


Bayesian Modeling and Forecasting of Intraday Electricity Load
Author(s): Remy Cottet and Michael Smith
Source: *Journal of the American Statistical Association*, Vol. 98, No. 464 (Dec., 2003), pp. 839-849
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/30045335
Accessed: 15-12-2016 16:45 UTC

Bayesian Modeling and Forecasting of Intraday Electricity Load

Remy COTTET and Michael SMITH

The advent of wholesale electricity markets has brought renewed focus on intraday electricity load forecasting. This article proposes a multiequation regression model with a diagonal first-order stationary vector autoregresson (VAR) for modeling and forecasting intraday electricity load. The correlation structure of the disturbances to the VAR and the appropriate subset of regressors are explored using Bayesian model selection methodology. The full spectrum of finite-sample inference is obtained using a Bayesian Markov chain Monte Carlo sampling scheme. This includes the predictive distribution of load and the distribution of the time and level of daily peak load, something that is difficult to obtain with other methods of inference. The method is applied to several multiequation models of half-hourly total system load in New South Wales, Australia. A detailed model based on 3 years of data reveals trend, seasonal, bivariate temperature/humidity, and serial correlation components that all vary intraday, justifying the assumption of a multiequation approach. Short-term forecasts from simple models highlight the gains that can be made if accurate temperature predictions are exploited. Bayesian predictive means for half-hourly load compare favorably with point forecasts obtained using iterated generalized least squares estimation of the same models.

KEY WORDS: Bayesian model selection; Electricity demand; Markov chain Monte Carlo; Peak load forecasting; Seemingly unrelated regression; Vector autoregression.

1. INTRODUCTION

To maintain a particular voltage throughout an electricity grid, at any instant in time the amount of electricity drawn from the grid and the amount generated should balance. This amount, usually called the electricity load, is, in the absence of blackouts or load-shedding, equal to the demand for electricity. Modeling and forecasting load at an intraday resolution has long been a key activity of the electricity industry. First, power systems management need short-term forecasts with horizons of 5 minutes to 1 week ahead to ensure system stability. Second, utilities need longer-term forecasts at an intraday resolution for maintenance scheduling, operations, and capital planning. Third, forecasts of both the level and time of daily peaks in load are important for scheduling and potential load-shedding plans by both power system management and utilities. It has long been known that even marginal improvements in such forecasts can represent multimillion dollar reductions in costs to network participants and end users through a reduction in spinning reserves and probabilities of supply disruptions, and also allow improved scheduling of generation capacity.

However, with the movement to wholesale electricity markets in many regions, the importance to utilities and energy traders of load forecasting at an intraday resolution has been reinforced. Although the design of such markets varies from market to market and over time, all provide a mechanism by which available generating capacity (supply) and load (demand) directly determine the spot price at which generators sell electricity to retailers or third-party energy traders. Therefore, load is a major determinant of the electricity spot price and accurate load forecasts are an essential part of day-to-day trading activity, and the financial cost of having poor load forecasts can be great. This is particularly important during peak load periods or during seasons of high variability in load. In many markets, including the Australian market, unanticipated load at these periods can result in extraordinarily high prices, because matching generating capacity is physically constrained in its response time. In some markets, the regulator has even introduced direct incentives for improved load forecasts. For example, under the New Electricity Trading Arrangements in the United Kingdom, since March 2001 market participants can pay a direct penalty for inaccurate forecasts (Ku 2002).

It has long been known that electricity load has a large predictable component due to its very strong daily, weekly, and yearly periodic behavior, along with meteorological-based variation. Athough the meteorological variables that affect load can differ according to region, temperature appears to be by far the most important meteorological factor in most locations (Al-Zayer and Al-Ibrahim 1996). The level of current interest in intraday load forecasting in the electricity and financial industries is high, with a bewildering array of different methods available to commercial customers. Bunn and Farmer (1985) and Moghram and Rahman (1989) summarized early analyses of load. However, more recent approaches to intraday load forecasting can be categorized into methods that treat load as a univariate time series and those that consider separate equations for each intraday period. The former includes artificial neural network (ANN)-based solutions, along with semiparametric regression approaches. Although various ANN solutions are available (see Darbellay and Slama 2000 for a recent summary), one of the most popular involves using back-propagation algorithms to select parsimonious radial basis function representations of the nonlinear periodic daily and weekly effects (Hsu and Yang 1995; Lee, Cha, and Park 1992; Peng, Hubele, and Karady 1992; Ranaweera, Hubele, and Papalexopoulos 1995). Temperature and lagged variables can also be used as inputs to the ANN. Because the level of noise in the data is low, relative to the nonlinear signal, these parsimonious interpolating methods produce quality forecasts for short-term hori-

Remy Cottet is a doctoral candidate and Michael Smith is Associate Professor, Discipline of Econometrics and Business Statistics, Faculty of Economics and Business, University of Sydney, Sydney, NSW 2006, Australia (E-mail: *mikes@econ.usyd.edu.au*). The work of Michael Smith was supported by an Australian Research Council Grant. Remy Cottet is currently receiving an Australian International Postgraduate Research Scholarship. This project made extensive use of the Australian Partnership for Advanced Computing's National Facility. The authors thank Gavin Dowdell and Eva Knight for discussions with regards to load forecasting, as well as Tony Baldwin and Dick Whitaker of the Australian Bureau of Meteorology for suppling meteorological data. They also thank Denzil Fiebig, Alan Woodland, three referees, an associate editor, and the coordinating editor for constructive criticism on earlier drafts of the manuscript.

Journal of the American Statistical Association
December 2003, Vol. 98, No. 464, Applications and Case Studies
DOI 10.1198/016214503000000774

zons of 5 minutes to several days, although their reliability for longer-term horizons is largely undocumented. The statistical equivalent is to estimate an additive semiparametric regression model, which also results in the estimation of smooth nonlinear periodic daily and weekly effects (Harvey and Koopman 1993; Smith 2000). However, it is difficult to capture the strong intraday dependencies that are apparent in almost all parameters when load is considered as a univariate time series.

Therefore, an alternative to modeling and forecasting electricity load is to use multi-equation regression models employed in econometric analysis (Fiebig, Bartels, and Aigner 1991; Peirson and Henley 1994; Ramanathan, Engle, Granger, Vahid-Araghi, and Brace 1997). Here, each intraday period is written down as a separate parametric regression, which allows the model for the signal and any dynamic specifications to depend on the time of day. Recent research points to the highly competitive nature of the resulting forecasts. For example, Ramanathan et al. (1997) provided an extensive comparison of several multiequation regression models with two ANN methods and the semiparametric regression method of Harvey and Koopman (1993), in which the performance of the multiequation regressions are consistently superior.

This article presents a Bayesian approach for estimating multiequation models for intraday load forecasting that extends previous work in two important ways. First, it refines the multiequation load models used in previous analysis. This includes using a first-order vector autoregression (VAR) for the errors, where stationarity is enforced through the prior for the autoregressive parameters. The bandwidth of each column of the Cholesky factor of the precision matrix of the disturbances can be determined using a Bayesian model selection framework. This corresponds to exploring the intraday autocorrelation structure in the regression errors. In addition, selection between subsets of regressors can be undertaken simultaneously for all intraday regression equations also using a Bayesian model selection approach. Second, such intraday multiequation models involve too many parameters to estimate directly using maximum likelihood and are estimated using either ordinary least squares (OLS) on each equation or iterated feasible generalized least squares (IGLS). This article develops a Markov chain Monte Carlo (MCMC) algorithm that provides the full spectrum of finite-sample inference from the joint posterior distribution of the parameters, as well as forecasts from the full predictive distributions of load and daily peak load. In this sense, this approach provides a more complete approach to inference in these models. Each step of the sampling scheme involves generation from conditional posteriors that are unrecognizable, so that carefully constructed Metropolis–Hastings steps must be used to obtain iterates.

The method is applied to estimate models for half-hourly total system load in New South Wales (NSW), Australia. Using 3 years of load data, the functional form of a bivariate temperature and humidity effect is investigated using a radial basis. The seasonal evolution in the load profile is captured using seasonal polynomials, and long-term changes in the load profile are also investigated using quadratic trend variables for each intraday period. Results reveal that the weekly, seasonal, meteorological, and dynamic effects differ substantially at different times of day, confirming the basic precept of a multiequation model. Forecasts at a half-hourly resolution with a 6-month horizon are obtained with an average absolute percentage error of 3.1%. A short-term model using a running window of 3 weeks of data is also estimated, and the predictive distribution of load with a horizon of 7 days is obtained. Comparisons with forecasts derived using IGLS show that the Bayesian marginal predictive means form competitive point forecasts. In addition, the intraday predictive distribution of load and daily peak levels and times, which are difficult to evaluate using previous methods of inference, are also determined.

The article is organized as follows. Section 2 contains a brief introduction to features of load data generally and the NSW data in particular. Section 3 outlines the multiequation model for electricity load, along with the associated likelihood. Section 4 discusses the Bayesian estimation methodology, with details of the derivation of the Metropolis–Hastings sampling steps deferred to the Appendix. Section 5 applies the method to the NSW electricity load data and reports the empirical results, and Section 6 concludes the article.

2. ELECTRICITY LOAD DATA

The high degree of predictability in load data is largely as a result of these data's very strong daily, weekly, and yearly periodic behavior. For example, Harvey and Koopman (1993) and Smith (2000) have provided plots of intraday system load for 1 week in each of the four seasons from the northwest United States and NSW, Australia. Apart from the strong daily and weekly periodicity, the load profiles also vary substantially across seasons. Many authors address this issue of seasonal variation in load profiles by using a short-term model with a 2- or 3-week moving window of data, over which the periodic profile is assumed to be static (e.g., Hsu and Yang 1995; Smith 2000). Alternatively, a long-term model with multiple years of intraday data can be used and the seasonal effect modeled (e.g., Harvey and Koopman 1993).

