Forecasting economic time series using unobserved components time series models

Siem Jan Koopman and Marius Ooms
VU University Amsterdam, Department of Econometrics
FEWEB, De Boelelaan 1105, 1081 HV Amsterdam
The Netherlands

May 26, 2010

1 Introduction

The forecasting of economic time series is a challenging problem. We approach the forecasting challenge from a model-based perspective and adopt the unobserved components time series model. The key feature of this class of models is the decomposition of a time series into trend, seasonal, cycle and irregular components. Each component is formulated as a stochastically evolving process over time. The decomposition of an observed time series into unobserved stochastic processes can provide a better understanding of the dynamic characteristics of the series and the way these characteristics change over time. The trend component typically represents the long-term developments of the time series of interest and is often specified as a smooth function of time. The recurring and persistently changing patterns within the year can be captured by the seasonal component. In economic time series, the cycle component can represent the dynamic features associated with the business cycle (or the output gap). In economic policy, the focus is often on forecasting the variable of interest, not its separate components. However, we argue that an understanding of the time series decomposition and the dynamic properties of the underlying components can benefit the forecasting of the variable of interest.

Unobserved components time series models have a natural state space representation. The statistical treatment can therefore be based on the Kalman filter and its related methods. The resulting modelling framework is particularly convenient for the problem of forecasting as we will illustrate in this Chapter. For example, it provides optimal point and interval forecasts but it also provides the observation weights for the associated forecasting function. In this way, forecasts can be expressed directly as functions of past observations.

We present a concise discussion of the forecasting of economic time series on the basis of a general class of unobserved components time series models. We first introduce the model with explicit specifications for the components: trend, season, cycle and irregular. The estimation of parameters is carried out by the method of maximum likelihood in which the likelihood is evaluated via the Kalman filter. The likelihood is maximized by means of a numerical optimization method. Based on the parameter estimates, the components can be estimated using the observed time series. The actual decomposition of the time series into trend, seasonal, cycle and irregular can then be visualized. Model adequacy can be diagnosed using the standard test statistics applied to the standardised one-step ahead prediction errors. This approach to time series analysis implies a specific approach to the modelling of time series. It is somewhat different compared to the Box-Jenkins analysis. For example, in the unobserved components time series approach we do not require the differencing of a time series towards a stationary process. The nonstationary properties of a time series are explicitly formulated by a selection of the components in the decomposition. The Box-Jenkins approach requires that the observed time series has been differenced into a stationarity process. Although the two resulting methodologies are distinct, the model classes both belong to the linear Gaussian family of models and both can be formulated as linear dynamic processes.

This Chapter is organised as follows. Section 2 provides a comprehensive review of unobserved components time series models. Section 3 discusses the methodology of state space

analysis. We introduce the state space model, we give illustrations of how unobserved component time series models can be formulated in state space, we present the Kalman filter, we discuss maximum likelihood estimation of parameters and we present some diagnostic checking statistics. In Section 4 we discuss how forecasts can be generated as part of a state space time series analysis and how observation weights of the forecast function are computed. Various multivariate extensions of the unobserved components time series model are discussed in Section 5. Specifically, we present multivariate time series models with common trends and cycles and we discuss how a dynamic factor analysis based on maximum likelihood can be carried out in a computationally efficient way. To illustrate the methodology, we present in Section 6 an empirical analysis for daily electricity spot prices based on a univariate and a bivariate model. We present some interesting features of this analysis but we focus primarily on the forecasting of daily spot prices. Section 7 concludes.

2 Unobserved components time series models

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

$$y_t = \mu_t + \gamma_t + \psi_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2), \qquad t = 1, \ldots, n, \tag{1}$$

where μ_t, γ_t, ψ_t, and ε_t represent trend, seasonal, cycle, and irregular components, respectively. The trend, seasonal, and cycle components are modelled by linear dynamic stochastic processes which depend on disturbances. The components are formulated in a 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) where model (1) is referred to as the "Structural Time Series Model". The effectiveness of structural time series models compared to ARIMA type models is discussed in Harvey, Koopman, and Penzer (1998). They stress that time series models based on unobserved components are particularly effective when messy features are present in the time series such as missing values, mixed frequencies (monthly and quarterly seasons of time series), outliers, structural breaks and nonlinear non-Gaussian aspects. An elementary introduction and a practical guide to unobserved component time series modeling is provided by Commandeur and Koopman (2007).

2.1 Trend component

The trend component can be specified in many different ways. A selection of trend specifications is given below.

Local level - $I(1)$ process: The trend component can simply be modelled as a random walk

process and is then given by

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

where $\text{NID}(0, \sigma^2)$ refers to a normally independently distributed series with mean zero and variance σ^2. The disturbance series η_t is therefore serially independent and mutually independent of all other disturbance series related to y_t in (1). The initial trend μ_1 is for simplicity treated as an unknown coefficient that needs to be estimated together with the unknown variance σ_η^2. The estimation of parameters is discussed in Section 3.4.

In specification (2) the trend component is an $I(1)$ process. When this trend is included in the decomposition of y_t, the time series y_t is at least $I(1)$ as well. Harvey (1989, §2.3.6) defines the *local level model* as $y_t = \mu_t + \varepsilon_t$ with μ_t given by (2). In case $\sigma_\eta^2 = 0$, the observations from a local level model are generated by a NID process with constant mean μ_1 and a constant variance σ^2.

Local linear trend - $I(2)$ process: An extension of the random walk trend is obtained by including a stochastic drift component

$$\mu_{t+1} = \mu_t + \beta_t + \eta_t, \qquad \beta_{t+1} = \beta_t + \zeta_t, \qquad \zeta_t \sim \text{NID}(0, \sigma_\zeta^2), \tag{3}$$

where the disturbance series η_t is as in (2). The initial values μ_1 and β_1 are treated as unknown coefficients. Harvey (1989, §2.3.6) defines the *local linear trend model* as $y_t = \mu_t + \varepsilon_t$ with μ_t given by (3).

In case $\sigma_\zeta^2 = 0$, the trend (3) reduces to an $I(1)$ process given by $\mu_{t+1} = \mu_t + \beta_1 + \eta_t$ where the drift β_1 is fixed. This specification is referred to as a *random walk plus drift* process. If in addition $\sigma_\eta^2 = 0$, the trend reduces to the deterministic linear trend $\mu_{t+1} = \mu_1 + \beta_1 t$. When $\sigma_\eta^2 = 0$ and $\sigma_\zeta^2 > 0$, the trend μ_t in (3) remains an $I(2)$ process and is known as the integrated random walk process which can be visualised as a smooth trend function.

Trend with stationary drift - $I(1)$ process: To extend the random walk trend with a drift component but to keep the trend as an $I(1)$ process, we can include a stationary stochastic drift component to obtain

$$\mu_{t+1} = \mu_t + \beta_t + \eta_t, \qquad \beta_{t+1} = (1 - \varphi_\beta)\bar{\beta} + \varphi_\beta \beta_t + \zeta_t, \tag{4}$$

with autoregressive coefficient $0 < \varphi_\beta \leq 1$ and where the disturbance series η_t and ζ_t are as in (3). The restriction of a positive and strictly less than unity value for φ_β is necessary to have a stationary process for the drift β_t. In the stationary case, the initial variable μ_1 is treated as an unknown coefficient while the initial drift is specified as $\beta_1 \sim \text{N}[\bar{\beta}\,,\, \sigma_\zeta^2 \,/\, (1 - \varphi_\beta^2)]$. However, when $\varphi_\beta \to 1$ we return to a nonstationary process for the drift β_t and the local linear trend model (3) for μ_t. The stationary drift process for β_t can be generalised to a higher order autoregressive process and can include moving average terms. However, in practice it may be difficult to empirically identify such drift processes without very large data samples.

Higher-order smooth trend - $I(k)$ process: The local linear trend (3) with $\sigma_\eta^2 = 0$ is a smooth $I(2)$ process. The smooth trend component can alternatively be specified as $\Delta^2 \mu_{t+2} = \zeta_t$ where the initial variables μ_1 and $\mu_2 = \mu_1 + \beta_1$ are treated as unknown coefficients. To enforce more smoothness in the trend component, we can generalise the smooth trend specification by $\Delta^k \mu_{t+k} = \zeta_t$ where the initial variables $\mu_1, \ldots, \mu_k$ are treated as unknown coefficients for $k = 1, 2, \ldots$. In the usual way, we can specify the higher-order smooth trend component by $\mu_t = \mu_t^{(k)}$ where

$$\mu_{t+1}^{(j)} = \mu_t^{(j)} + \mu_t^{(j-1)}, \qquad \mu_t^{(0)} = \zeta_t, \tag{5}$$

for $j = k, k-1, \ldots, 1$ and where the disturbance series ζ_t is as in (3). In case $k = 2$, we obtain the smooth trend model (3) with $\sigma_\eta^2 = 0$ where $\mu_t = \mu_t^{(2)}$ and $\beta_t = \mu_t^{(1)}$. This trend specification is considered and discussed in more detail by Gomez (2001).

Trend with smooth stationary drift - $I(1)$ process: Although the smoothness of a trend is a desirable feature for many economic time series, the fact that the smooth trend is an $I(k)$ process is less convincing. We therefore propose a smooth $I(1)$ trend as given by

$$\mu_{t+1} = \mu_t + \beta_t^{(m)}, \qquad \beta_{t+1}^{(j)} = \varphi_\beta \beta_t^{(j)} + \beta_t^{(j-1)}, \qquad \beta_t^{(0)} = \zeta_t, \tag{6}$$

for $j = m, m-1, \ldots, 1$ and where the disturbance series ζ_t is as in (3). In case $m = 1$, we obtain the trend with stationary drift model (4) with $\sigma_\eta^2 = \bar{\beta} = 0$ where $\beta_t = \beta_t^{(1)}$. The autoregressive coefficient $0 < \varphi_\beta < 1$ is the same for each $\beta_{t+1}^{(j)}$ with $j = m, m-1, \ldots, 1$. This restriction can be lifted by having different autoregressive coefficients for each j but generally the parsimonious specification (6) is preferred.

2.2 Seasonal component

To account for the seasonal variation in a time series, the component γ_t is included in model (1). More specifically, γ_t represents the seasonal effect at time t that is associated with season $s = s(t)$ for $s = 1, \ldots, S$ where S is the seasonal length ($S = 4$ for quarterly data and $S = 12$ for monthly data). The time-varying seasonal component can be established in different ways.

Fixed dummy seasonal: In case the seasonal pattern is fixed over time, we have a set of S seasonal effects $\gamma_1, \ldots, \gamma_S$ which are taken as unknown coefficients that need to be estimated together with the other coefficients in the model. The seasonal effects must have the property that they sum to zero over the full year to make sure that they are not confounded with the trend component, that is

$$\gamma_1 + \ldots + \gamma_S = 0, \qquad \gamma_t = \gamma_{t-S}, \qquad t = S+1, \ldots, n. \tag{7}$$

For example, when we consider the regression model $y_t = \mu_1 + \gamma_t + \varepsilon_t$, with fixed constant μ_1 and fixed seasonal effects, the summing-to-zero constraint is necessary to avoid the

multicollinearity problem. The constraint

$$\gamma_S = -\gamma_{S-1} - \ldots - \gamma_1$$

ensures that the S seasonal effects sum to zero. We have $S-1$ unknown seasonal coefficients that need to be estimated.

