

Calculation of a risk measure for the net system load profile

Ken Russell
University of Wollongong

Pietro Cerone
Victoria University

Vivien Challis
The University of Queensland

1 Introduction

Integral Energy is one of three franchises which provide retail electricity in New South Wales (NSW). Integral Energy purchases wholesale electricity from the National Electricity Market and sells this to retail customers. The electricity market is unusual because the price at which electricity is sold to retail customers is fixed, but the price that the electricity retailer must pay for electricity from the National Electricity Market varies. Thus price changes incurred by Integral Energy are not passed on to their customers, introducing electricity price risk. Integral Energy uses the electricity hedge market, customer contract management, and the Electricity Tariff Equalisation Fund, along with market forecasting, to reduce its exposure to electricity price risk. These contracts are used to hedge Integral Energy against risk for electricity demand within the 95% confidence intervals of their long range electricity demand forecasts.

This project is in part concerned with comparing Integral Energy's electricity price risk with that of the other two NSW electricity franchises: Energy Australia and Country Energy. Integral Energy want a measure of the relative volatility of their demand compared with that of the other two franchises. Integral Energy is most at risk during very high load periods, such as very hot days in summer and very cold days in winter. Integral Energy's customers are largely based in western Sydney, where it gets hotter in summer and cooler in winter than on the coast. Houses in this area also tend to be less energy efficient, and Integral Energy believe these factors combine to produce higher demand during peak periods. Energy Australia's customers are largely based on the coast around Sydney. Coastal temperatures are more stable than inland temperatures, which Integral Energy suggest results in less volatile demand. Country Energy largely provides electricity for the rest of the state and, from historical data, appears to have the least volatile demand.

Integral Energy provided data collected every 30 minutes from 1 January, 2002 to 31 December, 2006. The actual values for demand were given, as well as Integral Energy's long term forecast values and their 95% confidence intervals. The supplied data included the temperature at Bankstown (in western Sydney), the National Electricity Market electricity price, and various load measurements. For this project, Integral Energy wanted us to consider the *Net System Load Profiles* (NSLPs) for each franchise. The NSLP is a measure of the load

over which the retailer has no control; it is discretionary electricity consumption controlled by the customers. The NSW State Load was also found to be of use in the modelling.

The brief from Integral Energy was to find a model for their NSLP given the temperature in Bankstown, which is used as a proxy in lieu of detailed temperature variation for the franchise, to estimate discretionary Load. This model would, in effect, be used to calculate expected load based around temperature in the short term. The second major task was to quantify the difference in volatility between the three franchise NSLP profiles.

2 Determination of the model

The data that Integral Energy provided came as an Excel spreadsheet of 87 648 rows and 61 columns. The rows each represent one observation on the characteristics of the electricity system. These observations are taken every 30 minutes throughout the day (at 00:30, . . . , 24:00), for each day of the year. Data were provided for the calendar years 2002, 2003, . . . , 2006.

The decision was taken to find a model that explained the Historical NSLP for Integral Energy over the years 2002 to 2005, and then to use the variables that were selected for that model to predict NSLP during 2006. A resulting good fit would suggest that the basic form of the model is satisfactory. It is important to note that this will not give exactly the same model for 2006 as for 2002 to 2005. For example, suppose that the best-fitting model for 2002 to 2005 is Predicted NSLP = $a + b \text{temp} + c \text{temp}^2$. Then the model that is fitted for 2006 would be of the same form, but new coefficients a_1 , b_1 and c_1 would be calculated.

All statistical analyses were done using version 5.1 of the statistical package JMP [3]. As expected, the NSLP is strongly influenced by the temperature at Bankstown airport. It was quickly apparent that the type of day (working or non-working, the latter including weekends and public holidays), time of day, month of the year and year also had an influence, over and above the effect of temperature. Moreover, there was an interaction between the effects of type of day and month of year, meaning that the difference between the effects of two different months could vary according to whether it was a working day or non-working day.

It was considered that the load on the previous day at a given time might also be a good predictor of load on the current day at the same time. Unfortunately, Integral Energy does not learn what its load was for the previous day until several weeks subsequently. However, the load for the whole State can be found for the previous day, so this load was used as a proxy for the Integral Energy load of the previous day. As this reading from the previous day was being tried in a model, it was decided also to include the temperature at the same time of the day previously.

