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 Correlated Seasonality

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

December 2004

Working Paper 28/04

# Forecasting Time-Series with Correlated Seasonality

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

Rob J. Hyndman, [||]

December 7, 2004

**Abstract**

A new approach to forecasting seasonal data is proposed where seasonal terms can be updated using the most recent *relevant* information. It was developed to handle features encountered in hourly electricity load data and daily hospital admissions data. The associated state space model is estimated with methods adapted from exponential smoothing, although the Kalman filter may also be used. It nests existing seasonal models and outperforms them over a range of prediction horizons on the data. The approach is likely to be useful in a wide range of applications involving both high and low frequency data.

*Keywords:* exponential smoothing; Holt-Winters; seasonality; structural time series model (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)

# 1 INTRODUCTION

Time series may contain multiple seasonal cycles of different lengths. The hourly electricity demand shown in Figure 1, for example, exhibits both daily and weekly cycles. This contrasts with archetypical monthly data which normally displays only one cycle.

A particular feature of the data in Figure 1 is that the daily cycles of Monday to Friday are effectively the same, while the daily cycles for Saturday and Sunday are quite distinct. It would appear that the series can be modelled with only three distinct daily cycles instead of a potential seven. A second feature of the data is that the underlying levels of the daily cycles change from one week to the next. Moreover, these weekly movements of the three distinct daily cycles are highly correlated.

Existing approaches to modelling seasonal patterns in the data include the exponential smoothing approach of Winters (1960), the ARIMA approach of Box et al. (1993), and the multiple source of error state space model/Kalman filter approach of Harvey (1989). The first two approaches were not designed for the type of data in Figure 1. While Harvey's approach is able to handle different daily cycles with some common cycles, the approach introduced in this paper will differ in that it relies on a single source of error model. Although the Kalman filter may still be be used (Snyder, 1985), this difference permits exponential smoothing for the estimation, a method that can easily be implemented in a spreadsheet environment. It also allows for a choice of different correlations among the different cycles of the same length and its relationships with other seasonal exponential smoothing methods is more transparent.

The multi-equation regression approach of Cottet and Smith (2003) has been designed to forecast intra day electricity loads. It produces accurate forecasts by allowing for a relationship between load and temperature, while dummy variables are also used to capture day-of-the-week effects. Since, our focus is on developing an approach that transcends electricity demand forecasting, it has to be independent of variables like temperature which may not be relevant in other applications such as hospital admissions.

While Winters approach is not appropriate for the type of data depicted in Figure 1, a method called double exponential smoothing (Taylor (2003)) is suitable. Developed for application to half-hourly utility demand, it involved two unrelated cycles with periods of a day and a week respectively. Being an adaptation of the Winters method to include a second cycle, it led to better forecasts over a range of lead times. Nevertheless, it also leaves room for improvement. The intra day

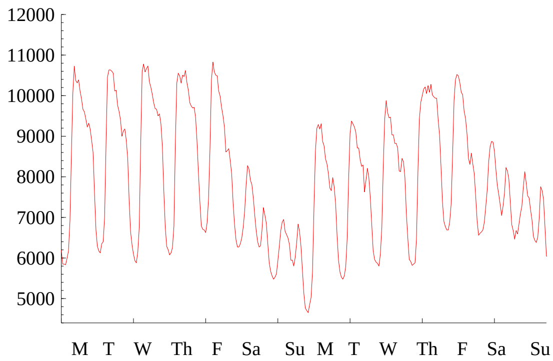



**Figure 1:** Sub-sample of hourly utility data

cycle is the same for all days of the week. Moreover, updates based on recent information (the intra day cycle) is the same for each day of the week.

An extension of the Winters method and double exponential smoothing is introduced in this paper. It allows for *separate* correlated sub-cycles within a larger cycle. To avoid the consequent explosion in quantities to be estimated, it also accommodates the possibility of repeated sub-cycles. Unlike its antecedents, this extension is based on a stochastic specification called here the correlated seasonality model (CSM). It is then possible to project the uncertainty surrounding predictions as well as the predictions themselves.

Higher frequency data is more likely to exhibit strong forms of multiple seasonality (see Figure 1). This is particularly true of data collected more frequently than once per day where time-of-day seasonality may dominate. The concepts behind the CSM, however, can also be used for yearly or monthly cycles, although sufficient data would be required to obtain meaningful estimates of the annual seasonal cycle.

The paper is structured as follows. The additive Winters method and double seasonal method are outlined in Section 2. The CSM is introduced and explored in Section 3. Applications to hourly and daily data are considered in Section 4. Concluding remarks and directions for further research are presented in Section 5.

