ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Forecasting Time-Series with Multiple Seasonal Patterns

Phillip G. Gould, Anne B. Koehler, Farshid Vahid-Araghi, Ralph D. Snyder,
J. Keith Ord and Rob J. Hyndman.

December 2004
Revised October 2005

Working Paper 28/04

Forecasting Time-Series with Multiple Seasonal Patterns

Phillip G. Gould [*] Anne B. Koehler [†]
Farshid Vahid-Araghi [‡] Ralph D. Snyder, [§] J. Keith Ord [¶]
Rob J. Hyndman, [||]

October 11, 2005

JEL CLASSIFICATION: C22

[*]Phillip G. Gould is a PhD candidate, Department of Business Statistics and Econometrics, Monash University, Melbourne, Australia (email:Phillip.Gould@general.monash.edu.au)

[†]Anne B. Koehler is Professor, Department of Decision Sciences and MIS, Miami University, Oxford, OH 45056 (email: KOEHLEAB@MUOHIO.EDU)

[‡]Farshid Vahid-Araghi is Professor, Faculty of Economics, Australian National University, Canberra, Australia (email:farshid.vahid@anu.edu.au)

[§]Ralph D. Snyder is Associate Professor, Department of Business Statistics and Econometrics, Monash University, Melbourne, Australia (email:Ralph.Snyder@buseco.monash.edu.au)

[¶]J. Keith Ord is Professor, McDonough School of Business, Georgetown University, Washington, DC20057 (email: ORDK@georgetown.edu)

[||]Rob J. Hyndman is Professor, Department of Business Statistics and Econometrics, Monash University, Melbourne, Australia (email:Rob.Hyndman@buseco.monash.edu.au)

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 single source of error approach which enables us to develop explicit models for both additive and multiplicative seasonality. Parameter estimates may be obtained using methods adapted from general exponential smoothing, although the Kalman filter may also be used. 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. 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: exponential smoothing; Holt-Winters; seasonality; structural time series model

1 INTRODUCTION

Time series may contain multiple seasonal cycles of different lengths. For example, the hourly electricity demand data shown in Figure 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.

There are several notable features in Figure 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 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.

Existing 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 Figure 1, but suffers from several important weaknesses. It would require 168 starting values (24 hours $\times 7$ 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. 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 reduced-form model compared to the specification afforded by Taylor's approach and by the methods we describe later in the paper.

Harvey (1989) provided an unobserved components approach to modeling multiple seasonal patterns and his approach is similar in some ways to that 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 & 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 single source of error 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).

There are, of course, other approaches to forecasting electricity load data, such as Ramanthan et al. (1997) and Cottet & Smith (2003). These authors rely upon short-term temperature predictions to improve the quality of their forecasts, which is clearly a desirable relationship to exploit in this context. As our interest lies more broadly in the development of single series methods for which suitable explanatory variables may not be available (e.g. hourly sales figures, volume of stock trading, traffic flows) we do not pursue this approach here. However, it is worth noting that regression components are readily added to such models (Ord 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 electricity 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.

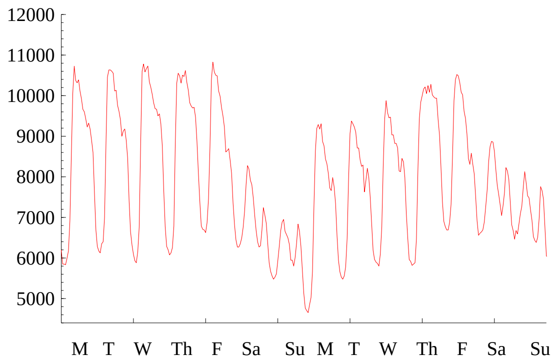

Figure 1: Sub-sample of hourly utility data

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 3.5. A model for the additive seasonal HW method will decompose the series value y_t into an error ε_t, a level ℓ_t, a trend b_t and a seasonal component (s_t). An underlying model based on the single source of error 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 $\varepsilon_t \sim NID(0, \sigma^2)$, and α, β and γ_w are smoothing parameters for the level, trend and seasonal terms, respectively. 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 hours. Or we might have an HW(168) model that has a cycle of length 168 (24 hours $\times 7$ 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).

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. The problem is resolved by specifying a single source of error 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 $\varepsilon_t \sim NID(0, \sigma^2)$, 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

$$\boldsymbol{c}_t^{(1)} = (s_t^{(1)}, s_{t-1}^{(1)}, \ldots, s_{t-m_1+1}^{(1)})' \tag{4}$$

and

$$\boldsymbol{c}_t^{(2)} = (s_t^{(2)}, s_{t-1}^{(2)}, \ldots, s_{t-m_2+1}^{(2)})' \tag{5}$$

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 $\boldsymbol{c}_t$ for HW(m_2) into k sub-cycles of length m_1 as follows:

$$\begin{aligned} \boldsymbol{c}_{it} &= (s_{i,t}, s_{i,t-1}, \ldots, s_{i,t-m_1+1})' \\ &= (s_{t-m_1(k-i)}, s_{t-m_1(k-i)-1}, \ldots, s_{t-m_1(k-i)-m_1+1})' \end{aligned} \tag{6}$$

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

$$\boldsymbol{c}_t = (\boldsymbol{c}'_{kt}, \boldsymbol{c}'_{k-1,t}, \ldots, \boldsymbol{c}'_{1t})' \tag{7}$$

For example with hourly data, we could divide the weekly cycle of length 168 into 7 daily sub-cycles of length 24.

Next we define a set of dummy variables that indicate the sub-cycle which 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:

$$x_{it} = \begin{cases} 1 & \text{if time } t \text{ occurs when sub-cycle } i \text{ (e.g.,day } i\text{) is in effect} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

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

$$y_t = \ell_{t-1} + b_{t-1} + \sum_{i=1}^{k} x_{it} s_{i,t-m_1} + \varepsilon_t \tag{9a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \tag{9b}$$

$$b_t = b_{t-1} + \beta\varepsilon_t \tag{9c}$$

$$s_{it} = s_{i,t-m_1} + \gamma_w x_{it}\varepsilon_t \qquad (i = 1, \ldots, k) \tag{9d}$$

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. Equation (9d) may also be written in a form that will be a special case of the MS model in the next section as follows:

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

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 modelling the electricity data in Figure 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, $r \leq k$ 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 & Koopman 1993).)

