Available online at www.sciencedirect.com

ScienceDirect

European Journal of Operational Research 191 (2008) 207–222

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/ejor

Stochastics and Statistics

Forecasting time series with multiple seasonal patterns

Phillip G. Gould [a], Anne B. Koehler [c,*], J. Keith Ord [d], Ralph D. Snyder [b], Rob J. Hyndman [b], Farshid Vahid-Araghi [e]

[a] *Monash University Accident Research Centre, Monash University, Melbourne, Australia*
[b] *Department of Business Statistics and Econometrics, Monash University, Melbourne, Australia*
[c] *Department of Decision Sciences and MIS, Miami University, Oxford, OH 45056, United States*
[d] *McDonough School of Business, Georgetown University, Washington, DC 20057, United States*
[e] *Faculty of Economics, Australian National University, Canberra, Australia*

Received 5 February 2007; accepted 8 August 2007
Available online 1 September 2007

Abstract

A new approach is proposed for forecasting a time series with multiple seasonal patterns. A state space model is developed for the series using the innovations approach which enables us to develop explicit models for both additive and multiplicative seasonality. Parameter estimates may be obtained using methods from exponential smoothing. The proposed model is used to examine hourly and daily patterns in hourly data for both utility loads and traffic flows. Our formulation provides a model for several existing seasonal methods and also provides new options, which result in superior forecasting performance over a range of prediction horizons. In particular, seasonal components can be updated more frequently than once during a seasonal cycle. The approach is likely to be useful in a wide range of applications involving both high and low frequency data, and it handles missing values in a straightforward manner.

Keywords: Forecasting; Time series; Exponential smoothing; Seasonality; State space models

1. Introduction

Time series may contain multiple seasonal cycles of different lengths. For example, the hourly utility demand data shown in Fig. 1 exhibit both daily and weekly cycles. Such a plot contrasts with the seasonal times series usually considered, which contain only an annual cycle for monthly or quarterly data. Note that we use the term "cycle" to denote any pattern that repeats (with variation) periodically rather than an economic cycle that has no fixed length. Our second example relates to traffic flows, and a similar pattern holds, see Fig. 8. It is easy to think of many other cases where daily and weekly cycles would occur, such as hospital admissions, demand for public transportation, calls to call centers, and requests for cash at ATM machines.

There are several notable features in Fig. 1. First, we observe that the daily cycles are not all the same, although it may reasonably be claimed that the cycles for Monday through Thursday are similar, and perhaps Friday also. Those for Saturday and Sunday are quite distinct. In addition, the patterns

* Corresponding author.
E-mail address: koehleab@muohio.edu (A.B. Koehler).

0377-2217/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2007.08.024

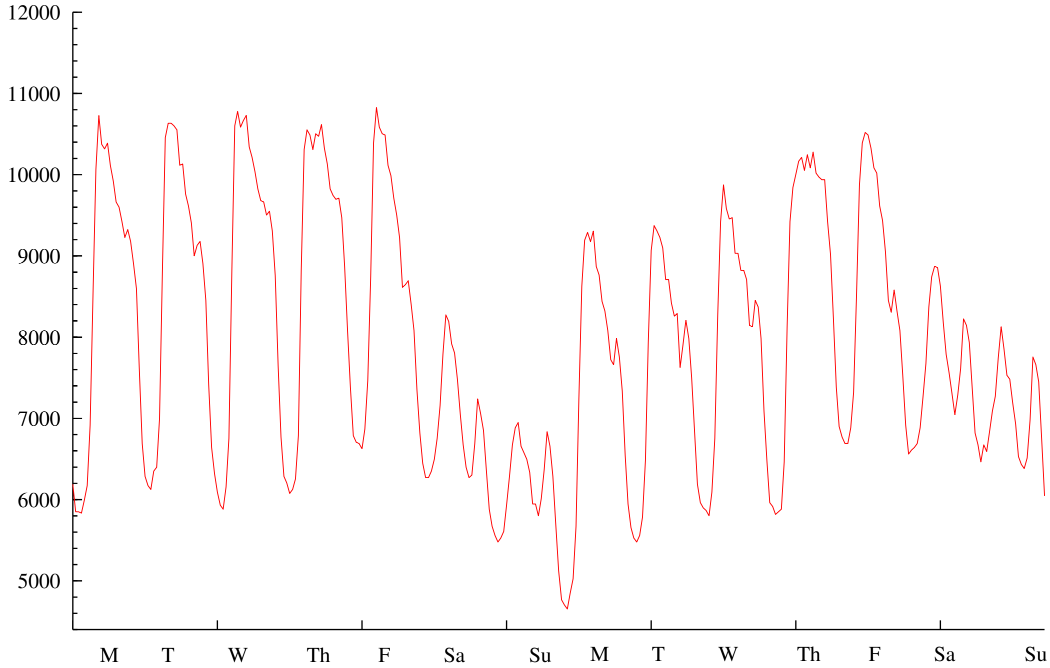

Fig. 1. Sub-sample of hourly utility data.

for public holidays are usually more similar to weekends than to regular weekdays. A second feature of the data is that the underlying levels of the daily cycles may change from one week to the next, yet be highly correlated with the levels for the days immediately preceding. Thus, an effective time series model must be sufficiently flexible to capture these principal features without imposing too heavy computational or inferential burdens.

The goal of this paper is to introduce a new procedure that uses innovations state space models to forecast time series with multiple seasonal patterns. The innovations state space approach provides a theoretical foundation for exponential smoothing methods (Ord et al., 1997; Hyndman et al., 2002). For a recent review of exponential smoothing methods, see Gardner (2006). This procedure improves on the current approaches by providing, in one methodology, a common sense structure to the models, flexibility in modeling seasonal patterns, a potential reduction in the number of parameters to be estimated, and model based prediction intervals.

The most commonly employed approaches to modeling seasonal patterns include the Holt–Winters exponential smoothing approach (Winters, 1960) and the ARIMA models of Box et al. (1993). The Holt–Winters approach could be used for the type of data shown in Fig. 1, but suffers from several important weaknesses. It would require 168 starting values (24 h $\times$ 7 days) and would fail to pick up the similarities from day-to-day at a particular time. Also, it does not allow for patterns on different days to adapt at different rates nor for the component for one day to be revised on another day. In a recent paper, Taylor (2003) has developed a double seasonal exponential smoothing method, which allows the inclusion of one cycle nested within another. His method is described briefly in Section 2.2. Taylor's method represents a considerable improvement, but assumes the same intra-day cycle for all days of the week. Moreover, updates based upon recent information (the intra-day cycle) are the same for each day of the week.

An ARIMA model could be established by including additional seasonal factors. Such an approach again requires the same cyclical behavior for each day of the week. Although the resulting model may provide a reasonable fit to the data, there is a lack of transparency in such a complex model compared to the specification provided by Taylor's approach and by the methods we describe later in this paper.

Harvey (1989) provided an unobserved components approach to modeling multiple seasonal patterns and his approach is similar in some ways to the innovations state space approach described in this paper. The principal difference is that Harvey's state equations use multiple (independent) sources of error in each state equation, whereas our scheme uses a single source of error, following Snyder (1985). At first sight, the multiple error model may seem to be more general, but this is not the case. As shown, for example, in Durbin and Koopman (2001), both sets of assumptions lead to the same class of ARIMA models, although the single source models typically have a larger parameter space. The innovations state space model has several advantages over the multiple source model (Ord et al., 2005):

(1) the parameters may be estimated directly by least squares without using the Kalman filter;
(2) the updating equations are identical in form to the model equations, making interpretation more straightforward;
(3) models for non-linear processes (e.g., the multiplicative Holt–Winters method) are readily formulated and easy to apply;
(4) it becomes feasible to develop prediction intervals for both linear and non-linear methods (Hyndman et al., 2005).

The paper is structured as follows: the additive Holt–Winters (HW) method and Taylor's double seasonal (DS) scheme are outlined in Section 2. Our multiple seasonal (MS) process is introduced and developed in Section 3; the primary emphasis is on the additive scheme, but the multiplicative version is also briefly described. Applications to hourly data on utility demand and on traffic flows are considered in Sections 4 and 5, respectively. Concluding remarks and directions for further research are presented in Section 6.

