

Dynamic Regression Intervention Modeling for the Malaysian Daily Load

Fadhilah Abd. Razak

Engineering Mathematics and Management Unit

Department of Mechanical Engineering

College of Engineering, Universiti Tenaga Nasional, Malaysia

fadhilah@uniten.edu.my

Mahendran Shitan

Institute for Mathematical Research

University Putra Malaysia

sarasmahen@gmail.com

Amir Hisham Hashim

Department of Electrical Power Engineering

College of Engineering, Universiti Tenaga Nasional, Malaysia

amir.hisham@tnb.com.my

Izham Zainal Abidin

Department of Electrical Power Engineering

College of Engineering, Universiti Tenaga Nasional, Malaysia

izham@uniten.edu.my

Abstract

Malaysia is a unique country due to having both fixed and moving holidays. These moving holidays may overlap with other fixed holidays and therefore, increase the complexity of the load forecasting activities. The errors due to holidays' effects in the load forecasting are known to be higher than other factors. If these effects can be estimated and removed, the behavior of the series could be better viewed. Thus, the aim of this paper is to improve the forecasting errors by using a dynamic regression model with intervention analysis. Based on the linear transfer function method, a daily load model consists of either peak or average is developed. The developed model outperformed the seasonal ARIMA model in estimating the fixed and moving holidays' effects and achieved a smaller Mean Absolute Percentage Error (MAPE) in load forecast.

Keywords: Dynamic Regression, Linear Transfer Function, Moving Holidays' Effects, SARIMA.

1. Introduction

Growing demands by stakeholders for efficient operation and dispatch of the power system has increased the need to have a more robust, reliable and accurate load forecasting. The forecasting errors for holidays are known to be higher than other factors. Malaysia has both fixed holidays such as Christmas and New Year, and moving holidays such as Aidil Fitri and Chinese New Year. The moving holidays are moving according to three different calendars. Chinese New Year follows Chinese lunar calendar, Aidil Fitri and Aidil Adha follow Islamic lunar calendar and Deepavalli follows Hindu calendar. Chinese New Year and Deepavalli move within two specific months of the Gregorian calendar but Aidil Adha and Aidil Fitri move forward from one period to the next over

the entire Gregorian calendar. As such the load will be affected more significantly by them compared to the fixed holidays and these effects are called ‘moving holidays’ effects’ (Lazim et al., 2006).

A statistically based technique which has been applied successfully for many years in this area and enjoyed great success in practice for short term forecast as discussed in (Ismail and Jamaludin, 2008; Ismail et al., 2009a) was related to these holidays effects. The Spanish system’s daily model was computed for forecasting the daily load up to ten days ahead, derived by starting from an ARIMA (auto-regressive integrated moving average) model with dummy variables to capture the influence of special days which include holidays was discussed in (Cancelo et al., 2008). The issue of identifying and eliminating deterministic seasonality due to holidays tied to the lunar calendar for Turkey was addressed in (Emre and Boragan, 2001). It was found out in (Lin and Liu, 2002) that adding holiday regressors can effectively control the impact of moving holidays and improve the seasonal decomposition of the selected series in Taiwan. The seasonal adjustment procedures using SEAM (Seasonal Adjustment for Malaysia) and Reg ARIMA (adjusted for Malaysia) were applied in (Lazim et al., 2006) to eliminate the presence of the moving holiday effects and seasonally adjusted the Malaysian economic time series data.

The effects of New Year holidays were estimated in (Razak et al., 2010) by the dynamic regression (DR) model based on the intervention analysis applied on to the daily peak load. The model achieved a smaller Mean Absolute Percentage Error (MAPE) value than the seasonal ARIMA (SARIMA) model based on seven days ahead and performed better in predicting the next New Year holiday ahead. The DR model was also discussed by Razak et al. (2011) to analyze the moving holidays’ effects and found that the observed series fit well with the predicted series. However the error made earlier by the DR model could be reduced if the sample size of the data is increased. Since the load will be affected more significantly by moving holidays compared to the fixed holidays, the DR model is developed further in this paper to deal with the sensitivity of the moving holidays using a larger sample size. A comparison is made between the DR model and SARIMA model in estimating the holidays’ effects and hence to forecast the daily peak and average load in Malaysia. The final selection of the appropriate models depends on the MAPE value and a variety of tools, which include the sample autocorrelation function (ACF), the sample partial autocorrelation function (PACF) and the Ljung-Box Chi-squared statistics.

The outline of the paper is as follows: Section 2 explains briefly the concepts of dynamic regression intervention model. Section 3 discusses the methodology of the modeling. Section 4 presents the results and discussion and Section 5 concludes.

2. Dynamic Regression Intervention Model

This paper aims to develop an improved load forecasting model that can capture the influence of Malaysian fixed and moving holidays. Thus, a dynamic regression intervention model is considered since holidays can be the intervention variables which may exert their influence on the daily load dynamically across several future time periods. The basic concepts and modeling of the model will be discussed in this section.