Based on the above considerations, a basic model of the following form is indicated:

$$\begin{aligned}
 & \text{NSLP for IE at time } t \\
 = & \beta_0 + \beta_1 \times \text{temperature at time } t + \beta_2 \times \text{previous day's State load at time } t \\
 & + \beta_3 \times \text{year} + \beta_4 \times \text{temperature on previous day at time } t \\
 & + \sum_{i \geq 5} \text{various parameters } \beta_i \text{ multiplied by "indicator variables"} \\
 & \text{to take account of the particular type of day, month of the year,} \\
 & \text{year, and the interaction between type of day and month of year} \\
 & + \text{random error.}
 \end{aligned} \tag{1}$$

Note that it is not necessary for the statistician to construct the indicator variables in the above model. Modern statistical computer packages can do this automatically, as long as a qualitative variable is provided that specifies the value of the predictor (e.g., the month) at each observation.

The “random error” term is included in the model to account for the fact that, if observations are made at two times with exactly the same values of each of the predictor variables, the NSLPs will almost certainly be different. Random error does not mean that any mistakes have been made; the term is simply used to account for the variation that cannot be explained by any (combinations of) variables used in the model. If hypotheses about the parameters need to be tested, it is usual to assume that the random errors are independent observations from a $N(0, \sigma^2)$ population; that is, they come from a bell-shaped distribution, and no random error is inherently likely to be more variable than another. However, if no hypotheses will be tested, and the chief aim is to estimate the parameters β_0, β_1, \dots , it suffices to assume that the random errors are uncorrelated and have a constant variance, σ^2 . In this case, the parameters are estimated by the method of *Ordinary Least Squares*. This method gives unbiased estimators of the parameters even when the data are correlated, which is an important point to consider, as we subsequently show that the data are not uncorrelated.

In assessing how well a model fits the data, it is common to use the *coefficient of determination*, R^2 . This measures what proportion of the variability in the values of NSLP can be predicted by the model being used. The higher that R^2 is, the smaller on average are the squared discrepancies between the actual values of NSLP and those predicted by the model. It is important to note that a high R^2 does not necessarily mean that we have the correct model; it simply indicates that the model is good at predicting the values of NSLP. For example, if there were only a slight quadratic trend to the NSLP over a limited range of time, and a straight line was fitted to the data, there might be a very high R^2 , but this would simply indicate the good predictive ability of the model, and not its appropriateness. A referee cautions that there will be some instances where a high value of R^2 will not be associated with good predictive ability. We believe that its use in this investigation to select a model has not misled us, as evidenced by the satisfactory manner in which a model based on 2002–2005 data is then applied to 2006 data.

When the model in (1) was fitted to the 2002 to 2005 data, the resulting value of R^2 was 0.8343, which is quite reasonable for an initial model. The standard deviation, σ , of the unexplained variation is estimated by the *Root Mean Square Error* (RMSE); for this model, the RMSE was 0.1831.

Subsequent investigations found that there was a benefit from including in the model the interaction between the effects of time of day and type of day, and also the interaction between the effects of time of day and month of the year. However, with these terms in the model, there was no significant value in including the temperature at the same time on the previous day, so it was deleted from the model.

A comparison of the values of NSLP predicted by this model with the actual values of NSLP indicated that the model was quite satisfactory for moderate values of temperature, but was failing to predict the NSLP adequately when the temperature was high. The investigators then considered the possibility of allowing an increment to the predicted value of NSLP when the temperature exceeded a certain figure (a ‘breakpoint’). They also considered the possibility of including the square of temperature in the model, to provide a quadratic expression in temperature.