The most important source of variation in load, over and above the strong periodicity, is meteorology, particularly temperature. The relationship between load and temperature is known to be approximately "V"-shaped, with a minimum around 18.3°C/65°F (see Engle, Granger, Rice, and Weiss 1986; and Al-Zayer and Al-Ibrahim 1996 for in-depth investigations into the functional form of the relationship). The relationship is also known to differ depending on the time of day (Peirson and Henley 1994). Ramanathan et al. (1997) incorporated this into their regression models for hourly load, whereas Smith (2000) used a nonparametric bivariate surface in temperature and time of day to model the effect of temperature on intraday load. In addition, in some climates humidity is thought to have a major effect on electricity load, but through an interaction with temperature and time of day in an unknown functional form. Along with day-type data, we use half-hourly observations on the following variables to capture these effects:

- $\text{LOAD}_{i,t}$: Total NSW electricity system load for day t and half hour i (MW hours).
- $\text{TEMP}_{i,t}$: Temperature at the Bankstown airport monitoring station in Sydney for day t and half hour i (degrees centigrade).

- $\text{HMD}_{i,t}$: Relative humidity at the Bankstown airport monitoring station in Sydney for day t and half hour i (in percent).
- $\text{TOY}_{i,t}$: Time of year of day t ($0 \le \text{TOY}_{i,t} < 1$).

3. MULTIEQUATION MODEL FOR ELECTRICITY LOAD

We model the half-hourly load observed over T days with $m = 48$ separate regressions for each half hour, so that

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\eta}_i + \mathbf{u}_i, \qquad \text{for } i = 1, \ldots, m,$$

is the regression model for half hour i. Here $\mathbf{y}_i = (\text{LOAD}_{i,1}, \text{LOAD}_{i,2}, \ldots, \text{LOAD}_{i,T})'$, $\mathbf{u}_i = (u_{i,1}, \ldots, u_{i,T})'$, and $\boldsymbol{\eta}_i = (\eta_{i,1}, \ldots, \eta_{i,p_i})'$ are the coefficients for the p_i variables in the regression for half hour i and $\mathbf{X}_i$ is the corresponding $T \times p_i$ design matrix. This design matrix can contain meteorological variables or periodic variables such as TOY, depending on whether a short-term or long-term model is used. The m regressions can be stacked, so that $\mathbf{y} = \mathbf{X}\boldsymbol{\eta} + \mathbf{u}$, where $\boldsymbol{\eta}' = (\boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_m')$, $\mathbf{y}' = (\mathbf{y}_1', \ldots, \mathbf{y}_m')$, $\mathbf{u}' = (\mathbf{u}_1', \ldots, \mathbf{u}_m')$, and $\mathbf{X} = \text{bdiag}(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ is a $(mT \times \sum_{i=1}^m p_i)$ block-diagonal matrix. The daily load profiles as a whole are thought to exhibit a degree of serial correlation. To model this, the regression errors are assumed to follow a simple VAR of lag length 1, so that $\mathbf{u}_{(t)} = \boldsymbol{\Phi}\mathbf{u}_{(t-1)} + \mathbf{e}_{(t)}$ for $t = 1, 2, \ldots, T$, where $\mathbf{u}_{(t)} = (u_{1,t}, u_{2,t}, \ldots, u_{m,t})'$ is the error vector for day t and $\boldsymbol{\Phi} = \text{diag}(\phi_1, \ldots, \phi_m)$ is a diagonal autoregressive parameter matrix (Judge et al. 1985, pp. 483–489). Stationarity is enforced in this interday VAR model by restricting each of the autoregressive coefficients to the unit interval, so that $-1 < \phi_i < 1$. Throughout the article, we use plain subscripts for vectors indexed by half hour and subscripts with parentheses to denote vectors indexed by day.

We capture any intraday serial correlation by assuming that the random disturbances of the VAR are independently and identically distributed $\mathbf{e}_{(t)} \sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Previous analyses (Smith and Kohn 2002) have suggested that this intraday residual correlation is likely to be well modeled by a short-lag autoregressive-style model with autoregressive coefficients that differ over half hours. Such models correspond to a parsimonious form for the Cholesky factorization $\boldsymbol{\Sigma}^{-1} = \mathbf{B}\mathbf{D}\mathbf{B}'$. Here $\mathbf{D}$ is a diagonal matrix and $\mathbf{B}$ is a banded lower triangular Cholesky factor, with 1's on the leading diagonal. Such parsimonious forms for correlation structures have been widely used in longitudinal modeling (see Pourahmadi 1999 and references therein). We denote the bandwidth of the ith column of $\mathbf{B}$ as k_i, which is equal to the length of the intraday lag at half hour i of the longitudinal model. In this article, we attempt to identify the lags $\mathbf{k} = (k_1, \ldots, k_{47})'$, given that $k_i \le k_{\max}$. In our analysis we consider $\text{LOAD}_{1,t}$ to correspond to the reading at 03:30, and $\text{LOAD}_{m,t}$ to correspond to the reading at 03:00. That is, a day is considered to start at 03:30 rather than at 00:30, which is a natural separation point for daily load profiles and corresponds approximately to the time of minimum daily usage.

The likelihood of the data y is

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\Sigma}, \boldsymbol{\Phi}, \boldsymbol{\eta}) &= p\big(\mathbf{u}|\boldsymbol{\Sigma}, \boldsymbol{\Phi}\big) \\ &= p\big(\mathbf{u}_{(1)}|\boldsymbol{\Sigma}, \boldsymbol{\Phi}\big)\, p\big(\mathbf{u}_{(2)}, \ldots, \mathbf{u}_{(T)}\big|\mathbf{u}_{(1)}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}\big). \end{aligned}$$

In the foregoing, the likelihood is decomposed into a product of the density of $\mathbf{u}_{(1)}$ and the likelihood conditional on $\mathbf{u}_{(1)}$. Assuming that the error process is stationary, it is well known (Judge, Griffiths, Hill, Lutkepohl, and Lee 1985, p. 485) that the distribution of the initial observation is $\mathbf{u}_{(1)}|\boldsymbol{\Sigma}, \boldsymbol{\Phi} \sim \text{N}(\mathbf{0}, \boldsymbol{\Gamma})$, where $\text{vec}(\boldsymbol{\Gamma}) = (\mathbf{I} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi})^{-1} \text{vec}(\boldsymbol{\Sigma})$. The conditional likelihood can be written as

$$p\big(\mathbf{u}_{(2)}, \ldots, \mathbf{u}_{(T)}\big|\mathbf{u}_{(1)}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}\big) = \prod_{t=2}^{T} p\big(\mathbf{u}_{(t)}\big|\mathbf{u}_{(t-1)}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}\big) \tag{1}$$

$$\propto |\mathbf{D}|^{(T-1)/2} \exp\Bigg\{-\frac{1}{2}\sum_{t=2}^{T}\big(\mathbf{u}_{(t)} - \boldsymbol{\Phi}\mathbf{u}_{(t-1)}\big)' \times \mathbf{B}\mathbf{D}\mathbf{B}'\big(\mathbf{u}_{(t)} - \boldsymbol{\Phi}\mathbf{u}_{(t-1)}\big)\Bigg\} \tag{2}$$

$$\propto \prod_{i=1}^{m} d_i^{(T-1)/2} \exp\Bigg\{-\frac{1}{2}\sum_{i=1}^{m} d_i \mathbf{b}_i' \mathbf{A} \mathbf{b}_i\Bigg\}, \tag{3}$$

where $\mathbf{b}_i$ is the ith column of $\mathbf{B} = \{b_{j,i}\}$ and $\mathbf{A} = \sum_{t=2}^{T}(\mathbf{u}_{(t)} - \boldsymbol{\Phi}\mathbf{u}_{(t-1)})(\mathbf{u}_{(t)} - \boldsymbol{\Phi}\mathbf{u}_{(t-1)})'$. Because $\mathbf{B}$ is a Cholesky factor, the quadratic form in (3) can be written in terms of the vector of nonfixed elements of $\mathbf{b}_i$, which we denote as $\boldsymbol{\beta}_i = \{b_{j,i} | i < j < \min(i + k_i, m)\}$, so that

$$\mathbf{b}_i' \mathbf{A} \mathbf{b}_i = \begin{cases} a_{i,i} + 2\boldsymbol{\beta}_i' \mathbf{a}_i + \boldsymbol{\beta}_i' \mathbf{A}_i \boldsymbol{\beta}_i & \text{for } i = 1, \ldots, m-1 \\ a_{m,m} & \text{for } i = m. \end{cases} \tag{4}$$

Here $a_{j,i}$ is the j,ith element of $\mathbf{A}$, the vector $\mathbf{a}_i = \{a_{j,i} | i < j < \min(i + k_i, m)\}$, and the matrix $\mathbf{A}_i = \{a_{l,j} | i < l < \min(i + k_i, m), i < j < \min(i + k_i, m)\}$.