Time-varying dummy seasonal: In time series applications, it is usually more appropriate to allow the seasonal pattern to change (slowly) over time. For this purpose we can relax the summing-to-zero constraint by replacing it with the stochastic equation given by

$$\gamma_{t+1} = -\gamma_t - \ldots - \gamma_{t-S+2} + \omega_t, \qquad \omega_t \sim \text{NID}(0, \sigma_\omega^2), \tag{8}$$

where the disturbance series ω_t is serially independent and mutually independent of all other disturbance series, for $t = S-1, \ldots, n$. The initial variables $\gamma_1, \ldots, \gamma_{S_1}$ are treated as unknown coefficients. When the disturbance variance $\sigma_\omega^2 = 0$, we return to the case of fixed dummy seasonal effects. When the variance σ_ω^2 is relatively large, the seasonal pattern will vary quickly over time.

Fixed trigonometric seasonal: A deterministic seasonal pattern can also be constructed from a set of sine and cosine functions. In this case the seasonal component γ_t is specified as a sum of trigonometric cycles with seasonal frequencies. Specifically, we have

$$\gamma_t = \sum_{j=1}^{\lfloor S/2 \rfloor} \gamma_{j,t}, \qquad \gamma_{j,t} = a_j \cos(\lambda_j t - b_j), \tag{9}$$

where $\lfloor \cdot \rfloor$ is the floor function, $\gamma_{j,t}$ is the cosine function with amplitude a_j, phase b_j, and seasonal frequency $\lambda_j = 2\pi j/S$ (measured in radians) for $j = 1, \ldots, \lfloor S/2 \rfloor$ and $t = 1, \ldots, n$. The seasonal effects are based on coefficients a_j and b_j. Given the trigonometric identities

$$\cos(\lambda \pm \xi) = \cos\lambda\cos\xi \mp \sin\lambda\sin\xi, \qquad \sin(\lambda \pm \xi) = \cos\lambda\sin\xi \pm \sin\lambda\cos\xi, \tag{10}$$

we can express $\gamma_{j,t}$ as the sine-cosine wave

$$\gamma_{j,t} = \delta_{c,j}\cos(\lambda_j t) + \delta_{s,j}\sin(\lambda_j t), \tag{11}$$

where $\delta_{c,j} = a_j \cos b_j$ and $\delta_{s,j} = a_j \sin b_j$. The reverse transformation is $a_j = \delta_{c,j}^2 + \delta_{s,j}^2$ and $b_j = \tan^{-1}(\delta_{s,j} \,/\, \delta_{c,j})$. The seasonal effects are alternatively represented by coefficients $\delta_{c,j}$ and $\delta_{s,j}$. When S is odd, the number of seasonal coefficients is $S-1$ by construction. For S even, variable $\delta_{s,j}$, with $j = S/2$, drops out of (11) since frequency $\lambda_j = \pi$ and $\sin(\pi t) = 0$. Hence for any seasonal length $S > 1$ we have $S-1$ seasonal coefficients as in the fixed dummy seasonal case.

The evaluation of each $\gamma_{j,t}$ can be carried out recursively in t. By repeatedly applying

the trigonometric identities (10), we can express $\gamma_{j,t}$ as the recursive expression

$$\begin{pmatrix} \gamma_{j,t+1} \\ \gamma^+_{j,t+1} \end{pmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{pmatrix} \gamma_{j,t} \\ \gamma^+_{j,t} \end{pmatrix}, \tag{12}$$

with $\gamma_{j,0} = \delta_{c,j}$ and $\gamma^+_{j,0} = \delta_{s,j}$ for $j = 1, \ldots, \lfloor S/2 \rfloor$. The variable $\gamma^+_{j,t}$ appears by construction as an auxiliary variable. It follows that the seasonal effect γ_t is a linear function of the variables $\gamma_{j,t}$ and $\gamma^+_{j,t}$ for $j = 1, \ldots, \lfloor S/2 \rfloor$ (in case S is even, $\gamma^+_{j,t}$, with $j = S/2$, drops out).

Time-varying trigonometric seasonal: The recursive evaluation of the seasonal variables in (12) allows the introduction of a time-varying trigonometric seasonal function. We obtain the stochastic trigonometric seasonal component γ_t by having

$$\begin{pmatrix} \gamma_{j,t+1} \\ \gamma^+_{j,t+1} \end{pmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{pmatrix} \gamma_{j,t} \\ \gamma^+_{j,t} \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega^+_{j,t} \end{pmatrix}, \quad \begin{pmatrix} \omega_{j,t} \\ \omega^+_{j,t} \end{pmatrix} \sim \text{NID}(0, \sigma^2_\omega I_2), \tag{13}$$

with $\lambda_j = 2\pi j/S$ for $j = 1, \ldots, \lfloor S/2 \rfloor$ and $t = 1, \ldots, n$. The $S-1$ initial variables $\gamma_{j,1}$ and $\gamma^+_{j,1}$ are treated as unknown coefficients. The seasonal disturbance series $\omega_{j,t}$ and $\omega^+_{j,t}$ are serially and mutually independent, and are also independent of all the other disturbance series. In case $\sigma^2_\omega = 0$, equation (13) reduces to (12). The variance σ^2_ω is common to all disturbances associated with different seasonal frequencies. These restrictions can be lifted and different seasonal variances for different frequencies λ_j can be considered for $j = 1, \ldots, \lfloor S/2 \rfloor$.

The random walk seasonal: The random walk specification for a seasonal component is proposed by Harrison and Stevens (1976) and is given by

$$\gamma_t = e_j' \gamma^\dagger_t, \qquad \gamma^\dagger_{t+1} = \gamma^\dagger_t + \omega^\dagger_t, \qquad \omega^\dagger_t \sim \text{NID}(0, \sigma^2_\omega \Omega), \tag{14}$$

where the $S \times 1$ vector $\gamma^\dagger_t$ contains the seasonal effects, e_j is the jth column of the $S \times S$ identity matrix I_S, $S \times 1$ disturbance vector $\omega^\dagger_t$ is normally and independently distributed with mean zero and $S \times S$ variance matrix $\sigma^2_\omega \Omega$. The seasonal effects evolve over time as random walk processes. To ensure that the sum of seasonal effects is zero, the variance matrix Ω is subject to restriction $\Omega\iota = 0$ with ι as the $S \times 1$ vector of ones. The seasonal index j, with $j = 1, \ldots, S$, corresponds to time index t and represents a specific month or quarter. A particular specification of Ω that is subject to this restriction is given by $\Omega = I_S - S^{-1}\iota\iota'$. Due to the restriction of Ω, the S seasonal random walk processes in $\gamma^\dagger_t$ are not evolving independently of each other. Proietti (2000) has shown that the time-varying trigonometric seasonal model with specific variance restrictions is equivalent to the random walk seasonal model (14) with $\Omega = I_S - S^{-1}\iota\iota'$.

Harvey (1989, §§2.3-2.5) studies the statistical properties of time-varying seasonal processes in more detail. He concludes that the time-varying trigonometric seasonal evolves more smoothly over time than time-varying dummy seasonals.

2.3 Cycle component

To capture business cycle features from economic time series, we can include a stationary cycle component in the unobserved components time series model. For example, for a trend-plus-cycle model, we can consider $y_t = \mu_t + \psi_t + \varepsilon_t$. Next we discuss various stochastic specifications for the cycle component ψ_t.

Autoregressive moving average process: The cycle component ψ_t can be formulated as a stationary autoregressive moving average (ARMA) process and given by

$$\varphi_\psi(L)\psi_{t+1} = \vartheta_\psi(L)\xi_t, \qquad \xi_t \sim \text{NID}(0, \sigma_\xi^2), \tag{15}$$

where $\varphi_\psi(L)$ is the autoregressive polynomial in the lag operator L, of lag order p with coefficients $\varphi_{\psi,1}, \ldots, \varphi_{\psi,p}$ and $\vartheta_\psi(L)$ is the moving average polynomial of lag order q with coefficients $\vartheta_{\psi,1}, \ldots, \vartheta_{\psi,q}$. The requirement of stationarity applies to the autoregressive polynomial $\varphi_\psi(L)$ and states that the roots of $|\varphi_\psi(L)| = 0$ lie outside the unit circle. The theoretical autocorrelation function of an ARMA process has cyclical properties when the roots of $|\varphi_\psi(L)| = 0$ are within the complex range. It requires $p > 1$. In this case the autocorrelations converge to zero when the corresponding lag is increasing, but the convergence pattern is cyclical. It implies that the component ψ_t has cyclical dynamic properties. Once the autoregressive coefficients are estimated, it can be established whether the empirical model with ψ_t as in (15) has detected cyclical dynamics in the time series. The economic cycle component in the model of Clark (1987) is specified as the stationary ARMA process (15) with lag orders $p = 2$ and $q = 0$.

Time-varying trigonometric cycle: An alternative stochastic formulation of the cycle component can be based on a time-varying trigonometric process such as (13) but with frequency λ_c associated with the typical length of an economic business cycle, say between 1.5 and 8 years, as suggested by Burns and Mitchell (1946). We obtain

$$\begin{pmatrix} \psi_{t+1} \\ \psi_{t+1}^+ \end{pmatrix} = \varphi_\psi \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{pmatrix} \psi_t \\ \psi_t^+ \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^+ \end{pmatrix}, \tag{16}$$

where the discount factor $0 < \varphi_\psi < 1$ is introduced to enforce a stationary process for the stochastic cycle component. The disturbances and the initial conditions for the cycle variables are given by

$$\begin{pmatrix} \kappa_t \\ \kappa_t^+ \end{pmatrix} \sim \text{NID}(0, \sigma_\kappa^2 I_2), \qquad \begin{pmatrix} \psi_1 \\ \psi_1^+ \end{pmatrix} \sim \text{NID}\left(0, \frac{\sigma_\kappa^2}{1-\varphi_\psi^2} I_2\right),$$

where the disturbances κ_t and κ_t^+ are serially independent and mutually independent, also with respect to disturbances that are associated with other components. The coefficients φ_ψ, λ_c and σ_κ^2 are unknown and need to be estimated together with the other parameters.

This stochastic cycle specification is discussed by Harvey (1989, §§2.3-2.5), where it is

argued that the process (16) is the same as the ARMA process (15) with $p = 2$ and $q = 1$ and where the roots of $|\varphi_\psi(L)| = 0$ are enforced to be within the complex range.

Smooth time-varying trigonometric cycle: To enforce a more smooth cycle component in the model, we can modify the cycle specification to let it have so-called bandpass filter properties. For this purpose, Harvey and Trimbur (2003) propose the specification $\psi_t = \psi_t^{(m)}$ where

$$\begin{pmatrix} \psi_{t+1}^{(j)} \\ \psi_{t+1}^{(j)+} \end{pmatrix} = \varphi_\psi \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{pmatrix} \psi_t^{(j)} \\ \psi_t^{(j)+} \end{pmatrix} + \begin{pmatrix} \psi_t^{(j-1)} \\ \psi_t^{(j-1)+} \end{pmatrix}, \tag{17}$$

for $j = m, m-1, \ldots, 1$ and where

$$\begin{pmatrix} \psi_t^{(0)} \\ \psi_t^{(0)+} \end{pmatrix} = \begin{pmatrix} \kappa_t \\ \kappa_t \end{pmatrix} \sim \text{NID}(0, \sigma_\kappa^2 I_2),$$

for $t = 1, \ldots, n$. The initial conditions for this stationary process need to be derived and are provided by Trimbur (2006). Although more stochastic variables are required for this specification, the number of coefficients remains three, that is φ_ψ, λ_c and σ_κ^2.