2.1 Dynamic Regression Model

Dynamic regression model provides all the advantages of a multiple regression model with the time series features of an ARIMA model. The basic form of the DR model in general is given by

$$Y_t = C + f(X_t) + N_t \quad (1)$$

where X_t is the explanatory or independent variable or input series, Y_t is the forecast variable or output series, C is the constant term, $f(X_t)$ is the transfer function which describes how a change in X_t is transferred to Y_t across several time periods and N_t is the stochastic disturbance series which is an ARMA process uncorrelated with X_t . The disturbance series N_t represents the effects of all excluded inputs on the variability of Y_t whilst the constant C captures the effects of excluded inputs on the overall level of Y_t (Makridakis et al., 1998).

2.1.1 Linear Distributed Lag

The effect of any holiday X_t on the electrical load lasts for some time due to many customers going for a long break until the load would return back to normal. So, the daily electrical load Y_t can be modeled as a transfer function $f(X_t)$ of the holiday and few days prior the holiday $X_t, X_{t-1}, X_{t-2}, \dots$. It can be written as a linear distributed lag of X_t values given by

$$f(X_t) = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots + v_k X_{t-k} \quad (2)$$

The coefficients v_0 through v_k are called transfer function weights and k is the longest lag of X used. Equation (2) can be written in a simpler form using the backward shift operator B defined as $B^k X_t = X_{t-k}$ (Pankratz, 1991).

$$\begin{aligned} f(X_t) &= v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots + v_k X_{t-k} \\ &= (v_0 + v_1 B + v_2 B^2 + \dots + v_k B^k) X_t \\ &= v(B) X_t \end{aligned} \quad (3)$$

2.1.2 Rational Distributed Lag

The lag k in Equation (3) can be a large number that involves a longer time in estimating the parameters v_0 through v_k . Thus, $v(B)$ is rewritten in terms of rational polynomial distributed lags defined as

$$v(B) = \frac{\omega(B)}{\delta(B)} B^b, \quad -1 < \delta < 1 \quad (4)$$

where

$$\omega(B) = \omega_0 + \omega_1 B + \omega_2 B^2 + \dots + \omega_h B^h \quad (5)$$

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r \quad (6)$$

and the dead time denoted as b is the delay periods before X_t influences Y_t , h is the order of the ω function that controls the number of transfer function coefficients before they begin to decay and r is the order of the δ function that controls the decay pattern. Hence, the DR model given in Equation (1) yields to a simpler form as

$$Y_t = C + \frac{\omega(B)}{\delta(B)} X_{t-b} + N_t \quad (7)$$

Rational forms $v(B)$ as given in Equation (4) can capture a wide variety of patterns with just a few parameters in most applications (Pankratz, 1991). Since holidays are the main contributing factors to the electrical load, they will be considered as interventions for the independent variable X_t in the DR model and their precise impact will be measured using the intervention analysis. Such model becomes a special case of a DR model and can be called DR intervention analysis model.

2.2 Intervention Analysis Model

The intervention may due to identified events or unexplained outliers (unusual observations) in a time series. There are two types of intervention analysis:

2.2.1 Step Intervention

Step intervention involves with a permanent change and expected to occur in a sudden drop or rise in the forecast variable (Makridakis et al., 1998). The step intervention occurred at time t is defined as the dummy deterministic variable X_t .

$$X_t = \begin{cases} 0 & t < i \\ 1 & t \geq i \end{cases} \quad (8)$$

2.2.2 Pulse Intervention

Pulse intervention involves with a temporary change. Interventions such as holidays yield temporary effects on the load and thus are under the pulse-based interventions. Suppose a particular holiday X_t occurs during time $t = i$ and does not occur at other times. Then it is assigned as a binary independent variable as

$$X_t = \begin{cases} 0 & t \neq i \\ 1 & t = i \end{cases} \quad (9)$$

If X_t is assumed to be any public holidays which occur at various time t , then the effect of the pulse intervention on series Y_t will involve a multi-period temporary response.

Each type of holidays is defined as in Equation (9), and denoted as $X_{i,t}$ for $i = 1, 2, \dots$. Since all these holidays are pulse intervention variables the model given in Equation (7) can be written compactly as

$$Y_t = C + \sum_{i=1}^j \frac{\omega_i(B) B^{b_i}}{\delta_i(B)} X_{i,t} + N_t \quad (10)$$

This model is also studied as the intervention model for analyzing the impact of extreme change on the arrival of tourist into Bali (Ismail et al., 2009b), to examine the impact of policy guidance on transportation demand (Chen et al., 2003) and to investigate the effects of the road safety operations on road accidents in Malaysia (Ya'akob et al., 2011).

2.3 ARIMA Model

Many non-stationary series are represented by an autoregressive integrated moving average (ARIMA) models. The ARIMA model or process $\{Y_t\}$, denoted by ARIMA (p, d, q) is given by

$$(1-B)^d \phi(B)Y_t = \theta(B)Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2) \quad (11)$$

where $\phi(B)$ is an autoregressive (AR) operator of order p and $\theta(B)$ is a moving average (MA) operator of order q as given by

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (12)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (13)$$

using the backward shift operator B to describe the process of differencing as defined by

$$BY_t = Y_{t-1}, \quad B^j Y_t = Y_{t-j}, \quad j = 0, 1, \dots \quad (14)$$

The integrated term of ARIMA is represented by d which is a degree of the first difference denoted as $(1-B)$. Thus, a non-stationary ARIMA reduces to stationary ARMA (p, q) model whenever $d = 0$ (Brockwell and Davis, 2002).

