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

A preliminary version, please do not quote

1 Introduction

The forecasting of seasonal 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 longer term developments of the time series of interest and is often specified as a smooth function of time. The recurring but persistently changing patterns within the years are 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 will show 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 contribution. 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 seasonal economic time series on the basis of a 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 decomposition approach we do not require the differencing of the time series to a stationary process. The non-stationary properties of a time series are explicitly formulated by a selection of the components in the decomposition. The Box-Jenkins approach requires that the series has been differenced to achieve stationarity. 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 autoregressive integrated moving average processes.

This chapter is organised as follows. Section 2 provides a comprehensive review of decomposition models. Section 3 discusses the methodology of state space analysis. We introduce

the state space model, give illustrations of how decomposition models can be represented in state space, present the Kalman filter, discuss maximum likelihood estimation of parameters and present some diagnostic checking statistics. In Section 4 we discuss how forecasts can be generated as part of the state space analysis and how observation weights of the forecast function are computed. Multivariate extensions of the decomposition model are discussed in Section 5. Section 6 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 and ε_t represent trend, seasonal and irregular components respectively. The trend and seasonal components are modelled by linear dynamic stochastic processes which depend on disturbances. The components are formulated in 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 the term "Structural Time Series Model" is used in this context. The effectiveness of structural time series models compared to ARIMA type models is discussed in Harvey, Koopman, and Penzer (1998). In particular, they stress that time series models based on unobserved components are effective when messy features are present such as missing values, mixed frequencies (monthly and quarterly frequencies of time series), outliers, structural breaks and nonlinear non-Gaussian aspects.

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

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

with autoregressive coefficient $0 < \varphi_\beta < 1$ and where the disturbance series η_t and ζ_t are as in (3). The restriction for φ_β is necessary to have a stationary process for the drift β_t. In this case, the initial variable μ_1 is treated as an unknown coefficient while the initial drift is specified as $\beta_1 \sim \text{N}(0, \sigma_\zeta^2 \,/\, 1 - \varphi_\beta^2)$. 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 $\sigma_\eta^2 = 0$ 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 $s = s(t)$ with $s = 1, \ldots, S$ where S is the seasonal length ($S = 4$ for quarterly data and $S = 12$ for monthly data). The time-varying seasonal can be established in different ways.

Fixed dummy seasonal: In case the seasonal pattern is fixed over time, we have 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

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

When we have 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 required to present multicollinearity. The constraint

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

ensures the seasonal effects sum to zero. In effect, we have $S - 1$ unknown seasonal coefficients that need to be estimated.

Time-varying dummy seasonal: 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

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

where the disturbance series ω_t is serially independent and mutually independent of all other disturbance series, for $t = S-1, \ldots, n$ and $j = 2, \ldots, S-1$. The initial variables $\gamma_{S-1}, \ldots, \gamma_1$ are treated as unknown coefficients. When the disturbance variance $\sigma_\omega^2 = 0$, we return to the case of fixed seasonal effects. When the variance is 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 $\delta_{s,j}$ drops out in (11) for $j = S/2$ 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, $\gamma^+_{j,t}$ drops out for $j = S/2$).

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 $\sigma_\omega^2 = 0$, equation (13) reduces to (12). The variance σ_ω^2 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 $j = 1, \ldots, \lfloor S/2 \rfloor$.

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. [SJ: expand on this ..., see work of Tommaso]

2.3 Cycle component

To capture the business cyclefeatures of a time series, we need to include a stationary cycle component in the model. Various stochastic specifications of the cycle component can be considered.

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

$$\varphi_\psi(L)\psi_{t+1} = \vartheta_\psi(L)\xi_t, \qquad \xi_t \sim \text{NID}(0, \sigma_\xi^2), \tag{14}$$

where $\varphi_\psi(L)$ is the autoregressive polynomial in the lag operator L of order p with coefficients $\varphi_{\psi,1}, \ldots, \varphi_{\psi,p}$ and $\vartheta_\psi(L)$ is the moving average polynomial of 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 with increasing lags but the convergence pattern is cyclical. It implies that the time series itself has cyclical dynamic properties. Once the autoregressive coefficients are estimated, it can be established whether the empirical model with ψ_t as in (14) has detected cyclical dynamics in the time series. The economic cycle decomposition model of Clark (1987) includes a stationary ARMA component with order $p = 2$ and $q = 0$.

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

$$\begin{pmatrix} \psi_{t+1} \\ \psi_{t+1}^+ \end{pmatrix} = \varphi_\psi \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{pmatrix} \psi_t \\ \psi_t^+ \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^+ \end{pmatrix}, \tag{15}$$

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 cyclical disturbance series κ_t and κ_t^+ are serially independent and mutually independent, also with respect to all other disturbance series. The coefficients φ_ψ, λ_c and σ_κ^2 are unknown and need to be estimated together with the other parameters.

The stochastic cycle specification is discussed by Harvey (1989, §2.3-2.5), where it is argued that the process (15) is the same as the ARMA process (14) with $p = 2$ and $q = 1$ when the roots are complex.

