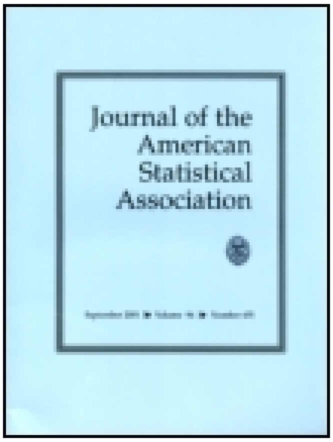

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Forecasting Hourly Electricity Demand Using Time-Varying Splines

Andrew Harvey [a] & Siem Jan Koopman [a]
[a] Department of Statistics, London School of Economics, London, WC2A 2AE, UK
Published online: 27 Feb 2012.

Forecasting Hourly Electricity Demand Using Time-Varying Splines

Andrew HARVEY and Siem Jan KOOPMAN*

A method for modeling a changing periodic pattern is developed. The use of time-varying splines enables this to be done relatively parsimoniously. The method is applied in a model used to forecast hourly electricity demand, with the periodic movements being intradaily or intraweekly. The full model contains other components, including a temperature response, which is also modeled using splines.

KEY WORDS: Cubic splines; Forecasting; Kalman filter; Load curve; Nonlinear regression; Seasonality; Structural time series model.

1. INTRODUCTION

Periodic movements are an important feature of many time series. With monthly or quarterly observations, seasonal effects tend to be the rule rather than the exception and, for observations at more frequent intervals, there may be periodic movements within the month, the week or the day. A fixed periodic pattern over s time intervals can be modeled with $s - 1$ dummy variables or, equivalently, with $s - 1$ trigonometric terms. But two important issues now arise. The first is that the periodic pattern may evolve over time. The second is that if s is large, then the model starts to become very cumbersome. The aim of this article is to provide a solution to these problems by developing the idea of time-varying splines. These ideas are then applied to the problem of modeling intradaily effects, and the technique is illustrated with hourly data from the Puget Sound Power and Light electricity company based in the northwest United States. The ultimate aim is to provide a method of making reliable short-term forecasts.

The collection of papers in Bunn and Falmer (1985) gives some indication of the methods that have been used in the short-term forecasting of energy. The main approaches seem to be based on autoregressive integrated moving average (ARIMA) models, regression, exponential smoothing, or some mixture of these. We argue here that an approach based on structural time series models lends itself much more readily to the kind of problems encountered. Such models are formulated in terms of unobserved components that represent the salient features of the series (see Harvey 1989). Thus for modeling hourly data trend, seasonal, daily, and intradaily components may be included. As indicated in the opening paragraph, the main issue is how to cut down on the number of terms used to model the periodic effect within each day or week.

With quarterly or monthly data, it is quite reasonable to include the full complement of seasonal dummies or trigonometric terms in the seasonal component. With weekly data this starts to become more difficult computationally, and it may well be less sensible statistically. Cutting down on the number of dummy variables will not usually be a good option because of the discontinuities that it entails. Reducing the number of trigonometric terms is a more viable proposition, and for a slowly changing seasonal effect, such as average weekly temperature, it may be entirely satisfactory. This is less likely to be the case when the seasonal pattern exhibits sharp peaks, as, for example, typically happens with sales of certain consumer goods, which tend to be concentrated in the weeks immediately before Christmas. Thus we seek a formulation that is parsimonious but flexible enough to capture marked variations in the pattern while retaining a reasonable degree of continuity.

In the context of electricity demand, the intradaily pattern is known as the load curve. A parsimonious way of modeling the load curve is highly desirable for hourly observations and becomes even more important when observations are made every half or quarter hour. Splines have already been used successfully to model intradaily effects in electricity demand (see Hendricks, Koenker, and Poirier 1979). But it is important to allow such splines to evolve over time. The intradaily pattern may change over a period of several years due to new technology. It certainly changes within the year, as can be seen in Figure 1, which shows the hourly electricity load pattern for Puget Sound Power and Light for typical weeks in each quarter. As one might expect, there are marked changes over the year, with the intradaily variations in demand being much greater in the winter.

2. TIME-VARYING PERIODIC SPLINES

A univariate structural time series model can be regarded as a regression model in which the explanatory variables are functions of time and their parameters are time-varying. Thus the first step in obtaining a time-varying periodic spline is to show how spline models in general, and periodic splines in particular, can be set up in terms of regression. The treatment is based on Poirier (1973, 1976). We then introduce time variation by allowing the parameters to follow stochastic processes.

2.1 Piecewise Regression Using Cubic Splines

Suppose that we have n pairs of observations (x_j, y_j), $j = 1, \ldots, n$, and that we wish to set up a nonlinear regression

* Andrew Harvey is Professor of Econometrics and Siem Jan Koopman is Research Officer, Department of Statistics, London School of Economics, London WC2A 2AE, U.K. The authors thank EPRI for financial support, Casey Brace for supplying the data and encouraging them to undertake this research, and the editors and referees for constructive comments.

Journal of the American Statistical Association
December 1993, Vol. 88, No. 424, Applications & Case Studies

model of the form

$$y_j = f(x_j) + \varepsilon_j, \qquad j = 1, \ldots, n, \tag{2.1}$$

where the ε_j's are mutually uncorrelated disturbances with 0 mean and constant variance, σ^2. In a cubic spline regression model, $f(x_j)$ is constructed by putting together polynomials of degree at most 3 in such a way as to preserve continuity in second derivatives. The h individual cubics are joined at the coordinates $(x_i^\dagger, \gamma_i^\dagger)$, $i = 0, \ldots, h$. The set of x values, $x_0^\dagger < x_1^\dagger < \cdots < x_h^\dagger$, is known as a *mesh;* the $h + 1 \geq 3$ individual points are called *knots.* The setup is completed by making assumptions about the spline at its end points.

Given the knots and the associated values of the ordinates, $x_0^\dagger, \ldots, x_h^\dagger$, it can be shown that any point on the spline function is a linear combination of the $\gamma_i^\dagger$'s. Thus at the observation points, we can write

$$f(x_j) = \mathbf{w}_j'\boldsymbol{\gamma}^\dagger, \qquad j = 1, \ldots, n, \tag{2.2}$$

where $\mathbf{w}_j$ is an $(h + 1) \times 1$ vector that depends on the position of the knots and the distance between them, as well as on the observed value x_j and $\boldsymbol{\gamma}^\dagger = (\gamma_0^\dagger, \gamma_1^\dagger, \ldots, \gamma_h^\dagger)'$ (see Poirier 1973, p. 517). If x_j corresponds to a knot, $x_j = x_i^\dagger$, then all the elements in $\mathbf{w}_j$ are 0, apart from the ith which is unity, and $f(x_j) = \gamma_i^\dagger$.

