

Analysis of Historical Time Series with Messy Features: The Case of Commodity Prices in Babylonia*

Siem Jan Koopman and Lennart Hoogerheide

VU University Amsterdam, The Netherlands

Tinbergen Institute, The Netherlands

This version: May 2013

Abstract

Time series data originates from the past; it is historical by construction. In most subjects of the social sciences, it is common to analyse time series data from the more recent past since it is most relevant to the present time. In historical research, one is particularly interested in data from the more remote past which can stretch out over hundreds of years. The statistical analysis of time series data should not be different when the data come from ancient times. However, the accuracy in recording the data will be lower. We can also expect that historical data may not be completely available for each year or each decennium. We show that statistical methodologies based on state space time series models can treat the typical messy features in historical data in an effective manner. An illustration is given for a multiple time series of commodity prices in the economy of Babylonia for the period from 385 to 61 BC. Although many stretches of observations can be observed at a daily frequency, most observations in this period need to be treated as missing. Our main interests center on the question whether commodity markets during these years have operated efficiently in the economy of Babylonia.

*We thank the participants of the KNAW Colloquium “The efficiency of Markets in Pre-industrial societies: the case of Babylonia (c. 400-60 BC) in comparative perspective”, 19-21 May 2011 in Amsterdam. The discussions have been very insightful.

1 Introduction

In this paper we discuss the multivariate analysis of historical time series that typically are subject to messy features such as missing observations and outlying observations. Our primary motivation is the multiple analysis of monthly commodity prices in Babylonia between 385 – 61 BC. We consider monthly price series for Barley, Dates, Mustard, Cress, Sesame and Wool. As it can be expected, the data set is far from complete. The monthly time series span over 300 years and hence we could have 3,888 observations for each price series. However for most series we only have around 530 monthly observations available. It means that the vast majority of data entries is missing. We need to treat 3,358 missing observations in the analysis. The available prices stretches over a long period of more than three centuries. The treatments of commodities and the market conditions have changed greatly in Babylonia during this long and hectic period. The time series of prices are subject to outliers and structural breaks due to periods of war and other disasters. In times of turmoil, the supply of commodities typically reduce and their availability becomes scarce. As a result, prices typically rise and often to very high levels. Other contributions of this Volume will discuss further particularities of this data set and related sets. Our contribution concentrates on the statistical treatment of these historical data sets.

Different statistical methodologies for time series analysis can be pursued. It is not our purpose to provide an overview of different time series methodologies. We can refer to reviews such as those of Ord (1990). In our contribution we discuss the class of Unobserved Components Time Series (UCTS) models for both univariate and multivariate analyses. A complete treatment is presented in Harvey (1989) who refers to such models as “Structural Time Series Models”. An up-to-date discussion of univariate UCTS models is presented in Section 2. The statistical analysis based on UCTS models relies mainly on the representation of the model in state space form. Once the model is framed in its state space form, the Kalman filter and related methods are used for the estimation of the dynamic features of the model but also for the computation of the likelihood function. We will introduce the state space form and the related methods in Section 3. We will argue that in particular the Kalman filter plays a central role in time series analysis as it is general and can handle missing entries in a time series as a routine matter. The generality of the UCTS model and the Kalman filter is further illustrated in Section 4 where we show how the univariate UCTS model and its treatment can be generalized towards a multivariate statistical analysis of a multiple time series.

The UCTS methodology is illustrated for our six monthly time series of commodity prices for Barley, Dates, Mustard, Cress, Sesame and Wool. We analyse first each time series first by an univariate UCTS model. It will be shown that we can obtain accurate estimates of the evolution of the price levels over a time span of 600 years. We discuss how outliers and structural breaks affect the analysis and how we can allow for these irregularities in the time series. A complete multivariate analysis is also considered. In particular we investigate how common the price evolutions have been for the different commodities.