# 2 MODELLING SEASONAL DATA

## 2.1 Structural Models and the Winters Method

The additive Winters method decomposes the series value $y_t$ into an error $\varepsilon_t$, a level $\ell_t$, a trend $b_t$ and a single seasonal component ($s_t$). It is based on the single-source of error model of Ord et al. (1997).

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t \tag{1a}$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \tag{1b}$$
$$b_t = b_{t-1} + \beta\varepsilon_t \tag{1c}$$
$$s_t = s_{t-m} + \gamma_W\varepsilon_t \tag{1d}$$
$$\varepsilon_t \sim NID(0, \sigma^2) \tag{1e}$$

where $\alpha$, $\beta$ and $\gamma_W$ are smoothing parameters for the level, trend and seasonal terms respectively. The smoothing parameters reflect the level of structural change in a series, and are related to the state variance estimates produced by Kalman filtering using the BSM (Harvey, 1989). The value of $m$ represents the lag length required to model seasonality. The Winters method only allows for one seasonal cycle. It must be assumed that daily cycles are identical ($m = 24$) or that they are totally unrelated ($m = 168$). Estimates of $m + 1$ different seed values for the recurrence relationships must be made. For longer lag lengths, the number of seeds can become prohibitively large. Similar structures to the Winters equations have been applied in a general state space framework by Durbin and Koopman (2001), among others.

## 2.2 Double Seasonal Exponential Smoothing

Double seasonal exponential smoothing (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.

Double seasonal exponential smoothing is a *method*. It was specified without recourse to a stochastic *model* and so it cannot be used in its current form to find estimates of the uncertainty surrounding predictions. The problem is resolved by specifying the model underpinning it. Letting $m_1$ and $m_2$ designate the periods of the two cycles, it is:

$$y_t = \ell_{t-1} + b_{t-1} + s_{1,t-m_1} + s_{2,t-m_2} + \varepsilon_t \tag{2a}$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \tag{2b}$$
$$b_t = b_{t-1} + \beta\varepsilon_t \tag{2c}$$
$$s_{1t} = s_{1,t-m_1} + \gamma_{D1}\varepsilon_t \tag{2d}$$
$$s_{2t} = s_{2,t-m_2} + \gamma_{D2}\varepsilon_t \tag{2e}$$
$$\varepsilon_t \sim NID(0, \sigma^2) \tag{2f}$$

The smoothing parameters for the two seasonal terms are $\gamma_{D1}$ and $\gamma_{D2}$.

The model has a long seasonal cycle consisting of $k = m_2/m_1$ shorter cycles. The short cycles are essentially the same except for shifts caused by structural change through the term $\gamma_{D2}\varepsilon_t$.

# 3 THE CORRELATED SEASONALITY MODEL

## 3.1 Structural Form

The fundamental principle of the CSM is to allow for a diversity of sub-cycles to provide a better fit to the data. At the same time, the number of seed seasonal values is reduced by replacing similar sub-cycles with common seasonal sub-cycles. When modelling the electricity data in Figure 1, for example, there are potentially seven distinct, but correlated cycles: one for each day of the week. A reduction in complexity is achieved by using the same sub-cycle for Monday to Friday.

The existence of common sub-cycles is is the key to reducing the number of seed values compared with both the Winters method and the double seasonal exponential smoothing. As described in Section 2.2, the longer cycle can be broken into $k = m_2/m_1$ shorter cycles of length $m_1$. Of these $k$ possible sub-cycles, $\kappa \leq k$ distinct cycles may be identified. For example, $m_1 = 24$ and $m_2 = 168$ for the hourly data. By assuming that Monday–Friday have the same seasonal pattern then $\kappa = 3$. The number of seed estimates required is reduced from $168$ for the Winters method to $72$. (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 $\kappa$ shorter cycles can be defined as follows:

$$x_{it} = \begin{cases} 1 & \text{if sub-cycle } i \text{ applies in day } t; \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

On any given day, only one of the $x_{it}$ values equals $1$. The one or two seasonal cycles in traditional approaches are replaced by $\kappa$ cycles. Let $\boldsymbol{x_t} = [x_{1t}, x_{2t}, x_{3t}, ..., x_{\kappa t}]'$ and $\boldsymbol{s_t} = [s_{1t}, s_{2t}, s_{3t}, ...s_{\kappa t}]'$.