2. Exponential smoothing for seasonal data

2.1. A structural model for the Holt–Winters (HW) method

The Holt–Winters (HW) exponential smoothing approach (Winters, 1960) includes methods for both additive and multiplicative seasonal patterns. Our primary development in Sections 2 and 3 is in terms of additive seasonality; the corresponding model for the multiplicative case is presented in Section 3.2. A model for the additive seasonal HW method decomposes the series value y_t into an error ε_t, a level ℓ_t, a trend b_t and a seasonal component (s_t). The underlying model based on the innovations state space model (Ord et al., 1997) is

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, \tag{1a}$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \tag{1b}$$
$$b_t = b_{t-1} + \beta\varepsilon_t, \tag{1c}$$
$$s_t = s_{t-m} + \gamma_w\varepsilon_t, \tag{1d}$$

where $\varepsilon_t \sim \mathrm{NID}(0, \sigma^2)$, and α, β and γ_w are smoothing parameters for the level, trend and seasonal terms, respectively. "NID(0, σ^2)" indicates that the errors are independent and identically distributed and that the common distribution is Gaussian with mean 0 and variance σ^2. The smoothing parameters reflect how quickly the level, trend, and seasonal components adapt to new information. The value of m represents the number of seasons in one seasonal cycle. We will denote this model by HW(m) and the seasonal cycle by

$$\boldsymbol{c}_t = (s_t, s_{t-1}, \ldots, s_{t-m+1})'. \tag{2}$$

Estimates of $m+2$ different seed values for the unobserved components must be made; one for the level, one for the trend, and m for the seasonal terms (although we constrain the initial seasonal components to sum to 0).

The HW method allows each of the m seasonal terms to be updated only once during the seasonal cycle of m time periods. Thus, for hourly data we might have an HW(24) model that has a cycle of length 24 (a daily cycle). Each of the 24 seasonal terms would be updated once every 24 h. Or we might have an HW(168) model that has a cycle of length 168 (24 h × 7 days). Although a daily pattern might occur within this weekly cycle, each of the 168 seasonal terms would be updated only once per week. In addition, the same smoothing constant γ_w is used for each of the m seasonal terms. We will show how to relax these restrictions by use of our MS model.

2.2. A structural model for the double seasonal (DS) method

Taylor's double seasonal (DS) exponential smoothing method (Taylor, 2003) was developed to forecast time series with two seasonal cycles: a short one that repeats itself many times within a longer one. It should not be confused with double exponential smoothing (Brown, 1959), the primary focus of which is on a local linear trend. Taylor (2003) developed a method for multiplicative seasonality (i.e., larger seasonal variation at higher values of y_t), which we adapt for additive seasonality (i.e., size of seasonal variation not affected by the level of y_t). We now present a description of the DS method, as a lead-in to the more general approach that we develop later in the paper.

Like the HW exponential smoothing methods, DS exponential smoothing is a *method*. It was specified without recourse to a stochastic *model*, and hence, it cannot be used in its current form to find estimates of the uncertainty surrounding predictions. In particular, a model is required to find prediction intervals. The problem is resolved by specifying an innovations state space model underpinning the additive DS method. Letting m_1 and m_2 designate the periods of the two cycles, this model is

$$y_t = \ell_{t-1} + b_{t-1} + s^{(1)}_{t-m_1} + s^{(2)}_{t-m_2} + \varepsilon_t, \tag{3a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \tag{3b}$$

$$b_t = b_{t-1} + \beta\varepsilon_t, \tag{3c}$$

$$s^{(1)}_t = s^{(1)}_{t-m_1} + \gamma_{d_1}\varepsilon_t, \tag{3d}$$

$$s^{(2)}_t = s^{(2)}_{t-m_2} + \gamma_{d_2}\varepsilon_t, \tag{3e}$$

where $\varepsilon_t \sim \mathrm{NID}(0, \sigma^2)$, and the smoothing parameters for the two seasonal components are γ_{d_1} and γ_{d_2}. We denote this model by DS(m_1, m_2) and the two seasonal cycles by

$$\boldsymbol{c}^{(1)}_t = \left(s^{(1)}_t, s^{(1)}_{t-1}, \dots, s^{(1)}_{t-m_1+1}\right)' \tag{4}$$

and

$$\boldsymbol{c}^{(2)}_t = \left(s^{(2)}_t, s^{(2)}_{t-1}, \dots, s^{(2)}_{t-m_2+1}\right)'. \tag{5}$$

Estimates for $m_1 + m_2 + 2$ seeds must be made for this model.

There are m_2 seasonal terms in the long cycle that are updated once in every m_2 time periods. There are an additional m_1 seasonal terms in the shorter cycle that are updated once in every m_1 time periods. It is not a requirement of the DS(m_1, m_2) model that m_1 is a divisor of m_2. However, if $k = m_2/m_1$, then there are k shorter cycles within the longer cycle. Hence for hourly data, there would be 168 seasonal terms that are updated once in every weekly cycle of 168 time periods and another 24 seasonal terms that are updated once in every daily cycle of 24 time periods. For the longer weekly cycle the same smoothing parameter, γ_{d_2}, is used for each of the 168 seasonal terms, and for the shorter daily cycle the same smoothing parameter, γ_{d_1}, is used for each of the 24 seasonal terms. In our MS model we will be able to relax these restrictions.

2.3. *Using indicator variables in a model for the HW method*

We now show how to use dummy variables to express the HW(m_2) model in two other forms when $k = m_2/m_1$. We do this to make it easier to understand the MS model and its special cases in the next section. First we divide the cycle $\boldsymbol{c}_0$ for HW(m_2) into k sub-cycles of length m_1 as follows

$$\begin{aligned}\boldsymbol{c}_{i,0} &= (s_{i,0}, s_{i,-1}, \dots, s_{i,-m_1+1})' \\ &= (s_{-m_1(k-i)}, s_{-m_1(k-i)-1}, \dots, s_{-m_1(k-i)-m_1+1})',\end{aligned} \tag{6}$$

where $i = 1, \dots, k$, and

$$\boldsymbol{c}_0 = (\boldsymbol{c}'_{k,0}, \boldsymbol{c}'_{k-1,0}, \dots, \boldsymbol{c}'_{1,0})'. \tag{7}$$

For example with hourly data, we could divide the weekly cycle of length 168 into $k = 7$ daily sub-cycles of length $m_1 = 24$. At each time period t, $\boldsymbol{c}_{it}$ contains the current values of the m_1 seasonal components for cycle i (i.e., day i) and is defined by

$$\boldsymbol{c}_{it} = (s_{i,t}, s_{i,t-1}, \dots, s_{i,t-m_1+1})' \quad i = 1, \dots, k. \tag{8}$$

Next we define a set of dummy variables that indicate which sub-cycle is in effect for time period t. For example, when using hourly data these dummy variables would indicate the daily cycle to which the time period belongs. The dummy variables are defined as follows

$$x_{it} = \begin{cases} 1 & \text{if time } t \text{ occurs when sub-cycle} \\ & i \text{ (e.g.day } i\text{)is in effect,} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Then the HW(m_2) model may be written as follows:

$$y_t = \ell_{t-1} + b_{t-1} + \sum_{i=1}^{k} x_{it}s_{i,t-m_1} + \varepsilon_t, \tag{10a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \tag{10b}$$

$$b_t = b_{t-1} + \beta\varepsilon_t, \tag{10c}$$

$$s_{it} = s_{i,t-m_1} + \gamma_w x_{it}\varepsilon_t \quad (i = 1, \dots, k). \tag{10d}$$

The effect of the x_{it} is to ensure that the m_2 ($=k \times m_1$) seasonal terms are each updated exactly once in every m_2 time periods. Eq. (10d) may also be written in a form that will be a special case of the MS model in the next section as follows

$$s_{it} = s_{i,t-m_1} + \left(\sum_{j=1}^{k} \gamma_{ij}x_{jt}\right)\varepsilon_t \quad (i = 1, \dots, k),$$

where

$$\gamma_{ij} = \begin{cases} \gamma_w & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

3. Multiple seasonal processes

3.1. A structural model for multiple seasonal (MS) processes

A fundamental goal of our new model for multiple seasonal (MS) processes is to allow for the seasonal terms that represent a seasonal cycle to be updated more than once during the period of the cycle. This goal may be achieved in two ways with our model. We start, as we did for the HW(m_2) model in the previous section, by dividing the cycle of length m_2 into k shorter sub-cycles of length m_1. Then we use a matrix of smoothing parameters that allows the seasonal terms of one sub-cycle to be updated during the time for another sub-cycle. For example, seasonal terms for Monday can be updated on Tuesday. Sometimes this goal can be achieved by combining sub-cycles with the same seasonal pattern into one common sub-cycle. This latter approach has the advantage of reducing the number of seed values that are needed. When modeling the utility data in Fig. 1, for example, there are potentially seven distinct sub-cycles; one for each day of the week. However, since the daily patterns for Monday through Thursday seem to be very similar, a reduction in complexity might be achieved by using the same sub-cycle for these four days. More frequent updates may also provide better forecasts, particulary when the observations m_1 time periods ago are more important than those values m_2 time periods earlier. It is also possible with our model to have different smoothing parameters for different sub-cycles (e.g., for different days of the week).

