European Journal of Operational Research 000 (2016) 1–9

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

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

A. E. Clements, A. S. Hurn, Z. Li*

School of Economics and Finance, Queensland University of Technology, Brisbane, QLD 4000, Australia

ARTICLE INFO

Article history:
Received 20 March 2015
Accepted 16 December 2015
Available online xxx

Keywords:
Short-term load forecasting
Modelling seasonality
Intra-day load correlation

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 method is 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.

1. Introduction

The national electricity market (NEM) in Australia, introduced in December 1998, operates one of the worlds largest interconnected power systems which comprise 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 12 July 1999 to 27 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 (Amaral, Souza, & Stevenson, 2008; Darbellay & Slama, 2000; Hagan & Behr, 1987; Hahn, Meyer-Nieberg, & Pickl, 2009; Pedregal & Young, 2008; Taylor, 2003). The efficacy of this approach derives from the strong seasonal patterns in electricity load. *Second*, semi-, or non-parametric methods which emphasise the non-linearity of load (Amjady, 2006, 2007; Fan & Hyndman, 2012; Hippert, Pedreira, & Souza, 2001; Park, El-Sharkawi, Marks, Atlas, & Damborg, 1991; Zhang, Eddy Patuwo, & Hu, 1998). *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 (Espinoza, Joye, Belmans, & De Moor, 2005; Peirson & Henley, 1994; Ramanathan, Engle, Granger, Vahid-Araghi, & Brace, 1997; Soares & 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 major advantage of the proposed approach is that the model remains linear in parameters, so that ordinary least squares can be used to estimate the parameters of the model and traditional test may be used to assess the significance of the explanatory variables. 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

* Corresponding author. Tel.: +61 7 3138 6673.
E-mail address: zili.li@connect.qut.edu.au (Z. Li).

http://dx.doi.org/10.1016/j.ejor.2015.12.030
0377-2217/

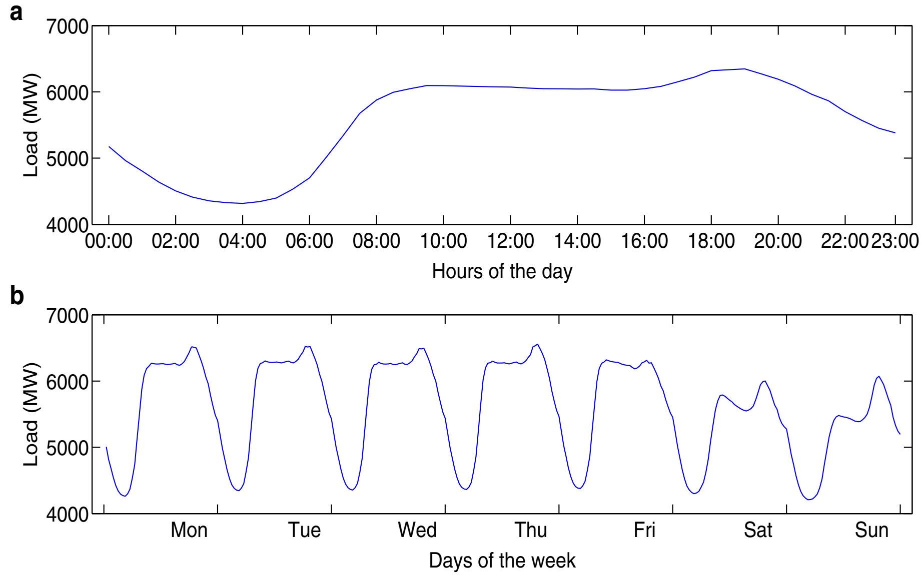

Fig. 1. Average half-hourly load over a day and average half-hourly load over a week in panels (a) and (b) respectively, for Queensland over the period from 12 July 1999 to 27 November 2013.

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, Espasa, and Grafe (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 comparison with two traditional ways of dealing with seasonality, namely the double seasonal ARIMA (Taylor & McSharry, 2007) and the double seasonal Holt-Winters exponential smoothing approach (Gould et al., 2008). After these refinements are incorporated into the final preferred multiple equation model, its superior forecast performance is demonstrated by comparison with the multiple equation model of Cancelo et al. (2008) and the semi-parametric approach used by AEMO.

2. A prototype multiple equation model

Fig. 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 12 July 1999 to 27 November 2013. Diurnal and weekly patterns, both well documented features of electricity load (Engle, Granger, & Hallman, 1989; Harvey & 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 Fig. 1 smooths out much of the half-hourly variation in load and it is this variation that a good forecasting model must capture.

A structure that captures half-hourly variability in load and respects the features of the load profile in Fig. 1 is one in which each half hour is modelled separately. Let the logarithm of the load at half hour h and day d be given by L_{hd}, then, the ARMA structure of the prototype model for a given half hour period is

$$L_{hd} = \theta_{h0} + \theta_{h1} L_{hd-1} + \theta_{h2} L_{hd-7} + \phi_{h1} \varepsilon_{hd-1} + \phi_{h2} \varepsilon_{hd-7} + \varepsilon_{hd},$$

in which $h = 1, \ldots, 48$ and ε_{hd} 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 for that half-hour interval. In this way, the partial correlation between load and lagged load is allowed to differ in a daily pattern by the different parameter values across equations. A minimal lag structure requires L_{hd} to be explained by load in the same half hour on the previous day, L_{hd-1} and the load in the same half hour of the same day in the previous week, L_{hd-7}. For the same reasoning, the unexpected changes in load in the same half hour on the previous day, ε_{hd-1} and the previous week, ε_{hd-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 be estimated, these special days are categorised into six distinct groups. Good Friday, Easter 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.

It is a well established fact that load is related to temperature. There is also 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: 9–15 degree centigrade, 9–20 degree centigrade, 22–26 degree centigrade and 22–30 degree centigrade. Temperatures between 20 degree centigrade and 22 degree centigrade are regarded as comfortable and having no extra effect on load. Also the temperatures beyond 9