The general summation form of the model is:

$$y_t = \ell_{t-1} + b_{t-1} + \sum_{j=1}^{\kappa} x_{jt} s_{j,t-m_1} + \varepsilon_t \tag{4a}$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \tag{4b}$$
$$b_t = b_{t-1} + \beta\varepsilon_t \tag{4c}$$
$$s_{it} = s_{i,t-m_1} + (\sum_{j=1}^{\kappa} \gamma_{ij} x_{jt})\varepsilon_t \qquad (i = 1, ..., \kappa) \tag{4d}$$

These equations can also be written in matrix form:

$$y_t = \ell_{t-1} + b_{t-1} + \boldsymbol{x}_t' \boldsymbol{s}_{t-m_1} + \varepsilon_t \tag{5a}$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \tag{5b}$$
$$b_t = b_{t-1} + \beta\varepsilon_t \tag{5c}$$
$$\boldsymbol{s}_t = \boldsymbol{s}_{t-m_1} + \boldsymbol{\Gamma}\boldsymbol{x}_t\varepsilon_t \tag{5d}$$

$\boldsymbol{\Gamma}$ is the seasonal smoothing matrix, which contains the smoothing parameters for each of the cycles.

## 3.2 First-Order Form of the Model

The CSM can be written in first-order form where the state variables are lagged by only one period in the state transition equation:

$$y_t = \boldsymbol{H}_t \boldsymbol{a}_{t-1} + \varepsilon_t \tag{6a}$$
$$\boldsymbol{a}_t = \boldsymbol{F}\boldsymbol{a}_{t-1} + \boldsymbol{G}_t\varepsilon_t \tag{6b}$$

where $\boldsymbol{\alpha}_t$ is the $1 \times (\kappa m_1 + 2)$ state vector containing level, trend and seasonal terms:

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

.

$\boldsymbol{H}_t$ is a $(1 \times (\kappa m_1 + 2))$ row vector containing values of $1$ and $0$ (depending on which subcycle $t$ corresponds to):

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

$\boldsymbol{F}$ is a block-diagonal $((2 + \kappa m_1) \times (2 + \kappa m_1))$ matrix of the form:

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

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 $(\kappa m_2 \times \kappa m_2)$ matrix defined by:

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

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

$$\boldsymbol{F}_1 = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \tag{7}$$

$\boldsymbol{G}_t$ is a $(2 + \kappa m_2) \times 1$ vector, the values of which are determined by $\boldsymbol{\Gamma}$, $\alpha$, $\beta$ and $\boldsymbol{x}_t$:

$$\boldsymbol{G_t} = \begin{pmatrix} \alpha \\ \beta \\ \sum_{i=1}^{\kappa}(\gamma_{1i}x_{it}) \\ 0 \\ \vdots \\ \sum_{i=1}^{\kappa}(\gamma_{2i}x_{it}) \\ 0 \\ \vdots \\ \sum_{i=1}^{\kappa}(\gamma_{\kappa i}x_{it}) \\ \vdots \\ 0 \end{pmatrix}$$

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).

## 3.3 Reduced Form of the CSM

The reduced form of the CSM may be derived from (5) by applying an appropriate transformation to $y_t$ to eliminate the state variables and achieve stationarity. The reduced form of the CSM is

$$\begin{aligned} \Delta\Delta_{m_2}y_t = &\left(\sum_{i=1}^{k}(\theta_{it}L^{i\times m_1} - \theta_{i,t+1}L^{i\times m_1+1})\right)\varepsilon_t \\ &+ \alpha\Delta_{m_2}\varepsilon_{t-1} + \beta\sum_{j=1}^{m_2}L^j\varepsilon_t + \Delta\Delta_{m_2}\varepsilon_t \end{aligned} \quad (8)$$

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:

$$\Delta_{m_2}y_t = \left(\sum_{i=1}^{k}\theta_{it}L^{im_1}\right)\varepsilon_t + \alpha\sum_{j=1}^{m_2}L^j\varepsilon_t + \Delta_{m_2}\varepsilon_t \quad (9)$$

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

For example, for the case with no trend and $m_1 = 4$, $m_2 = 12$ and $k = \kappa = 3$ (no repeating sub-cycles), then (9) can be written as:

$$\Delta_{12} y_t = \left( \sum_{i=1}^{3} \theta_{it} L^{4i} \right) \varepsilon_t + \alpha \sum_{j=1}^{12} L^j \varepsilon_t + \Delta_{12} \varepsilon_t \tag{10}$$

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 $\theta_{3t} = x_{1t}\gamma_{11} + x_{2t}\gamma_{22} + x_{3t}\gamma_{33}$.

The reduced form of the model verifies that the CSM 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 CSM is more logically derived and easily estimated than its reduced form counterpart. The ARIMA reduced forms can be used to show that the models underlying Winters' method and Taylor's double exponential smoothing are special cases of the CSM model and to see the exact relationships among the parameters. By placing restrictions on $\Gamma$ (and hence the $\theta_{it}$ values) both reduced forms are special cases of (8).

