Haiges R, Wang YD, Ghoshray A, Roskilly AP.

*Forecasting Electricity Generation Capacity in Malaysia: An Auto Regressive Integrated Moving Average Approach.*

*In: 8th International Conference on Applied Energy (ICAE2016)*

8-11 October 2016, Beijing, China: Elsevier Ltd.

Copyright:

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer-review under responsibility of the scientific committee of the 8th International Conference on Applied Energy

DOI link to article:

https://doi.org/10.1016/j.egypro.2017.03.795

Date deposited:

18/07/2017

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence.
The 8th International Conference on Applied Energy – ICAE2016

Forecasting electricity generation capacity in Malaysia:
An Auto Regressive Integrated Moving Average approach

Rina Haigesa, Y.D.Wanga, A. Ghoshrayb, A.P. Roskillya

a Sir Joseph Swan Centre for Energy Research, School of Mechanical & Systems Engineering, Newcastle University, 
b Newcastle University Business School, Newcastle upon Tyne, NE1 7RU, UK

Abstract

It is imperative for Malaysia to have a clear understanding of the future performance of its power sector with emphasis on the total installed capacity variable as this is integral to support the nation’s capacity succession planning over an intermediate to long term period in order to sustain the economy. This paper aims to deploy the Auto Regressive Integrated Moving Average (ARIMA) approach to fit the 40 years forecast up to 2053 by assessing 40 years of past data from 1973 until 2013. The different models will be evaluated using the Schwarz Bayesian Criterion (SBC). Validation was performed by comparison of forecast and actual data based on a five-year holdback period. Accuracy measures applied were the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). In this assessment, ARIMA(0,2) demonstrated a better forecast in terms of accuracy during the holdback period. However, the Diebold-Mariano (DM) test didn’t detect any differences between the ARIMA(1,0) and ARIMA(0,2) forecasts. Application of the forecast results was demonstrated as well.

1. Introduction

The power sector in Malaysia is dominantly dependent on conventional fossil sources, according to the energy balance, figures for installed capacity in 2013 indicated that 83.9% share comes from fossil fuel and 13.2% is derived from hydropower. To be distinct, the 83.9% accounts for 51.8% natural gas, 25.8% coal, 0.5% fuel oil and 5.8% diesel [1]. Peninsular Malaysia and Sabah have an electricity supply outlook with peak demand and generation forecasted until 2030 [2] and 2033 [3] respectively, however, this is not substantive since it does not cover the whole nation whereby Sarawak is not accounted. With the depletion of fossil resources facing many economies which effect Malaysia as well, sustainable power sector generation capacity planning has gained more emphasis. Therefore this paper aims to project the future power capacity requirements for Malaysia up to 2053 through analysing past 40 years (1973 -2013) of annual total installed capacity data by adapting the ARIMA approach. The application of the forecast results is demonstrated in order to estimate the future peak demand and gross electricity generation figures.

Nomenclature

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>Auto Regressive Integrated Moving Average</td>
</tr>
<tr>
<td>p</td>
<td>number of lags of the considered variable</td>
</tr>
</tbody>
</table>

1876-6102 © 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the scientific committee of the 8th International Conference on Applied Energy.
doi:10.1016/j.egypro.2017.03.795
2. Literature review

The literature on forecast validation methods became more noticeable in the 1850s with the onset of weather services in Europe and America [4]. There are numerous forecast approaches that could be applied depending on the research framework either from a quantitative or qualitative or even a mixed context [5]. Selection of forecast method can be based on a few considerations such as availability of data, the time frame to perform the analysis, ease of method, forecast period and prior research [6]. There are multiple models which can be used for forecasting energy demand such as time series, regression, econometric, decomposition, co-integration, ARIMA, artificial systems such as the Artificial Neural Network (ANN), Grey prediction, Input-output, Fuzzy-logic, and the bottom-up models [7].