4. BAYESIAN METHODOLOGY

4.1 Bayesian Prior Assumptions

The elements of the autoregressive coefficient matrix are assumed to be a priori independent and constrained to the stationary region, so that $p(\boldsymbol{\Phi}) = \prod_{i=1}^{m} \mathcal{I}(-1 < \phi_i < 1)$. Here the indicator function $\mathcal{I}(x) = 1$ if condition x is true and 0 otherwise. The elements of $\mathbf{D}$ are taken to be a priori independent $d_i \sim \text{gamma}(\zeta/\kappa, \kappa)$. We set $\zeta = 100$ and $\kappa = 1{,}000$, which is an uninformative prior compared with the scale of the data. The bandwidths of the columns of the Cholesky factor $\mathbf{B}$ are assumed to be a priori independent and flat, so that $p(k_i) = 1/\min(m - i, k_{\max})$. As is noted in the literature on Bayesian model uncertainty, a proper conditional prior is required for the nonfixed elements of $\mathbf{B}$. We use the same proper conditional prior $p(\mathbf{B}|\mathbf{k}, \mathbf{D}, \boldsymbol{\Phi})$ used to account for model uncertainty in $\mathbf{B}$, as suggested by Smith and Kohn (2002). This prior is taken proportional to $\frac{1}{c}$th of the conditional likelihood, so that $p(\mathbf{B}|\mathbf{k}, \mathbf{D}, \boldsymbol{\Phi}) \propto p(\mathbf{u}_{(2)}, \ldots, \mathbf{u}_{(T)}|\mathbf{u}_{(1)}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{\eta})^{1/c}$. Completing the square in $\boldsymbol{\beta}_i$ in the foregoing equation leads to the prior where the free elements of the columns of $\mathbf{B}$ are independently distributed $\boldsymbol{\beta}_i|\mathbf{k}, \mathbf{D}, \boldsymbol{\Phi}, \boldsymbol{\eta} \sim \text{N}(-\mathbf{A}_i^{-1}\mathbf{a}_i, \frac{c}{d_i}\mathbf{A}_i^{-1})$. In our empirical work, we chose $c = n$, which corresponds to a conjugate proper prior that is relatively uninformative compared with the likelihood. In some of the regression models that we fit, we consider model choice between subsets of the regressors simultaneously for all equations. To do this, we introduce a discrete

model variable, $\delta_i \in \{1, 2, \ldots, M\}$, for each equation, along with the proper prior, $\boldsymbol{\eta}_i|\delta_i = j \sim \mathrm{N}(\boldsymbol{\mu}_{ij}^0, \mathbf{V}_{ij}^0)$. Here $\boldsymbol{\mu}_{ij}^0$ is the OLS estimator of $\boldsymbol{\eta}_i$ under model $\delta_i = j$, whereas the matrix $\mathbf{V}_{ij}^0$ is the variance of this estimator scaled by a factor of 10. As with the fractional prior for $\mathbf{B}$, this conditional prior is located and scaled so that it is relatively uninformative compared with the likelihood. In cases where no model comparison is undertaken, for simplicity, a flat prior for $\boldsymbol{\eta}$ is used.

4.2 Sampling Scheme and Inference

It is difficult to compute analytically the posterior distributions of the parameters. Thus we use the following MCMC sampling scheme:

- Step 1. Generate from $p(\boldsymbol{\eta}_i, \delta_i|\tilde{\boldsymbol{\eta}}_{j\neq i}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}, \mathbf{y})$ for $i \in \{1, 2, \ldots, m\}$.
- Step 2. Generate from $p(\boldsymbol{\phi}_i|\boldsymbol{\phi}_{j\neq i}, \boldsymbol{\Sigma}, \boldsymbol{\eta}, \mathbf{y})$ for $i \in \{1, 2, \ldots, m\}$.
- Step 3. Generate from $p(\mathbf{D}|\mathbf{B}, \boldsymbol{\Phi}, \boldsymbol{\eta}, \mathbf{y})$.
- Step 4. Generate from $p(\boldsymbol{\beta}_i, k_i|\boldsymbol{\beta}_{j\neq i}, \mathbf{D}, \boldsymbol{\Phi}, \boldsymbol{\eta}, \mathbf{y})$ for $i \in \{1, \ldots, m-1\}$.

Because of the density of the initial observation $p(\mathbf{u}_{(1)}|\boldsymbol{\Sigma}, \boldsymbol{\Phi})$ and fractional prior $p(\mathbf{B}|\mathbf{D}, \boldsymbol{\Phi}, \boldsymbol{\eta}, \mathbf{k})$, the conditional distributions in these four steps are all unrecognizable in form. Therefore, to obtain iterates of the parameters, we construct four separate Metropolis–Hastings steps, which we outline in the Appendix.

For step 1, we generate the $\boldsymbol{\eta}_i$ and δ_i as a pair, because a sampling scheme that generated δ_i conditional on $\boldsymbol{\eta}_i$ would be reducible. When no model choice between subsets of regressors is required, the discrete variables δ_i are known, and we block together randomly selected sets of regressors to improve mixing. However, the sampling scheme demonstrates empirically reasonable mixing and convergence properties even without such blocking. The conditional distribution of $\boldsymbol{\Phi}$ is a multivariate constrained distribution from which it is difficult to generate, so we generate the $\boldsymbol{\phi}_i$ one at a time at step 2. The matrix $\mathbf{D}$ is a generated as a block from an approximation to the density at step 3 that is exact, except for the density of the initial observation. As with step 1, we generate the $\boldsymbol{\beta}_i$ and k_i as a pair in step 4 to avoid a reducible sampling scheme. In our empirical work, acceptance rates for these steps varied between 10% and more than 90% for step 1, 90% for step 2, more than 85% for step 3, and between 40% and 90% for step 4.

We run the sampling scheme for 10,000 iterations, after which we assume that it has converged to the joint posterior distribution. Then we generate further 10,000 iterations to use for Monte Carlo-based inference. We use histogram estimates of their marginal posterior means of the parameters as point estimates. Estimates of the Bayesian posterior probability intervals for each parameter can be obtained by ranking the iterates of the parameter and counting off an appropriate number of the largest and smallest values. Forecasts are obtained by using the predictive density of future load $\mathbf{y}^f$ over the forecast interval given by the integral $p(\mathbf{y}^f|\mathbf{y}) = \int p(\mathbf{y}^f|\mathbf{y}, \Theta)p(\Theta|\mathbf{y})\,d\Theta$, where $\Theta = \{\boldsymbol{\Phi}, \boldsymbol{\eta}, \boldsymbol{\Sigma}\}$. Using the iterates $\Theta^{[j]} \sim p(\Theta|\mathbf{y})$, the integral can be evaluated in a Monte Carlo fashion by sampling from $p(\mathbf{y}^f|\mathbf{y}, \Theta)$, which is straightforward because this density is a multivariate normal. The mixture estimate of the predictive mean, $E(\mathbf{y}^f|\mathbf{y}) = (1/J)\sum_{j=1}^{J} E(\mathbf{y}^f|\mathbf{y}, \Theta^{[j]})$, is used as the point forecast.

In the electricity industry, predictive probability intervals for future load at any half hour are of particular importance for scheduling of capacity and risk management. In many ways, they are of greater importance than point forecasts of future load. Monte Carlo estimates of these can be constructed by generating iterates from the predictive distribution $p(\mathbf{y}^f|\mathbf{y})$, ranking them, and counting off an appropriate number of the highest and lowest values. Using the Monte Carlo iterates of future load, it is also straightforward to identify the level and time of an intraday peak for each iterate, from which histogram estimates of the predictive distribution of the time and level of an intraday peak can be evaluated. Note that although predictive intervals for future load could be approximated using estimates of the second moment of $\mathbf{y}^f|\mathbf{y}$, this approach is not viable for estimates of the level and time of intraday peaks, because they both prove asymmetric, and the time of peak is occasionally multimodal.

5. APPLICATION TO NEW SOUTH WALES LOAD DATA

5.1 Long-Term Models

We fit the following model to 3 years of half-hourly NSW system load data, starting on 1/1/98 and finishing on 12/30/00 ($T = 1{,}095$):

$$\mathrm{LOAD}_{i,t} = \sum_{j=1}^{13} \eta_{i,j}^d \mathrm{DUM}_j(t) + f_i(\mathrm{TEMP}_{i,t}, \mathrm{HMD}_{i,t}) + s_i(\mathrm{TOY}_{i,t}) + r_i(t) + u_{i,t} \quad (5)$$

for $i = 1, \ldots, m$ and $t = 1, \ldots, T$. The function f_i is a bivariate temperature and humidity effect, s_i is a seasonal effect, and r_i captures the long-term trend in load. The day-type effect is a linear combination of 13 dummy variable terms, DUM_j, one for each of the 7 days of the week and six for public holidays. Here Christmas, Boxing Day, New Year's Day, Good Friday, and Easter are the most idiosyncratic public holidays and have separate dummies, whereas all others correspond to a single dummy. The 48 coefficients of each dummy variable parameterize basic daily load profiles associated with each day type. The seasonal devolution in load profiles has been recognized by many authors (e.g., Harvey and Koopman 1993; Engle, Granger, and Hallman 1989). Using several years of data, this seasonal effect s_i is modeled using a finite Fourier series with the 12 terms $\cos(2\pi j\mathrm{TOY})$, $\sin(2\pi j\mathrm{TOY})$ for $j = 1, \ldots, 6$. Because each half-hour is modeled with a separate seasonal effect, the estimates reflect how the typical daily load profile devolves throughout the year. Compare this with the time-varying smoothing splines of Harvey and Koopman (1993), which do not directly model the devolution but rather account for such variation by using a random walk on the coefficients of the smoothing splines. To model the trend, we consider a simple quadratic function for each r_i.