## 3.4 Model Restrictions

Because the general form of the CSM nests other seasonal models, some forms of the restricted model are equivalent or similar to the Winters and double Winters methods. In general, the number of smoothing parameters contained in $\mathbf{\Gamma}$ is equal to the square of the number of separate sub-cycles ($\kappa^2$) and is usually quite large. Restrictions can be imposed on $\mathbf{\Gamma}$ to reduce this problem.

The restriction to force common diagonal and off-diagonal elements:

$$\gamma_{ij} = \begin{cases} \gamma_1^*, & \text{if } i = j; \qquad \text{common diagonal} \\ \gamma_2^*, & \text{if } i \neq j. \qquad \text{common off-diagonal} \end{cases} \tag{11}$$

means that $\theta_{1t} = \theta_{2t} = ... = \theta_{\kappa-1,t} = \gamma_2^*$ and $\theta_{\kappa t} = \gamma_1^*$. This implies sub-cycles of different types relate via $\gamma_2^*$ and sub-cycles relate to their own past observations through $\gamma_1^*$ .

When appropriate seed values are chosen, $\kappa = k$ and the restriction of Equation (11) is imposed, the double seasonal Winters method is equivalent to the CSM (identical reduced forms). The models are only equivalent under the assumption of no repeating sub-cycles. The double seasonal Winters method cannot handle repeating sub-cycles.

- **Restriction 1:** $\gamma_1^* = \gamma_W$ , and $\gamma_2^* = 0$
  There is a common diagonal smoothing parameter equivalent to $\gamma_W$, the Winters smoothing parameter. This version of the model is equivalent to the Winters model described in (1) using a lag based on the long cycle ($m = m_2$). For hourly data, this is equivalent to a single seasonal term, updated using lagged values from $168$ hours ago. The cycles are uncorrelated.

- **Restriction 2:** $\gamma_1^* = \gamma_2^* = \gamma_W$
  Assumes all cycles are equally correlated with each other. This version of the model is equivalent to the Winters method based on the short cycle ($m = m_1$), although different days are allowed different seed values.

- **Restriction 3:** Equivalent to Equation (11)
  In general, this is not as restrictive as $1$ and $2$, as it allows for two levels of correlation between sub-cycles. However, there is an equivalent double seasonal Winters form if we assume $\kappa = k$. In this case $\gamma_1^* = \gamma_{D1} + \gamma_{D2}$ and $\gamma_2^* = \gamma_{D2}$.

The CSM allows us to explore a much broader range of assumptions than existing methods, while retaining parsimony. It nests the models underlying the Winters and double seasonal exponential smoothing methods. It contains other restricted forms that stand in their own right. All these possibilities are compared in Section 4 using information criteria to ensure that model complexity is suitably penalized.

## 3.5 Hyper Parameter and Model Estimation

Within the exponential smoothing framework, the CSM can be estimated by minimizing the one-step-ahead sum of squared errors (SSE) $\sum_{i=1}^{n} \varepsilon_t^2$, where $n$ is the number of observations in the series. The SSE is minimized with respect to the smoothing parameters and the seed states. For all models, the sum of the smoothing parameters was restricted to less than two ($\alpha + \beta + |\gamma_{ij}| < 2$). $\alpha$ and $\beta$ were also restricted to positive values. In practice, these restrictions were not binding.

It is necessary to estimate $168$ seasonal seed values when the standard Winters method is applied to hourly data with a weekly seasonal pattern. Such a large number of estimates is undesirable because it becomes a burden on any optimization routine. Often this load can be significantly reduced by seeking and exploiting common sub-cycles between days with similar seasonal patterns. For example, by assuming only two cycles (one for weekdays and one for weekends), estimation of only $48$ seasonal seed values is required, compared with $192$ for the double Winters method.

## 3.6 Model Selection

The various cases of the CSM are fitted to data by minimizing the sum of squared errors. This is equivalent to maximizing the conditional likelihood in the context of linear state space models. Models can then be selected an information criterion, the motivation being to penalize those cases with too many parameters. The exact likelihood is not used because for non-stationary data, Lindley's paradox precludes the use of a criterion such as the AIC.

The definition of the AIC used is

$$\text{AIC} = n \log \left( \frac{\sum_{i=1}^{n} \varepsilon_t^2}{n} \right) + 2p \tag{12}$$

where $p$ is the number of model parameters. Models with lower AIC are preferred. An alternative is the Schwarz Information Criterion (SIC),

$$\text{SIC} = n \log \left( \frac{\sum_{i=1}^{n} \varepsilon_t^2}{n} \right) + \log(n)p. \tag{13}$$

It imposes a larger penalty on over-parameterized models.

