Structural Time Series Models and the Kalman Filter: a concise review[*]

João Tovar Jalles[†]

Faculty of Economics and Politics, University of Cambridge, UK

19th of June, 2009

Abstract

The continued increase in availability of economic data in recent years and, more importantly, the possibility to construct larger frequency time series, have fostered the use (and development) of statistical and econometric techniques to treat them more accurately. This paper presents an exposition of structural time series models by which a time series can be decomposed as the sum of a trend, seasonal and irregular components. In addition to a detailed analysis of univariate specifications we also address the SUTSE multivariate case and the issue of cointegration. Finally, the recursive estimation and smoothing by means of the Kalman filter algorithm is described taking into account its different stages, from initialisation to parameter's estimation.

Keywords: SUTSE, cointegration, ARIMA, smoothing, likelihood

JEL Classification Numbers: C10, C22, C32

---

1 Introduction

The continued increase in availability of economic data in recent years and, more importantly, the possibility to construct larger frequency time series, have fostered the use (and development) of statistical and econometric techniques to treat them more accurately. In fact, given these new series it is becoming harder and harder to keep the assumption of a fixed pattern or behaviour through time - indeed, a statistical property of a large proportion of socio-economic series is its evolutive character - and, in this context, structural time series models constitute themselves as the appropriate technique, as they allow that each of the typical unobserved components within those series to possess a stochastic nature. In other words, the different components describing the evolution of a time series - trend, seasonality, cycle and irregular - have been traditionally modelled in a deterministic way; however, when these series are sufficiently large, one is able to assess the volatility of such components and, moreover, it may be reasonable to consider that they evolve randomly over time. This is the starting assumption of structural models.

Consider, for instance, a time series which can be decomposed as the sum of a trend, seasonal and an irregular component, then its behaviour could be captured by formulating a regression model whose explanatory variables were a deterministic trend and a set of seasonal qualitative variables. If these components are not stable then this formulation would be inadequate and it would be necessary for the regression coefficients to change over time. This flexibility is possible with structural models, such that "they are not more than regression models in which explanatory variables are a function of time and the parameters change with time" (Harvey, 1989).

In the next section we will introduce the basic formulation of structural time series models. Section 3 discusses the differences between Structural Time Series Models and ARIMA-type models. Finally, Section 4 presents a general overview of the Kalman filter algorithm.

2 Structural Time Series Models

2.1 Univariate time series models

A univariate time series model can be defined as follows:

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t{}^1 \tag{1}$$

---

[1]Even though a multiplicative model could be more appropriate in some cases, this could be treated as an additive one, just by taking a logarithmic transformation.

where μ_t represents the trend, ψ_t the cycle, γ_t the seasonal component and ε_t the irregular, which reflects the random movements. One peculiarity of a structural model is its flexibility in recognizing changes in the behaviour of a given series, by taking its different components as stochastic processes governed by random disturbances. As we will show from the different specifications described in this section, a structural model may not be necessarily defined in terms of all its unobserved components[2]. The following will describe some of the possible different formulations for each component or combinations among them.

The simplest formulation to the irregular component is to consider a white noise process, i.e., a sequence of serially uncorrelated random variables with constant (zero) mean and constant (σ_ε^2) variance. One can also consider more general models in which the stochastic process ε_t is a stationary autoregressive process of order p.

As far as the trend is concerned - i.e., the component that indicates the direction to which the series moves over the long run[3] - it is not specified as a deterministic function of time; on the contrary, one assumes that the trend component is a stochastic process whose value at a given point in time (stochastic level) results from adding to the value of the previous component a certain random amount (stochastic slope) as well as a random disturbance term. Certainly, it is possible that either the level or the slope do not possess a stochastic nature and, most importantly, the slope term may or not be present in the model.

The most simple example is a series whose observations move around an average level which is itself constant over time. If these movements are stationary - in the sense that although some values move away from the average level in the short run, they will always tend to that level in the long run - and if one assumes additionally that there are no correlation among them, then a formulation able to capture the such dynamic behaviour is $y_t = \mu_0 + \varepsilon_t$, for all $t = 1, ..., T$, where ε_t is a Gaussian white noise process with variance σ_ε^2. In this case we have a model in which the level component, μ_0, is deterministic.

The above formulation can be relaxed in order to allow the series' level to change through time, leading to the "local level model", in which the level at each moment in time is the sum of the previous period value and a random element. This model can be defined as:

$$y_t = \mu_t + \varepsilon_t \tag{2}$$

with $\mu_t = \mu_{t-1} + \eta_t$, such that $\eta_t \sim NID(0, \sigma_\eta^2)$, there is a random disturbance term around an underlying level which moves without a particular direction. This is also know

---

[2]In fact, the components to be considered depend, inter alia, on the size and time span of the sample under scrutiny.

[3]According to Harvey (2000) the trend is that part of the series which, when extrapolated, it allows for a clearer indication of the future long-run movement of the series. This definition is consistent with the idea of indicating a general direction.

as random walk with noise. If the variance of the level is zero, we obtain the deterministic level model as specified before. If the variance of the irregular component is zero but the variance of the level is different from zero, the series only possess a level component whose generating process follows a random walk[4].

If we add a slope to the level component as described above, then the average increase per unit of time is constant; alternatively, one can also refer to more flexible formulations such as the local linear trend model:

$$y_t = \mu_t + \varepsilon_t \tag{3}$$

with $\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$ and $\beta_t = \beta_{t-1} + \zeta_t$, where $\zeta_t \sim NID(0, \sigma_\zeta^2)$[5] The disturbance term ζ_t gives to the slope a random character, while given the level and slope of the previous period, the stochastic nature of the level at moment t derives from the presence of the disturbance term η_t.[6] Hence, if the variance of ζ_t is zero, such that the slope is constant, then the model becomes a "local level with drift" type, which takes a stochastic level together with a fixed slope, as follows:

$$\mu_t = \mu_{t-1} + \beta + \eta_t{}^7 \tag{4}$$

If the slope is not only constant but zero, the previous model collapses into the local level version, previously discussed. Finally, it is possible to keep the stochastic nature of the slope and, simultaneously, assume that, given the values of the level and slope at $t-1$, the value of the level at t is non-stochastic:

$$\mu_t = \mu_{t-1} + \beta_{t-1} \tag{5}$$
$$\beta_t = \beta_{t-1} + \zeta_t{}^8 \tag{6}$$

Moreover, the local level trend model can be modified by introducing a damping factor in the slope component of the trend, as follows:

---

[4]The inclusion of a random walk component means that the process is not stationary and, furthermore, it is equivalent to a ARIMA(0,1,1) process, where the MA parameter covers only half of its usual space, i.e., $-1 < \theta < 0$; if $\sigma_\eta^2 = 0$, then we obtain an ARIMA process in which the MA parameter takes the value -1.

[5]In these type of models, the local linear trend may be generalized to higher order polynomials and, though those trends are no longer linear in the sense that give rise to quadratic or cubic forecasting functions, they can still be formulated as time series linear models (Harvey, 1989 for the case of a quadratic linear trend model). The local linear trend model is equivalent to an ARIMA (0,2,2); moreover, it will be a non-invertible process when $\sigma_\eta^2 = 0$ and, contrarily, if $\sigma_\eta^2 > 0$ the process is invertible no matter the value of σ_η^2.

[6]Furthermore, one assumes that these random disturbance terms are mutually and serially uncorrelated.

[7]If $\sigma_\eta^2 = 0$ then we obtain a linear deterministic trend.

$$\beta_t = \rho\beta_{t-1} + \zeta_t \tag{7}$$

where $0 < \rho < 1$. We obtain therefore the damped local linear trend model.[9]

Now, at the time of choosing from and electing one of these specifications, it is important, a priori, to examine the behaviour of the series at hands and, a posteriori, check which has a better adjustment (although, you cannot ignore the implications of each specification and objectives one pursues with the chosen model).

In many economic time series it is important to distinguish between a long run trend and the cyclical or short-run movements, which admits as well a deterministic or stochastic specification. A deterministic cycle describes a perfectly periodic behaviour with a given frequency, which can be described as a function of sines and cosines of the following type:

$$\Psi_t = \alpha\cos(\lambda t) + \beta\sin(\lambda t) \tag{8}$$

where λ is the angle frequency of the cycle, in radians.[10] If we assume that the parameters α and β evolve randomly through time, then the cycle has a stochastic nature. But before incorporating a stochastic character, it is convenient to express the cyclical component in a recursive way such as:

$$\begin{pmatrix} \Psi_t \\ \Psi_t^* \end{pmatrix} = \begin{pmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{pmatrix} \begin{pmatrix} \Psi_{t-1} \\ \Psi_{t-1}^* \end{pmatrix}, t = 1, ..., T \tag{9}$$

with $\Psi_0 = \alpha$ and $\Psi_0^* = \beta$ (Harvey, 1993). The new parameters are Ψ_{t-1}, the value of the cycle at $t-1$, and Ψ_{t-1}^*, which appears by construction. Therefore, the cycle is stochastic when we introduce Gaussian white noise disturbance terms k_t and k_t^* such that:

$$\begin{pmatrix} \Psi_t \\ \Psi_t^* \end{pmatrix} = \begin{pmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{pmatrix} \begin{pmatrix} \Psi_{t-1} \\ \Psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} k_t \\ k_t^* \end{pmatrix} \tag{10}$$

Due to identification reasons and parsimony, we assume that k_t and k_t^* are mutually uncorrelated and possess the same variance σ_k^2. On the other hand, in order to recognize the pseudocyclical behaviour that characterizes many economic time series, it is convenient to include a damping factor ρ which gives for flexibility to the stochastic cycle, as follows:

---

[9]These models may be used to find global non-linear trends as, for example, those cases with a saturation level. This solution, although feasible for some periods of time, does not capture certain behavioural characteristics in an adequate way as other alternative formulations, such as growth curves (Harvey, 1989).

