be written as follows:

\[
\phi_p(B)\Phi_p(B^d)\Phi_p(B^s)y_t = \theta_q(B)\Theta_q(B^d)\Theta_p(B^s)a_t
\]

Where \(y_t\) and \(a_t\) are the actual value and random error at time period \(t\), respectively; \(\theta_p\) and \(\Phi_p\) is the order of autoregressive for non-seasonal and seasonal autoregressive; \(\theta_q\) and \(\Theta_q\) is the order of moving average for non-seasonal and seasonal.

2.5 Accuracy measures

The forecast performance was measured by using two accuracy measures which were Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The formula that was used to calculate MAE was given by;

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |e_i|
\]

Where \(e_i = A_i - F_i\), \(A_i\) is observed data and \(F_i\) is forecast data, \(n\) is the number of forecast data set and \(e_i\) is the error term. Meanwhile, the formula used to calculate MAPE was given by;

\[
MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{e_i}{A_i} \right|
\]

MAPE was used in most of the studies related to electricity load data due to its robustness [12].

3.0 Results and Discussion

In the case of data with seasonality, one week forecast was directly produced. However, for deseasonalisation data a single forecast for each day was produced and then it were compiled for a week forecast. Data of electricity load recorded in Central City Johor Bahru and Pasir Gudang are indicated as Station 1 and Station 2 respectively. In sample data for Station 1 was from 1st January 2003 until 1st February 2011. Meanwhile, data from 2nd February 2011 to 8th February 2011 were used as the out-sample data. Figure 1 shows the time series plot of this data.

In sample data for Station 2 was from 1st January 2003 until 1st March 2011 and the out sample was from 2nd March 2011 to 8th March 2011. The time series plot for this dataset is shown in Figure 2.

![Fig. 1: Time series plot of daily electricity load in Station 1](image-url)
The Box-Jenkins model of the seasonal data for each station is presented in Table 1. The Box-Jenkins model of each day is for Station 1 presented in Table 2.

**Table 1**: Box Jenkins model for seasonal data

<table>
<thead>
<tr>
<th>Station</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SARIMA(1,0,0)(0,1,1)$_{t}$</td>
</tr>
<tr>
<td>2</td>
<td>SARIMA(0,1,1)(0,1,1)$_{t}$</td>
</tr>
</tbody>
</table>
Table 2: Box-Jenkins models for deseasonalisation data

<table>
<thead>
<tr>
<th>Day</th>
<th>Station 1</th>
<th>Station 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARIMA(0,1,4)</td>
<td>ARIMA(1,1,1)</td>
</tr>
<tr>
<td>Sunday</td>
<td>ARIMA(0,1,1,)</td>
<td>ARIMA(2, 1, 1)</td>
</tr>
<tr>
<td>Monday</td>
<td>ARIMA(0,1,0)</td>
<td>ARIMA(1,1,1)</td>
</tr>
<tr>
<td>Tuesday</td>
<td>ARIMA(2,1,2)</td>
<td>ARIMA(0,1,1)</td>
</tr>
<tr>
<td>Wednesday</td>
<td>ARIMA(0,1,1)</td>
<td>ARIMA(1,1,1)</td>
</tr>
<tr>
<td>Thursday</td>
<td>ARIMA(1,1,1)</td>
<td>ARIMA(1,1,1)</td>
</tr>
<tr>
<td>Friday</td>
<td>ARIMA(2,1,1)</td>
<td>ARIMA(1,1,1)</td>
</tr>
<tr>
<td>Saturday</td>
<td>ARIMA(2,1,1)</td>
<td>ARIMA(1,1,1)</td>
</tr>
</tbody>
</table>

The performance of seasonal and non-seasonal data for exponential smoothing and Box-Jenkins method are presented in Table 3.

Table 3: Comparison of forecast performance

<table>
<thead>
<tr>
<th>Station</th>
<th>Seasonal (Exponential Smoothing)</th>
<th>Seasonal (Box-Jenkins)</th>
<th>Non-seasonal (Exponential Smoothing)</th>
<th>Non-seasonal (Box-Jenkins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3135.2</td>
<td>1348.7</td>
<td>1257.4</td>
<td>442.1</td>
</tr>
<tr>
<td></td>
<td>6.61</td>
<td>2.67</td>
<td>2.56</td>
<td>0.89</td>
</tr>
<tr>
<td>2</td>
<td>3537.3</td>
<td>3142.4</td>
<td>2636.5</td>
<td>1057.8</td>
</tr>
<tr>
<td></td>
<td>6.55</td>
<td>5.5</td>
<td>4.5</td>
<td>1.87</td>
</tr>
</tbody>
</table>

Results in the first row in the Table 3 of each station are the values for MAE while the results in second row are the MAPE (%). Box-Jenkins method outperformed the exponential smoothing method for both seasonal and non-seasonal data. Non-seasonal model for both methods has better accuracy measures as compared to seasonal model. In Station 1, deseasonalisation had increased the forecast accuracy up to 4.05% and 1.78% for exponential smoothing and Box Jenkins method respectively. Meanwhile, in Station 2, it increased the forecast accuracy up to 2.05% and 3.63%.

4.0 Conclusion

Two sets of daily electricity load recorded in Central City Johor Bahru and Pasir Gudang Station in Johor, Malaysia were used in this study. The daily seasonality component was removed from the data by dividing the data into seven different series according days. The performance of seasonal and non-seasonal models of exponential smoothing method and Box-Jenkins method were compared by using MAE and MAPE. It was found that the deseasonalisation data has improved the forecast accuracy of both methods for both datasets. This advantage is despite on the drawback in terms of the data length reduction when deseasonalisation data were employed. Moreover, the deseasonalisation data approach also may help to overcome complexity in modeling and forecasting data with double seasonality.

In terms of the comparison on the forecasting method, in overall Box-Jenkins produced better forecast compared with exponential smoothing method. As mentioned by [13], Box-Jenkins has good ability to forecast volatile data.

Acknowledgement

References

[5] Pardo, A., Meneu, V., & Valor, E. Temperature and seasonality influences on Spanish electricity