The ARIMA model can be extended to incorporate seasonality and known as seasonal ARIMA or SARIMA models. One can identify the existence of seasonality in the series from a large autocorrelation coefficient or partial correlation coefficient at the seasonal lags and significantly different from zero. It is a pattern that repeats itself over fixed intervals of time. So for a daily series, multiples of the seasonal lag 7, 14 ... might be seen such as illustrated by the sample autocorrelation function (ACF) in Figure 1.

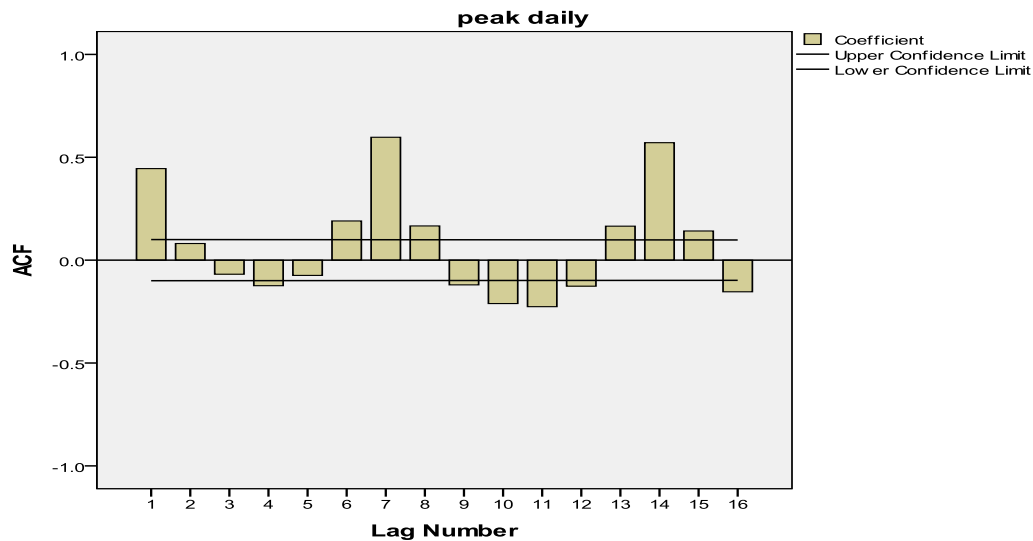


Figure 1: Seasonality apparent in a daily series

The general multiplicative SARIMA model $\{Y_t\}$ denoted by $ARIMA(p, d, q)(P, D, Q)_s$ where (p, d, q) refers to the non-seasonal parameters and $(P, D, Q)_s$ refers to the seasonal parameters with s as the number of seasons. The model $\{Y_t\}$ is defined as

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Y_t = \theta(B)\Theta(B^s)Z_t \quad (15)$$

where d and D are nonnegative integers (Brockwell and Davis, 2002; Shumway and Stoffer, 2006). The seasonal polynomials $\Phi(z)$ and $\Theta(z)$ to represent the orders P of AR and Q of MA models respectively are multiplied into Equation (15). Since ARIMA model has a variety of order terms, the Box-Jenkins methodology is used to decide on the most appropriate SARIMA model for the disturbance series N_t in the DR intervention model.

3. Methodology

The methodology of developing an appropriate dynamic regression intervention analysis model is based on the linear transfer function. The flowchart in Figure 2 summarizes the network of the stages involved in the method. The modeling activities for each stage will be presented in this section. The aim is to develop a daily load model that able to give the best estimate of the holidays' effects and hence to forecast a week ahead of the series. A benchmark of ARIMA model will be used.

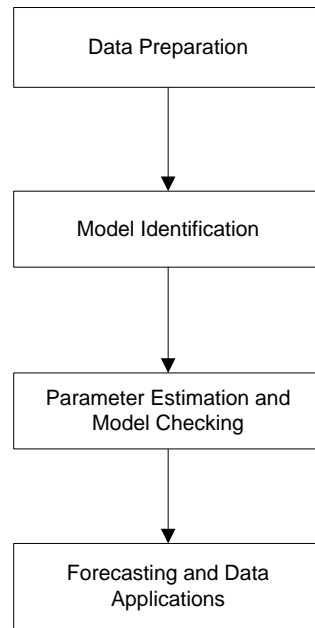


Figure 2: Flow Chart of Intervention Modeling Activities

3.1 Data Preparation

Stage 1 of the modeling activities is data preparation. The data used in this paper was a power load profile measured in Megawatts (MW) and the plot shown in Figure 3 represented the peak daily load series Y_t from January 1st, 2002 to December 31st, 2004.