The remainder of this paper is organised as follows. In Section 2 we discuss in detail our time series methodology based on UCTS models. The general state space methods are briefly discussed in Section 3. We introduce a number of interesting multivariate extensions for the UCTS methodology in Section 4. Finally, the empirical study for the six monthly time series of commodity prices for Barley, Dates, Mustard, Cress, Sesame and Wool in Babylonia between 385 – 61 BC is presented 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

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

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 The trend component can simply be modelled as a random walk process and is then given by

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

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. 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 An extension of the random walk trend is obtained by including a stochastic drift component

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

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 $\mu_{t+1} = \mu_t + \beta_1 + \eta_t$ 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 $\sigma_\xi^2 > 0$, the trend μ_t in (3) is known as the integrated random walk process which can be visualised as a smooth trend function.

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 $s = 1, \dots, S$ 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 trigonometric seasonal: A deterministic seasonal pattern can 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

$$\gamma_t = \sum_{j=1}^{\lfloor S/2 \rfloor} \gamma_{j,t}, \quad \gamma_{j,t} = a_j \cos(\lambda_j t - b_j), \quad (4)$$

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 $j = 1, \dots, \lfloor S/2 \rfloor$ and $t = 1, \dots, n$. The seasonal effects are based on coefficients a_j and b_j . Given the trigonometric identities

$$\cos(\lambda \pm \xi) = \cos \lambda \cos \xi \mp \sin \lambda \sin \xi, \quad \sin(\lambda \pm \xi) = \cos \lambda \sin \xi \pm \sin \lambda \cos \xi, \quad (5)$$

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

$$\gamma_{j,t} = \delta_{c,j} \cos(\lambda_j t) + \delta_{s,j} \sin(\lambda_j t), \quad (6)$$

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 (6) 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 (5), we can express $\gamma_{j,t}$ as the recursive expression

$$\begin{pmatrix} \gamma_{j,t+1} \\ \gamma_{j,t+1}^+ \end{pmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^+ \end{pmatrix}, \quad (7)$$

with $\gamma_{j,0} = \delta_{c,j}$ and $\gamma_{j,0}^+ = \delta_{s,j}$ for $j = 1, \dots, \lfloor S/2 \rfloor$. 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 $j = 1, \dots, \lfloor S/2 \rfloor$ (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 (7) allows the introduction of a time-varying trigonometric seasonal function. We obtain the stochas-

tic trigonometric seasonal component γ_t by having

$$\begin{pmatrix} \gamma_{j,t+1} \\ \gamma_{j,t+1}^+ \end{pmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^+ \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^+ \end{pmatrix}, \quad \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^+ \end{pmatrix} \sim \text{NID}(0, \sigma_\omega^2 I_2), \quad (8)$$

with $\lambda_j = 2\pi j/S$ for $j = 1, \dots, \lfloor S/2 \rfloor$ and $t = 1, \dots, n$. 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 (8) reduces to (7). 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 $j = 1, \dots, \lfloor S/2 \rfloor$.

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

$$\gamma_t = e_j' \gamma_t^\dagger, \quad \gamma_{t+1}^\dagger = \gamma_t^\dagger + \omega_t^\dagger, \quad \omega_t^\dagger \sim \text{NID}(0, \sigma_\omega^2 \Omega), \quad (9)$$

where the $S \times 1$ vector γ_t^\dagger contains the seasonal effects, e_j is the j th column of the $S \times S$ identity matrix I_S , $S \times 1$ disturbance vector ω_t^\dagger is normally and independently distributed with mean zero and $S \times S$ variance matrix $\sigma_\omega^2 \Omega$. 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 $j = 1, \dots, S$, 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 γ_t^\dagger 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 (9) 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

$$\varphi_\psi(L) \psi_{t+1} = \vartheta_\psi(L) \xi_t, \quad \xi_t \sim \text{NID}(0, \sigma_\xi^2), \quad (10)$$