A practical way of estimating the CSM is to start with a simple, restricted version of the model, and then relax restrictions on $\mathbf{\Gamma}$ and check their effect on the AIC and SIC. In many cases the improvement in model fit by estimating all elements of $\mathbf{\Gamma}$ will not justify the large increase in parameter numbers. Restrictions $2$ and $3$ from Section 3.4 are a simple way of allowing for correlations without requiring excessive numbers of parameter estimates.

Prediction intervals for multi-step-ahead forecasts can be calculated in the usual way for a state space model (see Section 3.2 for the structural form of the model).

# 4 Empirical Comparisons

## 4.1 Hourly Data

### 4.1.1 The Data

Hourly utility demand data can be used to demonstrate the forecasting and explanatory power of the CSM. The data set consists of $2250$ observations, dating

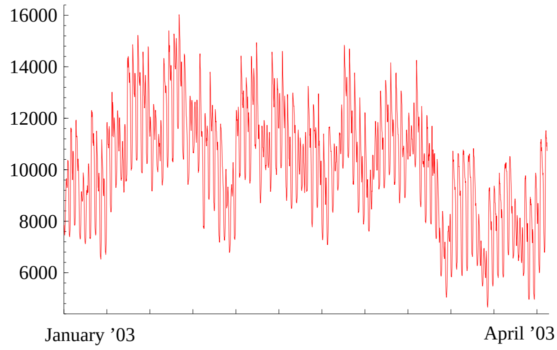



**Figure 2:** Full sample of hourly utility data

| Day of Week | $x_{1t}$ | $x_{2t}$ | $x_{3t}$ | $x_{4t}$ |
|---|---|---|---|---|
| Monday | 1 | 0 | 0 | 0 |
| Tuesday | 1 | 0 | 0 | 0 |
| Wednesday | 1 | 0 | 0 | 0 |
| Thursday | 1 | 0 | 0 | 0 |
| Friday | 0 | 1 | 0 | 0 |
| Saturday | 0 | 0 | 1 | 0 |
| Sunday or public holiday | 0 | 0 | 0 | 1 |

**Table 1:** Dummy variable specification for data and four distinct cycles

from January $2003$ to April $2003$. The full sample is shown Figure 2. The final $250$ observations were held back for evaluation of forecasting accuracy (see Section 4.1.4).

For this data it was decided that Monday–Thursday have a common daily cycle, but Friday, Saturday and Sunday each have their own distinct cycle. It is assumed that public holidays follow the Sunday daily cycle. Only $4$ distinct daily cycles are required. In formulating the model the indicator variable $x_{it}$ is needed to select the relevant daily cycle to predict $y_t$ in the measurement equation. The values for $x_{it}$ are defined by Table 1.

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. 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).

### 4.1.2 The Models

We compare four different model structures:

- **Model 1:** Linear Winters method with seasonal lag of $24$ hours (daily effects dominate).
- **Model 2:** Linear Winters method with seasonal lag of $168$ hours (weekly effects dominate).
- **Model 3:** Linear double seasonal Winters.
- **Model 4:** CSM with $4$ separate patterns (Monday–Thursday has a common pattern).

The models were estimated with some common features. The trend term, $b_t$, was dropped because the data, being collected over a short time interval, displayed no growth. Inclusion of a trend also had the effect of dramatically reducing the accuracy of forecasts for longer lead times.

### 4.1.3 Estimation Results

**Models 1–3** Models $1$–$3$ each had a good fit to the data. The smoothing parameter estimates, together with sum of squared errors, AIC and SIC values, are shown in Table 2. These results indicate that the double seasonal Winters (Model $3$) is the best of the three established methods. It is interesting to note the difference in parameter estimates among the traditional Winters models. Model $1$, which uses a $24$ hour lag, has much higher value for $\alpha$ than does Model $2$. This means that the level term is required to change more rapidly for the $24$ hour lag model. The cause is that the 24-hour model does not account for differences between certain days of the week.

**Model 4** Three different versions of the CSM are considered for Model $4$. They are the CSM under Restriction $2$, the CSM under Restriction $3$ where a separate smoothing parameter is used for the off-diagonal elements of $\mathbf{\Gamma}$, and the unrestricted CSM. The AIC's and SIC's for the three CSM models are presented in

| Model | $\alpha$ | $\gamma_1$ | $\gamma_2$ | SSE | AIC | SIC |
|---|---|---|---|---|---|---|
| 1 | 1.393 | -0.029 | na | $1.19 \times 10^8$ | 21988 | 22005 |
| 2 | 1.075 | 0.080 | na | $1.50 \times 10^8$ | 22451 | 22468 |
| 3 | 1.087 | 0.111 | 0.113 | $9.94 \times 10^7$ | 21637 | 21659 |