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

$$x_{it} = \begin{cases} 1 & \text{if time period } t \text{ occurs when sub-cycle } i \text{ is in effect;} \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

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 $r \leq k = m_2/m_1$ is:

$$y_t = \ell_{t-1} + b_{t-1} + \sum_{i=1}^{r} x_{it} s_{i,t-m_1} + \varepsilon_t \tag{11a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \tag{11b}$$

$$b_t = b_{t-1} + \beta\varepsilon_t \tag{11c}$$

$$s_{it} = s_{i,t-m_1} + \left(\sum_{j=1}^{r} \gamma_{ij} x_{jt}\right) \varepsilon_t \qquad (i = 1, \ldots, r) \tag{11d}$$

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

These equations can also be written in matrix form:

$$y_t = \ell_{t-1} + b_{t-1} + \boldsymbol{x}_t' \boldsymbol{s}_{t-m_1} + \varepsilon_t \tag{12a}$$

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

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

$$\boldsymbol{s}_t = \boldsymbol{s}_{t-m_1} + \boldsymbol{\Gamma} \boldsymbol{x}_t \varepsilon_t \tag{12d}$$

$\boldsymbol{\Gamma}$ 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 $\gamma_{ij}, i \neq j$, 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

$$\boldsymbol{c}_{it} = (s_{i,t}, s_{i,t-1}, \ldots, s_{i,t-m_1+1})' \qquad (i = 1, \ldots, r) \tag{13}$$

This model can also be written in a state space form (see Appendix A) and estimated using the Kalman filter (Snyder 1985). Here, as discussed in section 3.4, we estimate the model in the exponential smoothing framework of Ord et al., 1997.

3.2 Reduced Form of the MS Model

The reduced form of the MS($r; m_1, m_2$) model may be derived from (12) by applying appropriate transformations to y_t to eliminate the state variables and achieve stationarity. The reduced form of the MS model is

$$
\begin{aligned}
\Delta\Delta_{m_2} y_t = & \left[\sum_{j=1}^{r}\left(\theta_{jt}L^{jm_1} - \theta_{j,t+1}L^{i(m_1-1)}\right)\right]\varepsilon_t \\
& + \alpha\Delta_{m_2}\varepsilon_{t-1} + \beta\sum_{j=1}^{m_2} L^j\varepsilon_t + \Delta\Delta_{m_2}\varepsilon_t
\end{aligned} \tag{14}
$$

where L is the lag operator and $\Delta_i = (1 - L^i)$ takes the i^{th} difference. In the case where the trend b_t is omitted, the reduced form becomes:

$$
\Delta_{m_2} y_t = \left(\sum_{j=1}^{r}\theta_{jt}L^{jm_1}\right)\varepsilon_t + \alpha\sum_{j=1}^{m_2} L^j\varepsilon_t + \Delta_{m_2}\varepsilon_t \tag{15}
$$

The θ_{it} value will be a sum of r terms, each of which is a product of a value from $\boldsymbol{x_t}$ and a value from $\boldsymbol{\Gamma}$, but at any time t it will be equal to only one of the values from $\boldsymbol{\Gamma}$.

For example in the case with no trend, $m_1 = 4$, $m_2 = 12$ and $k = r = 3$ (no repeating sub-cycles), (15) can be written as:

$$
\Delta_{12} y_t = \left(\sum_{j=1}^{3}\theta_{jt}L^{4j}\right)\varepsilon_t + \alpha\sum_{j=1}^{12} L^j\varepsilon_t + \Delta_{12}\varepsilon_t \tag{16}
$$

In this case, $\theta_{1t} = x_{1t}\gamma_{13} + x_{2t}\gamma_{21} + x_{3t}\gamma_{32}$, $\theta_{2t} = x_{1t}\gamma_{12} + x_{2t}\gamma_{23} + x_{3t}\gamma_{31}$ and $\theta_{3t} = x_{1t}\gamma_{11} + x_{2t}\gamma_{22} + x_{3t}\gamma_{33}$. See Appendix B for the derivation of the reduced form.

The reduced form of the model verifies that the MS model has a sensible, though complex ARIMA structure with time dependent parameters at the seasonal and near seasonal lags. The advantage of the state space form is that the MS model is more logically derived, easily estimated, and interpretable than its reduced form counterpart. In the next section we give the specific restrictions on $\boldsymbol{\Gamma}$ (and hence the θ_{it} values) that may be used to show that the reduced forms of the HW(m_1), HW(m_2) and the DS(m_1, m_2) models are special cases of the reduced form for the MS($r; m_1, m_2$) model in (15).

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 $\mathbf{\Gamma}$ to reduce the number of parameters.

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

$$\gamma_{ij} = \begin{cases} \gamma_1^*, & \text{if } i = j \qquad \text{common diagonal} \\ \gamma_2^*, & \text{if } i \neq j \qquad \text{common off-diagonal} \end{cases} \tag{17}$$

In this case $\theta_{1t} = \theta_{2t} = \ldots = \theta_{r-1,t} = \gamma_2^*$ and $\theta_{rt} = \gamma_1^*$.

Within the type of restriction in (17), 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 $\gamma_1^* = \gamma_w$. 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 section 2.3 (where $t = 0$).

- **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 Equation (17)
  If $r = k$, this restricted model is equivalent to the DS(m_1, m_2) model in (3) 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

$$\begin{aligned} \boldsymbol{c}_{i0} = (&s_0^{(1)} + s_{-m_1(k-i)}^{(2)}, s_{-1}^{(1)} + s_{-m_1(k-i)-1}^{(2)}, \ldots, \\ &s_{-m_1+1}^{(1)} + s_{-m_1(k-i)-m_1+1}^{(2)})' \end{aligned} \tag{18}$$

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. 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

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

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

where n is the number of observations in the series. 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). The m_1 estimates for each of the k seasonal sub-cycles are then found by using the relationship between the cycles explained in 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 (11). The smoothing parameters are restricted to values between 0 and 1.

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 $\mathbf{\Gamma}$ 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) For reasonable values of r (e.g., using common sense and/or graphs), estimate the parameters in each model.
   (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

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

2. Choose the restrictions on $\mathbf{\Gamma}$.

(a) Using the model chosen in part 1 and the same $n - q$ time periods, compute the one-period-ahead forecast errors for Restrictions 1, 2, and 3, no restriction, and any other restrictions of particular interest.

(b) Choose the restriction with the smallest MSFE.

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

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

Prediction intervals for h periods in the future from time period n can be found by using the model in (11) as follows: simulate an entire distribution for y_{n+h} and pick the percentiles for the desired level of confidence (Ord et al. 1997). For an 80% prediction interval, one would used the tenth and ninetieth percentiles for the lower and upper limits of the prediction interval, respectively.

