MATEMATIKA, 2005, Jilid 21, Bil. 2, hlm. 143–152

SARIMA Model for Forecasting Malaysian Electricity Generated

[1]Zuhaimy Hj. Ismail & [2]Khairil Asmani Mahpol
[1,2]Faculty of Science, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
e-mail: [1]zhi@mel.fs.utm.my, [2]khairile7@yahoo.com

Abstract Time-series extrapolation which is also known as univariate time series forecasting relies on quantitative methods to analyse data for the variable of interest. Pure extrapolation is based only on values of variable being forecast. We are interested in forecasting the electricity generated for Malaysia. The Tenaga Nasional Berhad (TNB) operates an electricity network with the largest capacity of over 7100MW that accounts for over 62% of the total power generation of Peninsular Malaysia. The rest of the power is generated by other Independent Power Producer (IPP). A forecasting model has been developed which identifies seasonal factors in the time-series. Seasonality often accounts for the major part of time series data. In this paper we examine the forecasting performance of Box-Jenkins methodology for SARIMA models and ARIMA models to forecast future electricity generated for Malaysia. We employ the data on the electricity generated at Power Plant to forecast future electricity demand. The error statistics of forecast between the models for a month ahead are presented and the behaviour of data is also observed.

Keywords Forecasting, electricity demand, SARIMA, Box Jenkins, genetic algorithm.

1 Introduction

Forecasting activities has been going on since early days and there are records of various ways of how human based their forecast. As economic activities increase, forecasting becomes an important tool in economic forecasting through the century, particularly in the business cycles. An increase in the complexity of organisations and the expansion of computer technology has somehow increase the interest in forecasting. Globalization and liberalization further raise many issues and challenges confronting the market. The cost of supplying energy is very expensive and it will get more expensive in the future. At the same time, demand for energy will continue to increase as the economy keeps on expanding. It is very important to ensure both the service providers and customers share the benefits equitably. This is why forecasting tools are very useful in charting our future energy market for energy decision planning and development.

Malaysia is a developing nation and its development is highly dependent on the ability to produce and improve its productivity. Energy is recognized as one of the prime agents in the increase of productivity and a significant factor in economic development. There has been a strong relationship between the availability of energy, economic activity and improvements in standards of living and overall social well-being. One of the major electricity producers requires an accurate forecast for electrical energy demand for the country with data based on the standard format recommended in the energy sectors. A five year total monthly energy generated data in kWh unit is available for making future prediction. Many studies have pointed out the overwhelming sensitivity of electricity consumption to many variables, in particular focusing attention on the forecast limited to 24 hours ahead. The prediction of the system load over an interval ranging from one hour to one week ahead is known as short term forecasting (STF). Recent studies have also focused on the impacts of climatic changes both on supply and demand for energy. Weather sensitivity has also been examined in order to correlate electricity consumptions to the increases in market saturation of air conditioning induced by long term climatic changes [9].

Development of new forecasting methods and their practical application developed in the university environment will enable managers (the decision makers) to make better forecasts. More accurate forecast for electricity generated has become increasingly important to the energy provider and the government to assist in the decision processes in today's complex business environments. In practice, there are gaps between the conventional methods of forecasting and new, more powerful and precise methods through creative employment of the computer.

Various forecasting techniques have been utilized to provide electricity load forecast [4][6]Forecasting accuracy has been improved substantially especially in models such as the autoregressive and moving average models. However, the ARMA models show some deficiency in the modelling of seasonal changing periods. Recently, heuristic approaches have been proposed to complement ARMA model in forecast of load demand. Genetic Algorithm (GA) is one of the methods proposed. GA is widely known as a powerful searching procedure commonly used in optimization and approximation field. The increasing popularity of GA is due to their adaptability and simplicity as a problem solution especially when they are applied to several complex problems. GA's blind search is capable of learning the relationship based on the pattern-recognition of past data, a process that helps GA to make a prediction of future values. In this article, we explain the technique of developing a forecast model based on genetic algorithm.

