ARIMA with Regression Model in Modelling Electricity Load Demand

Nor Hamizah Miswan[1, 2], Rahaini Mohd Said[1,2] and Siti Haryanti Hairol Anuar[1,3]
[1]Faculty of Engineering Technology, Universiti Teknikal Malaysia Melaka (UTeM), 76100 Durian Tunggal, Melaka, Malaysia.
[2]Applied Mathematics Research Group, Advance Manufacturing Centre,
Universiti Teknikal Malaysia Melaka (UTeM), 76100 Durian Tunggal, Melaka, Malaysia.
[3] Optimisation, Modelling, Analysis, Simulation and Scheduling (OptiMASS),
Centre of Advance Computing Technology (C-ACT), Universiti Teknikal Malaysia Melaka (UTeM),
76100 Durian Tunggal, Melaka, Malaysia
norhamizah@utem.edu.my

***Abstract*—Electricity is among the most crucial needs for every people in this world. It is defined by the set of physical phenomena related with the flow of electrical charge. The importance of electricity itself leads to the increasing electricity load demand in the world including Malaysia. The purpose of the current study is to evaluate the performance of combined ARIMA with Regression model in forecasting electricity load demand in Johor Bahru. Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) and Regression models will be used as benchmark models since the model has been proven in many forecasting context. Using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as a forecasting accuracy criteria, the study concludes that the combined method is more appropriate model.**

***Index Terms*—ARIMA; ARIMA with Regression; Load Demand; Regression.**

I. INTRODUCTION

Load demand forecasting is a method used to predict the energy or power needed to meet the demand in a certain area. In forecasting electricity load demand data, Time Series approaches are among the widely used methods [1]. The electricity load data itself is univariate time series observations where the data can be obtained in every hour.

Nur Adilah et al. (2013) proposed Exponential Smoothing methods to forecast electricity load demand of Malaysia [2]. In their work, three Exponential Smoothing models were investigated which are Holt-Winters Taylor (HWT) Exponential Smoothing, standard Holt-Winters Exponential Smoothing and modified Holt-Winters Exponential Smoothing. They claimed that the HWT Exponential Smoothing is the ideal forecasting model.

Delson and Caston (2012) proposed the Regression-Seasonal ARIMA model in modelling daily peak electricity demand in South Africa whereby this model exhibited an outstanding result of RMSE and MAPE [3]. Caleb et al. (2013) conducted a case study on the Nigerian power sector [4]. In order to forecast the electric load demand in that country, they used least squares technique in four different regions.

Nataraja et al. (2012) conducted a study on short term load forecasting in Karnataka, India by using Time Series analysis [5]. Three types of ARIMA models were developed which are ARIMA, Autoregressive Moving Average (ARMA) and Autoregressive (AR) model. The result shows that ARIMA model is the reliable model.

On the other hand, Regression models are also common in load forecasting and used to model the relationship between the load demand and external factors [1]. Weather especially temperature and calendar information are among the common external factors used to model the relationship between predicted and respond variable [6].

M. A. Mahmud (2011) proposed Linear Regression Models to forecast load demand in isolated area of Bangladesh, called Swandip where a past history of load demand is unavailable together with no possibility of main land grid system connection. Linear Regression is done and based on the factors identification on which electrical growth depends.

N. Amral et al. (2007) used Multiple Linear Regression for short term load forecasting in South Sulawesi's power system. For a case study, historical data of hourly load demand and temperatures of South Sulawesi electrical system have been used [7]. Aayush Goel et. al. used Multiple Regression, Trend Seasonality Model and ARIMA modelling to forecast the electricity demand in New Delhi, India. For Multiple Regression, the independent variables used are temperature, mean humidity percentage, precipitation and time trend. The result shows that ARIMA perform better in New Delhi load forecasting [12].

From the literature, it shows that ARIMA and Regression are among the crucial models in load demand forecasting. Hence, the aim of this paper is to compare the performance of combined ARIMA with Regression models in comparison with ARIMA and Regression models itself to forecast the load demand in Johor Bahru.