Smooth time-varying trigonometric cycle: To enforce smoothness upon a 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) proposes the specification $\psi_t = \psi_t^{(m)}$ where

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

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 (2002).

Multiple cycles The dynamic specification of a cycle may be more intricate than the one given above. When a satisfactory description of the cycle dynamics is not found by a single component, a set of multiple cycle components can be considered, such as:

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

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, if the time-varying trigonometric cycle is adopted for each $\psi_{j,t}$, a different cycle frequency λ_c should be associated with $\psi_{j,t}$. In this way we may 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 which can affect

the dynamic behaviour in the series. In other cases it is evident that a set of explanatory variables need to be included in the model in order to capture specific (dynamic) variations in the time series. The explanatory variables can also be used to allow for outliers and breaks in the time series. Therefore, we extend the decomposition with a multiple regression effect,

$$y_t = \mu_t + \gamma_t + \psi_t + x_t'\delta + \varepsilon_t, \qquad e_t \sim \text{NID}(0, \sigma_\varepsilon^2), \tag{17}$$

for $t = 1, \ldots, n$, and where x_t is a K 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.

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), so that:

$$y_t = Z_t\alpha_t + G_t\epsilon_t, \qquad \alpha_{t+1} = T_t\alpha_t + H_t\epsilon_t, \qquad \epsilon_t \sim \text{NID}\left(0, I\right), \tag{18}$$

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 (18) 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 (18) and it is used to formulate the dynamic processes of the unobserved components in a companion form. The (possibly time-varying) deterministic matrices T_t, Z_t, H_t and G_t are referred to as system matrices and they will be often sparse selection matrices.

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 (18), we discuss a number of illustrations.

Basic decomposition 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 (18) 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 variances of the disturbances are unknown and need to be estimated. They are σ_η^2, σ_ζ^2, σ_ω^2 and σ_ε^2. It is common practive to transform the variance into logs for the purpose of estimation; we estimate log-variances which 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 §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; see §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 (14) 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 §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 (16) with $m = 2$, and irregular ε_t as in (1). The state and disturbance vectorsare 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 variables μ_1 is treated as an unknown coefficient. The initial conditions for the smooth cycle process can be obtained from Trimbur (2002). 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 (18). The predictive estimator of the state vector α_{t+1} is based on 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 and with $P_{t+1|t}$ as its state prediction error variance matrix. The Kalman filter is given by

$$\begin{aligned} v_t &= y_t - Z_t a_{t|t-1}, & F_t &= Z_t P_{t|t-1} Z_t' + G_t G_t', \\ & & M_t &= T_t P_{t|t-1} Z_t' + H_t G_t', & t &= 1, \ldots, n, \\ a_{t+1|t} &= T_t a_{t|t-1} + K_t v_t, & P_{t+1|t} &= T_t P_{t|t-1} T_t' + H_t H_t' - K_t M_t', \end{aligned} \tag{19}$$

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

$$a_{t|t} = a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} v_t, \qquad P_{t|t} = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}, \tag{20}$$

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 sympols 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, Section 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

$$\begin{aligned} \ell = \log p\left(y_1, \ldots, y_n\right) &= \sum_{t=1}^{n} \log p\left(y_t|y_1, \ldots, y_{t-1}\right) \\ &= -\frac{n}{2}\log\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{n} \log|F_t| - \frac{1}{2}\sum_{t=1}^{n} v_t' F_t^{-1} v_t. \end{aligned} \tag{21}$$

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 it explicit that the likelihood is a function of the parameter vector, we write $\ell = \ell(\theta)$.

3.4 Parameter estimation

In a state space analysis we are faced 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 §3.1 for an illustration. The second group are the initial variables for the unobserved (non-stationary) processes and the regression coefficients such as δ in (17). The initial conditions for unobserved stationary processes can be derived from the theoretical autocorrelation function.

Maximum likelihood estimation of θ: The log-likelihood function (21) 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 components of the joint density conditional on $y_1, \ldots, y_n$ and the Maximization (M) step maximizes the resulting expression with respect to θ. The E step mainly consists of evaluating the estimated state vector using a smoothing algorithm related to the Kalman filter. The M step is usually done 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 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 §3.1 it is shown that these initial and regression coefficients are collectively placed in the initial state vector α_1. The estimation of these coefficients is therefore subsumed in the estimation of the initial state vector α_1.

A straightforward approach to the estimation of α_1 is to estimate it 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 parameter vector of a high dimension. 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 (19) 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 that the maximum likelihood estimation of α_1 can affect the estimation of unknown variances in θ; for example, it increases 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. 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 and regression coefficients; see Ansley and Kohn (1985), de Jong (1991) and Koopman (1997).

Stationary conditions of initial state: When the state vector only represents 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

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

Since this solution also applies to α_1, we can initialize the Kalman filter (19) with $a_{1|0} = 0$ and $P_{1|0} = P^*$.