[10]That is, the parameter λ represents the number of times the cycle repeats itself over a period of 2π. Therefore, the ratio $\lambda/2\pi$ expresses the number of times the cycle repeats per unit of time.

$$\begin{pmatrix} \Psi_t \\ \Psi_t^* \end{pmatrix} = \rho \begin{pmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{pmatrix} \begin{pmatrix} \Psi_{t-1} \\ \Psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} k_t \\ k_t^* \end{pmatrix} \tag{11}$$

where $0 < \rho < 1$[11] If $\lambda = 0$ or $\lambda = \pi$, the stochastic cycle converts itself into a first order autoregressive process[12]. If $\lambda = 0$ and $\rho = 1$ the cyclical component transforms into a local level model specified for the trend[13].

The cyclical component may be combined with a trend in many different ways, but the most common formulations are: 1) the cyclical model,

$$y_t = \mu + \Psi_t + \varepsilon_t \tag{12}$$

with $0 < \rho < 1$[14]; 2) the trend cycle model,

$$y_t = \mu_t + \Psi_t + \varepsilon_t \tag{13}$$

where μ_t is a local linear trend[15]; a cyclical trend model,

$$y_t = \mu_t + \varepsilon_t \tag{14}$$

where $\mu_t = \mu_{t-1} + \Psi_{t-1} + \beta_{t-1} + \eta_t$ and $\beta_t = \beta_{t-1} + \zeta_t$[16]

Other specifications consist in adding an AR(2) component to $y_t = \mu_t + \varepsilon_t$ or $\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$[17]

If the observations in the series under scrutiny have a period smaller than one year, often seasonality effects may appear and, therefore, it is convenient to introduce such a component

---

[11]Notice that the incorporation of a damping factor smaller than one gives the cycle a representation equivalent to a VAR(1) process and then, in forecasting exercises, the cyclical fluctuation tends to disappear.

[12]Although $\rho \geq 0$, the parameter in the equivalent first order autoregressive formulation may be either positive ($\lambda = 0$) or negative ($\lambda = \pi$).

[13]Notice that the cycle variance, σ_Ψ^2, is related to the variance of the cycle disturbance term, σ_k^2, through the expression $\sigma_k^2 = (1 - \rho^2)\sigma_\Psi^2$, when $0 < \rho < 1$.

[14]This is equivalent to an ARMA (2,2) process with constant. The cycle is itself an ARMA (2,1). The MA part is subject to restrictions, however the most important ones refer to the AR part. If $0 < \lambda < \pi$, the roots of the AR polynomial are a pair of complete roots together with a module ρ^{-1} and a phase λ. If, on the other hand, $0 \leq \rho < \pi$, those roots lie outside the unit circle and the process is therefore stationary. Given that the roots of an AR(2) polynomial can be either real or complex, the formulation of the cyclical model is restricted to a feasible range of values for the AR coefficients such that it generates a pseudocyclical behaviour (Harvey, 1993).

[15]This representation is equivalent to an ARIMA(2,2,4) process (Harvey, 1989).

[16]If in this model one assumes $\lambda = 0$ and one eliminates the term β_{t-1}, then Ψ_t is a trend component and we obtain a damped local linear trend model. If we assume that $\rho = 1$ and $\sigma_k^2 = 0$ then we end up with a deterministic linear trend.

[17]Notice that if the time series presents a cyclical behaviour of frequency λ, such a behaviour can be captured by an autoregressive polynomial as $(1 - 2\cos(\lambda)L + L^2)$, with L being a lag operator.

into a model. If the seasonal fluctuations are stable, the seasonal component can be defined either as deterministic or fixed. If the number of seasons is s and, given that the sum of the seasonal effects throughout the year should equal zero, then the seasonal component at time t can be speficied as:

$$\gamma_t = \sum_{j=1}^{s-1} z_{j,t} \gamma_j \tag{15}$$

where $z_{j,t}$, $j = 1, ..., s - 1$, is equal to 1 if t belongs to season j, -1 if t belongs to season s, and 0 otherwise.

Formally, we have:

$$\gamma_t = \begin{cases} \gamma_t, \text{ if } t \text{ belongs to season } j, \\ -\sum_{j=1}^{s-1} \gamma_j, \text{ if } t \text{ belongs to season s} \end{cases} \quad j = 1, ..., s - 1 \tag{16}$$

Instead, if the seasonal pattern evolves over time, it seems more adequate to incorporate a stochastic seasonal, which admits, at least, two possible formulations.

One way of getting the seasonal pattern to evolve through time is to allow that the sum of the seasonal effects over the year does not exactly equal zero, but are equal to a random disturbance term. That is, the values of the seasonal component at times $t, t-1, ..., t-s+1$ are related through $\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t$, or $\gamma_t = \sum_{j=0}^{s-1} \gamma_{t-j} + \omega_t$, being ω_t a random disturbance term such that $\omega_t \sim NID(0, \sigma_\omega^2)$. Another possibility is to assume that each season's effect evolves according to a random walk (Harrison and Stevens, 1976). Hence, if t belongs to season j, $\gamma_t = \gamma_{t,,j}$, where

$$\gamma_{t,,j} = \gamma_{t,j-1} + \omega_{j,t}, \ j = 1, ..., s \tag{17}$$

and ω_j is a random disturbance term such that $\omega_t \sim NID(0, \sigma_\omega^2)$.

Another alternative is a trignometric formulation[18] in which

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t} \tag{18}$$

where $[s/2]$ denotes the integer part of $s/2$ and each $\gamma_{j,t}$ is generated by

---

[18]Modelling a fixed seasonal pattern through a set of trignometric terms was initially introduced by Hannan et al. (1970).

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} = \begin{pmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix} \tag{19}$$

for $j = 1, ..., [s/2]$ and $t = 1, ..., T$; $\lambda_j = 2\pi j/s$ is the frequency, in radians, and $\omega_{j,t}$ and $\omega_{j,t}^*$ are the random disturbance terms mutually uncorrelated according to $NID(0, \sigma_\omega^2)$ in both cases[19]. If s is even, $\gamma_{s/2,t} = -\gamma_{s/2,t-1} + \omega_{s/2,t}$. Notice that $\gamma_{j,t}^*$ exists by construction with the same purpose as the cyclical component[20].

Notice additionally that in both specifications the number of parameters involved in the seasonal component is the same and, if $\sigma_\omega^2 = 0$, the seasonal pattern is deterministic[21].

As the seasonal pattern changes relatively smoothly over time, some times it is reasonable to eliminate some of the trignometric terms corresponding to the highest frequencies (Anderson (1971) and Abraham and Box (1978)). However, when these terms are responsible for an important part of the seasonal variation, this would not be a adequate way to proceed. This happens when one has weekly observations in which there is an important intra-monthly effect, that is reflected in cycles with periods of 4 weeks (Pierce et al., 1984); in this case, the seasonal component could be formulated using trignometric terms associated with both low frequencies and frequencies of $\pi/2$ and π, as follows:

$$y_t = \mu_t + \gamma_t^{bf} + \gamma_t^{im} + \varepsilon_t \tag{20}$$

where γ_t^{bf} contains, e.g., the variations corresponding to frequencies $2\pi j/52$, $j = 1, ..., 6$, and γ_t^{im} the ones associated with frequencies $\pi/2$ and π[22]

Furthermore, in the case of weekly observations, the number of weeks per year varies between 52 and 53 and the same week does not correspond to the same period in a given year. Certain festivities, as the Easter week, do not coincide with the same moment in time from year to year. One way to address this problem is to consider that the seasonal component is the result of a combination of components, as follows:

---

[19] Assigning different variances to the random disturbance terms corresponding to trignometric terms associated to different frequencies, one can have distinct cyclical fluctuations with different degrees of variability. However, from a practical point of view it is desirable that they have the same variance because there is a small loss in terms of goodness of fit and there is a big gain in computation time. On the other hand, the specification for the seasonal component could be modified by replacing the white noise disturbance term by an ARMA process.

[20] This formulation may be more effective to model seasonal behaviour in economic time series, as it assumes a seasonal pattern that evolves over time according to an MA(s-2) process, which implies the fluctuations are smoother than the ones generated by a white noise disturbance term that characterizes formulations based on qualitative variables.