Since the available data is the installed capacity for Malaysia’s power generation since 1973 until 2013, this clearly indicates a univariate time series data, hence deliberation will be centred on time series univariate assessment. In general, time series can be analysed via a multivariate or univariate approach. A univariate analysis is dependent on one variable, whilst if the analysis involves correlation of more than one variable than it is considered a multivariate analysis. The univariate ARIMA model has gained extensive literature in energy demand projection [8-12] owing to its simple and reliable approach. It is also found suited for long term projection [13]. Furthermore, in order to avoid a spurious or invalid forecast, ARIMA is a recommended approach since it is widely established [14].

3. ARIMA approach

ARIMA modelling or the Box-Jenkins method was named after the two statisticians who introduced this approach in 1976. ARIMA is the combination of the autoregressive and moving average models [15]. The mathematical formula for an ARIMA model can be expressed as follows:

\[ y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \]  \hspace{1cm} (1)

There are a few underlying key essentials in ARIMA modelling, such as stationarity, invertibility and parsimony. Stationary means that the mean, variance and covariance of the series remains constant over time. This can be achieved by logarithmic transformation and by differencing either integrated to the order one I(1) or two I(2). Box and Jenkins believed that parsimonious models produce better forecast rather than an over-parameterised model with additional coefficients that would affect the degrees of freedom. Invertibility is another implicit requirement in ARIMA in which the measured variable \( y_t \) must exhibit a convergent autoregressive process or denoted by a finite order moving average. The three stages in ARIMA modelling as advocated by Box and Jenkins are (a) identification; (b) estimation; and (c) diagnostic checking [16]. The ARIMA forecast for installed capacity from 2014 up to 2053 was modelled by deploying the EViews software package which is primarily applied for time series econometric analysis.

4. Malaysia’s installed capacity historical time series data

Installed capacity data amplified 31 folds throughout the 40 year time frame from 1973 until 2013, which witnessed an increase from 951 MW to 29,729 MW at an average growth rate of 8.99%. There is a noticeable upward trend of the total installed capacity (cap) annual data from 1973 until 2013 [17-20]. In terms of installed capacity allocation, Peninsular Malaysia holds 81%, followed by Sarawak with 11.6% and Sabah has 7.4% share [1]. The noticeable increase in power demand is in line with Malaysia’s economic transformation from an agricultural commodity oriented to an
industrial manufacturing which later transcended into services based economy [2]. As a developing country, Malaysia experienced steady GDP growth, throughout 1973 until 2013 the average growth rate per year was 6.3% [21].

Since Malaysia is one of the emerging economies in South East Asia with a vision to achieve a developed nation status by 2020, GDP growth per annum is anticipated to be resilient at 5.9% from 2016 until 2020 and 6.2% during the course of 2021 until 2030 [2]. The average annual population growth rate is 2.3% over the 40 years period [22]. In 1973, Malaysia’s population was 11.7 million and in 2013 the figures expanded to 30.2 million. Malaysia’s reserve margin for electricity in 2013 stood at 30.1% [1], this specifies that the total installed capacity is able to cater for the peak demand. Malaysia’s electricity supply and demand growth alongside the GDP from 1973 until 2013 is depicted in Fig. 1.

![Fig. 1. Malaysia’s electricity supply and demand compared to the GDP from 1973 until 2013](image)

5. Results and discussion

The results will be presented in the following structure: identification, estimation, diagnostic check, forecasting, validation and application.

5.1 Identification

At this stage is the process to ensure data series is stationary, a statistical test known as the Augmented Dickie Fuller (ADF) test is performed to ensure the level and transformed series has achieved stationarity. The series are labelled as cap for the level series, lcap for the natural logarithmic series and dlcap after undergoing first differencing. The results of the ADF test on the three series are presented in Table 1. After being transformed to dlcap, the ADF test result showed that the series has achieved a stationary state. The results are considered stationary when the t stats value exceeds the 5% critical value.