The variance of the series is not constant and thus the natural logarithmic transformation is used to equalize the variability over the length of the series. The series also clearly shows an upward trend and a weekly seasonality with many troughs, possibly due to weekends and holidays. These patterns reveal that the series is not stationary and hence the data is differenced at lag 1 and 7 to obtain an approximate stationary time series. Thus, Equation (10) yields to

$$\Delta_7 \Delta \ln Y_t = \sum_{i=1}^j \frac{\omega_i(B) B^{b_i}}{\delta_i(B)} \Delta_7 \Delta X_{i,t} + \Delta_7 \Delta \ln N_t \quad (16)$$

where the constant term C is deleted due to its insignificance after differencing.

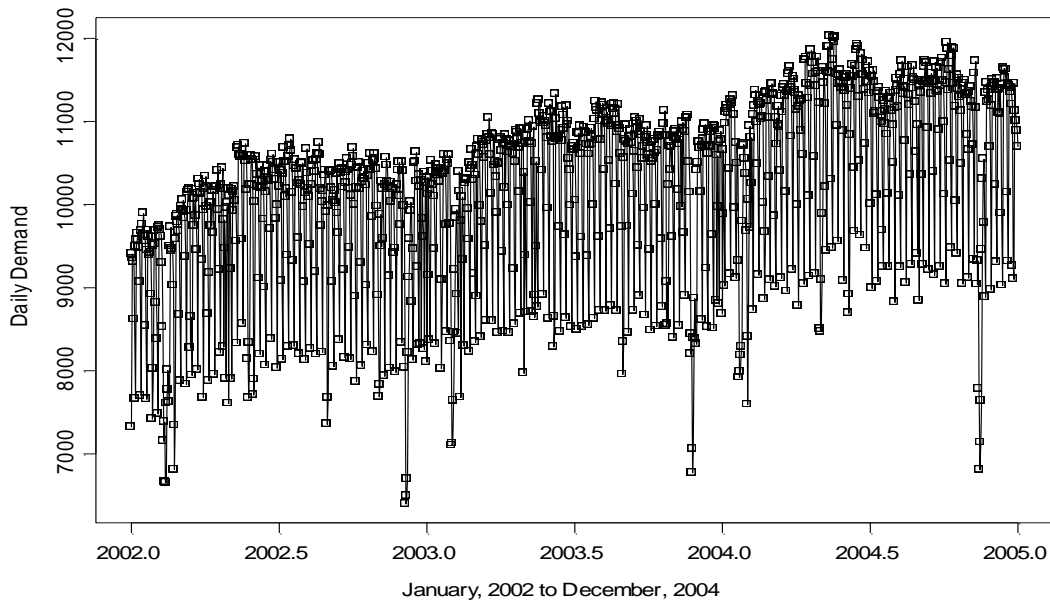


Figure 3: The three-year-series of Peak Daily Load

The next step of data preparation is to fit the model in Equation (16) to get the estimates of the regression parameters. An approximate low-order proxy autoregressive (AR) model such as AR (1) is used for the errors or disturbance series N_t in the model. The sample ACF of the residuals is studied to check how quickly the autocorrelation coefficients go to zero. Repeat the estimation of the preliminary model until the errors in the disturbance series become stationary.

3.2 Model Identification

Stage 2 of the intervention modeling is model identification. It is considered once the disturbance or frequently called residual series is stationary to identify the transfer function for the intervention variable and the ARMA model for the disturbance series based on the estimated parameters. The Box-Jenkins' principle of parsimony is applied in the identification of both models. More parsimonious form of models requires less parameter to estimate but able to adequately describe the data. Redundant parameters

may lead to unstable estimates in the models (Box et al., 1994). Linear transfer function method is used for the identification of both models as explained next.

3.2.1 Identification of Transfer Function

The linear transfer function method involves in selecting proper orders of one or more tentative rational form for the transfer function $v(B)$ for each intervention variable $X_{i,t}$. The transfer function can be identified initially by a hypothesis and then followed by an empirical approach. There are four commonly encountered types of pulse responses (Pankratz, 1991):

- a) $v(B) = \omega_0, b = 0, h = 0, r = 0$
- b) $v(B) = \frac{\omega_0}{1 - \delta_1 B}, b = 0, h = 0, r = 1$
- c) $v(B) = \omega_0 + \omega_1 B, b = 0, h = 1, r = 0$
- d) $v(B) = \frac{\omega_0 + \omega_1 B}{1 - \delta_1 B}, b = 0, h = 1, r = 1$ (17)

Since the intervention variables consist of different types of holidays, the dead time is $b = 0$ indicating an immediate or no delay period response.

If any intervention gives a one-period temporary effect at time $t = i$, its transfer function is function (a) where ω_0 is the size of the displacement of the load Y_t during time t . On the other hand, if the response is gradual then the function is in part (b) which occurs most.

The function (c) illustrates a dynamic response of the intervention on a series Y_t which rises by ω_0 units during period i and by added ω_1 units during period $i + 1$ and remains at the equilibrium level after $t = i + 1$. On the other hand, if the response is gradual after $t = i + 1$ with the function $|\delta| < 1$ that controls the decay pattern until the intervention slowly disappears, then the function is part (d). The application of these linear transfer functions will be discussed in later section.

