NCER Working Paper Series

Forecasting day-ahead electricity load using a multiple equation time series approach

A E Clements
A S Hurn
Z Li

Working Paper #103
September 2014
Revision May 2015

Forecasting day-ahead electricity load using a multiple equation time series approach

A.E. Clements, A.S. Hurn and Z. Li
School of Economics and Finance, Queensland University of Technology.

Abstract

The quality of short-term electricity load forecasting is crucial to the operation and trading activities of market participants in an electricity market. In this paper, it is shown that a multiple equation time-series model, which is estimated by repeated application of ordinary least squares, has the potential to match or even outperform more complex nonlinear and nonparametric forecasting models. The key ingredient of the success of this simple model is the effective use of lagged information by allowing for interaction between seasonal patterns and intra-day dependencies. Although the model is built using data for the Queensland region of Australia, the methods are completely generic and applicable to any load forecasting problem. The model's forecasting ability is assessed by means of the mean absolute percentage error (MAPE). For day-ahead forecast, the MAPE returned by the model over a period of 11 years is an impressive 1.36%. The forecast accuracy of the model is compared with a number of benchmarks including three popular alternatives and one industrial standard reported by the Australia energy market operator (AEMO). The performance of the model developed in this paper is superior to all benchmarks and outperforms the AEMO forecasts by about a third in terms of the MAPE criterion.

Keywords
Short-term load forecasting, seasonality, intra-day correlation, recursive equation system.

JEL Classification Numbers
C32; Q41; Q47.

Corresponding author
Zili Li `zili.li@connect.qut.edu.au`

1 Introduction

The national electricity market (NEM) in Australia, introduced in December 1998, operates one of the worlds largest interconnected power systems which comprises five regions, namely New South Wales, Victoria, Queensland, South Australia and Tasmania. The focus of this paper is short-term pre-dispatch (up to 24 hours ahead) load forecasts for the Queensland region of the NEM, using half hourly data for the period from 12th July 1999 to 27th November 2013. The reasons for the importance of accurate short-term load forecasting differ for each of the players in the market. From the perspective of the market operator (NEM), forecasting is crucial to the scheduling and dispatch of generation capacity; for the electricity generators, the strategic choices involved in bidding and rebidding of capacity depend critically on load forecasts; and for the electricity retailers, load forecasting affects decisions about the balance between hedging and spot acquisition of electricity. For these reasons, short-term load forecasting remains a problem of central interest and one which has generated a large literature.

Statistical models for short-term load forecasting fall very naturally into three main categories. *First*, single equation time series models model the trajectory of load using traditional time series methods (Hagan and Behr, 1987; Darbellay and Slama, 2000; Taylor and McSharry, 2007). The efficacy of this approach derives from the strong seasonal patterns in electricity load. *Second*, and probably the current method of choice for practitioners, is the neural network approach in which the trajectory of load is modelled semi-parametrically using basis functions with emphasis on the non-linearity of load (Park et al., 1991; Zhang et al., 1998; Hippert et al., 2001). *Third*, multiple equation time series models have enjoyed some popularity in the

literature but their influence has waned in recent years. In this approach, each period of the day (usually each half hour or hour) is treated as a separate forecasting problem with its own equation (Peirson and Henley, 1994; Ramanathan et al., 1997; Espinoza et al., 2005; Soares and Medeiros, 2008).

The central aim of this paper is to demonstrate that the multiple equation approach has the potential to achieve a very competitive forecast accuracy. The advantages of the approach are that the explanatory factors driving forecast performance are visible, testable using traditional tests and the fact that the model specification is linear in parameters meaning that ordinary least squares can be used to estimate the parameters rather than a numerical optimisation algorithm. The seminal paper on the multiple equation approach to load forecasting is that of Ramanathan et al. (1997) in which the advantage of the multiple equation approach was first demonstrated in the context of the Californian electricity market. In the Australian electricity market, a Bayesian approach is employed by Cottet and Smith (2003) to a multiple equation model in a case study of the regional market of New South Wales. Perhaps the most insightful multiple equation model is that of Cancelo et al. (2008) who build a model of load in the Spanish electricity market.

What distinguishes the proposed model in this paper from its predecessors in the multiple equation time series tradition is the way in which the daily and weekly patterns in electricity load interact and also the recognition of the importance of intra-day correlation in load. It turns out that allowing for a distinct weekly pattern in the coefficients governing one-day lagged load is a crucial advance on previous work. The efficacy of this innovation in dealing with seasonality is demonstrated by comparing with two tradi-

tional ways of dealing with seasonality, namely the double seasonal ARIMA, the double seasonal Holt-Winters exponential smoothing approach (Gould et al., 2008). Incorporating the proposed refinements into a multiple equation model, the forecasting performance of the final chosen model is shown by comparing with the multiple equation model of Cancelo et al. (2008) and a semi-parametric approach used by AEMO.

In Section 2, a prototype model representing the starting point for the modelling exercise is developed. This model includes a piecewise linear response of load to temperature and the development of load variations for special days (public holidays). In Section 3, the prototype model is expanded to capture detailed seasonality and intra-day dependency of load. Focusing on comparing the effectiveness of modelling the seasonality of load, Section 4 compares the proposed model with two other popular alternatives. Section 5 presents the important forecasting results. The 12-hour ahead forecast accuracy of the proposed model is compared with the forecasts from the AEMO and an alternative multiple equation model of Cancelo et al. (2008). Then section 6 is a brief conclusion.

2 A Prototype Multiple Equation Model

To provide a perspective on the forecasting problem addressed in this paper, Figure 1, plots the average half-hourly load over a day and average half-hourly load over the period of a week using Queensland data with the average taken over the entire sample period from 12th July 1999 to 27th November 2013. Diurnal and weekly patterns, both well documented features of electricity load (Engle et al., 1989; Harvey and Koopman, 1993; Taylor, 2010), are clearly evident. Load picks up very quickly between the

hours of 06:00 and 08:00 from the overnight low and remains high during the daylight hours. The daily peak in the load profile usually occurs at 18:00 before tailing off once more. The weekly pattern in load is also quite pronounced with a regular load profile evident from Monday through Thursday, but with significant differences on Friday, Saturday and Sunday. While it is tempting to seek to model the trajectory of load making use of these well defined features, in fact this turns out to be a sub-optimal strategy. The averaging process involved in computing the quantities in Figure 1 smooths out much of the half-hourly variation in load and it is this variation that a good forecasting model must capture.

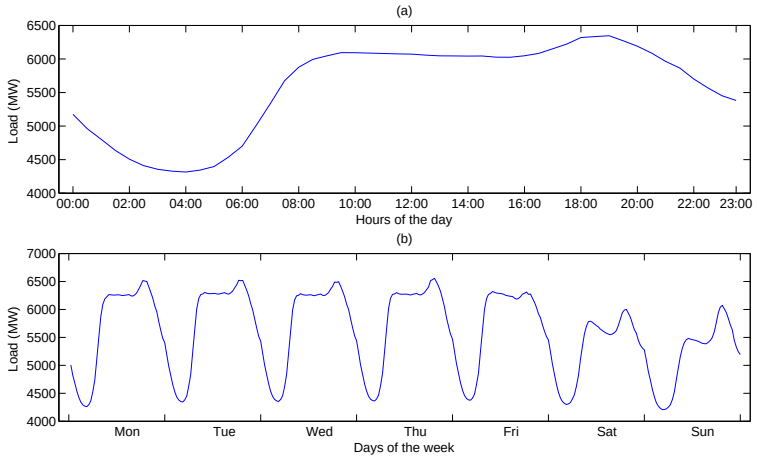

Figure 1: Averaged half-hourly load over a day and averaged half-hourly load over a week in panels (a) and (b) respectively, for Queensland over the period from 12th July 1999 to 27th November 2013.