On the last day of the MISG meeting, Professor David Griffiths suggested that we consider

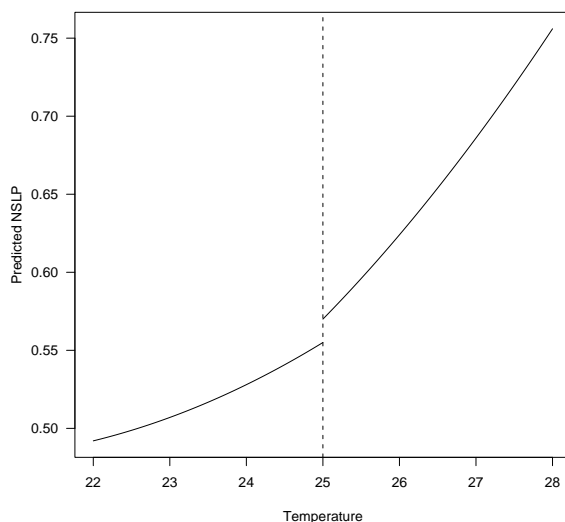


Figure 1: An illustration of a ‘jump’, and a different curve, on either side of a breakpoint at 25° .

the possibility of giving different weights to different observations when fitting the model. He subsequently also suggested that we consider the possibility of applying not just an increment when the temperature exceeded a breakpoint, but instead allowed a different model of the dependence on temperature. An illustration of this is provided in Figure 1, where there is both a ‘jump’, and a different quadratic curve, when the temperature exceeds 25° . [Please note that this Figure is for illustrative purposes only. The values on the y -axis are not indicative of any actual model.] However, not everyone agrees with this approach. A referee says ‘My gut reaction is of distaste for a curve which has a jump at a particular point.’

Breakpoints at 18° , 19° , \dots , 32° were investigated, and it was eventually determined that 20° yielded the best results. However, there is little to pick between any of these breakpoints. All resulted in R^2 values over 0.89, and the value for 20° was 0.9044, with an RMSE of 0.1391. If a differential weighting of various observations is ignored for the moment, the resulting optimal model is of the form:

$$\begin{aligned}
 & \text{NSLP for IE at time } t \\
 = & \beta_0 + \beta_1 \times \text{temperature at time } t + \beta_2 \times (\text{temperature at time } t)^2 \\
 & + I(\text{temp.} > 20)(\beta_0^* + \beta_1^* \times \text{temp. at time } t + \beta_2^* \times (\text{temp. at time } t)^2) \\
 & + \beta_3 \times \text{previous day's State load at time } t + \beta_4 \times \text{year} \\
 & + \sum_{i \geq 5} \text{various parameters } \beta_i \text{ multiplied by "indicator variables" to take} \\
 & \quad \text{account of the particular type of day, month of the year, year, interaction} \\
 & \quad \text{between type of day and month of year, interaction between type of day} \\
 & \quad \text{and time of day, interaction between month of year and time of day} \\
 & + \text{random error,} \tag{2}
 \end{aligned}$$

(a) Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	641	12719.143	19.8427	1025.172
Error	69438	1344.003	0.0194	Prob > F
C. Total	70079	14063.146		0.0000
(b) Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	641	13567.202	21.1657	1038.847
Error	69438	1414.745	0.0204	Prob > F
C. Total	70079	14981.947		0.0000
(c) Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	641	14225.576	22.1928	983.0530
Error	69438	1567.589	0.0226	Prob > F
C. Total	70079	15793.165		0.0000

Table 1: Basic Analysis of Variance tables when the same model is fitted to the 2002 to 2005 data, and (a) all observations are equally weighted, (b) those observations corresponding to temperatures greater than 28° have twice the weight of all other observations, and (c) those observations corresponding to temperatures greater than 32° have twice the weight of all other observations.

where $I(x)$ is the indicator function, taking the value 1 when x is true, and 0 otherwise.

The possibility was considered of giving more weight to observations associated with temperatures exceeding some value. Weighted regression is usually applied to data where it is thought that the variances of the observations are unequal (e.g., where the observations are the means of differing numbers of observations), with the aim of making the weighted observations have equal variance. In a situation such as the present one, where the data are regarded as having equal variances initially, the use of a weighted regression is a nonstandard attempt to make the model fit the NSLP values at higher temperatures. However, it is difficult to know what weights to use. Furthermore, the resulting R^2 values are not directly comparable. The value of R^2 is the ratio of the Model Sum of Squares (SS) to the Total SS. For unweighted regression, only the numerator changes when different models are fitted, so one can directly compare the values of R^2 for competing models. However, the effect of fitting a weighted regression is to alter both the numerator and denominator of the R^2 statistic, making a direct comparison of R^2 values less intuitive. This is illustrated in Table 1 by the basic ANOVA tables for a model with a breakpoint at 20° , for three sets of weightings: (a) equal weight for all observations, (b) double the weight for observations associated with temperatures greater than 28° , and (c) double the weight for observations associated with temperatures greater than 32° .