Substituting (2.2) in (2.1) gives the cubic spline regression model

$$y_j = \mathbf{w}_j'\boldsymbol{\gamma}^\dagger + \varepsilon_j, \qquad j = 1, \ldots, n \tag{2.3}$$

Given the assumptions on ε_j, $\boldsymbol{\gamma}^\dagger$ is estimated by ordinary least squares (OLS).

2.2 Periodic Splines

Now suppose that the explanatory variable is time and that there is a pattern repeated over a stretch of s observations, so that we have s periodic effects γ_j, $j = 1, \ldots, s$. We can model these effects by a spline of the form (2.2), in which $n = s$, $x_j = j$ and continuity from one period to the next is preserved by the condition

$$\gamma_0^\dagger = \gamma_h^\dagger \tag{2.4}$$

and the conditions that the first and second derivatives at 0 and h are the same. This removes the need for further assumptions about the end conditions. The implications for the $\mathbf{w}_j$ vectors, which are now $h \times 1$, corresponding to $\gamma_1^\dagger, \ldots, \gamma_h^\dagger$, are easily worked out (see Poirier 1976, pp. 43–47). Therefore, we have a periodic spline

$$\gamma_j = \mathbf{w}_j'\boldsymbol{\gamma}^\dagger, \qquad j = 1, \ldots, s, \tag{2.5}$$

where $\boldsymbol{\gamma}^\dagger$ is now $h \times 1$. If we have a set of observations, y_t, $t = 1, \ldots, T$, then we might posit a model

$$y_t = \gamma_t + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{2.6a}$$

where

$$\gamma_t = \gamma_j \tag{2.6b}$$

where an effect of the jth kind is appropriate.

It is convenient to have periodic effects summing to 0 over a complete period. Thus in terms of (2.5), we require

$$\sum_{j=1}^{s} \gamma_j = \sum_{j=1}^{s} \mathbf{w}_j'\boldsymbol{\gamma}^\dagger = \mathbf{w}_*'\boldsymbol{\gamma}^\dagger = 0, \tag{2.7}$$

where $\mathbf{w}_*$ is the $h \times 1$ vector

$$\mathbf{w}_* = \sum_{j=1}^{s} \mathbf{w}_j. \tag{2.8}$$

The restriction can be enforced by arbitrarily dropping one of the elements of $\boldsymbol{\gamma}^\dagger$. If $\gamma_h^\dagger$ is dropped, then substituting

$$\gamma_h^\dagger = -\sum_{i=1}^{h-1} (w_{*i}/w_{*h})\gamma_i^\dagger, \tag{2.9}$$

where w_{*i} is the ith element of $\mathbf{w}_*$, in (2.5) gives

$$\gamma_j = \sum_{i=1}^{h-1} (w_{ji} - w_{jh}w_{*i}/w_{*h})\gamma_i^\dagger, \qquad j = 1, \ldots, s, \tag{2.10}$$

where w_{ji} is the ith element of vector $\mathbf{w}_j$. In vector notation,

$$\gamma_j = \mathbf{z}_j'\boldsymbol{\gamma}^*, \tag{2.11a}$$

where the ith element of $\mathbf{z}_j$ is the bracketed expression in (2.10) and

$$\boldsymbol{\gamma}^* = (\gamma_1^\dagger, \ldots, \gamma_{h-1}^\dagger)'. \tag{2.11b}$$

2.3 Time-Varying Periodic Effects

As in the previous subsection, let $\boldsymbol{\gamma}$ be an $s \times 1$ vector of periodic effects, but suppose that these effects now evolve over time. However, at any point in time they are still subject to the zero-sum restriction, so that

$$\sum_{j=1}^{s} \gamma_{jt} = \mathbf{i}'\boldsymbol{\gamma}_t = 0, \tag{2.12}$$

where $\mathbf{i}$ is a vector of 1s. If no further restrictions are put on $\boldsymbol{\gamma}_t$, time variation may be introduced by letting $\boldsymbol{\gamma}_t$ follow a multivariate random walk,

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_{t-1} + \boldsymbol{\chi}_t, \tag{2.13}$$

where $\boldsymbol{\chi}_t$ is an $s \times 1$ vector of serially uncorrelated disturbances with mean 0 and covariance matrix

$$E(\boldsymbol{\chi}_t\boldsymbol{\chi}_t') = \sigma_\chi^2\left[\mathbf{I} - \frac{1}{s}\mathbf{i}\mathbf{i}'\right]. \tag{2.14}$$

The restriction (2.12) implies a similar restriction on $\boldsymbol{\chi}_t$ (i.e., $\mathbf{i}'\boldsymbol{\chi}_t = 0$), and this is enforced by (2.14). Of course, (2.12) implies that one element can be dropped from $\boldsymbol{\gamma}_t$. The time-varying periodic effect then enters into a state space model by including $s - 1$ elements of $\boldsymbol{\gamma}_t$ in the state vector and letting the measurement equation pick out the appropriate periodic effect at time t. Harrison and Stevens (1976) have modeled seasonality in this way.

Now suppose that we wish to model periodic effects using less than $s - 1$ parameters. To do this, suppose that $\boldsymbol{\gamma}_t$ is related to $h \leq s - 1$ time-varying parameters in a vector $\boldsymbol{\gamma}_t^\dagger$. Thus

$$\gamma_{jt} = \mathbf{w}_j'\boldsymbol{\gamma}_t^\dagger, \qquad j = 1, \ldots, s, \qquad t = 1, \ldots, T, \tag{2.15}$$

where $\mathbf{w}_j$ is a known $h \times 1$ vector of spline weights and γ_{jt} is the jth seasonal effect at time t. The zero-sum constraint, (2.12), implies a constraint on $\gamma_t^\dagger$ that is of exactly the same form as in (2.7). Furthermore, if $\gamma_t^\dagger$ evolves according to

$$\gamma_t^\dagger = \gamma_{t-1}^\dagger + \chi_t^\dagger, \tag{2.16}$$

where $\chi_t^\dagger$ is an $h \times 1$ vector of serially uncorrelated random disturbances with 0 mean, the implied constraint on these disturbances,

$$\mathbf{w}_*' \chi_t^\dagger = 0, \tag{2.17}$$

is reflected in the covariance matrix

$$E(\chi_t^\dagger \chi_t^{\dagger\prime}) = \sigma_\chi^2 \left[\mathbf{I} - \frac{1}{\mathbf{w}_*'\mathbf{w}_*} \mathbf{w}_* \mathbf{w}_*' \right], \tag{2.18}$$

where σ_χ^2 is a nonnegative scalar.

