

Forecasting daily time series using periodic unobserved components time series models

Siem Jan Koopman and Marius Ooms

Free University Amsterdam, Department of Econometrics
De Boelelaan 1105, NL-1081 HV Amsterdam

July 16, 2004

Abstract

We explore a periodic analysis in the context of unobserved components time series models that decompose time series into components of interest such as trend, seasonal and irregular. Periodic time series models allow dynamic characteristics such as autocovariances to depend on the period of the year, month, week or day. In the standard multivariate approach one can interpret periodic time series modelling as a simultaneous analysis of a set of, traditionally, yearly time series where each series is related to a particular season, and the time index is in years. The periodic analysis in this paper applies to a monthly vector time series related to each day of the month.

Particular focus is on forecasting performance and therefore on the underlying periodic forecast function, defined by the in-sample observation weights for producing (multi-step) forecasts. These weight patterns facilitate the interpretation of periodic model extensions. We take a statistical state space approach to estimate our model. In this way we can identify stochastic unobserved components and we can deal with irregularly spaced daily time series. We extend existing algorithms to compute observation weights for forecasting based on state space models with regressor variables.

Our methods are illustrated by an application to a time series of clearly periodic daily Dutch tax revenues. The dimension of our periodic unobserved components model is relatively large as we allow the time series for each day of the month to be subject to a changing seasonal pattern. Nevertheless, even with only five years of data we find that the increased periodic flexibility can help in simulated out-of-sample forecasting for two extra years of data.

1 Introduction

One of the main problems in forecasting seasonal time series is to estimate the recurring but persistently changing patterns within the years. Some patterns are more variable than others and imply different forecasting functions. Fixed patterns can be used for long forecast horizons, whereas variable patterns are only relevant for short term forecasts. Unobserved components (UC) time series models provide a convenient statistical tool to approach this problem. Unobserved components time series models decompose a time series into trend, seasonal and irregular components. For the problem at hand, this approach suits three aims: firstly, it decomposes the observed series into unobserved stochastic processes with component specific error terms, which provide (after estimation) a better understanding of the dynamic characteristics of the series and the changes in the components; secondly, it straightforwardly generates optimal point- and interval forecasts using the Kalman filter; thirdly, the algorithms associated with the Kalman Filter provide the observation weights of the forecasting functions, directly expressing forecasts as functions of past observations.

The analysis in this paper is motivated by the results in Koopman and Ooms (2003), who discussed an unobserved components time series model for daily national Dutch tax revenues, where the main periodic movement is a monthly repeating pattern. Their main focus was on modelling the time series for the 5 days around the turn of the month, when most of the taxes are collected. They developed a model for all 23 monthly time series for the separate banking days of the month. The model for the days with low revenues was based on stochastic cubic splines to interpolate monthly pattern. Some evidence of periodic residual correlation remained. Some of these periodic correlations may be captured by a flexible periodic analysis with separate univariate unobserved component models for each day of the month.

Koopman and Ooms (2002) compared different theoretical approaches to periodic unobserved component time series modelling of seasonal data. The first approach is standard nonperiodic unobserved component time series modelling, where only the mean and variance depend on the season and the time index is in seasons. The second method is fully periodic univariate unobserved component modelling, where in principle all parameters depend on the season. The third approach is univariate unobserved components modelling of seasonal subseries, where the time index is in years and where correlations across the series are unmodelled. In this paper we explore the third method in the context of daily time series modelling. Section 5 discusses other methods which are, as yet, less suitable for daily time series modelling.

Proietti (2004) also explores a periodic analysis in the context of unobserved components models. He focuses on signal extraction for monthly data. He achieves an adequate decomposition of the time series by specific restrictions on periodic disturbance covariances relating periodic trend components for the different seasons one year. We discuss his approach in more detail in Section 5.4 below. We explore different periodic models and concentrate on interpreting the observation weights of forecasts from periodic models adopted for the analysis of daily time series.

In the context of integrated autoregressive moving average models (ARIMA) and dynamic econometric models, extensive studies using the periodic approach are carried out by Osborn and Smith (1989), Osborn (1991) and in the monograph by Franses and Paap (2004). The consequences of a periodic approach for seasonal long-memory or fractional ARMA models have been explored by Ooms and Franses (2001) and Carnero, Koopman, and Ooms (2003).

The main purpose of this paper is to examine the possibilities of a periodic analysis of

seasonal daily time series within the class of unobserved components time series models for nonstationary processes. Periodic unobserved components (PUC) models are implemented using state space methods and empirical results are obtained by estimating and forecasting different models for daily time series. The forecast performances of these models are interpreted using new algorithms to compute observations weights and compared for the example series of daily tax revenues.

The paper is organised as follows. Section 2 describes the series of daily tax revenues and the main periodic characteristics. Section 3 introduces the notation for the standard unobserved component and defines the state space representation. Section 4 describes the computation of weight functions for forecasting in state space models with possible regressor components. Section 5 discusses four periodic unobserved components (PUC) models. Section 6 discusses empirical results for the daily tax series and Section 7 concludes.

2 Daily tax revenues

2.1 Data description

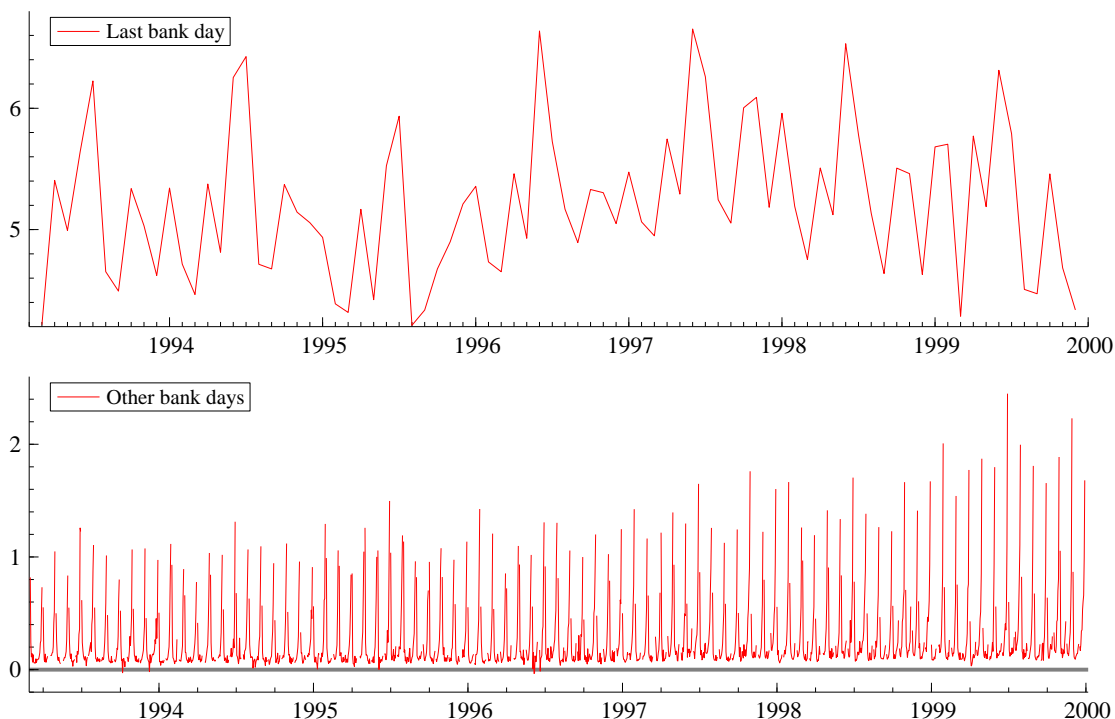
We illustrate daily time series features using a series for Dutch aggregate tax revenues in billions of Euros, described in more detail in Koopman and Ooms (2003). The series starts in March 1993. It contains a (negative) component of tax restitutions up to 1997 which means that values close to zero and even negative values can occur. Tax revenues are only received on bank days: Mondays to Fridays. On a daily basis only total gross tax revenues are available. Yesterday's figures can be used to forecast today's revenues. The Treasury lacks relevant information from tax assessments on taxes that are due on a daily basis, so short term forecasting is based on a pure time series approach.

Figure 1 presents daily Dutch central tax revenues from March 1993 to December 1999. The high revenues on the last bank day of each month stand out and are therefore presented in the top panel on a separate scale. Many taxes are due on the last bank day of the month. The majority of revenues is collected on the last bank day, but the revenues on the four days leading up to this day are also substantial, as seen in the bottom panel. The top panel of Figure 1 also shows a yearly seasonal pattern. June and July display high end-of-the-month revenues compared to other months of the year.

The mean of income clearly depends on the number of bank days that remain until the turn of the month and on the number of bank days after the turn of the month. Tax revenues on the first day of the month are also important. The basic intramonthly pattern in the middle of each month seems similar across the years. The dominating effects are due to the the bank day of the month and to the month of the year. The bank-day-of-the-month effect dominates and is mainly determined by the number of bank days before the turn of the month when most taxes are due. The month-of-the-year effects are partly caused by a quarterly effect from corporate tax revenues. This leads to a higher average for January, April, July and October. In addition there is an effect due to extra salary payments prior to the summer holidays. This is most clearly observed for the month of June.

The original data, indexed by the calendar-day of the month are irregularly spaced because of weekends and holidays. Simply deleting the weekends and holidays and working with bank days does not lead to parsimonious models. Straightforward explanation of the revenues by a smooth function of the calendar-day-of-the-month or the bank-day-of-the-month does not work

Figure 1: Daily Dutch national tax Revenues in billions of Euro



March 1993-December 1999. Top: last bank day of each month. Bottom: other days of the month.

well in that case. The irregular occurrence of holidays and varying numbers of bank days leads to an exaggerated time-variation in the patterns across years, which unnecessarily complicates time series modelling.

2.2 An earlier model

Koopman and Ooms (2003) consider an unobserved components time series model for regularly spaced observations that share the basic pattern within the month. The time distance between two turns of the month has been made constant. This transformation of the time variable enables them to model the data for months with varying numbers and spacing of bank days in a parsimonious way. The transformed variable is called ‘model time’ and denoted by t . The model day-of-the-month index, $p(t)$, runs from 1 to 23. An index $p(t) = 1$ indicates the first bank day of the month. An index 23 indicates the last bank day of the month. A maximum of 5 artificial missing values is introduced around the middle of the month. Apart from these middle sections of each month, the model time data are regularly spaced. In model time, the pattern in the data does not vary as much across years as in calendar time. For example, in model time the penultimate banking day of June is always lagging the previous penultimate banking day of June by $23 \cdot 12 = 276$ days, exactly 12 model months, or 1 model year. The model time index is also used in the lower panel of Figure 1.

A possible unobserved components time series model with intramonthly and monthly effects for daily tax revenues is given by

$$y_t = \mu_t + \psi_t + \lambda_t + x_t' \delta + \varepsilon_t, \quad t = 1, \dots, n,$$

where t is measured in days (model time) and n is the number of days. The unobservable components μ_t , ψ_t and γ_t represent the overall trend, the monthly and the intramonthly effects as stochastic functions of time. The intra-month seasonality is captured by the stochastic cubic spline λ_t whereas the irregular ε_t represents the irregular. The variances of the disturbances driving the components λ_t and ε_t depend on the day of the month. The regression effect, $x_t'\delta$, includes deterministic explanatory variables in the vector x_t and consists of particular dummy indicators for the month of year, the day of week, number of working days in the month, etcetera. A more detailed discussion on unobserved components time series models is presented in Section 3.

The time series of daily taxes contains some extreme outliers due special events in 1999. This paper does not deal with the topic of influential observations and therefore we will be using outlier corrected data. The main outlier correction concerns the effect of an unexpected discontinuation of the banking holiday at the end of April 1999, which most tax payers did not take into account in their planning.

2.3 Periodic descriptive statistics

The analysis in Koopman and Ooms (2003) suggested that even with 5 years of data increased flexibility in their model could improve its forecasting performance. In particular more independently evolving μ_t and ψ_t for different days of the month could be introduced. In practice this can be implemented by specifying and estimating separate trends and seasonals $\mu_{1,t}, \dots, \mu_{23,t}$ and $\psi_{1,t}, \dots, \psi_{23,t}$ with day-of-the month dependent variances, thereby capturing day-of-the-month dependent serial correlation patterns.

A periodic data analysis can reveal which correlations are sufficiently important for modelling. Our periodic analysis is performed as follows. We analyse the daily tax revenue data according to 23 monthly subseries, y_{p,t^*} , where our model-day-of-the-month index, $p(t)$, runs from 1 to 23 and where time t^* is measured in months. The first 9 bank days of each month correspond to $p(t) = 1, \dots, 9$, whereas $p(t) = 15, \dots, 23$ for the last 9 bank days. In our sample each month has at least 18 bank days. Missing observations occur only for $p(t) = 10, \dots, 14$. The year is denoted by $j(t)$ and the month of the year by $s(t)$. Our sample for tax revenues $y_{j(t),s(t),p(t)}$ runs from 1993.3.1 to 1999.12.23. More details on the time transformation are presented in Koopman and Ooms (2003).

These monthly subseries are nonstationary because of trends and a seasonal pattern visible in Figure 1. One can take account of this nonstationarity by conditioning on time trends and seasonal dummies, or by taking monthly or yearly differences. Here we construct yearly differences $\Delta_{276}y_t$, or equivalently, $\Delta_{12}y_{p,t^*}$, $p = 1, 2, \dots, 23$, to make the series stationary. The resulting series run from 1994.3.1 to 1999.12.23.