Multiple cycles The dynamic specification of a cycle may be more intricate than the specifications given above. When a satisfactory description of cyclical dynamics cannot be provided by a single component, a set of multiple cycle components can be considered, that is

$$\psi_t = \sum_{j=1}^{J} \psi_{j,t}, \tag{18}$$

where each $\psi_{j,t}$ can be modelled as an independent cycle process, which is specified as one of the cycle processes described above. For example, when a time-varying trigonometric cycle is adopted for each $\psi_{j,t}$ in (18) with a different cycle frequency λ_c, the model can be used to empirically identify shorter and longer cyclical dynamics from a time series simultaneously.

2.4 Regression component

The basic model (1) may provide a succesful description of the time series, although it may sometimes be necessary to include additional components in (1). For example, seasonal economic time series are often affected by trading day effects and holiday effects. In other cases it is evident that a set of explanatory variables need to be included in the model for capturing specific (dynamic) variations in the time series. Explanatory variables can also be used to allow for outliers and breaks in the model. Therefore, we extend the decomposition with a multiple regression effect,

$$y_t = \mu_t + \gamma_t + \psi_t + x_t'\delta + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2), \tag{19}$$

for $t = 1, \ldots, n$, and where x_t is a $K \times 1$ vector of predetermined covariates and δ is a $K \times 1$ vector of regression coefficients. Since all components are allowed to change over time, elements of δ can also be allowed to change over time. A typical specification for a time-varying element in δ is one of those discussed as a time-varying trend function. However, when we aim to establish stable relationships between a dependent variable and a set of explanatory variables, we should keep δ constant for the full sample or, for at least, a large part of the sample.

3 Linear Gaussian state space models

The state space form provides a unified representation of a wide range of linear time series models, see Harvey (1989), Kitagawa and Gersch (1996) and Durbin and Koopman (2001). The linear Gaussian state space form consists of a transition equation and a measurement equation. We formulate the model as in de Jong (1991), that is

$$y_t = Z_t \alpha_t + G_t \epsilon_t, \qquad \alpha_{t+1} = T_t \alpha_t + H_t \epsilon_t, \qquad \epsilon_t \sim \text{NID}\,(0, I)\,, \tag{20}$$

for $t = 1, \ldots, n$, and where ϵ_t is a vector of serially independent disturbance series. The $m \times 1$ state vector α_t contains the unobserved components and their associated variables. The measurement equation is the first equation in (20) and it relates the observation y_t to the state vector α_t through the signal $Z_t \alpha_t$. The transition equation is the second equation in (20) and it is used to formulate the dynamic processes of the unobserved components in a companion form. The deterministic matrices T_t, Z_t, H_t and G_t, possibly time-varying, are referred to as system matrices and they will often be sparse and known matrices. Specific elements of the system matrices may be specified as functions of an unknown parameter vector.

3.1 Unobserved component models in state space form

To illustrate how the unobserved components discussed in Section 2 can be formulated in the state space form (20), we present a number of illustrations.

Basic structural model Consider the model $y_t = \mu_t + \gamma_t + \varepsilon_t$ with trend component μ_t as in (3), seasonal component γ_t as in (8) with seasonal length $S = 4$ (quarterly data) and irregular ε_t as in (1). We require a state vector of five elements and a disturbance vector of four elements; they are given by

$$\alpha_t = (\mu_t, \beta_t, \gamma_t, \gamma_{t-1}, \gamma_{t-2})'\,, \qquad \epsilon_t = (\varepsilon_t, \eta_t, \zeta_t, \omega_t)'\,.$$

The state space formulation of the basic decomposition model is given by (20) with the

system matrices

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

$$Z_t = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \end{pmatrix}, \qquad G_t = \begin{pmatrix} \sigma_\varepsilon & 0 & 0 & 0 \end{pmatrix}.$$

Here the system matrices T_t, H_t, Z_t and G_t do not depend on t; the matrices are time-invariant. The standard deviations of the disturbances in H_t and G_t are fixed, unknown and need to be estimated. The corresponding variances are σ_η^2, σ_ζ^2, σ_ω^2 and σ_ε^2. It is common practice to transform the variances into logs for the purpose of estimation; the log-variances can be estimated without constraints. The unknown parameters are collected in the 4×1 parameter vector θ. Estimation of θ can be carried out by the method of maximum likelihood; see Section 3.4.

For the trend component μ_t in (3) the initial variables μ_1 and β_1 are treated as unknown coefficients. For the dummy seasonal component γ_t in (8) with $S = 4$, the initial variables γ_1, γ_0 and γ_{-1} are also treated as unknown coefficients. Given the composition of the state vector above, we can treat α_1 as a vector of unknown coeffients. We can estimate α_1 simultaneously with θ by the method of maximum likelihood or we can concentrate α_1 from the likelihood function. We discuss the initialization issues further in Section 3.4.

Smooth trend plus ARMA model Consider the model $y_t = \mu_t + \psi_t + \varepsilon_t$ with trend component μ_t as in (5) with $k = 3$, cycle component γ_t as the ARMA process (15) with $p = 2$ and $q = 1$ and irregular ε_t as in (1). The state and disturbance vectors are given by

$$\alpha_t = \left(\mu_t^{(3)}, \mu_t^{(2)}, \mu_t^{(1)}, \psi_t, \alpha_{5,t}\right)', \qquad \epsilon_t = (\varepsilon_t, \zeta_t, \xi_t)'.$$

The state space formulation has the system matrices

$$T_t = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \varphi_{\psi,1} & 1 \\ 0 & 0 & 0 & \varphi_{\psi,2} & 0 \end{bmatrix}, \quad H_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \sigma_\zeta & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \vartheta_{\psi,1}\sigma_\xi \end{bmatrix},$$

$$Z_t = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \end{pmatrix}, \qquad G_t = \begin{pmatrix} \sigma_\varepsilon & 0 & 0 \end{pmatrix}.$$

The system matrices are time-invariant. The unknown disturbance variances need to be estimated together with the ARMA coefficients $\varphi_{\psi,1}$, $\varphi_{\psi,2}$ and $\vartheta_{\psi,1}$. In particular, we estimate three log-variances and transform the ARMA coefficients such that ψ_t is

stationary and the moving average polynomial $\vartheta_\psi(L)$ is non-invertible. The unknown parameters are collected in the 6×1 parameter vector θ.

For the smooth trend component μ_t in (5) the initial variables $\mu_1^{(3)}$, $\mu_1^{(2)}$ and $\mu_1^{(1)}$ are treated as unknown coefficients. The initial conditions for the ARMA process can be determined from the unconditional autocovariance function; see Section 3.4. In this case we treat a part of the initial state vector as unknown coeffients (first three elements) while for the remaining part we need to derive its statistical properties.

Random walk plus smooth cycle model Consider the model $y_t = \mu_t + \psi_t + \varepsilon_t$ with the random walk process μ_t as in (2), cycle component γ_t as the smooth cycle process (17) with $m = 2$, and irregular ε_t as in (1). The state and disturbance vectors are given by

$$\alpha_t = \left(\mu_t, \psi_t^{(2)}, \psi_t^{(2)+}, \psi_t^{(1)}, \psi_t^{(1)+}\right)', \qquad \epsilon_t = \left(\varepsilon_t, \eta_t, \kappa_t, \kappa_t^+\right)',$$

and the corresponding system matrices by

$$T_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & T_t^{22} & T_t^{23} & 1 & 0 \\ 0 & -T_t^{23} & T_t^{22} & 0 & 1 \\ 0 & 0 & 0 & T_t^{22} & T_t^{23} \\ 0 & 0 & 0 & -T_t^{23} & T_t^{22} \end{bmatrix}, \quad H_t = \begin{bmatrix} 0 & \sigma_\eta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\kappa & 0 \\ 0 & 0 & 0 & \sigma_\kappa \end{bmatrix},$$

$$Z_t = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \end{pmatrix}, \qquad G_t = \begin{pmatrix} \sigma_\varepsilon & 0 & 0 & 0 \end{pmatrix},$$

where $T_t^{22} = \varphi_\psi \cos \lambda_c$ and $T_t^{23} = \varphi_\psi \sin \lambda_c$. The system matrices are time-invariant. The unknown disturbance variances need to be estimated together with the discount factor φ_ψ and cycle frequency λ_c. In particular, we consider three log-variances and enforce restrictions $0 < \varphi_\psi < 1$ and $0 < \lambda_c < \pi$ via transformations. The unknown parameters are collected in the 5×1 parameter vector θ.

For the random walk component μ_t in (2), the initial variable μ_1 is treated as an unknown coefficient. The initial conditions for the smooth cycle process can be obtained from Trimbur (2006). The first element of the initial state vector is treated as unknown while the remaining part has known statistical properties.

3.2 Kalman filter

Consider the linear Gaussian state space model (20). The predictive estimator of the state vector α_{t+1} is a linear function of the observations $y_1, \ldots, y_t$. The Kalman filter computes the minimum mean square linear estimator (MMSLE) of the state vector α_{t+1} conditional on the observations $y_1, \ldots, y_t$, denoted by $a_{t+1|t}$, together with its mean square error (MSE) matrix, denoted by $P_{t+1|t}$. We will also refer to $a_{t+1|t}$ as the state prediction estimate with $P_{t+1|t}$ as its

state prediction error variance matrix. The Kalman filter is given by

$$\begin{aligned} v_t &= y_t - Z_t a_{t|t-1}, & F_t &= Z_t P_{t|t-1} Z_t' + G_t G_t', \\ & & M_t &= T_t P_{t|t-1} Z_t' + H_t G_t', \qquad t = 1, \ldots, n, \\ a_{t+1|t} &= T_t a_{t|t-1} + K_t v_t, & P_{t+1|t} &= T_t P_{t|t-1} T_t' + H_t H_t' - K_t M_t', \end{aligned} \tag{21}$$

with Kalman gain matrix $K_t = M_t F_t^{-1}$, and for particular initial values $a_{1|0}$ and $P_{1|0}$. The one-step ahead prediction error is $v_t = y_t - \mathrm{E}(y_t|y_1, \ldots, y_{t-1})$ with variance $\mathrm{Var}(v_t) = F_t$. The innovations have mean zero and are serially independent by construction so that $E(v_t v_s') = 0$ for $t \neq s$ and $t, s = 1, \ldots, n$.

Before the MMSLE $a_{t+1|t}$ and the MSE $P_{t+1|t}$ are computed in the Kalman filter, the MMSLE of the state vector α_t conditional on $y_1, \ldots, y_t$, denoted by $a_{t|t}$, and its corresponding MSE matrix, denoted by $P_{t|t}$, can be computed as

$$a_{t|t} = a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} v_t, \qquad P_{t|t} = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}, \tag{22}$$

It then follows that

$$a_{t+1|t} = T_t a_{t|t}, \qquad P_{t+1|t} = T_t P_{t|t} T_t' + H_t H_t'.$$

Formal proofs of the Kalman filter can be found in Anderson and Moore (1979), Harvey (1989) and Durbin and Koopman (2001). However, the proof of the Kalman filter and related results can be derived by the use of the following basic lemma.