3.5 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. One way that many forecasters change a multiplicative seasonal pattern into an additive one is to apply a transformation to the time series, such as a logarithmic transformation. Another approach is to build the multiplicative seasonality directly into the model itself. 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 $r \leq k = m_2/m_1$ is:

$$y_t = (\ell_{t-1} + b_{t-1})(\sum_{i=1}^{r} x_{it} s_{i,t-m_1})(1 + \varepsilon_t) \tag{19a}$$

$$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t) \tag{19b}$$

$$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t \tag{19c}$$

$$s_{it} = s_{i,t-m_1}\left[1 + \left(\sum_{j=1}^{r} \gamma_{ij} x_{jt}\right)\varepsilon_t\right] \qquad (i = 1, \ldots, r) \tag{19d}$$

where $\varepsilon_t \sim NID(0, \sigma^2)$. The interested reader should see Ord et al., 1997 for information on how to estimate the parameters in this model.

4 An Application to Utility Data

In this empirical example, we show that the MS model performs best within the class of exponential smoothing models. Other approaches to forecasting electricity demand may be more appropriate in particular circumstances: see, for example, Ramanathan et al., 1997.

4.1 The Study

The sample consists of 2520 observations (15 weeks) of hourly utility demand, beginning on January 1, 2003, from a utility company in the Midwestern area of the United States. A graph of the data is shown in Figure 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 appears 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 is on additive models, although a multiplicative version of our model is also tested. In addition to the fitting sample ($n = 2520$), 504 observations (3 weeks) are available as post-sample data ($p = 504$) for the comparison of forecasts from different models. 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

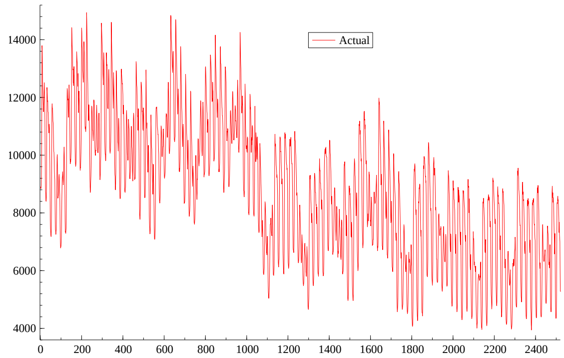

Figure 2: Hourly utility demand

will only seek to capture the daily and weekly seasonal patterns. The yearly pattern can be modelled using a trigonometric specification (Proietti 2000) or by explicitly including temperature as an explanatory variable (Ramanthan et al. 1997).

A number of competing seasonal structures are examined. The HW(24), HW(168), DS($24, 168$) and MS($7; 24, 168$) versions are fitted to the sample of size $n = 2520$. The first four weeks of the 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 twenty-four 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.

Next the best MS model for the utility data is chosen by using the procedure described in section 3.4 where the last 504 observations ($q = 504$) are withheld from the 2520 values in the fitting sample. This additional MS model is compared with the previously considered models.

Then 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:

$$\text{MSFE}(h) = \frac{\sum_{t=n}^{n+p-h} (y_{t+h} - \hat{y}_{t+h}(t))^2}{p - (h-1)} \tag{20}$$

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. $p = 504$ hours) of post-sample data and lead times h of 1 up to 48 hours.

Model	Restriction	MSFE(1)	Parameters
HW(24)	na	278.50	2
HW(168)	na	278.04	2
DS($24, 168$)	na	227.09	3
MS($7; 24, 168$)	none	208.45	50
MS($2; 24, 168$)	3	206.45	3

Table 1: Comparison of post-sample forecasts for the utility data

4.2 Comparison of the Full MS Model with HW and DS Models

In the first four lines of Table 1, the post-sample MSFE(1) can be compared for each of the following models: HW(24), HW(168), DS($24, 168$), and full MS($7; 24, 168$). The fifth model MS($2; 24, 168$) with Restriction 3 will be discussed later. The comparison of the first four models shows that the MS($7; 24, 168$) model has the lowest MSFE(1).

In Figure 3, we use the post-sample data to compare MSFE(h) for these four models over longer lead-times; specifically for lead-times of $h = 1$ through 48 hours (i.e. one hour to forty-eight hours). As we did for Table 1, we will delay the discussion of MS($2; 24, 168$) model with Restriction 3 until later. Most of the models are reasonably accurate for shorter lead-times, although the HW(24) model is consistently the worst and shows the greatest deterioration over the longer lead times. None of HW or DS models does as well as the MS($7; 24, 168$) model, especially at lead-times other than those where h is near a multiple of 24.

4.3 Selecting an MS Model

In this section we select the best MS model, based on the procedure for model selection described in section 3.4. 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 sample ($n = 2520$) . 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-hour pattern for each day of the week on the same horizontal axis. In figure 4 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 Figure 4 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 Figure 4 where Friday starts to behave more like Saturday in the evening hours. We list these three choices below.

- $r = 4$ 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. The MSFE(1) values are computed for the withheld time periods $n - q + 1$ to n (i.e. 2017 to 2520) and are shown in Table 2. In Table 2 we say this MSFE(1) compares 'withheld-sample' forecasts to distinguish it from the MSFE(1) in Table 1, which is computed for the p post-sample values (i.e. 2521 to 3024) that are not part of the original sample. The model with the smallest MSFE(1) in Table 2 is the MSFE$(2; 24, 168)$ model. Thus, we choose this model in the first step.

In the second step of the process from section 3.4, we compare Restrictions 1, 2, and 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.

Model	Restriction	MSFE(1)	Parameters
MS($7;24,168$)	none	234.72	50
MS($4;24,168$)	none	239.67	17
MS($4(2);24,168$)	none	250.34	17
MS($2;24,168$)	none	225.51	5
MS($2;24,168$)	1	246.51	2
MS($2;24,168$)	2	234.49	2
MS($2;24,168$)	3	225.31	3

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

4.4 Forecasting with the MS, HW and DS Models

In general, the MS models provide better point forecasts than the HW and DS models. To see this, we return to the examination of Table 1 and Figure 3, both of which contain values for post-sample forecasting. In Table 1, the model with the smallest MSFE(1) is the MS($2;24,168$) model with Restriction 3, and MS($7;24,168$) is second best. The MS($2;24,168$) model with Restriction 3 also has far fewer parameters to estimate than MS($7;24,168$); 3 rather than 50.