As before, one element can be dropped from $\gamma_t^\dagger$. The choice is arbitrary, but if the hth is dropped, the elements of γ_t can be obtained from the first $(h - 1)$ elements of the vector $\gamma_t^\dagger$ *via* an expression of the form (2.10). For a model of the form (2.6), the state space formulation is straightforward with the transition equation as in (2.16), with one element dropped and the measurement equation

$$y_t = \mathbf{z}_t' \gamma_t^* + \varepsilon_t, \tag{2.19}$$

where $\mathbf{z}_t$ is $\mathbf{z}_j$ of (2.11a) for a type j periodic effect at time t and γ_t^* is the time-varying version of γ^* in (2.11).

As the notation implies, the stochastic periodic model of (2.15) and (2.16) applies directly to the periodic cubic spline of the previous subsection. In the special case when $h = s$, the $s \times s$ matrix $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_s)'$ is the identity matrix, and the model reduces to the Harrison–Stevens seasonal model, (2.13) and (2.14), with $\gamma_t = \gamma_t^\dagger$.

3. INTRADAILY EFFECTS

Intradaily effects change over the year with the seasons. On the assumption that these patterns are repeated each year, one approach would be to estimate the intradaily effect at different times of the year by constructing a model that effectively allows one to average over past years. But such an approach is cumbersome to implement and may not be very satisfactory when there are only a few years of data.

Our preferred approach is to allow the intradaily effect to accommodate seasonal and other changes by a slow movement over time. The time-varying spline technique enables this to be done with a relatively small number of parameters. With hourly data it is important to economize on the number of parameters, because without any restrictions, a stochastic intradaily effect will contribute 23 elements to the state vector. Obviously this problem becomes more acute if observations are available at more frequent intervals. It also becomes more acute if different days display different intradaily patterns. For example, if Saturday and Sunday are both different than weekdays, 69 state elements are required with hourly data.

This section describes the modelling of intradaily effects in the light of our experience with hourly electricity demand.

3.1 Standard Model

The standard intradaily model is as described in Subsection 2.3. The main practical problem is determining the position of the knots. Generally speaking, sections of the periodic pattern displaying sharp peaks require relatively more knots than do less variable sections. For the Puget data, the main peaks are around breakfast time and, to a lesser extent, in the early evening. In our preferred model we placed knots at 3 AM, 6 AM, 7 AM, 8 AM, 9 AM, 11 AM, 2 PM, 5 PM, 6 PM, 8 PM, 11 PM, and midnight (cf. Hendricks et al. 1979). This requires 11 parameters, representing a considerable savings over the 23 needed with an unrestricted model. Furthermore, our spline gave a better fit than a trigonometric model, with the same number of parameters. Nevertheless it must be stressed that a certain amount of experimentation was needed to determine a good mesh. There is really no systematic way of going about this problem, although the starting point is to form an idea of the intradaily pattern from prior knowledge, an examination of unrestricted estimates, or a graph like that in Figure 1.

3.2 Correction Factors for Nonstandard Days

Unfortunately, the same intradaily model will not normally apply to all days of the week. In particular, Saturdays and Sundays may be different than weekdays, and in the case of electricity demand there is no doubt that they are different. One way of handling this problem is to set up a time-varying spline to give an intradaily correction factor to atypical days. Continuity is enforced by setting up splines that are constrained to be 0 at the beginning and end of the day. This constraint is reflected in the $\mathbf{w}_j$ vectors and replaces the zero-sum constraint that applies to the intradaily effect for typical days. The salient feature of the correction factor for Saturday is that the morning peak is less pronounced and occurs an hour or so later. This is exactly what we would expect, given our prior knowledge of behavior on Saturday mornings.

Using a correction factor is likely to be particularly appealing when the difference between the intradaily pattern and the standard intradaily pattern is relatively smooth. This implies that less knots are needed as compared to a fully intradaily model. The same argument also leads one to consider modeling Sunday by a Saturday correction factor plus a further correction factor, again constrained to be 0 at its end points, for the difference between Saturday and Sunday. In the case of our electricity demand study, we were able to base an additional Sunday correction factor on only eight knots, whereas Saturday required ten knots.

The use of correction factors means that the intradaily effects do not sum to 0 over a nonstandard day. Instead, the sum of the correction factors gives the total amount by which the day in question differs from a nonstandard day.

There are two disadvantages to using correction factors. The first is that they lead to discontinuity in first and second derivatives at the point where they join the standard intradaily spline. The second is that when one nonstandard day follows another, as with Sunday and Saturday, the 0 end point constraints imply that the intradaily effect at the point