The existence of common sub-cycles is the key to reducing the number of seed values compared to those required by the HW method and DS exponential smoothing. As described in Section 2.3, it may be possible for a long cycle to be broken into $k = m_2/m_1$ shorter cycles of length m_1. Of these k possible sub-cycles, $r \leqslant k$ distinct cycles may be identified. For example, consider the case when $m_1 = 24$ and $m_2 = 168$ for hourly data. By assuming that Monday–Friday have the same seasonal pattern, we can use the same sub-cycle for these 5 days. We can use the same sub-cycle for Saturday and Sunday, if they are similar. Thus, we might be able to reduce the number of daily sub-cycles from $k = 7$ to $r = 2$. The number of seed estimates required for the seasonal terms would be reduced from 168 for the HW method and 192 for the DS method to 48 for the new method. (A similar quest formed the motivation for developing cubic spline models for hourly utility data (Harvey and Koopman, 1993).)

A set of dummy variables based on the r shorter cycles can be defined by

$$x_{it} = \begin{cases} 1 & \text{if time period } t \text{ occurs when} \\ & \text{sub-cycle } i \text{ is in effect,} \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

On any given day, only one of the x_{it} values equals 1. Let $\boldsymbol{x_t} = [x_{1t}, x_{2t}, x_{3t}, \ldots, x_{rt}]'$ and $\boldsymbol{s_t} = [s_{1t}, s_{2t}, s_{3t}, \ldots, s_{rt}]'$.

The general summation form of the MS model for $r \leqslant k = m_2/m_1$ is

$$y_t = \ell_{t-1} + b_{t-1} + \sum_{i=1}^{r} x_{it} s_{i,t-m_1} + \varepsilon_t, \tag{12a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \tag{12b}$$

$$b_t = b_{t-1} + \beta\varepsilon_t, \tag{12c}$$

$$s_{it} = s_{i,t-m_1} + \left(\sum_{j=1}^{r} \gamma_{ij} x_{jt}\right)\varepsilon_t \quad (i = 1, \ldots, r), \tag{12d}$$

where $\varepsilon_t \sim \mathrm{NID}(0, \sigma^2)$.

These equations can also be written in matrix form:

$$y_t = \ell_{t-1} + b_{t-1} + \boldsymbol{x}_t' \boldsymbol{s}_{t-m_1} + \varepsilon_t, \tag{13a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \tag{13b}$$

$$b_t = b_{t-1} + \beta\varepsilon_t, \tag{13c}$$

$$\boldsymbol{s}_t = \boldsymbol{s}_{t-m_1} + \boldsymbol{\Gamma} \boldsymbol{x}_t \varepsilon_t, \tag{13d}$$

$$\hat{y}_t(1) = \ell_{t-1} + b_{t-1} + \boldsymbol{x}_t' \boldsymbol{s}_{t-m_1}, \tag{13e}$$

where $\hat{y}_t(1)$ is the one-period ahead prediction.

$\boldsymbol{\Gamma}$ is the seasonal smoothing matrix, which contains the smoothing parameters for each of the cycles. The parameter γ_{ii} is used to update seasonal terms during time periods that belong to the same sub-cycle (e.g., days that have the same daily pattern). The parameter γ_{ij}, $i \neq j$, is used to update seasonal terms belonging to a sub-cycle during the time periods that occur during another sub-cycle (e.g., seasonal terms for one day can be updated during a day that does not have the same daily pattern). We will denote this model by MS(r; m_1, m_2) and the seasonal cycles by

$$\boldsymbol{c}_{it} = (s_{i,t}, s_{i,t-1}, \ldots, s_{i,t-m_1+1})' \quad (i = 1, \ldots, r). \tag{14}$$

While (13) is helpful when programming with matrix software, (12) is better for understanding how the model works. For example, suppose we have hourly data that has a different daily pattern for each of the 7 days of the week to give a weekly pattern that repeats every 168 h. If time period t occurs say on day 3 of the week, then cycle $\boldsymbol{c}_{3t}$ is in effect. Hence, $x_{3t}=1$, and $x_{jt}=0$ for all $j\neq 3$. If $\gamma_{3,3}\neq 0$, then the seasonal component $s_{3,t}$ would be updated. This is the seasonal component for the day of the week that corresponds to time t. In addition, if say $\gamma_{2,3}\neq 0$, then the seasonal component $s_{2,t}$ would be revised, and it corresponds to the same time of the day as time t, but on day 2. Furthermore, if days 2 and 3 have the same pattern, we can reduce the number of cycles by one with cycle $c_{2,t}$ representing both days. Then if time t is on either day 2 or 3, the dummy variable is defined to pick this cycle.

3.2. A model for multiplicative seasonality

Thus far, we have concentrated upon models for time series that exhibit additive, rather than multiplicative seasonal patterns. In the additive case the seasonal effects do not depend on the level of the time series, while for the multiplicative case the seasonal effects increase at higher values of the time series. We can adapt the MS(r; m_1, m_2) model to account for a multiplicative seasonal pattern using the approach of Ord et al. (1997) for the multiplicative HW method.

The general multiplicative form of the MS model for $r\leqslant k=m_2/m_1$ is

$$y_t=(\ell_{t-1}+b_{t-1})\left(\sum_{i=1}^{r}x_{it}s_{i,t-m_1}\right)(1+\varepsilon_t), \tag{15a}$$

$$\ell_t=(\ell_{t-1}+b_{t-1})(1+\alpha\varepsilon_t), \tag{15b}$$

$$b_t=b_{t-1}+\beta(\ell_{t-1}+b_{t-1})\varepsilon_t, \tag{15c}$$

$$s_{it}=s_{i,t-m_1}\left[1+\left(\sum_{j=1}^{r}\gamma_{ij}x_{jt}\right)\varepsilon_t\right]\quad (i=1,\ldots,r), \tag{15d}$$

$$\hat{y}_t(1)=(\ell_{t-1}+b_{t-1})\left(\sum_{i=1}^{r}x_{it}s_{i,t-m_1}\right), \tag{15e}$$

where $\varepsilon_t\sim \mathrm{NID}(0,\sigma^2)$.

3.3. Model restrictions

In general, the number of smoothing parameters contained in $\boldsymbol{\Gamma}$ is equal to the square of the number of separate sub-cycles (r^2) and can be quite large. In addition to combining some of the sub-cycles into a common sub-cycle, restrictions can be imposed on $\boldsymbol{\Gamma}$ to reduce the number of parameters. We shall see that some of these restrictions produce the HW(m_1), HW(m_2), and DS(m_1, m_2) models as special cases of the MS(r; m_1, m_2) model in (13).

One type of restriction is to force common diagonal and common off-diagonal elements as follows

$$\gamma_{ij}=\begin{cases}\gamma_1^*, & \text{if } i=j \quad \text{common diagonal},\\ \gamma_2^*, & \text{if } i\neq j \quad \text{common off-diagonal}.\end{cases} \tag{16}$$

Within the type of restriction in (16), there are three restrictions of particular interest. We will refer to them as

- *Restriction 1*: $\gamma_1^*\neq 0$, and $\gamma_2^*=0$
  If $r=k$, this restricted model is equivalent to the HW(m_2) model in (1) where $\gamma_1^*=\gamma_{\rm w}$. The seed values for the k seasonal cycles in this MS(k; m_1, m_2) model and the one seasonal cycle in the HW(m_2) model are related as shown in Eqs. (6) and (7) of Section 2.3.
- *Restriction 2*: $\gamma_1^*=\gamma_2^*$
  If the seed values for the r seasonal sub-cycles in the MS(r; m_1, m_2) model are identical, this restricted model is equivalent to the HW(m_1) model in (1) where $\gamma_1^*=\gamma_w$. Normally in the MS(r; m_1, m_2) model, the different sub-cycles are allowed to have different seed values. Hence, this restricted model will only be exactly the same as the HW(m_1) model, if we also restrict the seed values for the sub-cycles to be equal to each other.
- *Restriction 3*: Equivalent to Eq. (16)
  If $r=k$, this restricted model is equivalent to the DS(m_1, m_2) model in (2) where $\gamma_1^*=\gamma_{d_1}+\gamma_{d_2}$ and $\gamma_2^*=\gamma_{d_1}$. The seed values for the k seasonal cycles in this MS(k; m_1, m_2) model and the two seasonal cycles in the DS(m_1, m_2) model are related by

$$\boldsymbol{c}_{i0}=\left(s_0^{(1)}+s_{-m_1(k-i)}^{(2)},s_{-1}^{(1)}+s_{-m_1(k-i)-1}^{(2)},\ldots,s_{-m_1+1}^{(1)}+s_{-m_1(k-i)-m_1+1}^{(2)}\right)'. \tag{17}$$

The MS(r; m_1, m_2) model allows us to explore a much broader range of assumptions than existing methods, while retaining parsimony. It nests the models underlying the additive HW and DS methods. It contains other restricted forms that stand in their own right. Table 1 presents the number of parameters and seed values that require estimates for the MS(r:m_1, m_2) model and some of its restric-

Table 1
Number of smoothing parameters and seed values

Model	Parameters	Seed values
MS(r; m_1, m_2)	r^2+2	rm_1+2
MS(r; m_1, m_2)-Rstr. 1	3	rm_1+2
HW(m_2)	3	$m_2+2=km_1+2$
MS(r; m_1, m_2)-Rstr. 2	3	rm_1+2
HW(m_1)	3	m_1+2
MS(r; m_1, m_2)-Rstr. 3	4	rm_1+2
DS(m_1, m_2)	4	$m_2+2=km_1+2$[a]

[a] Short cycle seed values may be started at 0.

tions. A procedure for choosing among the possible MS(r; m_1, m_2) models with and without these restrictions is described in the next section.