<table>
<thead>
<tr>
<th>Series</th>
<th>t stats</th>
<th>5% cv</th>
<th>Analysis</th>
<th>Stationarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>cap</td>
<td>-1.2709</td>
<td>-3.5266</td>
<td>t stats &lt; 5% cv</td>
<td>Not stationary</td>
</tr>
<tr>
<td>lcap</td>
<td>-0.9562</td>
<td>-3.5266</td>
<td>t stats &lt; 5% cv</td>
<td>Not stationary</td>
</tr>
<tr>
<td>dlcap</td>
<td>-5.1500</td>
<td>-2.9389</td>
<td>t stats &gt; 5% cv</td>
<td>Stationary</td>
</tr>
</tbody>
</table>

Visual assessment for stationarity can be done by analysing the plots of the actual and transformed series. The upward trend line as per Fig.2(a) is no longer visible as it has been detrended through the transformation process. While the plots for the first differenced series as illustrated in Fig. 2(b) may indicate signs of stationarity due to a reverting mean and constant variance.
5.2 Estimation

Different ARIMA models are estimated at this stage to obtain the Schwarz Bayesian Criterion (SBC) for comparison. The SBC is utilised as a measure to identify the plausible model with the best fit since SBC is found to be more consistent in selecting a parsimonious model [16, 23]. After running few estimations based on the parsimonious model condition where \( p + q \leq 6 \), three models were selected based on the minimum SBC value as in Table 2.

Table 2. Schwarz criterion value

<table>
<thead>
<tr>
<th></th>
<th>(1,0)</th>
<th>(0,1)</th>
<th>(0,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBC</td>
<td>-2.034</td>
<td>-2.032</td>
<td>-1.952</td>
</tr>
</tbody>
</table>

5.3 Diagnostic check

This to check if the residuals have the white noise characteristics. To avoid a spurious forecast, residuals must not be correlated and it is a good practice to include a Chow test [24] to rule out structural breaks. Both the Auto Correlation Function (ACF) and the Partial Auto Correlation Function (PACF) for all three models showed no signs of correlation since all of the spikes were within the standard error bars as shown in Fig.3. Furthermore, as a cut-off point to maintain a meaningful forecast by having at least 30 actual observed data, 1984 was the identified breakpoint test year for the Chow test. The Chow test results for the three identified models indicated that there was no occurrence of a breakpoint for the chosen test year since the p-value exceeded the F-statistic value or the 0.05 significance level as indicated in Table 3.

Table 3. Chow test results

<table>
<thead>
<tr>
<th>F-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,0)</td>
<td>0.355</td>
</tr>
<tr>
<td>(0,1)</td>
<td>0.441</td>
</tr>
<tr>
<td>(0,2)</td>
<td>0.588</td>
</tr>
</tbody>
</table>
5.4 Forecasting

At this stage, the forecast data needs to be transformed back to level data. The forecast results for all three models are shown in Fig. 4. The highest forecast was derived from ARIMA(1,0), whilst ARIMA(0,2) produced the lowest forecast and ARIMA(0,1) formed a slightly higher forecast than ARIMA(0,2).

![Fig. 4. Installed capacity annual forecast 2014 until 2053](image)

Measures such as the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used to isolate the more reasonable forecast model. The results in Table 4 presents that ARIMA(1,0) followed by ARIMA(0,2) are the more reliable forecast out of the three models since it bears lesser MAE and RMSE values.

Table 4. Forecast performance measures

<table>
<thead>
<tr>
<th></th>
<th>(1.0)</th>
<th>(0.1)</th>
<th>(0.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.0609</td>
<td>0.0614</td>
<td>0.0612</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0808</td>
<td>0.0810</td>
<td>0.0810</td>
</tr>
</tbody>
</table>

5.5 Validation

In order to measure forecast accuracy, the forecast results can be held back to a period with known actual observation for comparison. In this exercise, we implemented a five-year holdback period from 2009 until 2013. The plots for the ARIMA(1,0) and ARIMA(0,2) forecast denoted CAPF were paralleled alongside the actual data CAP as per Fig. 5.