Recursive lemma: Suppose that x, y and z are random vectors of arbitrary orders that are jointly normally distributed with means μ_p and (co)variances $\Sigma_{pq} = \mathrm{E}\left[(p - \mu_p)(q - \mu_q)'\right]$ for $p, q = x, y, z$ and with $\mu_z = 0$ and $\Sigma_{yz} = 0$. The symbols x, y, z, p, q, μ and Σ are employed for convenience and these definitions hold only here. Then:

$$\mathrm{E}(x|y, z) = \mathrm{E}(x|y) + \Sigma_{xz}\Sigma_{zz}^{-1} z, \qquad \mathrm{Var}(x|y, z) = \mathrm{Var}(x|y) - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}.$$

The proof of this lemma can be obtained from multivariate normal regression theory; see, for example, Anderson (1984). The elementary nature of this lemma drives home the point that the theoretical basis of state space analysis is simple.

3.3 Likelihood evaluation

The Kalman filter can be used to evaluate the Gaussian likelihood function via the prediction error decomposition, see Schweppe (1965), Jones (1980) and Harvey (1989, §3.4). Given a model as described in Section 2 for y_t, we denote the joint density of $y_1, \ldots, y_n$ by $p(y_1, \ldots, y_n)$ and the prediction error decomposition is then given by

$$p(y_1, \ldots, y_n) = p(y_1) \prod_{t=2}^{n} p(y_t|y_1, \ldots, y_{t-1}).$$

The predictive density $p(y_t|y_1,\ldots,y_{t-1})$ is Gaussian and has mean $\mathrm{E}(y_t|y_1,\ldots,y_{t-1}) = Z_t a_{t|t-1}$ and variance $\mathrm{Var}(y_t|y_1,\ldots,y_{t-1}) = Z_t P_{t|t-1} Z_t' + G_t G_t' = F_t$. For a realized time series $y_1,\ldots,y_n$, the log-likelihood function is given by

$$\begin{aligned}\ell = \log p\left(y_1,\ldots,y_n\right) &= \sum_{t=1}^{n} \log p\left(y_t|y_1,\ldots,y_{t-1}\right) \\ &= -\frac{n}{2}\log\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{n}\log|F_t| - \frac{1}{2}\sum_{t=1}^{n} v_t' F_t^{-1} v_t. \end{aligned} \tag{23}$$

The one-step ahead prediction errors v_t and their variances F_t are computed by the Kalman filter for a given value of the parameter vector θ. To make the dependence of the likelihood function on the parameter vector θ explicit, we can write $\ell = \ell(\theta)$.

3.4 Parameter estimation

In a state space analysis we are concerned with two groups of parameters that need to be estimated for a given model specification. The first group is contained in parameter vector θ, see Section 3.1 for an illustration. The second group consists of initial variables for the unobserved (non-stationary) processes and the regression coefficients such as δ in (19). The initial conditions for unobserved stationary processes can be derived from the theoretical autocovariance function.

Maximum likelihood estimation of θ: The log-likelihood function (23) can be maximised with respect to θ numerically using a numerical quasi-Newton method. For example, the method of Broyden-Fletcher-Goldfarb-Shanno (BFGS) is generally regarded as computationally efficient in terms of convergence speed and numerical stability; see Nocedal and Wright (1999). The BFGS iterative optimization method is based on information from the gradient (or score). Analytical and computationally fast methods for computing the score for a current value of θ in a state space analysis are developed by Koopman and Shephard (1992). The BFGS method is terminated when some pre-chosen convergence criterion is satisfied. The convergence criterion is usually based on the gradient evaluated at the current estimate, the parameter change compared to the previous estimate or the likelihood value change compared to the previous estimate. The number of iterations required to satisfy these criteria depends on the choice of the initial parameter values, the tightness of the chosen criterion and the shape of the likelihood surface.

An alternative method for maximum likelihood estimation is the EM-algorithm; see Shumway and Stoffer (1982) and Watson and Engle (1983) in the context of a state space analysis. The basic EM procedure works roughly as follows. Consider the joint density $p(y_1,\ldots,y_n,\alpha_1,\ldots,\alpha_n)$. The Expectation (E) step takes the expectation of the state vectors conditional on $y_1,\ldots,y_n$ and the Maximization (M) step maximizes the resulting expression with respect to θ. The E step requires the evaluation of the estimated state vector using a smoothing algorithm related to the Kalman filter, see de Jong (1989). The M step is usually carried out analytically and is simpler than maximizing the full

likelihood function directly. Given the "new" estimate of θ from the M step, we return to the E step and evaluate the smoothed estimates based on the new estimate. This iterative procedure converges to the maximum likelihood estimate of θ. Under fairly weak conditions it can be proven that each iteration of the EM algorithm increases the value of the likelihood. The EM converges to a maximum of the likelihood as a result. In practice it is often found that while the EM gets to a neighbourhood of the maximum quickly, it converges to the maximum slowly. Therefore a mix of EM and direct maximization is often advocated. In case θ only contains parameters in G_t and H_t, Koopman (1993) shows that the EM can be modified toward a fast and simple procedure.

Estimation of initial states and regression coefficients: The non-stationary trend and seasonal components, as discussed in Section 2, rely on initial variables that are treated as fixed unknown coefficients. When regression effects are added to the model, we also have fixed unknown regression coefficients. In the illustrations in Section 3.1 it is shown that these initial states and regression coefficients are collectively placed in the initial state vector α_1. We can therefore concentrate on the estimation of the initial state vector α_1.

Preferably we estimate α_1 jointly with θ by the method of maximum likelihood as discussed above. However, numerical problems may arise when the likelihood function is maximised with respect to a high-dimensional parameter vector that joins θ and α_1. Fortunately, the direct maximization with respect to α_1 can be avoided since the one-step ahead prediction error v_t is a linear function of the initial state α_1, that is $v_t = v_t^o + v_t^\alpha \alpha_1$ where v_t^o is equal to v_t when the Kalman filter (21) is started with $a_{1|0} = 0$ and $P_{1|0} = 0$ and v_t^α is a function of the system matrices Z_t, T_t, G_t and H_t. Given this linear dependence, the initial state vector can be concentrated out from the log-likelihood function in the usual way. We then maximize the concentrated likelihood with respect to θ. The implementation of this approach is developed by Rosenberg (1973).

Tunnicliffe-Wilson (1989) and Harvey and Shephard (1990) argue convincingly that the maximum likelihood estimation of α_1 can lead to bias in the estimation of unknown variances in θ; for example, it can increase the probability that a variance is estimated as zero while the true variance is not zero. They advocate the estimation of θ via the maximimization of a *marginal* or *diffuse* likelihood function with respect to initial state α_1. In a state space analysis, this approach can be embedded within a unified treatment for the initialization of the Kalman filter with respect to initial states and regression coefficients; see Ansley and Kohn (1985), de Jong (1991) and Koopman (1997). It is recently argued by Francke, Koopman, and de Vos (2010) that the strict implementation of the marginal likelihood function for models with initial states and regression coefficients is preferred for parameter estimation.

Stationary conditions for the initial state: When the state vector only contains stationary variables, the initial conditions for α_1 can be obtained from the theoretical autocovariance function. In a time-invariant stationary state space model we have $\alpha_{t+1} = T\alpha_t + H\epsilon_t$

with $\mathrm{E}(\alpha_t) = 0$ and $P = \mathrm{Var}(\alpha_t)$ for $t = 1, \ldots, n$. It follows that $P = TPT + HH'$ with solution

$$\mathrm{vec}(P^*) = (I - T \otimes T)^{-1}\mathrm{vec}(HH').$$

Efficient algorithms for solving Riccati equations can be used to compute P^* when its dimension is large, as discussed in Anderson and Moore (1979) and Hindrayanto, Koopman, and Ooms (2010). Since this solution also applies to α_1, we can initialize the Kalman filter (21) with $a_{1|0} = 0$ and $P_{1|0} = P^*$.

In most models, the initial state vector α_1 contains initial variables and regression coefficients as well as stationary variables; see also the illustrations in Section 3.1. The Kalman filter initialization methods of de Jong (1991) and Koopman (1997) account for such general model specifications.

3.5 Diagnostic checking

The assumptions underlying the models in Section 2 are that all disturbances, such as ε_t, η_t and κ_t, are normally distributed, are serially and mutually independent and have constant variances. Under these assumptions the standardised one-step ahead prediction errors (or *prediction residuals*) are given by

$$e_t = \frac{v_t}{\sqrt{F_t}}, \qquad t = 1, \ldots, n. \tag{24}$$

The prediction residuals are also normally distributed and serially independent with unit variance. We can investigate whether these properties hold by means of the following large-sample diagnostic tests:

Normality: The first four moments of the standardised forecast errors are given by

$$m_1 = \frac{1}{n}\sum_{t=1}^{n} e_t, \qquad m_q = \frac{1}{n}\sum_{t=1}^{n}(e_t - m_1)^q, \qquad q = 2, 3, 4.$$

Skewness and kurtosis are denoted by M_3 and M_4, respectively, and when the model assumptions are valid they are asymptotically normally distributed as

$$M_3 = \frac{m_3}{\sqrt{m_2^3}} \sim \mathrm{N}\left(0, \frac{6}{n}\right), \qquad M_4 = \frac{m_4}{m_2^2} \sim \mathrm{N}\left(3, \frac{24}{n}\right).$$

see Bowman and Shenton (1975). Standard statistical tests can be used to check whether the observed values of M_3 and M_4 are consistent with their asymptotic densities. They can also be combined as

$$M_N = n\left\{\frac{S^2}{6} + \frac{(K-3)^2}{24}\right\},$$

which asymptotically has a χ^2 distribution with two degrees of freedom under the null hypothesis that the normality assumption is valid. The *QQ plot* is a graphical display

of ordered residuals against their theoretical quantiles. The 45 degree line is taken as a reference line (the closer the residual plot to this line, the better the match).

Heteroscedasticity: A simple test for heteroscedasticity is obtained by comparing the sum of squares of two exclusive subsets of the sample. For example, the statistic

$$H(h) = \frac{\sum_{t=n-h+1}^{n} e_t^2}{\sum_{t=1}^{h} e_t^2},$$

is $F_{h,h}$-distributed for some preset positive integer h, under the null hypothesis of homoscedasticity.

Serial correlation: The correlogram of the prediction residuals should not reveal significant serial correlation. A standard portmanteau test statistic for serial correlation is based on the Box-Ljung statistic suggested by Ljung and Box (1978). This is given by

$$Q(k) = n(n+2)\sum_{j=1}^{k} \frac{c_j^2}{n-j},$$

for some positive integer k, where c_j is the jth correlation:

$$c_j = \frac{1}{nm_2}\sum_{t=j+1}^{n} (e_t - m_1)(e_{t-j} - m_1).$$

Although these statistics can be used for formal hypothesis testing, in practice they are used as diagnostic tests. Diagnostic graphic tools can be even more informative and they include a time series plot, a histogram and a correlogram of the prediction residuals.

4 Forecasting

The unobserved component time series model and its state space analysis is used for a model-based approach to the forecasting of economic time series. A convenient property of the Kalman filter and related methods is their ability to account for missing observations in a data set. In a relatively straightforward manner, the filter can be amended when it is confronted with missing data. Some calculations are skipped while other calculations do not need to be changed. This feature is of high practical relevance as many data-sets have at least some data points not available. In our context, it also offers a solution to the forecasting problem since we can regard the future observations as a set of missing observations. As a consequence, the Kalman also delivers all necessary computations for forecasting.