We propose the possibility of using GA's approach as one of the unique forecasting method in making forecast of the electricity generated by Tenaga Nasional Berhad (TNB) and other independent power producers. Tenaga Nasional Berhad (TNB) is the largest electric utility in Malaysia. It was established in September 1990 through a corporatisation and privatization exercise by the Malaysian government. The corporatisation is regulated under the Electricity Supply Department under the Ministry of Energy, Telecommunications and Posts, Malaysia. TNB has more than RM 74 billion in assets and serving over five million customers throughout the Peninsular, Sabah and Sarawak. Its core activities are in the generation, transmission and distribution of electricity. Until now TNB remains the main provider in electricity generation and through its wholly owned subsidiary TNB Generation Sdn. Bhd. TNB has the largest generation capacity of over 7100 MW that accounts for over 62

To support the goal of modelling energy generated, we collected data on the power generated at Power Plants and carry out the analysis. The aim is to forecast the future electricity generated using the time series Box-Jenkins Methodology namely the use of ARIMA model and to examine the forecasting performance of SARIMA models to forecast electricity generated for Malaysia.

2 Problem Description

Economic and reliable operation of electricity utility power system depend on a significant extent of the accuracy of load demand forecast. Load forecasting problems may be divided into two main categories;

(i) long-term forecast used for system planning and purchasing of generation units by TNB,

(ii) short-term forecasts used for optimal generator unit commitment, economic dispatching, security assessment, fuel allocation, maintenance scheduling and buying and selling of power.

It is important to understand the exact nature and extent of the electricity generated based on the demand of the country. The accuracy of load forecasts depends on many factors, which are of course more accurate for one day ahead than for seven days ahead. At present, it is the practice at TNB that the demand or load is forecasted manually by operation planning (dispatching) engineers, through searching for a similar day, making adjustments to that day's load to account for differences in weather conditions and for other factors that may affect the instantaneous load. They typically forecast for a maximum of 24 hours ahead, with forecasts being re-evaluated during the day. An experienced planner can estimate the peak demands to within 100 MW of actual demand, with a lead-time of 12 hours. This is what TNB classifies as a short term forecast ranging from an hour to a 24 hr forecast. Short term forecast (STF) plays a key role in the formulation of economic, reliable and secure operating strategies for the power system. The principal objective of the STF function is to provide the load predictions for basic generation scheduling, assessing the security of the power systems at any time point and as timely dispatcher information. It is used to drive the scheduling functions that determine the most economic commitment of generation sources consistent with reliability requirements, operational constraints and policies, and physical, environmental and equipment limitations. Therefore the principle issue here is to improve forecast accuracy.

3 Method Based on Fitting a SARIMA Model

Autoregressive model of order p or briefly $AR(p)$ of a time series y_t could be represented by the general equation:

$$y_t = \alpha_0 + \sum_{t=1}^{p} \alpha_t (y_{t-1} - \mu) + \varepsilon_t$$

where α_0 and α_i ; $i = 1, 2, ..., p$ are autoregressive parameters to be estimated and μ refers to the mean value of y_t, and ε_t represents random errors with zero mean and finite variances.

The theory of Autoregressive Integrated Moving Average (ARIMA) models and their applications in time series forecasting are discussed in many textbooks ([2],[4]) In the traditional Box-Jenkins methodology, the main tool for specifying a suitable ARIMA model is the sample autoregression function. In 1970, George Box and Gwilym Jenkins proposed ARIMA as a model building methodology comprising several stages; identification, estimation, diagnostic checking and forecasting. They effectively put together in a comprehensive manner the relevant information required to understand and use univariate time series ARIMA models. The ARIMA procedure utilizes the Box-Jenkins approach to time series data. The ARIMA procedure produces forecasts for autoregressive models. One advantage of ARIMA procedure is its ability to identify the error process underlying the time series. Additionally, ARIMA allows us to fit a more complex model. The forecasts produced by the ARIMA procedure depend upon the type of model and the estimation method used to fit the model. In some cases, it may be more efficient to use ARIMA to identify models of unknown processes.

The extension to the $AR(p)$ and MA(q) models is the mixed model that is of the form:

$$z_i = c + \phi_1 z_{i-1} + \phi_2 z_{i-2} + ... + \phi_p z_{i-p} + a_i - \theta_1 a_{i-1} - \theta_2 a_{i-2} - .... - \theta_q a_{i-q} \tag{3.1}$$

called $ARMA(p,q)$ model. Many observed non-stationary time series exhibit a certain homogeneity and can be accounted by making a simple modification on the $ARMA$ model, called integrated processes. By adding an integrated process the $ARMA(p,q)$ model will become an ARIMA (p,d,q) namely autoregressive integrated moving average model. It can be written as

$$\phi(B)\, z_i = \theta(B)\, a_i \tag{3.2}$$

where $\varphi(B)$, the generalized autoregressive operator, is a polynomial of degree $p+d$ with exactly d zeros equal to unity. Therefore

$$\varphi(B) = \phi_p(B)(1-B)^d = \phi_p(B)\,\nabla^d \tag{3.3}$$

where $\phi_p(B)$ is a stationary autoregressive operator of order p.