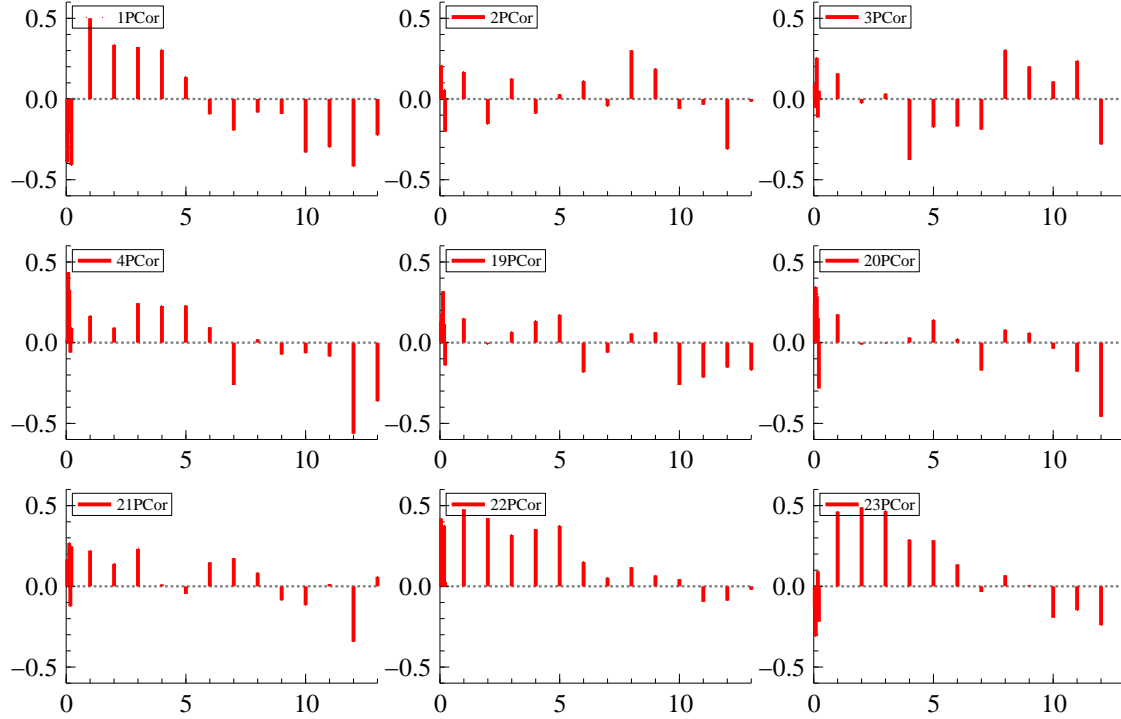
Table 1 presents descriptive statistics for the yearly changes in tax revenues per model day of the month. Figure 2 shows corresponding periodic autocorrelations. It is clear that both variances and autocorrelations depend on the day of the month. Periodic autocorrelations are computed as sample covariances between standardised subseries $(\Delta_{12}y_{p,t^*} - \overline{\Delta_{12}y_p})/\text{s.d.}(\Delta_{12}y_p)$. For example, $r(1)$ for $p = 23$ is the sample covariance between the standardised subseries for $p = 23$ and the standardised subseries for $p = 22$. The variances are higher near the turn of the month. The fifth and sixth column of Table 1 show that short run autocorrelations are negative for the first and last day of the month. Positive shocks in revenues on the last day of the month tend to be followed by negative ones the next day. The short run correlations are positive in

Table 1: *Descriptive statistics yearly changes in tax revenues, $\Delta_{276}y_t$, by model day of the month*

$p(t)$	n_p	lag in months		1 2 3 4 7 8 9 10 11 12 13												
		mean	s.d.	$r(1)$	$r(2)$	$r(23)$	$r(46)$	$r(69)$	$r(92)$	$r(161)$	$r(184)$	$r(207)$	$r(230)$	$r(253)$	$r(276)$	$r(299)$
1	70	10	275	-0.39	-0.21	0.50	0.34	0.32	0.30	0.14	-0.08	-0.09	-0.33	-0.30	-0.42	-0.22
2	70	-6	58	0.21	0.05	0.17	-0.15	0.13	-0.09	0.03	0.30	0.19	-0.06	-0.04	-0.31	-0.02
3	70	0	45	0.09	-0.05	0.16	-0.03	0.03	-0.38	-0.17	0.30	0.20	0.11	0.24	-0.28	-0.09
4	70	2	45	0.02	0.44	0.17	0.09	0.25	0.23	0.23	0.02	-0.07	-0.06	-0.09	-0.56	-0.36
5	70	6	48	0.42	0.21	0.03	0.24	-0.12	-0.09	-0.02	-0.06	0.22	-0.27	-0.12	-0.37	-0.16
6	70	8	39	0.16	-0.13	-0.23	0.02	0.23	-0.31	0.10	0.08	0.15	-0.19	0.30	-0.18	-0.33
7	70	12	51	-0.22	-0.05	0.07	-0.20	-0.01	-0.06	0.03	-0.01	0.13	0.08	0.06	-0.44	-0.10
8	70	12	50	0.08	0.04	-0.10	0.03	-0.19	-0.13	0.06	0.15	0.29	-0.07	-0.15	-0.49	0.02
9	70	15	45	0.12	0.23	-0.16	0.15	-0.08	-0.07	-0.08	-0.05	0.18	-0.12	0.19	-0.33	0.22
10	59	18	86	0.43	0.25	0.10	-0.02	0.15	-0.40	-0.16	0.22	0.27	-0.15	0.05	-0.26	-0.25
11	21	21	57	0.49	0.50	1.53	-0.07	0.08	-0.26	-1.14	0.09	0.23	0.23	0.04	0.28	.NaN
12	6	38	60	0.05	-0.20	.NaN	.NaN	.NaN	.NaN	.NaN	.NaN	.NaN	.NaN	.NaN	0.24	.NaN
13	44	11	61	-0.20	0.19	-0.49	-0.06	0.44	-0.53	0.19	0.43	0.22	-0.41	0.40	-0.08	-0.18
14	68	9	52	0.32	-0.31	-0.10	-0.03	0.10	0.04	0.04	0.01	0.10	-0.03	0.01	-0.42	0.19
15	70	12	62	0.21	-0.13	-0.03	-0.13	-0.17	-0.04	0.00	0.11	0.12	0.04	-0.00	-0.42	0.04
16	70	11	38	0.32	-0.01	-0.10	-0.01	0.08	0.14	0.19	-0.01	0.37	-0.07	-0.04	-0.26	0.26
17	70	9	65	-0.04	-0.22	-0.01	0.06	0.11	0.21	0.14	0.08	0.13	-0.02	0.11	-0.28	0.11
18	70	20	63	0.28	0.34	0.06	-0.02	0.07	0.20	0.23	-0.10	0.08	-0.14	0.03	-0.37	-0.02
19	70	21	71	0.13	0.18	0.15	-0.01	0.07	0.13	0.17	0.06	0.07	-0.26	-0.22	-0.15	-0.17
20	70	46	103	0.25	0.35	0.18	-0.01	-0.00	0.03	0.14	0.08	0.06	-0.04	-0.18	-0.46	-0.08
21	70	45	177	0.17	0.16	0.22	0.14	0.23	0.01	-0.05	0.08	-0.09	-0.12	0.01	-0.34	0.06
22	70	153	213	0.42	0.38	0.48	0.42	0.32	0.35	0.38	0.12	0.07	0.04	-0.10	-0.09	-0.02
23	70	26	420	-0.13	-0.31	0.46	0.49	0.47	0.29	0.29	0.07	0.01	-0.19	-0.15	-0.24	-0.05

Sample: 1994.3.1-1999.12.23. $p(t)$: model day of the month. n_p : number of observations for each day of the month. Mean and standard deviations (s.d) measured in 10^6 Euro. $r(\tau)$: periodic autocorrelation with lag τ measured in model days. Lag $k \cdot 23$ corresponds to a lag of k months. .NaN: no observations to compute these autocorrelations. See Figure 2 for a graphical representation of the numbers.

Figure 2: Selected Periodic Correlations of yearly changes in tax revenues



Horizontal axis: lag in months. Correlations for day of the month 1,2,3,4,19,20,21,22,23 for lags of 1 to 4 days and 1 to 13 Months. Sample: 1994.3.1-1999.12.23. See Table 1 for a graphical representation of the numbers.

the days leading up to the end of the month. Also, the longer run autocorrelations at monthly lags, which are related to dynamics of the seasonal pattern, clearly depend on the day of the month. Autocorrelations at the yearly lag are negative, and vary from -0.5 for $p(t) = 4$ to -0.1 for $p(t) = 22$.

In sum, the descriptive statistics suggest that a periodic time series analysis allowing for different day-of-the-month means, variances and autocovariances is worthwhile. We note that periodicity in the autocorrelation function is also obtained when the subseries of y_t are corrected for deterministic trends and seasonal dummies. Koopman and Ooms (2003) introduced a number of periodic elements into an unobserved components models with a stochastic cubic spline. In this paper we explore a more flexible periodic approach and investigate how this increased flexibility affects the actual forecasting function $y_t = f_t(y_{t-1}, y_{t-2}, \dots)$ and we examine whether this helps in improving forecasting accuracy. We present the flexible periodic model in Section 5, after introducing unobserved components models and the state space form in Section 3 and after discussing the computation of observation weights for forecasting in Section 4.

3 Unobserved components time series models

In this section we set up the notation for unobserved component time series models in state space form, following Harvey (1989) and Durbin and Koopman (2001).

3.1 Basic unobserved components model

The univariate unobserved components time series model that is particularly suitable for many economic data sets is given by

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \quad t = 1, \dots, n, \quad (1)$$

where μ_t , γ_t and ε_t represent trend, seasonal and irregular components respectively. The trend and seasonal components are modelled by linear dynamic stochastic processes which depend on disturbances. The components are formulated in flexible way and they are allowed to change over time rather than being deterministic. The disturbances driving the components are independent of each other. The definitions of the components are given below, but a full explanation of the underlying rationale can be found in Harvey (1989, Chapter 2), who used the term 'Structural Time Series Model' in this context. The effectiveness of structural time series models compared to ARIMA type models, especially when messy features in time series are present, is shown in Harvey, Koopman, and Penzer (1998).

The trend component can be defined as

$$\begin{aligned} \mu_{t+1} &= \mu_t + \beta_t + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2), \\ \beta_{t+1} &= \beta_t + \zeta_t, & \zeta_t &\sim NID(0, \sigma_\zeta^2), \end{aligned} \quad (2)$$

where the level and slope disturbances, η_t and ζ_t , are mutually uncorrelated. When σ_ζ^2 is zero, we have a *random walk plus drift*, and when σ_η^2 is zero as well, a deterministic linear trend is obtained.

To take account of the seasonal variation the component γ_t is included. A deterministic seasonal component should have the property that it sums to zero over the previous year to ensure that it cannot be confounded with the trend. Flexibility of the seasonal component is achieved when it is allowed to change over time. This can be established by adding a disturbance term to the sum of the Q seasonal effects over the past year. This is the stochastic dummy variable form of the seasonal component:

$$S_Q(L)\gamma_{t+1} = \omega_t \quad \omega_t \sim NID(0, \sigma_\omega^2) \quad (3)$$

where $S_Q(L)$ is the summation operator defined as $S_Q(L)\gamma_t = \gamma_t + \gamma_{t-1} + \dots + \gamma_{t-Q+1}$.

Alternatively, a deterministic seasonal pattern is obtained by a set of sine and cosine functions. Allowing these to be time-varying leads to the trigonometric form of the stochastic seasonal component γ_t . The trigonometric seasonal process evolves more smoothly over time than the dummy seasonal process.

3.2 State space representation

The state space form provides a unified representation of a wide range of linear Gaussian time series models including the structural time series model; see, for example, Harvey (1993,

Chapter 4), Kitagawa and Gersch (1996) and Durbin and Koopman (2001). The Gaussian state space form consists of a transition equation and a measurement equation; we formulate the model as in de Jong (1991), that is

$$\alpha_{t+1} = T_t \alpha_t + H_t \varepsilon_t, \quad \alpha_1 \sim N(a, P_0), \quad t = 1, \dots, n, \quad (4)$$

$$y_t = Z_t \alpha_t + G_t \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, I), \quad (5)$$

where $\text{NID}(\mu, \Psi)$ indicates an independent sequence of normally distributed random vectors with mean μ and variance matrix Ψ , and, similarly, $N(\cdot, \cdot)$ indicates a normally distributed variable. The N observations at time t are placed in the vector y_t and the $N \times n$ data matrix is given by (y_1, \dots, y_n) . The standard structural time series model (1)-(2) has a univariate measurement equation: $N = 1$. The periodic models discussed below in Section 5 are based on a multivariate measurement equation with N equal to the number of model days per month. In the following we use a single time index t which may refer to a time scale measured either in model days or in months, or even years, while the number of observations n can refer to the number of model days or the number of months in the sample.

The $m \times 1$ state vector α_t contains unobserved stochastic processes and unknown fixed effects. The initial state vector is assumed to be random with mean a and $m \times m$ variance matrix P but some elements of the state can be diffuse which means that they have mean zero and variance κ where κ is large.

The measurement equation (5) relates the observation vector y_t to the state vector α_t through the signal $Z_t \alpha_t$ and the vector of disturbances ε_t . The initial variance matrix P is given by

$$P = P_* + \kappa P_\infty,$$

where $\kappa \rightarrow \infty$. The nonzero elements of the matrix P_* contain the variances and covariances of the stationary elements of the state vector and P_∞ is a diagonal matrix with ones for nonstationary and deterministic elements of the state and zeroes elsewhere. If $P_\infty \neq 0$ we refer to (4) and (5) as a state space model with (partially) diffuse initial conditions. Finally, the deterministic matrices T_t , Z_t , H_t and G_t are referred to as system matrices and they are usually sparse selection matrices.