[21] This option is recommended when the number of years in the sample is small such that there aren't enough observations to allowing a changing seasonal pattern.

[22] An alternative simplified formulation of the seasonal component can be obtained through polynomial functions of order three define in branches (cubic spline).

$$\gamma_t = \gamma_t^p + \gamma_t^f \tag{21}$$

where γ_t^p represents the periodic component, which is a function of the number of days in a given year that have passed until the observation is recorded[23], and γ_t^f is the calendar component, which recognizes the effects associated with festivities that always take place at the same day of the week although at different dates in two distinct years[24].

If one has daily observations it is possible to incorporate into the model a component θ_t to pick up the behaviour attributed to the weekly day to which belongs the observation recorded at time t. Denote by w the number of different types of days within a week and by k_j the number of $j - type$ days in some given week, for $j = 1, ..., w$, then the effect associated to the jth day-type is $\theta_{j,t}$ where

$$\theta_{j,t} = \theta_{j,t-1} + \chi_{j,t}, \ j = 1, ..., w-1 \tag{22}$$

being $\chi_{j,t}$ the zero mean disturbance term with variance $\sigma_\chi^2(1 - k_j^2/K)$, $j = 1, ..., w-1$, with $K = \sum_{j=1}^w k_j^2$ and disturbance covariances, $\chi_{j,t}$ and $\chi_{h,t}$ respectively equal to $-\sigma_\chi^2 k_j k_h/K$, $h, j = 1, ..., w-1, j \neq h$. Although each one of the w daily effects, $\theta_{j,t}$, evolves over time, the value of the daily component at some point in time, θ_t, is given by the effect value associated to the day of the week to which the observation recorded at time t belongs, such that, $\theta_t = \theta_{j,t}$, $t = 1, ..., T$ where j is the type of day to which the observation belongs to. As the model is formulated to the first $w - 1$ daily effects, the effect for the day of type w is obtained through the condition that, at each moment in time, the sum of the daily effects corresponding to different days of the week equals zero, that is, $\theta_{w,t} = -k_w^{-1} \sum_{j=1}^{w-1} k_j \theta_{j,t}$[25]. The application scope of this last model is limited to daily effects, but may be interpreted as a generalization of the model formulated for the seasonal component in which one assumes that not all seasons are equally weighted.

Lastly, if the frequency of observations is of order less than annual and the analysis reveals the absence of seasonal variations or the data has been seasonally adjusted[26], one can apply a structural model without a seasonal component.

---

[23]The most common specifications used to recognize this periodic component are the well known spline functions (Koopman, 1992).

[24]This approach requires working on a daily basis such that one knows the number of days passed since the beginning of a given year until each observation is recorded (Harvey, Koopman and Riani, 1995).

[25]$\theta_{w,t}$ has exactly the same statistical properties as $\theta_{j,t}$, $j = 1, ..., w-1$.

[26]Notice that seasonally adjusted data not always present the desired properties, especially if the seasonal pattern changes in such a way that it is not captured by usual adjustment methods. In general, it is recommended not to use data that has been previously transformed aiming at eliminating some of its components; hence, the advantage of the structural approach is that it allows to jointly estimate the different elements composing a time series (Harvey, 1991).

9

Taking the trend to be formulated as a local linear trend model and the seasonal component as one of the several basic stochastic formulations, i.e., without allowing possible intra-monthly nor calendar effects, we obtain the Basic Structural Model[27]. If we select a trignometric formulation for the seasonals, then the Basic Structural Model is specified as follows:

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \ \varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \tag{23}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \ \eta_t \sim NID(0, \sigma_\eta^2) \tag{24}$$

$$\beta_t = \beta_{t-1} + \zeta_t, \ \zeta_t \sim NID(0, \sigma_\zeta^2) \tag{25}$$

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t} \tag{26}$$

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} = \begin{pmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix}, \ j = 1, ..., [s/2], \ \lambda_j = 2\pi j/s, \tag{27}$$

$$\omega_{j,t} \sim NID(0, \sigma_\omega^2), \ \omega_{j,t}^* \sim NID(0, \sigma_\omega^2), \ Cov\left\{\omega_{j,t}, \omega_{j,t}^*\right\} = 0 \tag{28}$$

The model specification is closed assuming the inexistence of correlation between the distinct disturbance terms due to identification reasons[28]. Given that the statistical properties of the model depends on the complete set of disturbances, then the autocovariance matrix of the series is simply the sum of the autocovariances of the different components.

Lastly, as a natural extension, the model may include lagged values of a dependent variable as well as other exogenous regressors:

$$y_t = \mu_t + \gamma_t + \psi_t + \sum_{\tau=1}^{p} \varphi_\tau y_{t-\tau} + \sum_{i=1}^{k} \sum_{\tau=0}^{q_k} \delta_{i\tau} x_{i,t-\tau} + \varepsilon_t \tag{29}$$

where $x_{i,t}$ are exogenous variables[29], and φ_τ, and $\delta_{i\tau}$ are unknown parameters. In case

---

[27]This is equivalent to a process $\Delta\Delta_s y_t \sim MA(s+1)$ or alternatively to $y_t = \zeta_t/\Delta^2 + S^{-1}(L)\omega_t + \varepsilon_t$. Given that the operators $\Delta^2 y S(L)$ do not possess common roots, changes in the seasonal pattern are not misunderstood with changes in the trend (Harvey, 1989). A Basic Structural Model with qualitative seasonal variables constitutes an equivalent formulation in forecasting terms to Box and Jenkin's (1976) proposal to analyse the series "number of passengers".

[28]The independence of innovations, which implies independence of their respective components, has been commonly established in the literature. The justification for the decomposition rests, indeed, on the need to break down different types of evolutions, although they are observed all together. According to Harvey and Koopman (1992) and Koopman (1992), estimates of these innovations can present high correlation levels. In general, one may assume that there are correlation structures between disturbance terms and, therefore, the relationship between distinct components is more adequate when dealing with economic time series.

[29]$x_{i,t}$ is exogenous if it does not give any information about the non-observed components at time $t+s$, for $s = 0, 1, 2, ...,$ over and above the information contained in $y_{t-1}, ..., y_1$ (Hamilton, 1994).

the model has some lagged values of the dependent variable, e.g.:

$$y_t = \phi y_{t-1} + ... + \phi_p y_{t-p} + \mu_t + \varepsilon_t \tag{30}$$

then the trend is redefined as:

$$\mu_t^+ = \Phi_p^{-1}(L)\mu_t \tag{31}$$

such that,

$$y_t = \mu_t^+ + \Phi_p^{-1}(L)\varepsilon_t \tag{32}$$

being $\Phi_p(L) = (1 - \phi_1 L - ... - \phi_p L^p)$. In this case, both the trend and irregular components are subjected to similar autoregressive effects.

When one has high frequency data, usual specifications for the seasonal component require the incorporation of a large number of parameters which may bring additional estimation problems. In this context, an alternative is to use spline functions, which constitute more parsimonious specifications that assume the use of qualitative variables.

A case of particular interest is when explanatory variables are intervention variables which allow to pick up outlier observations and/or structural breaks (Harvey 1989, and Harvey and Koopman 1992). In the simplest case, i.e., an outlier observation at some place in time is dealt with an intervention or impulse variable at that point. Similarly, a structural break characterized by a permanent increase or decrease of the level component of a given series from a certain point in time, may be captured through an intervention variable as a step or, alternatively, by adding an impulse intervention variable in the level equation. If, on the contrary, there is a permanent change in the slope of the trend component at some moment in time, then this change may be modelled as an intervention variable taking values 1,2,3... from that point onwards, or alternatively, by adding an impulse intervention variable into the slope equation[30].

2.2 Multivariate time series models and Cointegration

When we have more than one time series each of them may be analysed independently, however, as they might be subjected to similar influences and, therefore, be correlated, then a joint estimation should be considered to capture such interaction and common effects. Hence, the need to refer to multivariate types of time series models, as such correlation between

---

[30]Although impulse or step-type intervention variables are appropriate to model sudden changes in the level or slope, one can also use them to pick up unexpected changes in the seasonal component.

different series is introduced through the non-zero elements outside the main diagonal of the disturbances' covariance matrix.

In this context, the focus on multivariate analysis means a greater approximation to reality, as one is able to capture dynamic interactions between series; however, in any case, the underlying main assumptions rest on the univariate ones. The latter can be easily generalized to a multivariate version, in which we analyse a vector of $N \times 1$ observable variables $\mathbf{y}_t = (y_{1,t}, ..., y_{N,t})'$. As an example, the local level version of the multivariate model is $y_{i,t} = \mu_{i,t} + \varepsilon_{i,t}$, with $\mu_{i,t} = \mu_{i,t-1} + \eta_{i,t}$, for all $i = 1, ..., N$ and $t = 1, ..., T$, as follows:

$$
\begin{pmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t} \\ \vdots \\ \mu_{N,t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{N,t} \end{pmatrix}
\tag{33}
$$

with

$$
\begin{pmatrix} \mu_{1,t} \\ \vdots \\ \mu_{N,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t-1} \\ \vdots \\ \mu_{N,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \vdots \\ \eta_{N,t} \end{pmatrix}
\tag{34}
$$

such that $\varepsilon_t = (\varepsilon_{1,t}, ..., \varepsilon_{N,t})' \sim NID(\mathbf{0}, \mathbf{\Sigma}_\varepsilon)$ and $\eta_t = (\eta_{1,t}, ..., \eta_{N,t})' \sim NID(\mathbf{0}, \mathbf{\Sigma}_\eta)$. Each of the variables in vector $\mathbf{y}_t$ has its own stochastic level and the matrix $\mathbf{\Sigma}_\eta$ contains the covariances between different levels. This model is commonly referred as SUTSE model (seemingly unrelated time series equations), as series are only linked through contemporaneous correlations established between matrices $\mathbf{\Sigma}_\varepsilon$ and $\mathbf{\Sigma}_\eta$.