3.4. Model estimation, selection, and prediction

The estimation, model selection, and prediction described in this section apply to both the additive and multiplicative MS models.

3.4.1. Estimation

Within the exponential smoothing framework, the parameters in an MS(r; m_1, m_2) model can be estimated by minimizing the one-step-ahead sum of squared errors

$$\mathrm{SSE}=\sum_{i=1}^{n}(y_t-\hat{y}_t)^2,$$

where n is the number of observations in the series, and $\hat{y}_t=\hat{y}_{t-1}(1)$. The seed states for the level, trend and seasonal components may be estimated by applying the procedures for HW(m_2) in Hyndman et al. (2002) to the time periods that represent four completions of all the sub-cycles (e.g., the first four weeks for hourly data). In principle, maximum likelihood could be used to estimate these seed values along with the smoothing parameters (c.f. Hyndman et al., 2002; Bermudez et al., 2006a,b), but a very large number of states (e.g., 168 states for the seasons when each day has a different pattern) would make this approach very cumbersome and potentially very unstable numerically. The m_1 estimates for each of the k seasonal sub-cycles are then found by using the relationship between the cycles explained in Eqs. (6) and (7) of Section 2.3. If $r<k$, the estimates for the sub-cycles with the same seasonal pattern are averaged. Then the SSE is minimized with respect to the smoothing parameters by using the exponential smoothing equations in (12). The smoothing parameters are restricted to values between 0 and 1.

3.4.2. Model selection

We have seen that various special cases of the MS(r; m_1, m_2) model may be of interest. We may wish to choose the number of seasonal sub-cycles r to be less than k, restrict the values of the seasonal parameters, or use a combination of the two. We employ a two-step process to make these decisions. First we choose r, and then we determine whether to restrict Γ as follows:

(1) Choose the value of r in MS(r; m_1, m_2).
   (a) From a sample of size n, withhold q time periods, where q is the last 20% of the data rounded to the nearest multiple of m_2 (e.g., whole number of weeks).
   (b) Select a set of values of interest for r (e.g., using common sense and/or graphs), and estimate the parameters for each model using observations $4m_1+1$ to $n-q$.
   (c) For each of the models in 1(b), find one-period-ahead forecasts for time periods $n-q+1$ to n without re-estimating.
   (d) Pick the value of r with the smallest mean square forecast error.

$$\mathrm{MSFE}(1)=\sum_{t=n-q+1}^{n}(y_t-\hat{y}_t)^2/q$$

(2) Choose the restrictions on $\boldsymbol{\Gamma}$.
   (a) Using the value of r selected in part 1 and the same $n-q$ time periods, compute the one-period-ahead forecast errors for Restrictions 1–3, no restriction, and any other restrictions of particular interest over $[n-q+1, n]$.
   (b) Choose the restriction with the smallest MSFE.

3.4.3. Prediction

A point forecast for y_{n+h} at time period n is the conditional expected value

$$\hat{y}_{n+h}(n)=E(y_{n+h}|\boldsymbol{a}_0,y_1,\ldots,y_n),$$

where

$$\boldsymbol{a}_0=(\ell_0,b_0,s_{1,0},\ldots,s_{1,-m_1+1},s_{2,0},\ldots,s_{2,-m_1+1},\ldots,\\ s_{r,0},\ldots,s_{r,-m_1+1})'=(\ell_0,b_0,\boldsymbol{c}'_{1,0},\boldsymbol{c}'_{2,0},\ldots,\boldsymbol{c}'_{r,0})'.$$

Prediction intervals for h periods in the future from time period n can be found by using the model in (12) as follows: simulate an entire distribution for y_{n+h} and pick the percentiles for the desired level of confidence (Ord et al., 1997).

4. An application to utility data

Forecasting demand is an important issue for utility companies. For example, electricity suppliers need to compute short-term hourly demand forecasts for operational planning. Turning generators on and off is very expensive, but so are black-outs and brown-outs. Consequently, short-term demand forecasting is an essential part of energy management systems and is needed for control and scheduling of power systems, and for the optimal utilization of generators and power stations. Accurate demand forecasting provides system dispatchers with timely information to operate economically and reliably (Bunn and Farmer, 1985). In addition, analyzing forecast demand shapes is needed to set pricing. Peak-load is always a big concern and utilities with limited capacity will often negotiate demand-management strategies with large commercial users involving lower prices but low usage during forecast peaks. Similar comments apply to other utility companies such as those supplying natural gas or water.

Utility demand data was selected to illustrate our MS procedure because it clearly has multiple seasonal cycles. In this simple example, we do not discuss the role of important covariates (particularly temperature-related variables) which are often useful in utility demand forecasting; see, for example, Ramanthan et al. (1997) and Cottet and Smith (2003). The MS model could be extended to include such variables, but we leave that for later research. In this empirical example, we show that the MS model performs best within the class of exponential smoothing models.

4.1. The study

The data set plotted in Fig. 1 consists of 3024 observations (18 weeks) of hourly demand, beginning on January 1, 2003. These data are from a utility company in the Midwestern area of the United States. The details of the company and nature of the utility are commercially confidential.

A graph of the data is shown in Fig. 2. This utility data appears to have a changing level rather than a trend so the growth rate b_t is omitted. The data also appear to exhibit an additive seasonal pattern, that is, a seasonal pattern for which the variation does not change with the level of the time series. For this reason the main focus of this application

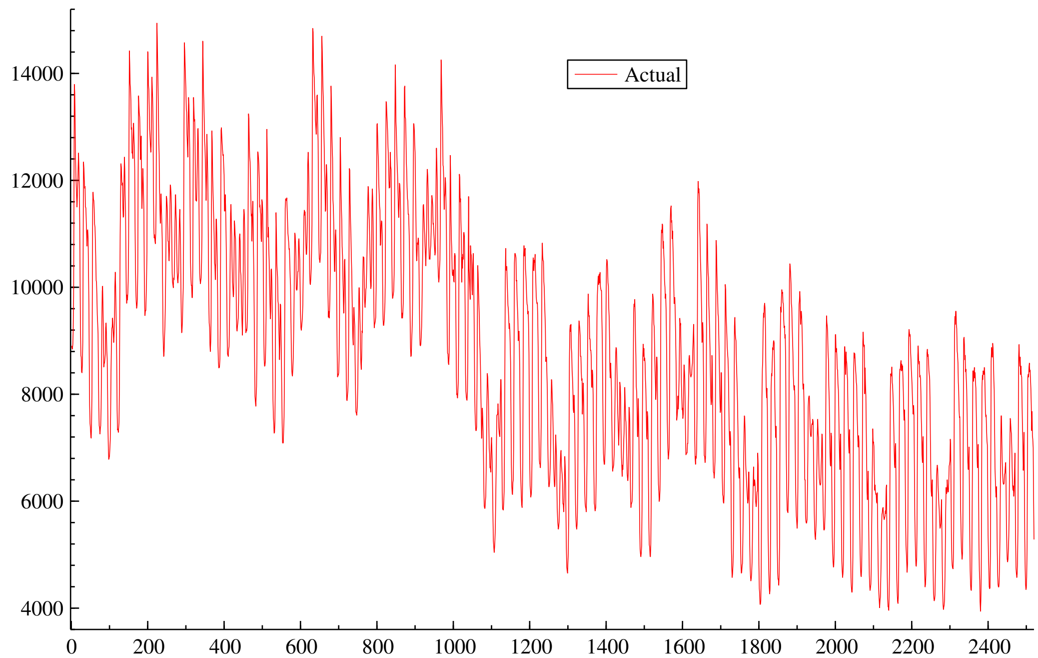

Fig. 2. Hourly utility demand.

is on additive models, although a multiplicative version of our model is also tested. The data are split into two parts: a fitting sample of ($n = 2520$) observations (i.e., 15 weeks) and a post-sample data of ($p = 504$) observations (i.e., 3 weeks). There are no weekday public holidays during the period of this post-sample data.

The data have a number of important features that should be reflected in the model structure. There are three levels of seasonality: yearly effects (largely driven by temperatures), weekly effects and daily effects. For this case study, we will only seek to capture the daily and weekly seasonal patterns.