Although a trend, seasonal and day-type effects are widely recognized components of intraday load in the load forecasting literature, the form and importance of any meteorological-based

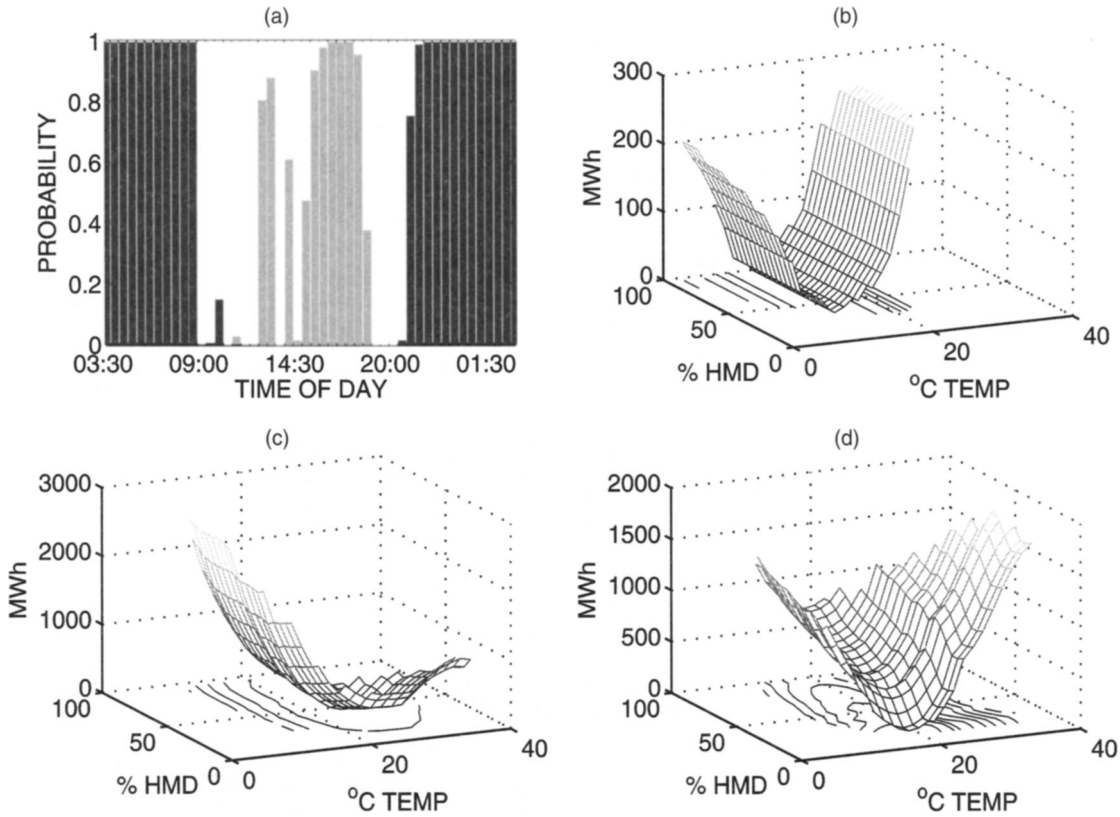

Figure 1. (a) Barplot of Estimated Marginal Posterior Probabilities $p(\delta_i|\mathbf{y})$ Against Time of Day, and (b, c, and d) Estimates of $E(f_i|\mathbf{y})$ Over the Convex Hulls of the Observed Humidity and Temperature Variables for 5:30, 12:00 and 17:30.

component is less certain. Here, we decompose the bivariate effect into the addition of univariate main effects, plus another univariate effect in the multiplication of TEMP and HMD, so that $f(\text{TEMP}, \text{HMD}) = f_1(\text{TEMP}) + f_2(\text{HMD}) + f_{12}(\text{TEMP} * \text{HMD})$. This can prove to be a reasonable approximation to smooth nonlinear bivariate functions; Shively and Sager (1999) took such an approach applied to the effect of temperature and humidity on ozone maxima. Each of these univariate functions are modeled using a radial basis with 11 knots placed within the domain of the observations (Powell 1987). Following previous authors, we augment the basis with a linear term. Collecting the effects together, (5) can be written as a linear model where $\boldsymbol{\eta}$ is a vector with 48×63 elements. For each half-hour, we consider the following five models for the bivariate meteorological effect: no effect, f_1, f_2, $f_1 + f_2$, and $f_1 + f_2 + f_{12}$. These are denoted as models $\delta_i = 1, \ldots, 5$.

The sampling scheme was run and the long-term model at (5) estimated. Estimates of the marginal posterior model probabilities $p(\delta_i = 1|\mathbf{y})$ and $p(\delta_i = 3|\mathbf{y})$ are 0 up to five decimal places for all half hours, so that temperature is found to relate to load throughout the day. The remaining probabilities vary according to half hour, and Figure 1(a) shows these in barplot format. Each bar corresponds to a separate half hour, with the probability $p(\delta_i = 2|\mathbf{y})$ as black, $p(\delta_i = 4|\mathbf{y})$ as white, and $p(\delta_i = 5|\mathbf{y})$ as gray. Humidity has zero probability of affecting load during the low temperature/load overnight period of 21:30 to 8:00. Figure 1(b) shows the surface estimate $E(f_i|\mathbf{y})$ for 5:30; the univariate temperature effect has a minimum around 11°C and a relatively minor relationship with load. During the afternoon period 15:00 to 18:00, the full bivariate model ($\delta_i = 5$), including the interaction effect f_{12}, is selected, although outside these periods, the relationship is additive. Figures 1(c) and (d) provide estimates of the bivariate surfaces at 12:00 and 15:30. The minimum of these functions with respect to temperature is close to 18.3°C/65°F, which is the point of inversion for parametric models widely used in the electricity industry. Overall, it is possible to conclude that temperature is a prime meteorological driver of NSW load and has a substantial effect during peak hours, although much less so during overnight periods. Humidity also has an effect during peak hours, although as an interaction with temperature during the afternoon. The implication for utilities is that incorporation of accurate meteorological forecasts (particularly temperature) has the potential to improve load forecasts overall, especially during peak hours. This issue is explored further in forecasting short-term load.

The estimates of the dummy variable coefficients $\eta^d_{i,j}$ were similar for the working days Monday to Friday ($j = 1, \ldots, 7$), but with a slightly lower load on Friday afternoon and Monday morning, reflecting spillover from the weekends in NSW working patterns. Loads on Saturday are higher than those on Sunday up until around 17:00, which reflects retail trading during Saturday. Loads on the public holidays vary depending on the holiday type, although they follow a profile similar to that of Sunday. The seasonal effects, $s_1, \ldots, s_{48}$, can be considered as a bivariate surface over time of year and day. Figure 2 summarizes their estimates by four cross-sections in both dimensions. Panel (a) plots the estimates against the time of year for four half hours, whereas (b) plots the estimates

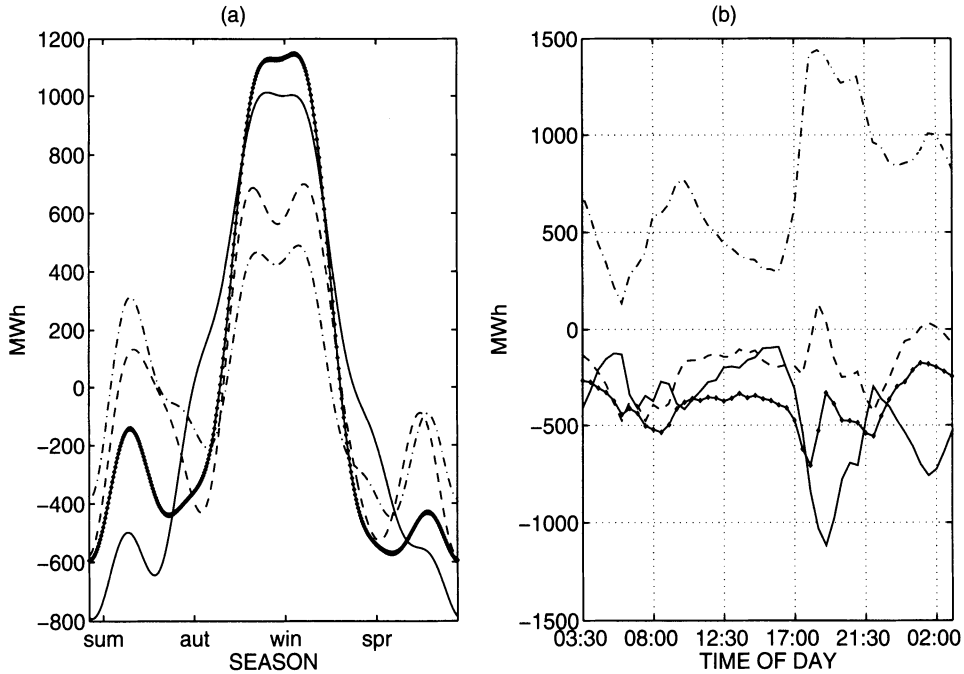

Figure 2. Cross-Sections of Estimated Seasonal Effects for (a) Season and (b) Time of Day [(a) —, 02:00; - - -, 08:00; ·-·-, 12:30; -●-, 21:30. (b) —, summer; - - -, autumn; ·-·-, winter; -●-, spring].

against time of day evaluated at four times of the year: January 15 (summer), April 15 (autumn), July 14 (winter), and October 12 (spring). Panel (a) shows that load peaks in NSW in winter (June, July, and August), most prominently in the evening hours (18:00–03:00), likely due to residential heating-based load. Two smaller peaks occur in December and from late January to February. December to February are the hottest months of the year in NSW, and these two smaller peaks correspond to summer air-conditioning load, so that it is not surprising that these effects are most notable during the daylight hours (06:00–18:00). Load dips between Christmas and mid-January, because traditionally during this period many firms close for periods up to 1 month and nonresidential demand is low. Panel (b) shows that for mid-summer, the seasonal effect on load is highest during the period 10:30–17:00, probably due to preprogrammed office air-conditioning. The seasonal effect on the load profile during mid-spring and autumn are similar, not surprising, because meteorological conditions are similar in these two seasons in NSW.