Figure 3 shows MSFE(h) values for post-sample lead-times of 1 hour to 48 hours (i.e. $h=1,...,48$). The MSFE(h) values for the MS models are consistently lower than for the HW and DS alternatives with the values for the MS model chosen by our selection process being much lower. The more accurate forecasts are the result of the MS models offering a more realistic structure to capture changes in seasonality. An interesting feature of Figure 3 is the way in which the models have clearly lower MSFE(h) values when h is a multiple of 24. Thus, forecasts are generally more accurate for all methods when they are made for the same time of day as the last observation.

Figure 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 eight hours of the post-sample period because the intervals become extremely wide as the time horizon increases. Furthermore, temperature becomes an important factor in utility demand when forecasting more than a few hours into the future. As noted in the introduction, explanatory variables can be added to state space models. During the eight hour period in Figure 5, the forecasts are very good. The 80% prediction intervals were calculated via simulation.

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

Table 3: Utility data smoothing parameter estimates

4.5 Further Comments

The parameter α was estimated to be between 0 and 1, and this constraint was binding in most cases (i.e. $\hat{\alpha} = 1$). Note that an $\alpha = 1$ corresponds to a purely seasonal model for the differences. 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. However, if one's goal is to forecast more than a few hours ahead, it would be appropriate to estimate the parameters by minimizing the sum of squared h-step-ahead errors instead of the usual one-step-ahead errors. Table 3 shows the effect on the estimates for the parameters when the estimation criterion is altered. When the sum of squares is minimized for 24-step-ahead or 168-step-ahead errors, the estimate for α is 0. This smaller value for $\hat{\alpha}$ will reduce the width of the prediction intervals at the longer lead times.

In addition to examining the utility data in Figure 2 when deciding that additive seasonal models were appropriate, we tested the multiplicative MS($7; 24, 168$) model in (19) 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 An Application to Traffic Data

In this section we investigate an application of the MS approach to hourly vehicle counts and compare forecasts with those from HW and DS approaches.

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 Figure 6. The observation series has missing values. The gaps in the data are for periods of days (i.e. multiples of 24) and can be handled using unconditional updating of the states. When y_t is not observed, the error cannot be calculated. The error is still unknown and governed by an $N(0, \sigma^2)$ distribution. The best prediction of the uncertain error is then the mean of its distribution, namely 0. Hence we use $e_t = 0$ to predict the next state vector from the old state vector using the system equation. Such an approach can be applied to any single source of error model. 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. 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 ($h = 1, ..., 336$), 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;

- $r = 3$ 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.

Model	Restriction	MSFE(1)	Parameters
MS($7;24,168$)	none	498.31	50
MS($4;24,168$)	none	428.88	17
MS($3;24,168$)	none	394.42	10
MS($2;24,168$)	none	308.84	5
MS($2;24,168$)	1	310.94	2
MS($2;24,168$)	2	333.85	2
MS($2;24,168$)	3	310.94	3
MS($2;24,168$) public holidays	none	228.68	5

Table 4: Comparison of withheld-sample forecasts for the traffic data

In Table 4 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 5, the post-sample MSFE(1) can be compared for each of the following six models: HW(24), HW(168), DS($24,168$), full MS($7;24,168$) (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 one hour, offering a reduction of approximately 15% in MSFE over the HW and DS models.

In Figure 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 hours. The values of MSFE(h) when $h>1$ in this figure give a different ordering to forecasting accuracy than those in Table 5. The estimate of γ_{d_1} when $m_1=24$ for DS($24,168$) is effectively zero, meaning it is

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

Table 5: Comparison of post-sample forecasts for the traffic data

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 2 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 Figure 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 partially 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 single source of error models that allow us to forecast time series with either additive (linear) or multiplicative (nonlinear) 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. For the additive case, we have looked at this relationship with the HW and DS methods explicitly in two ways; though the reduced ARIMA form of the MS model and though the use of indicator variables. In this development, we have provided single source of error models for HW, DS, and MS approaches. The parameter estimates for our MS model can be found using a Kalman filter or by the more intuitive 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 allows 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 were easily 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.

A First-Order Form of the Model

The MS($r; m_1, m_2$) model can be written in first-order form where the state variables are lagged by only one period in the state transition equation:

$$y_t = \boldsymbol{H}_t \boldsymbol{a}_{t-1} + \varepsilon_t \tag{21a}$$

$$\boldsymbol{a}_t = \boldsymbol{F} \boldsymbol{a}_{t-1} + \boldsymbol{G}_t \varepsilon_t \tag{21b}$$

where $\boldsymbol{a}_t$ is the $1 \times (rm_1 + 2)$ state vector containing level, trend and seasonal terms:

$$\boldsymbol{a}_t = (\ell_t, b_t, s_{1,t}, \ldots, s_{1,t-m_1+1}, s_{2,t}, \ldots, s_{2,t-m_1+1}, \ldots, s_{r,t}, \ldots, s_{r,t-m_1+1})'$$

.

$\boldsymbol{H}_t$ is a $1 \times (rm_1 + 2)$ row vector containing values of 1 and 0 (depending on which subcycle t corresponds to):

$$\boldsymbol{H}_t = (1, 1, 0, \dots, 0, x_{1t}, 0, \dots, 0, x_{2t}, 0, \dots, 0, x_{rt})$$

$\boldsymbol{F}$ is a block-diagonal $(2 + rm_1) \times (2 + rm_1)$ matrix of the form:

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}_\ell & \vdots & \boldsymbol{0} \\ \dots & \vdots & \dots \\ \boldsymbol{0} & \vdots & \boldsymbol{F}_s \end{pmatrix}$$

where

$$\boldsymbol{F}_{\boldsymbol{\ell}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is the level and trend component of $\boldsymbol{F}$. The seasonal component is the $rm_2 \times rm_2$ matrix defined by:

$$\boldsymbol{F}_s = \boldsymbol{I} \otimes \boldsymbol{F}_1$$

where $\boldsymbol{I}$ is the $r \times r$ identity matrix and $\boldsymbol{F}_1$ is the $m_2 \times m_2$ matrix of the form

$$\boldsymbol{F}_1 = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \tag{22}$$

$\boldsymbol{G}_t$ is a $(2 + rm_2) \times 1$ vector, the values of which are determined by $\boldsymbol{\Gamma}$, α, β and $\boldsymbol{x}_t$:

$$\boldsymbol{G}_{\boldsymbol{t}} = \begin{pmatrix} \alpha \\ \beta \\ \sum_{i=1}^r \gamma_{1i} x_{it} \\ 0 \\ \vdots \\ \sum_{i=1}^r \gamma_{2i} x_{it} \\ 0 \\ \vdots \\ \sum_{i=1}^r \gamma_{ri} x_{it}) \\ \vdots \\ 0 \end{pmatrix}$$

The first-order form of the model can be estimated using general state space methods such as the Kalman filter (Snyder 1985) or by exponential smoothing (Ord et al. 1997).