Since the model is linear, the forecast function for y_{n+h} is a linear function of the observations $y_1, \ldots, y_n$, for $h = 1, 2, \ldots$. We first consider forecasting as filling in missing values, and then discuss the computation of observation weights for forecasting.

4.1 Missing values and forecasting

The Kalman filter produces one-step ahead predictions of the state vector as denoted by $a_{t+1|t}$ with its error variance matrices $P_{t+1|t}$ for $t = 1, \ldots, n$. In the Kalman filter, if y_τ is missing, we do not know its value or its one-step ahead prediction error v_τ. The missing information on v_τ can be reflected by having $F_\tau \to \infty$ as it indicates that we have no information about v_τ. The consequences of having $F_\tau \to \infty$ in the Kalman filter is that $K_\tau \to 0$ while the remaining computations in the Kalman filter can still be carried out. The prediction step of the Kalman filter reduces to

$$a_{t+1|t} = T_t a_{t|t-1}, \qquad P_{t+1|t} = T_t P_{t|t-1} T_t' + H_t H_t', \tag{25}$$

for $t = \tau$ as $F_\tau \to \infty$. Note that $a_{t|t} = a_{t|t-1}$ and $P_{t|t} = P_{t|t-1}$ for $t = \tau$. The implementation of a Kalman filter with missing data entries is straightforward and relies simply on a conditional statement: if y_t is observed, carry out the Kalman filter as in (21); if y_t is missing, carry out the prediction step (25). Missing entries are allowed throughout the data sample $y_1, \ldots, y_n$, individually and in blocks.

The treatment of missing values can be adopted to the computation of forecasts and their forecast error variances. After the last observation, we add a series of missing values to the data set and carry on with the Kalman filter. It treats the future observations as missing values in the way described above. We then effectively obtain the state prediction estimates $a_{n+h|n}$ and its prediction error variance matrix $P_{n+h|n}$ for $h = 1, 2, \ldots$. The observation forecasts $\hat{y}_{n+h|n} = \mathrm{E}(y_{n+h}|y_1, \ldots, y_n)$ and its error variance matrix $V_{n+h|n} = \mathrm{Var}(y_{n+h} - \hat{y}_{n+h}|y_1, \ldots, y_n)$ are then computed by

$$\hat{y}_{n+h|h} = Z_{n+h} a_{n+h|n}, \qquad V_{n+h|n} = Z_{n+h} P_{n+h|n} Z_{n+h}' + H_{n+h} H_{n+h}',$$

for $h = 1, 2, \ldots$. This simple treatment of missing observations and forecasting is one of the attractions of state space analysis.

4.2 Observation weights of forecast function

It is of interest to know how observations are weighted when forecasting future observations. For a linear time series model with time-varying components, the forecasting weights should gradually decline for observations further from the forecast origin as they become less relevant. When time-variations are more volatile, weights should decline more rapidly compared to cases where time-variations are smooth. A special algorithm for computing forecasting weights in a state space analysis is discussed in Koopman and Harvey (2003). Here we concentrate on the observation weights for $y_1, \ldots, y_n$ when forecasting observation y_{n+h}. For the Gaussian linear state space model (20), the MMSLE of y_{n+h} is a linear function of the observations $y_1, \ldots, y_n$, that is

$$\mathrm{E}(y_{n+h}|y_1, \ldots, y_n) = \hat{y}_{n+h|n} = \sum_{j=1}^{n} w_j(\hat{y}_{n+h|n}) y_j, \tag{26}$$

where $w_j(\hat{x})$ represents the weight associated with observation y_j for the computation of $\hat{x}$.

The predicted state estimate for α_t is based on the observations $y_1, \ldots, y_{t-1}$ and can be written as

$$a_{t|t-1} = \sum_{j=1}^{t-1} w_j(a_{t|t-1}) y_j. \tag{27}$$

The Kalman filter (21) provides the means to compute these weights by storing the Kalman gain matrices. Then the following backward recursion is implemented,

$$w_j(a_{t|t-1}) = B_{t,j} K_j, \qquad B_{t,j-1} = B_{t,j} T_j - w_j(a_{t|t-1}) Z_j, \qquad j = t-1, t-2, \ldots, 1, \tag{28}$$

with initialization $B_{t,t-1} = I$. It computes the observations weights for the state prediction $a_{t|t-1}$. 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

$$w_j(\hat{y}_{t|t-1}) = b_{t,j} K_j, \qquad b_{t,j-1} = b_{t,j} T_j - w_j(\hat{y}_{t|t-1}) Z_j, \qquad j = t-1, t-2, \ldots, 1, \tag{29}$$

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

When missing values are present in the data set, the Kalman filter can still be applied. Assume that y_τ is missing, recursion (28) at time $j = \tau$ reduces to

$$w_\tau(a_{t|t-1}) = 0, \qquad B_{t,\tau-1} = B_{t,\tau} T_\tau. \tag{30}$$

Hence missing value can be accommodated in algorithms for computing observation weights in a state space analysis. The forecasting observation weights can therefore be obtained from (30). Using the initialisation of (28) to time $t = n + h$, that is $B_{n+h,n+h-1} = I$, and applying (30) for $j = n + h - 1, \ldots, n + 1$ we obtain $w_n(a_{n+h|n}) = T_{n+h-1} \ldots T_{n+1} K_n$. The weights for $w_j(a_{\tau|n})$ are computed by (28) when y_j is observed and by (30) when y_j is missing, for $j = n-1, \ldots, 1$. The forecasting observation weights are then computed by $Z_t w_j(a_{t|t-1})$ or directly via (29).

4.3 Autoregressive representation of model

The weights also lead to the autoregressive representation of any unobserved components time series model in Section 2. It follows from the Kalman filter prediction error equation that $y_t = Z_t a_{t|t-1} + v_t$ where $v_t \sim \text{NID}(0, F_t)$. Furthermore, it follows from (27) and (28) that we can express y_t as

$$y_t = \sum_{i=0}^{\infty} \phi_i y_{t-i} + v_t, \qquad \phi_i = b_{t,t-i} K_{t-i}, \qquad v_t \sim \text{NID}(0, F_t). \tag{31}$$

This specification is effectively a valid autoregressive model representation of a time series y_t that is generated by an unobserved components time series model. The infinite lag order is in practice not necessary as $b_{t,t-i} \to 0$ as $i \to \infty$ with t fixed. We therefore conclude that $\phi(L) y_t = v_t$ with autoregressive lag polynomial $\phi(L)$ provides a valid representation of any unobserved components time series model for y_t.

5 Multivariate components

In Section 2 we have set out a comprehensive class of unobserved components time series models. In economic theory one focuses on the dynamic relationships between variables. Hence the need of econometricians to simultaneously analyze and model a multiple set of related time series. The multivariate analysis of time series is a challenging task because the dynamic interactions between time series can be intricate and the number of parameters in a model can increase rapidly. In this section we will highlight a number of multivariate extensions of decomposition models together with a number of applications.

5.1 Multivariate trend model

The decomposition models can easily be extended for the modelling of multivariate time series. For example, letting y_t denote a $p \times 1$ vector of observations, the multivariate local level model for y_t is given by

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \Sigma_\varepsilon), \\ \mu_{t+1} &= \mu_t + \xi_t, \qquad \xi_t \sim \text{NID}(0, \Sigma_\xi), \end{aligned} \tag{32}$$

for $t = 1, \ldots, n$, where μ_t, ε_t, and ξ_t are $p \times 1$ vectors and Σ_ε and Σ_ξ are $p \times p$ variance matrices. In what is known as the *seemingly unrelated time series equations model* (32), the series are modelled as in the univariate situation, but the disturbances driving the level components are allowed to be instantaneously correlated across the p series. When slope, seasonal, or cycle components are involved, each of these three components also has an associated $p \times p$ variance matrix allowing for correlated disturbances across series.

The dynamic properties implied by the trend decomposition model (32) further depend on the specifications of the variance matrices Σ_ε and Σ_ξ. When both variance matrices are of full rank, the dynamic interactions between the time series can alternatively represented by

$$y_t = \Lambda_\xi \mu_t^\dagger + \Lambda_\varepsilon \varepsilon_t^\dagger, \qquad \mu_{t+1}^\dagger = \mu_t^\dagger + \xi_t^\dagger, \qquad \varepsilon_t^\dagger \sim \text{NID}(0, \mathcal{D}_\varepsilon), \qquad \xi_t^\dagger \sim \text{NID}(0, \mathcal{D}_\xi), \tag{33}$$

where the various terms are defined implicitly by relating the terms in (32) with those in (33) via

$$\mu_t = \Lambda_\xi \mu_t^\dagger, \qquad \varepsilon_t = \Lambda_\varepsilon \varepsilon_t^\dagger, \qquad \Sigma_\varepsilon = \Lambda_\varepsilon \mathcal{D}_\varepsilon \Lambda_\varepsilon', \qquad \Sigma_\xi = \Lambda_\xi \mathcal{D}_\xi \Lambda_\xi',$$

where $\mathcal{D}_\varepsilon$ and $\mathcal{D}_\xi$ are $p \times p$ variance matrices. Since we have assumed full rank variance matrices, it is also true that $\mu_t^\dagger = \Lambda_\xi^{-1} \mu_t$ and, similarly, $\varepsilon_t^\dagger = \Lambda_\varepsilon^{-1} \varepsilon_t$. The representation (33) shows in a more transparent, direct way how the time series relate to each other. The loading matrix Λ_ξ typically determines the long-term movements or dynamics between the variables whereas the loading matrix Λ_ε links the contemporaneous shocks in the time series.

The matrices Λ_x and $\mathcal{D}_x$ can be regarded as the result of the variance matrix decomposition of Σ_x, for $x = \varepsilon, \xi$. The variance decomposition $\Sigma_x = \Lambda_x \mathcal{D}_x \Lambda_x'$ is not unique, for $x = \varepsilon, \xi$. Since the number of coefficients in Σ_x is $\frac{1}{2}p(p+1)$, all elements in the $p \times p$ matrices Λ_x and $\mathcal{D}_x$ cannot be identified in the model. An appropriate set of identification restrictions are obtained

by assuming that Λ_x is a lower (or upper) triangular matrix with unit values on the diagonal and that $\mathcal{D}_x$ is a diagonal matrix consisting of positive values. The restrictions imply the Cholesky decomposition of Σ_x. For given values of Λ_x and $\mathcal{D}_x$, the trend can still be transformed without affecting the model for y_t itself. For all orthonormal $p \times p$ matrices B and C, such that $B'B = I_p$ and $C'C = I_p$, we can reformulate the model as

$$y_t = \Lambda_\xi^* \mu_t^* + \Lambda_\varepsilon^* \varepsilon_t^*, \qquad \mu_{t+1}^* = \mu_t^* + \xi_t^*, \qquad \varepsilon_t^* \sim \text{NID}(0, C\mathcal{D}_\varepsilon C'), \qquad \xi_t^* \sim \text{NID}(0, B\mathcal{D}_\xi B'), \quad (34)$$

where

$$\Lambda_\xi^* = \Lambda_\xi B', \qquad \mu_t^* = B\mu_t^\dagger, \qquad \Lambda_\varepsilon^* = \Lambda_\varepsilon C', \qquad \varepsilon_t^* = C\varepsilon_t^\dagger,$$

for $t = 1, \ldots, n$. The transformations based on B and C can be exploited to obtain a loading structure that suits an economic interpretation. We emphasize that the statistical dynamic properties of y_t are the same for all model specifications (32), (33) and (34).