Henceforth, these multivariate models can incorporate much more complete specifications for the trend component, as well as both cycles and seasonals. As far as the cyclical component is concerned, in particular when dealing with economic time series, it seems reasonable to assume that movements in different series are related to a common economic cycle. Similar considerations can be formulated to seasonals.

Under certain assumptions, the multivariate local level structural model allows for the following representation of the irregular:

$$
y_{i,t} = \mu_{i,t} + \varepsilon_t + \varepsilon_{i,t}^*
\tag{35}
$$

$$
\mu_{i,t} = \mu_{i,t-1} + \eta_t + \eta_{i,t}^*
\tag{36}
$$

being $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$, $\varepsilon_{i,t}^* \sim NID(0, \sigma_{\varepsilon^*}^2)$, $\eta_t \sim NID(0, \sigma_\eta^2)$ and $\eta_{i,t}^* \sim NID(0, \sigma_{\eta^*}^2)$, for

all $i = 1, ..., N$ and $t = 1, ..., T$[31] and where ε_t and η_t capture the common effects, whereas ε_i, t^* and η_i, t^* capture the especific effects. A especially interesting case is when $\sigma_{\eta>0}^2$, but $\sigma_{\eta^{*2}=0}$; then the N series have trend components that move in a parallel fashion and therefore, they are cointegrated.

In general, multivariate models should admit the possibility of common trends to some variables composing the vector $\mathbf{y}_t$[32]. In this case, there exist cointegration between series, which can be incorporated in a model such as:

$$\mathbf{y}_t = \Theta \mu_t^* + \mu_0 + \varepsilon_t \tag{37}$$

$$\mu_t^* = \mu_{t-1}^* + \eta_t^* \tag{38}$$

with $\varepsilon_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ and $\eta_t^* \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_{\eta^*})$ for all $t = 1, ..., T$, being $\boldsymbol{\Sigma}_{\eta^*}$ a diagonal matrix $k \times k$, μ_0 a $N \times 1$ vector whose first k elements are non-zero, and Θ is a $N \times k$ fixed matrix such that $\Theta_{ij} = 0, \forall j > i$ and $\Theta_{ii} = 1, \forall i$. If there are k common trends, then there will exist $N - k$ cointegrating relations. The $N - k$ correspondent cointegrating vectors define the $N - k$ rows in the matrix $\mathbf{A}_{(N-k) \times N}$, such that $\mathbf{A}\Theta = 0$, i.e., $\mathbf{A}\mathbf{y}_t = \mathbf{A}\mu_0 + \mathbf{A}\varepsilon_t$, such that $\mathbf{A}\mathbf{y}_t$ is a stationary process of dimension $(N - k) \times 1$. If $k = N$, there are no trends, therefore, no cointegration[33]. As $\boldsymbol{\Sigma}_{\eta^*}$ is diagonal, then common trends (the elements of μ_t^*) are independent.

For a model with local level trend with drift, the common trend specification is:

$$\mathbf{y}_t = \Theta \mu_t^* + \mu_0 + \varepsilon_t \tag{39}$$

$$\mu_t^* = \mu_{t-1}^* + \beta + \eta_t^* \tag{40}$$

for all $t = 1, ..., T$, $\varepsilon_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ and $\eta_t^* \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_{\eta^*})$, being β a $k \times 1$ vector of slopes. As an example, take a bivariate local level model a follows:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \tag{41}$$

---

[31] Note that $\Sigma_\varepsilon = \sigma_{\varepsilon^*}^2 \mathbf{I} + \sigma_\varepsilon^2 \mathbf{i}\mathbf{i}'$ and $\Sigma_\eta = \sigma_{\eta^*}^2 \mathbf{I} + \sigma_\eta^2 \mathbf{i}\mathbf{i}'$, where $\mathbf{i}$ is a unit vector of dimension N and $\mathbf{I}$ is the identity matrix.

[32] Those series which have the same source of growth are an example in which it makes sense to model common trends.

[33] One way to determine k is based on estimating the original model as if no common factors exist, apply a principal components analysis on the estimated matrix Σ_η and then take the value k that captures an important part of total variability. A more formal approach was suggested by Stock and Watson (1988).

13

$$\begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \tag{42}$$

for all $t = 1, ..., T$, where

$$\boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & \sigma_{\varepsilon_{12}}^2 \\ \sigma_{\varepsilon_{21}}^2 & \sigma_{\varepsilon_2}^2 \end{pmatrix} \tag{43}$$

and

$$\boldsymbol{\Sigma}_\eta = \begin{pmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_{12}}^2 \\ \sigma_{\eta_{21}}^2 & \sigma_{\eta_2}^2 \end{pmatrix} \tag{44}$$

Alternatively, its representation with an error components model is:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \end{pmatrix} + \varepsilon_t \mathbf{i}_2 + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \tag{45}$$

$$\begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{pmatrix} + \eta_t \mathbf{i}_2 + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \tag{46}$$

for all $t = 1, ..., T$, being $\mathbf{i}_2 = (1, 1)'$, $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$, $\varepsilon_t^* \sim NID(0, \sigma_{\varepsilon^*}^2)$ for $i = 1, 2$, $\eta_t \sim NID(0, \sigma_\eta^2)$ and $\eta_t^* \sim NID(0, \sigma_{\eta^*}^2)$, for $i = 1, 2$. If $\sigma_{\eta^*}^2 = 0$ we obtain $\mu_{1,t} = \mu_{1,t-1} + \eta_t$ and $\mu_{2,t} = \mu_{2,t-1} + \eta_t$ such that trend components in both series follow parallel paths. It follows that processes $y_{1,t}$ and $y_{2,t}$ are non-stationary, hence they don't move around a constant level, however, there is a common trend shared by the two series. That is, $y_{1,t}$ and $y_{2,t}$ are I(1) and there exists a stationary linear relation, i.e., the series are cointegrated. If $\mu_{1,0} = \mu_{2,0}$ such a relation is in fact $y_{1,t} - y_{2,t}$.

In general, one can check the presence of cointegration in the bivariate case through a common trend model given by:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 1 \\ \theta_{2,1} \end{pmatrix} \mu_t^* + \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \tag{47}$$

where $\mu_t^* = \mu_{t-1}^* + \eta_t^*$.

In this context we have the following representation for the two series:

$$y_{1,t} = \mu_t^* + \varepsilon_{1,t} \tag{48}$$
$$y_{2,t} = \theta_{2,1} \mu_t^* + \mu_2 + \varepsilon_{2,t} \tag{49}$$

where $\mu_t^* = \mu_{t-1}^* + \beta + \eta_t^*$.

So,

$$y_{1,t} - \frac{1}{\theta_{2,1}} y_{2,t} = -\frac{\mu_2}{\theta_{2,1}} + \varepsilon_{1,t} - \frac{1}{\theta_{2,1}} \varepsilon_{2,t} \tag{50}$$

is a stationary linear combination which implies cointegration. The cointegrating vector $\begin{pmatrix} 1 & -1/\theta_{2,1} \end{pmatrix}$ is such that $\begin{pmatrix} 1 & -1/\theta_{2,1} \end{pmatrix} \begin{pmatrix} 1 \\ \theta_{2,1} \end{pmatrix} = 0$, and the element causing the non-stationarity is eliminated.

Common trends as specified before imply cointegration at frequency zero, however all these concepts can be extended to other frequencies, including additionally the existence of cyclical and seasonal common components, which may be interpreted in the sense that changes in cyclical or seasonal patterns in different time series are driven from a common source.

The properties of a multivariate model and, more specifically, the multivariate local level model depend on the structure of the variance-covariance matrices $\boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Sigma}_\eta$. Under the constraint of homogeneity, which establishes that $\boldsymbol{\Sigma}_\eta = q\boldsymbol{\Sigma}_\varepsilon$ (Harvey, 1989), all linear combinations of elements of $\mathbf{y}_t$ have the same stochastic properties[34].

When in a multivariate model all but one variables composing $\mathbf{y}_t$ are exogenous, it is not necessary to formulate a N-equations system, as one can simply formulate a single-equation dynamic regression model. In the context of structural models, one can obtain a single-equation model with exogenous variables of the type:

$$\mathbf{y}_t = \mu_t + \mathbf{X}_t \delta + \varepsilon_t \tag{51}$$

with $\mu_t = \mu_{t-1} + \eta_t$, $\mathbf{X}_t = (x_{1,t}, ..., x_{N-1,t})$ and $\delta = (\delta_1, ..., \delta_{N-1})$.