where $\varphi_\psi(L)$ is the autoregressive polynomial in the lag operator L , of lag order p with coefficients $\varphi_{\psi,1}, \dots, \varphi_{\psi,p}$ and $\vartheta_\psi(L)$ is the moving average polynomial of lag order q with coefficients $\vartheta_{\psi,1}, \dots, \vartheta_{\psi,q}$. 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 (10) 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 (10) 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 (8) 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

$$\begin{pmatrix} \psi_{t+1} \\ \psi_{t+1}^+ \end{pmatrix} = \varphi_\psi \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{pmatrix} \psi_t \\ \psi_t^+ \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^+ \end{pmatrix}, \quad (11)$$

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

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

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 (11) is the same as the ARMA process (10) with $p = 2$ and $q = 1$ and where the roots of $|\varphi_\psi(L)| = 0$ are enforced to be within the complex range.

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 (2012). 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

$$y_t = Z_t \alpha_t + G_t \epsilon_t, \quad \alpha_{t+1} = T_t \alpha_t + H_t \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, I), \quad (12)$$

for $t = 1, \dots, n$, 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 (12) 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 (12) 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 (12), we consider the basic structural model as given by

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2), \quad (13)$$

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

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

The state space formulation of the basic decomposition model is given by (12) with the system matrices

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

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

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 coefficients. 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.

3.2 Kalman filter

Consider the linear Gaussian state space model (12). The predictive estimator of the state vector α_{t+1} is a linear function of the observations y_1, \dots, 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, \dots, 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

$$\begin{aligned} v_t &= y_t - Z_t a_{t|t-1}, & F_t &= Z_t P_{t|t-1} Z_t' + G_t G_t', \\ a_{t+1|t} &= T_t a_{t|t-1} + K_t v_t, & M_t &= T_t P_{t|t-1} Z_t' + H_t G_t', \\ & & P_{t+1|t} &= T_t P_{t|t-1} T_t' + H_t H_t' - K_t M_t', \end{aligned} \quad t = 1, \dots, n, \quad (14)$$

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 $v_t = y_t - E(y_t | y_1, \dots, y_{t-1})$ with variance $\text{Var}(v_t) = F_t$. The innovations have mean zero and are serially independent by construction so that $E(v_t v_s') = 0$ for $t \neq s$ and $t, s = 1, \dots, n$.

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, \dots, y_t , denoted by $a_{t|t}$, and its corresponding MSE matrix, denoted by $P_{t|t}$, can be computed as

$$a_{t|t} = a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} v_t, \quad P_{t|t} = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}, \quad (15)$$

It then follows that

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

Formal proofs of the Kalman filter can be found in Anderson and Moore (1979), Harvey (1989) and Durbin and Koopman (2012).

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, \dots, y_n by $p(y_1, \dots, y_n)$ and the prediction error decomposition is then given by

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

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

$$\begin{aligned} \ell = \log p(y_1, \dots, y_n) &= \sum_{t=1}^n \log p(y_t | y_1, \dots, y_{t-1}) \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \log |F_t| - \frac{1}{2} \sum_{t=1}^n v_t' F_t^{-1} v_t. \end{aligned} \quad (16)$$

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 (13). The initial conditions for unobserved stationary processes can be derived from the theoretical autocovariance function.

Maximum likelihood estimation of θ : The log-likelihood function (16) 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 $p(y_1, \dots, y_n, \alpha_1, \dots, \alpha_n)$. The Expectation (E) step takes the expectation of the state vectors conditional on y_1, \dots, 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: The non-stationary trend and seasonal components, as discussed in Section 2, rely on initial variables that are treated as fixed unknown coefficients. In the illustrations in Section 3.1 it is shown that these initial states are collectively placed in 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 (14) 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).

Tunncliffe-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 maximization 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; 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 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 $E(\alpha_t) = 0$ and $P = \text{Var}(\alpha_t)$ for $t = 1, \dots, n$. It follows that $P = TPT + HH'$ with solution

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

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 (14) with $a_{1|0} = 0$ and $P_{1|0} = P^*$.