4.2. Selecting an MS model

In this section we follow the procedure for model selection described in Section 3.4 to select the best MS model. The first step is to choose r in MS(r;24,168). To start this step we withhold 504 observations or three weeks of data ($q = 504$) from the fitting sample ($n = 2520$). The value of q is 20% of n and is the same as that of p in this example. Then, we need to re-examine the data to look for common daily patterns for different days of the week. One way to look for potential common patterns is to graph the 24-h pattern for each day of the week on the same horizontal axis. In Fig. 3, we plot the seasonal terms that are estimated for the seven sub-cycles in the MS(7;24,168) model during the last week of the sample ($t = n - 168, \ldots, n$). The plot suggests that $r = 7$ may use more daily patterns than is required. The similarity of some weekday sub-cycles indicates that alternative structures could be tested.

Visual inspection of Fig. 3 shows that the Monday – Friday sub-cycles are similar, and Saturdays and Sundays are similar. Closer inspection shows that the Monday–Tuesday patterns are similar, Wednesday–Friday patterns are similar, and Saturdays and Sundays display some differences from each other. A third possible approach is to assume Monday–Thursday have a common pattern and Friday, Saturday and Sunday have their own patterns. This choice is plausible because Fridays should have a different evening pattern to other weekdays as consumers and industry settle into weekend routines. Support for this choice of common sub-cycles can also be seen in Fig. 3 where Friday starts to behave more like Saturday in the evening hours. We list these three choices below.

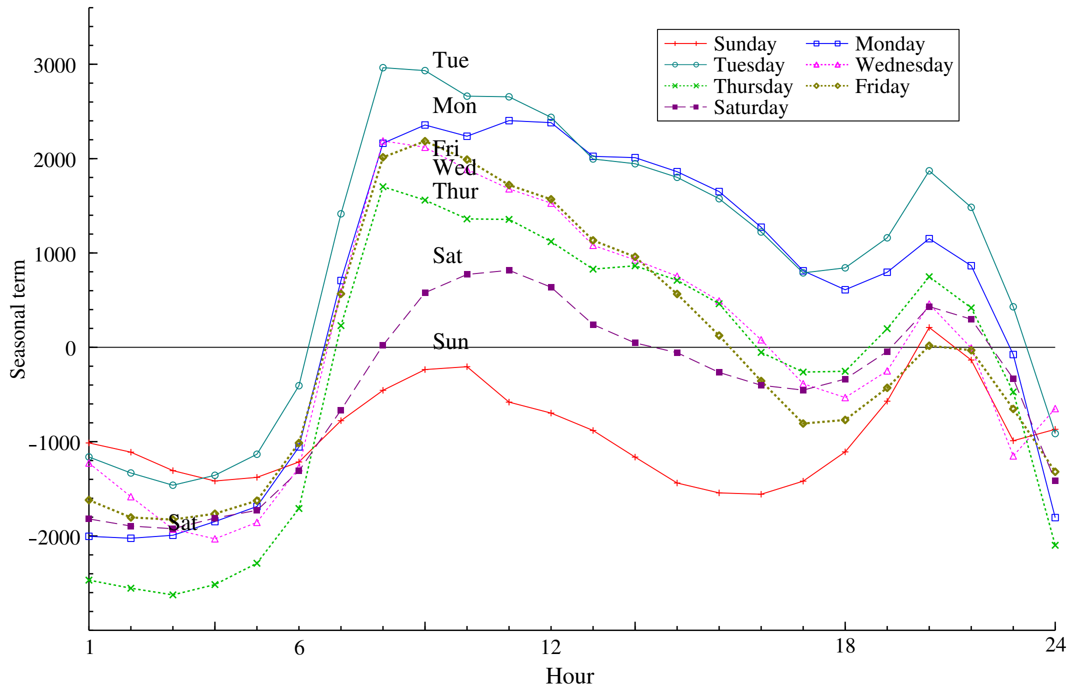

Fig. 3. MS(7;24,168): hourly sub-cycles by day, based on the last 168 observations ($t = 2353, \ldots, 2520$).

- $r = 4$ Version 1 MS(4;24,168): common Monday–Thursday sub-cycle, separate Friday, Saturday and Sunday sub-cycles;
- $r = 4$ Version 2 MS(4(2);24,168): common Monday–Tuesday, Wednesday–Friday sub-cycles, separate Saturday and Sunday sub-cycles;
- $r = 2$ MS(2;24,168): common Monday–Friday sub-cycle, common weekend sub-cycle.

We finish the first step of the model selection process by comparing the value of MSFE(1) for the MS(7;24,168) model to the values for the three sub-models listed above. Of these four models, MS(2;24,168) has the smallest MSFE(1), as shown in Table 2. Thus, we choose this model in the first step. The MSFE(1) values in Table 2 are computed for the withheld time periods $n - q + 1$ to n (i.e., 2017–2520). In Table 2, we say this MSFE(1) compares 'withheld-sample' forecasts to distinguish it from the MSFE(1) in Table 3, which will be computed for the p post-sample values (i.e., 2521–3024) that are not part of the fitting sample.

In the second step of the process from Section 3.4, we compare Restrictions 1–3 from Section 3.3 for the MS(2;24,168) model that was chosen in the first step. The MSFE(1) values for these three additional models are also shown in Table 2. The model with the smallest MSFE(1) for the withheld-sample test is the MS(2;24,168) model with Restriction 3. Hence, this model is our selection for the best MS model for forecasting.

Table 2
Withheld-sample MSFE in MS model selection for utility data

Model	Restriction	MSFE(1)	Parameters	Seed values
MS(7;24,168)	None	234.72	50	168
MS(4;24,168)	None	239.67	17	96
MS(4(2);24,168)	None	250.34	17	96
MS(2;24,168)	None	225.51	5	48
MS(2;24,168)	1	246.51	2	48
MS(2;24,168)	2	234.49	2	48
MS(2;24,168)	3	225.31	3	48

Table 3
Comparison of post-sample forecasts for the utility data

Model	Restriction	MSFE(1)	Parameters	Seed values
HW(24)	na[a]	278.50	2	24
HW(168)	na	278.04	2	168
DS(24,168)	na	227.09	3	168
MS(7;24,168)	None	208.45	50	168
MS(2;24,168)	3	206.45	3	48

[a] Restrictions not applicable to this model.

4.3. Forecasting with the MS, HW and DS models

In general, the MS models provide better point forecasts than the HW and DS models. The forecasting accuracy of the models is compared by using the mean square forecast error for h periods ahead over p post-sample values. The mean square forecast error is defined as

$$\mathrm{MSFE}(h) = \frac{\sum_{t=n}^{n+p-h} (y_{t+h} - \hat{y}_{t+h}(t))^2}{p - (h-1)}, \tag{18}$$

where $\hat{y}_{t+h}(t)$ is the forecast of y_{t+h} at time t. In this application, the MSFE(h) values are averages based on 3 weeks (i.e., $p = 504$ h) of post-sample data and lead-times h of 1–48 h. Table 3 contains the post-sample MSFE(1) for the two HW models of HW(24) and HW(168), the double seasonal model DS(24,168), the full unrestricted multiple seasons model MS(7;24,168), and the selected multiple seasons model MS(2;24,168) with Restriction 3 from Section 4.2. Fig. 4 contains the MSFE(h) for these same five models where the lead-time, h ranges from 1 to 48 (i.e., 1–48 h).

The estimation of the parameters and seed values for these five models is done using the fitting sample of size $n = 2520$. The first four weeks of the fitting sample are used to find initial values for the states. For the HW method these values are found by using the approach in Hyndman et al. (2002). The 24 additional initial values for daily seasonal components in the DS method are set equal to 0. The initial values for MS models are found as described in Section 3.4. Smoothing parameters for all the models are estimated by minimizing the SSE for the fitting sample of length $n = 2520$, and all parameters are constrained to lie between 0 and 1.