If the level variance equals zero, we obtain the classical basic regression model. When the irregular component variance equals zero, but the level component is different from zero, we obtain a regression in first differences. Finally, when both variances are different from zero, we get a transference function model. Cases with the inclusion of explanatory variables in multivariate structural models include intervention analysis and the error correction mechanism model (Harvey and Scott, 1984).

---

[34]That is, for a local level model, $\mathbf{y}_t$ is homogenous if the stationary process $\Delta \mathbf{w}' \mathbf{y}_t$ has the same spectral density and the same autocorrelation function for all vector $\mathbf{w}$ of fixed parameters (non stochastic). In that case, the structural model for any combination $\mathbf{w}' \mathbf{y}_t$ verifies the relation $\Sigma_\eta = q\Sigma_\varepsilon$ and the autocorrelation function and spectral density depend solely on q.

3 Structural Models versus ARIMA

Structural models have several advantages when compared with ARIMA models, some of which will be discussed in this section. ARIMA models can be formulated in state space form[35] and many structural models admit an ARIMA representation. In fact, when time series have a simple underlying structure both formulations are basically equivalent to each other; however, when such structure is more complex, then differences between both approaches become more evident[36].

In a time series structural model each component, such as the trend, cycle or seasonal changes, is explicitly formulated and, therefore, it is possible to get specific information about them. Moreover, the researcher can identify and introduce any characteristic which might require a particular treatment in special circumstances. This is perhaps, the main advantage of a structural model vis-a-vis an ARIMA approach, in which the trend and seasonal components are eliminated by applying convenient differences to the original series before carrying out the analysis. In sum, the ARIMA methodology constitutes itself as a kind of black-box in which the adopted model depends entirely from the data, without a prior analysis of the structure underlying the generating system. Structural models are, from this point of view, more transparent as they allow to check if the predicted behaviour by the model for each component corresponds to what is expected from the data.

One should take into account that eliminating both the trend and seasonality may have strong inconveniences if , as it happens with official statistics or in many econometric applications, those components have interest in themselves in addition to forecasting exercises. Furthermore, the requirement of stationarity in Box and Jenkins' (1976) approach implies differentiating the series, however not always one is able to decide the right integration order. In fact, basic tools to identify ARIMA models, i.e., autocorrelation functions, are quite imprecise and very often do not allow to opt for a unique model[37].

State space models are also more flexible. The recursive nature of the model and the computation techniques used for its analysis allow the direct incorporation of known breaks in the system structure over time. On the contrary, Box and Jenkins' models are based on the assumption that differenced series are stationary, which incorporates more rigidity. In this

---

[35]Some examples of ARIMA models in State Space Form can be found in Harvey (1989), Brockwell and Davis (1991) and Hamilton (1994).

[36]Harvey (1985) shows some of these differences by analysing annual macroeconomic time series. In particular, he observes that ARIMA models selected under a parsimonious principle may give place to formulations which, though providing good short-term forecasts, do not recognize the most important characteristics of the series under scrutiny and they have little economic interpretation. Some of these differences are also discussed in Durbin and Koopman (2001).

[37]Under the ARIMA methodology it is common to find that economic magnitudes behave as integrated processes of order one, i.e., processes that are characterized by persistent changes.

context, with the structural approach forecasting is relatively straightforward[38] and missing observations are easier to treat[39]. Brockwell and Davis (1991) consider that state-space representation and recursive equations which characterize the Kalman Filter are ideal to analyse series with missing observations.[40]. Using an ARIMA approach, however, is more difficult to handle missing observations, and the incorporation of explanatory variables, calendar effects and structural breaks is not always are immediate as with state space representation.

On the other hand, observations corresponding to multivariate series can be manipulated by direct extension of the univariate structural formulation; which does not happen with ARIMA. Moreover, the Markovian nature of state space models allows the necessary computations to be implemented in a recursive way; this, in fact, allows manipulation of high dimensional models without an overwhelming increase of the computational task.

In this sense, multivariate structural models of economic time series have some advantages over the VAR representation, which in general required a large number of parameters. In particular, when some series are non-stationary, a VAR in differences might not be an adequate specification if different series have distinct integration orders. So, the application of differences would allow the appearance of strictly non-invertible MA components. In these cases, it seems more likely to formulate the model in levels; however, when series are non-stationary it may be necessary to use an higher order VAR. A structural model offers a better alternative formulation.

Furthermore, these last models allow the imposition of cointegration restrictions as well as consider explanatory variables (Harvey, 1989). Given that in a multivariate structural model the trend is explicitly modelled, short-term movements can be captured by a low order VAR and cointegration will appear when there are common trends. Hence, although structural models are the adequate way to provide the most important characteristic in observed series, they convert into behavioural models once one incorporates restrictions introducing short-run dynamic movements and including exogenous variables in some of their equations.

According to Durbin and Koopman (2001) the main advantage of structural models compared with ARIMA models rests of the fact that the former are taught and used in few universities, there is scarce literature about them in statistical-econometric terms and only recently they were provided with specific software. The complete reverse is true for the Box and Jenkins methodology.

---

[38]Harvey and Todd (1983) compare forecasts from ARIMA and structural models and they conclude that there are strong arguments in favour of the structural approach from a practical point of view.

[39]In order to get filtered estimates corresponding to missing observations, we simply assume that $e_t = 0$ and $K_t = 0$ in those missing values (Harvey and Koopman, 1992 and Durbin and Koopman, 2001).

[40]The authors show several advantages of this approach vis-a-vis ARIMA models by evaluating a Gaussian likelihood function at the observed values together with a minimum mean squared error for missing observations. Jones (1993) also presents a procedure to deal with missing observations in state space representation.

4   The Kalman Filter - an introductory note

The purpose of this section is to provide an introduction to the Kalman filter as well as to establish the relation between this filter and the state-space form representation. The importance in studying the Kalman algorithm[41] is based on the fact it constitutes the main procedure to estimate dynamic systems represented in state-space form, which have many interesting econometric applications (as seen in the previous sections).

The filter has its origin in Kalman's (1960) paper who describes a recursive solution to the linear filtering problem of discrete data. Kalman's derivation took place within the context of state-space models, whose core is the recursive least squares estimation. Since the 1960s, greatly due to the computer and digital revolution, the Kalman filter has been object of extensive study and application, especially in the fields of autonomous and assisted navigation, missile's tracking and economics.

The state-space representation is essentially a convenient notation to estimate stochastic models in which one assumes measurement errors in the system which allows handling a large set of time series models. Among its uses one can refer to unobserved components modelling and time varying parameters, as well as representation of ARIMA-type models and some others which may be approximated by maximum likelihood.

The filter is a mathematical tool which operates by means of a prediction and correction mechanism. Essentially, this algorithm predicts the new state (which contains all information up to that point in time) starting from a previous estimation and adding a proportional correcting term to the prediction error, such that the latter is statistically minimized.

Within the state-space notation, the Kalman filter derivation rests on the assumption of normality of the initial state vector and well as the disturbances of the system. Hence, it is possible to compute the likelihood function of the prediction error, which is then used to estimate the unknown parameters of the system.

The complete estimation procedure is as follows: the model is formulated in state-space form and for a set of initial parameters, the model prediction errors are generated from the filter. These are then used to recursively evaluate the likelihood function until it is maximized.

---

[41] An algorithm is a specific set of instructions to solve a problem, usually required that such procedure comes to an end eventually. Specific algorithms are sometimes named method, procedure or technique. The word algorithm is a deformation of "al-Khwarizmi", a Persian mathematician who wrote an important treaty on algebrical methods.

4.1 Recursive Estimation and the Kalman Filter

Beginning with a bit of history, the concept of least squares regression is partly associated with Legendre (1805) who first published this theory in 1805 and actually named it "least squares". Nevertheless, it was Gauss (1809) who developed the method as a statistical method by incorporating least squares in a context in which there is a probabilistic treatment of the error terms (D.S.G. Pollock, The Kalman Filter). The first exposition of the least squares method by Gauss is intrinsically linked to the estimation of six coefficients to determine the elliptical orbit of a planetary body, when the number of observations exceed the number of parameters. The second exposition has been presented in a series of papers published in 1821, 1823 and 1826, under the title Theoria Combinatonis Observationum Erroribus Minimis Obnoxiae (1823), which contains the famous theorem that states the amongst all the unbiased linear estimators, the ordinary least squares has the smaller mean squared error - the Gauss Markov Theorem.

The relevance of Gauss' second exposition to the estimation theory of recursive least squares and to the concept of the Kalman filter is found in a brief paragraph in which Gauss shows that is possible to detect, with increased probability, changes of an unknown event when a new equation is incorporated with some (ex-ante) computed weights. In fact, Gauss developed the estimation algorithm of recursive least squares.