The maximum *residual* (discrepancy between observed and predicted values of NSLP) is minimized when the weight is doubled for observations associated with temperatures above 32° , but this does not have the greatest (weighted) R^2 or least (weighted) RMSE. This makes it difficult to find a consistent set of criteria with which to select a weighting scheme. Does one select the model that gives the least “weighted R^2 ”, or the least RMSE, or the least maximum residual? The authors’ suspicion is that the selection of the appropriate breakpoint at which

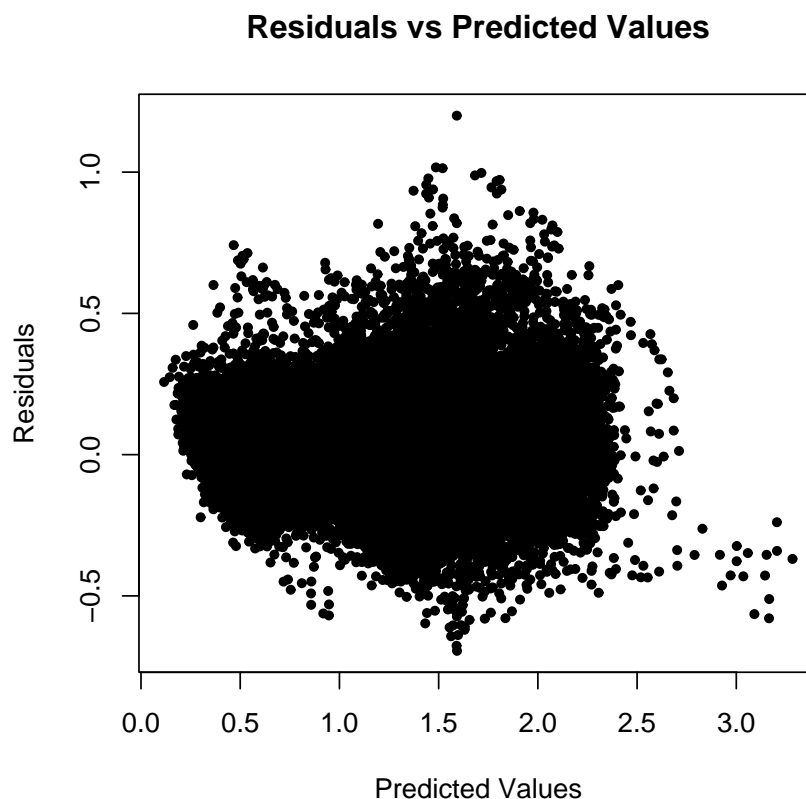


Figure 2: Residuals (on the Y-axis) vs Predicted Values for the 2002–2005 data under model (2).

to institute a greater weight is very much data-driven, and that a breakpoint appropriate for one data set (e.g., 2002 to 2005) may not be appropriate for a different data set (e.g., 2002 to 2006).

An assessment of model (2), when used with equally-weighted data, includes an examination of whether the assumptions of the analysis are met. For Ordinary Least Squares estimation, the main assumption is that the observations have equal variance. The usual way to examine this is to plot the residuals against the corresponding predicted values, and to see if there is any systematic pattern in the vertical spread of the residuals. (The *predicted values* are the values which the model predicts for each observation, and the residuals are the differences between the actual observations and the predicted values. The residuals are the best estimates of the error terms.) The lack of a systematic pattern will suggest that the variance is relatively constant.

Figure 2 displays the plot of the residuals against the predicted values for the model given in (2).