Figure 3 plots the long-term trend components $r_1, r_2, \ldots, r_{48}$ evaluated on 1/1/99, 1/1/00, and 12/30/00, compared with a base of $r_i = 0$ at the start of our data in 1/1/98. This plot reveals an increase in basic load of 650–1,150 MWh over the 3-year period. This increase was greatest around 10:00 and 19:30, corresponding to morning and evening peak periods. The plot shows that attempts to push demand into the off-peak period (22:00–6:00) and produce a more uniform daily load profile appears to have had effect during the 3-year period. Such a trend is disturbing for long-term energy planning in NSW, where increasing demand for electricity needs to be matched by long-term investment in additional generating capacity.

We reestimate the model (5) with the meteorological components omitted and $k_{\max} = 6$, and calculate the predictive distribution of future half-hourly load for a horizon of 6-month horizon. The meteorological effects were omitted, because it is not feasible to obtain reliable forecasts of meteorological conditions at a half-hourly resolution for long horizons. The model was estimated using data from 1/1/98 to 2/5/00 and the predictive distributions calculated for the period 2/6/00 to 7/12/00, which was selected to avoid the one-off effect of the Sydney 2000 Olympic games. The estimates of the seasonal, day type, and long-term trend components of the model were similar to those obtained previously. Much of the now-omitted meteorological effect is accounted for by the intraday autocorrelation structure found in $\mathbf{\Sigma}$, because meteorological conditions tend to persist for much of the day. Figure 4 contains the mode of the posterior distribution of k_i for each half hour. Longer bandwidths ($k_i = 5$ or 6) are prominent for most columns, except for the peak period 9:00–22:00, where shorter lags occur and intraday serial correlation has a shorter memory. Examining the

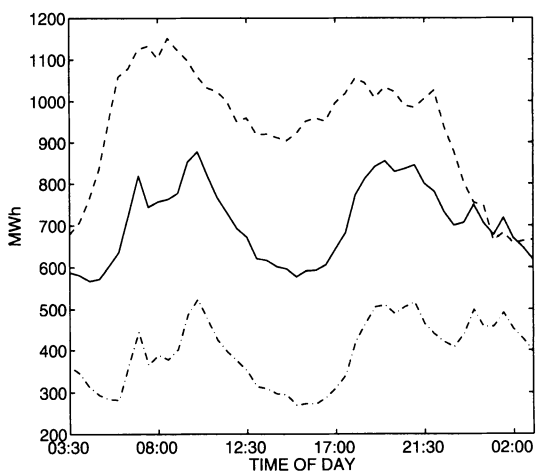

Figure 3. Estimated Quadratic Trend From the Long-Term Model (·-·-, 1/1/99; —, 1/1/00; - - -, 12/29/01).

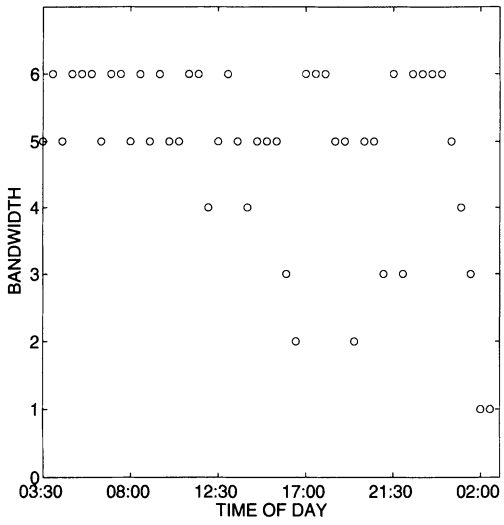

Figure 4. Mode of Distribution of $p(k_i|\mathbf{y})$ Against Time of Day for Long-Term Model.

estimate of $E(\mathbf{B}|\mathbf{y})$ reveals that the first band features values close to -1. In terms of an antedependent model, this corresponds to positive first-order intraday autocorrelations. Subsequent bands have values close to 0 and a much lower impact, except around 5:30 and 22:00, where the elements of the second and third band have absolute value of around .5. Note that these two time points are where the daily load profile switches between day and overnight usage patterns.

To assess the accuracy of our forecasts, we calculate the absolute percentage error (APE) for each half hour in the 6-month forecast period, and then evaluate the daily mean APE (MAPE) for each of the days in our forecast horizon. Table 1 contains summary statistics of the daily MAPE during the 6-month forecast period, broken down by day type in the first four rows and by month in the horizon in the last six rows. Public holidays were excluded in the monthly breakdown, because there were only seven in the forecast period. Notice that the forecasts do not degrade during the 6-month period, but are relatively stable for such a horizon. This is because the underlying periodicity and trend that dominate load is captured by the model, whereas any residual serial correlation associated with the VAR has an effect only on the predictive distribution of the first week in the horizon. Forecasts prove less accurate on weekends than on weekdays, because usage patterns are more volatile on these day types. The figures show that the public holidays can be poorly forecast. Utilities often consider forecasting load on public holidays using separate tailored models that include many years of data. June is poorly forecast because it was abnormally cold and wet compared with June in the previous 3 years.

Table 1. Summary of the Daily MAPE's for Half-Hourly Load Forecasts With a 6-Month Horizon

Summary	*Public Holiday*	*Weekdays*	*Weekends*
Mean	$8.37_{(11.67)}$	$3.10_{(3.08)}$	$3.43_{(3.59)}$
Median	$7.50_{(10.44)}$	$3.10_{(3.07)}$	$3.33_{(3.56)}$
Min	$.21_{(.08)}$	$2.31_{(2.32)}$	$2.55_{(2.57)}$
Max	$20.09_{(29.33)}$	$3.82_{(3.79)}$	$4.54_{(4.77)}$
Mean daily MAPE's categorized by day type and month			
February		$3.97_{(3.69)}$	$4.00_{(4.67)}$
March		$3.17_{(3.16)}$	$3.44_{(3.46)}$
April		$2.31_{(2.14)}$	$3.06_{(3.16)}$
May		$3.14_{(3.25)}$	$3.74_{(3.83)}$
June		$4.45_{(4.55)}$	$4.53_{(4.61)}$
July		$2.17_{(2.16)}$	$2.33_{(2.41)}$

NOTE: The first four rows give summary statistics of the daily MAPE's broken down by day type; The next six rows give the mean MAPE broken down by day type and month in the forecast horizon. The figures in parentheses correspond to the results from IGLS.

Computing the direct maximum likelihood estimator of the model in (5) is difficult because of the large size of the parameter vector. However, an IGLS estimator can be computed as in Judge et al. (1985, pp. 483–490) which we iterate 20 times for efficiency. The summaries of the daily MAPE's from this model are given in parentheses in Table 1. Compared with the Bayesian results, weekday forecasts are almost identical, whereas weekends and public holidays are somewhat worse. We also estimated each equation using OLS and found that the resulting point forecasts underperformed those of the Bayesian estimator and, in particular, had higher maximum daily MAPE's. This suggests that accounting for the time series error structure improves the estimate of the mean. However, the improvement of the Bayesian forecasts over the IGLS becomes more notable for shorter-term forecasts where there is a larger VAR component. Moreover, the Bayesian framework allows for the full spectrum of inference, including finite-sample predictive probability intervals. For example, Figure 5(a) plots the predictive mean for 1 full week 6 months ahead, along with 90% probability limits for the predictive distribution. Also plotted is the percentage error when the predictive mean is used as a forecast. It can be seen that the predictive interval is wider during peak midday hours, reflecting the greater load variation during this period. This intraday heteroscedasticity is omitted from many univariate intraday time series and regression models (e.g., Smith 2000; Harvey and Koopman 1993).

5.2 Short-Term Models

Energy utilities are also interested in short-term forecasts with horizons between a few hours and 1 week ahead. Although these can be obtained from long-term models as before, a common alternative is to use a running window of data to obtain forecasts (see Ho, Hsu, and Yang 1992; Smith 2000). We investigate the applicability of our approach in obtaining forecasts using a simple short-term linear model.

We use $T = 21$ days of data for the in-sample windows, a period over which the long-term trend and seasonal devolution in the load profile is small enough to be considered approximately constant. We model load at each half hour as a linear model with 5 day-type dummy variables for Monday, Tuesday to Thursday, Friday, Saturday, and Sunday, with public holidays coded as Sundays. The parametric temperature effect, $|\text{TEMP}_{i,t} - 18.3|$, is also included for each half hour. We consider only temperature in our model, because half-hourly temperature forecasts are more readily available than other meteorological variables, including humidity. The effect of temperature is modeled as an inverted "V," with the apex at 18.3°C, which is a widely used parametric approximation for the univariate response of load to temperature (see Engle et al. 1986 for a discussion on the effect of temperature on load). A parametric form is chosen, because 3 weeks of data are unlikely to provide sufficient temperature variation to allow reliable estimation of the form of the effect.