In most models, the initial state vector α_1 contains initial stationary and nonstationary 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

$$e_t = \frac{v_t}{\sqrt{F_t}}, \quad t = 1, \dots, n. \quad (17)$$

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

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

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

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

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 j th 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.

3.6 Missing values

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.

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 $t = 1, \dots, n$. 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 $F_\tau \rightarrow \infty$ as it indicates that we have no information about v_τ . The consequences of having $F_\tau \rightarrow \infty$ in the Kalman filter is that $K_\tau \rightarrow 0$ while the remaining computations in the Kalman filter can still be carried out. The prediction step of the Kalman filter reduces to

$$a_{t+1|t} = T_t a_{t|t-1}, \quad P_{t+1|t} = T_t P_{t|t-1} T_t' + H_t H_t', \quad (18)$$

for $t = \tau$ as $F_\tau \rightarrow \infty$. 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 (14); if y_t is missing, carry out the prediction step (18). Missing entries are allowed throughout the data sample y_1, \dots, 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 $h = 1, 2, \dots$. The observation forecasts $\hat{y}_{n+h|n} = E(y_{n+h}|y_1, \dots, y_n)$ and its error variance matrix $V_{n+h|n} = \text{Var}(y_{n+h} - \hat{y}_{n+h|n}|y_1, \dots, y_n)$ are then computed by

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

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

4 Multivariate extensions

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.

4.1 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

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \text{NID}(0, \Sigma_\varepsilon), \\ \mu_{t+1} &= \mu_t + \xi_t, & \xi_t &\sim \text{NID}(0, \Sigma_\xi), \end{aligned} \quad (19)$$

for $t = 1, \dots, n$, 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* (19), 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 (19) 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 be represented by

$$y_t = \Lambda_\xi \mu_t^\dagger + \Lambda_\varepsilon \varepsilon_t^\dagger, \quad \mu_{t+1}^\dagger = \mu_t^\dagger + \xi_t^\dagger, \quad \varepsilon_t^\dagger \sim \text{NID}(0, \mathcal{D}_\varepsilon), \quad \xi_t^\dagger \sim \text{NID}(0, \mathcal{D}_\xi), \quad (20)$$

where the various terms are defined implicitly by relating the terms in (19) with those in (20) via

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

where \mathcal{D}_ε and \mathcal{D}_ξ 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 (20) 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

$$y_t = \Lambda_\xi^* \mu_t^* + \Lambda_\varepsilon^* \varepsilon_t^*, \quad \mu_{t+1}^* = \mu_t^* + \xi_t^*, \quad \varepsilon_t^* \sim \text{NID}(0, C \mathcal{D}_\varepsilon C'), \quad \xi_t^* \sim \text{NID}(0, B \mathcal{D}_\xi B'), \quad (21)$$

where

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

for $t = 1, \dots, n$. 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 (19), (20) and (21).

4.2 Common trends and cycles

When the variance matrix of the trend disturbance Σ_ξ has not full rank, the multivariate local level model (19) 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 (20), the dimensions of the matrices Λ_ξ and \mathcal{D}_ξ are $p \times r$ and $r \times r$, respectively. Hence, the trend vector μ_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 μ_t^\dagger . The locations of the remaining $p - r$ time series in y_t are then adjusted by the constant

vector $\bar{\mu}$ in

$$y_t = \bar{\mu} + \Lambda_\xi \mu_t^\dagger + \varepsilon_t, \quad \mu_{t+1}^\dagger = \mu_t^\dagger + \xi_t^\dagger, \quad (22)$$