5.2 Common trends and cycles

When the variance matrix of the trend disturbance Σ_ξ has not full rank, the multivariate local level model (32) implies a common trend component for y_t. In other words, when $\text{rank}(\Sigma_\xi) = r < p$, the underlying trends of the p time series in y_t depend on a smaller set of r common trends. In terms of the model representation (33), the dimensions of the matrices Λ_ξ and $\mathcal{D}_\xi$ are $p \times r$ and $r \times r$, respectively. Hence, the trend vector $\mu_t^\dagger$ represents the common trends and has dimension $r \times 1$. Since the time series in y_t can all have different locations. The locations of r time series can be determined by the r trends in $\mu_t^\dagger$. The locations of the remaining $p - r$ time series in y_t are then adjusted by the constant vector $\bar{\mu}$ in

$$y_t = \bar{\mu} + \Lambda_\xi \mu_t^\dagger + \varepsilon_t, \qquad \mu_{t+1}^\dagger = \mu_t^\dagger + \xi_t^\dagger, \quad (35)$$

where $\bar{\mu}$ consists of r zero and $p - r$ non-zero values. Common trends in a model allows interesting economic relations and are related to the concept of cointegration, see Stock and Watson (1988) and Anderson and Vahid (2011, this volume) where common cycles and trends are studied using vector autoregressive models.

Common dynamics can also be introduced for other unobserved components in the model. In particular, common drifts and common cycles are of interest in economic time series. The basic formulation of a model with common trends and cycles is given by

$$y_t = \bar{\mu} + \Lambda_\xi \mu_t^\dagger + \Lambda_\kappa \psi_t^\dagger + \varepsilon_t, \quad (36)$$

where $\mu_t^\dagger$ is the $r_\mu \times 1$ vector of common trends and vector $\psi_t^\dagger$ contains the r_ψ common cycles. The loading matrices Λ_ξ and Λ_κ have dimensions $p \times r_\mu$ and $p \times r_\psi$, respectively. We can adopt one of the cycle specifications discussed in Section 2.3 and generalize these to multivariate processes. For example, a multivariate version of the ARMA process (15) can be considered, see Shumway and Stoffer (2006, Chapter 5.7). The multivariate version of the cycle process (16) is known as the similar cycle since the discount factor φ_ψ and the cycle frequency λ_c are

common to all individual cycles, see the discussion in Harvey and Koopman (1997). We define the similar cycle process for $\psi_t^\dagger$ in (36) by

$$\begin{pmatrix} \psi_{t+1}^\dagger \\ \psi_{t+1}^+ \end{pmatrix} = \varphi_\psi \left\{ \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \otimes I_{r_\psi} \right\} \begin{pmatrix} \psi_t^\dagger \\ \psi_t^+ \end{pmatrix} + \begin{pmatrix} \kappa_t^\dagger \\ \kappa_t^+ \end{pmatrix}, \tag{37}$$

where the auxiliary cycle vector ψ_t^+ has dimension $r_\psi \times 1$, the discount factor φ_ψ and cycle frequency λ_c remain scalars and $\otimes$ is the Kronecker matrix product operator. The $r_\psi \times 1$ disturbance vectors $\kappa_t^\dagger$ and κ_t^+ together with the initial conditions for the cycle vectors are given by

$$\begin{pmatrix} \kappa_t^\dagger \\ \kappa_t^+ \end{pmatrix} \sim \text{NID}(0, I_2 \otimes \mathcal{D}_\kappa), \qquad \begin{pmatrix} \psi_1^\dagger \\ \psi_1^+ \end{pmatrix} \sim \text{NID}\left(0, \frac{1}{1-\varphi_\psi^2} I_2 \otimes \mathcal{D}_\kappa\right),$$

and the cyclical disturbance series $\kappa_t^\dagger$ and κ_t^+ are serially independent and mutually independent. It follows for the cycle component $\psi_t = \Lambda_\kappa \psi_t^\dagger$ in (36) that

$$\text{E}(\psi_t) = 0, \qquad \text{Var}(\psi_t) = \Lambda_\kappa \mathcal{D}_\kappa \Lambda_\kappa',$$

for $t = 1, \ldots, n$. The individual cycle processes in $\psi_t^\dagger$ are mutually independent of each other while those in ψ_t are correlated with each other. This multivariate generalization of the cycle process (16) can also be applied to the smooth cycle process (17).

In the decomposition model (36) for y_t with trend and cycle components, only time series with coincident cycles are viable candidates to be included in the model for y_t. It can be of economic interest to investigate whether leads or lags of economic variables are appropriate for its inclusion in y_t. For this purpose, the model can be modified to allow the base cycle ψ_t to be shifted for each time series. The phase shift mechanism proposed by Rünstler (2004) allows the cycle process ψ_t to be shifted ν time periods to the right (when scalar $\nu > 0$) or to the left (when $\nu < 0$) by considering

$$\cos(\nu\lambda_c)\psi_t + \sin(\nu\lambda_c)\psi_t^+, \qquad t = 1, \ldots, n.$$

The shift ν is measured in real-time so that $\nu\lambda_c$ is measured in radians and due to the periodicity of trigonometric functions the parameter space of ν is restricted within the range $-\frac{1}{2}\pi < \nu\lambda_c < \frac{1}{2}\pi$. Individual cycles in ψ_t can be shifted differently by having different ν values. For the ith equation of (36), we may have

$$y_{it} = \bar{\mu}_i + \Lambda_{\xi,i}\mu_t^\dagger + \cos(\nu_i\lambda_c)\Lambda_{\kappa,i}\psi_t^\dagger + \sin(\nu_i\lambda_c)\Lambda_{\kappa,i}\psi_t^+ + \varepsilon_{it},$$

where z_{it} is the ith element of z_t for $z = y, \varepsilon$, $\bar{\mu}_i$ is the ith element of $\bar{\mu}$ and $\Lambda_{x,i}$ is the ith row of Λ_x for $x = \xi, \kappa$ with $i = 1, \ldots, p$. For identification purposes, we assume that a specific equation j contains the contemporaneous base cycle with $\nu_j = 0$. The remaining $p-1$ ν_i's can be determined uniquely and their corresponding cycles then shift with respect to the

base cycle $\Lambda_{\kappa,j}\psi_t^\dagger$. More discussions on shifted cycles together with an empirical illustration for constructing a business cycle from a panel of macroeconomic time series are provided in Azevedo, Koopman, and Rua (2006).

5.3 State space representation, estimation and forecasting

The unobserved components time series models discussed here can be represented in state space form including their multivariate versions. The multivariate trend and cycle decomposition model with common components and possibly with shifted cycles remains linear with respect to the time-varying unobserved components and can therefore be represented in state space form. Kalman filter and related methods discussed in Section 3 are applicable to multivariate time series models. The methodology of estimation and forecasting remains as for the univariate model. However, the dimensions for both the state vector α_t and the parameter vector θ are typically larger and computations are more time-consuming. It is therefore important that all necessary computations are implemented in a numerically stable and efficient manner; see the discussions in Koopman, Shephard, and Doornik (1999, 2008).

The forecasting methodology discussed in Section 4 remains applicable in the multivariate case. In many different settings it is empirically interesting to investigate whether the extra effort of analysing a set of time series simultaneously also lead to more precise forecasts in comparison to forecasts based on univariate models. A possible way to diagnose whether other time series have a high impact in the forecasting of particular series of interest in a multivariate analysis is to study the observation weights $w_j(\hat{y}_{n+h|n})$ in (26). However, even if the observation weights imply a high impact of various time series on forecasting, it remains to be seen whether empirically more precise forecasts are obtained in a multivariate setting.

5.4 Dynamic factor analysis

The multivariate model (36) for y_t can be interpreted as a dynamic factor model when the dimension of y_t is large and the dimension of $\mu_t^\dagger$ is small. In many economic applications, the dimension of y_t can be potentially very large. Stock and Watson (2011, this volume) provide an excellent survey of dynamic factor models. The task of maximum likelihood estimation and forecasting is challenging in models with high-dimensional observation vectors. Sargent and Sims (1977) and Geweke (1977) are the earliest references discussing maximum likelihood methods for dynamic factor models. For a relatively low-dimensional model for wage rates, Engle and Watson (1981) consider maximum likelihood estimation using Fisher scoring to maximize the likelihood while Watson and Engle (1983) and Shumway and Stoffer (1982) develop a particular EM method. Methods related to principal components analysis can be adopted as an alternative to maximum likelihood, see Forni, Hallin, Lippi, and Reichlin (2000), Stock and Watson (2002) and Bai (2003). The principal component method is computationally fast and easy to implement compared to maximum likelihood. Recently, there has been a renewed interest in the use of maximum likelihood estimation for high-dimensional models. Doz, Giannone, and Reichlin (2006) show that in various cases a maximum likelihood analysis produces more precise estimates of the factors than a principal component method.

Jungbacker and Koopman (2008) present results that lead to a computationally efficient procedure for parameter estimation by maximum likelihood. They argue that the observation vector y_t can be split into a low-dimensional vector and a high-dimensional vector series. For the evaluation of the likelihood function $\ell(\theta)$, the Kalman filter is only applied to the low-dimensional series while standard regression calculations suffice for the high-dimensional part of y_t. As a result, large computational gains can be achieved.

6 Empirical illustration: daily spot electricity prices

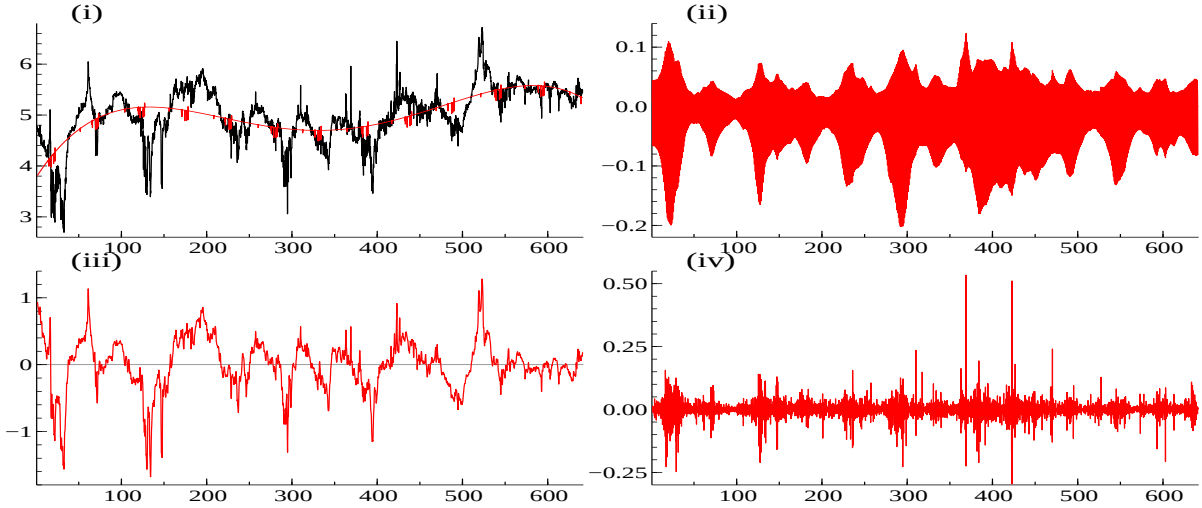
In this illustration we examine the time series of daily spot electricity prices from the Nord Pool exchange market in Norway. Nord Pool was established in 1991 as a market for trading hydroelectric power generated in Norway. Sweden joined in 1996, Finland joined in 1998 and Denmark joined in 1999. We consider only the Norwegian electricity prices. Most of this electricity is generated in hydroelectric power stations, and thus supply depends heavily on weather conditions. The average production capability of Norway's hydro power plants is about 113 Terawatt hours (1 TWh = 109 KWh) per year. Nord Pool operates as a day-ahead market and concentrates on daily trades for electricity delivered on the next day. Daily series are constructed as the average of 24 price series for the different hours of the day. The resulting prices are referred to as spot prices and are measured in Norwegian Kroner (NOK, 8 NOK $\approx$ 1 Euro). In the analysis we consider logs of spot prices from the Nord Pool electricity market from January 4, 1993 to April 10, 2005. This period covers 640 weeks or 4480 days. The daily spot prices vary over the years and are subject to yearly cycles, weekly patterns, persistent level changes, and jumps. This dataset has been analysed extensively by Koopman, Ooms, and Carnero (2007).