In examining Table 3, we see that MS(2;24,168) with Restriction 3 has the smallest MSFE(1), and MS(7;24,168) is second best. The MS(2;24,168) model also has far fewer parameters (3 versus 50) and seed values (48 versus 168) than the MS(7;24,168) model. In Fig. 4, the MSF(h) values are consistently lower for the MS models than for the HW and DS alternatives with the MS model chosen by our selection process being much lower. The more accurate forecasts are the result of the

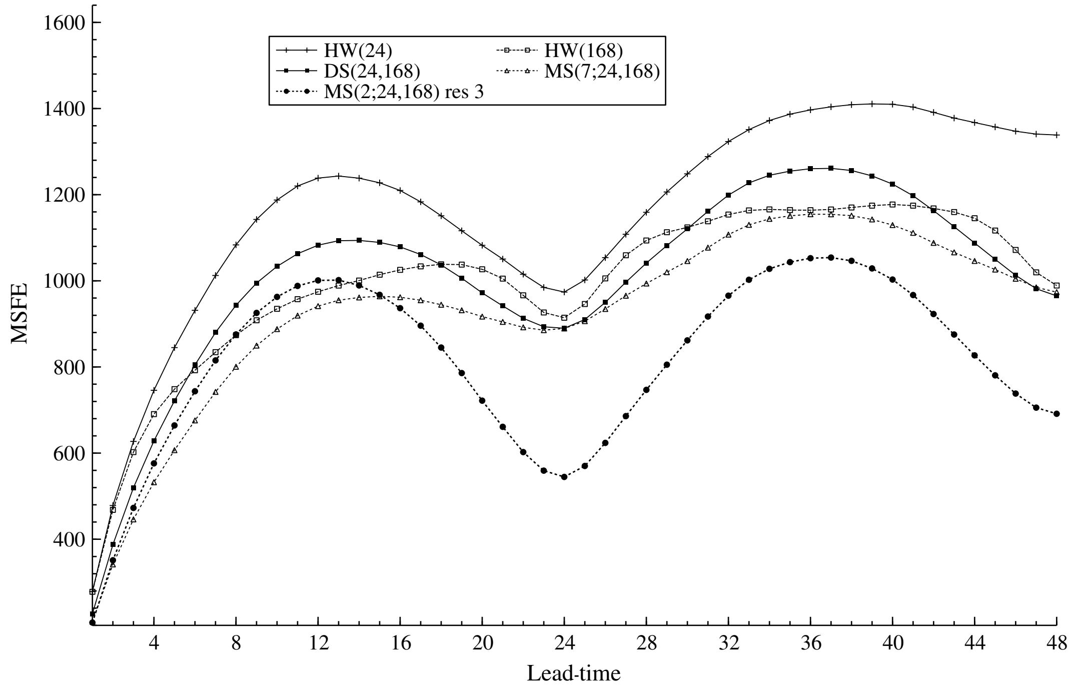

Fig. 4. Forecasting accuracy (MSFE) for lead-times from 1 to 48 h (i.e., 2 days).

MS models offering a more reliable structure to capture the changes in seasonality.

Fig. 5 shows post-sample forecasting accuracy of the MS(2;24,168) model with Restriction 3. Forecasts and 80% prediction intervals are provided only for the first 8 h of the post-sample period because the intervals become extremely wide as the time horizon increases. During the 8 h period in Fig. 5,

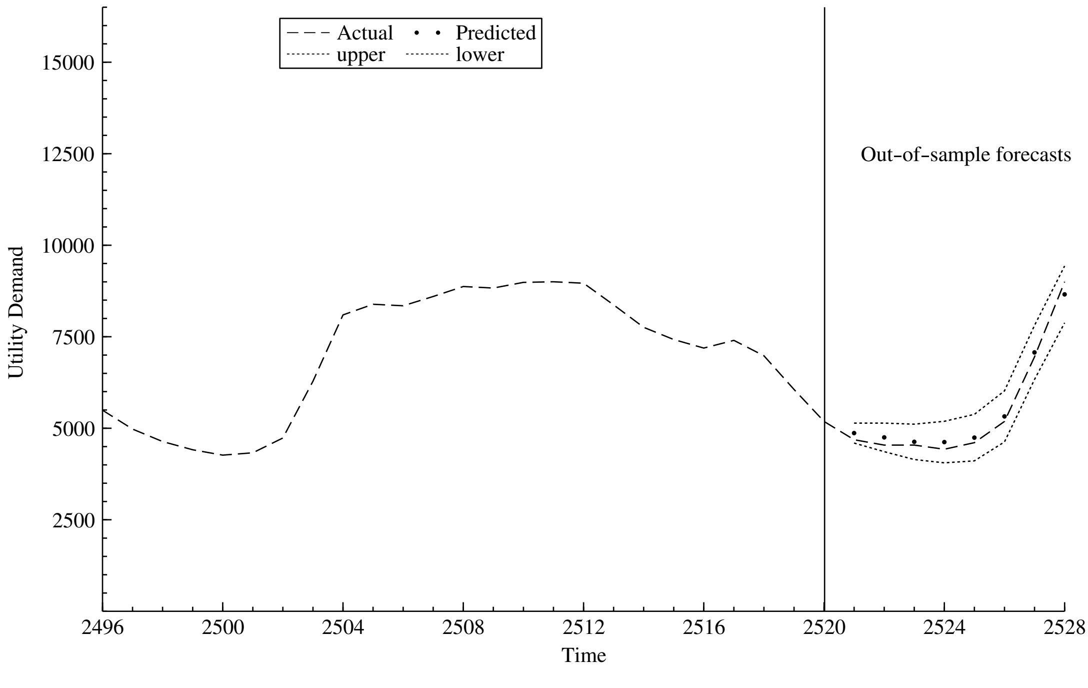

Fig. 5. MS(2;24,168) Restriction 3: Point forecasts and 80% prediction intervals for the first 8 h in the next week ($t = 2521, \ldots, 2528$) of the utility demand.

Table 4
Utility data smoothing parameter estimates

Estimation done for	$\hat{\alpha}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
$h = 1$	1	0.12	0.084
$h = 24$	0	0.83	0.83
$h = 168$	0	0.13	0.11

the forecasts are very good. The 80% prediction intervals are calculated via simulation. One cause for the wide prediction intervals at longer time horizons is the large estimate for α. Large structural change will require wide prediction intervals. For the utility data the parameter α was estimated to be between 0 and 1, and this constraint was binding in most cases (i.e., $\hat{\alpha} = 1$). In this case, the resulting model corresponds to a purely seasonal model for first differences.

4.4. Further comments

The wide prediction intervals that were found when forecasting the utility data can sometimes be avoided, if one's goal is to forecast more than a few hours ahead, For the longer time horizons, the parameters can be estimated by minimizing the sum of squared h-step-ahead errors instead of the usual one-step-ahead errors. Table 4 shows the effect on the estimates for the parameters when the estimation criterion is altered for the utility data in our study. When the sum of squares is minimized for 24-step-ahead or 168-step-ahead errors, the estimate for α is 0, so that the local level becomes a constant. This smaller value for $\hat{\alpha}$ will reduce the width of the prediction intervals at the longer lead-times. Examination of the equations in (12), without (12c) and when $\alpha = 0$, reveals that the prediction intervals will only increase in width every m_1 periods rather than every period. Fig. 4 suggests that narrower prediction intervals become possible, especially for $(1/2)m_1 < h \leqslant m_1$.

An interesting feature of Fig. 4 is the way in which the models have clearly lower MSFE(h) values when h is a multiple of 24. This pattern has been seen in the other studies of seasonal series (e.g., Makridakis and Hibon, 2000) and indicates some degree of model mis-specification. The implication of the numerical results in this case is that the forecasts are more accurate when they are made for a full day ahead at the same time of day (i.e., a purely seasonal model).

In addition to examining the utility data in Fig. 2 to decide that additive seasonal models were appropriate, we tested the multiplicative MS(7;24,168) model in (15) with no trend. We found that the withheld-sample MSFE(1) was 271.48, which is larger than the MSFE(1) of 234.72 for the additive MS(7;24,168) model. This provides further support for our choice of additive seasonality. An advantage of the single source of error models is that such nonlinear models can be included in a study.

Since Taylor (2003) found that adding an AR(1) term improved the forecasting accuracy of the DS model for his utility data, we also examined whether adding an AR(1) would help for our data. We found that forecasts at lead-times longer than one time period are worse when the AR(1) term is included.

5. Analysis of traffic data

Federal and state government agencies of transportation are vitally interested in traffic flow along the roadways. They must provide the information on traffic flow for the scheduling of road maintenance with minimal disruption to motorists and business. Since the maintenance work often spans a period of a few hours, the forecasts must be based on relatively high frequency data to aid in the selection of optimal times of the day. In order to forecast traffic flow, the government agencies collect data on vehicle counts. When combined with transportation models, the forecasts of the vehicle counts enable the planning and scheduling that is necessary. These traffic flow (vehicle count) forecasts can also be used by road freighting companies, who wish to minimize travel times for their vehicle fleets. In practice, the counting equipment is not always reliable and results in missing data. Our MS approach can readily handle missing data, and thus the MS model can be used to predict missing values with one-period-ahead forecasts. Providing values for the missing data is important for models in road safety research.

Hence, for our second application of the MS procedure we have chosen vehicle count data. This data, like the utility demand data, clearly have multiple seasonal cycles. We will show that the forecasts from the MS procedure are more accurate than those from the HW and DS methods.