The model used here was Seasonal Autoregressive Integrated Moving Average, SARIMA. The SARIMA model with degree p, q, P and Q can be written as SARIMA $(p,d,q)(P,D,Q)$. Usually seasonal ARIMA models can be detected by looking at their seasonal and trend pattern which circulate at time T, repeat their pattern for the next period either yearly or quarterly.

Such a series exhibits periodic behaviour with period T just like the time series data in this research which contain $T = 12$ with the basic time interval being 1 month. Therefore, the stationary seasonal autoregressive operator of order P, the $AR(P)_T$ operator is modelled as

$$\Phi_P\left(B^T\right) \equiv 1 - \Phi_1 B^T - .... - \Phi_P B^{TP} \tag{3.4}$$

and similarly the invertible $MA(Q)_T$ operator

$$\Theta_Q\left(B^T\right) \equiv 1 + \Theta_1 B^T + ... + \Theta_Q B^{TQ} \tag{3.5}$$

and the seasonal difference operator

$$\nabla_T \equiv 1 - B^T; \tag{3.6}$$

if we applied D times it can be written as

$$\nabla_T^D \equiv \left(1 - B^T\right)^D. \tag{3.7}$$

Therefore the model is of the form

$$\Phi_P\left(B^T\right)\nabla_T^D\tilde{z}_i=\Theta_Q\left(B^T\right)a_i \tag{3.8}$$

called as SARIMA will be produced.

By denoting the ordinary of stationary $AR(p)$ and $MA(q)$ operators, $\phi_p(B)$ and $\theta_q(B)$ respectively, the general Box-Jenkins model that allows seasonality is

$$\phi_p(B)\ \Phi_P\left(B^T\right)\nabla^d\nabla_T^D\tilde{z}_i=\theta_q(B)\ \Theta_Q\left(B^T\right)\ a_i \tag{3.9}$$

and is referred to as the multiplicative $(p,d,q)\times(P,D,Q)_T$ model, where

Z_t	: time series values,
ϕ_p	: unknown parameter for AR,
θ_q	: unknown parameter for MA,
Φ_P	: unknown parameter for SAR,
Θ_Q	: unknown parameter for SMA,
$a_{i-1},a_{i-2},\ldots,a_{i-q}$	: the error variables at time $t-q$,
a_i	: the error term at time I,
c	: constant term.

4 Data and Data Analysis

The applications chosen for this study is the forecasting of Malaysian electricity energy generated at power plant. Data used were collected from 1996 to 2003, which has been classified into total monthly electricity generated in kWh unit. The flow of electricity is represented by the following equation

$$\begin{aligned}\text{Electricity Generated} &= \text{Gross Inland Consumption}\\ &= FEC+CET+DL\end{aligned}$$

where FEC is the Final Electricity Consumption, CET is the Consumption of the Energy Transformed and DL is the Distribution Losses. The final electricity consumption refers to the total quantity of electricity delivered to final user. The DL refers to losses of electrical energy, which occur outside the utilities and plants, and the consumption of electricity by utilities and plants for operating their installation.

Graph 1.0 shows the data movement from September 1996 to January 2003. We can obtain the addictive seasonal pattern with seasonal lag, $T=12$ and linear trend on this series. Because of the variables constraint this series need transformation using natural logarithm, ln defined as follows

$$z_t^*=\begin{cases} z_t^\lambda, \\ \ln z_t, \end{cases}$$

where $\lambda\neq 0$ (power transformation) and $\lambda=0$ (logarithm transformation).

Further data analyses must be done to describe their behaviour before suitable model can be developed. We need to examine the model output by looking at the process converged, error term and parameters estimate. In this research, data will be examined using Box-Jenkins Methodology for SARIMA models. The final result analyses will be compared and discussed.