2.1 The basic model structure

A model structure that captures half-hourly variability in load while respecting the features of the load profile in Figure 1 is one in which each half

hour is modelled separately, but also uses the diurnal and other seasonal information in the load series. Let the logarithm of the load at half hour h and day d be given by $L_{h\,d}$, then, the ARMA structure of the prototype model for a given half hour period is

$$L_{h\,d} = \theta_{h\,0} + \theta_{h\,1} L_{h\,d-1} + \theta_{h\,2} L_{h\,d-7} + \phi_{h\,1} \varepsilon_{h\,d-1} + \phi_{h\,2} \varepsilon_{h\,d-7} + \varepsilon_{h\,d}\,,$$

in which $h = 1, \cdots, 48$ and $\varepsilon_{h\,d}$ is the disturbance term. So for each half-hour, h, the parameters are estimated based on a subset of the data which only contains the observations at that interval. In this way, the partial correlation between load and lagged load are allowed to differ in a daily pattern by the different parameter values across equations. A minimal lag structure requires $L_{h\,d}$ to be explained by load in the same half hour on the previous day, $L_{h\,d-1}$ and the load in the same half hour of the same day in the previous week, $L_{h\,d-7}$. For the same reasoning, the unexpected changes in load in the same half hour on the previous day, $\varepsilon_{h\,d-1}$ and the previous week, $\varepsilon_{h\,d-7}$, are included.

It is important to factor in the effects of public holidays into the load forecasting equation, something which is accomplished quite parsimoniously using dummy variables following Cottet and Smith (2003) and Espinoza et al. (2005). To economise on the number of parameters to estimate, these special days are categorised into six distinct groups. Good Friday, Ester Monday, Christmas Day and New years are the four unique special days. The remaining two groups are a local Brisbane (the capital city of Queensland) only holiday and all the single day public holidays. Including special day

variables, the prototype model becomes:

$$\begin{aligned} L_{h\,d} =& \theta_{h\,0} + \theta_{h\,1} L_{h\,d-1} + \theta_{h\,2} L_{h\,d-7} + \phi_{h\,1}\varepsilon_{h\,d-1} + \phi_{h\,2}\varepsilon_{h\,d-7} + \varepsilon_{h\,d} \\ &+ \sum_{j=1}^{6} \left(\alpha_{j\,h\,1}\mathbb{S}_{j\,h\,d} + \alpha_{j\,h\,2}\mathbb{S}_{j\,h\,d-1}\right), \end{aligned}$$

where, $\mathbb{S}_{j\,h\,d}$ is the jth type of special day at half-hour interval h of day d. Following (Ramanathan et al., 1997), the effect of one day lagged special days, $\mathbb{S}_{j\,h\,d-1}$, is also considered. The reasoning is that when the load on special days (which is typically lower than on a normal day) is used as one day lagged load, $L_{h\,d-1}$, to infer load on normal days, the effect can be suitably adjusted. This adjustment is found to be significant and is therefore maintained. The effect of one week lagged special holidays is also investigated but discounted because the improvement was found to be insignificant.

2.2 Dealing with the effect of temperature

There is some evidence to suggest that the response of load to temperature is nonlinear in nature and the challenge is to model this nonlinear response but at the same time maintain a model specification that is linear in parameters. A piecewise linear specification following Cancelo et al. (2008) is adopted with linear responses in four different temperature ranges: 9°C - 15°C, 9°C - 20°C, 22°C - 26°C and 22°C - 30°C. Temperatures between 20°C and 22°C are regarded as comfortable and having no extra effect on load. Also the temperature beyond 9°C and 30°C are also treated as having no extra effect since the demand is ultimately limited by the capacity of temperature controlling devices, an effect termed as exhaustion. If temperature in half hour h on day d is denoted $T_{h\,d}$, then to implement the piecewise linear specification four variables must be constructed which represent the changes

in the relevant ranges of temperature. For the cooling degree temperature ranges the following two variables are defined:

$$\mathbb{C}_{1\,h\,d} = \begin{cases} 0 & T_{h\,d} \leq 22 \\ T_{h\,d} - 22 & 22 < T_{h\,d} \leq 30 \\ 30 - 22 & 30 < T_{h\,d} \end{cases}, \quad \mathbb{C}_{2\,h\,d} = \begin{cases} 0 & T_{h\,d} \leq 22 \\ T_{h\,d} - 26 & 26 < T_{h\,d} \leq 30 \\ 30 - 26 & 30 < T_{h\,d}\,. \end{cases}$$

Similarly, for the heating degree temperatures another two variables are defined:

$$\mathbb{H}_{1\,h\,d} = \begin{cases} 0 & 15 \leq T_{h\,d} \\ 15 - T_{h\,d} & 9 \leq T_{h\,d} < 15 \\ 15 - 9 & T_{h\,d} < 9 \end{cases}, \quad \mathbb{H}_{2\,h\,d} = \begin{cases} 0 & 20 \leq T_{h\,d} \\ 20 - T_{h\,d} & 9 \leq T_{h\,d} < 20\,. \\ 20 - 9 & T_{h\,d} < 9 \end{cases}$$

These variables together admit a piecewise linear response of load to temperature as illustrated in Figure 2 which is similar in spirit to the flexible spline method used by Harvey and Koopman (1993). The ranges of defined temperature variables in which they takes non-zero values are denoted by solid lines with arrows indicating the direction of the values which deviate positively from zero. Also shown is a nonparametric kernel regression of the conditional expectation of load given temperature. The nonlinear nature of the relationship is apparent, but the piecewise linear fit appears almost identical to the nonparametric regression. The advantage of the piecewise linear specification is that it accommodates the nonlinearity but does so within a model that remains linear in parameters. It should be noted that different combinations of knots for specifying temperature variable were tried in the final version of model, but discarded in favour of the current specification. Although, the temperature variables are included in the model, the actual load plot in Figure 2 suggests that load varies quite widely for any given temperature. This may be a consequence of the diverse climate in Queensland and the non-representative temperature record which is taken at only one specific location.

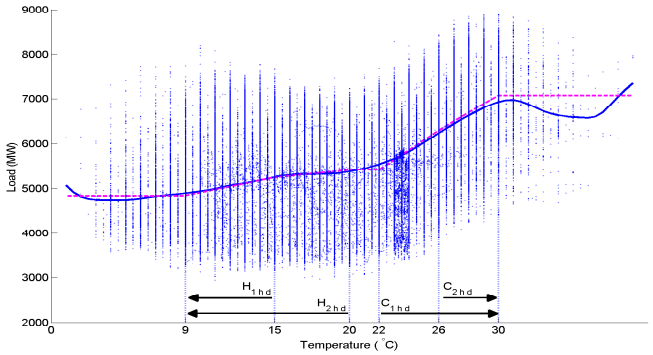

Figure 2: Queensland load and temperatures, from July 1999 to December 2013. Solid line denotes a nonparametric regression fit with normal kernel and bandwidth 1. Dashed line is the ordinary least squares fit with the four temperature variables $\mathbb{C}_{1\,h\,d}$, $\mathbb{C}_{2\,h\,d}$, $\mathbb{H}_{1\,h\,d}$ and $\mathbb{H}_{2\,h\,d}$. The ranges of defined temperature variables in which they deviate positively from zero are indicated by the arrows.