Gauss' algorithm has been ignored for more than one and a half century by the time it was rediscovered by two different people. The first one was Plackett (1950) before the computing age and who also ended up forgotten. The second rediscover of the recursive algorithm was R. E. Kalman in 1960 in this context of control theory[42]. Since Kalman's (1960) and Kalman and Bucy's (1961) papers, which described a recursive solution to the problem of linear filters in discrete data, this algorithm has gained an extensive attention in research and subsequent applications.

Plackett's exposition of the recursive ordinary least squares rests on algebra employing statistical concepts from classical linear regression models. Kalman's derivation was made within a wider context of state-space models with time-varying parameters. Hence, the core of the Kalman filter is the Gauss-Plackett estimation algorithm of recursive least squares, however in a context where the extension and algebrical complexity increases, Kalman has based the construction of the filter in probabilistic theory, more specifically, on the conditionally Gaussian properties of random variables. The proposed criterion was to minimize the norm of the state vector covariance matrix generating the classical recursion: the estimation of the new state is computed by adding the previous estimate to a correction term

---

[42]It is the mathematical study of how parameter's manipulation affect the behaviour of a given system in order to reach an optimal or desired result.

proportional to the forecast error.

From Kalman's original paper many other derivations have followed. The great majority of those intent to reduce the terminology to something close to the OLS regression theory. Others have arisen: from maximum likelihood functions to Bayesian econometrics. The truth is that the Kalman filter is a complex topic: its derivation, by any method, is lengthy and its equations difficult to obtain. Nevertheless, it is this complexity with grants to the Kalman filter its enormous power to solve a wide range of problems of statistical inference.

The Kalman filter comprises a set of mathematical equations which result from an optimal recursive solution given by the least squares method. The purpose of this solution consists in computing a linear, unbiased and optimal estimator of a system's state[43] at time t, based on information available at $t-1$ and update, with the additional information at t, these estimates (Clar et al. 1998). The filter's performance assumes that a system can be described through a stochastic linear model with an associated error following a normal distribution with mean zero and known variance. The solution is optimal provided the filter combines all observed information and previous knowledge about the system's behaviour such that the state estimation minimizes the statistical error. The recursive term means that the filter re-computes the solution each time a new observation is incorporated into the system.

The Kalman filter is the main algorithm to estimate dynamic systems in state-space form. This representation of the system is described by a set of state variables. The state contains all information relative to that system at a given point in time. This information should allow to infer about the past behaviour of the system, aiming at predict its future behaviour. What makes the Kalman filter so interesting is its ability to predict the past, present and future state of a system, even when the precise nature of the modelled system is unknown. In practical terms, the individual state variables of a dynamic system cannot be exactly determined through a direct measurement. In this context, their measurement is done by means of stochastic processes involving some degree of uncertainty.

4.2   Kalman Filter and Smoothing

Take a simple model in which trend $\mu_t = \alpha_t$, where α_t is a random walk, with no seasonals and all variables normally distributed. Assume that the error term ε_t has constant variance σ_ε^2. This model is given by:

---

[43]The state should contain the most relevant information of the system at each moment in time, trying to consider the lowest possible number of variables.

$$y_t = \alpha_t + \varepsilon_t, \ \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{52}$$

$$\alpha_{t+1} = \alpha_t + \eta_t, \ \eta_t \sim N(0, \sigma_\eta^2) \tag{53}$$

for $t = 1, ..., n$ where ε_t and η_t are mutually independent and independent from α_1. Despite its simplicity, the model - known as Gaussian linear State Space model - provides the basis for several practical problems when analysing time series. It presents a structure characteristic in state space models in which there is a set of non observed values $\alpha_1, ..., \alpha_n$, which represents the development of the system over time, and a set of observations $y_1, ..., y_n$ that relate with state α, by means of the state space model above. The purpose of the methodology is to infer the relevant properties about the $\alpha's$ through ex-ante knowledge about the available observations.

We assume that initially $\alpha_1 \sim N(\alpha_1, P_1)$ where α_1 and P_1 are known as well as σ_ε^2 and σ_η^2. Given that random walks are non-stationary, the model is also non-stationary. In order to apply the model to observed series we need to compute values such as the mean of α_t given $y_1, ..., y_{t-1}$ or the mean of α_t given $y_1, ..., y_n$ together with their variances; we need additionally to adjust the model to the data by computing the maximum likelihood estimates of parameters σ_ε^2 and σ_η^2. The proposed methodology is to employ filtering and smoothing techniques as outlined below.

The main purpose of filtering is to update our knowledge about the system each time a new observation y_t is known. As all distributions are Gaussian, conditional joint distributions of a set of observations given another set, also follow a Normal distribution. Take Y_{t-1} the set of past observations $\{y_1, ..., y_{t-1}\}$ and assume that the conditional distribution of α_1 given Y_{t-1} is $N(a_t, P_t)$, where a_t and P_t are known. Given that the latter two are known, the aim is to compute a_{t-1} and P_{t-1} when y_t becomes known. This is done by means of well known regression results.

The set of updating relations of $t + 1$ starting from t, the Kalman filter, is given by:

$$v_t = y_t - a_t, \ F_t = P_t + \sigma_\varepsilon^2, \ K_t = P_t/F_t \tag{54}$$

$$a_{t+1} = a_t K_t v_t, P_{t+1} = P_t(1 - K_t) + \sigma_\eta^2 \tag{55}$$

for $t = 1, ..., n$. Note that a_1 and P_1 are assumed to be known and P_t depends solely on σ_ε^2 and σ_η^2 and not on Y_{t-1}.

The Kalman filter residuals, $v_t = y_t - a_t$, and their variance F_t are respectively the

21

forecast errors and the one-step-ahead error variance of y_t given Y_{t-1}. The forecast errors are known as innovations as they represent the new part of y_t which cannot be predicted from the past $t = 1, ..., n$. Define the state estimation error as being

$$x_t = \alpha_t - a_t, \ Var(x_t) = P_t \tag{56}$$

It follows from the Kalman filter that

$$v_t = x_t + \varepsilon_t, \ x_{t+1} = L_t x_t + \eta_t - K_t \varepsilon_t, \ L_t = 1 - K_t = \sigma_\varepsilon^2 / F_t \tag{57}$$

Analogously, these relations are valid for the local level model Eq. 52 and 53, $t = 1, ..., n$ with $x_1 = \alpha_1 - a_1$.

Now consider the estimation of $\alpha_1, ..., \alpha_n$ given the entire sample, Y_n. As long as all distributions are normally distributed, the conditional density of α_t given $y = (y_1, ..., y_n)$ is $N(\widehat{\alpha}, V_t)$, where $\widehat{\alpha}_t = E(\alpha_t \mid y)$ and $V_t = (\alpha_t \mid y)$. $\widehat{\alpha}_t$ is the smoothed state and V_t is the variance of the smoothed state, and the operation of computing $\widehat{\alpha}_1, ..., \widehat{\alpha}_n$ is known as smoothing.

The forecast errors $v_1, ..., v_n$ are mutually independent and are a linear transformation of $y_1, ..., y_n$ and the $v_i's$ are independent of $y_1, ..., y_{t-1}$, with mean zero. Furthermore, when $y_1, ..., y_n$ are fixed, Y_{t-1} and $v_1, ..., v_n$ are fixed and vice-versa. Using the properties of the multivariate normal regression theory we obtain:

$$\widehat{\alpha}_t = E(\alpha_t \mid y) = a_t + \sum_{j=t}^{n} Cov(\alpha_t, v_j) F_j^{-1} v_j \tag{58}$$

We arrive at the recursions of the smoothed state

$$r_{t-1} = F_t^{-1} v_t + L_t r_t, \ \widehat{\alpha}_t = a_t + P_t r_{t-1}, \ t = n, ..., 1 \tag{59}$$

with $r_n = 0$, for $t = n, n_{-1}, ..., 1$ and

$$r_{t-1} = \frac{v_t}{F_t} + L_t \frac{v_{t+1}}{F_{t+1}} + L_t L_{t+1} \frac{v_{t+2}}{F_{t+2}} + L_t L_{t+1}...L_{n-1} \frac{v_n}{F_n} \tag{60}$$

is a weighted sum of innovations after $t - 1$. Note that the smoothed state is computed by backward recursion.

The variance of the smoothed states can be computed by backward recursion procedure given by

$$N_{t-1} = F_t^{-1} + L_t^2 N_t, \ V_t = P_t - P_t^2 N_{t-1}, \ t = n, ..., 1 \tag{61}$$

with $N_n = 0$ and

$$N_{t-1} = \frac{1}{F_t} + L_t^2 \frac{1}{F_{t+1}} + L_t^2 L_{t+1}^2 \frac{1}{F_{t-2}} + \dots + L_t^2 L_{t-1}^2 \dots L_{n-1}^2 \frac{1}{F_n} \tag{62}$$

is a weighted sum of the inverted variances of innovations after $t - 1$. From the above recursions we have that $Var(r_t) = N_t$ as the forecast errors are independent. With $\widehat{\alpha}_t's$ squared roots $V_t^{1/2}$ one can construct confidence intervals for α_t, for $t = 1, \dots, n$.