The state equation (4) has a Markovian structure to model the serial correlation in the time series y_t . In matrix notation, the univariate observations y_t can be stacked into the vector $y = (y_1, \dots, y_n)'$. To indicate that vector y is modelled by the state space model (4) and (5), we use the notation

$$(y - \xi) \sim \text{ssf},$$

with mean vector $\xi = E(y)$. The variance matrix $\Omega = \text{var}(y)$ is implied by the state space model and depends on the system matrices and the initial state conditions. Since the state space model is Gaussian, it follows that $(y - \xi) \sim N(0, \Omega)$. Note that when the model is (partially) diffuse, that is when $P_\infty \neq 0$, the variance matrix Ω is not properly defined.

3.3 State space formulation of a basic UC model

The unobserved components time series model (1)-(2)-(3) for quarterly series ($Q = 4$), with trend, seasonal and irregular components, requires a state vector of five elements which is given by $\alpha_t = (\mu_t, \beta_t, \gamma_{1,t}, \gamma_{2,t}, \gamma_{3,t})'$. The model specifications for trend and seasonal are given by (2)

and (3), respectively. The state space formulation of the full model relies on time-invariant system matrices, which are given by

$$T_t = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad H_t = \begin{bmatrix} 0 & \sigma_\eta & 0 & 0 \\ 0 & 0 & \sigma_\zeta & 0 \\ 0 & 0 & 0 & \sigma_\omega \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (6)$$

$$Z_t = (1 \ 0 \ 1 \ 0 \ 0), \quad G_t = (\sigma_\varepsilon \ 0 \ 0 \ 0). \quad (7)$$

Here T_t , H_t , Z_t and G_t do not depend on t , so the index t just reflects the standard univariate model index t . Note that the introduction of deterministic time dependent regressors leads to a time varying Z_t in the state space form of the model, as we discuss in the next subsection. The variances of the disturbances are unknown and need to be estimated. They are σ_ε^2 , σ_η^2 , σ_ζ^2 and σ_ω^2 , which we transform to logs, so that we can estimate them without constraints. These unknown parameters are collected in the 4×1 vector φ .

3.4 Regression effects in state space

The basic model (1) can be extended by including fixed regression effects, that is

$$y_t = x_t' \delta + \mu_t + \gamma_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \quad t = 1, \dots, n, \quad (8)$$

where x_t is a $k \times 1$ vector of explanatory variables and δ is a $k \times 1$ vector of coefficients. In matrix notation we may represent model (8) as

$$y = X\delta + u, \quad \xi = E(y) = X\delta, \quad u = y - \xi, \quad u \sim \text{ssf} \Rightarrow u \sim N(0, \Omega),$$

where $X = (x_1, \dots, x_n)'$. The (recursive) generalised least squares estimation of δ can be carried out using state space methods in various ways. More details are discussed in Section 4. It should be noted that when model (8) is used for forecasting, future values of x_t need to be known or need to be computed by special algorithms. Many seasonal effects can be accounted for within the state space model. Other seasonal explanatory variables can be computed by deterministic calendar functions. For example, an algorithm to compute the timing of Easter in every year exists, but is rather complicated. In the case of non-seasonal explanatory variables, future values need to be collected and are considered as given and known.

In several interesting applications, e.g. for time varying cubic splines of Harvey and Koopman (1993), one may want to model changes in the coefficient vector of the regression effects using the random walk specification,

$$\delta_{t+1} = \delta_t + \nu_t \quad \nu_t \sim NID(0, \Sigma_\nu) \quad (9)$$

where ν_t is a $k \times 1$ vector of innovations to the random coefficient vector δ_t and Σ_ν is a $k \times k$ covariance matrix, possibly of a rank smaller than k . It is assumed that ν_t is uncorrelated with the other error terms in the model. Although a time-varying regression component is more difficult to estimate than its constant coefficient counterpart, it does not complicate forecasting as such. This will be explored further in Section 4.

3.5 State space methods

The Kalman filter is a recursive algorithm for the evaluation of moments of the normal distribution of state vector α_{t+1} conditional on the data set $Y_t = \{y_1, \dots, y_t\}$, that is

$$a_{t+1} = E(\alpha_{t+1}|Y_t), \quad P_{t+1} = \text{var}(\alpha_{t+1}|Y_t),$$

for $t = 1, \dots, n$. The Kalman recursions can be used to compute the Gaussian log-likelihood function via the prediction error decomposition for models in state space form, see Schweppe (1965), Jones (1980) and Harvey (1989, Section 3.4). The log-likelihood function is given by

$$\begin{aligned} l = \log p(y_1, \dots, y_n; \varphi) &= \sum_{t=1}^n \log p(y_t|y_1, \dots, y_{t-1}; \varphi) \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n (\log |F_t| + v_t' F_t^{-1} v_t), \end{aligned} \quad (10)$$

where φ is the parameter vector for the state space form (4)-(5). The one-step ahead prediction errors or innovations $v_t = y_t - E(y_t|Y_{t-1})$ and their variances F_t are computed within the Kalman filter for a given vector φ . The innovations have mean zero and are serially independent by construction so that $E(v_t v_s) = 0$ for $t \neq s$ and $t, s = 1, \dots, n$. Note that the summation in (10) is from 1 to n , but usually the first summations will be approximately zero in models with (partially) diffuse initial conditions. The Kalman filter recursions are given in the Appendix. A detailed discussion on the treatment of diffuse initial conditions within the Kalman filter is given in Koopman (1997). When observations y_t for $t = \tau, \dots, \tau^* - 1$ are missing, the Kalman filter is easily adapted to compute

$$a_{\tau^*|\tau-1} = E(\alpha_{\tau^*}|Y_{\tau-1}), \quad P_{\tau^*|\tau-1} = \text{var}(\alpha_{\tau^*}|Y_{\tau-1}).$$

The same principle can be used to generate forecasts and theoretical forecast root mean squared errors since future observations can be regarded as missing values, see Durbin and Koopman (2001) for further details. This simple treatment of missing observations and forecasting is one of the attractions of state space methods in the analysis of daily tax data.

The Kalman filter and likelihood evaluation can also be represented in matrix form. Define vector $v = (v_1, \dots, v_n)'$ and diagonal matrix $F = \text{diagonal}(F_1, \dots, F_n)$. Since the Kalman filter is linear, the prediction error decomposition in matrix form is given by

$$(y - \xi) \sim \text{ssf} \Rightarrow (y - \xi) \sim N(0, \Omega), \quad C(y - \xi) = v, \quad C\Omega C' = F,$$

where matrix C is a unity lower triangular matrix and represents the Kalman filter that effectively carries out a Choleski decomposition of Ω^{-1} , see Durbin and Koopman (2001, Section 4.11). It follows that the loglikelihood function of y is given by

$$\begin{aligned} l &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \{ \log |\Omega| + (y - \xi)' \Omega^{-1} (y - \xi) \} \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} (\log |F| + v' F^{-1} v) \end{aligned} \quad (11)$$

since $|C| = 1$, C^{-1} exists and $\Omega^{-1} = C' F^{-1} C$. The loglikelihood expression is clearly consistent with (10).

The standardised prediction error vector $F^{-\frac{1}{2}}v$ can be used for testing the assumptions underlying the analysis using diagnostic statistics and graphs. Under the assumption that the model is specified correctly, the standardised errors should be normally and independently distributed with zero mean and constant unit variance.

The prediction error v is conditional on the past data only, and useful in an online analysis. For a historical diagnostic analysis of the model, additional information can be obtained from smoothed estimates of disturbances. The smoothed estimate of the sequence of disturbance vectors is the expectation of $\varepsilon = (\varepsilon'_1, \dots, \varepsilon'_n)'$ conditional on all observations $y = Y_n$. It follows from model (4)-(5) that

$$\begin{pmatrix} y \\ \varepsilon \end{pmatrix} \sim N \left\{ \begin{pmatrix} \xi \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega & \text{cov}(\varepsilon, y)' \\ \text{cov}(\varepsilon, y) & I \end{pmatrix} \right\}.$$

By applying a standard lemma of the multivariate normal distribution, we have

$$\hat{\varepsilon} = E(\varepsilon|y) = \text{cov}(\varepsilon, y)\text{var}(y)^{-1}(y - \xi) = \text{cov}(\varepsilon, y)\Omega^{-1}(y - \xi),$$

where $\text{cov}(\varepsilon, y)$ is the covariance matrix of the vectors ε and y . Given the earlier definition of C and F , the smoothed disturbance equals $\text{cov}(\varepsilon, y)e$ where e is defined by

$$e = C'F^{-1}v.$$

The smoothed error e can be computed efficiently using a backward smoothing recursion as detailed in the Appendix. The structure of $\text{cov}(\varepsilon, y)$ is explored in Durbin and Koopman (2001, Section 4.11, p. 98). Note that they use $\hat{\varepsilon}$ for our $G\hat{\varepsilon}$, $\hat{\eta}$ for our $H\hat{\varepsilon}$ and u for our e . The smoothed disturbances, $\hat{\varepsilon}$, are of general interest. However, we show in Section 4.3 that the smoothed error vector, e , is also required for the computation of observation weights for regression estimates in a state space model.

4 Forecast functions and observation weights

4.1 Introduction

It is often of interest to know how observations are weighted when, for example, a forecast of the observation at time $n + 1$ is constructed. It is well known that unobserved component and time-varying coefficient models imply gradually declining observation weights for observations in the past, whereas stationary fixed coefficient models often imply nondeclining weight patterns for past observations. In this section we use a framework in which the forecasting weights of both types of models can be computed and interpreted.

When the model is in state space form, observation weights can be obtained as discussed in Koopman and Harvey (2003). Here we concentrate on the observation weights for y when the aim is to forecast the next observation y_{n+1} given a realisation for y . For the Gaussian linear state space model (4) and (5), the minimum mean square estimator of y_{n+1} is a linear function of realisation y , that is

$$E_y(y_{n+1}|y) = \hat{y}_{n+1|n} = w(\hat{y}_{n+1|n})'y \tag{12}$$

where E_x refers to expectation with respect to density of stochastic variable x and $w(\hat{\theta})$ is an $n \times 1$ vector of weights for the estimator of the quantity of interest θ . Possible quantities for θ are a (future) observation y_t , a state vector α_t , or a disturbance vector ε_t for $t = 1, \dots, n, n + 1, \dots$

4.2 Observation weights for forecasting

In the following we extend the algorithms of Koopman and Harvey (2003) for computing observation weights by considering state space models with explanatory variables. The main motivation is to determine the observation weights for forecasting a set of n_* observations y_* (typically $y_* = y_{n+1}$ and $n_* = 1$) as part of the model given by

$$\begin{pmatrix} y \\ y_* \end{pmatrix} = \begin{pmatrix} X \\ X_* \end{pmatrix} \delta + \begin{pmatrix} u \\ u_* \end{pmatrix}, \quad \begin{pmatrix} u \\ u_* \end{pmatrix} \sim N \left\{ 0, \begin{pmatrix} \Omega & \Omega_* \\ \Omega'_* & \Omega_{**} \end{pmatrix} \right\}, \quad (13)$$

where X is an $n \times k$ matrix of known time dependent regressors representing the deterministic part of y and where δ is a fixed but unknown $k \times 1$ vector of coefficients. The $n_* \times k$ matrix X_* is the equivalent matrix for y_* . The model in (13) can be regarded as the standard linear regression model with “disturbance” u that contains stationary or possibly nonstationary stochastic components of y . For example, u may represent a stationary autoregressive process, *a regression model with AR disturbances*, but it may also account for stochastic trend and seasonal components, *a basic structural time series model with regression effects*. The $n_* \times 1$ vector u_* is associated with y_* and is correlated with u . In a general framework, we model the elements of vector $(u', u'_*)'$ by the state space model (4) and (5) with y replaced by $(u', u'_*)'$, that is

$$\begin{pmatrix} u \\ u_* \end{pmatrix} \sim \text{ssf}.$$

The conditional minimum mean squared error estimator of y_* given y is equivalent to $E(y_*|y)$ for a linear Gaussian model, and can be derived from the standard lemma of the multivariate normal distribution. In the case $\delta = 0$, we have $u = y$ and $u_* = y_*$. It follows that

$$\hat{u}_* = E_u(u_*|y) = E_u(u_*|u) = \Omega'_* \Omega^{-1} u = w(\hat{u}_*) u,$$

where $w(\hat{u}_*) = \Omega'_* \Omega^{-1}$ is the $n_* \times n$ matrix of observation weights for the computation of $\hat{u}_* = E_u(u_*|y)$. In practice, the computation of \hat{u}_* is done via Kalman filtering based on the state space model for $u \sim \text{ssf}$. An algorithm for the direct evaluation of $w(\hat{u}_*)$ is given by Koopman and Harvey (2003).

For known δ , we have

$$\begin{aligned} E_u(y_*|y) &= X_* \delta + \Omega'_* \Omega^{-1} (y - X \delta) \\ &= (X_* - \Omega'_* \Omega^{-1} X) \delta + \Omega'_* \Omega^{-1} y \\ &= (X_* - w(\hat{u}_*) X) \delta + w(\hat{u}_*) y, \end{aligned} \quad (14)$$

where the matrix of weights $w(\hat{u}_*)$ is the same as for the case of $\delta = 0$. For the case δ is fixed and unknown, we replace it by the minimum mean squared error linear estimator $\hat{\delta}$, given by the generalised least squares estimator (or Gaussian Maximum Likelihood estimator)

$$\begin{aligned} \hat{\delta} &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \\ &= w(\hat{\delta}) y, \end{aligned} \quad (15)$$

where the $k \times n$ matrix of weights $w(\hat{\delta})$ is given by

$$w(\hat{\delta}) = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1}. \quad (16)$$

The weight matrix for $\hat{y}_* = E(y_*|y) = w(\hat{y}_*) y$ is obtained by

$$w(\hat{y}_*) = (X_* - w(\hat{u}_*) X) w(\hat{\delta}) + w(\hat{u}_*). \quad (17)$$

Methods for computing the regression weight matrix $w(\hat{\delta})$ are discussed in the next section.