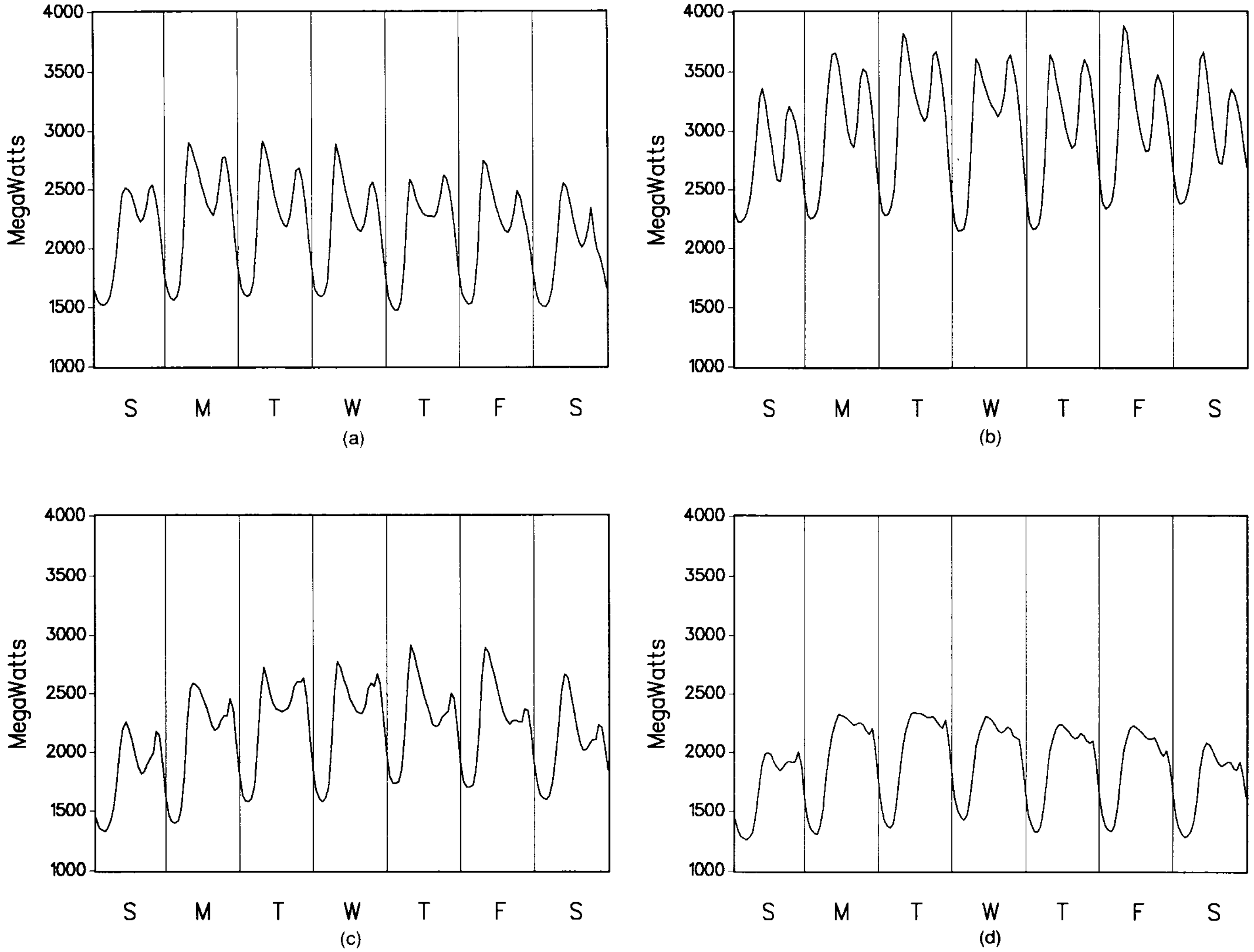

Figure 1. Weekly Load Patterns in (a) Autumn 1990; (b) Winter 1991; (c) Spring 1991; and (d) Summer 1991.

where the correction factors meet should be as for a standard day. This second problem may be solved by having one single correction factor for the whole weekend (which may include Friday afternoon and Monday morning). But in the electricity demand application, the problem is not particularly important in practice when the join point is in the early hours of the morning. Similarly, the discontinuities in first and second derivatives are not very important when demand is low.

3.3 Weekly Model

Another way of dealing with the problem of different patterns within different days is to set up a periodic spline for the whole week. The zero-sum constraint is now imposed over the whole week, and with hourly observations, s is 168. This approach does not suffer from the disadvantages noted at the end of the previous subsection, but at first sight it would appear to be very inefficient if the patterns in some of the days are the same. Fortunately, it is possible to take this point into account. The vector $\gamma_t^\dagger$ contains sets of parameters corresponding to the knots in each of the 7 days. If some of the intradaily patterns are the same, then the corresponding elements in $\gamma_t^\dagger$ will be the same. Thus we can work with a parameter vector of reduced dimension that contains only the elements that are different. For example, we might have parameters for Saturdays, Sundays, and weekdays. If this new parameter vector is denoted by $\bar{\gamma}_t^\dagger$, then we may write

$$\gamma_t^\dagger = \mathbf{S}\bar{\gamma}_t^\dagger, \tag{3.1}$$

where $\mathbf{S}$ is a selection matrix of 0s and 1s. Now if $\mathbf{w}_j$ is defined as in Section 2, then the effect associated with the jth hour in the week is

$$\mathbf{w}_j'\mathbf{S}\bar{\gamma}_t^\dagger = \bar{\mathbf{w}}_j'\bar{\gamma}_t. \tag{3.2}$$

But because some of the $\bar{\mathbf{w}}_j$'s will be the same, corresponding to some of the elements in $\gamma_t^\dagger$ being the same, less than the full complement of s $\bar{\mathbf{w}}_j$'s need to be stored. Furthermore, although the matrix $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_s)'$ is quite large, it needs to be computed only once and then be postmultiplied by $\mathbf{S}$ to yield the $\bar{\mathbf{w}}_j$'s. The net effect of this modification is that days depending on the same elements in $\gamma_t^\dagger$ will be the same at the knots. This does not mean that the patterns in these days will be exactly the same, because the patterns will partly depend on adjacent days. Thus the realized pattern on Friday may be different than that on Wednesday because of the influence of Saturday.

4. A SHORT-TERM FORECASTING MODEL FOR ELECTRICITY DEMAND

Suppose that we have observations within the day and we wish to forecast up to 2 or 3 days ahead. We may set up a structural time series model of the form

$$y_t = \mu_t + \gamma_t + \sum_{i=1}^{k} \delta_i x_{it} + \varepsilon_t, \qquad t = 1, \ldots, T, \quad (4.1)$$

where μ_t denotes a stochastic trend component as in Harvey (1989, chap. 2); γ_t is the intradaily effect, modeled either with correction factors as in subsection 3.2 or as part of a weekly effect as in subsection 3.3; x_{it} is an observable exogenous variable; δ_i is its coefficient; and ε_t is a stationary disturbance term. In a Gaussian model, the disturbances driving the various components are assumed to be normally distributed, and the notation NID(0, σ^2) will denote normally and independently distributed with mean 0 and variance σ^2.

The model can be handled by putting it in state space form. The Kalman filter and associated recursive algorithms then provide the basis for updating, predicting, and smoothing. In addition, the Kalman filter may be used to construct the likelihood function. Maximization of the likelihood function yields estimators of the hyperparameters; that is, the variance parameters such as σ_χ^2.

The treatment of explanatory variables is quite straightforward if they enter linearly. In the context of short-term electricity demand, the explanatory variables are weather variables, such as temperature, wind speed, and humidity. The effects of these variables tend to be nonlinear (see Bunn and Falmer 1985). Because the functional form is not always known a priori, a flexible modeling procedure is needed. We follow Engle, Granger, Rice, and Weiss (1986) in using splines. Suppose that we have a single explanatory variable. Let the effect of this variable in (4.1) be denoted by δ_t, so that

$$y_t = \mu_t + \gamma_t + \delta_t + \varepsilon_t, \qquad t = 1, \ldots, T \quad (4.2)$$

and

$$\delta_t = \tilde{\mathbf{x}}_t'\boldsymbol{\delta}, \quad (4.3)$$

where $\tilde{\mathbf{x}}_t$ is a vector constructed in the same way as the $\mathbf{w}_j$ of (2.2) and $\boldsymbol{\delta}$ is a corresponding vector of unknown parameters. The knots are now at particular values of the explanatory variable, x_t. To avoid multicollinearity between the constant in the trend and the spline for the explanatory variable, a restriction must be introduced into (4.3). This can be done by setting one of the elements in $\boldsymbol{\delta}$ equal to 0. Thus if there are $h + 1$ knots in the spline, then there are only h unknown parameters. Alternatively, the initial value of the trend may be set to 0; that is, $\mu_0 = 0$.