3.2.2 Identification of ARMA Disturbance

The next step is to identify the appropriate ARMA model for the disturbance series or to calculate the residuals N_t from the DR model after been captured by the intervention variables. The estimated residual series N_t for daily load model can be calculated to get

$$\Delta_7 \Delta \ln N_t = \Delta_7 \Delta \ln Y_t - \sum_{i=1}^j \frac{\omega_i(B) B^{h_i}}{\delta_i(B)} \Delta_7 \Delta X_{i,t} \quad (18)$$

Plots of the sample ACF and partial autocorrelation function (PACF) of residuals are examined for the potential seasonal ARIMA $(p, d, q) (P, D, Q)_s$ model for N_t .

3.3 Parameter Estimation and Model Checking

Stage 3 of the intervention modeling is parameter estimation and model checking. It involves in refitting the complete DR intervention analysis model for daily load series using the new fitted ARMA model for N_t from Equation (18). Each parameter in both potential models of ARMA disturbance and the transfer function $v_i(B)$ for $i = 1, 2, \dots, j$ of the intervention variables are re-estimated using the conditional least squares and maximum likelihood estimation method. Besides being significant parameters for the potential model, these parameters has to satisfy certain requirements of stationarity, invertibility and stability for both seasonal and non-seasonal parameters (See Pankratz, 1991). Any insignificant parameters will be deleted if they are any.

The usual diagnostic tests for model checking were performed to validate the final derivation of both the intervention part and the disturbance part in the full complete DR intervention model before forecasting. The residual plots of ACF and PACF are observed to make sure that there are no autocorrelations significantly different from zero. These plots should behave as the white noise for an adequate disturbance ARIMA model. White noise data have a sampling distribution that can be approximated by a normal curve with mean zero and standard error $1/\sqrt{n}$ where n is the number of observations in the series.

There is a Ljung-Box χ^2 statistic test to detect an inadequate disturbance ARIMA model. If the test statistic value is less than the χ^2 critical value at the 5% level, then the approximate χ^2 test does not suggest inadequacy in the disturbance ARIMA model.

3.4 Forecasting and Model Applications

The final stage of the intervention modeling is forecasting and model applications. Once the complete DR model is identified with all the estimated parameters, the forecasts of future values of the load are produced. The DR intervention daily peak load model will be used to forecast seven days ahead based on public holidays' effects. Five-year series of daily average load model will be applied to forecast other scenarios.

The fitted model can be assessed further by looking into the forecasting accuracy criterion. The criterion chosen to measure the accuracy of the forecast in this study is the Mean Absolute percentage Error (MAPE) which is given by;

$$\text{MAPE} = \frac{\sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right|}{n} \cdot 100\% \quad (19)$$

where x_i and \hat{x}_i are the actual observed values and the predicted values respectively while n is the number of predicted values. The rational justification of using the MAPE compared to the other criteria is due to its scale free and relative or percentage error measures.

4. Results and Discussion

The results obtained based on the linear transfer function methodology of developing an appropriate dynamic regression intervention analysis model is presented in this section. Two cases are considered as follows:

4.1 Three-Year Series of Daily Peak Load

The 3-year series daily peak load was prepared as in Section 3.1. The potential models fitted for the disturbance series $\Delta_7 \Delta \ln N_t$ in Equation (18) are seasonal ARIMA (p, d, q) $(P, D, Q)_s$ models with three different parameters as $(1, 1, 1)$ $(0, 1, 1)_7$, $(2, 1, 1)$ $(0, 1, 1)_7$ and $(1, 1, 2)$ $(0, 1, 1)_7$. As such there will be three respective potential DR intervention models for the load $\Delta_7 \Delta \ln Y_t$ as given in Equation (16). The goodness of fit of the potential models is justified by the significant Ljung-Box chi-squared test at 5% level shown by Table 1. Besides the MAPE measures shown by the table as the model fit statistics, the normalized Bayesian Information Criterion (BIC) measures are also recorded. The BIC criterion is an order selection criteria for ARIMA models. Similarly to MAPE, the smallness of the criterion value is indicative of a more appropriate model. However, both model fit statistical measures of BIC and MAPE recorded for seasonal ARIMA and DR intervention models are about the same for each respective parameter.

Table 1: Model Fit Statistics for the potential Models

Model	Parameters	Model Fit Statistics		Ljung-Box Q(18)	
		BIC	MAPE	statistics (df=15)	Sig
Seasonal ARIMA (No predictors)	(111)(011) ₇	12.439	3.159	16.68	0.34
	(112)(011) ₇	12.442	3.147	14.39	0.42
	(211)(011) ₇	12.442	3.146	14.87	0.39
DR Intervention (4 predictors)	(111)(011) ₇	11.989	2.49	16.68	0.34
	(112)(011) ₇	11.991	2.49	14.39	0.42
	(211)(011) ₇	11.992	2.49	14.87	0.39