![Fig. 5. Five year holdback validation for (a) ARIMA(1.0); (b) ARIMA(0.2)](image)

The forecast accuracy measures namely the MAE and RMSE were calculated based on errors between forecasted and actual data [25] as in Table 5. Results showed that ARIMA(0,2) gave a more accurate forecast during the 5 years
holdback period. To confirm this, the Diebold-Mariano (DM) test was conducted to check which model forecast was better between two competing models. The null hypothesis ($H_0$) for the DM test suggests that both forecasts have the same accuracy. The alternative hypothesis ($H_1$) states that the forecasts have different levels of accuracy, one model is better than the other. The null hypothesis cannot be rejected when there is a p-value. The findings of the DM test is shown in Table 6 conclude that there isn’t any significant difference between ARIMA(1,0) and ARIMA(0,2) for the overall forecast period.

Table 5. Forecast accuracy measures

<table>
<thead>
<tr>
<th></th>
<th>(1,0)</th>
<th>(0,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>1,588.7</td>
<td>1,326.7</td>
</tr>
<tr>
<td>RMSE</td>
<td>1,967.3</td>
<td>1,532.9</td>
</tr>
</tbody>
</table>

Table 6. Diebold-Mariano test

<table>
<thead>
<tr>
<th>$H_1$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,0) &gt; (0,2)</td>
<td>0.360</td>
</tr>
<tr>
<td>(0,2) &gt; (1,0)</td>
<td>0.639</td>
</tr>
</tbody>
</table>

5.6 Application

By normalising the forecast results to the 2013 demand profile as displayed in Fig. 6., the peak demand (MW) and gross electricity generation (GWh) of a specified reference year can be established. Table 7 shows the future power supply and demand estimates deduced from the ARIMA(0,2) forecast. The forecast results were directly applied to generate the installed capacity data, whilst available capacity was obtained by multiplying the installed capacity figures with the capacity factor value of 0.83. Successively, average power was calculated by dividing electricity generation with 8760 total hours per year, while the reverse function was carried out to obtain electricity generation. Finally, peak demand was derived by normalising average power against the 2013 peak demand profile.

![Fig. 6. Malaysia’s peak demand profile for 2013](image_url)
Based on these results as illustrated in Fig. 7, an upward development is identified for installed capacity, peak demand and electricity generation which aligns well with the historical data trend. The annual growth rate throughout the forecast period is 8.78% for all three estimated variables. The assumption applied here, is that the power sector continues to operate under the business as usual scenario, whereby the Malaysian economy continues to experience steady GDP growth at 6.3%. The Malaysian government needs to prepare a sound plan to accommodate the increase in generation capacity requirements. As a counter measure to the energy security and climate change challenges that lies ahead, it is high time for Malaysia to proactively plan ahead and restructure her electricity mix into a more sustainable portfolio.

6. Conclusion

ARIMA approach is one of the more widely used approaches in energy demand projection. Forecasting with ARIMA method provides a projection which relies on past historical data, in which the data has been modified to reach a state of statistical equilibrium. In this evaluation ARIMA(0,2) yielded the more accurate forecast over the five-year holdback period, however, for the overall forecast period, the DM test didn’t detect any significant differences between ARIMA(1,0) and ARIMA(0,2) forecasts. We acknowledge that uncertainty develops as the forecast period extends longer into the future, this stands true for forecast accuracy as well. It is expected that this data could be applied by fellow researchers, power utility companies, regulators and policy makers for the benefit of Malaysia’s intermediate to long-term power capacity succession planning.

7. Copyright

Authors keep full copyright over papers published in Energy Procedia
Acknowledgements

Utmost appreciation is accorded to the Public Service Department of Malaysia for funding the first author’s doctoral works of which this paper forms part of her studies.

References


Biography

Rina Haiges holds a Master’s degree in Conservation Biology. She worked with Shell Malaysia Limited and was involved in the management of science and technology policies and programs for more than 8 years. Her current PhD studies with Swan Energy institute, Newcastle University is related to sustainable power sector modelling.