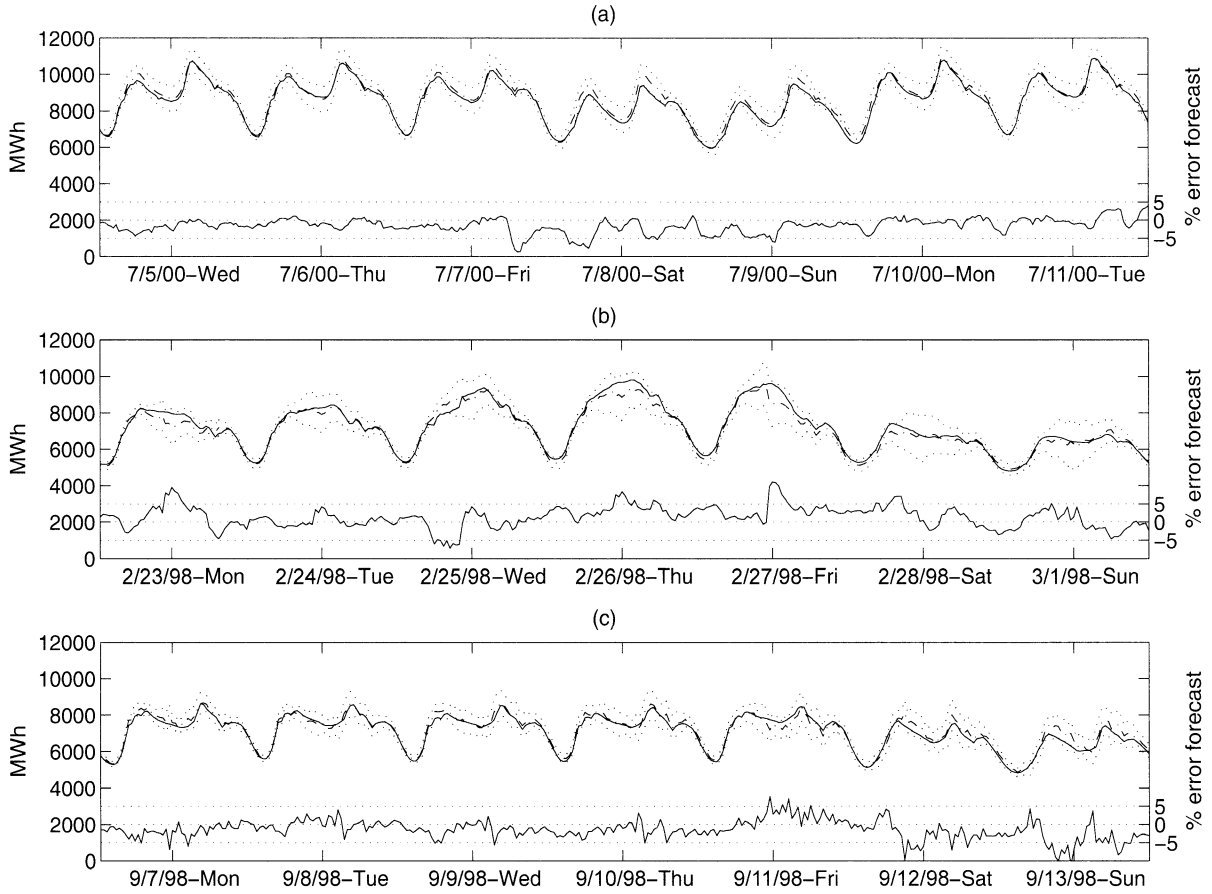

Figure 5. Plots of Weekly Half-Hourly Forecasts. The true load is represented by the bold line; the forecast, by the dashed line; and the upper and lower 90% predictive intervals, are by with dotted lines; (a) a week six months ahead from the long-term model; (b) and (c) seven-day-ahead forecasts from the short-term model, with (b) corresponding to a summer week and (c) correspondong to a winter week. The APE is plotted below, with a scale of ± 5%.

We also set $\mathbf{\Sigma}$ to be diagonal (so that $k_i = 0$), because it is hard to reliably distinguish intraday correlation from the temperature effect.

We estimate the model daily during 1998, and produce updated forecasts with 7 day horizons. The estimates of the coefficients of the parametric temperature effect are significantly positive between 9:00 and 24:00, but are close to 0 between 0:30 and 8:30. This confirms the previous analysis that sensitivity of load to temperature differs throughout the day. The estimates of the posterior means $E(\phi_i|\mathbf{y})$ vary, but are almost always positive and generally greater than .5, representing a strong interday first-order autocorrelation. Figures 5(b) and (c) plot 7-day-ahead forecasts made at 3:00 for 1 week in summer (2/23/98–3/1/98) and 1 week in winter (9/7/98–9/13/98). These forecasts are based on true temperature, because it is useful to separate the load forecasting method from the temperature forecasting method, which is a complex spatiotemporal problem. Note that the summer forecasts have wider prediction intervals, which is due to the higher degree of variability in general meteorological conditions in the NSW summer.

Table 2 presents summaries of the daily MAPE's for forecasts calculated daily with a 7-day horizon during the period 1/26/98–12/24/98. Utilities are interested in these forecasts for daily and weekly load management and short-term energy trading decisions. The forecasts degrade over the horizon, which is attributable to a measurable VAR component to the forecasts. Table 3 presents summary statistics of the daily MAPE's of 7-day-ahead forecasts made at 03:00 on the 48 Tuesdays between 1/27/98–12/22/98. Both tables also include, in parentheses, the forecast summaries obtained using IGLS estimation of the same model with a parametric temperature term. The fore-

Table 2. Summary Statistics of the Daily MAPE's of Half-Hourly Load Forecasts Produced From the Short-Term Model for a 7-Day Horizon During 1998

Days ahead	*+1*	*+2*	*+3*	*+4*	*+5*	*+6*	*+7*
Mean	$2.27_{(2.33)}$	$2.67_{(2.88)}$	$2.86_{(3.07)}$	$3.00_{(3.25)}$	$3.10_{(3.33)}$	$3.19_{(3.31)}$	$3.29_{(3.41)}$
Median	$2.21_{(2.10)}$	$2.73_{(2.87)}$	$2.91_{(3.14)}$	$3.02_{(3.21)}$	$2.99_{(3.09)}$	$3.00_{(3.20)}$	$3.25_{(3.22)}$
Min	$1.11_{(1.10)}$	$1.42_{(1.49)}$	$1.46_{(1.54)}$	$1.52_{(1.76)}$	$1.64_{(1.71)}$	$1.83_{(1.75)}$	$1.82_{(1.71)}$
Max	$4.42_{(4.54)}$	$4.63_{(4.65)}$	$4.89_{(4.75)}$	$4.85_{(5.47)}$	$5.17_{(6.06)}$	$5.06_{(5.29)}$	$5.27_{(5.60)}$

NOTE: The corresponding statistics from IGLS estimation of the same model are given in parentheses.

Table 3. Summary of Daily MAPE's of 7-Day-Ahead Forecasts Made From Simple Multiequation Short-Term Models With a Temperature Term on Tuesdays During 1998

Summary	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday
Mean	2.56(2.72)	2.75(3.12)	2.81(3.08)	2.81(3.06)	3.39(3.73)	3.73(4.00)	2.97(3.21)
Median	2.40(2.72)	2.74(3.16)	2.84(3.13)	2.85(3.05)	3.33(3.65)	3.58(4.02)	3.00(3.22)
Min	1.55(1.74)	1.64(1.97)	1.57(1.80)	1.61(1.94)	2.39(2.65)	2.76(2.75)	1.83(1.97)
Max	3.52(3.59)	3.91(4.15)	3.48(4.58)	3.83(4.27)	4.47(5.14)	5.03(5.16)	4.11(4.31)

NOTE: The corresponding statistics from IGLS estimation of the same model are given in parentheses.

casts made using the Bayesian predictive mean as a point forecast prove marginally superior. To demonstrate the advantage of using a temperature input, we reestimated the simple dummy variable model without the parametric temperature term and with $k_{\max} = 6$, so that $\mathbf{B}$ captures the serial correlation due to the omitted temperature variable. The corresponding point forecasts are 25% to 30% degraded compared with the model with the parametric temperature effect.

Utilities, traders, and system management are also interested in forecasting the time and level of the daily load peak. Spot prices are usually at their maximum at the peak and, given daily supply constraints, if peak load exceeds a certain level, then there is a high probability that prices will spike. Figure 6(a) provides the 1-day-ahead predictive forecast of load for Monday, February 23, 1998. The figure also contains a histogram of the predictive distribution of the time of the daily peak, with a vertical line marking the actual peak. Between 9:00 and 17:30 during the summer, load can plateau at near the daily peak; the distribution of time of peak reflects this uncertainty, with two modes at 10:00 and 14:00. In comparison, in winter the predictive distribution of the daily peak is more tightly concentrated round 18:00. Seasonal factors in the distribution of the level of the daily peak are also strong. Figure 6(b) plots density estimates of the 1-day-ahead predictive distribution of the level of the daily peak for four separate days in the year. These days fall in the middle of each of the four NSW seasons, but the distributions are all skewed to the right. Note that methods that do not explore the full predictive density of time of peak and peak level, but are based around symmetric approximations, would not provide adequate prediction inference.