Based on smaller MAPE relative error measures for the model fitting, DR intervention models may perform better than seasonal ARIMA models in forecasting public holidays. A comparison is done further by calculating the peak daily load forecasts. The significant intervention variables considered in each potential DR intervention model which act as predictors are 3 moving holidays: Aidil Adha $X_{1,t}$, Aidil Fitri $X_{2,t}$, Chinese New Year $X_{3,t}$; and New Year $X_{4,t}$ as a selected public holiday. The linear transfer functions $v(B)$ given in Equation (17) are referred when estimating each holiday's effect. It is expected that Aidil Fitri holidays give the most impact on the load, followed by Chinese New Year, Aidil Adha and finally New Year. Due to their different holiday patterns, the function defined in part (d) is fitted to both Aidil Fitri $X_{2,t}$ and Chinese New Year $X_{3,t}$ while the same rational forms of function in part (a) are fitted to Aidil Adha $X_{1,t}$ and New Year $X_{4,t}$.

The MAPE is calculated based on seven-day-ahead forecasts each starting with four different festival events. The results are presented in Table 2. DR intervention models achieved less than 4% for the MAPE measures of accuracy compared to Seasonal ARIMA models due to their better recognition of the interventions' effects. Thus, Seasonal ARIMA models failed to explain well the variation in the electrical load due to these festival events especially dealing with moving holidays.

Table 2: The MAPE for Peak Daily Load Forecasts

Festival Events	Seasonal ARIMA			DR Intervention		
	(111) (011) ₇	(112) (011) ₇	(211) (011) ₇	(111) (011) ₇	(112) (011) ₇	(211) (011) ₇
New Year 2005	3.45	3.28	3.26	1.48	1.5	1.5
Chinese New Year 2004	7.61	7.9	7.91	3.01	2.99	3.01
Aidil Adha 2003	8.07	8.22	8.22	2.21	2.34	2.29
Aidil Fitri 2003	9.22	9.68	9.69	3.27	3.38	3.38

Among the three potential DR intervention models, seasonal ARIMA (111) (011)₇ model is chosen for the disturbance series $\Delta_7 \Delta \ln N_t$. The reason is due to a smaller MAPE value as highlighted in Table 2 and also due to a less significant parameter estimate of both non-seasonal MA₂ and AR₂ in ARIMA (112) (011)₇ and ARIMA (211) (011)₇ respectively. The chosen model is defined as

$$\Delta_7 \Delta \ln N_t = \frac{(1 - \Theta_1 B^7)(1 - \theta_1 B)}{(1 - B^7)(1 - \phi_1 B)} a_t \quad (20)$$

where a_t is the white noise residuals.

Table 3 presents all the parameter estimates obtained in the full DR intervention model Y_t satisfying the required conditions of invertibility, stationarity and stability. These parameters are also significant at better than 5 % level. The absolute t -values are all greater than 2 indicating that the model Y_t is appropriate and it is given by

$$\begin{aligned} \Delta_7 \Delta \ln Y_t = & \omega_{1,0} \Delta_7 \Delta X_{1,t} + \frac{(\omega_{2,0} + \omega_{2,1} B)}{1 - \delta_{2,1} B} \Delta_7 \Delta X_{2,t} + \frac{\omega_{3,0} + \omega_{3,1} B}{1 - \delta_{3,1} B} \Delta_7 \Delta X_{3,t} \\ & + \omega_{4,0} \Delta_7 \Delta X_{4,t} + \frac{(1 - \Theta_1 B^7)(1 - \theta_1 B)}{(1 - B^7)(1 - \phi_1 B)} a_t \end{aligned} \quad (21)$$

where $X_{1,t}$, $X_{2,t}$, $X_{3,t}$ and $X_{4,t}$ represent the holidays with their fitted transfer functions.

Table 3: Estimated Parameters for DR Intervention Model of Peak Daily Load

Variable	Parameter	Estimate	Std. Error	t Value	p-Value
AR ₁	ϕ_1	0.336	0.034	9.796	0.000
MA ₁	θ_1	0.961	0.010	91.856	0.000
MA _{S1}	Θ_1	0.996	0.026	38.359	0.000
Aidil Adha	$\omega_{1,0}$	-0.223	0.022	-10.183	0.000
Aidil Fitri	$\omega_{2,0}$	-0.320	0.024	-13.057	0.000
Chinese New Year	$\omega_{2,1}$	0.175	0.029	6.078	0.000
	$\delta_{2,1}$	0.661	0.035	19.000	0.000
	$\omega_{3,0}$	-0.255	0.024	-10.609	0.000
	$\omega_{3,1}$	0.070	0.029	2.416	0.016
	$\delta_{3,1}$	0.735	0.040	18.359	0.000
New Year	$\omega_{4,0}$	-0.181	0.011	-16.461	0.000

The MAPE errors recorded by DR intervention model can be reduced by increasing the sample size of observations and including other significant factors. A similar application on a five-year series of daily average load is discussed next.