There is no difficulty in principle in letting the parameters, $\boldsymbol{\delta}$, change over time according to a multivariate random walk. This can still be handled within the state space framework. But we did not find it necessary to pursue this possibility in the current application.

Public holidays can sometimes be treated as though they were Sundays; however, this is unlikely to be satisfactory for some holidays, such as Christmas. The best thing to do in such cases is probably to have a special intradaily correction factor that is separate from the rest of the model and is just estimated from past observations, in terms of deviations from the underlying level, on that particular holiday. Forecasts of future values can then be made by adding this correction factor to the predicted underlying level. The observed holiday values are probably best treated as missing observations, because to include them could easily introduce distortions into future estimates.

4.1 Specification of Model for Puget

Models of the kind described earlier were fitted to observations on hourly load, in megawatts, for Puget Sound Power and Light over the period of January 1, 1985 to September 9, 1991. The preferred model is of the form

$$y_t = \mu_t + \psi_t + \gamma_t + \delta_t + \varepsilon_t, \qquad t = 1, \ldots, T, \quad (4.4)$$

where μ_t is a random walk

$$\mu_t = \mu_{t-1} + \eta_t, \qquad \eta_t \sim \text{NID}(0, \sigma_\eta^2);$$

ψ_t is a deterministic cycle of period 1 year, or 365 × 24 hours, that is

$$\psi_t = \alpha \cos \lambda t + \beta \sin \lambda t,$$

where $\lambda = 2\pi/8{,}760$; γ_t is a time-varying weekly spline; δ_t is a temperature spline, as in (4.3); and ε_t is a random disturbance term, that is, $\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$. Taking logarithms to give a multiplicative model gave no improvement in forecasting performance.

The random walk is a special case of a stochastic trend component designed to pick up underlying changes in the level. The seasonal movement, ψ_t, is not primarily due to temperature, the effect of which is modeled directly, but rather arises from other seasonal changes, such as the change in the number of hours of daylight.

Regarding the weekly spline, the knots for a standard weekday were set as in subsection 3.1. There was some modification for Monday morning, with knots at 3 AM, 6 AM, 7

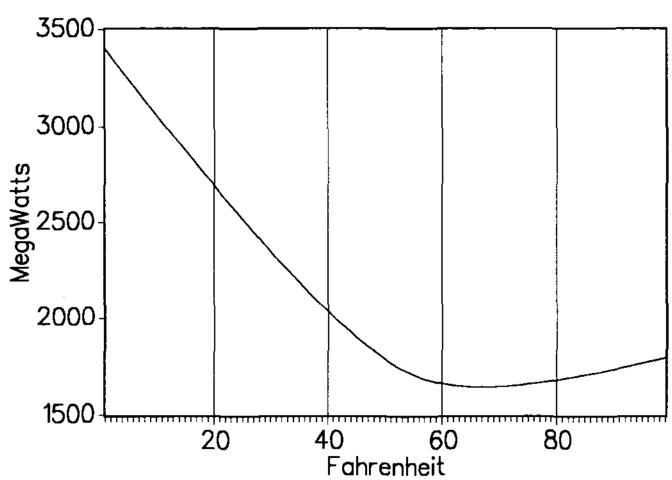

Figure 2. Estimated Temperature Response.

AM, 8 AM, and 9 AM, and Friday evening, with knots at 5 PM, 6 PM, 8 PM, 11 PM and midnight. For Saturdays and Sundays, the knots were chosen as follows:

Saturday: 3 AM 7 AM 8 AM 9 AM 11 AM 2 PM 6 PM 7 PM 9 PM 10 PM midnight

Sunday: 3 AM 6 AM 7 AM 8 AM 9 AM noon 4 PM 6 PM 8 PM 10 PM midnight

After allowing for the fact that some weekdays have the same knot parameters, and that there is a zero-sum constraint, the total number of distinct unknown parameters in $\gamma_t^\dagger$ is 43.

The temperature response was modeled by a natural cubic spline with knots at 0°, 40°, 65°, and 99° Fahrenheit. Other explanatory variables, such as humidity and wind speed, were included in preliminary versions of the model, but no specification in which they had a significant influence could be found. Lagged values of temperature were included in the manner suggested in Appendix A, but the main effect seemed to come from current temperature. Some evidence was found of different responses to temperature at different times of the day, but again the additional complexity of the model was not justified in terms of forecasting performance; see Appendix B.

4.2 Estimation Results

Estimation was carried out in the time domain via the prediction error decomposition. The relative hyperparameters, $q_\eta = \sigma_\eta^2/\sigma_\varepsilon^2$ and $q_\chi = \sigma_\chi^2/\sigma_\varepsilon^2$, were estimated to be .0002 and .0012. These hyperparameters give the extent to which the level of the series and the weekly pattern can change over time.

The temperature effect was highly significant. The shape of the temperature response is shown in Figure 2. The changing intraweekly pattern is shown in Figure 3. The contrast between winter and spring is particularly marked; compare Figure 1.

The model fitted better in the warmer months than in the winter. In particular, the prediction errors were much larger for temperatures below freezing. The reason is that the climate in Seattle is rather mild, and so there are relatively few data points for very cold days.

Minor modifications may be made to the model. The first is to generalize the irregular term, ε_t, so that it becomes a first-order autoregressive process. Given the estimated hyperparameters, an autoregressive parameter of .85 was found to maximize the likelihood function when a grid search was carried out. This procedure was adopted since attempting to maximize the likelihood function with respect to all the pa-