4.3 Computing observation weights in regression context

The weight matrix $w(\hat{\delta})$ can be computed in various ways depending on the treatment of the fixed regression effects in the state space model. Direct implementation of formulas like (16) is infeasible, unless n is small as it involves the Choleski decomposition of $n \times n$ matrices which don't have a simple Toeplitz form. In interesting applications to daily data n is larger than 1000 as we need at least three years of data. The application of the Kalman filter, see Section 3.5, provides convenient solutions to this computational problem. One can add the regression effects to the state space form or one can use the augmented Kalman Filter. We discuss these two options in turn.

- *Adding δ to the state vector:*

Regression coefficients can be placed in the state vector α_t and embedded in the state space model (4) and (5), see equation (8) above. For example, consider the local linear trend model with regression effects

$$y_t = \mu_t + x_t' \delta + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2),$$

where the trend component μ_t is given by (2) and x_t is a $k \times 1$ vector of exogenous variables. The state vector is $\alpha_t = (\mu_t, \beta_t, \delta)'$, the transition equation is given by

$$\alpha_{t+1} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I_k \end{bmatrix} \alpha_t + \begin{pmatrix} \eta_t \\ \zeta_t \\ 0 \end{pmatrix}, \quad t = 1, \dots, n,$$

with a fully diffuse initial state vector, and the measurement equation follows immediately from the model. Computing the weights matrix $w(\hat{y}_*)$ of (17) in this way, requires two applications of the Kalman filter. The computation of $w(\hat{\delta})$ is based on the Kalman filter for δ , which is included in the state space model as indicated above. The computation of $w(\hat{u}_*)$ requires the Kalman filter based on a state space model without δ since it should only reflect properties of u consisting of the elements $\mu_t + \varepsilon_t$. The observation weights are then derived from the algorithm to compute filtering weights, obtained by Koopman and Harvey (2003) and implemented in version 3 in the `0x/SsfPack` environment, see the discussion in the Appendix. The same technique can be used if δ follows a multivariate random walk process as defined in (9).

- *Computing δ weights by the augmented Kalman Filter:*

Fixed regression coefficients can also be treated separately from the stochastic elements that are placed in the state vector α_t and the irregular ε_t . In that case we consider again the model in matrix notation as given by (13). The (filtering) weight matrix $w(\hat{u}_*)$ is again obtained from the Kalman filter for $u \sim \text{ssf}$, but now the computation of $w(\hat{\delta})$ is based on the augmented Kalman filter and smoother as developed by de Jong (1991). We can write these computations in a simple form using the notation of Section 3.5. Consider the prediction error decomposition with v the vector of independent prediction errors v_i with variances F_i , as defined in Section 3.5.

$$Cy = v, \quad C\Omega C' = F, \tag{18}$$

where v and F derive from the Kalman filter applied to y . Similarly, we define the $n \times k$ matrix V

$$V = CX, \tag{19}$$

which is computed by applying the Kalman filter to each column of X , exactly as if it was y . By standard matrix algebra it follows that

$$\begin{aligned} w(\hat{\delta}) &= (V'F^{-1}V)^{-1}(V'F^{-1}C) \\ &= (V'F^{-1}V)^{-1}E', \end{aligned} \tag{20}$$

where $E = C'F^{-1}V$. Therefore, the $k \times n$ matrix E can be evaluated by applying the error smoothing algorithm to each column of V . The corresponding weights can be derived from the algorithm to compute observation weights for error smoothing, developed in Koopman and Harvey (2003). In the augmentation approach, one does not need to add k elements to the state. This can be a considerable advantage when k is large and when the diffuse initialisation of δ in the previous algorithm presents numerical difficulties, for example because of collinearity of columns of X in the beginning of the sample. In contrast to the first algorithm, the augmentation approach is not extendible to models with time varying δ as defined in Equation (9) of Section 3.4.

Note that stationarity is not a necessary condition to derive the weights discussed in this section. The matrix Ω can model both trends and periodicities in the variances and covariances of the process. The algorithms have been implemented in the `Ox/SsfPack` environment for which the details are provided in the Appendix. The implementation and the computation work satisfactory in the empirical example of Dutch tax revenues, where $k = 20$ and $n = 1334$. We present those results below in Section 6. First we discuss general periodic unobserved components (PUC) time series models in section 5.

5 Periodic unobserved components time series models

5.1 Motivation

The periodic data analysis of Section 2.3 clearly shows that the autocorrelation function of daily tax revenues varies across the days of the month. In particular, the autocorrelation function of the yearly changes varies at monthly and yearly lags. We also note in time series plots that different trend patterns and changing seasonal patterns emerge for the different days of the month. In Koopman and Ooms (2003) we introduced periodic features to a standard unobserved components model to incorporate some of these periodicities. Here we start with a fully periodic unobserved components approach where all parameters are allowed to depend on the day of the month. We focus in particular on the the independent periodic unobserved components (PUC) model, which is both very flexible and easy to estimate. In the empirical Section 6 we examine whether this flexible model outperforms the existing model in simulated out-of-sample forecasting. As the answer is positive we conclude that the fully periodic unobserved components methodology can be important in practice, even when the sample contains only five years of daily data. In this section we introduce the notation of our fully periodic UC time series model that we apply to the daily Dutch tax revenues and we discuss some extensions that might be relevant for other daily periodic time series.

5.2 Notation

In a periodic analysis the time series are re-ordered in P subseries which we denote by y_{p,t^*} ; it represents observation y_t with time index $t = (t^* - 1)P + p$ for year $t^* = 1, \dots, T^*$ and subperiod $p = 1, \dots, P$ where T^* is the number of observations for the subseries; see Tiao and Grupe (1980). The stack of the P yearly time series is represented by

$$y_{t^*}^* = \begin{pmatrix} y_{1,t^*} \\ \vdots \\ y_{p,t^*} \end{pmatrix}, \quad t^* = 1, \dots, T^*.$$

For the daily tax revenues, t is measured in days, $P = 23$ and t^* is measured in months. In our general notation, $x_{t^*}^*$ refers to the stack of the seasonal values $x_{1,t^*}, \dots, x_{p,t^*}$, that is

$$x_{t^*}^* = \begin{pmatrix} x_{1,t^*} \\ \vdots \\ x_{p,t^*} \end{pmatrix},$$

where x can refer to different variables $\mu, \varepsilon, \beta, \eta, \zeta$ and ω .

5.3 Independent periodic unobserved components (PUC) models

A simple periodic approach is to consider independent basic UC models with stochastic dummy variable seasonality for the P subseries, that is

$$\begin{aligned} y_{p,t^*} &= \mu_{p,t^*} + \gamma_{p,t^*} + \varepsilon_{p,t^*}, \\ \mu_{p,t^*+1} &= \mu_{p,t^*} + \beta_{p,t^*} + \eta_{p,t^*}, & p = 1, \dots, P, \\ \beta_{p,t^*+1} &= \beta_{p,t^*} + \zeta_{p,t^*}, & t^* = 1, \dots, T^*, \\ S_Q(L)\gamma_{p,t^*+1} &= \omega_{p,t^*} \end{aligned} \quad (21)$$

where $S_Q(L)$ is the summation operator defined as $S_Q(L)\gamma_{p,t^*} = \gamma_{p,t^*} + \gamma_{p,t^*-1} + \dots + \gamma_{p,t^*-Q+1}$. For the monthly subseries of daily tax data, the period of the yearly seasonality, Q , is 12. The disturbances ε_{p,t^*} , η_{p,t^*} , ζ_{p,t^*} and ω_{p,t^*} are serially and mutually uncorrelated over both indices. Effectively, P separate basic structural time series models are introduced which can be represented in one multivariate model by

$$\begin{aligned} y_{t^*}^* &= \mu_{t^*}^* + \varepsilon_{t^*}^*, & \varepsilon_{t^*}^* &\sim N(0, \Sigma_\varepsilon), \\ \mu_{t^*+1}^* &= \mu_{t^*}^* + \beta_{t^*}^* + \eta_{t^*}^*, & \eta_{t^*}^* &\sim N(0, \Sigma_\eta), \\ \beta_{t^*+1}^* &= \beta_{t^*}^* + \zeta_{t^*}^*, & \zeta_{t^*}^* &\sim N(0, \Sigma_\zeta), \\ S_Q(L)\gamma_{t^*+1}^* &= \omega_{t^*}^* & \omega_{t^*}^* &\sim N(0, \Sigma_\omega), \quad t^* = 1, \dots, T^*, \end{aligned} \quad (22)$$

where the variance matrices $\Sigma_\varepsilon, \Sigma_\eta, \Sigma_\zeta, \Sigma_\omega$ are restricted to be diagonal.

The simplest example of an independent PUC model is the seasonal random walk, where P is the number of observations per year, and $\Sigma_\varepsilon = 0$, $\Sigma_\eta = \sigma_\eta^2 \cdot I_P$, $\beta_{t^*}^* = 0$, $\gamma_{t^*}^* = 0$. A combination of random walk behaviour for three subseries with one white noise subseries is also a possible in this independent periodic model. In the general case, the number of unknown variance parameters adds up to $4P$ parameters which is 92 in the our case. The periodic forecast function of the model follows from its state space representation, discussed below. From a

practical forecasting point of view one can also interpret the forecasting functions of this model as a statistical approach to additive seasonal Holt-Winters forecasting where the Holt-Winters recursions, see Holt (2004), are applied to the P separate subseries. McKenzie (1984) discussed the relationship between Seasonal Holt-Winters (SHW) procedures and ARIMA modelling in order to give a statistical interpretation to the seasonal Holt-Winters procedure. We compare the forecasts of independent PUC model and Periodic Seasonal Holt-Winters (PSHW) forecasts in Section 6 below.

The variance parameters in the diagonal matrices Σ_ε , Σ_η , Σ_ζ and Σ_ω can be estimated separately for each subseries. The periodic dummy seasonal process can be replaced by periodic versions of the trigonometric seasonal process in a straightforward way.

5.4 Dependent PUC models

If P is not too large one can consider several extensions that allow for nondiagonal Σ_ε , Σ_η , Σ_ζ and Σ_ω in (22). Harvey (1989, Section 8.3) discusses several structures in the context of multivariate structural time series models. *Homogeneous* structural time series models are obtained by replacing the diagonal variance matrix restrictions for model (22) with the homogeneity restrictions

$$\Sigma_\varepsilon = q_\varepsilon \Sigma_*, \quad \Sigma_\eta = q_\eta \Sigma_*, \quad \Sigma_\zeta = q_\zeta \Sigma_*, \quad \Sigma_\omega = q_\omega \Sigma_*. \quad (23)$$

The variance matrices of the disturbances of the different components are therefore equal up to scalar factors q_ε , q_η , q_ζ and q_ω . The variance ratios of trend, slope, and seasonal (normalised on the irregular) are the same for each linear combination of the subseries and the correlation function at lags which are multiples of P do not depend on the index p . This aspect of homogeneous the model is therefore nonperiodic. Periodicity is confined to the innovation variances and to correlations at other other lags. Homogeneous models are used for forecasting multiple time series for the purpose of inventory control of ranges of products. However, such models have not yet been used in the context of periodic time series. As the correlation function of the daily tax revenues is clearly periodic at monthly lags, this series is not well suited for homogeneous periodic modelling.

Model (22) can also be considered without restrictions on the variance matrices. It can then be interpreted as a set of seemingly unrelated time series equations or, more appropriately, seemingly unrelated periodic structural time series equations (SUPSE). All parameters are identified theoretically but in practice the estimation of parameters may be difficult due to an insufficient number of observations and numerical convergence to the maximum loglikelihood can be slow when P is large. This is not feasible for the daily tax data.

A final set of dependent PUC models that can be considered obtains by rank restrictions on the variance matrices in model (22). It is shown by Harvey and Koopman (1997) that rank restrictions on variance matrices Σ_η and Σ_ζ lead to the notion of common trends which can be interpreted as stochastic cointegration for a vector of time series. For PUC models, common trends can occur as common stochastic slopes, common stochastic levels, common yearly seasonality, or as combinations of both.

Proietti (2004) considers related dependent PUC models in the context of monthly data, where $P = 12$. He specifies η_{t*}^* and ζ_{t*}^* as yearly sums of monthly disturbances η_t^* and ζ_t^* . He selects a linear transformation of μ_t^* to represent the “nonperiodic” trend component. Our independent PUC model is not obtained as a special case, as Proietti (2004) specifies

the covariance matrices for the monthly disturbance vector η_t^* , following Harrison and Stevens (1976), whereas we take a fully periodic approach and consider the covariance matrix of the yearly disturbance vector $\eta_{t^*}^*$.

5.5 State space formulation of periodic UC models

A periodic unobserved components (PUC) time series models is represented by a multivariate state space model with $N = P$. In the case of the periodic local linear trend model the state vector is of size $2P$ and is given by $\alpha_{t^*} = (\mu_{t^*,1}, \beta_{t^*,1}, \dots, \mu_{p,t^*}, \beta_{p,t^*})'$ for $t^* = 1, \dots, T^*$. The number of observations T^* for the periodic model will be approximately n/P where n is the number of observations for the nonperiodic time series model. The system matrices for the state space representation (4)-(5) are given by

$$T_{t^*} = I_P \otimes \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad Z_{t^*} = I_P \otimes \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad (24)$$

where I_P is the $P \times P$ identity matrix and \otimes denotes the Kronecker matrix product. The dimension of the state vector therefore equals $2P$ in this periodic model and $(2 + Q - 1) \cdot P$ in model (22). The specifications for H_{t^*} and G_{t^*} vary depending on the type of periodic model.

As mentioned above, in the independent periodic model there is no need to compute all parameters in H_{t^*} and G_{t^*} simultaneously. By the assumption of independence, the information matrix is block diagonal across the parameters for the different equations for y_t^* .