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

$$y_t = \bar{\mu} + \Lambda_\xi \mu_t^\dagger + \Lambda_\kappa \psi_t^\dagger + \varepsilon_t, \quad (23)$$

where μ_t^\dagger is the $r_\mu \times 1$ vector of common trends and vector ψ_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 (10) can be considered, see Shumway and Stoffer (2006, Chapter 5.7). The multivariate version of the cycle process (11) 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 ψ_t^\dagger in (23) by

$$\begin{pmatrix} \psi_{t+1}^\dagger \\ \psi_{t+1}^\dagger \end{pmatrix} = \varphi_\psi \left\{ \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \otimes I_{r_\psi} \right\} \begin{pmatrix} \psi_t^\dagger \\ \psi_t^\dagger \end{pmatrix} + \begin{pmatrix} \kappa_t^\dagger \\ \kappa_t^\dagger \end{pmatrix}, \quad (24)$$

where the auxiliary cycle vector ψ_t^\dagger 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 κ_t^\dagger and κ_t^+ together with the initial conditions for the cycle vectors are given by

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

and the cyclical disturbance series κ_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 (23) that

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

for $t = 1, \dots, n$. The individual cycle processes in ψ_t^\dagger are mutually independent of each other while those in ψ_t are correlated with each other.

In the decomposition model (23) 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

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

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 i th equation of (23), 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 i th element of z_t for $z = y, \varepsilon$, $\bar{\mu}_i$ is the i th element of $\bar{\mu}$ and $\Lambda_{x,i}$ is the i th row of Λ_x for $x = \xi, \kappa$ with $i = 1, \dots, p$. 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).

4.3 State space representation and parameter estimation

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).

5 A time series analysis of commodity prices in Babylonia

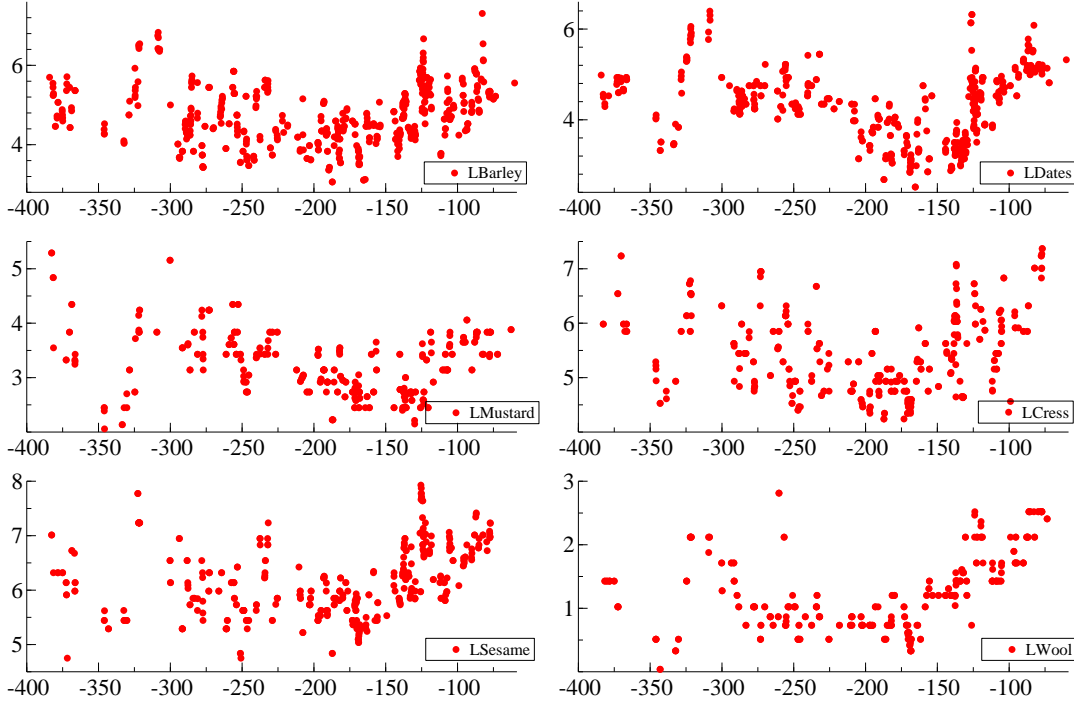
In Figure 1 six monthly time series of commodity prices for Barley, Dates, Mustard, Cress, Sesame and Wool in Babylonia between 385 – 61 BC. The series length is for 3,888 observations but we only have 530 prices available. The vast majority of the data is therefore missing. The data is transformed in logs in the graphs and for all analyses. In the earlier years of our sample, less data is available while in the later years more prices are observed. Some common features in the evolution of the log-prices emerge. For all commodities, the prices are increasing in the years towards 300 BC while in the following years upto 150 BC the prices are slowly falling. In the last century of our sample, say from 150 BC, the prices are increasing for most commodities. All time series are subject to outliers; some of them can be identified with specific historical events.