In order for tests of hypotheses about the terms in the model to be valid, the residuals should also appear to come from a single Normal distribution. However, the residuals from model (2) failed the *Shapiro-Wilk test* of Normality. A histogram of the residuals in Figure 3 indicates that there are an abnormally large number of residuals in each tail of the distribution

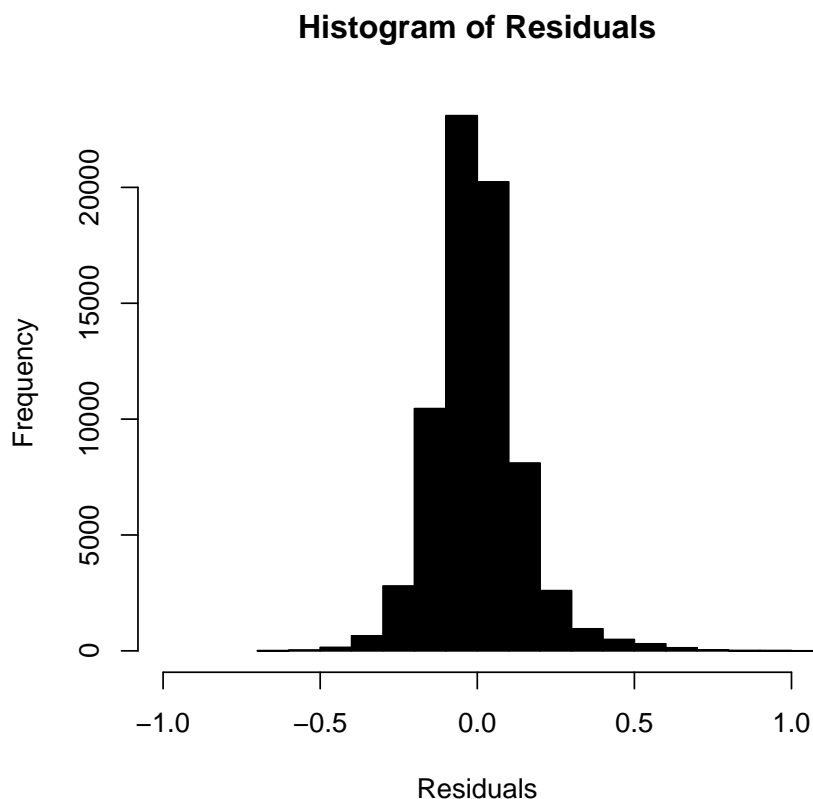


Figure 3: Histogram of the Residuals for the 2002–2005 data under model (2).

(particularly the upper tail). One must therefore view any tests of hypotheses with caution. Notwithstanding this, the Analysis of Variance for the terms in model (2) gave p -values of “ <0.0001 ” for every term in the model (counting an interaction, e.g. type of day by time of day, as one term), which suggests strongly that each term does belong in the model.

An additional assumption of the tests of hypotheses is that the error terms are uncorrelated. This assumption is examined by seeing whether the residuals are essentially uncorrelated. As the data take the form of successive observations over time (a *time series*), this is best considered by examining the *autocorrelation* between observations one, two, ... time units apart. When the temperature at the same time on the previous day was included in the model, there was negligible autocorrelation between residuals. However, when this term was deleted, the autocorrelation became pronounced. Figure 4 shows the values of the autocorrelation function, and partial autocorrelation function, of the residuals for lags of $0, 1, \dots, 48$ units of 30 minutes. The autocorrelation decreases, and then steadily increases as the lag approaches 48 (i.e., where we consider the correlation between a reading and the reading taken at the same time on the previous day), as would be expected. The partial autocorrelation, which shows the remaining autocorrelation at lag $(x + 1)$ after accounting for the autocorrelation at lags $1, \dots, x$, is small after a lag of 1. This suggests that the times series pattern in the residuals might be modelled by an autoregressive function with a lag of 1: AR(1). The existence of this autocorrelation is another reason to view any tests of hypotheses with

Supplier	R^2	RMSE
Integral Energy	0.8934	0.1666
Energy Australia	0.9171	0.1299
Country Energy	0.9379	0.0832

Table 2: Results of applying (2) to the 2006 data of each energy supplier individually.

caution.

3 Validation of the model

The terms in model (2) were then used to fit a model to the 2006 data, with all observations being equally weighted. Note that the year effect is omitted, as it is constant for all observations in the data set for 2006; any effect due to 2006 is automatically included in the initial constant in the model.