**Table 2:** Estimation results for Models 1–3.

| Restriction | $\alpha$ | $\gamma_1^* = \gamma_2^*$ | SSE | AIC | SIC |
|---|---|---|---|---|---|
| Restriction 2 | 1.113 | 0.090 | $7.42 \times 10^7$ | 21050 | 21066 |
| Restriction 3 | 1.117 | na | $7.34 \times 10^7$ | 21028 | 21051 |
| None | 1.124 | na | $7.23 \times 10^7$ | 21027 | 21128 |

**Table 3:** Estimation results for different versions of the CSM (Model 4)

Table 3. For Restriction $3$, $\gamma_1^* = 0.074$ and $\gamma_2^* = 0.099$. For the unrestricted model, similar values were found:

$$\mathbf{\Gamma} = \begin{pmatrix} 0.061 & 0.116 & 0.124 & 0.096 \\ 0.084 & 0.086 & 0.130 & 0.093 \\ 0.084 & 0.126 & 0.091 & 0.097 \\ 0.071 & 0.120 & 0.128 & 0.055 \\ \end{pmatrix}$$

Restriction $3$ gives the lowest SIC (21051), and is the preferred version of the model as it has far fewer parameters than the full model (which has a lower AIC of 21027). Forcing a common value for all elements of $\mathbf{\Gamma}$ does not allow for enough flexibility in the model. The smoothing parameter estimates for Restriction $3$ and the unrestricted CSM both show larger smoothing parameters for off-diagonal elements of $\mathbf{\Gamma}$. This suggests, for example, that the changes in the Saturday pattern are more strongly dependent on recent information (weekday information) than the previous Saturday. This is in keeping with the results of Table 2, which show the $24$ hour lagged version of the Winters model is better suited than the $168$ hour lag version. Figure 3 shows that the CSM provides good out-of-sample forecasts, up to one week ahead.

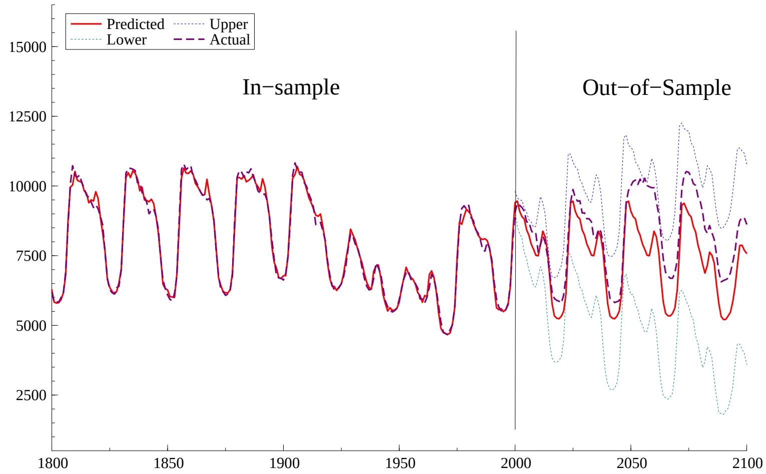



**Figure 3:** In and out-of-sample forecasts for hourly Utility Data with 95% prediction intervals

#### 4.1.4 Comparing MAPES

The forecasting accuracy of the different models is assessed using the mean absolute percentage error (MAPE):

$$MAPE = \sum_{i=1}^{n} \frac{|\varepsilon_i|}{y_i} \times \frac{100}{n} \tag{14}$$

The usual problems associated with using MAPEs with zero-valued data are not encountered due to the nature of the series examined in this paper (although occasional low counts are found).

Because some models are better suited to certain prediction lead times than others, the models are evaluated over a range of lead times. For example, when using hourly data, the Winters model based on a lag of $24$ hours is likely to perform better than the $168$ hour version over shorter lead times. The comparison of MAPEs over a range of lead times should reveal which model is best for certain types of forecasts. The MAPEs presented represent an average of values taken from $250$ different starting points (different hours of the week).

Figure 4 shows the MAPEs for the four competing models between lead times of one hour and one week. Although the graphs are too close to see, the values of the MAPEs show that the $24$ hour lag Winters model outperforms the $168$ hour lag