Consider now the computations of the smoothed errors of observations $\widehat{\varepsilon}_t = E(\varepsilon_t \mid y) = y_t - \widehat{\alpha}_t$ and the smoothed disturbances of the states $\widehat{\eta}_t = E(\eta_t \mid y) = \widehat{\alpha}_{t+1} - \widehat{\alpha}_t$, together with their respective variances. The smoothed disturbances of the observations are computed as

$$\widehat{\varepsilon}_t = \sigma_\varepsilon^2 u_t, \ u_t = F_t^{-1} v_t - K_t r_t, \ t = n, \dots, 1 \tag{63}$$

and its smoothed variance given by

$$Var(\varepsilon_t \mid y) = \sigma_\varepsilon^2 - \sigma_\varepsilon^4 D_t, \ D_t = F_t^{-1} + K_t^2 N_t, \ t = n, \dots, 1 \tag{64}$$

where the recursions of N_t are given by Eq. 61 and 62. Given that v_t is independent of r_t and $Var(r_t) = N_t$ we have that $Var(u_t) = D_t$.

The state smoothed error is computed by

$$\widehat{\eta}_t = \sigma_\eta^2 r_t, \ t = n, \dots, 1 \tag{65}$$

where the r_t recursion is given by Eq. 59. Its smoothed variance is given by:

$$Var(\eta_t \mid y) = \sigma_\eta^2 - \sigma_\eta^4 N_t, \ t = n, \dots, 1 \tag{66}$$

where the N_t recursions are given by Eq. 61 and 62. As $Var(r_t) = N_t$, we have that $Var(\widehat{\eta}_t) = \sigma_\eta^4 N$. These results are interesting as they provide an interpretation for the values of r_t and N_t; these are respectively the smoothed estimates of $\eta_t = \alpha_{t+1} - \alpha$ and of its non-conditional variance.

4.3 Initialization

Assume at first that the distribution of the initial state α_1 is $N(a_1, P_1)$, where a_1 and P_1 are known. Now consider starting the filter Eq. 54 and 55 when nothing is known ex-ante about the distribution of α_1, which in practice is the usual situation. In that case, we represent α_1 as having a diffuse prior density, that is, we fix α_1 to an arbitrary value and let $P_1 \to \infty$. From Eq. 54 and 55 we have:

$$v_1 = y_1 - a_1, \ F_1 = P_1 + \sigma_\varepsilon^2 \tag{67}$$

and substituting a_2 and P_2 in (Eq. 54 and 55) it follows:

$$P_2 = \frac{P_1}{P_1 + \sigma_\varepsilon^2}\sigma_\varepsilon^2 + \sigma_\eta^2 \tag{68}$$

Making $P_1 \to \infty$ we obtain $a_2 = y_1$, $P_2 = \sigma_\varepsilon^2 + \sigma_\eta^2$ and then we proceed as usual with the Kalman filter in Eq. 54 and 55 for $t = 2, ..., n$. This procedure is known as diffuse initialization of the Kalman filter and the resulting filter is called the diffuse filter. Note that the same values of a_1 and P_1 for $t = 2, ..., n$ can be obtained by taking y_1 as fixed and taking $\alpha_1 \sim N(0, \sigma_\varepsilon^2)$.(Durbin and Koopman, 2004).

It is necessary to take into account the diffuse distribution of the initial state α_1 in the smoothing recursions. From the previous paragraphs we noted that the filtering equations "aren't" affected by making $P_1 \to \infty$. Therefore, the state equations and the smoothing disturbances aren't affected as well for $t = 2, ..., n$, as they depend solely on the results from the Kalman filter. From Eq. 59 the smoothed mean of the state α_1 when $P_1 \to \infty$ is given by $\widehat{\alpha}_1 = a_1 + v_1 + \sigma_\varepsilon^2 r_1$ and substituting v_1 one obtains $\widehat{\alpha}_1 = y_1 + \sigma_\varepsilon^2 r_1$.

The conditional smoothed variance of state α_1, given y when $P_1 \to \infty$ is equal to $\widehat{\varepsilon}_t = -\sigma_\varepsilon^2 r_1$. Note that r_1 depend on the results of the Kalman filter for $t = 2, ..., n$ and the variance of these disturbances for $t = 1$ by Eq. 63 depends on D_1 and N_1, being D_1 the only affected by $P_1 \to \infty$; hence we obtain $D_1 = N_1$ and then $Var(\widehat{\varepsilon}_t) = \sigma_\varepsilon^4 N_1$. The variance of the smoothed estimates of η_1 remains unchanged with $Var(\widehat{\eta}_t) = \sigma_\varepsilon^4 N_1$.

4.4   Parameters estimation

The model is now defined. The filtering and smoothing algorithms are associated with the state space form and can be conditionally applied to the states and the system of matrices of known errors. The unknown values in these matrices are treated as parameters to be estimated. The next step is to estimate these parameters using maximum likelihood methods.

Consider briefly the issue of adjust the local level model to the data from the point of view of classical inference (Durbin and Koopman, 2004). Consequently, this involves deriving a number of formulae under the hypothesis that the parameters are known, and then substitute them by the maximum likelihood estimates. Parameters in state space models are known as hyperparameters.

The computation of the log-likelihood considers the joint density:

$$p(y_1, ..., y_t) = p(Y_{t-1} = p(y_t \mid Y_{t-1}) \tag{69}$$

$t = 2, ..., n$ and can be expressed as $P(Y) = \prod_{t=1}^{n} p(y_t \mid Y_{t-1})$, where $p(y_1 \mid Y_0) = p(y_1)$.
Now $p(y_t \mid Y_{t-1}) = N(a_t, F_t)$ and $v_t = y_t - a_t$ and taking logs and assuming that both a_1 and P_1 are known, the log-likelihood function is given by

$$\log L = \log p(y) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}(\log F_t + \frac{v_t^2}{F_t}) \tag{70}$$

The exact log-likelihood can be constructed from the Kalman filter Eq. 54 and 55.

Alternatively, the log-likelihood may be derived from the local level model in matrix form, which gives us the following expression

$$\log L = \log p(y) = -\frac{n}{2}\log \mid \Omega \mid -\frac{1}{2}(y - a_1 1)'\Omega^{-1}(y - a_1 1) \tag{71}$$

which follows the multivariate normal distribution $y \sim N(a_1 1, \Omega)^{44}$.

The log-likelihood can then be defined for the diffuse case by keeping all terms finite as we make $P_1 \to \infty$ with y fixed, except for $t = 1$. In order to remove the influence of P_1 when $P_1 \to \infty$, define the diffuse log-likelihood as

$$\log L_d = \lim_{P_1 \to \infty} (\log L + \frac{1}{2}\log P_1) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}(\log F_t + \frac{v_t^2}{F_t}) \tag{72}$$

as long as $F_1/P_1 \to 1$ and $v_1^2/F_1 \to 0$ when $P_1 \to \infty$. Note that v_t and F_t remain finite when $P_1 \to \infty$ for $t = 2, ..., n$.

Given that P_1 does not depend neither on σ_ε^2 nor on σ_η^2, the values of these parameters that maximize the $\log L$ are identical to the ones maximizing $\log L + (1/2)\log P_1$. When $P_1 \to \infty$ these values converge to the values the maximize $\log L_d$ as the first and second derivatives with respect to σ_ε^2 and σ_η^2 are finite and strictly negative. It follows that the maximum likelihood estimators of σ_ε^2 and σ_η^2 obtained by maximizing Eq. 70 converge to the values obtain with the maximization of Eq. 72 when $P_1 \to \infty$ (Durkin and Koopman, 2004).

The unknown parameters σ_ε^2 and σ_η^2 are estimated by numerical maximization of expressions Eq. 70 and 72, taking into account is a_1 and P_1 are known or not. In practice it is more convenient to numerical maximize with respect to the quantities $\psi_\varepsilon = \log \sigma_\varepsilon^2$ and $\psi_\eta = \log \sigma_\eta^2$ (see KHDS, 2000). The optimization procedure is based on the quasi-Newton scheme known

---

[44]Note that $1 = (1, ..., 1)'$.

as BFGS (Durbin and Koopman, 2004).