Incorporating the temperature variables into the prototype model yields

$$\begin{aligned} L_{h\,d} =& \theta_{h\,0} + \theta_{h\,1} L_{h\,d-1} + \theta_{h\,2} L_{h\,d-7} + \phi_{h\,1} \varepsilon_{h\,d-1} + \phi_{h\,2} \varepsilon_{h\,d-7} + \varepsilon_{h\,d} \\ &+ \sum_{j=1}^{6} \left(\alpha_{j\,h\,1} \mathbb{S}_{j\,h\,d} + \alpha_{j\,h\,2} \mathbb{S}_{j\,h\,d-1} \right) \\ &+ \sum_{k=1}^{2} \left(\beta_{k\,h\,1} \mathbb{H}_{k\,h\,d} + \beta_{k\,h\,2} \mathbb{C}_{k\,h\,d} + \beta_{k\,h\,3} \mathbb{H}_{k\,h\,d-1} + \beta_{k\,h\,4} \mathbb{C}_{k\,h\,d-1} \right). \quad (1) \end{aligned}$$

This is the preferred specification for the prototype model against which all the refinements in later sections will be judged.

2.3 Estimating and forecasting the prototype model

The prototype model in (1) can be estimated equation-by-equation using iterative ordinary least squares (Spliid, 1983). In the estimation, each equation is initially estimated ignoring the moving-average error terms and the regression residuals stored. The equations are then re-estimated using the regression residuals from the previous step as observed moving average error

terms. This process is then iterated until convergence which is defined as the difference in parameter values in successive iterations being less than a user supplied tolerance, in this case the square root of machine precision for floating-point arithmetic.

To assess forecast performance, a 3-year rolling window of data is used for model estimation. The day-ahead forecast is produced starting from 00:00 and uses the information available at the time of making the forecast with the exception of the temperature variables. To avoid having to provide forecasts for temperature, the actual data are used in all forecasting evaluations unless specified otherwise. Moreover, as the next-day temperature forecasts are very accurate in general, any loss in accuracy of load forecast is expected to be very small when the actual temperature is replaced with a forecast. The models are re-estimated every week. In total, a period of over 11 years from July 2002 to December 2013 is used for forecast evaluation. MAPE is used as the main criterion for assessing forecast accuracy.

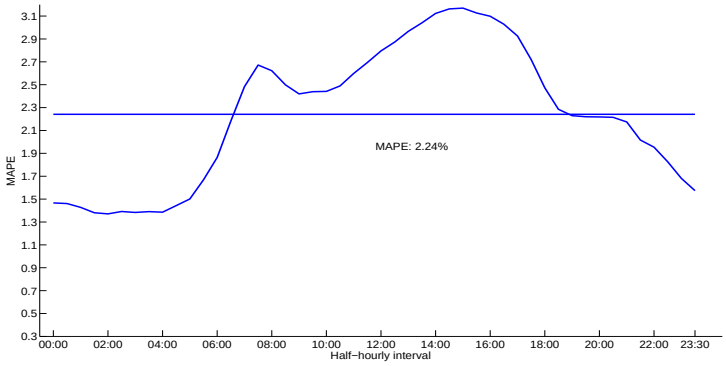

Figure 3: Half-hourly MAPEs and overall MAPE for the prototype model, equation (1). The overall MAPE is denoted as the solid horizontal line with its value indicated below.

A summary of the forecasting results for the prototype model are reported in Figure 3. The overall MAPE obtained is 2.24% with half-hourly MAPEs during the daily peak period slightly over 3%. Figure 3 also shows a clear daily pattern in half-hourly MAPE in which it reaches its lowest point during the night hours, increases to a small peak at around 08:00 and then rises continually to the daily maximum at around 16:00.

3 Extensions to the prototype model

The importance of seasonal patterns in load for accurate load forecasting is apparent and well documented in the literature (Engle et al., 1989; Harvey and Koopman, 1993; Taylor, 2010). In this section two extensions to the prototype model in (1) are proposed. The first extension addresses the important interaction between daily and weekly load patterns, and the second deals with intra-day load dependency by treating the equations as a recursive system.

3.1 Addressing Seasonality

Although the design of the lag structure in equation (1) is based on observed load profile, it does not capture completely its seasonal features. Figure 4 plots the weekly pattern in the forecast errors from the prototype model in (1), computed by averaging the half-hourly forecasting errors over a week. It is particularly evident that load in the half-hour intervals on Saturday and Sunday is significantly over-predicted (negative bias in the errors). This stems from the fact that the generally higher load on a weekday is being used as one-day lagged load in generating the forecast for weekends. Similarly, when Sunday load is used in generating the forecast for Monday, significant under-prediction occurs (positive bias in the errors). Essentially

this bias is due to the fact that the coefficients on one-day lagged load do not differentiate between days of the week. A simple way to deal with this issue is to interact the one-day lagged load with day-of-the-week dummy variables, $\mathbb{W}_{dp}$, $p = 1, \cdots, 7$. Attempts to reduce the number of dummy variables in the specification, for example by using one for weekdays and one for weekends, or defining the dummy variables in terms of whether the day before and after is a weekday or in weekend, produced inferior results.

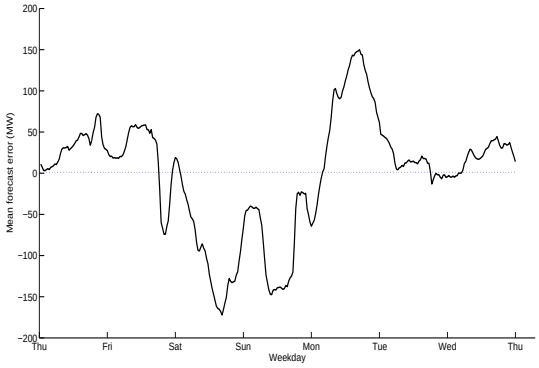

Figure 4: The mean half-hourly forecast errors in days of a week over all weeks for the prototype model, equation (1).