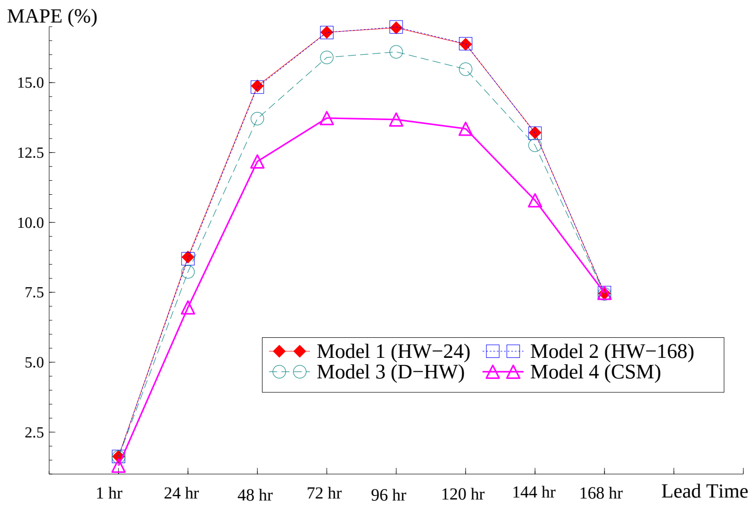



**Figure 4:** Hourly MAPE values for a range of lead times. The values for the two Winters methods are very similar, which means that they have similar accuracy, not that they make the same forecasts.

model over a lead time of up to 24 hours, as one would expect. The CSM significantly outperforms the other models over the full range of lead times. The double seasonal Winters model, which accounts for both daily and weekly dependencies is the next best model for a lead time up to 168 hours. The CSM has the advantage of using all available data from the last week to update seasonal estimates. Figure 4 clearly shows that such information significantly improves forecasting accuracy over a range of lead times.

Another important feature of Figure 4 is the way in which forecasts improve as the lead-time increases beyond 96 hours. All of the models presented contain some sort of dependence on the previous day's observations. When making forecasts that begin on a weekday, forecasted values for a similar type of day at shorter and longer lead-times (i.e. 1 day or 7 days) are likely to be more accurate. For example, suppose a forecast begins on Wednesday. Forecasts made for other weekdays (particularly Wednesdays) will tend to be more accurate than those for weekends, as the last available information is from a weekday. The two Winters methods only allow for one cycle, and as a result forecasts made over lead times involving different types of days are the poorest. The CSM significantly outperforms the other models over these more difficult lead times, as it allows for more subtle dependencies between different days of the week.

## 4.2 Daily Data

### 4.2.1 The Data

In this section the CSM is applied to daily hospital admission data. Figure 15 shows that the data contains more noise and lower counts (around 12 per day) than the hourly utility demand data. The series consists of $1250$ observations, dating from August $1989$ through to December $1992$. The weekly data consists of two levels of seasonality: a weekly pattern and a yearly pattern. No attempt is made to model the yearly seasonality, the purpose being to illustrate the application of the CSM to data with only one seasonal pattern. It is assumed that all seven days of the week have different seasonal patterns, consisting of one observation each. In this case $m_1 = 1$ and $m_2 = 7$.

Removing the level term and ignoring the possibility of similar daily terms (i.e. a separate seasonal effect is used for each of the $7$ days of the week), the updating equations become:

$$y_t = \sum_{j=1}^{7} x_{jt} s_{j,t-1} + \varepsilon_t \tag{15a}$$

$$s_{it} = s_{i,t-1} + \sum_{j=1}^{7} (\gamma_{ij} x_{jt}) \varepsilon_t \tag{15b}$$

Because the seasonal terms are updated using information at time $t-1$, the level term is now redundant, so this version of the CSM can be considered a "level in seasonal" model.

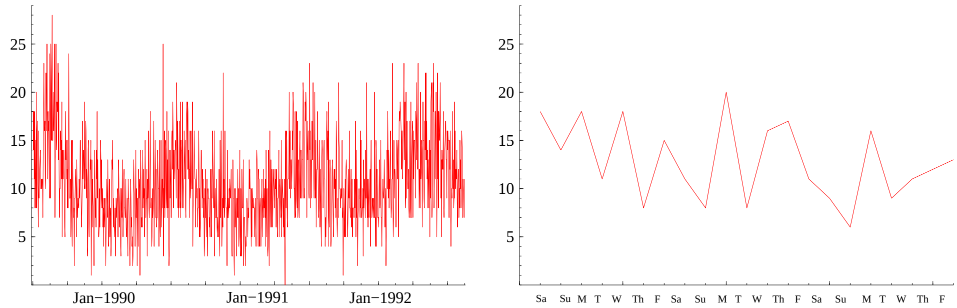
### 4.2.2 The Models

The two models used to forecast daily hospital admissions are:

- **Winters:** Single seasonal Winters with level but no trend (1).
- **CSM:** CSM model (15) using seven daily patterns ($m_1 = 1, m_2 = 7$).

### 4.2.3 Model Estimation

Three versions of the CSM were tested: the full unrestricted model, the model with Restrictions $2$ and and the model with Restrictions $3$. The unrestricted model requires the estimation of $49$ seasonal smoothing parameters.