6 A periodic analysis of daily tax revenues

In this section we investigate the performance of the fully periodic unobserved components in comparison to the existing model for Dutch daily tax revenues. We estimate the independent PUC model using period 1993.3.1-1997.12.23 and we produce simulated out-of-sample forecasts for the period 1998.1.1-1999.12.23. We present one-step-ahead observation weights for forecasting tax revenues on selected days in December 1997, both for the fully periodic model and for the existing model. Finally, we compare the simulated one-step-ahead out-of-sample forecasts of the the independent PUC model, the existing model and two benchmark forecasting procedures.

6.1 Independent periodic unobserved components model

We estimate the independent PUC model (22) on 23 monthly time series for the period 1993.3-1997.12 by maximising the state space log likelihood function (10), so that missing values are automatically accommodated. Table 2 presents the estimation results for the independent periodic model (22). These estimates of periodic variances, in particular $\sigma_{\eta,p}$ and $\sigma_{\omega,p}$, and the resulting periodic variance ratios $\sigma_{\eta,p}/\sigma_{\varepsilon,p}$ $\sigma_{\omega,p}/\sigma_{\varepsilon,p}$, reflect the periodic variances and correlation functions for the yearly changes presented in Table 1. The parameters clearly vary across the days of the month. Different stochastic components are of differing relative importance for the different days of the month.

It is not directly obvious what these parameters imply for the forecasting functions. Interpretation of the forecasting implications is easier by looking at the observation weights for the periodic forecast function which we present in Figure 4.

Table 2: *Estimation results independent PUC model 1993.3-1997.12*

p	n_p	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_\eta$	$\hat{\sigma}_\zeta$	$\hat{\sigma}_\omega$	$\hat{\sigma}_\eta/\hat{\sigma}_\varepsilon$	$\hat{\sigma}_\omega/\hat{\sigma}_\varepsilon$
1	58	1.06	9.33	0.00	8.75	8.79	8.25
2	58	4.53	0.01	0.07	0.01	0.00	0.00
3	58	3.05	0.00	0.02	0.01	0.00	0.00
4	58	3.11	0.45	0.04	0.06	0.14	0.02
5	58	3.65	0.35	0.02	0.05	0.10	0.01
6	58	3.05	0.00	0.00	0.02	0.00	0.01
7	58	3.61	0.00	0.00	0.02	0.00	0.01
8	58	3.27	0.00	0.00	0.00	0.00	0.00
9	58	2.83	0.00	0.00	0.01	0.00	0.00
10	51	5.19	0.47	0.00	0.86	0.09	0.17
11	25	2.24	0.16	0.02	0.12	0.07	0.05
12	10	4.69	1.00	0.05	0.22	0.21	0.05
13	43	3.85	0.55	0.01	0.01	0.14	0.00
14	57	3.37	0.41	0.02	0.39	0.12	0.12
15	58	4.36	0.00	0.00	0.00	0.00	0.00
16	58	2.14	0.00	0.05	0.19	0.00	0.09
17	58	2.29	0.09	0.07	1.34	0.04	0.58
18	58	4.78	0.54	0.06	0.08	0.11	0.02
19	58	2.80	2.14	0.00	3.10	0.77	1.11
20	58	7.48	0.93	0.00	0.01	0.12	0.00
21	58	13.23	0.66	0.00	0.63	0.05	0.05
22	58	9.87	0.05	0.29	3.33	0.00	0.34
23	58	15.56	9.40	0.05	9.76	0.60	0.63

Sample: 1993.3.1-1997.12.23. ML estimation of independent periodic model (22). $p(t)$: model day of the month. n_p : number of observations for each day of the month. Standard deviations $\hat{\sigma}_p$ measured in 10^7 Euro.

Figure 4 presents the observation weight distributions for forecasting important days of December 1997, one-step-ahead, obtained by the algorithm of Koopman and Harvey (2003) and discussed in the Appendix. In this case, one-step-ahead means one month ahead. The scale of the horizontal axis in Figure 4 is therefore measured in months. The weights show how the parameter estimates of Table 2 should be interpreted for forecasting. The importance of changes in the seasonal component shows as a relatively large estimate $\hat{\sigma}_\omega$ in Table 2 for $p = 1, 19, 22, 23$ and in Figure 4 as large weights for the observation of exactly one model year before the day to be forecast, in combination with lower weights for observations 2, 3 and 4 years before. In contrast, the very small value for $\hat{\sigma}_\omega$ for $p = 4$ is reflected in large and equal weights for observations exactly 2, 3 and 4 years before, as one would expect in a model with fixed seasonal means. The importance of changes in the level component shows in large estimates for $\hat{\sigma}_\eta$ and large weights for recent months for the forecasts for $p = 1, 19, 23$.

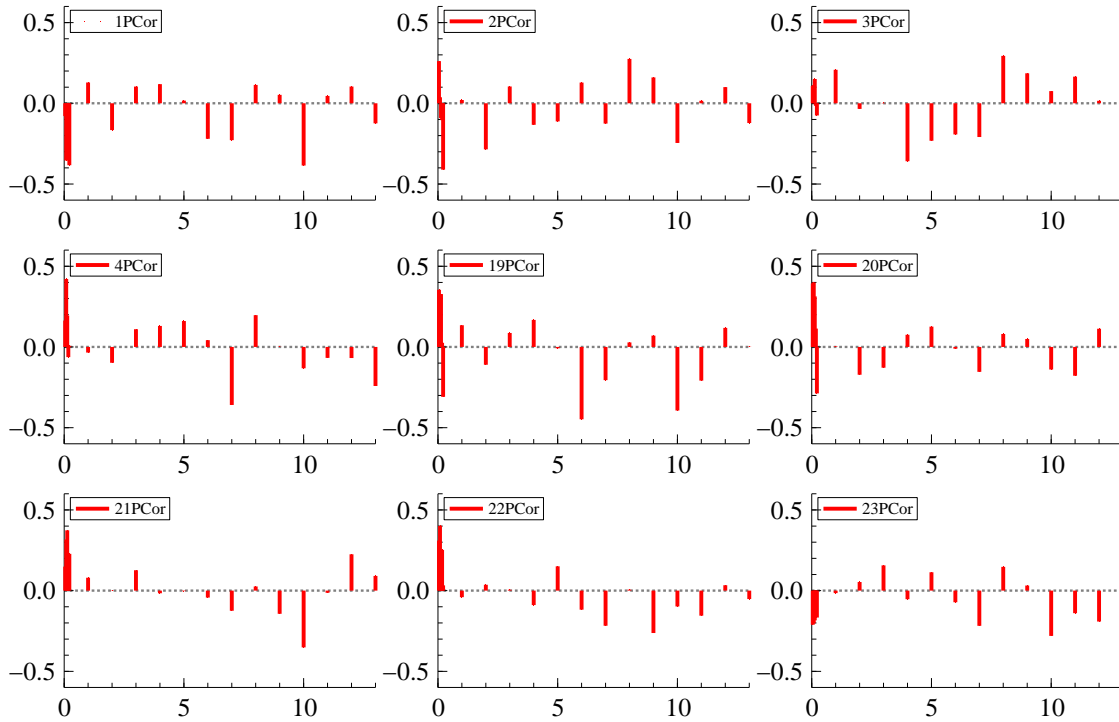
It is worth investigating which part of the periodic correlations in the data are picked up by the independent PUC model. Therefore we reestimate the model for the entire sample and look at the residual periodic autocorrelations. Table 4 presents the periodic autocorrelations for the one-step-ahead forecast errors of 1995.1.1-1999.12.23 for Model (22) reestimated for the entire sample 1993.3.1-1999.12.23.

The estimation results for the full sample are presented in Table 3. Most estimation results are qualitatively similar to the subsample results of Table 2, with two exceptions. First, the importance of changes in the seasonal component, σ_ω , has disappeared for $p = 1$. Second, a stochastic seasonal component appears for $p = 13$. As indicated in Koopman and Ooms (2003), a new monthly pattern seems to arise in the middle of the last three months of the year in the last part of the sample.

As the state vector of this model is of large dimension and diffuse in the beginning of the

sample, we require quite a number of observations to identify the conditional mean of the $y_{p,s}$. Therefore we evaluate the forecast errors starting from 1995.1, i.e. conditional on the first 22 months of data. Figure 3 presents selected periodic autocorrelations in graphical form. This figure can be compared with Figure 2 which shows the periodic correlations for the "raw" yearly changes, i.e. the forecast errors of a seasonal random walk with period 276. At monthly lags the absolute values of the correlations in Figure 3 are clearly smaller than in Figure 2. The correlations for the first daily lags are barely affected by the independent periodic model (22) as correlations between forecast errors of different days within the month are not taken into account. This can be a serious disadvantage for short run forecasting with this model when these short run correlations are well identified and stable over time.

Figure 3: Periodic Correlations of one-step-ahead forecast errors Model (22)



Horizontal axis: lag in months. Correlations for day of the month 1,2,3,4,19,20,21,22,23 for lags of 1 to 4 days and 1 to 13 Months. Sample: 1995.1.1-1999.12.23.

6.2 Evaluating out-of-sample forecasts against alternative methods

Table 5 presents a forecast comparison, building on Koopman and Ooms (2003). It compares the precision of the forecasts of the independent PUC model and a model presented in Koopman and Ooms (2003) as available at the end of 1997. Before discussing the results of the forecast comparison we present the observation weights of the UC model of Koopman and Ooms (2003). As this unobserved components model involves regressors with fixed parameters and regressors with time-varying coefficients, we applied the technique of Section 4.3 to compute the weights.

Figure 5 presents observation weights for the basic parsimonious UC model of Koopman and Ooms (2003), which involves 9 cubic spline regressors with time varying coefficients to capture

Table 3: *Estimation results independent PUC model, full sample*

p	n_p	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_\eta$	$\hat{\sigma}_\zeta$	$\hat{\sigma}_\omega$	$\hat{\sigma}_\eta/\hat{\sigma}_\varepsilon$	$\hat{\sigma}_\omega/\hat{\sigma}_\varepsilon$
1	82	14.51	8.06	0.00	0.01	0.56	0.00
2	82	4.12	0.34	0.05	0.01	0.08	0.00
3	82	3.33	0.00	0.00	0.01	0.00	0.00
4	82	2.97	0.43	0.00	0.00	0.14	0.00
5	82	3.44	0.17	0.00	0.00	0.05	0.00
6	82	2.89	0.01	0.01	0.00	0.00	0.00
7	82	3.56	0.01	0.02	0.01	0.00	0.00
8	82	3.69	0.00	0.01	0.01	0.00	0.00
9	82	3.35	0.11	0.01	1.01	0.03	0.30
10	74	6.85	0.25	0.00	0.00	0.04	0.00
11	38	5.40	0.48	0.02	0.03	0.09	0.00
12	12	4.29	1.00	0.04	0.16	0.23	0.04
13	62	2.73	0.26	0.00	3.81	0.10	1.39
14	81	3.70	0.18	0.00	0.00	0.05	0.00
15	82	4.35	0.01	0.02	0.00	0.00	0.00
16	82	2.27	0.01	0.03	1.19	0.00	0.52
17	82	4.14	0.23	0.07	1.38	0.05	0.33
18	82	4.08	0.02	0.04	1.77	0.00	0.43
19	82	4.29	0.78	0.01	2.74	0.18	0.64
20	82	7.46	0.00	0.05	0.13	0.00	0.02
21	82	11.58	2.65	0.03	0.66	0.23	0.06
22	82	12.73	2.59	0.22	0.37	0.20	0.03
23	82	10.71	11.36	0.01	13.31	1.06	1.24

Sample: 1993.3.1-1997.12.23. ML estimation of independent periodic model (22). $p(t)$: model day of the month. n_p : number of observations for each day of the month. Standard deviations $\hat{\sigma}_p$ measured in 10^7 Euro.

the pattern within the month, one stochastic level to capture the trend, and 11 deterministic regressor variables to capture seasonal effects for particular days of the week, quarter and year. As this model involves regressors with time varying coefficient we used equation (17) to compute forecasting weights, where we put the regression coefficient vector δ in the state. This model is formulated on a daily basis, so the lags on the horizontal axis are measured in model days.

Figure 5 shows that weights for monthly and yearly lags dominate the pattern. The effect of the intramonthly pattern is hardly visible for Day 23 ($p = 23$). Moreover, the pattern differs importantly across days of the month. The pattern for Day 23 is dominated by the effect of the dummy for the ultimate day of June. A high observation for the last day of June increases the value for its coefficient and decreases the coefficient of the end-of-quarter dummy. As the end-of-quarter dummy is very important for predicting the last day of December ($p = 23$), there is a large negative weight of the last day of June for forecasting the last day of December, which shows up by the large negative observation weight at a lag of six months. The corresponding beginning-of-July effect is less relevant for forecasting the first day in December 1997 ($p = 1$). The forecast for Day 4 ($p = 4$) is not significantly influenced by the effect of such special dummy regressors and displays a more standard weight pattern where the first lags receive a large weight and the remaining pattern is dominated by the intramonthly pattern.