6. CONCLUSION

In this article we have discussed how a multiequation regression model, coupled with estimation using MCMC, can be used as a powerful modeling and forecasting method for intraday electricity load. A particular advantage of the Bayesian approach is that the full spectrum of finite-sample inference, including the entire predictive distribution of load and time and level of daily peak, can be obtained. This is of particular importance for maintaining system stability, scheduling efficient generating capacity, making load-shedding decisions, and energy trading in wholesale electricity markets. Another advantage is that Bayesian model selection methodology can be used to explore the form of an intraday correlation structure and allow for selection between various subsets of regressors simultaneously for all intraday regressions. In comparison, direct maximum likelihood estimation of such models is difficult because of the high dimension of the parameter space. Nevertheless, feasible generalized least squares estimators are easy to implement, and point forecasts of future load can be obtained. However, it is difficult to obtain full predictive distributions for load and peak load, which are important in load forecasting practice.

When the Bayesian method is applied to 3 years of recent half-hourly NSW total system load data, it is possible to decompose load into meteorological, seasonal, trend, and day-type effects, along with a stationary diagonal first-order VAR error process. All parameters and components of the model vary over the half hour, confirming a multiequation approach as appropriate. The bivariate meteorological effects are statistically significant throughout the day, but with functional forms that vary depending on the time of day. Forecasts with an average MAPE

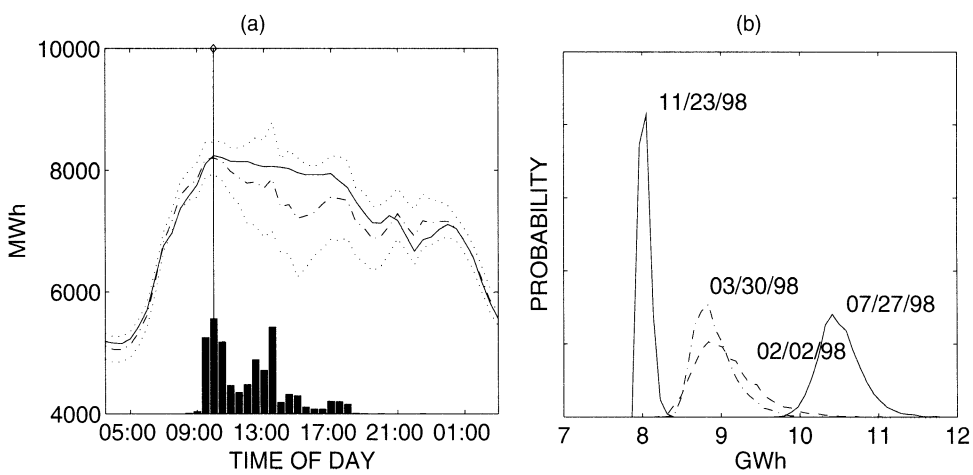

Figure 6. One-Day-Ahead Predictive Distributions of (a) Load and Time of Peak for Monday, February 23 (· · · ·, 90CiLow; -·-·, estimate; · · · ·, 90CiHigh; —, true) and (b) Level of Daily Peak for Four Mondays in 1998.

of 3.1% over a 6-month horizon can be obtained from a model without meteorological inputs estimated from 3 years of data. Estimating this model takes approximately 20 hours on a contemporary CPU. However, short-term models based on 3 weeks of data can be estimated in approximately 10 minutes, so that rapidly updated forecasts can be obtained. Daily updated short-term forecasts on running windows of 3 weeks of data show that including a temperature component can greatly improve forecasts.

APPENDIX: THE FOUR METROPOLIS–HASTINGS STEPS USED IN THE SAMPLING SCHEME IN SECTION 4.2

Generating From $\eta_i, \delta_i|\eta_{j\neq i}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}, \mathbf{y}$

Generating the pair (η_i, δ_i) is done by generating first from the flat distribution $q_2(\delta_i) = 1/M$ and then from $q_1(\eta_i|\delta_i)$. To develop a proposal density q_1, first note that, following in Judge et al. (1985, p. 488), there is a simple transformation matrix $\mathbf{H}(\boldsymbol{\Phi})$, such that if $\mathbf{u}^-$ is the vector of errors, $u_{i,t}$, omitting the first observation, $\mathbf{u}_{(1)} = (u_{1,1}, u_{2,1}, \ldots, u_{m,1})'$, then $\mathbf{u}^\star = \mathbf{H}\mathbf{u}^- \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_{T-1})$. Similarly, let $\mathbf{y}^-, \mathbf{X}^-$ be $\mathbf{y}, \mathbf{X}$ with the observations at the first time period omitted and $\mathbf{y}^\star = \mathbf{H}\mathbf{y}^-, \mathbf{X}^\star = \mathbf{H}\mathbf{X}^-$. Then the proposal density is based on the transformed conditional likelihood, so that

$$q_1(\eta_i|\delta_i = k) \propto \exp\left\{-\frac{1}{2}\left[\left(\mathbf{y}^\star - \mathbf{X}_i^\star\eta_i - \sum_{j\neq i}\mathbf{X}_j^\star\eta_j\right)' \times (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_{T-1})\left(\mathbf{y}^\star - \mathbf{X}_i^\star\eta_i - \sum_{j\neq i}\mathbf{X}_j^\star\eta_j\right) + (\eta_i - \boldsymbol{\mu}_{ik}^0)'(\mathbf{V}_{ik}^0)^{-1}(\eta_i - \boldsymbol{\mu}_{ik}^0)\right]\right\},$$

where $\mathbf{X}_j^\star$ is a matrix comprising the columns of $\mathbf{X}^\star$ that correspond to the subvector η_j. This can be recognized as a normal density in η_i from which a new iterate, η_i^n, can be generated. This is accepted over the old iterate η_i^o with probability $\min(1, \alpha)$, with

$$\alpha = \frac{p(\eta_i^n, \delta_i^n|\eta_{j\neq i}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}, \mathbf{y})q_1(\eta_i^o|\delta_i^o)}{p(\eta_i^o, \delta_i^o|\eta_{j\neq i}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}, \mathbf{y})q_1(\eta_i^n|\delta_i^n)}.$$

Generating From $\phi_i|\phi_{j\neq i}, \boldsymbol{\Sigma}, \eta, \mathbf{y}$

$$p(\phi_i|\boldsymbol{\Sigma}, \phi_{j\neq i}, \eta, \mathbf{y}) \propto\propto p\big(\mathbf{u}_{(1)}|\boldsymbol{\Phi}, \boldsymbol{\Sigma}\big)\prod_{t=2}^{T} p\big(\mathbf{u}_{(t)}|\mathbf{u}_{(t-1)}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}\big) \times p(\mathbf{B}|\boldsymbol{\Phi}, \mathbf{D}, \eta, \mathbf{k})\mathcal{I}(-1 < \phi_i < 1),$$

which follows from the prior for ϕ_i and (1). Setting the proposal density $q(\phi_i)$ proportional to the conditional likelihood constrained to the stationarity region, we have

$$q(\phi_i) \propto \exp\left(-\frac{1}{2}\sum_{t=2}^{T}\big(\mathbf{u}_{(t)} - \boldsymbol{\Phi}\mathbf{u}_{(t-1)}\big)'\boldsymbol{\Sigma}^{-1}\big(\mathbf{u}_{(t)} - \boldsymbol{\Phi}\mathbf{u}_{(t-1)}\big)\right) \times \mathcal{I}(-1 < \phi_i < 1),$$

which follows from (2). This density is univariate normal constrained to $(-1, 1)$ with variance $\tau_i^2 = (\sum_{t=2}^{T} u_{i,t-1}^2\sigma^{ii})^{-1}$ and mean

$$\mu_i = \tau_i^2\left[\sum_{t=2}^{T}\left\{u_{i,t-1}\left(\sum_{j=1}^{m}\sigma^{ij}u_{j,t} - \sum_{j\neq i}u_{j,t-1}\phi_j\sigma^{ij}\right)\right\}\right],$$

where $\boldsymbol{\Sigma}^{-1} = \{\sigma^{ij}\}$. An iterate, ϕ_i^n, is generated from the proposal and accepted over the old iterate, ϕ_i^o, with probability $\min\{1, \alpha\}$, where

$$\alpha = \frac{p(\mathbf{u}_{(1)}|\phi_i^n, \phi_{j\neq i}, \boldsymbol{\Sigma})p(\mathbf{B}|\mathbf{k}, \phi_i^n, \phi_{j\neq i}, \mathbf{D}, \eta)}{p(\mathbf{u}_{(1)}|\phi_i^o, \phi_{j\neq i}, \boldsymbol{\Sigma})p(\mathbf{B}|\mathbf{k}, \phi_i^o, \phi_{j\neq i}, \mathbf{D}, \eta)}.$$

Generating From $\mathbf{D}|\mathbf{B}, \boldsymbol{\Phi}, \eta, \mathbf{y}$

$$p(\mathbf{D}|\mathbf{B}, \boldsymbol{\Phi}, \eta, \mathbf{y}) \propto p\big(\mathbf{u}_{(1)}|\boldsymbol{\Sigma}, \boldsymbol{\Phi}\big)\prod_{t=2}^{T} p\big(\mathbf{u}_{(t)}|\mathbf{u}_{(t-1)}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}\big) \times \prod_{i=1}^{m-1} p(\boldsymbol{\beta}_i|\mathbf{D}, \boldsymbol{\Phi}, \eta, \mathbf{k})p(\mathbf{D}).$$