It is more advantageous to reparameterize the model prior to the maximization in order to reduce the dimension for the numerical iteration. For example, in the local level model, one can define $q = \sigma_\eta^2/\sigma_\varepsilon^2$ and obtain

$$y_t = \alpha_t + \varepsilon_t, \ \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{73}$$

$$\alpha_{t-1} = \alpha_t + \eta_t, \ \eta_t \sim N(0, q\sigma_\varepsilon^2) \tag{74}$$

and estimate the pair σ_ε^2, q instead of σ_ε^2 and σ_η^2. By making $P_t^* = P_t/\sigma_\varepsilon^2$ and $F_t^* = F_t/\sigma_\varepsilon^2$, from Eq. 54 and 55 and the previous section, the diffuse Kalman filter for the local level model is reparameterized as follows

$$v_t = y_t - a_t, \ F_t^* = P_t^* + 1 \tag{75}$$

$$a_{t+1} = a_t + K_t v_t, \ P_t^* = (1 - K_t) + q \tag{76}$$

where $K_t = P_t^*/F_t^* = P_t/F_t$ for $t = 2, ..., n$ and it is initialized with $a_2 = y_1$ and $P_2^* = 1+q$. Note that F_t^* depends on q but not on σ_ε^2. The log-likelihood Eq. 72 becomes:

$$\log L_d = -\frac{n}{2}\log(2\pi) - \frac{n-1}{2}\log\sigma_\varepsilon^2 - \frac{1}{2}\sum_{t=1}^{n}(\log F_t^* + \frac{v_t^2}{\sigma_\varepsilon^2 F_t^*}) \tag{77}$$

Maximizing Eq. 76 with respect to σ_ε^2 given $F_2^*, ..., F_n^*$ we obtain

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{n-1}\sum_{t=2}^{n}\frac{v_t^2}{F_t^*} \tag{78}$$

The value of $\log L_d$ obtained by substituting σ_ε^2 by $\widehat{\sigma}_\varepsilon^2$ in Eq. 76 is called the concentrated diffuse log-likelihood denoted by $\log L_{cd}$ yielding

$$\log L_{cd} = -\frac{n}{2}\log(2\pi) - \frac{n-1}{2} - \frac{n-1}{2}\log\sigma_\varepsilon^2 - \frac{1}{2}\sum_{t=1}^{n}(\log F_t^*) \tag{79}$$

This function is maximized with respect to q by a unidimensional numerical iteration.

4.5 Steady-State

Consider the convergence of the Kalman filter to a steady-state value when $n \to \infty$. This would be the case if P_t converges to a positive values $\overline{P}$. Then we would have $F_t \to \overline{P} + \sigma_\varepsilon^2$ and $K_t \to \overline{P}/(\overline{P} + \sigma_\varepsilon^2)$. In order to check if there exists a steady-state, make $P_{t+1} = P_t = \overline{P}$ in Eq. 54 and 55 and verify than the resulting equation in $\overline{P}$ has a positive solution. The equation itself is

$$\overline{P} = \overline{P}\left(1 - \frac{\overline{P}}{\overline{P} + \sigma_\varepsilon^2}\right) + \sigma_\varepsilon^2 \tag{80}$$

which can be reduced to a quadratic form $x^2 - xh - h = 0$, where $x = \overline{P}/\sigma_\varepsilon^2$ and $h = \sigma_\eta^2/\sigma_\varepsilon^2$, with solution $x = (h + (h^2 + 4h)^{1/2}/2$.

This expression is positive when $h > 0$ which is true for non-trivial models. When it takes negative values, it is inapplicable for $h > 0$. Hence, every non-trivial local level model has a steady-state solution.

The practical advantage of the model having a steady-state solution is that, once it is verified that convergence between P_t and $\overline{P}$ is close, one can stop the computation of F_t and K_t. Therefore, the Kalman filter is reduced to a single relation $a_{t+1} = a_t + \overline{K}v_t$, with $\overline{K} = \overline{P}/(\overline{P} + \sigma_\varepsilon^2)$ and $v_t = y_t - a_t$. This is an important property for more complicated models, where P may be a very large matrix.

5 The Kalman Filter: advantages and disadvantages

At this stage it is relevant to distinguish this filter with the Hodrick-Prescott and Baxter-King ones. The latter two filters are useful tools to analyse business cycles and extract trends. On the other hand, the Kalman filter consists of a set of equations which provide an optimal recursive solution, by means of least squares methodology, to a dynamic linear system.

5.1 Advantages

It avoids the influence of possible structural breaks during the estimation. The recursive estimation starts from an initial sample and it then updates the estimates by successive incorporation of new observations until all data are covered. This means that the most recent estimates are affected by the distant history of the series, which under the presence of struc-

tural breaks could end up biased. This bias may be corrected by sequential estimations[45], however, at the cost of a higher standard error. Hence, the Kalman filter, as the recursive methods, uses the entire history of the series with the advantage that it aims to estimate the stochastic path of the coefficients instead of a deterministic one (which solves any possible estimation bias in the presence of structural breaks).

The Kalman filter uses the least squares method to generate recursively a state estimator at time k, which is linear, unbiased and with minimum variance. The filter is in line with the Gauss-Markov theorem and this gives the Kalman filter a great power to solve a large range of statistical inference problems. The filter distinguishes itself by its ability to predict the state of a model at the past, present and future, even when the precise natural of the modelled system is not known. The dynamic modelling of a system is one of the main characteristics of this filter. Dynamic linear models are models with linear transition from one period to the next, which in fact are able to describe the majority of models commonly used in time series applications.

5.2   Disadvantages

Amongst the disadvantages of the Kalman filter one can refer the need for initial conditions for the mean and variance of the state vector in order to start the algorithm. About the way to determine these initial conditions there is still no consensus. For example, a Bayesian focus of this filter requires a ex-ante set of values for initial coefficients and their respective variances. One way to obtain such information is through the estimation of a similar model as desired but with fixed coefficients for a sample sub-period. On the other hand, it is necessary to specify the variances which according to Doan et al. (1984) should be small and proportional relative to the ones obtained for the initial coefficients.

The application of the Kalman filter, as it is presented in the original document, assumes a large knowledge of probabilistic theory, specifically Gaussian conditional properties of random variables, which may limit its study and application scope. When it is applied to autoregressive models, then the results are conditional on past information of the variable under scrutiny. In this sense, the prognosis with time series represents the strength or inertia that currently characterizes the system and are efficient solely in the short term.

---

[45]Given a time series with T observations, a sequential estimation of a given model is simply a sequence of estimates for a set of samples of size n, such that the first sample contains observations from 1 to n; then, from 2 to $n + 1$; etc., until the last sample covers the range $T - n + 1$ till T (with $n < T$). This implies a sequence of $T - n + 1$ estimators associated to the model under scrutiny. E.g., take $T = 50$ and $n = 20$, then one would end up with 31 sequential estimates. The sequential estimation procedure keeps the sample size constant.

References

Abraham, B. and G.E.P. Box (1978), "Deterministic and forecast-adaptive time dependent models", Applied Statistics 27, 120-130.

Anderson, T.W. (1971), "The Statistical Analysis of Time Series", Wiley.

Box, G.E.P. and G.M. Jenkins (1976), "Time Series Analysis: forecasting and control", Holden Day.

Brockwell, P.J. and R.A. Davis (1991), "Time Series: theory and methods", Springer.

Durbin, J. and S.J. Koopman (2001), "Time Series Analysis by State Space Models", Oxford University Press.

Gauss, K. F. (1809), "Theoria Motus Corporum Celestium".

Hamilton, D. (1994), "Time Series Analysis", Princeton University Press.

Hannan, E.J., R.D. Terrell and N. Tuckwell (1970), "The seasonal adjustment of economic time series", International Economic Review 11:24-52.

Harrison, P.J. and C.F. Stevens (1976), "Bayesian forecasting", Journal of the Royal Statistical Society series B 38:205-247.

Harvey, A.C. (1989), "Forecasting, Structural Time Series and the Kalman Filter", Cambridge University Press.

Harvey, A.C. (1993), "Time Series Model", Harvester Wheatsheaf.

Harvey, A.C. (2000), "Trend Analysis". Mimeo.

Harvey, A.C. and A. Scott (1994), "Seasonality in dynamic regression models", The Economic Journal 104:1324-1345.

Harvey, A.C. and A. Jaeger (1991), "Detrending, stylized facts and the business cycles", Mimeo.

Harvey, A.C. and S.J. Koopman (1992), "Diagnostic checking of unobserved components time series models", Journal of Business and Economic Statistics 10:377-389.

Harvey, A.C. and S.J. Koopman (19) Structural Time Series Models,...

Harvey, A.C., S.J. Koopman and M. Riani (1995), "The modelling and seasonal adjustment of weekly observations", London School of Economics, Discussion Paper EM/95/284.

Jones, R.H. (1993), "Longitudinal Data with Serial Correlation: A state-space approach", Chapman & Hall.

Kalman, R.E. (1960), "A new approach to linear filtering and prediction problems", Trans. ASME, J. Basic Engineering, 82, 94-135.

Kalman, R.E., Bucy, R.S (1961), "New results in linear filtering and prediction theory", Trans. ASME, J. Basic Engineering, 83, 95-107.

Koopman, S.J. (1992), "Diagnostic checking and intra-daily effects in time series models", Doctoral Thesis, Tinbergen Institute research series, 27.

Legendre, A.M. (1805), "Nouvelles Methodes pour la Determination des Orbites des Comets".

Pierce, D.A., M.R. Gruppe and W.P. Clevelland (1984), "Seasonal adjustment of the weekly monetary aggregate: a model-based approach", Journal of Business and Economic Statistics 2:260-270.

Plackett, R.L. (1950), "Some theorems in least squares", Biometrika, 37, 149-157.

Stock, J.H. and M.W. Watson (1988), "Testing for common trends", Journal of the American Statistical Association 83:1097-1107.

Wecker, W.E. and C.F. Ansley (1983), "The signal extraction approach to nonlinear regression and spline smoothing", Journal of the American Statistical Association 78:81-89.