In most models, the initial state vector α_1 contains initial and regression coefficients as well as stationary variables; see also the illustrations in §3.1. The Kalman filter initialization treatments 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 and serially and mutually independent with constant variances. Under these assumptions the standardised one-step ahead prediction errors (or *prediction residuals*) are given by

$$e_t = \frac{v_t}{\sqrt{F_t}}, \qquad t = 1, \ldots, n, \tag{22}$$

are also normally distributed and serially independent with unit variance. We can check that 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:

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

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 may have some data points missing in time. In our context, it also offers a solution to the forecasting challenge since we can regard the future observations as a set of missing observations. As a consequence, the Kalman also delivers the 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 forecasts as filling in missing values, and then describe observation weights.

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

$$a_{t+1|t} = T_t a_{t|t-1}, \qquad P_{t+1|t} = T_t P_{t|t-1} T_t' + H_t H_t', \tag{23}$$

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 (19); if y_t is missing, carry out the prediction step (23). 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 (18), the MMSLE of y_{n+h} is a linear function of the observations $y_1, \ldots, y_n$, that is

$$\mathrm{E}(y_{n+h}|y_1, \ldots, y_n) = \hat{y}_{n+h|n} = \sum_{j=1}^{n} w_j(\hat{y}_{n+h|n})y_j, \tag{24}$$

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

$$a_{t|t-1} = \sum_{j=1}^{t-1} w_j(a_{t|t-1})\ y_j. \tag{25}$$

The Kalman filter (19) provides the means to compute these weights by storing the Kalman gain matrices. Then the following backward recursion is implemented,

$$w_j(a_{t|t-1}) = B_{t,j}K_j, \qquad B_{t,j-1} = B_{t,j}T_j - w_j(a_{t|t-1})Z_j, \qquad j = t-1, t-2, \ldots, 1, \tag{26}$$

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

$$w_j(\hat{y}_{t|t-1}) = b_{t,j}K_j, \qquad b_{t,j-1} = b_{t,j}T_j - w_j(\hat{y}_{t|t-1})Z_j, \qquad j = t-1, t-2, \ldots, 1, \tag{27}$$

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 (26) at time $j = \tau$ reduces to

$$w_\tau(a_{t|t-1}) = 0, \qquad B_{t,\tau-1} = B_{t,\tau}T_\tau. \tag{28}$$

Hence missing value can be accommodated in algorithms for computing observation weights in a state space analysis. Forecasting is treating future observations as missing. The forecasting

observation weights can therefore be obtained from (28). Using the initialisation of (26) to time $t = n + h$, that is $B_{n+h,n+h-1} = I$, and applying (28) 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 $w_j(a_{\tau|n})$ can be obtained by (26) when y_j is observed and by (28) 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 (27).

4.3 Autoregressive representation of model

The weights also lead to the autoregressive representation of any decomposition 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 (25) and (26) that we can express y_t as

$$y_t = \sum_{i=0}^{\infty} \phi_i y_{t-i} + v_t, \qquad \phi_i = b_{t,t-i} K_{t-i},$$

with $n \to \infty$. This specification is effectively an autoregressive model representation of y_t that is originally formulated as a decomposition model. The infinite lag order is in practice not necessary as $b_{t,t-i} \to 0$ as $i \to \infty$. We therefore conclude that $\phi(L) y_t = v_t$ with an effectively finite lag polynomial $\phi(L)$ is a valid representation of y_t.

5 Multivariate components

In Section 2 we set out a comprehensive class of unobserved components time series models. In economic theory one focuses on the dynamic relationships between variables.

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, a multivariate local level model can be applied to the p time series simultaneously:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \Sigma_\varepsilon), \\ \mu_{t+1} &= \mu_t + \xi_t, \qquad \xi_t \sim \text{NID}(0, \Sigma_\xi), \end{aligned} \tag{29}$$

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* (29), 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. If it is found that the rank r of Σ_ξ in (29) is smaller than p, then this indicates that the p series have r *common levels*. Such common factors may not only have a nice interpretation, but may also result in more efficient inferences and forecasts.

6 Conclusion

In this paper we have discussed the state space analysis of decomposition models for univariate time series. This class of linear time series models are composed from a set of unobserved components which represent time-varying trend, seasonal, cycle and regression effects. Decomposition models require the explicit formulation of stochastically time-varying unobserved components; each component captures a specific dynamic feature of the time series. Although these models can be used for different purposes, their main application is forecasting. The state space analysis provides practical tools to build a model for forecasting. The measure of success for a model depends on the task for which the model is used. When its purpose is forecasting, it is important that the model provides an effective description of all dynamic features in the data. For this purpose, diagnostic statistics based on one-step ahead prediction errors should be satisfactory. When the final model (with unknown parameters estimated by maximum likelihood) is found to be satisfactory, it can be used for forecasting. The empirical model ideally implies a forecast function that can be justified by common sense. For this purpose, particular attention is given to the construction of observation weights used in the forecast function.

References

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

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.

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.

de Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics 19*, 1073–83.

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

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.

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

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.

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

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

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika 80*, 117–26.

Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for non-stationary time series models. *J. American Statistical Association 92*, 1630–8.

Koopman, S. J. and A. C. Harvey (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control 27*, 1317–1333.

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

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

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

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.

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.

Trimbur, T. (2002). Properties of a general class of stochastic cycles. Mimeo, University of Cambridge.

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.