It was found that the same model as in (2) gave an R^2 value of 0.8934, and an RMSE of 0.1666. Of course, the actual values of the coefficients of the various terms will be different from those for the 2002 to 2005 data, but the same form of model was used. The plot of Residuals vs Predicted Values had the same basic shape as in Figure 2. The Residuals again failed a test of Normality, and their histogram was similar to the one in Figure 3. All terms in the Analysis of Variance table, which tests the statistical significance of these terms, had p -values of “< 0.0001”. As before, the lack of Normality means that these p -values can only be taken as indicative.

We conclude from this validation that the model in (2) will be suitable for the combined 2002 to 2006 data, and should also be suitable for any additional data, provided that no systematic change occurs in the nature of the electrical operations or weather being measured.

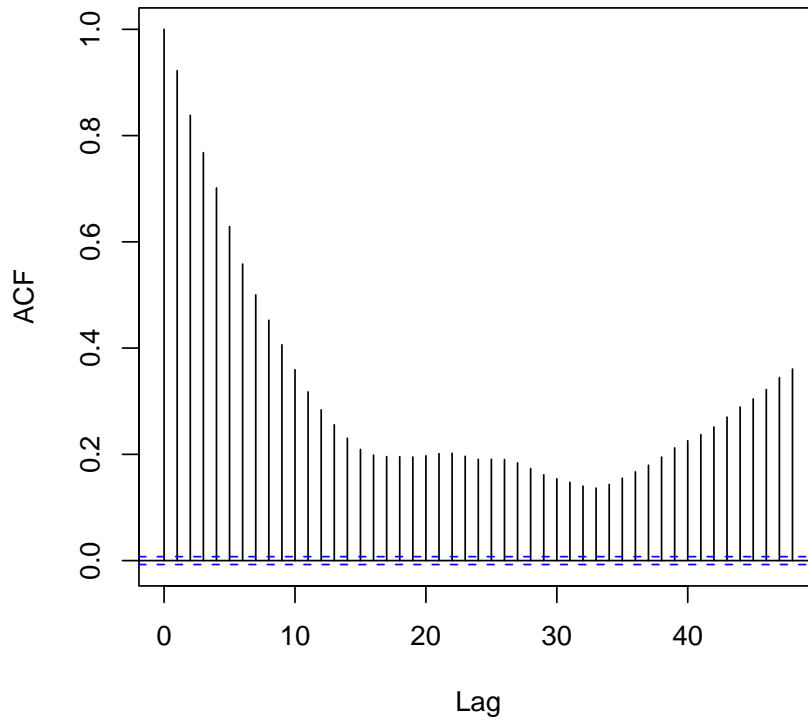
4 Comparison of relative volatilities

The same form of model (2) (i.e., the same terms were in the model, but the coefficients were calculated from the data for the particular supplier) was applied to the 2006 data of the other two suppliers. All observations were equally weighted. Table 2 gives the results for all three electricity suppliers.

Thus, although the model had been determined from Integral Energy data, it provided a better fit (as measured by R^2) for the other two suppliers, and the measure of the variability of the unexplained variation was smaller for the other two suppliers than for Integral Energy. This suggests that the relative volatility in demand is greater for Integral Energy than for the other two suppliers. However, as the NSLP data has been scaled for each supplier so as to have a mean of 1, it will be necessary for Integral Energy to rescale it in order to get a true figure for the RMSE. [Note: if the raw data are multiplied by a factor of x (say), this will also multiply the RMSE by a factor of x .]

As the temperature was measured in Bankstown, it is likely to be more applicable to the Integral Energy data than to the data for each of the other suppliers. Despite this, the NSLPs of the other two suppliers were better forecast by the common form of the model.