5.1 Univariate decompositions

In our first set of analyses, we aim to decompose the dynamic features of the time series in level, cycle, seasonal and irregular processes. It provides an insight into the major properties of the monthly time series. The unobserved components time series model that we consider is given by

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t,$$

FIGURE 1: MONTHLY COMMODITY PRICES IN BABYLONIA



Monthly time series of commodity prices (in logs) for Barley, Dates, Mustard, Cress, Sesame and Wool in Babylonia between 385 – 61 BC. The x-axis displays a year BC as a negative number.

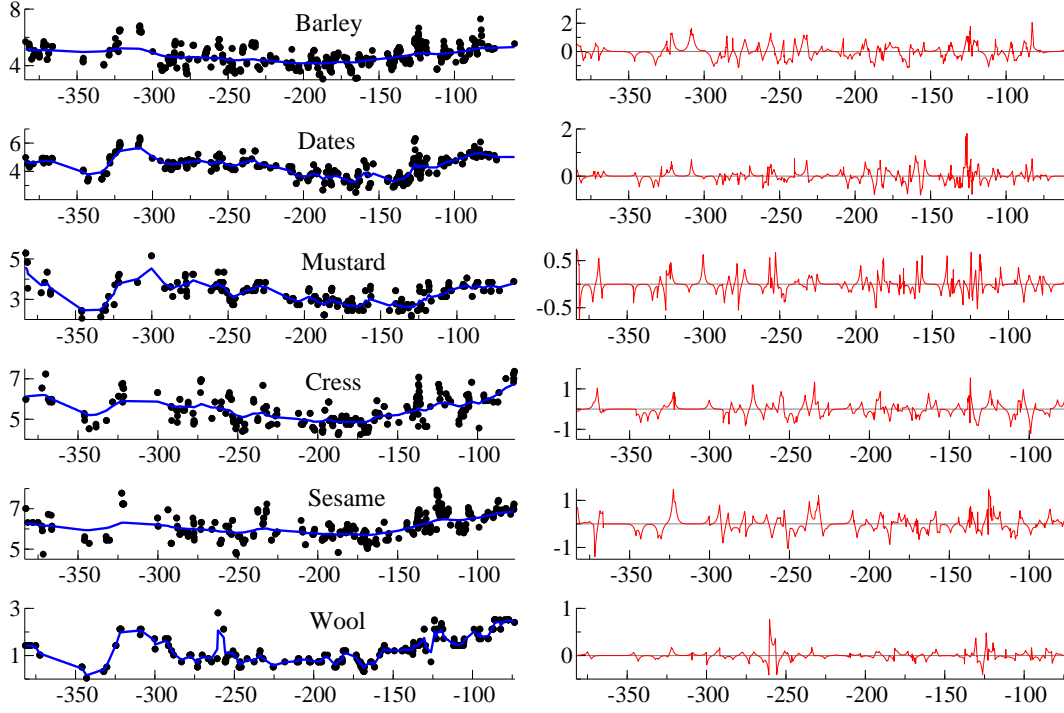
where y_t is an univariate time series of monthly commodity prices and where we treat the level component μ_t as the random walk process (2), the cycle ψ_t as the trigonometric cycle process (11), the seasonal γ_t as the sum of time-varying seasonal trigonometric terms (8) and the irregular ε_t as Gaussian white noise. The model is framed in state space and the Kalman filter plays a central role in the analysis. Hence the treatment of missing values is straightforward and we can carry out the analysis in a standard fashion.