5.1. The study

The fitting sample consists of 1680 observations (10 weeks) of hourly vehicle counts for the Monash Freeway, outside Melbourne in Victoria, Australia, beginning August, 1995. A graph of the data is shown in Fig. 6. The observation series has missing values when the data recording equipment was not operational. The gaps in the data are for periods of days (i.e., multiples of 24) and can be handled easily in exponential smoothing. Since y_t is not observed, the error ε_t cannot be estimated by $y_t - \hat{y}_{t-1}(1)$. The error is still unknown and governed by an $N(0, \sigma^2)$ distribution. Hence, in this case we use 0 to estimate ε_t in model (12) when estimating the components at time t from the old components at time $t-1$. Such an approach can be applied to any innovations state space model. This procedure for missing values ensures that the exponential smoothing algorithm continues to yield minimum mean squared error predictors (conditional means). It is the analogue of the procedure in Kalman filtering with missing values (Koopman, 1997; Cipra and Romera, 1997). In many traffic applications this ability to handle missing values is particularly useful when counting equipment has to be taken off-line for maintenance.

Apart from the missing observations, the traffic data share the same features as the utility data, although yearly effects are less pronounced. As before, we seek only to capture the daily and weekly seasonal patterns. Since this data appears to have no trend and to exhibit an additive seasonal pattern, we use additive seasonality for the HW, DS, and MS approaches and omit the equation for the growth rate b_t. Victorian public holidays appear throughout the sample and follow a similar daily pattern to Sundays.

This study of vehicle flows includes the HW(24), HW(168), DS(24,168) and MS models. Models are compared by using the MSFE for h periods ahead over a post-sample of length $p = 504$. We examine lead-times of up to two weeks ($h = 1, \dots, 336$), which can be relevant for planning road works. Smoothing parameters and seeds are estimated using the same procedures as the previous section.

An MS model is chosen using the method in Section 3.4 with $q = 336$ (i.e., two weeks of data). Based on visual inspection of the raw data and plots of the seasonal terms for the MS(7;24,168) model, three candidates were tested along with the full MS model.

- $r = 4$ Version 1 MS(4;24,168): common Monday–Thursday sub-cycle, separate Friday, Saturday and Sunday sub-cycles;

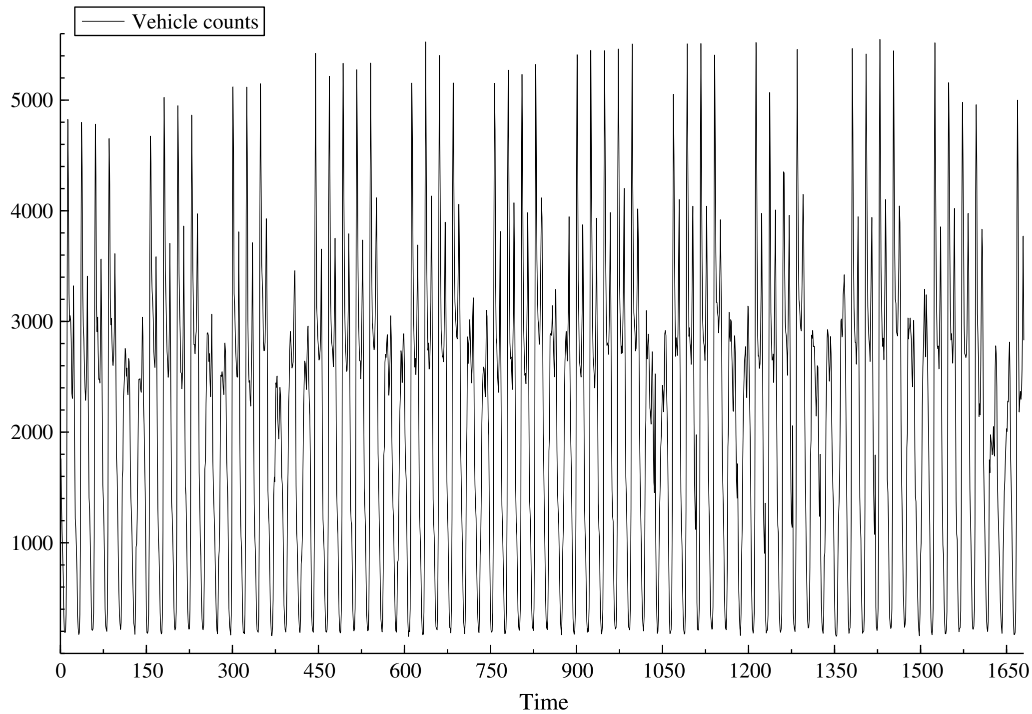

Fig. 6. Hourly vehicle counts.

Table 5
Comparison of withheld-sample forecasts for the traffic data

Model	Restriction	MSFE(1)	Parameters	Seed values
MS(7;24,168)	None	498.31	50	168
MS(4;24,168)	None	428.88	17	96
MS(3;24,168)	None	394.42	10	72
MS(2; 24, 168)	None	308.84	5	48
MS(2;24,168)	1	310.94	2	48
MS(2;24,168)	2	333.85	2	48
MS(2;24,168)	3	310.94	3	48
MS(2;24,168) public holidays	None	228.68	5	48

Table 6
Comparison of post-sample forecasts for the traffic data

Model	Restriction	MSFE(1)	Parameters	Seed values
HW(24)	na	365.09	2	24
HW(168)	na	228.60	2	168
DS(24,168)	na	228.59	3	168
MS(7; 24, 168)	None	238.33	50	168
MS(7; 24, 168) public holidays	None	245.25	50	168
MS(2;24,168) public holidays	None	203.64	5	48

- $r = 3$ MS(3;24,168): common Monday–Friday sub-cycle, separate Saturday and Sunday sub-cycles;
- $r = 2$ MS(2;24,168): common Monday–Friday sub-cycle, common weekend sub-cycle.

In Table 5, we see that, among the first four models, MS(2;24,168) has the smallest MSFE(1), where this MSFE is computed using the withheld values within the original sample. Thus, we choose $r = 2$ in step 1. None of the restrictions are supported. However, if we account for public holidays by using the same indicator as the one for the Saturday/Sunday sub-cycle, the one-period-ahead forecasts for the withheld data are greatly improved. Hence, we choose MS(2;24,168) with *public holidays* for our best MS model.

5.2. Comparison of the MS models with the HW and DS models

In Table 6, the post-sample MSFE(1) can be compared for each of the following six models: HW(24), HW(168), DS(24,168), full MS(7;24,168)

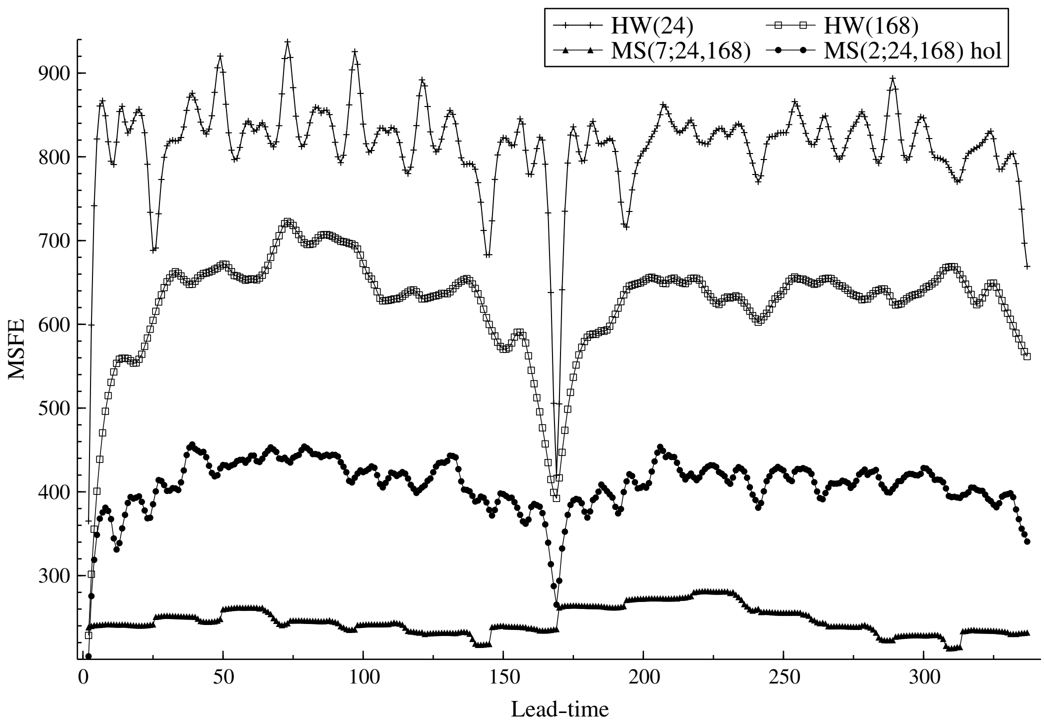

Fig. 7. Forecasting accuracy (MSFE) for lead-times from 1 to 336 h (i.e., 2 weeks).