5 Implementation at Case Study

The model development process begins by making transformation on the data series using natural logarithms. In order to produce the stationary series we have to make a differencing ∇^d or ∇^D. After the data is stationary, check their ACF and $PACF$. Seasonality may suggest itself if the autocorrelations and/or partial autocorrelations at the seasonal lags are large and significantly different from zero. Only the converged process that produces forecast graph will be finalized. The model that gives the least MSE is considered to be the best model.

In this study we have selected the multiplicative SARIMA$(1,1,0) \times (1,0,1)_{12}$ as the best final model for electricity generated forecast with Sum Squared Error, S.S.E, 0.13441, Mean Squared Error, M.S.E, 0.00184, parameters estimate for ϕ_1, -0.5781, $\Phi_1, 0.91075$ and $\Theta_1, 0.33655$. The diagram 1.0, diagram (a) shows the transformation for Y_t using natural logarithms, diagram (b) is the differencing of order-1, diagram (c) is the result of Autocorrelation Function (ACF), diagram (d) is the Partial Autocorrelation Function (PACF), diagram (e) is the forecast graph for ARIMA$(1,1,1)$ model and (f) is the forecast graph for SARIMA$(1,1,0) \times (1,0,1)_{12}$ model. The trial and error method used gives the process convergence result, the error term and the parameter estimate as given in Table 1.0 and Table 1.1.

6 Discussion, Conclusion and Further Research

We investigated the robustness properties of several forecasting methods for seasonal time series. The development of parameter estimation of Holt-Winters methods as developed by Zuhaimy and Nizam, ([9]) using GA and the ARIMA model fitting approach based on suitable parsimonious models possess satisfactory robustness for a wide range of time series. Comparing the results for the ARIMA model approach, the SARIMA model gave much better results. The results using SARIMA model also demonstrate the ability to predict the next forecast value based on past observation. This model can be used to predict the next value, and it also follows their trend and seasonality otherwise ARIMA model failed to make accurate forecast. The SARIMA$(1,1,0) \times (1,0,1)_{12}$ has been selected as the best

forecast model for this particular Electricity Generated Data provided by TNB. It recorded the lowest mean square error (MSE) of 0.00184.

The forecasting performance of SARIMA modelling approach for electricity generated is very sensitive and has significant impact on the forecasting accuracy. For further improvement, we continue the search for a much better forecast using Genetic Algorithm which is a useful tool for searching solution in optimization. GA may be used as the forecast optimizer. It can be done by integrating GA searching method into improving parameters estimates for SARIMA model.

Acknowledgement

The research is supported by IRPA Grant Vot. 74047 and Biasiswazah UTM-PTP for the postgraduate student organized by the Pusat Pengajian Siswazah, UTM and Department of Mathematics, Faculty of Science UTM. The authors would like to thank lecturers and friends for many helpful ideas and discussions. We also wish to thank TNB for providing necessary data in this study.

References

[1] B.L. Bowermann & R.T. O′ Connel, *Forecasting and Time Series: An Applied Approach,* 3rd Ed. Duxubury Press, Belmont, 1990.

[2] G. E. P Box & G. M. Jenkins, *Time Series Analysis: Forecasting and Control Revised Edition,* Holden-Day, San Francisco, 1976.

[3] G. E. P Box & G. M. Jenkins, *Time Series Analysis: Forecasting and Control,* Prentice Hall, New Jersey, 1970.

[4] P.J. Brookwell & R.A.Davis, *Time Series:Theory and Methods 2nd ed.,* New York, 1991.

[5] C. Chatfield, *The Analysts of Time Series : An Introduction,* Chapmann and Hall, 1984.

[6] Marek Obitko, *Genetic Algorithm, Hochschule fur Technik und Wirtschaft Dresden (FH),* Czech Technical University, 1998.

[7] O.D. Anderson, *Time Series Analysis and Forecasting: The Box-Jenkins Approach,* Butterworths & Co., 1976.

[8] Zuhaimy Hj. Ismail & Maizah Hura Ahmad, *Surveys On Forecasting Used In Commercial Enterprises,* Laporan Teknik FSUTM, LT/M Bil 1(2000).

[9] Zuhaimy Hj. Ismail & Mohd. Nizam Kahar, *Preliminary Study On Using Genetic Algorithm In Triple Smoothing Method of Forecasting,* Laporan Teknik FSUTM, LT/M Bil 3 (2000).