The resulting decomposition for all six time series are displayed in Figure 2. We only present the estimated level (or trend) and the cycle components for each monthly price series. The seasonal component has only been estimated significantly for the Barley prices but it is found that these seasonal effects are not varying over time. We have kept the seasonal component inside the model for Barley but for all other series we have removed the seasonal component γ_t from the model specification. The estimated irregular components are not displayed. In some series the irregular is estimated to have a very small variance and in other series it only has captured some outliers in the series. The estimated cycle component is of key interest and we find that this effect clearly takes out movements in the time series that last for a small number of years. We may want to interpret these dynamic features as economic cyclical effects in commodity prices.

When we take a closer look at the estimated levels and cycles, we can identify some common movements. In particular the evolution of the level effect in the series appear to be common across Dates, Mustard, Cress and Sesame while Barley and Wool appear to be subject to somewhat different dynamics. The estimated cycle component also presents some common features among the different commodities but they are less strong and convincing. The cycle for Wool is certainly somewhat

different. In order to investigate the commonalities in the price series more accurately, we consider a multivariate analysis next for Barley, Dates, Mustard, Cress and Sesame. We exclude Wool in the multivariate analysis.

FIGURE 2: UNIVARIATE DECOMPOSITIONS OF COMMODITY PRICES IN BABYLONIA



The lefthandside graphs present the estimated level (or trend) components for the monthly time series of commodity prices (in logs) for Barley, Dates, Mustard, Cress, Sesame and Wool in Babylonia between 385 – 61 BC. The righthandside graphs present the estimated cycle components. The x-axis displays a year BC as a negative number.

5.2 Multivariate decomposition

In the multivariate analysis, we simultaneously decompose the dynamic features in the time series of Barley, Dates, Mustard, Cress and Sesame. We include the components for level, cycle and irregular while for Barley we include fixed seasonal dummies to capture the significant seasonal effects in Barley. In this joint analysis of the series we may provide an insight into the common properties of the monthly time series. The multivariate unobserved components time series model that we consider is then given by