4.2 Five-Year Series of Daily Average Load

The average daily load instead of the peak daily load data is studied next for the predictions of seven-days prior to, during and ahead of moving holidays. The observed data Y_t is a 5-year series taken from 1st September 2000 to 31st August 2005. The daily data still behave similarly as before and as such the same ARIMA(1,1,1)(0,1,1)₇ model is fitted successfully for the differenced disturbance series N_t . After considering all i significant moving and fixed holidays as intervention variables $X_{i,t}$ at different time t , the DR intervention model for the average daily load is obtained in the same form as in Equation (16). The parameters of each transfer function for $X_{i,t}$ and parameters for ARIMA model are optimally estimated. The adequacy of the model is justified significantly at 5% level with the Ljung-Box chi-squared statistic value $\chi^2 = 24.26$ and the MAPE records 1.595. Two different predictions are considered as follows:

4.2.1 Predictions based on Moving Holidays' Effects

The DR intervention model is used to predict seven days before, during and after each moving holiday from Year 2003 to 2004. Table 4 presents the MAPE from 1.02 % to 3.75% based on each moving holiday's effect on the electrical load occurring in a fortnight period. The highest errors corresponding to Aidil Fitri and Deepavalli in Year 2004 are due to their overlapping holidays which make it difficult to get good predictions. However, the MAPE is less than 3% on the average and it shows that the model is satisfactory.

Table 4: The MAPE based on The Effects of Moving Holidays

Year	Aidil Fitri	Aidil Adha	Chinese New Year	Deepavalli
2003	1.36	1.12	2.66	1.02
2004	3.75	2.73	1.79	3.62
Average MAPE	2.56	1.92	2.22	2.32

4.2.2 Prediction based on 2005 Chinese New Year's Effects

The DR intervention model is applied again to the daily average data based on a 5-year series. The predictions are based on seven days before, during and after Chinese New Year 2005 which fall on Wednesday 9th February. The same predictions using Seasonal ARIMA (1, 1, 1) (0, 1, 1)₇ model (SARIMA) are also produced. Figure 4 illustrates the comparison between these two models with the actual daily load (the line with circles). It is apparent from the figure that SARIMA model (the line with crosses) failed to capture the critical-days' effect from Monday 7th to Sunday 13th February due to the holiday compared to the DR model (the line with squares).

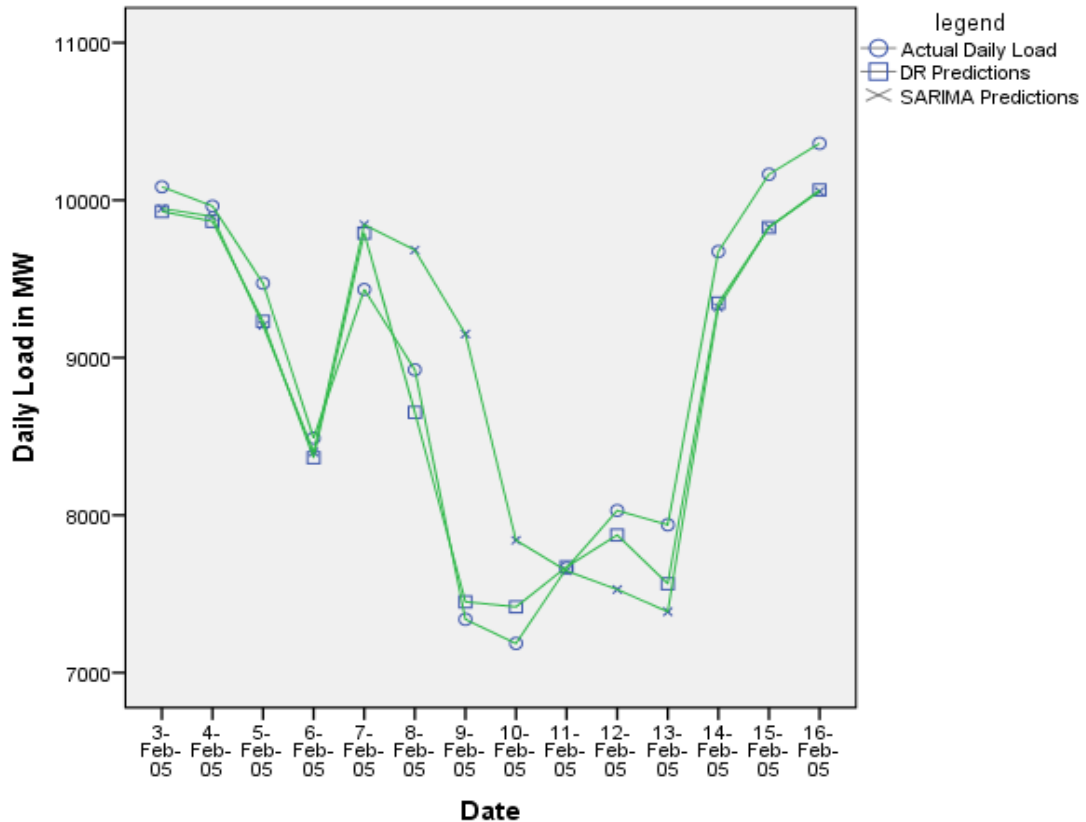


Figure 4: Comparison of models for Chinese New Year 2005

This period is the most challenging to predict as the long weekend effect is uncertain to occur either before or after such Wednesday holiday. However, DR intervention model still records a good MAPE measure of 2.46% compared to SARIMA model with more than 5 % error. The DR intervention model is expected to be improved better in detecting such occurrence if more samples are available.