6.1 Univariate analysis

To accomodate the long-term movements, intra-yearly and intra-weekly (day of week) effects and the irregular jumps in daily spot prices, we include a smooth trend, a smooth cycle, a trigonometric seasonal and an irregular component (respectively) as part of the univariate unobserved components time series model. Our model of interest is (19) where $\mu_t = \mu_t^{(k)}$ is specified as the smooth trend (5) with $k = 4$, the time-varying trigonometric seasonal component is $\gamma_t = \sum_{j=1}^{3} \gamma_{j,t}$ and $\gamma_{j,t}$ is specified by (13) with $S = 7$, the smooth time-varying trigonometric cycle is $\psi_t = \psi_t^{(m)}$ as in (16) with $m = 2$, and the irregular component is $\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$. All disturbances driving the unobserved components in (19) are mutually and serially uncorrelated. By setting $k = 4$ for the trend and $m = 2$ for the cycle component, we have imposed smoothness restrictions on the decomposition. After limited prior examination, these settings produced the most satisfactory decomposition. The regression effect $x_t'\delta$ in the model consists of a set of constructed dummies for holidays and related special days; they are documented in Koopman, Ooms, and Carnero (2007). Their model includes an additional six explanatory variables (constant, time index, yearly and half-yearly cosine and sine terms) which we have excluded from the current analysis since these variables are effectively replaced by the time-varying trend and

FIGURE 1: TIME SERIES DECOMPOSITION OF DAILY SPOT ELECTRICITY PRICES

The estimated components from the univariate unobserved components time series model (19) applied to daily spot electricity prices from the Nord Pool market are presented by four plots: (i) daily spot prices and the composite estimate of the smooth trend plus regression effects, $\mu_t + x_t'\delta$; (ii) the estimated seasonal component ($S = 7$, the day-of-week effect); (iii) the estimated yearly cyclical component; (iv) estimated irregular component. The estimates are based on all observations (smooth estimates). The sample is from January 4, 1993 to April 10, 2005. The horizontal axis displays the week number for the 640 weeks in the sample. Week 1 refers to the first full week of 1993 starting on Monday.



cycle components. The resulting model (19) is extensive and requires a high-dimensional state vector when the model is placed in state space form. The STAMP 8 software package of Koopman, Harvey, Doornik, and Shephard (2008) is used for parameter estimation and forecasting. The statistical methodology adopted in STAMP is based on maximum likelihood estimation as discussed in Sections 3 and 4 while further details are provided in Durbin and Koopman (2001).

We will not discuss the estimation results in much detail. We focus on the forecasting results which are reported in Section 6.3. However, to provide an insight to whether the time series decomposition for daily spot prices implied by model (19) is appropriate, we present the estimated components in Figure 1. The estimated components are based on model (19) with the unknown coefficients replaced by their maximum likelihood estimates. The estimated components are obtained from the Kalman filter and the additional smoothing algorithm for computing estimates of the state vector based on all observations, that is $\mathrm{E}(\alpha_t|y_1, \ldots, y_n)$; see de Jong (1989) for more details.

The trend estimate is rather smooth and follows the long-term movements in the daily spot prices. The sample covers a period of more than 12 years and it includes recession periods and periods of economic growth. As electricity prices depend partly on economic conditions, we may suspect that the trend component includes the influence of economic activity on spot prices. The estimated regression effects (dummies for holidays and related days) are visualised in panel (i) of Figure 1 as shock deviations from the estimated trend. Although these shocks appear to be small in comparison with the large swings of the estimated trend, most holiday dummies are estimated as significant. The overall magnitude of the estimated day-of-week seasonal effects

is similar to the holiday effects as can be concluded from a visual inspection of panel (ii). The size of the estimated seasonal effect varies strongly over time. It appears that when the overall price level is high, the estimated seasonal day-of-week effect is less pronounced. Panel (iii) of Figure 1 presents the estimated cyclical effect. The cycle is included in the model to capture the recurring yearly effects due to winter/summer differences but also due to short-term changes in market conditions which cannot be captured by well-defined explanatory variables. The cycle estimate may have captured some yearly recurring effects but these appear to be quite erratic. The cycle estimate clearly captures other short-term effects as well. The estimate of the irregular component displayed in panel (iv) contains all sudden jumps which are typical in daily electricty spot prices. A major problem for the irregular component is that we assume it is a Gaussian disturbance. It is unlikely that the tails of the Gaussian distribution is sufficiently fat to include all jumps in the irregular. We should consider a more heavy-tailed distribution for the irregular; for example, the Student's t-distribution. Despite this last deficiency of our model specification (19), it appears that the implied model decomposition is appropriate for Nord Pool daily electricity spot prices.

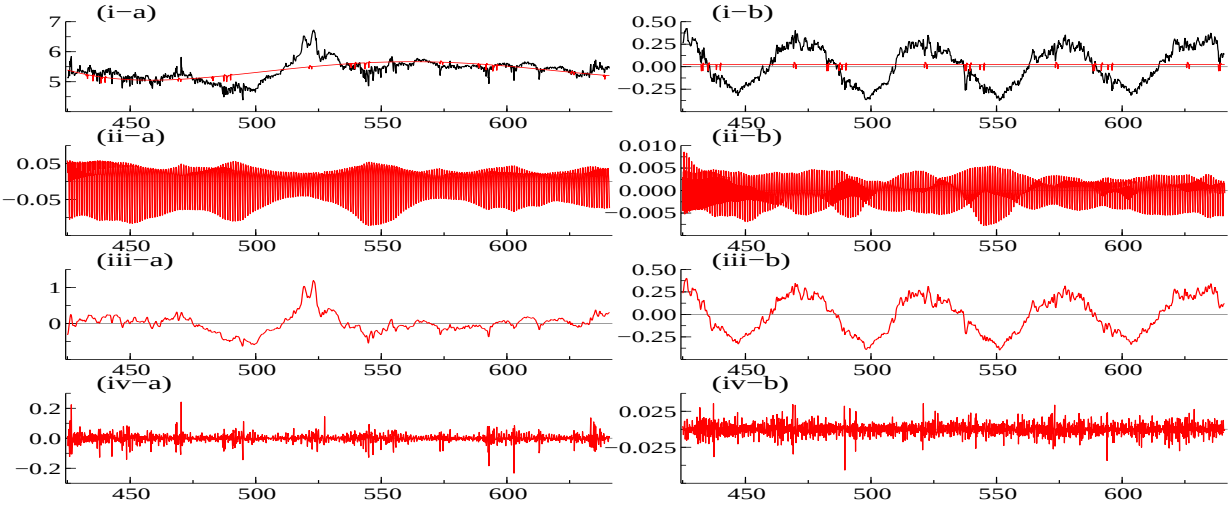
6.2 Bivariate analysis

Although univariate time series modelling of electricity prices is important in its own right, it is also interesting to extend the analysis using publicly available data on the determinants of power demand. The most relevant and closely watched variable for the hydropower market of Nord Pool is the daily data on Norwegian power consumption. We therefore will attempt to model the electricity price and consumption series jointly and investigate whether forecast precision can be improved. The unobserved components model specification remains but we let each component be a bivariate vector in (19) with full variance matrices for their corresponding disturbances. In effect, we consider a bivariate seemingly unrelated time series equation model. We further assume that all explanatory variables (for holidays and special days) apply to both equations. Since the yearly average of electricity consumption is rather constant over a number of years, we replace the stochastic trend by a constant level. The daily consumption data is only available from February 19, 2001 onwards, that is from week 425 onwards. We therefore need to reduce the sample by approximately 70%. The dimension of the state vector for the state space form is twice as large as the high-dimensional state vector for the univariate model. The STAMP 8 software package of Koopman, Harvey, Doornik, and Shephard (2008) is able to handle the bivariate model (19) with its modification for a constant level in the consumption equation. Once the model is formulated, STAMP carries out parameter estimation by maximum likelihood and forecasting of future y_t's.

The time series decomposition of the bivariate time series is presented in Figure 2. The decomposition for spot prices based on the bivariate model has not changed much compared to the decomposition based on the univariate model when we consider the estimates from week 425 onwards in Figure 1. This is a convenient result, in particular for the price forecast comparisons in the next section. Since the univariate and bivariate decompositions for the spot prices are similar, the dynamic properties of prices implied by both model specifications are also similar.

FIGURE 2: TIME SERIES DECOMPOSITION OF ELECTRICITY PRICES & CONSUMPTION

The estimated components from the bivariate unobserved components time series model (19) applied to (a) daily spot electricity prices and (b) electricity consumption, both for the Nord Pool market, are presented by two columns of four plots: (i-a) daily spot prices and the composite estimate of the smooth trend plus regression effects; (i-b) daily consumption and the composite estimate of the constant level plus regression effects; (ii-a,b) the estimated seasonal component ($S = 7$, the day-of-week effect); (iii-a,b) the estimated yearly cyclical component; (iv-a,b) estimated irregular component. The estimates are based on all observations (smooth estimates). The sample is from February 19, 2001 to April 10, 2005. The x-axis displays the week number for the 215 weeks in the sample. Week 425 refers to the first full week of 2001.



We can then measure the effect of including consumption in our analysis accurately. The decomposition for consumption seems appropriate as well. The time series is decomposed into a constant and a prominent smooth intra-yearly cycle. In addition, the time-varying seasonal (for day-of-week effect), the regression effects and the irregular component are clearly present in the series.

6.3 Forecasting results

Here we present and compare forecasting results for the univariate and bivariate models (19). These results are given for illustrative purposes only. Serious conclusions cannot be taken from these results. We show that forecasting results can vary and that a single model is not necessarily providing the most accurate results in all cases. The design of our limited forecasting study is as follows. For the univariate model, we consider four estimation samples which all start at week 1 but end at different weeks: (1) 636, (2) 637, (3) 638 and (4) 639. For these four different samples, we estimate the parameters in our model by maximum likelihood and we forecast the next seven days. Since the actual observations in our forecast horizon are available for all four samples, we can compute the (one-step, . . . , seven-step ahead) forecast errors. The weeks 638 and 639 are non-standard weeks as they are subject to calendar effects due to the special days of Maundy Thursday (March 24, 2005), Easter Monday (March 28, 2005) and those in between. It is anticipated that electricity prices and consumption drop during such periods.