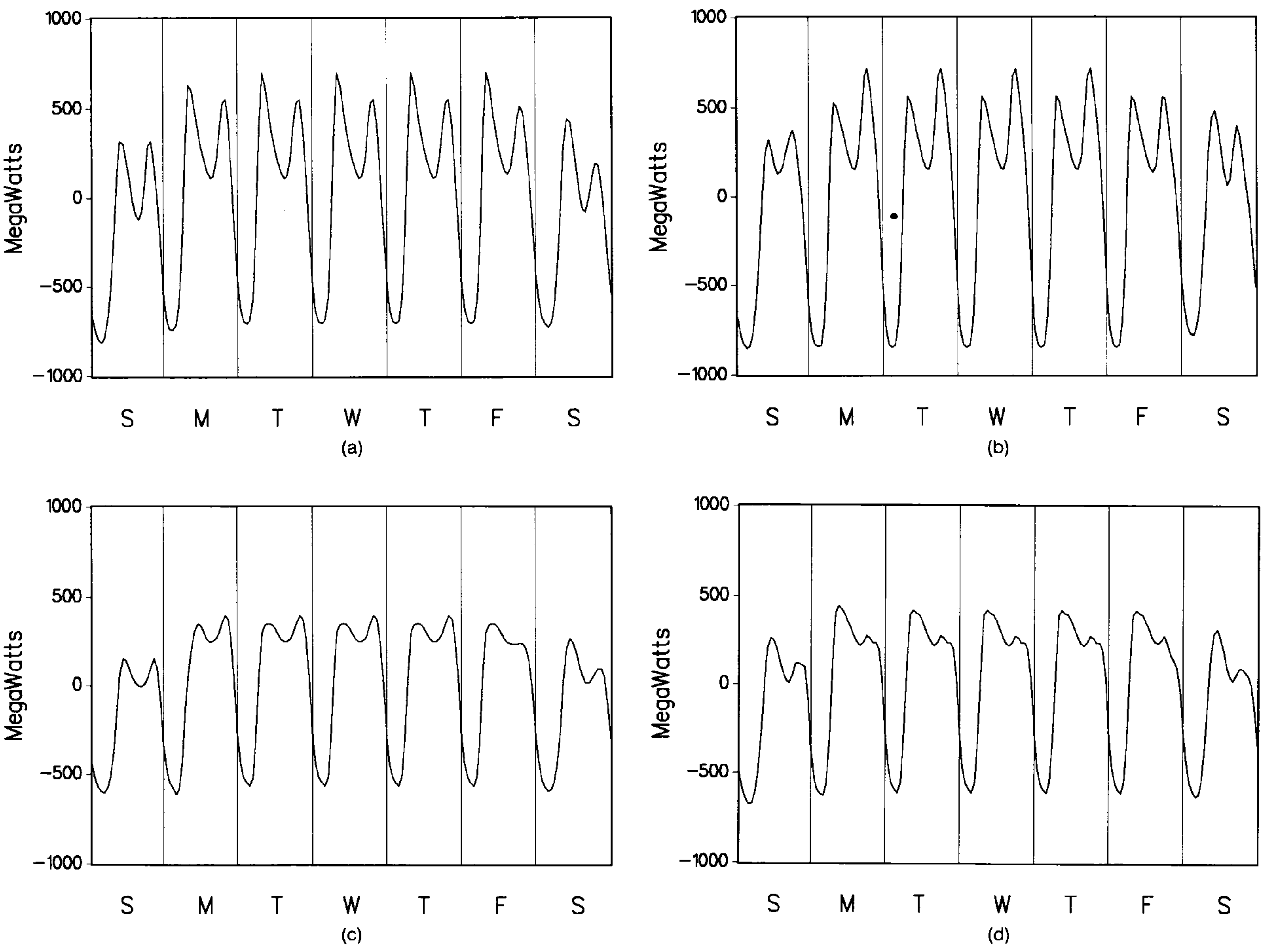

Figure 3. Estimated Intraweekly Splines in (a) Autumn; (b) Winter; (c) Spring; and (d) Summer.

Table 1. Forecasting Performance of SHELF Model in Terms of MAPE

	Mon.	*Tue.*	*Wed.*	*Thu.*	*Fri.*	*Sat.*	*Sun.*	*Weekday*	*Weekend*
a. Average	4.3	3.7	3.2	4.1	3.7	3.9	5.1	3.8	4.5
b. Min.	0.5	.4	.2	.8	.5	.9	1.0	.5	.9
c. Max.	10.5	8.7	8.2	9.8	8.9	8.5	10.8	9.2	9.6
d. AM	4.9	3.9	2.7	3.2	4.1	4.6	5.7	3.7	5.2
e. PM	4.7	4.6	4.2	4.7	4.2	3.9	5.9	4.5	4.9

rameters in the model led to computational difficulties. The autoregression removes most of the residual serial correlation, though it has little impact on the forecasting performance of the model. The second modification is to make more specific allowance for public holidays.

4.3 Forecasting

The electricity company makes forecasts at 9 AM on Monday through Thursday, for the next day, based on information up to 8 AM. On Friday morning forecasts are made for Saturday, Sunday, and Monday. For the period November 7, 1990 through March 31, 1991, the forecasting performance of our model—SHELF (Structural Hourly Electricity Load Forecaster)—is shown in Table 1. The following indicators, all based on the mean absolute percentage error (MAPE), are presented:

a. the average MAPE over all hours
b. the MAPE for the smallest error each day
c. the MAPE for the largest error each day
d. the average of the MAPE's for the peak morning hours of 7, 8, and 9 AM
e. the average of the MAPE's for the peak afternoon hours of 4, 5, and 6 PM.

Figures 4 and 5 are typical examples of forecasts made on Tuesday and Friday mornings. They show the absolute fore-

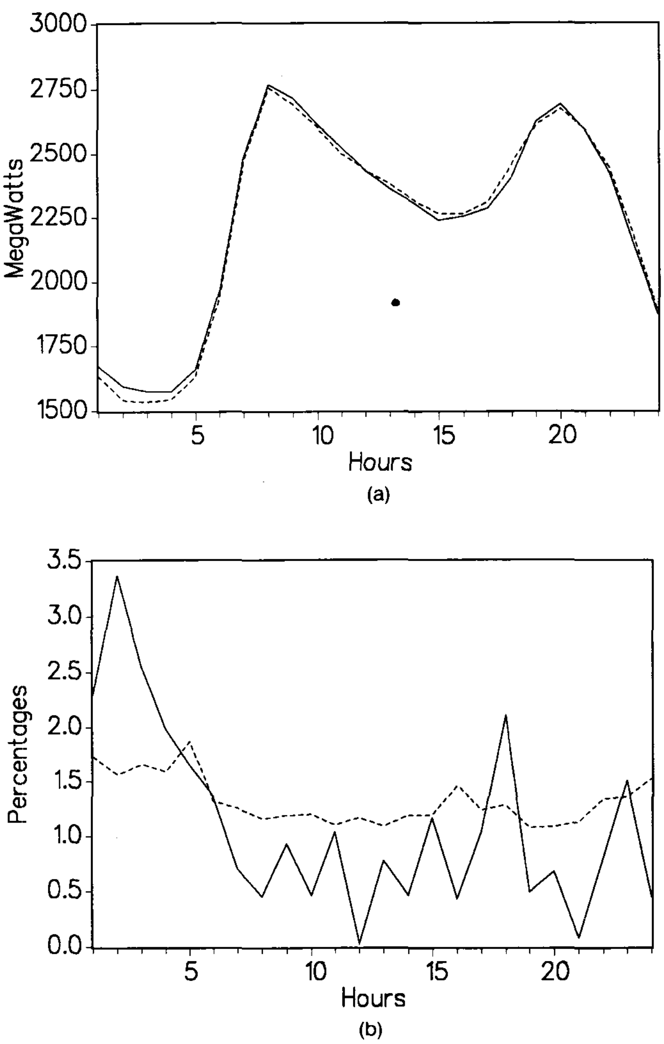

Figure 4. Weekday Forecasts. (a) Load forecasts on a typical Wednesday (——, actual; – – –, forecast). (b) Forecast error on a typical Wednesday (——, absolute % of forecast error; – – –, % RMSE).