5. Conclusion

Two types of dynamic regression (DR) intervention model forecasts which are equally good in their performance are discussed in this chapter for the power system company to deal with the sensitivity of the Malaysian public holidays. The peak and average daily load are aimed at producing forecasts up to seven days ahead for network outage planning. The modeling is based on the linear transfer function method which able to explain well the variation in the electrical load due to any interventions such as holidays. The forecasting performance was evaluated by analyzing the errors based on the effects of holidays. Since these effects reduce significantly the routine activities for the periods in the vicinity of the holiday dates, the predictions for seven days before holidays are also calculated.

DR intervention model for the peak daily load achieved less than 4% for the MAPE value compared to SARIMA model based on seven-day-ahead forecasts. Similarly as for the average daily load DR intervention model. It recorded the average MAPE of less than 3 % based on the effects of Chinese New Year 2005 compared to SARIMA which record the MAPE measure of more than 5%. Thus, the role of the intervention variables in the DR model is significant in explaining the variation in the electrical load especially dealing with moving holidays.

DR intervention model can be proposed for forecasting the electrical load during any public holidays in Malaysia. The result of this study may help the utility company to predict the future demand of the customers better and to minimize inaccuracies with respect to over or under forecast when predicting for any future load forecasts especially dealing with overlapping or consecutive holidays. The model will be designed to capture the load characteristics using the predefined intervention variables.

References

1. Box, G.E.P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, Third Edition, Prentice Hall, Upper Saddle River, New Jersey.
2. Brockwell, P.J. and Davis, R.A. (1996). *Introduction to Time Series and Forecasting*, Springer Texts in Statistics, Springer-Verlag New York.
3. Cancelo, J. R., Espasa, A. and Grafe, R. (2008). Forecasting the Electricity Load from One Day to One Week Ahead for the Spanish System Operator, *International Journal of Forecasting*, 24, 588-602.

4. Chen, M. C., Chen, K. K. and Pan, A. C. M. (2003). Establishing an Intervention Model to Examine the Impact of Policy Guidance on Transportation demand, *Journal of the Eastern Asia Society for Transportation Studies*, 5, 1565-1575.
5. Emre, A. C. and Boragan S. A. (2001). Moving Holidays and Seasonality: An application in the Time and Frequency Domains for Turkey, Working paper, Department of Economics, Bogazici University, 07.
6. Ismail, Z. and Jamaluddin, F. (2008). Time Series Regression Models for Forecasting Malaysian Electricity Load Demand, *Asian Journal of Mathematical Statistics*, 1, 139-149.
7. Ismail, Z., Yahaya, A. and Mahpol, K.A. (2009a). Forecasting Peak Load Electricity Demand Using Statistics and Rule Based Approach, *American Journal of Applied Sciences*, 6(8), 1618-1625.
8. Ismail, Z., Suhartono, Yahaya, A. and Effendi, R. (2009b). Intervention Model for Analyzing the Impact of terrorism to Tourism Industry, *Journal of Mathematics and Statistics*, 5(4), 322-329.
9. Lazim, M. A., Shuja, N. and Wah, Y.B. (2006). Deseasonalising Economic Time Series for Non-Fixed Seasonal Effect, National Statistics Conference, Putrajaya International Convention Centre, Putrajaya, Malaysia.
10. Lin, J.L. and Liu, T. S. (2002). Modeling Lunar Calendar Holiday Effects in Taiwan, Presented at workshops on Macroeconomic Modelling, Institute of Economics, Academia Sinica, Taipei.
11. Makridakis, S., Wheelwright, S.C. and Hyndman, R. J. (1998). *Forecasting: Methods and Applications*, 3rd Edition. J.W. Wiley.
12. Pankratz, A. (1991). *Forecasting with Dynamic Regression Models*, John Wiley and Sons Inc.
13. Razak, F. A., Hashim, A. H., Abidin, I. Z. and Shitan, M. (2010). Moving Holidays' Effects on the Malaysian Peak Daily Load, IEEE Power and Energy International Conference (PECON), 906-910.
14. Razak, F. A., Hashim, A. H. and Abidin, I. Z. (2011). Dynamic Regression Model for Analyzing the Holidays' Effects to the Malaysian Load, Y1215, 517-520,
15. Proceedings of the 2011 4th International Conference on Advanced Computer Theory and Engineering (ICACTE), ISBN: 978-0-7918-5993-3.
16. Shumway, R.H. and Stoffer, D.S. (2006). *Time Series Analysis and Its Applications: With R Examples*, USA, Springer.
17. Yaacob, W.F.W, Husin, W.Z.W, Aziz, N. A. and Nordin, N. I. (2011). An Intervention Model of Road Accidents: The Case of OPS Sikap Intervention, *Journal of Applied Sciences*, 11 (7), 1105-1112, ISSN 1812-5654, Asian Network for Scientific Information.