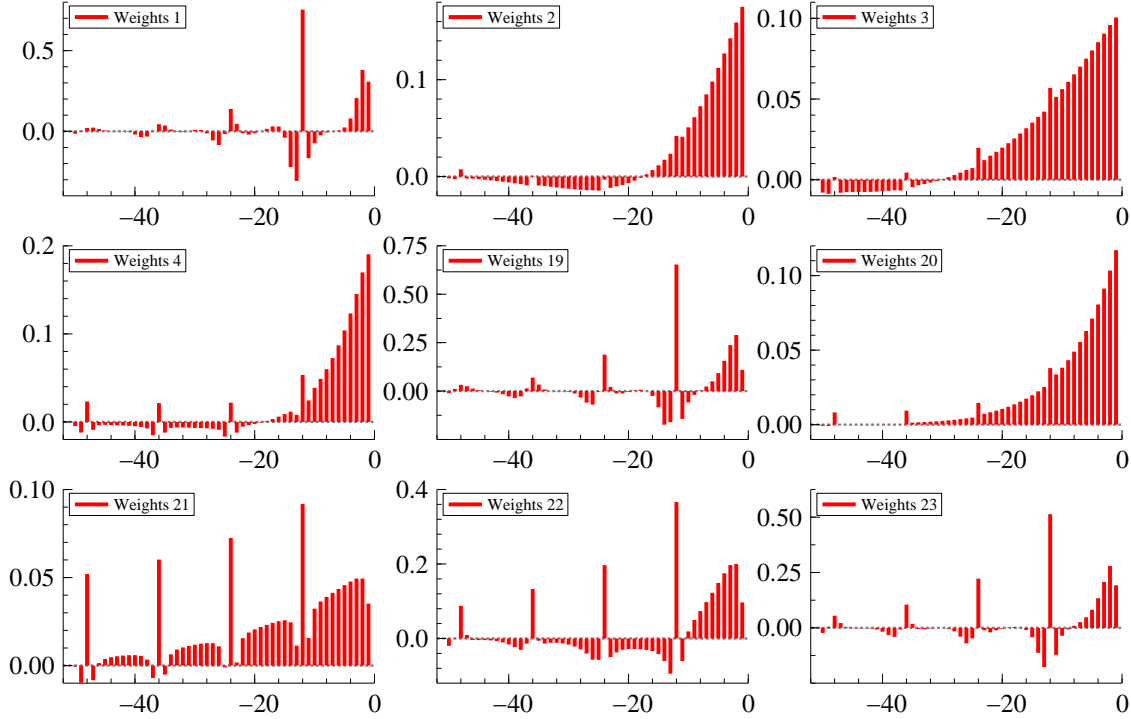
The plots in Figure 5 make clear that one has to take care in interpreting components when a few additional periodic dummy variables are present as they can dominate the forecasting function for related days much more than expected. A simultaneous analysis of the regression part and the purely stochastic part of the model is necessary for a sensible interpretation of periodic models. As the number of observations per day of the month is low, small sample effects can play an important role in the regression, and thereby in the forecasting weights.

Table 4: *Periodic Correlations one step ahead forecasts errors independent PUC Model (22)*

$p(t)$	n_p	lag in months		1 2 3 4 7 8 9 10 11 12 13												
		mean	s.d.	$r(1)$	$r(2)$	$r(23)$	$r(46)$	$r(69)$	$r(92)$	$r(161)$	$r(184)$	$r(207)$	$r(230)$	$r(253)$	$r(276)$	$r(299)$
1	60	7	227	-0.08	-0.35	0.13	-0.17	0.10	0.12	-0.23	0.11	0.05	-0.39	0.05	0.10	-0.12
2	60	8	52	0.26	0.04	0.02	-0.29	0.10	-0.13	-0.13	0.27	0.16	-0.25	0.02	0.10	-0.12
3	60	5	40	0.11	0.10	0.21	-0.03	0.00	-0.36	-0.21	0.29	0.19	0.08	0.17	0.02	-0.08
4	60	2	39	0.16	0.42	-0.03	-0.10	0.11	0.13	-0.36	0.20	0.00	-0.13	-0.07	-0.07	-0.24
5	60	5	41	0.46	0.28	-0.09	0.24	-0.16	-0.17	0.06	0.01	0.11	-0.31	-0.15	-0.02	-0.17
6	60	4	36	0.07	-0.18	-0.33	-0.03	0.29	-0.43	-0.13	0.01	0.04	-0.15	0.18	0.13	-0.39
7	60	3	41	-0.12	0.10	0.07	-0.16	0.23	-0.13	-0.12	-0.12	0.03	-0.18	0.11	0.16	-0.10
8	60	7	44	0.20	-0.02	-0.21	-0.06	-0.08	-0.15	0.01	0.19	0.21	-0.15	-0.20	0.00	0.01
9	60	3	45	0.30	0.29	0.10	0.00	-0.12	-0.04	-0.12	0.01	0.11	-0.19	0.12	0.09	0.32
10	55	5	86	0.44	0.17	0.06	-0.10	0.17	-0.38	-0.28	0.02	0.21	-0.19	0.20	0.29	-0.33
11	26	17	71	0.27	0.35	-0.38	-0.23	-0.20	-0.16	0.52	-0.50	0.09	-0.60	-0.16	0.01	1.24
12	8	-37	76	0.41	0.17	.NaN	.NaN	0.21	.NaN	.NaN	.NaN	0.29	.NaN	.NaN	0.05	.NaN
13	45	7	87	0.13	-0.03	0.09	-0.01	0.07	0.06	-0.08	0.41	0.02	-0.25	0.30	-0.11	0.19
14	59	6	46	0.42	0.18	-0.03	-0.03	-0.06	-0.01	0.00	-0.01	0.20	-0.04	0.02	-0.03	0.19
15	60	8	51	0.23	-0.04	-0.05	-0.12	-0.28	-0.13	0.01	0.04	-0.09	0.01	0.04	-0.06	-0.05
16	60	6	35	0.33	-0.02	-0.12	-0.16	-0.03	-0.01	-0.13	-0.05	0.29	-0.20	-0.18	-0.11	0.28
17	60	3	60	0.04	-0.26	-0.06	-0.10	-0.06	0.15	-0.10	0.06	-0.02	-0.11	0.08	-0.03	0.02
18	60	11	59	0.35	0.34	0.03	-0.08	-0.02	-0.05	-0.19	-0.06	0.04	-0.24	-0.06	-0.01	-0.01
19	60	12	68	0.35	0.27	0.13	-0.11	0.09	0.17	-0.21	0.03	0.07	-0.39	-0.21	0.12	0.00
20	60	11	89	0.39	0.40	0.00	-0.17	-0.13	0.08	-0.15	0.08	0.05	-0.14	-0.18	0.11	0.05
21	60	15	123	0.15	0.32	0.08	-0.00	0.13	-0.02	-0.12	0.03	-0.14	-0.35	-0.01	0.23	0.09
22	60	27	169	0.31	0.40	-0.04	0.04	0.01	-0.09	-0.22	0.01	-0.26	-0.10	-0.15	0.03	-0.05
23	60	-31	330	-0.21	-0.20	0.02	0.05	0.16	-0.05	-0.22	0.15	0.03	-0.28	-0.14	-0.19	0.03

Estimation Sample: 1993.3.1-1999.12.23. Sample for recursive one-step-ahead forecast errors: 1995.1.1-1999.12.23.
 $p(t)$: model day of the month. n_p : number of observations for each day of the month. mean and standard deviations (s.d) measured in 10^6 Euro. $r(\tau)$: periodic autocorrelation with lag τ measured in model days. Lag $k \cdot 23$ corresponds to a lag of k months. .NaN: no observations to compute these autocorrelations.

Figure 4: Forecasting weight distributions of independent PUC model

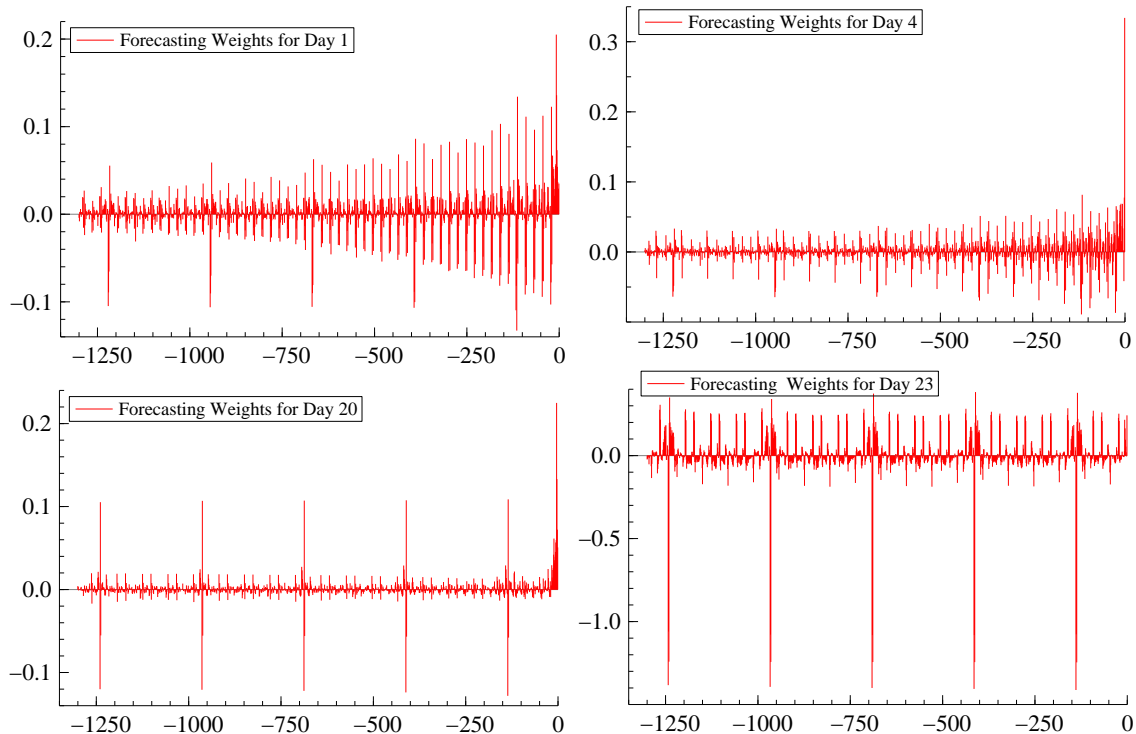


Periodic weight distributions on past observations, implied by an independent periodic model, for forecasting observations in December 1997. Parameters estimates for model (22) given in Table 2. Horizontal axis: distance of observations from 1997.12 measured in months. Figures present weights for forecasting the days $p=1,2,3,4,19,20,21,22,23$ in December 1997, using data from March 1993 until November 1997.

Figure 6 presents the observation weights for forecasting the same days in December 1997 as in Figure 5, but this time one month ahead, instead of one day ahead. The weights for the last 23 observations are now zero by construction, and the remaining observations receive a higher average weight. The effect of the yearly and quarterly regression dummies with fixed coefficients resembles the pattern for the one-day-ahead forecasts in Figure 5 very closely. The time-varying component for the intramonthly pattern turns out to be much more important for forecasting Day 20 ($p = 20$) one month ahead than for forecasting one day ahead. A similar difference between the monthly patterns of the weights for one-day-ahead and those for one-month-ahead forecasting does not occur for Day 1 ($p = 1$) and Day 4 ($p = 4$).

Supplied with this background information on the forecasting function of the different models, we turn to the out-of-sample forecasting results. Table 5 shows that the independent PUC model (22) which has a state vector of dimension $23 \cdot 13$ and $23 \cdot 4$ unknown variances outperforms the basic unobserved components model in terms of simulated one-step-ahead forecast RMSE, also for the last two days of the month, for which 4 separate dummy variables were introduced in the basic model (2 quarterly and 2 yearly dummies). However, the existing model outperforms the independent PUC model for the first day of the month ($p = 1$, for which the basic unobserved components model also contains an end-of-quarter and a beginning-of-July dummy). This bad result is due to the extreme estimate for σ_ω , presented in Table 2 which did not reappear in the full sample estimate presented in Table 3. A similarly bad performance of

Figure 5: Forecasting one-day-ahead weight distributions of UC model Koopman and Ooms (2003)



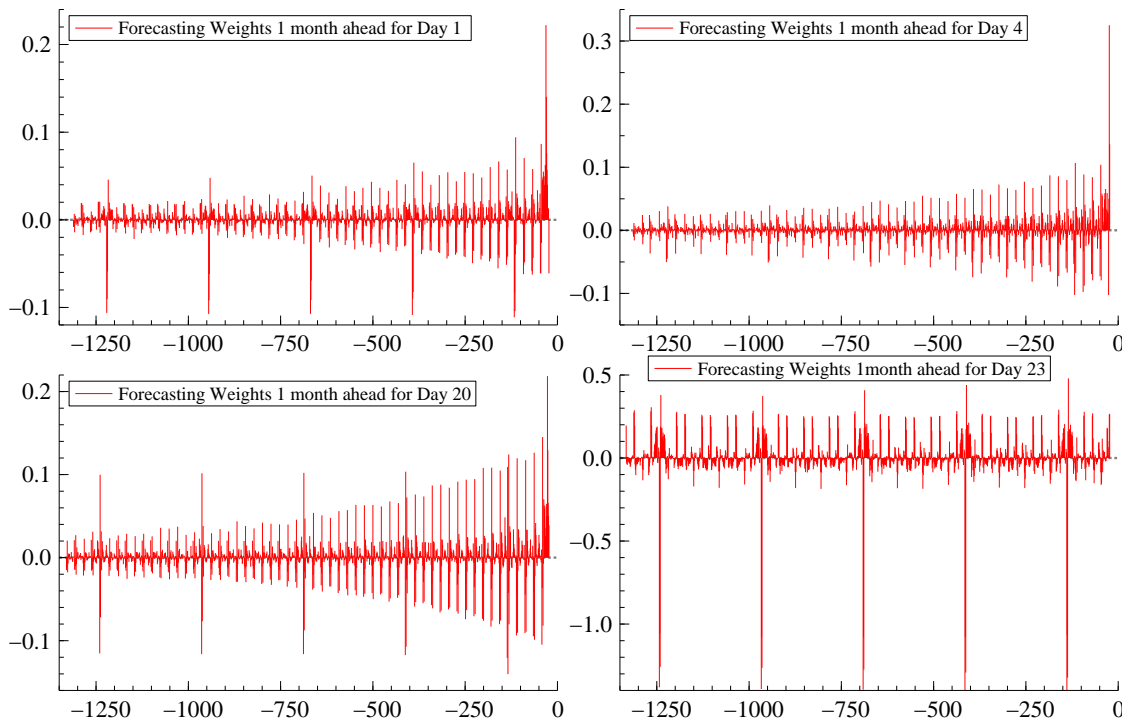
Periodic observation weight distributions implied by UC model Koopman and Ooms (2003), for forecasting observations in December 1997. Horizontal axis: distance of observations from 1997.12.p measured in days. Figures present weights for forecasting the days $p=1,4,20,23$ in December 1997, using data from March 1993 until 1997.12.($p - 1$).

the PUC model is obtained for Day 13 of the month ($p = 13$).