This density is unrecognizable in $\mathbf{D}$ due only to the density of the initial observation, so that it is well approximated by $q(\mathbf{D}) \propto \prod_{t=2}^{T} p(\mathbf{u}_{(t)}|\mathbf{u}_{(t-1)}, \boldsymbol{\Phi}, \boldsymbol{\Sigma})\prod_{i=1}^{m-1} p(\boldsymbol{\beta}_i|\mathbf{D}, \boldsymbol{\Phi}, \eta, \mathbf{k})p(\mathbf{D})$. Up to proportionality, this can be recognized as the product of independent gamma(α_i, θ_i) densities in $d_1, \ldots, d_m$, where

$$\alpha_i = \begin{cases}(T - 1 + \tilde{k}_i)/2 + \zeta/\kappa & \text{if } i < m\\ (T-1)/2 + \zeta/\kappa & \text{if } i = m,\end{cases}$$

$$\theta_i = \begin{cases}[(\mathbf{b}_i'\mathbf{A}\mathbf{b}_i + \mathbf{h}_i)/2 + 1/\kappa]^{-1} & \text{if } i < m\\ [\mathbf{b}_i'\mathbf{A}\mathbf{b}_i/2 + 1/\kappa]^{-1} & \text{if } i = m,\end{cases}$$

$\tilde{k}_i = \min(k_i, m - i)$, and $\mathbf{h}_i = \frac{1}{c}(\boldsymbol{\beta}_i + \mathbf{A}_i^{-1}\mathbf{a}_i)'\mathbf{A}_i(\boldsymbol{\beta}_i + \mathbf{A}_i^{-1}\mathbf{a}_i)$. After generating a new iterate, $\mathbf{D}^n$, from the proposal $q(\cdot)$, this is accepted over the old iterate, $\mathbf{D}^o$, with probability $\min\{1, p(\mathbf{u}_{(1)}|\boldsymbol{\Phi}, \mathbf{B}, \mathbf{D}^n)/p(\mathbf{u}_{(1)}|\boldsymbol{\Phi}, \mathbf{B}, \mathbf{D}^o)\}$.

Generating From $\boldsymbol{\beta}_i, \mathsf{k}_i|\boldsymbol{\beta}_{j\neq i}, \mathbf{D}, \boldsymbol{\Phi}, \eta, \mathbf{y}$

Generating the pair $(\boldsymbol{\beta}_i, k_i)$ is done by generating first from the flat distribution $q_2(k_i) = 1/\min(m - i, k_{\max})$ and then from $q_1(\boldsymbol{\beta}_i|k_i)$. The exact conditional posterior is

$$p(\boldsymbol{\beta}_i|\boldsymbol{\beta}_{j\neq i}, k_i, \mathbf{D}, \boldsymbol{\Phi}, \mathbf{u}) \propto p\big(\mathbf{u}_{(1)}|\boldsymbol{\Sigma}, \boldsymbol{\Phi}\big)p\big(\mathbf{u}_{(2)}, \ldots, \mathbf{u}_{(T)}\big|\mathbf{u}_{(1)}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}\big)p(\boldsymbol{\beta}_i|\mathbf{D}, \boldsymbol{\Phi}, \eta, k_i).$$

The density q_1 is obtained by omitting the density of the first observation, so that

$$q_1(\boldsymbol{\beta}_i|k_i) \propto p\big(\mathbf{u}_{(2)}, \ldots, \mathbf{u}_{(T)}\big|\mathbf{u}_{(1)}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}\big)p(\boldsymbol{\beta}_i|\mathbf{D}, \boldsymbol{\Phi}, \eta, k_i) \propto \prod_{i=1}^{m-1}\exp\left\{-\frac{d_i}{2}(2\boldsymbol{\beta}_i'\mathbf{a}_i + \boldsymbol{\beta}_i'\mathbf{A}_i\boldsymbol{\beta}_i)\right\} \times \exp\left\{-\frac{d_i}{2c}(\boldsymbol{\beta}_i + \mathbf{A}_i^{-1}\mathbf{a}_i)'\mathbf{A}_i(\boldsymbol{\beta}_i + \mathbf{A}_i^{-1}\mathbf{a}_i)\right\},$$

which follows from (4) and the fractional prior for $\mathbf{B}$ in Section 4.1. Examining this density, $\boldsymbol{\beta}_i|k_i \sim \mathrm{N}(-\mathbf{A}_i^{-1}\mathbf{a}_i, \frac{c}{d_i(1+c)}\mathbf{A}_i^{-1})$, so that generating an iterate of $\boldsymbol{\beta}_i$ is straightforward. Once a new iterate, $(k_i^n, \boldsymbol{\beta}_i^n)$, is generated, it is accepted over the old iterate, $(k_i^o, \boldsymbol{\beta}_i^o)$, with probability $\min(1, \alpha)$, where

$$\alpha = \frac{p(k_i^n, \boldsymbol{\beta}_i^n|\boldsymbol{\beta}_{j\neq i}, \mathbf{D}, \boldsymbol{\Phi}, \mathbf{u})q_2(k_i^o)q_1(\boldsymbol{\beta}_i^o|k_i^o)}{p(k_i^o, \boldsymbol{\beta}_i^o|\boldsymbol{\beta}_{j\neq i}, \mathbf{D}, \boldsymbol{\Phi}, \mathbf{u})q_2(k_i^n)q_1(\boldsymbol{\beta}_i^n|k_i^n)}$$

and $p(k_i, \boldsymbol{\beta}_i|\boldsymbol{\beta}_{j\neq i}, \mathbf{D}, \boldsymbol{\Phi}, \mathbf{u}) \propto p(\mathbf{u}|\boldsymbol{\Sigma}, \boldsymbol{\Phi})p(\boldsymbol{\beta}_i|\mathbf{D}, \boldsymbol{\Phi}, \eta, k_i)$.

[Received October 2001. Revised September 2003.]

REFERENCES

Al-Zayer, J., and Al-Ibrahim, A. (1996), "Modelling the Impact of Temperature on Electricity Consumption in the Eastern Province of Saudi Arabia," *Journal of Forecasting*, 15, 97–106.

Bunn, D., and Farmer, E. (1985), *Comparative Models for Electrical Load Forecasting*, New York: Wiley.

Darbellay, G., and Slama, M. (2000), "Forecasting the Short-Term Demand for Electricity: Do Neural Networks Stand a Better Chance?" *International Journal of Forecasting*, 16, 71–83.

Engle, R., Granger, C., and Hallman, J. (1989), "Merging Short- and Long-Run Forecasts: An Application of Seasonal Cointegration to Monthly Electricity Sales Forecasting," *Journal of Econometrics*, 40, 45–62.

Engle, R., Granger, W., Rice, J., and Weiss, A. (1986), "Semiparametric Estimates of the Relationship Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81, 310–320.

Fiebig, D., Bartels, R., and Aigner, D. (1991), "A Random Coefficient Approach to the Estimation of Residential End-Use Load Profiles," *Journal of Econometrics*, 50, 297–327.

Harvey, A., and Koopman, S. (1993), "Forecasting Hourly Electricity Demand Using Time-Varying Splines," *Journal of the American Statistical Association*, 88, 1228–1253.

Ho, K., Hsu, Y., and Yang, C. (1992), "Short-Term Load Forecasting Using a Multilayer Neural Network With an Adaptive Learning Algorithm," *IEEE Transactions on Power Systems*, 7, 141–149.

Hsu, Y., and Yang, C. (1995), "Electrical Load Forecasting," in *Applications of Neural Networks*, ed. A. Murray, Boston: Kluwer, pp. 157–189.

Judge, G., Griffiths, W., Hill, R., Lutkepohl, H., and Lee, T. (1985), *The Theory and Practice of Econometrics* (2nd ed.), New York: Wiley.

Ku, A. (2002), "The Art of Forecasting Demand," *Global Energy Business*, March/April, 21–23.

Lee, K., Cha, Y., and Park, J. (1992), "Short-Term Load Forecasting Using an Artificial Neural Network," *IEEE Transactions on Power Systems*, 7, 124–132.

Moghram, I., and Rahman, S. (1989), "Analysis and Evaluation of Five Short-Term Load Forecasting Techniques," *IEEE Transactions on Power Systems*, 4, 1484–1491.

Peirson, J., and Henley, A. (1994), "Electricity Load and Temperature," *Energy Economics*, 16, 235–243.

Peng, T., Hubele, N., and Karady, G. (1992), "Advancement in the Application of Neural Networks for Short-Term Load Forecasting," *IEEE Transactions on Power Systems*, 7, 250–257.

Pourahmadi, M. (1999), "Joint Mean-Covariance Models With Applications to Longitudinal Data: Uncontsrained Parameterisation," *Biometrika*, 86, 677–690.

Powell, M. (1987), "Radial Basis Functions for Multivariate Interpolation: A Review," in *Algorithms for Approximation*, eds. J. Mason and M. Cox, New York: Oxford University Press, pp. 143–167.

Ramanathan, R., Engle, R., Granger, C., Vahid-Araghi, F., and Brace, C. (1997), "Short-Run Forecasts of Electricity Loads and Peaks," *International Journal of Forecasting*, 13, 161–174.

Ranaweera, D., Hubele, N., and Papalexopoulos, A. (1995), "Application of Radial Basis Function Neural-Network Model for Short-Term Load Forecasting," *IEEE Proceedings—Generation, Transmission and Distribution*, 142.1, 45–50.

Shively, T., and Sager, T. (1999), "A Semiparametric Regression Approach to Adjusting for Meteorological Variables in Air Pollution Trends," *Environmental Science and Technology*, 33, 3873–3880.

Smith, M. (2000), "Modeling and Short-Term Forecasting of New South Wales Electricity System Load," *Journal of Business & Economic Statistics*, 18, 465–478.

Smith, M., and Kohn, R. (2002), "Parsimonious Covariance Matrix Estimation for Longitudinal Data," *Journal of the American Statistical Association*, 97, 1141–1153.