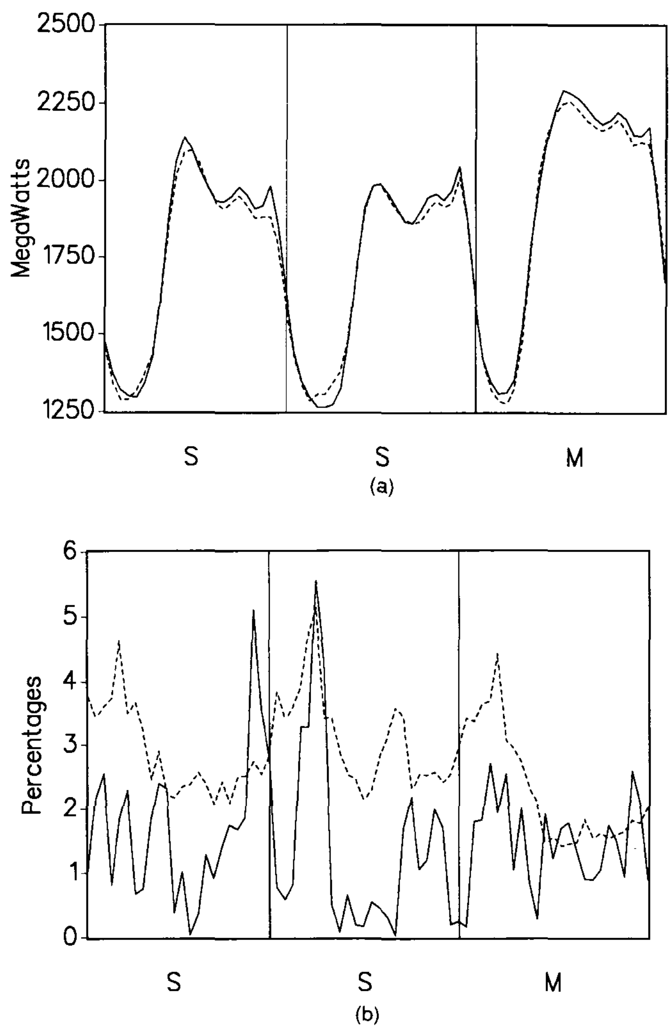

Figure 5. Weekend Forecasts. (a) Load forecasts on a typical Saturday, Sunday, and Monday (——, – – –, forecast). (b) Forecast error on a typical Saturday, Sunday, and Monday (——, absolute % of forecast error; – – –, % RMSE).

cast error expressed as a percentage of the corresponding observation. The broken line is the theoretical root mean square error (RMSE) as a percentage of the corresponding observation. This is calculated by the Kalman filter prediction equations, conditional on the estimated values of the hyperparameters. These RMSE's can be used to construct prediction intervals if desired.

To put the forecast results of Table 1 in perspective, a simple (naive) forecast equation was applied to the same data, and its forecast performance was evaluated for the same period. The forecast function is based on the set of 24 dynamic regression models

$$y_{i,t} = y_{i,t-7} + \beta_i(x_{i,t} - x_{i,t-7}) + \varepsilon_{i,t}, \quad i = 1, \ldots, 24, \quad (4.5)$$

where $y_{i,t}$, $x_{i,t}$, and $\varepsilon_{i,t}$ are the load demand, the temperature, and the Gaussian noise term at hour i and day t. The regression parameters, β_i, $i = 1, \ldots, 24$, are estimated by treating the model as a system of seemingly unrelated regression equations (SURE). Excluding public holidays, the average hourly MAPE's are 7.31 (Sunday), 5.60 (Monday), 5.22 (Tuesday), 5.39 (Wednesday), 5.70 (Thursday), 6.93 (Friday), 6.74 (Saturday), 5.77 (Weekday), and 7.03 (Weekend). It can be seen that the forecasts of model (4.4) are much more precise.

5. CONCLUSION

A time-varying periodic spline component appears to provide a good way of modeling the changing electricity load pattern within the week. The effect of the nonlinear response to temperature may be captured by a fixed spline, and the overall forecasts are relatively accurate.

APPENDIX A: DISTRIBUTED LAGS

Suppose that the nonlinear response to the explanatory variable is modeled by a spline, as in (4.3), but is subject to lags. We then have

$$\delta_t = \sum_{\tau=0}^{m} \tilde{\mathbf{x}}'_{t-\tau}\boldsymbol{\delta}_\tau. \quad (A.1)$$

Once the $\tilde{\mathbf{x}}_t$'s have been constructed, the treatment of this model is no different from that of (4.1). The only problem is the large number of parameters, arising from the fact that the parameters of the spline are different at every lag. But a more parsimonious representation may be obtained by supposing that the shape of the nonlinear response is the same at all lags and all that varies is its intensity. In other words,

$$\boldsymbol{\delta}_\tau = \alpha_\tau\boldsymbol{\delta}, \qquad \tau = 0, \ldots, m, \quad (A.2a)$$

where $\boldsymbol{\delta}$ is an $(h+1) \times 1$ vector, but the α_τ's are scalars subject to the (arbitrary) normalizing restriction

$$\sum_{\tau=0}^{m} \alpha_\tau = 1. \quad (A.2b)$$

The number of parameters to be estimated is now reduced from $(h+1)(m+1)$ to $h + m + 1$.

Regarding estimation, the easiest way to proceed is by a stepwise procedure in which the α_τ's are held constant while $\boldsymbol{\delta}$ is estimated, and then the α_τ's are estimated while $\boldsymbol{\delta}$ is held constant. This process may then be iterated to convergence; compare Sargan (1964), Oberhofer and Kmenta (1974), and the Cochrane–Orcutt procedure for first-order autoregressive disturbances. A reparameterization is useful. Instead of (A.1), write

$$\delta_t = \tilde{\mathbf{x}}'_t\boldsymbol{\delta}^* + \sum_{\tau=0}^{m-1} \Delta\tilde{\mathbf{x}}'_{t-\tau}\boldsymbol{\delta}^*_\tau, \quad (A.3)$$

where Δ is the first difference operator,

$$\boldsymbol{\delta}^* = \sum_{\tau=0}^{m} \boldsymbol{\delta}_\tau,$$

and

$$\boldsymbol{\delta}^*_\tau = -\sum_{j=\tau+1}^{m} \boldsymbol{\delta}_j, \qquad \tau = 0, 1, \ldots, m-1.$$