It is also possible that there is an annual pattern in the load, despite the sub-tropical nature of the Queensland climate. Similar to the treatment of the weekly pattern where the effect is channelled via the coefficient on $L_{h\,d-1}$, the annual pattern is specified in such a way that it enters the model via the coefficient on $L_{h\,d-7}$. Accordingly Fourier polynomials with annual cycles are interacted with the one-week lagged load, $L_{h\,d-7}$. The degree of the Fourier polynomials in the series expansion is four. While this choice is not tested formally, experimentation showed that little is to be gained by

increasing the degree of the polynomials.[1]

Incorporating the adjustments for the weekly and annual cycles gives the extended model

$$\begin{aligned} L_{h\,d} =& \theta_{h\,0} + \theta_{h\,d\,1} L_{h\,d-1} + \theta_{h\,d\,2} L_{h\,d-7} + \phi_{h\,1}\varepsilon_{h\,d-1} + \phi_{h\,2}\varepsilon_{h\,d-7} + \varepsilon_{h\,d} \\ &+ \sum_{j=1}^{6} \left(\alpha_{j\,h\,1}\mathbb{S}_{j\,h\,d} + \alpha_{j\,h\,2}\mathbb{S}_{j\,h\,d-1}\right) \\ &+ \sum_{k=1}^{2} \left(\beta_{k\,h\,1}\mathbb{H}_{k\,h\,d} + \beta_{k\,h\,2}\mathbb{C}_{k\,h\,d} + \beta_{k\,h\,3}\mathbb{H}_{k\,h\,d-1} + \beta_{k\,h\,4}\mathbb{C}_{k\,h\,d-1}\right), \quad (2) \end{aligned}$$

in which:

$$\theta_{h\,d\,1} = \sum_{p=1}^{7} \eta_{h\,p}\mathbb{W}_{d\,p}\,,$$

$$\theta_{h\,d\,2} = \tau_{h\,1} + \sum_{q=1}^{4} \left[\tau_{h\,2\,q}\sin\left(2q\pi\left(\frac{hd}{17472}\right)\right) + \tau_{h\,3\,q}\cos\left(2q\pi\left(\frac{hd}{17472}\right)\right)\right].$$

Forecasts obtained from this model are now evaluated using exactly the same procedure as outlined in Section 2.3 .

The half-hourly MAPEs are shown in Figure 5 together with the MAPEs of the prototype model. The extended model shows a significant improvement over the prototype model in every half hour period and for the overall MAPE recorded (1.56% versus 2.24%). Interestingly, it appears to be the weekly pattern rather than the annual cycle which drives this improvement. An overall MAPE of 1.61% was obtained from an alternative model with only specifying the weekly interactive dummy variables. The mean half-hourly forecast errors in days of a week obtained from the model in (2) are shown in panel (b) of Figure 5. The weekly pattern in the forecast errors has been largely eliminated.

[1] In principle, the weekly pattern previous discussed can also be modelled using Fourier polynomials. The dummy variable specification is preferred because it allows a natural interpretation of the coefficient estimates.

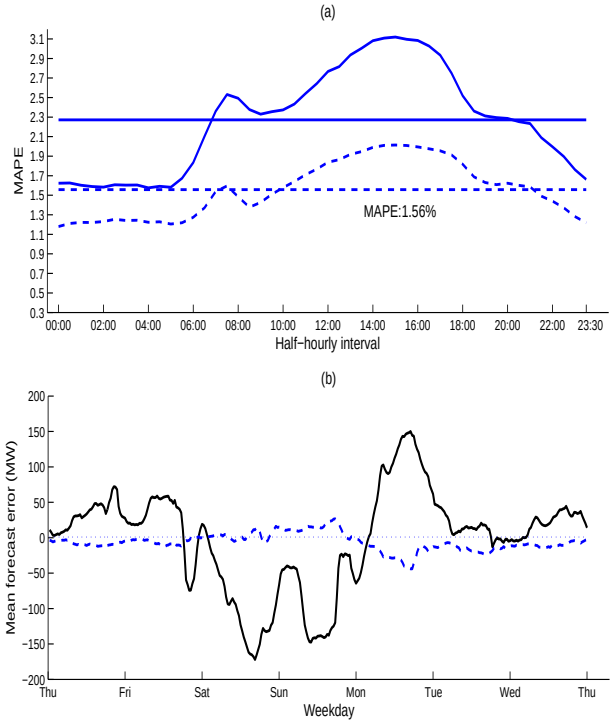

Figure 5: In panel (a), the half-hourly MAPEs and overall MAPE for the prototype model (solid lines) in (1) are compared to the model with seasonal patterns (dashed lines) in the parameters given in (2). The overall MAPEs are shown as horizontal lines with the value for equation (2) indicated below. In panel (b), the mean half-hourly forecast errors in days of a week over all weeks from the prototype model (equation (1), solid line) and the model with seasonal patterns in the parameters (equation (2), dashed line).

A more detailed breakdown of the forecast performance is provided in Table 1. It is apparent that the most significant improvements achieved using the extended model in (2) are found in the forecasts on normal days and weekends. The total number of large forecast errors defined as an absolute percentage error (APE) greater than 5% is reduced by more than 10,000 instances (a 68% improvement). Overall, by interacting $L_{h\,d-1}$ and $L_{h\,d-7}$ with the weekly dummy variables and annual cycles, respectively, the overall

Table 1: The forecast comparison between the prototype model in (1) and the model with seasonal patterns in the parameters (equation (2)).

		Overall MAPE	Maximum APE	No. APE $\geq 5\%$	No. APE $\geq 10\%$	No. APE $\geq 15\%$	No. APE $\geq 25\%$	Obs.
Overall	Eq. (1)	2.24%	33.58%	19702	2630	430	33	199584
	Eq. (2)	1.56%	26.86%	6303	542	107	4	
Normal days	Eq. (1)	2.12%	33.58%	13629	1940	322	16	137232
	Eq. (2)	1.50%	24.68%	3671	266	33	0	
Weekend	Eq. (1)	2.48%	24.75%	4912	517	61	0	57024
	Eq. (2)	1.62%	21.83%	2056	156	15	0	
Special days	Eq. (1)	2.89%	32.31%	1390	243	70	17	6384
	Eq. (2)	2.54%	26.86%	798	186	70	4	

MAPE of the forecast improves by 0.68% in comparison with the prototype model in Section 2.

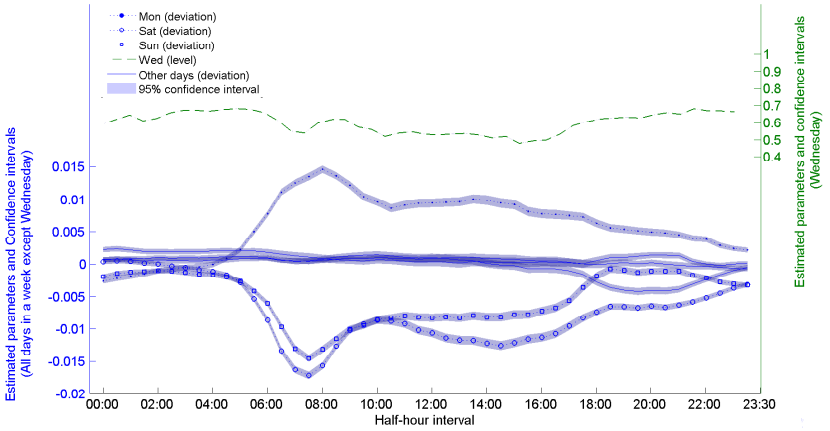

Figure 6: Estimated parameters and 95% confidence intervals (shaded areas) for weekly dummy variables in the model with seasonal patterns in the parameter (equation (2)). On the left vertical axis, the deflections of the parameter estimates of Monday, Saturday, Sunday and other weekdays from Wednesday are denoted by dotted line with dots, dotted line with circles, dotted line with squares and solid lines respectively. On the right axis, the level of parameter estimates for Wednesday are plotted in dashed line.