We then compute the mean absolute percentage error (MAPE) for the one- to seven-days

ahead forecasts, cumulatively within every single week only. For a formal definition of MAPE and related precision measures, we refer to Makridakis, Wheelwright, and Hyndman (1998). In this study, the MAPE for day d is computed as the average of d absolute percentage errors, for $d = 1, \ldots, 7$. In Table 1, we report the MAPE values for the four different samples. We also report the MAPE values that are obtained from the spot price forecasts implied by the bivariate model. All computations are carried out by the STAMP software package.

Table 1: Forecasting results

We present the mean absolute percentage errors (MAPE) for the forecasting of one- to seven-days ahead of log average daily electricity spot prices. We have carried out the forecasting analysis based on the univariate (uni) model (19) and its bivariate (biv) counterpart with full variance matrices for each component. In the bivariate model, the second time series is log daily electricity consumption and its trend component is replaced by a fixed constant. The parameters are estimated by maximum likelihood using four different samples. The one- to seven-days ahead forecasts of the log spot prices are computed for the next seven days. The first forecast is for Monday, March 14, 2005 in Week 637. The last forecast is for Sunday, April 10, 2005 in Week 640. Weeks 638 and 639 contain calendar effects for Maundy Thursday (March 24, 2005) and the days until Easter Monday (March 28, 2005). The MAPE for Monday is based on the single error for Monday and is therefore equal to its absolute percentage error. The MAPE for Sunday summarizes the average forecasting performance for the full week (seven days).

estimation sample up to	week 636		week 637		week 638		week 639	
forecast target	week 637		week 638		week 639		week 640	
considered model is	uni	biv	uni	biv	uni	biv	uni	biv
horizon – day								
1 – Monday	0.83	1.11	0.15	0.07	0.83	1.01	0.92	0.27
2 – Tuesday	0.86	0.94	0.51	0.53	1.20	1.36	0.74	0.20
3 – Wednesday	1.43	1.55	0.67	0.79	1.40	1.52	0.62	0.16
4 – Thursday	1.94	2.09	0.64	0.88	1.71	1.75	0.60	0.14
5 – Friday	1.69	1.93	0.65	0.72	2.01	2.00	0.60	0.30
6 – Saturday	1.62	1.95	0.58	0.69	2.26	2.17	0.67	0.43
7 – Sunday	1.61	2.05	0.68	0.90	2.44	2.27	0.79	0.56

The main conclusion is that the bivariate model does not produce more precise forecasts necessarily. Although an important related variable is included in the analysis, it does not give a guarantee that the implied forecasts from a more extended model is more precise in all cases. A clear distinction between the two models is only obtained for the last sample in which the daily spot prices in week 640 are forecast. For this sample, the spot price forecasts from the bivariate model are much more precise. In the first two samples, the univariate forecasts are somewhat more precise. In the third sample, the univariate forecasts for the first three days are more precise while the forecasts from the bivariate model are more precise for the last days. We should emphasize, however, that the estimation sample for the bivariate analysis is much shorter (starting at week 425) than the estimation sample for the univariate analysis (starting at week 1). Finally, when considering multivariate models for analyzing time series, it should be kept in mind that no guarantees can be given about increases in forecast precision, relative to the forecast precision obtained from a univariate analysis.

7 Conclusion

In this paper we have discussed the state space analysis of unobserved components time series models for univariate and multivariate time series. This class of linear time series models consist of components that stochastically evolve over time and can represent trend, seasonal, cycle and other effects in a time series. Each unobserved component can capture a specific dynamic feature from the time series. It is important that the model provides an effective description of all dynamic features in the data. Although the model can be used for different purposes, an important application is forecasting. The state space analysis provides practical tools to construct an appropriate model, to estimate unknown parameters and to compute the forecasts. Diagnostic statistics and graphics based on one-step ahead prediction errors are of key importance for the formulation of a satisfactory model. The empirical model ideally implies a forecast function that can be justified by common sense. For this purpose, the observation weights in the forecast function are of interest. The illustration concerns daily electricity spot prices and it shows how an analysis based on unobserved components time series models can lead to the generation of optimal forecasts.

Although the linear Gaussian class of unobserved components time series models is general in many ways, extensions towards nonlinear and non-Gaussian features in time series are often required. Nonlinearities can be introduced in various ways. The measurement and transition equations of the state space formulation may depend on past observations. This extension opens up a wide range of models with parameters that depend on past observations. Since the Kalman filter operates by conditioning on past observations, it still yields optimal estimates. Nonlinearity also arises when the measurement and transition equations include nonlinear functions of the state vector. For example, in the macroeconomic literature much attention is given to the dynamic stochastic general equilibrium (DSGE) model which is effectively a nonlinear state space model as shown by Galí (2008). A model becomes non-Gaussian when disturbances need to be generated by heavy-tailed distributions; see the discussion of our illustration in Section 6. Nonnormality may also be intrinsic to the observations. For example, when we need to analyse a time series of small counts, a normal approximation is unreasonable and we may need to account for this by considering a Poisson distribution. Extending the unobserved component models to a non-Gaussian class is highly relevant for the modelling of economic and financial risk; for example, a credit risk application is considered by Koopman and Lucas (2008). Parameter estimation and forecasting future observations require simulation-based methods when time series models include nonlinear and/or non-Gaussian features; an introductory treatment of such methods is provided by Durbin and Koopman (2001, Part II).

References

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Englewood Cliffs: Prentice-Hall.

Anderson, O. D. and F. Vahid (2011). Common trends and cycles. See Clements and Hendry (2011).

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (2nd ed.). New York: John Wiley & Sons.

Ansley, C. F. and R. Kohn (1985). Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. *Annals of Statistics 13*, 1286–1316.

Azevedo, J. V., S. J. Koopman, and A. Rua (2006). Tracking the business cycle of the Euro area: a multivariate model-based band-pass filter. *J. Business and Economic Statist. 24*, 278–290.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–72.

Bowman, K. O. and L. R. Shenton (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika 62*, 243–50.

Burns, A. and W. Mitchell (1946). Measuring Business cycles. Working paper, NBER, New York.

Clark, P. K. (1987). The cyclical component of U.S. economic activity. *The Quarterly Journal of Economics 102*, 797–814.

Clements, M. P. and D. F. Hendry (Eds.) (2011). *Oxford Handbook on Economic Forecasting*. Oxford: Oxford University Press.

Commandeur, J. J. F. and S. J. Koopman (2007). *An Introduction to State Space Time Series Analysis*. Oxford: Oxford University Press.

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.

Doz, C., D. Giannone, and L. Reichlin (2006). A quasi maximum likelihood approach for large approximate dynamic factor models. Discussion paper, CEPR.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Engle, R. F. and M. W. Watson (1981). A one-factor multivariate time series model of metropolitan wage rates. *J. American Statistical Association 76*, 774–81.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic factor model: Identification and estimation. *Rev. Economics and Statistics 82*, 540–54.

Francke, M., S. J. Koopman, and A. de Vos (2010). Likelihood functions for state space models with diffuse initial conditions. *Journal of Time Series Analysis 31*, forthcoming.

Galí, J. (2008). *Monetary Policy, Inflation, and the Business Cycle*. Princeton: Princeton University Press.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. J. Aigner and A. S. Goldberger (Eds.), *Latent variables in socio-economic models*. Amsterdam: North-Holland.

Gomez, V. (2001). The use of Butterworth filters for trend and cycle estimation in economic time series. *J. Business and Economic Statist. 19*, 365–73.

Harrison, J. and C. F. Stevens (1976). Bayesian forecasting (with discussion). *J. Royal Statistical Society B 38*, 205–47.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge, UK: Cambridge University Press.

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.

Harvey, A. C. and N. Shephard (1990). On the probability of estimating a deterministic component in the local level model. *J. Time Series Analysis 11*, 339–47.

Harvey, A. C. and T. Trimbur (2003). Generalised model-based filters for extracting trends and cycles in economic time series. *Rev. Economics and Statistics 85*, 244–55.

Hindrayanto, I., S. J. Koopman, and M. Ooms (2010, in press). Exact maximum likelihood estimation for non-stationary periodic time series models. *Computational Statistics & Data Analysis*.

Jones, R. H. (1980). Maximum likelihood fitting of ARIMA models to time series with missing observations. *Technometrics 22*, 389–95.

Jungbacker, B. and S. J. Koopman (2008). Likelihood-based analysis for dynamic factor models. Working paper.

Kitagawa, G. and W. Gersch (1996). *Smoothness Priors Analysis of Time Series*. New York: Springer Verlag.

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 & Control 27*, 1317–1333.

Koopman, S. J., A. C. Harvey, J. A. Doornik, and N. Shephard (2008). *Stamp 8.0: Structural Time Series Analyser, Modeller and Predictor*. London: Timberlake Consultants.

Koopman, S. J. and A. Lucas (2008). A non-gaussian panel time series model for estimating and decomposing default risk. *J. Business and Economic Statist. 26*, 510–25.

Koopman, S. J., M. Ooms, and A. M. Carnero (2007). Periodic Seasonal Reg-ARFIMA-GARCH Models for Daily Electricity Spot Prices. *Journal of the American Statistical Association 102*, 16–27.

Koopman, S. J. and N. Shephard (1992). Exact score for time series models in state space form. *Biometrika 79*, 823–6.

Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal 2*, 113–66. http://www.ssfpack.com/.

Koopman, S. J., N. Shephard, and J. A. Doornik (2008). *Statistical Algorithms for Models in State Space Form: SsfPack 3.0*. London: Timberlake Consultants.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika 66*, 67–72.

Makridakis, S., S. C. Wheelwright, and R. J. Hyndman (1998). *Forecasting: Methods and Applications* (3rd ed.). New York: John Wiley and Sons.

Nocedal, J. and S. J. Wright (1999). *Numerical Optimization*. New York: Springer Verlag.

Proietti, T. (2000). Comparing seasonal components for structural time series models. *International Journal of Forecasting 16*, 247–260.

Rosenberg, B. (1973). Random coefficients models: the analysis of a cross-section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement 2*, 399–428.

Rünstler, G. (2004). Modelling phase shifts among stochastic cycles. *Econometrics Journal 7*, 232–248.

Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a priori economic theory. In C. A. S. et al. (Ed.), *New methods in business cycle research*. Minneapolis: Federal Reserve Bank of Minneapolis.

Schweppe, F. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory 11*, 61–70.

Shumway, R. H. and D. S. Stoffer (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Analysis 3*, 253–64.

Shumway, R. H. and D. S. Stoffer (2006). *Time Series Analysis and Its Applications: With R Examples* (2 ed.). New York: Springer-Verlag.

Stock, J. H. and M. Watson (1988). Testing for common trends. *J. American Statistical Association 83*, 1097–107.

Stock, J. H. and M. Watson (2002). Macroeconomic forecasting using diffusion indexes. *J. Business and Economic Statist. 20*, 147–162.

Stock, J. H. and M. Watson (2011). Dynamic factor models. See Clements and Hendry (2011).

Trimbur, T. (2006). Properties of higher order stochastic cycles. *J. Time Series Analysis 27*, 1–17.

Tunnicliffe-Wilson, G. (1989). On the use of marginal likelihood in time series model estimation. *J. Royal Statistical Society B 51*, 15–27.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression. *J. Econometrics 23*, 385–400.