B The MS($r; m_1, m_2$) Model in Reduced Form

Consider the MS$(r; m_1, m_2)$ model in (11). We restrict attention to the case where $r = \frac{m_2}{m_1}$. This does not entail any loss of generality because all cases in which the number of distinct cycles of period m_1 is less than $\frac{m_2}{m_1}$ can be considered as special cases of the case $r = \frac{m_2}{m_1}$ with some equality constraints on the smoothing parameters and seed values.

The non-seasonal part of the series is a "local linear trend" process, which is an ARIMA(0,2,2) , see e.g. Snyder(1985). Hence, the important thing to establish is the reduced form of the seasonal component. Since $s_{i,t-m_1}$ rather than $s_{i,t}$ appears in the y_t equation, We start by lagging equation (11d) m_1 periods. Then the following is true for $i = 1, ..., r$.

$$s_{i,t-m_1} = s_{i,t-2m_1} + (\sum_{j=1}^{r} \gamma_{ij} x_{j,t-m_1}) \varepsilon_{t-m_1}$$

Repeated substitution r times leads to:

$$\begin{aligned} s_{i,t-m_1} &= s_{i,t-m_2-m_1} + (\sum_{j=1}^{r} \gamma_{ij} x_{j,t-m_1}) \varepsilon_{t-m_1} + (\sum_{j=1}^{r} \gamma_{ij} x_{j,t-2m_1}) \varepsilon_{t-2m_1} + \cdots \\ &\quad + (\sum_{j=1}^{r} \gamma_{ij} x_{j,t-(r-1)m_1}) \varepsilon_{t-(r-1)m_1} + (\sum_{j=1}^{r} \gamma_{ij} x_{j,t}) \varepsilon_{t-m_2}. \end{aligned}$$

The last term has $x_{j,t}$ rather than $x_{j,t-m_2}$ because $x_{j,t} = x_{j,t-m_2}$. For each j, one and only one of the r dummy variables $x_{j,t}, x_{j,t-m_1}, ..., x_{j,t-(r-1)m_1}$ is equal to one and the rest are zero, and as j changes, a different one of these indicator variables switches to one. Hence the r terms $(\sum_{j=1}^{r} \gamma_{ij} x_{j,t-m_1})$, $(\sum_{j=1}^{r} \gamma_{ij} x_{j,t-2m_1})$, ..., $(\sum_{j=1}^{r} \gamma_{ij} x_{j,t})$ are a circular backward rotation of $\gamma_{i1}, \gamma_{i2}, ..., \gamma_{ir}$, the r^{th} row of the matrix of smoothing parameters. An example of such backward rotation would be $\gamma_{i2}, \gamma_{i1}, \gamma_{ir}, \gamma_{ir-1}, ..., \gamma_{i4}, \gamma_{i3}$. Depending on which sub-cycle t belongs to, the rotation starts from a different point. However, since $s_{i,t-m_1}$ is added to y_t only when $x_{i,t} = 1$, i.e. when t belongs to sub-cycle i, the relevant rotation starts from $\gamma_{i,i-1}$ (or $\gamma_{i,r}$ if $i = 1$) and circles back and ends with $\gamma_{i,i}$.

Hence, we have

$$(1 - L^{m_2})\, s_{i,t-m_1} = \sum_{j=1}^{r} \gamma_{ij}^{c} \varepsilon_{t-jm_1} \quad \text{for } t \in \text{sub-cycle } i \tag{23}$$

where $\gamma_{i1}^c, ..., \gamma_{ir}^c$ is the particular backward rotation of $\gamma_{i1}, \gamma_{i2}, ..., \gamma_{ir}$ described above and L is the 'lag' or the 'backshift' operator. This shows that each of the r seasonal factors is a seasonal ARIMA $(0,1,0)_{m_2} \times (0,0,r-1)_{m_1}$. Using equation (23) and noting that $x_{i,t} = x_{i,t-m_2}$, the seasonal component of y_t can be written as:

$$\begin{aligned}
\sum_{i=1}^{r} x_{i,t} s_{i,t-m_1} &= \sum_{i=1}^{r} x_{i,t} s_{i,t-m_1-m_2} + \sum_{i=1}^{r} x_{i,t} \sum_{j=1}^{r} \gamma_{ij}^{c} \varepsilon_{t-jm_1} \\
&= \sum_{i=1}^{r} x_{i,t-m_2} s_{i,t-m_1-m_2} + \sum_{j=1}^{r} \left(\sum_{i=1}^{r} x_{i,t} \gamma_{ij}^{c} \right) \varepsilon_{t-jm_1} \\
&= \sum_{i=1}^{r} x_{i,t-m_2} s_{i,t-m_1-m_2} + \sum_{j=1}^{r} \theta_{j,t} \varepsilon_{t-jm_1}
\end{aligned}$$

where $\theta_{j,t} \equiv \sum_{i=1}^{r} x_{i,t} \gamma_{ij}^{c}$. This shows that the seasonal component in y_t is a seasonal ARIMA $(0,1,0)_{m_2} \times (0,0,r-1)_{m_1}$ with periodic moving average parameters. Hence y_t is the sum of an ARIMA(0,2,2) and a seasonal ARIMA $(0,1,0)_{m_2} \times (0,0,r-1)_{m_1}$ with moving average parameters that depend on which sub-cycle t belongs to.

To find the reduced form, we subtract y_{t-m_2} from y_t first:

$$y_t - y_{t-m_2} = \ell_{t-1} - \ell_{t-1-m_2} + b_{t-1} - b_{t-1-m_2} + \sum_{j=1}^{r} \theta_{j,t} \varepsilon_{t-jm_1} + \varepsilon_t - \varepsilon_{t-m_2}$$

Repeated substitution in equation (11b) yields:

$$\ell_{t-1} - \ell_{t-1-m_2} + b_{t-1} - b_{t-1-m_2} = \sum_{j=1}^{m_2} b_{t-j} + \alpha \sum_{j=1}^{m_2} \varepsilon_{t-j}$$

which leads to:

$$\triangle_{m_2} y_t = \sum_{j=1}^{m_2} b_{t-j} + \alpha \sum_{j=1}^{m_2} \varepsilon_{t-j} + \sum_{j=1}^{r} \theta_{j,t} \varepsilon_{t-jm_1} + \triangle_{m_2} \varepsilon_t.$$