II. METHODOLOGY

A. ARIMA Modelling

ARIMA is a broadly used among the statistical models for time series analysis and forecasting applications [9], which was introduced by George box and Gwilym Jenkins [10]. ARIMA models contain three parts, which are Autoregressive part of order p, AR(p), differencing part of order d, I(d) and Moving Average part of order q, MA(q) [8]. There are 3 steps in the ARIMA model, which are model identification, model estimation and model application.

a. Model Identification

Firstly, we need to identify the order of differencing (d) to stationaries the data. Before any analysis of Time Series being conducted, stationary checking is the most crucial step. Let y be the d^{th} difference of order Y, which means:

$$\text{If } d=0\text{:}\quad y_t = Y_t \tag{1}$$

$$\text{If } d=1\text{:}\quad y_t = Y_t - Y_{t-1} \tag{2}$$

Equation (1) indicates that there is no difference on Y which means the data is stationary. Equation (2) indicates that it is a first difference of Y in order to make the data stationary. After the data was stationaries by differencing, the next step is using ACF and PACF plot to determine the potential model.

From the autocorrelation function (ACF) and partial auto correlation function (PACF) plots of the difference series, we can identify the number of terms for AR(p) and MA(q) that are required. The combinations of AR and MA term will give the entire potential model for our forecasting. Note that the ACF plot is given by MA(q) model and the PACF plot is given by AR(p) model [9].

b. Model Estimation and Validation

From the term obtained from AR(p), I(d), and MA(q), we have the different combination between that 3 terms. Hence, using the combination, the best model to forecast is measured by using AIC (Akaike's Information Criterion) criterion. The lowest value indicates the best modelling model [10], which is given by:

$$AIC = -2 \ln(\text{maximum likelihood}) + 2p \tag{3}$$

where p = The number of parameters involved.

c. Model Application

The most satisfactory model for ARIMA forecasting can be confirmed using the forecast accuracy criteria such as MAE and RMSE [11], which are given by the respective equations.

$$MAE = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{n} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}} \tag{5}$$

where: y_t = The original data

$\hat{y}_t$ = The predicted data

n = Total observations

B. ARIMA Model with Regression

The ARIMA time series with the regression model is one of the method for time series analysis, where there are combining the features of ARIMA and Linear Regression models. The best ARIMA model will be combined with the regression in order to improve the accuracy of the model. Hence, the model combination can be explained as the following steps:

Step 1: Fit the best an ARIMA (p,d,q) model with no regressors. Then the forecasting equation as follows:

$$\hat{Y}_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \vartheta_i y_{t-i} \tag{6}$$

Step 2: Find the Equation of Regression Model.

$$\hat{Y}_t = \beta_0 + \beta_1 x + \varepsilon \tag{7}$$

Step 3: Add the regressor X to the forecasting ARIMA model.

$$\hat{Y}_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \vartheta_i y_{t-i} + \beta_0 + \beta_1 x + \varepsilon \tag{8}$$

Hence the equation can be simplified as follows:

$$\hat{Y}_t = \varphi + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \vartheta_i y_{t-i} + \beta_1 x + \varepsilon \tag{9}$$

where φ is constant and ε is error term.

III. RESULT AND DISCUSSION

The data used in this study are load profiling data of Johor Bahru obtained from Tenaga Nasional Berhad and temperature of Johor Bahru obtained from Malaysian Department of Meteorology.

A. ARIMA Modelling

Figure 1 shows the linear trend analysis for the original electricity load demand data. The plot suggests that the original data set is not stationary.

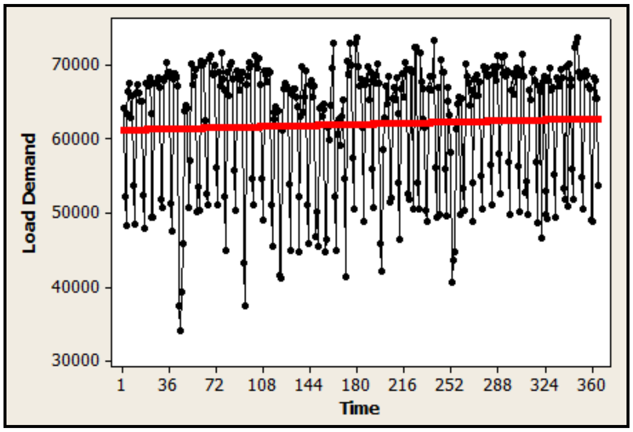

Figure 1: Time series plot of the original data

Therefore, we analyse the first difference of the series, $\Delta y = y_t - \hat{y}_t$ and we can conclude that the data is stationary. The plot of the first difference data is shown by Figure 2.