A set of representative parameter estimates for the interactive dummy

variables $\mathbb{W}_{dp}$, $p = 1, \ldots, 7$ and their 95% confidence intervals from a 3-year rolling window estimation are plotted in Figure 6. The largest coefficient values are seen to occur on Monday because the weekend load being used as one day lagged load in forecasting weekday load is substantially lower that the observed Monday load. The smallest coefficient values are found on the weekends. This is exactly the opposite effect to that noted for Monday; higher weekday loads are now being used to generate forecasts of lower weekend loads. More interestingly, the values of the coefficients vary in different half hours of a day. Another discernible pattern is to be found in the coefficients for different weekdays. In off-peak half-hourly intervals, the coefficients have a very similar magnitude with, in some instances, overlapping confidence intervals across different weekdays. During peak load half-hourly intervals, however, the values of the coefficients are substantially different across different weekdays. This indicates clearly that there is an interaction between daily and weekly patterns in load, a characteristic which tends to be ignored in the load forecasting literature.

3.2 Intra-day Correlations

In the models studied thus far, the information set is defined at a daily resolution at day $d-1$. One important piece of information which is ignored is the observed load in last half-hour period of the day prior to making a forecast, $L_{48\,d-1}$. This is particularly important for the first half hour period of the forecast, as this lagged load is observed in the immediately preceding

half hour. Making this adjustment yields the model

$$\begin{aligned} L_{h\,d} =& \theta_{h\,0} + \theta_{h\,d\,1} L_{h\,d-1} + \theta_{h\,d\,2} L_{h\,d-7} + \theta_{h\,4} L_{48\,d-1} \\ &+ \phi_{h\,1}\varepsilon_{h\,d-1} + \phi_{h\,2}\varepsilon_{h\,d-7} + \varepsilon_{h\,d} \\ &+ \sum_{j=1}^{6} \left(\alpha_{j\,h\,1}\mathbb{S}_{j\,h\,d} + \alpha_{j\,h\,2}\mathbb{S}_{j\,h\,d-1}\right) \\ &+ \sum_{k=1}^{2} \left(\beta_{k\,h\,1}\mathbb{H}_{k\,h\,d} + \beta_{k\,h\,2}\mathbb{C}_{k\,h\,d} + \beta_{k\,h\,3}\mathbb{H}_{k\,h\,d-1} + \beta_{k\,h\,4}\mathbb{C}_{k\,h\,d-1}\right), \quad (3) \end{aligned}$$

in which

$$\begin{aligned} \theta_{h\,d\,1} =& \sum_{p=1}^{7} \eta_{h\,p}\mathbb{W}_{d\,p}\,, \\ \theta_{h\,d\,2} =& \tau_{h\,1} + \sum_{q=1}^{4} \left[\tau_{h\,2\,q}\sin\left(2q\pi\Big(\frac{hd}{17472}\Big)\right) + \tau_{h,3,q}\cos\left(2q\pi\Big(\frac{hd}{17472}\Big)\right)\right]. \end{aligned}$$

Figure 7 compares the MAPE of the model in (3) with the model (2) in Section 3.1. Not surprisingly, the biggest improvement is found in the first half-hour interval. Moreover, the substantial improvements in the half-hourly MAPEs in the first 20 half-hour intervals indicates that this idea is well worth pursuing a little further. Indeed, it is reasonable to posit that the load in consecutive half hours will be correlated so that in addition to observed load in last half-hour period of the day prior to the making a forecast, $L_{48\,d-1}$, each equation contains the lagged load from the immediately preceding half hour, $L_{h-1\,d}$. Additional lags of consecutive half-hour periods were tried but the improvement in forecast performance was minimal.

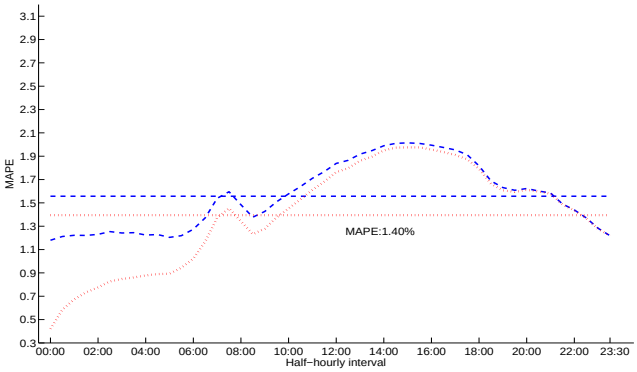

Figure 7: The half-hourly MAPEs and overall MAPE for the mode with seasonal pattern in the parameters (equation (2), dashed lines) and the model with the most recent load information in (3) (dotted lines). The overall MAPEs are shown as horizontal lines with the value of which for equation (3) indicated below.

Consequently, the preferred multiple equation time series model is now

$$\begin{aligned} L_{h\,d} =& \theta_{h\,0} + \theta_{h\,d\,1} L_{h\,d-1} + \theta_{h\,d\,2} L_{h\,d-7} + \theta_{h\,4} L_{48\,d-1} + \theta_{h\,5} I_{h>1} L_{h-1\,d} \\ &+ \phi_{h\,1} \varepsilon_{h\,d-1} + \phi_{h\,2} \varepsilon_{h\,d-7} + \varepsilon_{h\,d} \\ &+ \sum_{j=1}^{6} \left(\alpha_{j\,h\,1} \mathbb{S}_{j\,h\,d} + \alpha_{j\,h\,2} \mathbb{S}_{j\,h\,d-1} \right) \\ &+ \sum_{k=1}^{2} \left(\beta_{k\,h\,1} \mathbb{H}_{k\,h\,d} + \beta_{k\,h\,2} \mathbb{C}_{k\,h\,d} + \beta_{k\,h\,3} \mathbb{H}_{k\,h\,d-1} + \beta_{k\,h\,4} \mathbb{C}_{k\,h\,d-1} \right), \quad (4) \end{aligned}$$

in which

$$\theta_{h\,d\,1} = \sum_{p=1}^{7} \eta_{h\,p} \mathbb{W}_{d\,p}\,,$$

$$\theta_{h\,d\,2} = \tau_{h\,1} + \sum_{q=1}^{4} \left[\tau_{h\,2\,q} \sin\left(2q\pi \left(\frac{hd}{17472} \right) \right) + \tau_{h\,3\,q} \cos\left(2q\pi \left(\frac{hd}{17472} \right) \right) \right],$$

and $I_{h>1}$ denotes an indicator function which is equal to 1 when $h > 1$ and 0 otherwise. This modification turns the 48 equations for the half hours of a day into a recursive system. Once again, repeated application of ordinary

least squares can be used to estimate the system, it provides a parsimonious way of capturing the intra-day load correlation without increasing computational complexity significantly. Experimentation indicates that the more efficient estimation method with taking into account of intra-day error correlation does not generally improve forecast accuracy.

The forecast results using (4) are plotted in Figure 8. Overall, the results show that half-hourly day-ahead MAPEs are all below 2%, with an overall MAPE of 1.36%. The improvement in forecast accuracy from using the recursive system is mainly for the daily peak intervals between 14:00 and 18:00.