If there was no trend in the model, the reduced form would be the above equation without the first term on the right hand side. With the trend, because b_t is integrated, we still have to take an extra round of first differencing to achieve stationarity. Using the facts that $\triangle b_{t-j} = \beta\varepsilon_{t-j}$ and $\sum_{j=1}^{m_2}\varepsilon_{t-j} - \sum_{j=1}^{m_2}\varepsilon_{t-j-1} = \varepsilon_{t-1} - \varepsilon_{t-m_2-1}$, we get

$$\triangle\triangle_{m_2} y_t = \beta\sum_{j=1}^{m_2}\varepsilon_{t-j} + \alpha\triangle_{m_2}\varepsilon_{t-1} + \sum_{j=1}^{r}(\theta_{j,t}\varepsilon_{t-jm_1} - \theta_{j,t-1}\varepsilon_{t-jm_1-1}) + \triangle\triangle_{m_2}\varepsilon_t.$$

This shows that after first and m_2 differencing, y_t is a moving average of order $m_2 + 1$ with non-zero, but periodic, moving average parameters about seasonal lags corresponding to a sub-cycle of period m_1.

Acknowledgement

The authors would like to thank Julie Maley from VicRoads for providing the traffic volume data.

References

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*, New York: McGraw-Hill.

Cottet, R. & Smith, M. (2003), ‘Bayesian modeling and forecasting of intraday electricity load’, *Journal of the American Statistical Association* **11**, 839–849.

Durbin, J. & Koopman, S. J. (2001), *Time series analysis by state space methods*, Oxford statistical science series, Oxford University Press, Oxford.

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., Snyder, R. D. & Grose, S. (2002), 'A state space framework for automatic forecasting using exponential smoothing methods', *International Journal of Forecasting* **18**, 439–454.

Hyndman, R., Koehler, A., Ord, J. & Snyder, R. (2005), 'Prediction intervals for exponential smoothing using two new classes of state space models', *Journal of Forecasting* **24**, 17–37.

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* pp. 1621–1629.

Ord, J., Snyder, R., Koehler, A., Hyndman, R. & 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.* .

Proietti, T. (2000), 'Comparing seasonal components for structural time series models', *International Journal of Forecasting* **16**, 247–260.

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.

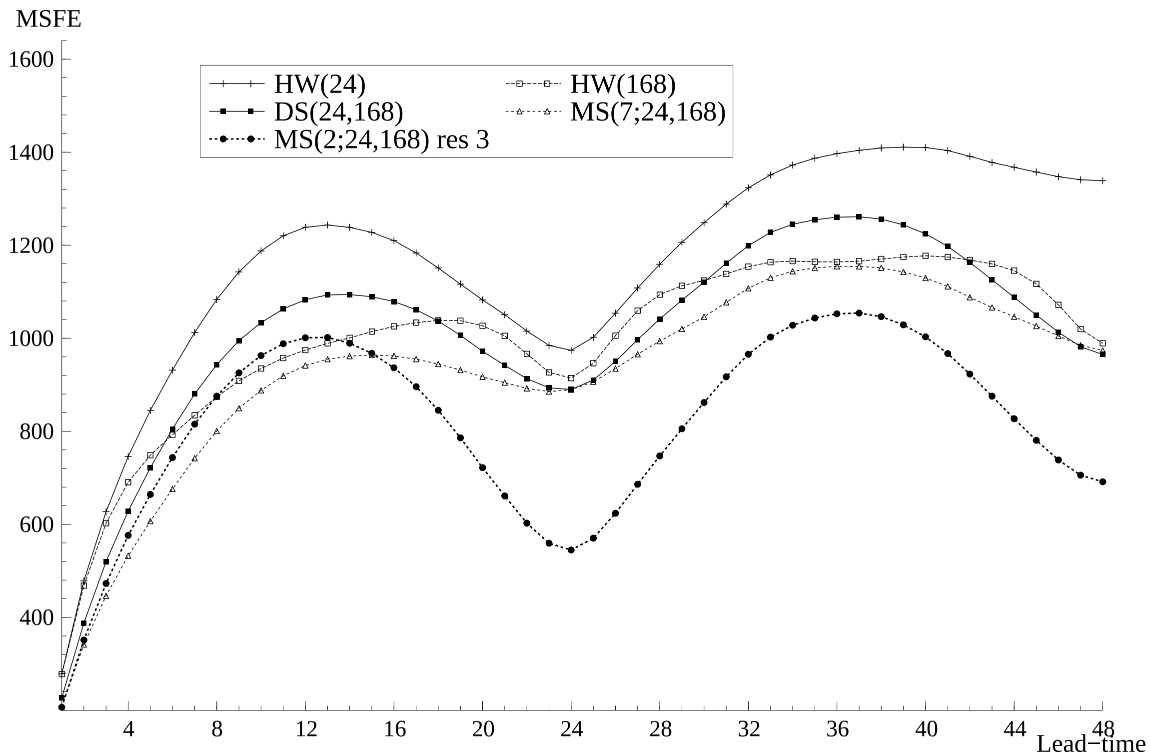

Figure 3: Forecasting accuracy (MSFE) for lead-times from 1 to 48 hours (i.e. 2 days)