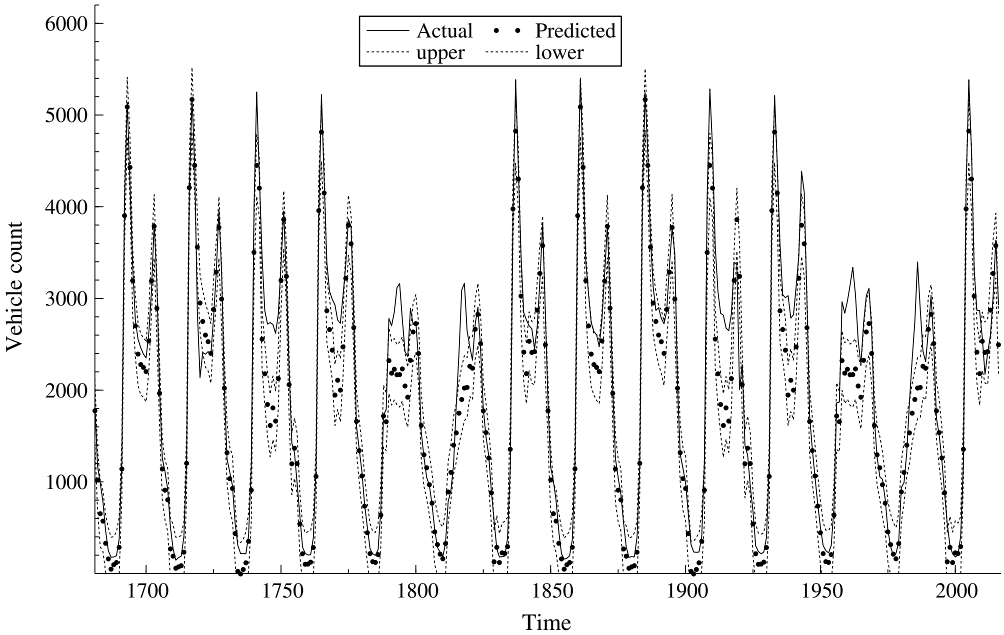

Fig. 8. MS(7;24,168): multi-step-ahead point forecasts and 80% prediction intervals for the vehicle counts for each hour in the last two weeks of the evaluation sample ($t = 1681, \ldots, 2016$).

(with and without public holidays), and MS(2; 24,168) model with public holidays. We see that the MS(2;24,168) model that accounts for public holidays has the smallest MSFE(1), while the MSFE(1) for the MS(7;24,168) model is slightly larger than the essentially common value for HW(168) and DS(24,168). The MS(2;24,168) model with public holidays is clearly the best model for forecasting ahead 1 h, offering a reduction of approximately 15% in MSFE over the HW and DS models.

In Fig. 7, we can compare the HW(24) model, the HW(168) model, the MS(7;24,168)model, and the MS(2;24,168) model with public holidays over lead-times of 1 through 336 h. The values of MSFE(h) when $h > 1$ in this figure give a different ordering to forecasting accuracy than those in Table 6. The estimate of γ_{d_1} when $m_1 = 24$ for DS(24,168) is effectively zero, meaning it is equivalent to HW(168). Thus, the DS(24,168) model is not included, as it is indistinguishable from HW(168). The model selected by our MS selection process, MS(2;24,168) with public holidays, is no longer best, but it still provides far more accurate forecasts than the HW and DS models. Clearly, the MS(7;24,168) produces the best forecasts (i.e., the smallest MSFE(h)) for forecasting horizons of two or more hours ahead.

The unconditional updating of the states during periods of missing data proves to be effective for all models. Generally jumps are observed in the level ℓ_t after periods of missing data. The jumps are more pronounced for the MS(7;24,168) model, which has a relatively stable level during periods of no missing data.

Multi-step-ahead forecasts and 80% prediction intervals for the post-sample data using the MS(7;24,168) model can be found in Fig. 8. The forecasts follow the observed series closely and the prediction intervals are not as wide as those for the utility data. These narrower intervals can be explained by the extremely small estimate for α. For MS(7;24,168), $\hat{\alpha} = 0.01$.

6. Conclusions

A new approach for forecasting a time series with multiple seasonal patterns has been introduced. The framework for this approach employs innovations state space models that allow us to forecast time series with either additive (linear) or multiplicative (non-linear) seasonal patterns. For both additive and multiplicative seasonality, the Holt–Winters (HW) methods and Taylor's double seasonal (DS) methods are special cases of our new multiple seasonal (MS) process. In this development, we have provided innovations state space models for the HW, DS, and MS approaches. The estimation of the parameters for all the models can be done using exponential smoothing. Since we have models rather than just methods, we were able to produce prediction intervals as well as point forecasts.

We applied our MS model to utility demand and vehicle flows. In each case the data had been collected by the hour and displayed daily and weekly seasonal effects. The MS model provided more accurate forecasts than the HW and DS models because of its flexibility. The MS model allows for each day to have its own hourly pattern or to have some days with the same pattern. In addition, the model permits days with different patterns to affect one another. By identifying common seasonal patterns for different sub-cycles, the MS structure makes it possible to greatly reduce the number of parameters and seeds required by the full MS model. We found in both examples that we could use two sub-cycles. Public holidays and missing values are readily handled by the MS model in the examples.

There are some interesting possibilities for future research. Investigation of the effect of longer lead-times on model selection and parameter estimation would be valuable. Our multiple seasonal approach should also be helpful on lower frequency observations when one does not want to wait to update a seasonal factor. A key aim of the MS model is to allow for the seasonal terms to be updated more than once during the period of the long cycle of the data, which was 168 in both of our examples.

Note: Ox code and a Microsoft Excel spreadsheet for the multiple seasonal (MS) model are available at http://www.robhyndman.info/papers/multiseasonal. htm.

References

Bermudez, J.D., Segura, J.V., Vercher, E., 2006a. A decision support system methodology for forecasting of time series based on soft computing. Computational Statistics and Data Analysis 51, 177–191.

Bermudez, J.D., Segura, J.V., Vercher, E., 2006b. Improving demand forecasting accuracy using nonlinear programming software. Journal of the Operational Research Society 57, 94–100.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1993. Time series analysis: forecasting and control. Prentice Hall, New Jersey.

Brown, R., 1959. Statistical Forecasting for Inventory Control. McGraw-Hill, New York.

Bunn, D.W., Farmer, E.D., 1985. Economical and operational context of electric load prediction. In: Bunn, D.W., Farmer, E.D. (Eds.), Comparative Models for Electrical Load Forecasting. Wiley, pp. 3–11.

Cipra, T., Romera, R., 1997. Kalman filter with outliers and missing observations. Test 6, 379–395.

Cottet, R., Smith, M., 2003. Bayesian modeling and forecasting of intraday electricity load. Journal of the American Statistical Association 98 (464), 839–849.

Durbin, J., Koopman, S.J., 2001. Time series analysis by state space methodsOxford Statistical Science Series. Oxford University Press, Oxford.

Gardner Jr., E.S., 2006. Exponential smoothing: the state of the art – part ii. International Journal of Forecasting 22, 637–666.

Harvey, A.C., 1989. Forecasting, structural time series models, and the Kalman filter. Cambridge University Press, Cambridge.

Harvey, A.C., Koopman, S.J., 1993. Forecasting hourly electricity demand using time-varying splines. Journal of the American Statistical Association 88 (424), 1228–1236.

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D., 2005. Prediction intervals for exponential smoothing using two new classes of state space models. Journal of Forecasting 24, 17–37.

Hyndman, R.J., Koehler, A.B., Snyder, R.D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 18, 439–454.

Koopman, S.J., 1997. Exact initial kalman filtering and smoothing for nonstationary time series models. Journal of the American Statistical Association 92 (440), 1630–1638, fLA Theory and Methods 01621459 American Statistical Association.

Makridakis, S., Hibon, M., 2000. The M3-competition: results, conclusions, and implications. International Journal of Forecasting 16, 451–476.

Ord, J.K., Koehler, A.B., Snyder, R.D., 1997. Estimation and prediction for a class of dynamic nonlinear statistical models. Journal of the American Statistical Association 92, 1621–1629.

Ord, J.K., Snyder, R.D., Koehler, A.B., Hyndman, R.J., Leeds, M., 2005. Time series forecasting: the case for the single source of error state space approach. Working Paper Series, Department of Econometrics and Business Statistics, Monash University, Australia.

Ramanthan, R., Engle, R., Granger, C.W., Vahid-Araghi, F., Brace, C., 1997. Short-run forecasts of electricity loads and peaks. International Journal of Forecasting 13, 161–174.

Snyder, R.D., 1985. Recursive estimation of dynamic linear models. Journal of the Royal Statistical Society: Series B 47, 272–276.

Taylor, J., 2003. Short-term electricity demand forecasting using double seasonal exponential smoothing. Journal of Operational Research Society 54, 799–805.

Winters, P.R., 1960. Forecasting sales by exponentially weighted moving averages. Management Science 6, 324–342.