Then, substituting from (5.5a),

$$\delta_t = \tilde{\mathbf{x}}'_t\boldsymbol{\delta} + \sum_{\tau=0}^{m-1} \alpha^*_\tau\Delta\tilde{\mathbf{x}}'_{t-\tau}\boldsymbol{\delta}, \quad (A.4)$$

where

$$\alpha^*_\tau = -\sum_{j=\tau+1}^{m} \alpha_j$$

and the coefficient of $\tilde{\mathbf{x}}_t$ is just $\boldsymbol{\delta}$ in view of (A.2b); that is, $\boldsymbol{\delta}^* = \boldsymbol{\delta}$. Estimation proceeds as follows. Suppose initially that the only stochastic component in (4.1) is ε_t, which is NID$(0, \sigma^2)$. Then follow these steps:

1. Carry out an unrestricted regression based on (5.6) to obtain an initial consistent estimator of $\boldsymbol{\delta}$ from the coefficient vector of $\tilde{\mathbf{x}}_t$.
2. Regress $y_t - \tilde{\mathbf{x}}'_t\hat{\boldsymbol{\delta}}$ on $\Delta\tilde{\mathbf{x}}'_{t-\tau}\hat{\boldsymbol{\delta}}$, $\tau = 0, \ldots, m-1$, where $\hat{\boldsymbol{\delta}}$ is the estimator from step 1, to obtain estimators of the α^*_τ's, denoted $\hat{\alpha}^*_\tau$.
3. Regress y_t on

$$\tilde{\mathbf{x}}_t + \sum_{\tau=0}^{m-1} \hat{\alpha}^*_\tau\Delta\tilde{\mathbf{x}}_{t-\tau} \qquad \text{to reestimate } \boldsymbol{\delta}.$$

4. Repeat steps 2 and 3 until convergence.

In the context of the more general model in (4.1), the dependent and explanatory variables must first be transformed so as to give a scalar covariance matrix. If the hyperparameters are unknown, the preceding stepwise procedure must be embedded within a search procedure to estimate them.

There are two ways in which the transformation of the explanatory variables may be carried out. Both require the application of the Kalman filter appropriate for the stochastic part of the model, $\mu_t + \gamma_t + \varepsilon_t$, so that quantities such as the Kalman gain need be computed only once, and repeated application of the Kalman filter does not involve the equations for the state covariance matrix; see Kohn and Ansley (1985) for a discussion of the principles involved. In the first method, the variables defined in steps (2) and (3) are transformed each time the estimates $\hat{\boldsymbol{\delta}}$ or $\hat{\boldsymbol{\alpha}}^* = (\hat{\alpha}^*_0, \ldots, \hat{\alpha}^*_{m-1})$, change. Thus the filter has to be applied $m + 1$ times at step 2 and h times at step 3. The second method entails a single transformation, for given hyperparameters, of all $h(m+1)$ variables contained in the vectors $\mathbf{x}_t$, $\Delta\mathbf{x}_{t-\tau}$, $\tau = 0, \ldots, m-1$. The transformed variables required at steps 2 and 3 are then formed by constructing the appropriate linear combination based on $\hat{\boldsymbol{\delta}}$ and the $\hat{\alpha}^*_\tau$'s. This method requires more storage, and if the stepwise procedure converges in a small number of iterations, it is also computationally more demanding.

APPENDIX B: INTRADAILY RESPONSE PATTERN

It is possible that the response to a particular explanatory variable will depend on the time of day. Thus we might expect a low temperature to result in a much greater demand for heating if it occurs in the day rather than the night.

If the jth time of day, $j = 1, \ldots, s$, is prevailing, then

$$\delta_t = \tilde{\mathbf{x}}'_{t,j}\boldsymbol{\delta}_j, \qquad (A.5)$$

where $\boldsymbol{\delta}_j$ may be reparameterized as

$$\boldsymbol{\delta}_j = (\delta_j^* + 1)\boldsymbol{\delta}, \qquad j = 1, \ldots, s, \qquad (A.6)$$

such that

$$\Sigma\delta_j^* = 0 \quad \text{and} \quad \Sigma\boldsymbol{\delta}_j = s\boldsymbol{\delta}.$$

Thus if $\delta_j^* = 0$ for all j, then the response is the same throughout the day.

As with lags, it may be reasonable to assume that the shape of the response does not change. Thus $\delta_j^* = \mathbf{w}'_j\boldsymbol{\gamma}^\dagger$ and (A.5) becomes

$$\delta_t = \tilde{\mathbf{x}}'_{t,j}(\mathbf{w}'_j\boldsymbol{\gamma}^\dagger + 1)\boldsymbol{\delta}, \qquad (A.7)$$

where $\mathbf{w}'_j\boldsymbol{\gamma}^\dagger$ denotes the jth point on a periodic spline constrained to sum to 0 over the day. Although the notation is the same, this spline will be different than the one used to model the intradaily level effects, γ_t, in (4.1). Estimation proceeds in much the same way as for distributed lags. The initial estimator of $\boldsymbol{\delta}$ is obtained by averaging the unrestricted estimators of the $\boldsymbol{\delta}_j$'s.

Without the restriction in (A.7), we might use a bivariate spline to allow the shape of the response to change over the day. This would require many more parameters and could become quite complex.

[Received February 1992. Revised March 1993.]

REFERENCES

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

Engle, R. F., Granger, C. W. J., Rice, J. E., and Weiss, A. (1986), "Semi-parametric Estimates of the Relation Between Weather and Between Weather and Electricity Sales," *Journal of the American Statistical Association,* 81, 310–320.

Harrison, P. J., and Stevens, C. F. (1976), "Bayesian Forecasting," *Journal of the Royal Statistical Society,* Ser. B, 38, 205–247.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models, and the Kalman Filter,* Cambridge, U.K.: Cambridge University Press.

Hendricks, W., Koenker, R., and Poirier, D. J. (1979), "Residential Demand for Electricity: An Econometric Approach," *Journal of Econometrics,* 9, 33–57.

Kohn, R., and Ansley, C. F. (1985), "Efficient Estimation and Prediction in Time Series Regression Models," *Biometrika,* 72, 694–697.

Oberhofer, W., and Kmenta, J. (1974), "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models," *Econometrica,* 42, 579–590.

Poirier, D. J. (1973), "Piecewise Regression Using Cubic Splines," *Journal of the American Statistical Association,* 68, 515–524.

——— (1976), *The Econometrics of Structural Change: With Special Emphasis on Spline Functions,* Amsterdam: North-Holland.

Sargan, J. D. (1964), "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in *Econometric Analysis for National Economic Planning,* eds. P. E. Hart, G. Mills and J. K. Whitaker, London: Butterworth, pp. 25–54.