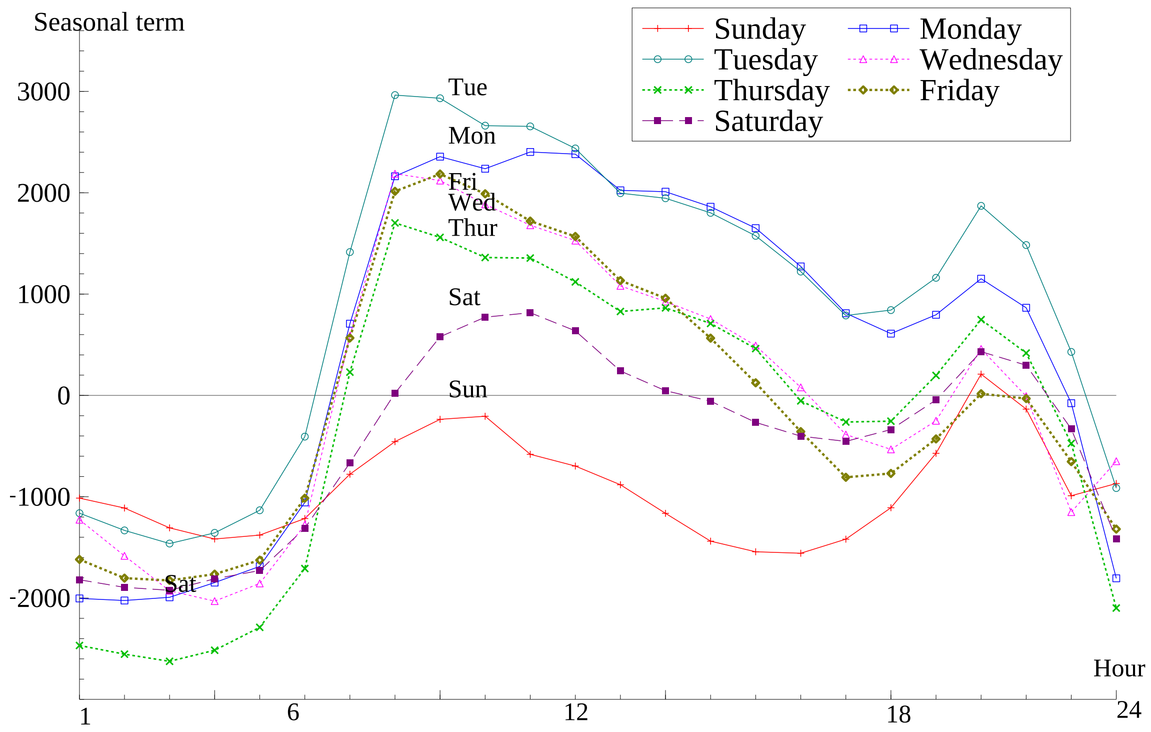

Figure 4: MS($7;24,168$): Hourly sub-cycles by day, based on the last 168 observations ($t = 2353, \ldots, 2520$)

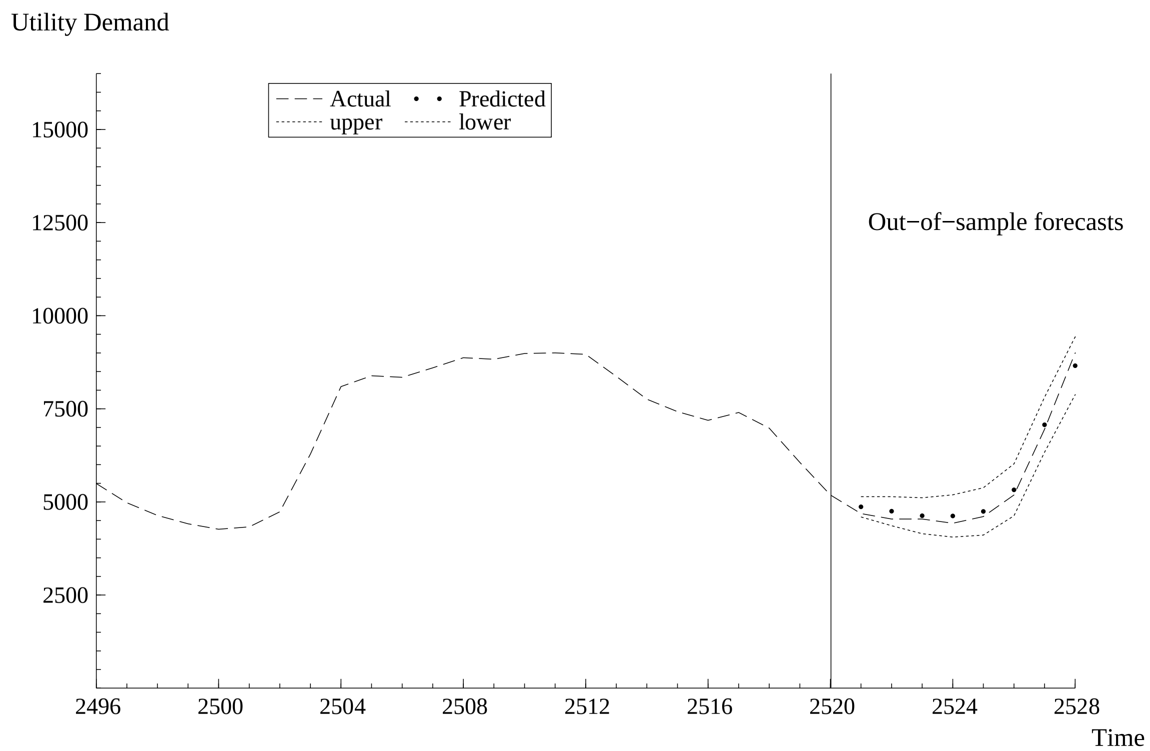

Figure 5: MS($2; 24, 168$) Restriction 3: Point forecasts and 80% prediction intervals for the first 8 hours in the next week ($t = 2521, ..., 2528$)