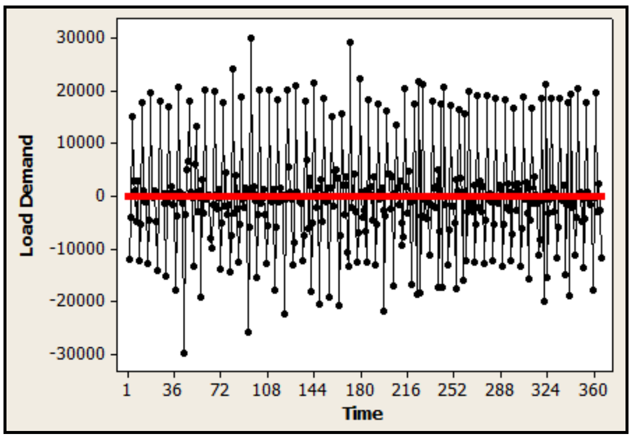

Figure 2: Linear trend plot of the first difference data

The next steps are to plot ACF and PACF as shown in Figure 3 and Figure 4 respectively.

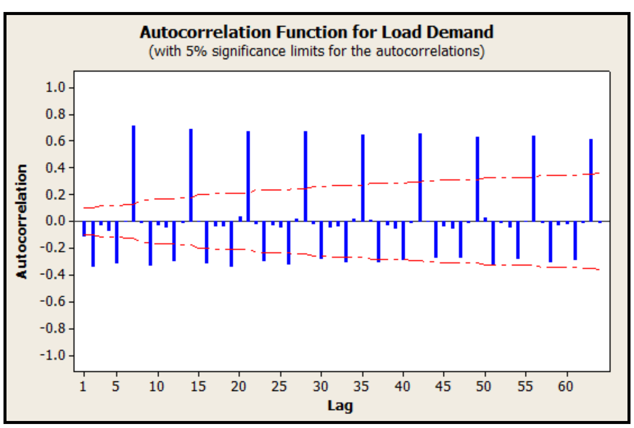

Figure 3: Collelogram for ACF plot

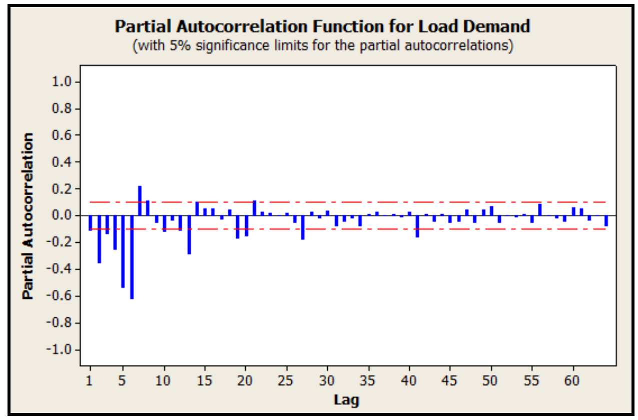

Figure 4: Collelogram for PACF plot

The graph of the plot in Figure 3 and Figure 4 suggests an appropriate ARIMA model of the data. AR (p) model is represented by the PACF and the MA (q) model is represented by the ACF. Based on the plot and the significant spike, the following nine models have been estimated and identified using Eviews software. The estimated ARIMA model for forecasting the electrical load with their corresponding AIC values is given in Table 1.

Table 1
The list of potential ARIMA models

Models	AIC
ARIMA(2,1,2)	21.13434
ARIMA(2,1,4)	21.13119
ARIMA(2,1,5)	21.17026
ARIMA(5,1,2)	21.12331
ARIMA(5,1,5)	21.20484
ARIMA(5,1,4)	21.11131
ARIMA(7,1,2)	20.57594
ARIMA(7,1,4)	20.57717
ARIMA(7,1,5)	20.58752

From Table 1, ARIMA(7,1,2) has the minimum AIC indicates that ARIMA(7,1,2) is the best model among the other ARIMA models. The best model will be used with regressor later.

B. ARIMA Model with Regression

After obtained the best ARIMA model for forecasting, the next step is to find the equation for regression line. For this study, temperature of Johor Bahru will be used as an independent variable, *x* while load demand for Johor Bahru will be used as a dependent variable, *y*.

By using Excel, the equation for Regression line is as follows:

$$Y = 42733.58 + 670.5616x \qquad (10)$$

where Y is load demand and x is temperature.