(a) Full Sample (b) Sub-Sample

**Figure 5:** Full and sub-sample of daily hospital admission data

| Model | $\gamma_1^*$ | $\gamma_2^*$ | $\gamma_W$ | $\alpha$ | SSE | AIC | SIC |
|---|---|---|---|---|---|---|---|
| CSM Unrestricted | na | na | na | na | 15416 | 2833 | 3074 |
| CSM Restriction 2 | 0.091 | 0.091 | na | na | 17078 | 2839 | 2845 |
| CSM Restriction 3 | 0.041 | 0.085 | na | na | 15588 | 2751 | 2760 |
| Winters | na | na | 0.000 | 0.089 | 16694 | 2819 | 2829 |

**Table 4:** Estimation Summary for Daily Hospital Admissions Models

The estimation results (Table 4) show that the Restriction $3$ CSM outperforms the other models in terms of AIC and SIC. The differences between the models are relatively small. No model provides accurate forecasts. All models have one-step ahead MAPES of about $25\%$ to $27\%$. For the unrestricted version of the CSM, the $\gamma$-values along the diagonal are all similar (around $0.03$) as are the off-diagonal values (around $0.09$), showing that the model did not move far from the values imposed by Restriction $3$. The estimate for $\gamma_W$ is zero indicating that the Winters method does not capture any structural change in the seasonal pattern.

#### 4.2.4 Comparing MAPES

As in Section 4.1.4, the models are evaluated over lead times of $1$ day to $1$ week. Because only one form of seasonality is assumed (by ignoring annual effects), the single seasonal Winters model should provide forecasts of similar accuracy over a one week period. The MAPE values presented in Figure 6 are an average of values taken from $250$ different starting points (different days of the week).

Figure 6 shows that the CSM models outperform the Winters model across

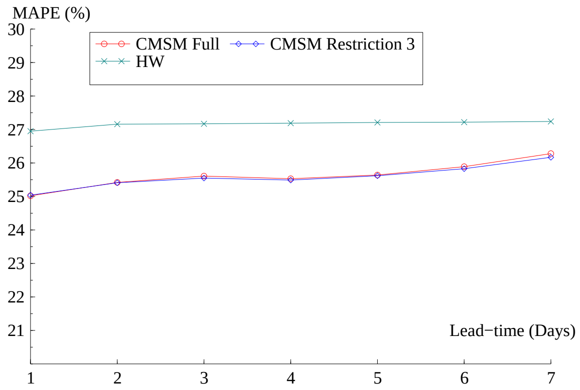



**Figure 6:** Daily Data MAPE values for a range of lead times

all lead times. The difference is smaller than for the hourly data. This is to be expected, as the daily data is modelled with only one form of seasonality. None of the forecasts are very accurate, a reflection of the noise in the data. The results show that the CSM is better in most cases as it is more general.

# 5 Summary and Concluding Comments

Many types of data display one or more seasonal patterns. This is particularly true of high frequency data: daily and weekly cycles become apparent. The Winters method does not allow for multiple seasonal patterns, something that motivated the development of double seasonal Winters forecasting.

A new approach to modelling seasonality was introduced in this paper. Based on what we call the CSM, it breaks the data into a number of correlated seasonal patterns. By allowing for correlations between the seasonal terms, the most recent relevant available information is used. The CSM was estimated by exponential smoothing, an approach that is very simple to implement in a spreadsheet. Its structure, however, can easily be adapted to allow estimation in a general state space or ARIMA environment.

Using hourly data, the CSM was compared to the single and double seasonal versions of the Winters method. The CSM was found to provide significantly better forecasts over a range of lead times. The CSM was able to capture the best features of the other models by providing an updating structure which uses as much recent information as possible, without removing the effects of longer-term patterns. Restricted versions of the CSM gave the lowest SIC values and was preferred over the full model for their parsimony.

The CSM was also evaluated on daily data, which was assumed to contain only one seasonal pattern. The purpose of this was to demonstrate that it can be adapted for lower frequency data. Although not specifically designed for this type of data, it's generality means that forecasts are at least as accurate as from the Winters model (generally slightly better).

The CSM is a general structure which nests the single and double seasonal versions of the Winters method. The generality of the CSM means that a wider range of restrictions can be applied. The empirical evidence presented in this paper shows that structures which can only be estimated within the CSM framework produce the lowest AIC and SIC values. The restrictions implied by the other methods do not match the data, although they provide for simpler reduced forms.

The structure of the CSM can be extended to include covariates. For example, temperature forecasts are widely used to improve load forecasts for electricity data. Such an application should form the basis of future work, to compare the CSM with other, more widely applied methods.

In this paper, the data had one or two levels of seasonality. The CSM can be extended to model more levels of seasonality. Such models would be expected to provide better forecasts over longer lead times.

# References


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

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

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

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

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. and 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.