of the utility demand

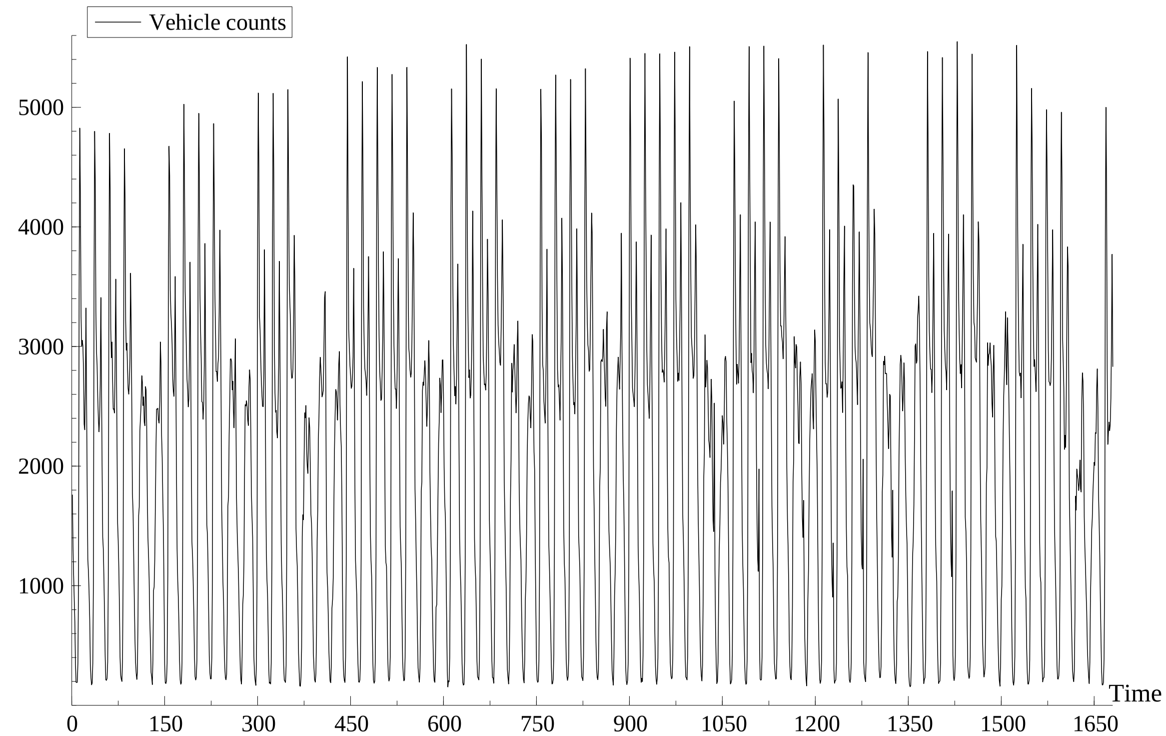

Figure 6: Hourly vehicle counts

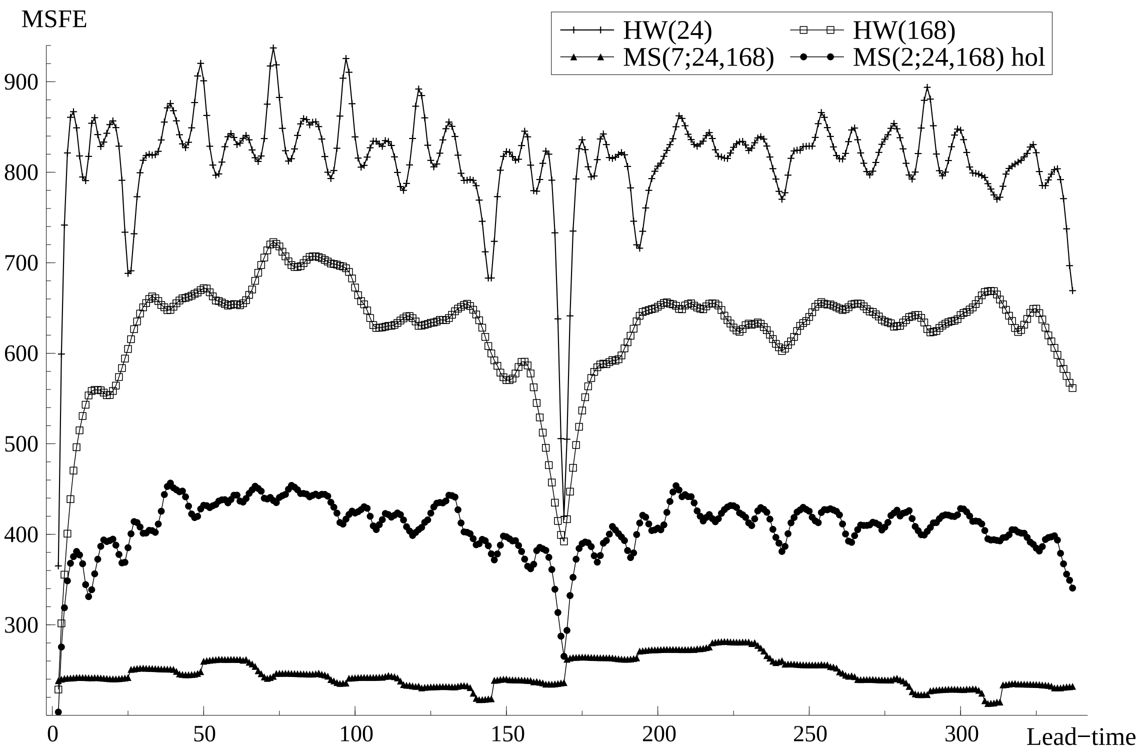

Figure 7: Forecasting accuracy (MSFE) for lead-times from 1 to 336 hours (i.e. 2 weeks)

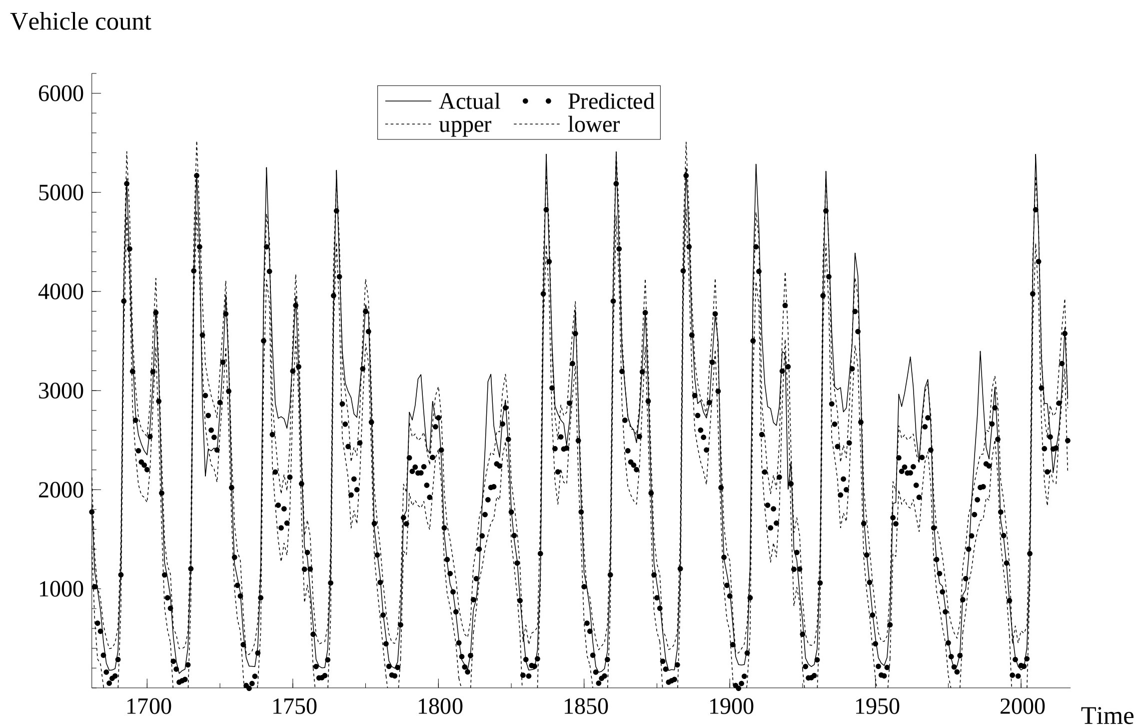

Figure 8: MS$(7; 24, 168)$: Multi-step-ahead point forecasts and 80% prediction intervals for the vehicle counts for each hour in the last two weeks of the evaluation sample ($t = 1681, ..., 2016$)