$$y_t = \mu_t + \psi_t + \varepsilon_t,$$

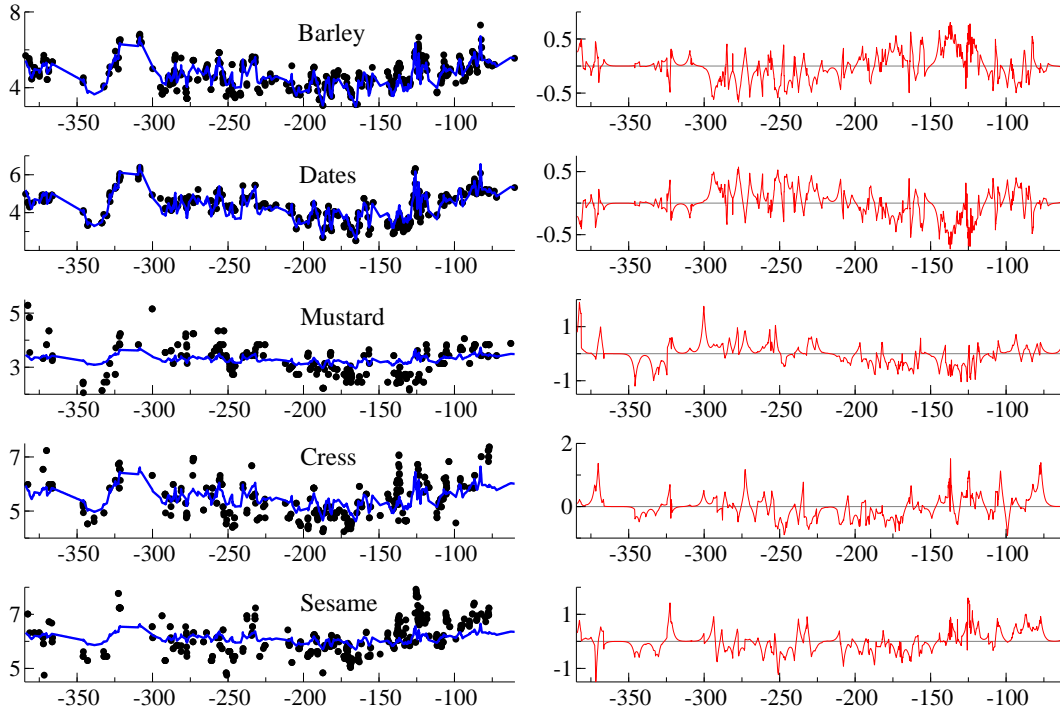
where y_t is the 5×1 vector of monthly commodity prices for Barley, Dates, Mustard, Cress and Sesame, while we treat the level component μ_t as the multivariate random walk process in (19), the cycle vector ψ_t as the similar cycle process (24) and the irregular ε_t as the vector of Gaussian white noise processes. The model is framed in state space and the Kalman filter is adapted to this multivariate setting as discussed in Section 4.3. The treatment of missing values remains straightforward and our analysis remains to rely on standard state space methods.

After a first analysis based on our multivariate decomposition, it has become very clear that

the estimated variance matrix of the level component μ_t , that is Σ_η , has a rank close to one. The four smallest eigenvalues of this variance matrix are very close to zero in comparison to the largest eigenvalue. Hence we have changed the decomposition model towards a model with only a single common level in its specification; see the discussion in Section 4.2. The results of the analysis based on this model are discussed below.

The common features in the level component appear most strongly for Barley, Dates and Cress although the common level captures the main movement in all commodity prices. The cycle components do not appear to share many commonalities. In other words, the deviations from the time-varying level component are mostly adjusting differently in the shorter term. However, a strong and persistent negative correlation appears in the estimated cycle components for the price series of Barley and Dates. This is an interesting finding as it suggests that in Babylonia Barley and Dates have been food substitutes in a much stronger way than they are for Mustard, Cress and Sesame.

FIGURE 3: MULTIVARIATE DECOMPOSITIONS OF COMMODITY PRICES IN BABYLONIA



The lefthandside graphs present the estimated elements of the vector level component for the monthly time series of commodity prices (in logs) for Barley, Dates, Mustard, Cress and Sesame in Babylonia between 385 – 61 BC. The righthandside graphs present the estimated elements of the vector cycle component. The x-axis displays a year BC as a negative number.

6 Conclusion

We have reviewed time series analyses based on unobserved components time series models, both in univariate and multivariate versions. The methodology is general and enables the handling of messy features in time series. An empirical analysis is presented for a set of historical time series from Babylonia. The six time series consist of monthly commodity prices, in logs. In the vast majority of

months, the prices have not been available. The Kalman filter can handle such missing observations as part of the analysis when it needs to treat both univariate and multivariate models. We may conclude that a time series analysis based on unobserved components is effective in identifying the main dynamic features in the commodity prices for Barley, Dates, Mustard, Cress and Sesame in Babylonia between 385 – 61 BC.

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. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook on Economic Forecasting*. Oxford: Oxford University Press.
- 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.
- 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.
- 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.
- Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford: Oxford University Press.
- 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.
- 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.

- 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.
- 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 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.
- Nocedal, J. and S. J. Wright (1999). *Numerical Optimization*. New York: Springer Verlag.
- Ord, J. K. (1990). *Time Series*. Sevenoaks: Edward Arnold.
- 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.
- 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.
- 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.