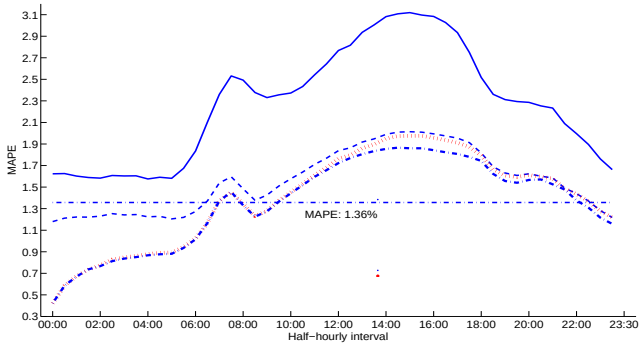

Figure 8: Forecast comparison on the half-hourly MAPEs for all the four models (solid line for equation (1), dashed line for equation (2), dotted line for equation (3) and dash-dot line for equation (4)) studied and the overall MAPE for the model using recursive system in (4) (dash-dot horizontal line with its value indicated below).

More detailed results are reported in Table 2, in which models from (1) to (4) are compared. It can be seen that the most significant improvement is obtained due to the introduction of the weekly dummy variables interacting with the lagged load, $L_{h\,d-1}$, in equation (2). In particular, the number of

instances of large errors (APE $\geq$ 5%) decreases by nearly 10,000 on normal days when moving from the specification in the prototype model (1) to the weekly dummy variable specification in (2). In addition, incorporating the most recent information, equation (3), and using a recursive system for intra-day correlation, equation (4), also improve accuracy but the size of the improvement is not as large. Overall, comparing the final model in (4) with the prototype model in (1), the reduction in the number of large APE is over 70% in all bands, and overall MAPE drops from 2.24 % to 1.36%, results which vindicate the modifications proposed in this section.

Table 2: The forecasting accuracy for the models studied, from equation (1) to (4).

		Overall MAPE	Maximum APE	No. APE $\geq$ 5%	No. APE $\geq$ 10%	No. APE $\geq$ 15%	No. APE $\geq$ 25%	Obs.
Overall	Eq. (1)	2.24%	33.58%	19702	2630	430	33	199584
	Eq. (2)	1.56%	26.86%	6303	542	107	4	
	Eq. (3)	1.40%	25.98%	5130	467	93	4	
	Eq. (4)	1.36%	25.70%	4499	451	95	1	
Normal days	Eq. (1)	2.12%	33.58%	13629	1940	322	16	137232
	Eq. (2)	1.50%	24.68%	3671	266	33	0	
	Eq. (3)	1.35%	24.00%	2982	216	30	0	
	Eq. (4)	1.31%	21.62%	2544	203	31	0	
Weekend	Eq. (1)	2.48%	24.75%	4912	517	61	0	57024
	Eq. (2)	1.62%	21.83%	2056	156	15	0	
	Eq. (3)	1.44%	21.55%	1663	143	10	0	
	Eq. (4)	1.41%	21.77%	1521	142	13	0	
Special days	Eq. (1)	2.89%	32.31%	1390	243	70	17	6384
	Eq. (2)	2.54%	26.86%	798	186	70	4	
	Eq. (3)	2.28%	25.98%	697	168	59	4	
	Eq. (4)	2.25%	25.70%	641	167	58	1	

4 A Comparison of Approaches to Modelling Seasonality

The extensions proposed in Section 3 are designed to effectively model the detailed seasonality in electricity load. In this section, the extended model

in equation (4) is compared with two popular methods commonly used in the literature for dealing with seasonality. In order to focus the comparison on modelling seasonality alone, the models in this section use only lagged load information and all other information, such as temperature and special days, are ignored. The two models used for comparative purposes are now outlined.

Single equation double seasonal ARIMA model

ARIMA type models for load forecasting are widely used in the literature (Taylor, 2012; Kim, 2013). The single equation double seasonal ARIMA model is specified as:

$$\phi_p(B)\Phi_{P_1}(B^{S_1})\Phi_{P_2}(B^{S_2})(1-B)^d(1-B^{S_1})^{D_1}(1-B^{S_2})^{D_2}(L_t - c - bt) = \theta_q(B)\Theta_{Q_1}(B^{S_1})\Theta_{Q_2}(B^{S_2})\varepsilon_t, \tag{5}$$

where, B is the back shift operator. $\phi_p(B)$, $\Phi_{P_2}(B^{S_1})$, $\Phi_{P_2}(B^{S_2})$ and $\theta_q(B)$, $\Theta_{Q_2}(B^{S_1})$, $\Theta_{Q_2}(B^{S_2})$ denote the autoregressive and moving average parts respectively, with back shift polynomials of degree p, P_1, P_2 and q, Q_1, Q_2 respectively and seasonal factors S_1 and S_2. D_1, D_2 are the orders of differencing. The parameter c is the constant term and b is the parameter for the time trend t. The model can be written as

$$ARIMA(p,q,d) \times (P_1,Q_1,D_1)_{S_1} \times (P_2,Q_2,D_2)_{S_2}\,.$$

Focusing on comparing the effectiveness of the models for modelling the seasonality and to make the model in (5) comparable in a sense that it uses approximately the same amount of information as used by the multiple equation model, the specification

$$ARIMA(1,1,0) \times (1,1,1)_{48} \times (1,1,1)_{336}\,,$$

is chosen. Depending on specific case, both the proposed multiple equation model and the single equation double seasonal ARIMA in (5) can be easily expanded to accommodate more distant lags and other explanatory variables.

Double seasonal Holt-Winters exponential smoothing model

In short-term load forecasting, the seasonal Holt-Winters exponential smoothing (HWES) is another common choice for modelling seasonality in load (Gould et al., 2008; De Livera et al., 2011; Taylor, 2012). An intra-day cycles double seasonal HWES approach of (Gould et al., 2008) is implemented here, which includes an unconstrained seasonal updating scheme with 7 daily sub-cycles in a week and additive seasonal components. As suggested by Taylor (2012), an AR(1) term for the residual is included for better forecast accuracy. The model is specified as:

$$\begin{aligned} L_t &= l_{t-1} + b_{t-1} + \mathbf{x}_t' \mathbf{s}_{t-48} + \phi r_{t-1} + \varepsilon_t, \\ r_t &= L_t - l_{t-1} - b_{t-1} - \mathbf{x}_t' \mathbf{s}_{t-48}, \\ l_t &= l_{t-1} + b_{t-1} + \alpha r_t, \\ b_t &= b_{t-1} + \beta r_t, \\ \mathbf{s}_t &= \mathbf{s}_{t-48} + \Gamma \mathbf{x}_t r_t, \end{aligned} \tag{6}$$

where l_t and b_t are the level and trend at time t, respectively. The variable $\mathbf{x}_t$ is a 7×1 vector of day-of-the-week dummy variables, $\mathbf{s}_t$ is a 7×1 vector of seasonal components for the same half-hour intervals for the 7 days in a week, r_t and ε_t are, respectively, the residual term and an independent and identically distributed error term with zero mean. The constants α, β are smoothing parameters for the level and the trend, respectively, and ϕ is the AR(1) parameter for the residual. The matrix Γ has dimension 7×7 and

contains the smoothing parameters for the seasonal components.