The regressor will be added to the ARIMA(7,1,2) models and the equation are as follows:

$$Y = 42733.58 + 670.5616x + ARIMA(7,1,2) \quad (11)$$

C. Comparative Performance for ARIMA, Regression and ARIMA with Regression Models

RMSE and MAE will be used as a forecast accuracy criterion in order to measure the performance of each models. The results are tabulated in Table 2.

Table 2
Comparative performance for each Time Series and Regression models

Model	RMSE	MAE
ARIMA(7,1,2)	7084.934	4900.439
Regression	8801.006	7642.711
ARIMA(7,1,2) with Regression	6455.5146	4807.954

From Table 2, the lowest RMSE and MAE values are from ARIMA (7,1,2) with regressor model. Hence, ARIMA model with regressor is the best models for modelling and forecasting electricity load demand data in Johor Bahru as compared to ARIMA and Regression itself.

IV. CONCLUSION

The forecasting of electricity load demand has become one of the essential areas of research in recent years. This paper presents an attempt to compare the existing single forecasting models with the combined model in forecast the load demand of Johor Bahru. ARIMA and Regression models have been chosen as benchmark since the models have been widely utilized in many areas in Time Series, especially for load forecasting. ARIMA with Regression model has been considered as the best model as compared to ARIMA and Regression model itself due to the lowest RMSE value. This model can be used in forecasting the electricity load demand in Johor Bahru for the future.

ACKNOWLEDGMENT

We wish to thank the faculty of Engineering Technology and UTeM for the help and support in carrying this research under the grant vote number RAGS/1/2014/SG04/FTK/B00087.

REFERENCES

[1] Heiko Hahn, Silja Meyer-Nieberg and Stefan Pickl, Electric Load Forecasting Methods : Tools for Decision Making, *European Journal of Operational Research*, **199**, 902-907 (2009)

[2] Nur Adilah Abd Jalil, Maizah Hura Ahmad and Norizan Mohamed, Electricity Load Demand Forecasting Using Exponential Smoothing Methods, *World Applied Sciences Journal* **22**, 1540-1543 (2013)

[3] Delson Chikobvu and Caston Sigauke, Regression-SARIMA Modelling of Daily Peak Electricity Demand in South Africa, *Journal of Energy in Southern Africa*, **23**, 23-30 (2012)

[4] Caleb A. Amlabu, Jonathan U. Agber, Cletus O. Onah and Suberu Y. Mohammed, Electric Load Forecasting : A Case Study of the Nogerian Power Sector, *International Journal of Engineering and Innovative Technology*, **2**, 23-27 (2013)

[5] Nataraja. C, M. B. Gorawar, Shilpa. G. N. and Shri Harsha. J, Short Term Load Forecasting Using Time Series Analysis : A Case Study for Kartanaka, India, *International Journal of Engineering Science and Innovative Technology*, **1**, 45-53 (2012)

[6] M. Mahmud, Isolated Area Load Forecasting using Linear Regression Analysis : Practical Approach, Energy and Power Engineering, **3**, 547-550 (2011)

[7] N. Amral, C. S. Ozveren and D. King, Short Term Load Forecasting using Multiple Linear Regression. *42nd Universities Power Engineering International Conference (UPEC 2007)*, 1192-1198 (2007)

[8] Kavasseri, R. G. and Seetharaman, K., Day-Ahead Wind Speed Forecasting using f-ARIMA Models, *Renewable Energy*, **34**, 1388-1393 (2009)

[9] Lee, C. M. and Ko, C. N., Short Term Load Forecasting Using Lifting Scheme and ARIMA Models, *Expert Systems with Applications*, **38**, 5902-5911 (2011)

[10] Nor Hamizah Miswan, Rahaini Mohd Said, Nor Hafizah Hussin, Khairum Hamzah and Emy Zairah Ahmad, Comparative Performance of ARIMA and DES Models in Forecasting Electricity Load Demand in Malaysia, *International Journal of Electrical & Computer Sciences IJECS-IJENS*, **16**, 6-9 (2015)

[11] Nie, H., Liu, G., Liu, X. and Wang, Y., Hybrid of ARIMA and SVMs for Short Term Load Forecasting, *Energy Procedia*, **16**, 1455-1460 (2011)

[12] Aayush Goel and Agam Goel, Regression Based Forecast of Electricity Load Demand of New Delhi, *International Journal of Scientific and Research Publications*, **4**, 1-7 (2014)