Series Residuals



Series Residuals

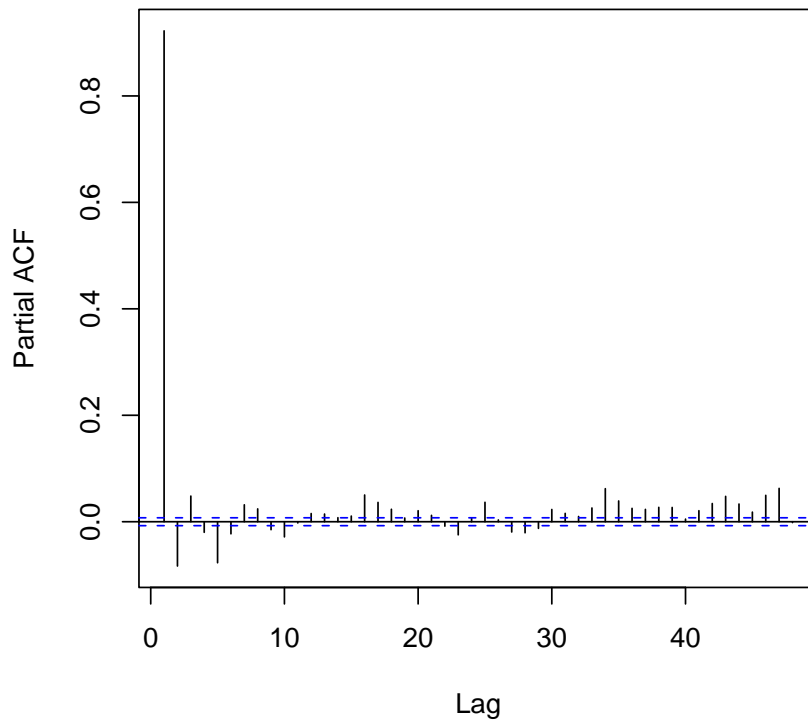


Figure 4: The autocorrelation function (higher figure), and (b) partial autocorrelation function, of the Residuals for the 2002–2005 data under model (2).

This demonstrates further that the Integral Energy data have greater volatility than the data from Energy Australia or Country Energy.

5 Discussion

The general form of the model given in (2) provides a quite satisfactory fit to the data, and we confidently recommend the use of this model to predict values of the NSLP. We have not specified the exact equation for the model, as its nature depends upon how any particular statistical software package deals with qualitative variables. For example, some software packages would treat the first month as a ‘baseline’ month, and measure the effect of month i ($i = 2, \dots, 12$) upon the NSLP by the departure of its effect from the effect of month 1; another package might treat the last month as the baseline. Moreover, as there are over 640 individual terms in the model, specifying the estimated value of each individual parameter would not be practical. A prospective user should set up a model of the form in (2), and then apply it to all the data (e.g., from 2002 to the present) that will be used to construct the model.

From consideration of the problem and the data, other predictors may prove useful to Integral Energy in modelling the NSLP load. In particular, using temperature measurements from around the franchise may provide more insight than just using the Bankstown temperature. Different observations may also be weighted (given more importance) on the basis of customer or usage density. Humidity may be a key predictor, as suggested in [1]. There may be information from the psychology literature, or from studies in physiology, to suggest at what temperature band people tend to change their behaviour, or to indicate whether there is an interaction between the effects of temperature and humidity. Also, other socioeconomic factors and demographic factors may be useful in predicting NSLP more accurately.

We have restricted the analysis to a conventional multiple regression that could be done in any of the standard commercially-available statistical software packages. However, a recent statistical innovation known as *semiparametric regression* is likely to be of use in trying to model the NSLP. An early reference on this is [2]. The R software [4] could be used to carry out any analyses that were planned. The software is free, readily downloaded, and widely used by statisticians. However, it is a package that is not easy for statistical novices to use. A referee strongly argues for the use of semiparametric regression and the R package in this problem, as they will provide a more robust analysis, not dependent on the normality and independence of the error terms.

Acknowledgment

We acknowledge the suggestions and statistical advice of Ray Chambers, Walter Davis, Kaye Marion, David Steel, and Jing You in this investigation. We especially thank a referee for very useful and helpful comments and suggestions.

References

- [1] Doulai, P. & Cahill, W. (2001) Short-term price forecasting in electric energy market. *Proceedings of the International Power Engineering Conference (IPEC2001), Singapore, May 17-19, 2001*, 749-754.

- [2] Engle, R.F., Granger, C.W.J., Rice, J. & Weiss, A. (1986) Semiparametric estimates of the relation between weather and electricity sales. *J. of the American Statistical Association* **81**, 310-319.
- [3] SAS. (2003) JMP version 5.1. SAS Institute Inc., SAS Campus Drive, Cary, NC, USA.
- [4] Development Core Team (2007) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL <http://www.R-project.org>.