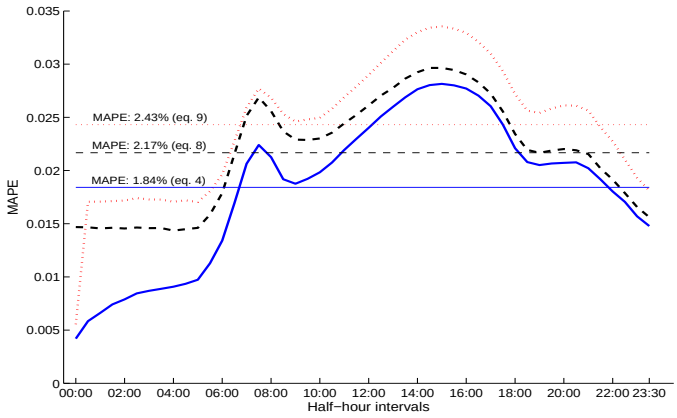

Figure 9: The half-hourly MAPEs of the one day ahead forecast produced by the proposed method (equation (4) without temperature and special days, denoted by solid lines), the single equation double seasonal ARIMA (equation (5), denoted by dashed lines), and the unconstrained intra-day cycles double seasonal HWES (equation (6), denoted by dotted lines) from July 2002 to December 2013. The overall MAPEs are shown as the horizontal lines with the values indicated above.

Figure 9 plots the half-hourly MAPEs of the one-day-ahead forecasts produced by the three models for the period from July 2002 to December 2013. The efficacy of the proposed multiple equation model for modelling seasonality in the load is obvious. The half-hourly MAPEs and overall MAPE for this approach are clearly lower than the corresponding forecasting statistics produced by the two competitor approaches. In short, the proposed methodology is flexible in accommodating not only daily and weekly patterns of load, but also the interaction between the two in a way that leads to a significantly improved accuracy in forecast performance as shown in Figure 9. In the double seasonal ARIMA, neither daily nor weekly patterns are allowed in the parameter for lagged load. In the double seasonal HWES,

the unconstrained seasonal component smoothing parameters, Γ allow the seasonal component for a half-hour interval in a day of a week to be updated based on the observed load at the same half-hour interval in other days of a week, but the intra-day smoothing parameter is assumed to be fixed.

5 Assessing Forecast Performance of the Full Model

In this section, the forecast performance of the preferred model in (4) is compared against the industry standard reported by the market operator AEMO. AEMO as the operator of the NEM, provides short-term load forecasts in pre-dispatch IS reports for the next trading day.[2] Among the horizons of the load forecast, 12-hour ahead forecasts provide important information for dispatch planning for the next day. To monitor 12-hour ahead load forecast accuracy, the monthly averaged MAPE of the 12-hour ahead forecasts is reported by AEMO as a benchmark for assessing the forecasting performance.[3] Although the details of the specification of the AEMO forecasting procedure are not available, it is known to be based on the semi-parametric specification of Fan and Hyndman (2012) and as the main forecasting model chosen by the market operator, may be taken to be representative of the state of art performance of load forecasting models.[4]

The model is also compared with the multiple equation model proposed by Cancelo et al. (2008), hereafter CEG. In this model, the seasonality of load is dealt with using a seasonal ARIMA process, which results in a non-linear model specification requiring estimation by maximum likelihood.

[2] See, `http://www.nemweb.com.au/REPORTS/CURRENT/PreDispatchIS_Reports/`.

[3] See,`http://www.aemo.com.au/Electricity/Data/PreDispatch-Demand-Forecasting-Performance`

[4] See, `http://www.aemo.com.au/Electricity/Planning/Forecasting/National-Electricity-Forecasting-Report-2012`

Forecasting of the models is implemented using an identical procedure and the same set of variables defined in Section 2.3. To align with the 12-hour ahead forecast accuracy reported by AEMO, the accuracy of the proposed model (4) and CEG are assessed using 12-hour ahead forecasts.

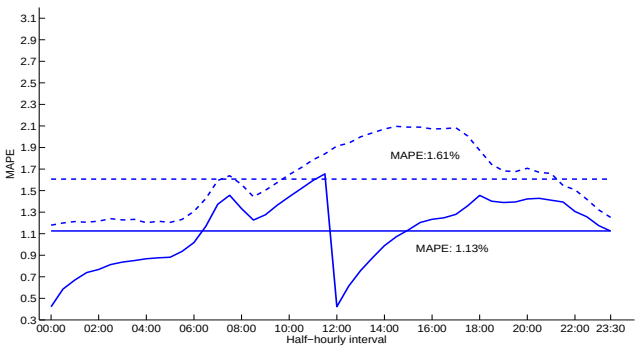

Figure 10: The half-hourly MAPEs of 12-hour ahead forecast by equation (4) (solid lines) and CEG (dashed lines) from July 2002 to December 2013. The overall MAPEs are shown as the horizontal lines.

A first comparison involves only the preferred model, (4), and CEG given that the AEMO forecast errors are only available for a shorter period. Forecasts of the two multiple equation models are generated using the same procedure as in Section 2.3 and the results are illustrated in Figure 10. It can been seen that the forecast accuracy of proposed model (4) is superior to that of CEG. An important anomaly in the CEG model is that it only utilizes information available 24 hours previously in making a forecast. This is clearly a flaw because it does not allow the model to be flexible in terms of forecasting for periods less than 24 hours. Even in the first 12 hours when forecasts from the two models are based on the same available information, the lower MAPEs obtained from proposed model (4) shows the advantages

of using the latest observed load together with the recursive structure developed in Section 3.2. Note that in the case of 12-hour ahead forecasts, the variable $L_{48\,d-1}$ in (4) is replaced with $\mathbb{L}_{h\,d} = I_{h\leq 24}L_{48\,d-1} + I_{h>24}L_{24\,d}$. This is responsible for the marked decrease in half-hourly MAPEs shown in Figure 10 starting from 12:00 when the most recent load information is updated. A more detailed comparison of the performance of the two models is shown in columns 2 and 3 of Table 3 where CEG produces inferior forecasts under all criteria. Since CEG only utilize information at a daily resolution, the results shown in column 2 for the CEG forecasts over the whole period can also be compared with the 24-hour ahead forecast from model (4) shown in row 5 of Table 2. The 1.36% overall MAPE of proposed model (4) is 0.25% lower than the one obtained from CEG (1.61%) and similar superior performance of the former is observed in all the criteria.

Table 3: Summary comparison of 12-hours ahead forecast by equation (4), CEG and the AEMO forecasts.