Table 5 also shows forecasting results for the Seasonal Random Walk Model (SRW) and for the Periodic (additive) Seasonal Holt-Winters procedure (PSHW), applied separately to each monthly series for each model day of the month. The popular Seasonal Holt-Winters procedure, Holt (2004), also involves the additive decomposition of a time series into trend and seasonal, allowing for a time-varying slope, but it does not identify separate error terms for the trend and the seasonal component. The procedure is easy to implement and widely available in various packages. Yet, standard methods to initialise the components and standard methods to deal with missing observations are harder to come by. We estimated the unknown smoothing parameters of the Holt-Winters procedures by the nonlinear least squares estimator for the equivalent ARIMA process, following McKenzie (1984). We initialised the Holt-Winters components (trend, slope and seasonals) from coefficients of an OLS regression for the first two years of data on a constant trend and seasonals. Missing observations were replaced by one-step-ahead forecasts.

The forecasting results for the periodic seasonal Holt-Winters procedure are also competitive compared with the existing unobserved components model. This shows that also the Seasonal Holt-Winters procedure is able to identify reasonable stochastic trends and seasonals for the different days of the month, so the applicability of our extension is not completely model specific.

Figure 6: Forecasting one-month-ahead weight distributions of UC model Koopman and Ooms (2003)



Periodic observation weight distributions implied by UC model Koopman and Ooms (2003), for forecasting observations in December 1997. Horizontal axis: distance of observations from 1997.12. p measured in days. Figures present weights for forecasting the days $p=1,4,20,23$ in December 1997, using data from March 1993 until 1997.11. (p) .

We do not present component estimates here, but in general the estimated periodic trends and seasonals of the PSHW procedure were comparable to those obtained by the independent PUC model. However, the independent PUC model, which has a rigorous statistical foundation with unambiguous optimal procedures for diffuse component initialisation and for dealing with missing observations, outperforms the periodic seasonal Holt-Winters procedure in terms of forecasting.

It is surprising and encouraging that the large increase in flexibility associated with the independent PUC model, (22), does result in good simulated out-of-sample forecasts for most days of the month, even when this model predicts one month ahead and the existing unobserved components model of Koopman and Ooms (2003) predicts one day ahead. Apparently, allowing for more independent movements in the trends and seasonal patterns of the different days of the month can be valuable, even if the estimation sample contains less than five years.

7 Conclusion

In this paper we have investigated periodic unobserved components (PUC) time series models for daily data and compared them with a basic unobserved component time series model, with the purpose of short term forecasting. We used a statistical state space approach for estimation

Table 5: *Simulated out-of-sample forecast error comparison 1998-1999 based on RMSE*

$p(t)$	IPUC	PSHW	KO2003	SRW
1	288	278	208	275
2	37	37	65	46
3	43	42	41	54
4	31	31	45	34
5	33	34	28	38
6	30	34	42	43
7	43	42	37	49
8	49	47	43	55
9	56	58	46	54
10	100	102	85	116
11	97	93	108	95
12	78	13	16	5
13	94	90	54	78
14	47	47	56	61
15	51	62	55	60
16	47	48	71	53
17	84	86	92	95
18	62	59	72	56
19	74	85	73	71
20	84	100	97	122
21	141	138	172	181
22	209	255	318	377
23	348	374	370	396
1-23	123	132	134	153

Estimation sample 1993.3.1-1997.12.23. Forecast period 1998.1.1-1999.12.23: Root mean squared errors. IPUC: Independent periodic model (22). PSHW: Periodic Additive Seasonal Holt-Winters forecasting. KO2003: Unobserved Component model of Koopman and Ooms (2003). SRW: Yearly seasonal random walk. $p(t)$: model day of the month. 1 – 23 : overall RMSE. 513 forecast observations for forecast errors, but 500 for SRW. RMSE $\hat{\sigma}_p$ measured in 10^6 Euro.

and forecasting. We discussed the computation of observation weights for forecasting in periodic unobserved component models in the presence of regressor variables, where we extended results obtained for state space models derived by Koopman and Harvey (2003).

The empirical results were based on a series of daily tax revenues, where the main periodic features depend on the day of the month. We found convincing empirical evidence that the daily time series is subject to periodic correlations. We successfully modelled periodic correlations at monthly and yearly lags using an independent PUC model with strongly periodic forecasting functions as a result. The independent PUC model, estimated using less than five year of data, outperformed competing models in a simulated out-of-sample forecasting exercise for two years of daily data.

We showed that a fully periodic approach to unobserved components time series modelling can be a viable method to model daily data with strong periodic features. We confined ourselves to the independent PUC model, which is easy to estimate using standard software. In future research we shall investigate the viability of dependent PUC models for forecasting daily data. This provides a handle to systematically exploit periodic correlations at lags shorter than a month.

Appendix: computing observation weights

In this appendix we provide a detailed description of the algorithms that are discussed and used in the Section 3 and Section 4. We assume that y_t is a scalar observation such that y is a $n \times 1$ observation vector. Further we define the set of observations $Y_t = \{y_1, \dots, y_t\}$. The set of past observations at time t is therefore referred to as Y_{t-1} for $t = 1, \dots, n$ and it follows that $y = Y_n$. First, we introduce the notation for the Kalman Filter and we discuss state and error smoothing. Next, we present algorithms for computing filtering weights for multi-step forecasting. Finally, we discuss implementation of the algorithms in `Ox/SsfPack`.

Kalman filter

Consider the linear state space model (4) and (5). A predictive filtered estimator is an estimator of (a function of) the state vector at time $t + 1$, α_{t+1} , based on observations up to and including time t . The Kalman filter computes the predictive filtered estimator $a_{t+1|t}$, the minimum mean square linear estimator (MMSLE) of the $m \times 1$ state vector α_{t+1} conditional on the observations Y_t , together with its mean square error (MSE) matrix, that is the $m \times m$ variance matrix of the estimation error, $P_{t+1|t}$. The Kalman filter is given by

$$\begin{aligned} v_t &= y_t - Z_t a_{t|t-1}, & F_t &= Z_t P_{t|t-1} Z_t' + G_t G_t', \\ a_{t+1|t} &= T_t a_{t|t-1} + K_t v_t, & M_t &= T_t P_{t|t-1} Z_t' + H_t G_t', \\ & & P_{t+1|t} &= T_t P_{t|t-1} T_t' + H_t H_t' - K_t M_t', \end{aligned} \quad t = 1, \dots, n, \quad (25)$$

with Kalman gain matrix $K_t = M_t F_t^{-1}$ and initialisation $a_{1|0} = a$ and $P_{1|0} = P$. The one-step ahead prediction error is v_t with variance $\text{var}(v_t) = F_t$. The filtered estimator $a_{t|t}$, the MMSLE of the state vector α_t conditional on Y_t , and its MSE matrix $P_{t|t}$ can be computed by

$$a_{t|t} = a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} v_t, \quad P_{t|t} = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}, \quad t = 1, \dots, n. \quad (26)$$

Derivations of the Kalman filter can be found in Anderson and Moore (1979), Harvey (1989) and Durbin and Koopman (2001). Our notation follows de Jong (1991).

State and error smoothing

Smoothed estimators of the state vector, or estimators of α_t based on all observations Y_n , are computed by first running the Kalman filter and, subsequently, running the backward recursion

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \quad N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \quad t = n, \dots, 1, \quad (27)$$

where $L_t = T_t - K_t Z_t$, with initialisations $r_n = 0$ and $N_n = 0$. We note that r_t is an $m \times 1$ vector with $\text{var}(r_t) = N_t$ for $t = 1, \dots, n$. The Kalman filter output of v_t , F_t^{-1} and K_t must be stored for $t = 1, \dots, n$, before applying backward recursion (27). The MMSLE of the state vector α_t conditional on Y_n , i.e. the smoothed state vector $a_{t|n}$, with its MSE matrix $P_{t|n}$ is then computed by

$$a_{t|n} = a_{t|t-1} + P_{t|t-1} r_{t-1}, \quad P_{t|n} = P_{t|t-1} - P_{t|t-1} N_{t-1} P_{t|t-1}, \quad t = n, \dots, 1. \quad (28)$$

Additional memory is required to store $a_{t|t-1}$ and $P_{t|t-1}$ for $t = 1, \dots, n$.

The smoothed error is defined in Section 3.5 as $e = \Omega^{-1}y = C'F^{-1}v$. The t th element of e , that is e_t , and its variance D_t can be computed by the backward recursion (27) and

$$e_t = F_t^{-1}v_t - K_t'r_t, \quad D_t = F_t^{-1} + K_t'N_tK_t, \quad t = n, \dots, 1.$$

It can be shown that the smoothed disturbances are simple functions of e_t and r_t , when the disturbances for the state and for the measurement equations are mutually uncorrelated at all times, that is when $H_tG_t' = 0$ for $t = 1, \dots, n$. In that frequently encountered case it holds that $G_t\hat{\varepsilon}_t = G_tG_t'e_t$ and $H_t\hat{\varepsilon}_t = H_tH_t'r_t$. Derivations of state and disturbance smoothing can be found in Kohn and Ansley (1989), de Jong (1989) and Koopman (1993).

Algorithms for computing filtering weights

Here we present details of the observation weights discussed in Section 4. The filtered estimator of the state vector, that is the estimator of α_t based on information available at time $t - 1$, can be written as

$$a_{t|t-1} = \sum_{j=1}^{t-1} w_j(a_{t|t-1}) y_j, \quad (29)$$

where $w_j(a_{t|t-1})$ is the $m \times 1$ weight vector for observation y_j . In matrix notation we have $a_{t|t-1} = w(a_{t|t-1})y$ where the $m \times n$ matrix $w(a_{t|t-1})$ is given by

$$w(a_{t|t-1}) = \{w_1(a_{t|t-1}), \dots, w_n(a_{t|t-1})\}.$$

After the Kalman filter (25) is applied up to time $t - 1$ and after the matrices K_j are stored for $j = 1, \dots, t - 1$, the weight vectors can be computed by the backward recursion

$$w_j(a_{t|t-1}) = B_{t,j}K_j, \quad B_{t,j-1} = B_{t,j}T_j - w_j(a_{t|t-1})Z_j, \quad j = t - 1, t - 2, \dots, 1, \quad (30)$$

with initialisation $B_{t,t-1} = I$.

The observation weights for $\hat{y}_{t|t-1} = Z_t a_{t|t-1}$ are given by $Z_t w_j(a_{t|t-1})$ but they can also be directly computed from the backward recursion

$$w_j(Z_t a_{t|t-1}) = b_{t,j}K_j, \quad b_{t,j-1} = b_{t,j}T_j - w_j(Z_t a_{t|t-1})Z_j, \quad j = t - 1, t - 2, \dots, 1, \quad (31)$$

starting with $b_{t,t-1} = Z_t$ and where $b_{t,j} = Z_t B_{t,j}$.

The weights for contemporaneous filtering are closely related to the weights for predictive filtering $w_j(a_{t|t-1})$. We have

$$\begin{aligned} w_t(a_{t|t}) &= P_{t|t-1}Z_t'F_t^{-1}, \\ w_j(a_{t|t}) &= (I - P_{t|t-1}Z_t'F_t^{-1}Z_t)w_j(a_{t|t-1}), \quad j = t - 1, t - 2, \dots, 1. \end{aligned}$$

Therefore, recursion (30) can be used, starting with $B_{t,t-1} = I - P_{t|t-1}Z_t'F_t^{-1}Z_t$. The weights for the filtered estimator $\hat{y}_{t|t} = Z_t a_{t|t}$ are given by $w_j(\hat{y}_{t|t}) = Z_t w_j(a_{t|t})$, for $j = t, t - 1, \dots, 1$. The derivations and details of the filtering weight algorithms are given by Koopman and Harvey (2003). They also developed an algorithm for computing the weights for smoothed estimates.

Missing values and multi-step forecasting

When missing values are present in the time series, the Kalman filter can still be applied. When y_s is missing, the Kalman quantities v_s , F_s and K_s get the values

$$v_s = 0, \quad F_s = \infty, \quad K_s = 0, \quad 1 < s < n,$$

as discussed in Durbin and Koopman (2001). Recursion (30) at time $j = s$ reduces to

$$w_s(a_{t|t-1}) = 0, \quad B_{t,s-1} = B_{t,s}T_s. \quad (32)$$

In this way a missing value can be accommodated in the algorithms for computing filtering weights. Future observations are missing by definition. Multi-step forecasts of y_τ for $\tau > n + 1$ can therefore be computed by the Kalman filter and taking y_{n+1}, \dots, y_τ as missing. The observation weights of these multi-step forecasts can be obtained from (32). Using the initialisation of (30) to time $t = \tau$, so that $B_{\tau,\tau-1} = I$, and then applying (32) for $s = \tau - 1, \dots, n + 1$ we obtain $w_n(a_{\tau|n}) = T_{\tau-1} \dots T_{n+1}K_n$. Note that $a_{\tau|n} = a_{\tau|n}$ for any $\tau > n$ since y_{n+1}, \dots, y_τ are missing. The weights for $w_j(a_{\tau|n})$ can be obtained by (30) for $j = n - 1, \dots, 1$ (or by (32) when an observation is missing).

Implementation in Ox/SsfPack

The algorithm for computing observation weights is implemented in the library of state space functions **SsfPack** of Koopman, Shephard, and Doornik (1999) for the object-oriented matrix programming system **Ox** of Doornik (2001). The **SsfPack** function **SsfWeights** is introduced by Koopman and Harvey (2003). Observation weights of the estimated state vector at time j can be obtained via the function

```
mw = SsfWeights(imode, index, {ssf});
```

where the first argument indicates type of estimate (for prediction **ST_PRED**, for filtering **ST_FIL** and for smoothing **ST_SMO**) and the second argument is a $1 \times n$ vector of zeroes except at time t where it is unity. The third argument represents the **SsfPack** system matrices $\{T_t, Z_t, H_t, G_t\}$ and the initial conditions $\{a, P\}$, see Koopman, Shephard, and Doornik (1999). For example, we have

```
imode = ST_PRED;
index = zeros(1,100); index[99] = 1;
```

for a sample size of $n = 100$. Then function **SsfWeights** computes the observation weights for the predictive estimate of the state and observation vectors at time index $t = 100$. Note that in **Ox** the indices of vectors and matrices start at 0 rather than 1. In the case of the example, the function **SsfWeights** returns the matrix

$$mw = \begin{bmatrix} w_1(a_{t|t}) & w_2(a_{t|t}) & \cdots & w_{100}(a_{t|t}) \\ w_1(\hat{y}_{t|t}) & w_2(\hat{y}_{t|t}) & \cdots & w_{100}(\hat{y}_{t|t}) \end{bmatrix}, \quad \text{with } t = 100.$$

To indicate that observations are missing, elements of **index** can take the “not available” constant **M_NAN** in **Ox**. When the weights of the state and observation vectors at time index t are required while observation y_t itself is missing, the index associated with time t should take the value -1 rather than 1. It follows that the weights for the five-step ahead forecast of the state and observation vectors can be obtained by

```
imode = ST_PRED;
index = zeros(1,100) ~ constant(M_NAN, 1, 5); index[104] = -1;
mw = SsfWeights(imode, index, {ssf});
```

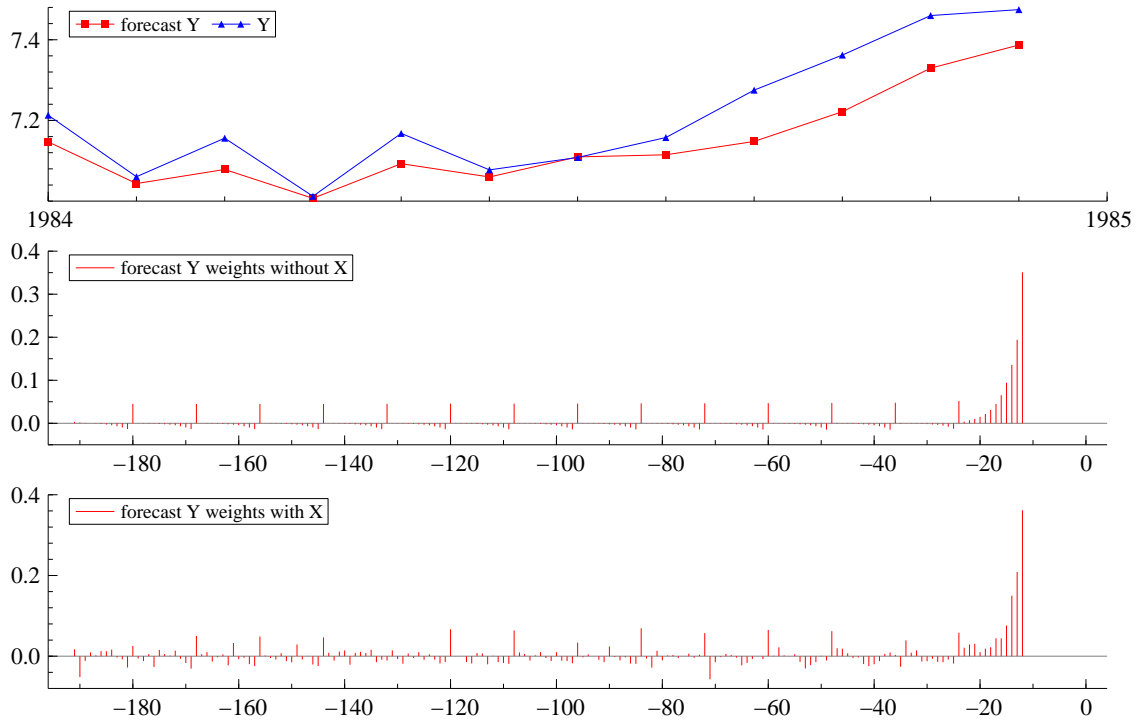
Illustration

To illustrate the use of the `SsfPack` function `SsfWeights`, we consider the basic unobserved components (UC) model $y_t = \mu_t + \gamma_t + \varepsilon_t$ as given by (1) where the trend μ_t is given by (2), with $\beta_t = 0$ for all t , and seasonal γ_t is given by (3), with $Q = 12$ for monthly data. The state space form of this class of models is given in Section 3.3. The seatbelt data-set of Harvey and Durbin (1986) is considered for this illustration. Their basic UC model describes the log number of drivers killed and seriously injured in road accidents in Great Britain per month between 1969 and 1984. The model also includes explanatory variables in the form of (8). Their four explanatory variables are the number of kilometres travelled, the real price of petrol, the tax rate on cars and a dummy for the start of the seatbelt law on January 31st, 1983. Due to the inclusion of explanatory variables, adjustments as described in Section 4.3 need to be applied for the computation of the observations weights. The actual implementation of the calculations is carried out in an `Ox` program that is presented in Listings 1 and 2. The graphical output of the program is presented in Figure 7. The program computes the weights for the forecast of December 1984 based on the observations up to December 1983. The difference in the weight patterns for a model with explanatory variables and without explanatory variables is clearly noticeable.

References

- Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Englewood Cliffs: Prentice-Hall.
- Carnero, A., S. J. Koopman, and M. Ooms (2003). Periodic heteroskedastic RegARFIMA models for daily electricity spot prices. Technical Report 03-071/4, Tinbergen Institute, Amsterdam, The Netherlands.
- de Jong, P. (1989). Smoothing and interpolation with the state space model. *J. American Statistical Association* 84, 1085–8.
- de Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics* 19, 1073–83.
- Doornik, J. A. (2001). *Object-Oriented Matrix Programming using Ox 3.0*. London: Timberlake Consultants Press.
- Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- Franses, P. H. and R. Paap (2004). *Periodic Time Series Models*. Oxford, U.K.: Oxford University Press.
- Harrison, J. and C. F. Stevens (1976). Bayesian forecasting. *Journal of the Royal Statistical Society, Series B* 38, 205–247.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge, UK: Cambridge University Press.

Figure 7: Multi-step forecasts and weight function for model of Harvey and Durbin (1986) with and without regression effects



Forecast period: January 1984–December 1984. Top: Multi-step forecast with realisation. Middle: Weights $w(\hat{u}_*)$, see(14). Bottom: Weights $w(\hat{y}_*)$, see (17). The weights are computed as detailed in Section 4.2 and are for the 12-steps ahead forecast. Horizontal axis of Top: time-index of months in 1984. Horizontal axis of Middle and Bottom: distance of observations from 1984.12 measured in months.

Harvey, A. C. (1993). *Time Series Models, second edition*. New York, London: Harvest Wheatsheaf.

Harvey, A. C. and J. Durbin (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling, (with discussion). *J. Royal Statistical Society B* 149, 187–227.

Harvey, A. C. and S. J. Koopman (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association* 88, 1228–1237.

Harvey, A. C. and S. J. Koopman (1997). Multivariate structural time series models. In C. Heij, H. Schumacher, B. Hanzon, and C. Praagman (Eds.), *Systematic dynamics in economic and financial models*, pp. 269–98. Chichester: John Wiley and Sons.

Harvey, A. C., S. J. Koopman, and J. Penzer (1998). Messy time series: A unified approach. In T. B. Fomby and R. Carter Hill (Eds.), *Advances in Econometrics, Volume 13*, pp. 103–143. New York, NY, USA: JAI Press.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* 20, 5–10. Reprinted version of the 1957 report to the Office of Naval Research (ONR 52).

- Jones, R. H. (1980). Maximum likelihood fitting of ARIMA models to time series with missing observations. *Technometrics* 22, 389–95.
- Kitagawa, G. and W. Gersch (1996). *Smoothness Priors Analysis of Time Series*. New York: Springer Verlag.
- Kohn, R. and C. F. Ansley (1989). A fast algorithm for signal extraction, influence and cross-validation. *Biometrika* 76, 65–79.
- Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* 80, 117–26.
- Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for non-stationary time series models. *J. American Statistical Association* 92, 1630–8.
- Koopman, S. J. and A. C. Harvey (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control* 27, 1317–1333.
- Koopman, S. J. and M. Ooms (2002). Periodic unobserved component time series models: estimation and forecasting with applications. Paper presented at the 57th European Meeting of the Econometric Society (ESEM2002) held in Venice, August 25-28, 2002, Department of Econometrics, Vrije Universiteit Amsterdam.
- Koopman, S. J. and M. Ooms (2003). Time series modelling of daily tax revenues. *Statistica Neerlandica* 57, 439–469.
- Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space using SsfPack 2.2. *The Econometrics Journal* 2, 107–160, www.ssfpack.com.
- McKenzie, E. (1984). General exponential smoothing and the equivalent ARMA process. *Journal of Forecasting* 3, 333–344.
- Ooms, M. and P. H. Franses (2001). A seasonal periodic long memory model for monthly river flows. *Environmental Modelling & Software* 16, 559–569.
- Osborn, D. R. (1991). The implications of periodically varying coefficients for seasonal time-series processes. *Journal of Econometrics* 48, 373–384.
- Osborn, D. R. and J. P. Smith (1989). The performance of periodic autoregressive models in forecasting seasonal UK consumption. *Journal of Business and Economic Statistics* 7, 117–127.
- Proietti, T. (2004). Seasonal specific structural time series. *Studies in Nonlinear Dynamics and Econometrics* 8(2), Article 16, <http://www.bepress.com/snnde/vol8/iss2/art16>.
- Schweppe, F. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory* 11, 61–70.
- Tiao, G. C. and M. R. Grupe (1980). Hidden periodic autoregressive-moving average models in time series data. *Biometrika* 67, 365–373.

```

static decl s_mStsm, s_vVarCmp;          // matrix for unobserved components model

SetStsmModel(const vP)
{
    // map to sts model with level, slope and trig seasonal
    s_mStsm = <CMP_LEVEL,      1, 0, 0;
              CMP_SEAS_DUMMY, 1, 12, 0; // 12 for monthly data
              CMP_IRREG,      1, 0, 0>;
    s_mStsm[][1] = exp(vP);
    decl vr = exp(2.0 * vP);           // from log(s.d.) to variance
    s_vVarCmp =                       // s_vVarCmp is diagonal(Omega)
    vr[0] | vr[1] | zeros(10,1) | vr[2];
}

ComputeForecastY(const mY, const mX, const mXF)
{
    decl mphi, momega, msigma, mjphi = <>, md, mp, mf;
    decl cst, cf = columns(mXF);

    GetSsfStsm(s_mStsm, &mphi, &momega, &msigma);
    if (mX != <>) AddSsfReg(mX, &mphi, &momega, &msigma, &mjphi);
    cst = columns(mphi);
    mp =
    SsfMomentEstEx(ST_PRED, &md, mY, mphi, momega, msigma, <>, mjphi, <>, <>, mX);

    return
    SsfForecast(zeros(1, cf), mphi, momega, mp, <>, mjphi, <>, <>, mXF)[cst] [];
}

ComputeForecastYWeights(const ctf, const mY, const mX, const mXF)
{
    decl cst, ct = columns(mY), ck = rows(mX);
    decl mphi, momega, msigma, mjphi = <>;
    decl index = zeros(1,ct)~constant(M_NAN,1,ctf); index[0][ctf+ct-1] = -1;
    decl mw, mwx;

    GetSsfStsm(s_mStsm, &mphi, &momega, &msigma);
    // computing weights for u
    cst = columns(mphi);
    mw = SsfWeights(ST_PRED, index, mphi, momega, msigma)[cst] [];
    if (mX != <>)
    {
        AddSsfReg(mX ~ mXF, &mphi, &momega, &msigma, &mjphi);
        // computing weights for regressor coefficients
        mwx =
        SsfWeights(ST_PRED, index, mphi, momega, msigma, <>, mjphi, <>, <>, mX~mXF);

        mw += (mXF[][ctf - 1]' - mw * (mX ~ mXF)') * mwx[:ck-1] [];
    }
    return mw;
}

```

Listing 1: Ox code for computing multi-step forecasts and observation weights for multi-step forecasts within a UC time series model with explanatory variables

```

#include <oxstd.h>
#include <oxfloat.h>
#include <oxdraw.h>
#include <packages/ssfpack/ssfpack.h>

static decl s_mStsm, s_vVarCmp;          // matrix for unobserved components model

main()
{
  decl data = loadmat("seatbelt.in7"); // load seatbelt data
  decl cf   = 12;                       // number of forecasts
  decl ct   = columns(data)-cf;         // insample size
  decl my   = data[0][:ct-1];           // Y
  decl mx   = data[3][:ct-1];           // X
  decl ck   = rows(mx);                 // no of regressors
  decl mxf  = data[3:][ct:];            // X forecasts

  SetStsmModel(<-4.; -7.; -3.>);
  decl mf, mw;

  mf = ComputeForecastY(my, mx, mxf);
  DrawTMatrix(0, mf | data[0][ct:], {"forecast Y", "Y"}, 1984, 1, 12, 0, 2);
  mw = ComputeForecastYWeights(cf, my, <>, <>);
  DrawTMatrix(1, mw, {"forecast Y weights without X"}, 1-(ct+cf), 1, 1, 0, 2);
  mw = ComputeForecastYWeights(cf, my, mx, mxf);
  DrawTMatrix(2, mw, {"forecast Y weights with X"}, 1-(ct+cf), 1, 1, 0, 2);
  ShowDrawWindow();
}

```

Listing 2: Ox code for program that outputs Figure 7