	Jul 2002 - Dec 2013		Jul 2012 - Nov 2013			
	CEG	Eq. (4)	CEG	Eq. (4)	Eq. (4) without temperature	AEMO forecasts
Overall MAPE	1.61%	1.13%	1.67%	1.21%	1.37%	1.88%
Max. APE	27.99%	25.68%	20.89%	20.21%	20.26%	-
No. APE $\geq$ 5%	6981	2009	1092	384	585	-
No. APE $\geq$ 10%	645	205	128	38	44	-
No. APE $\geq$ 15%	120	45	27	7	7	-
No. APE $\geq$ 25%	3	3	0	0	0	-
Max. monthly MAPE	-	-	2.99%	1.84%	2.02%	3.2%
Obs.	199584	199584	24864	24864	24864	-

Given the limited historical data publicly available from AEMO, the period from July 2012 to November 2013 is used for subsequent comparison. Although this period is only 17 months, the advantage of the proposed model is shown clearly in Figure 11 and columns 4 to 7 of Table 3, with the

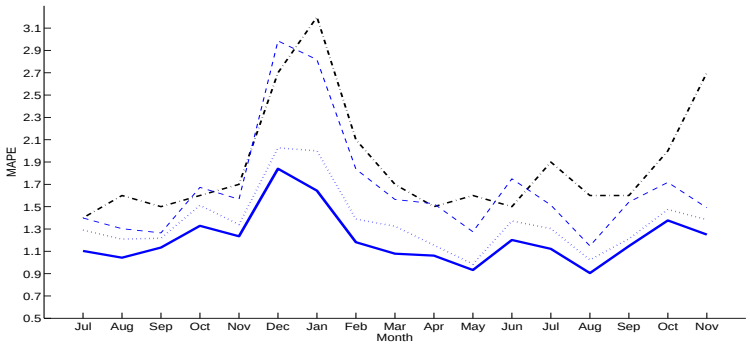

Figure 11: 12-hours ahead forecasts comparison of monthly MAPE between equation (4) (solid line), equation (4) without the future temperature (dotted line), CEG (dashed line) and the AEMO forecast (dot-dash line), from July 2012 to November 2013.

monthly MAPEs well below the AEMO forecasts and an improvement of around 0.67% in the overall MAPE over the AMEO forecasts. Since AEMO forecasts are based on temperature forecasts instead of real temperature, the results from the proposed model obtained by omitting the variables for current temperature are also reported. While there is a fall in accuracy relative to the situation when actual temperature is used, Figure 11 demonstrates that this effect is very small and the model is still more accurate than the AEMO forecast under all criteria (0.51% lower in the overall MAPE). The advantage of model (4) over CEG (which uses actual temperature data in the forecast) is also shown in Figure 11 and columns 4 to 6 of Table 3, where either with or without actual temperature, the preferred model is seen to outperform CEG under all criteria.

6 Conclusion

The problem of forecasting load is an important one for all electricity market participants because it informs their strategic decisions about dispatch (market operators), bidding and rebidding (generators) and trading activity (retailers). In recent times a consensus seems to have developed that neural network or non-parametric based forecasts of load, with their inherently nonlinear structure, offer the best alternative for accurate forecasting. This paper has demonstrated that a traditional time-series approach, in which an equation is specified for each half hour of the day, provides a viable alternative method which produces very competitive results if implemented carefully.

The multiple equation load forecasting model in this paper pays particular attention to the interaction between daily and weekly load patterns. Probably the most important distinguishing factor in the proposed model relative to others in the literature is the flexibility built into the influence of load from the same half hour on the previous day. Allowing the strong weekly pattern to interact with the daily pattern in coefficients on lagged load yields important improvements in short-term forecast performance. Another innovative dimension of the current model is the use of the inherent recursive structure of the model to capture the intra-day load correlation. The effectiveness of the proposed approach on modelling the seasonal features of electricity load is demonstrated by comparing with two popular alternatives, double seasonal ARIMA and Holt-Winters exponential smoothing. Despite these modifications to the preferred model, it remains linear in parameters and can be estimated equation-by-equation by ordinary least squares.

Overall, the forecasting performance of the preferred model is impressive

and significantly out-performs two benchmarks with which it is compared. In particular, the model improves on the mean average percentage error of 12-hour ahead forecast reported by the Australian energy market operator by about a third. For the entire 11 year period, the model returns a mean average percentage error of 1.36% on half-hourly day-ahead forecasts, a figure is lower than most (if not all) comparable average error statistics reported in the literature. Of course, the simple computation of an error metric does not really encapsulate the economic advantage to market participants of providing accurate load forecasts. The challenge for future work is to devise a metric that is capable of measuring economic gains to more accurate load forecasting.

References

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**(4), 588–602.

Cottet, R., and Smith, M. 2003. Bayesian Modeling and Forecasting of Intraday Electricity Load. *Journal of the American Statistical Association*, **98**(464), 839–849.

Darbellay, G. A., and Slama, M. 2000. Forecasting the short-term demand for electricity: Do neural networks stand a better chance? *International Journal of Forecasting*, **16**(1), 71–83.

De Livera, A. M., Hyndman, R. J., and Snyder, R. D. 2011. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, **106**(496), 1513–1527.

Engle, R. F., Granger, C. W. J., and Hallman, J. J. 1989. Merging short-and long-run forecasts: An application of seasonal cointegration to monthly electricity sales forecasting. *Journal of Econometrics*, **40**(1), 45–62.

Espinoza, M., Joye, C., Belmans, R., and De Moor, B. 2005. Short-Term Load Forecasting, Profile Identification, and Customer Segmentation: A Methodology Based on Periodic Time Series. *IEEE Transactions on Power Systems*, **20**(3), 1622–1630.

Fan, S., and Hyndman, R.J. 2012. Short-Term Load Forecasting Based on a Semi-Parametric Additive Model. *IEEE Transactions on Power Systems*, **27**(1), 134–141.

Gould, Phillip G., Koehler, Anne B., Ord, J. Keith, Snyder, Ralph D., Hyndman, Rob J., and Vahid-Araghi, Farshid. 2008. Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, **191**(1), 207–222.

Hagan, M. T., and Behr, S. M. 1987. The Time Series Approach to Short Term Load Forecasting. *IEEE Transactions on Power Systems*, **2**(3), 785–791.

Harvey, A., and Koopman, S. J. 1993. Forecasting Hourly Electricity Demand Using Time-Varying Splines. *Journal of the American Statistical Association*, **88**(424), 1228–1236.

Hippert, H. S., Pedreira, C. E., and Souza, R. C. 2001. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, **16**(1), 44–55.

Kim, M. S. 2013. Modeling special-day effects for forecasting intraday electricity demand. *European Journal of Operational Research*, **230**(1), 170–180.

Park, D. C., El-Sharkawi, M. A., Marks, R. J., Atlas, L. E., and Damborg, M. J. 1991. Electric load forecasting using an artificial neural network. *IEEE Transactions on Power Systems*, **6**(2), 442–449.

Peirson, J., and Henley, A. 1994. Electricity load and temperature: Issues in dynamic specification. *Energy Economics*, **16**(4), 235–243.

Ramanathan, R., Engle, R., Granger, C. W. J., Vahid-Araghi, F., and Brace, C. 1997. Shorte-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, **13**(2), 161–174.

Soares, L. J., and Medeiros, M. C. 2008. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. *International Journal of Forecasting*, **24**(4), 630–644.

Spliid, H. 1983. A Fast Estimation Method for the Vector Autoregressive Moving Average Model With Exogenous Variables. *Journal of the American Statistical Association*, **78**(384), 843–849.

Taylor, J. W. 2010. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, **204**(1), 139–152.

Taylor, J. W., and McSharry, P. E. 2007. Short-Term Load Forecasting Methods: An Evaluation Based on European Data. *IEEE Transactions on Power Systems*, **22**(4), 2213–2219.

Taylor, J.W. 2012. Short-Term Load Forecasting With Exponentially Weighted Methods. *IEEE Transactions on Power Systems*, **27**(1), 458–464.

Zhang, G., Eddy Patuwo, B., and Hu, M. Y. 1998. Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, **14**(1), 35–62.