Forecasting with Structural Time Series Models

Tommaso Proietti
Dipartimento di Scienze Statistiche, Università di Udine

In *A Companion to Economic Forecasting*,
M. Clements e D. Hendry (Eds.), 1991,
Blackwell Publishers, Oxford.

1 Introduction

Structural time series models are formulated directly in terms of unobserved components, such as trends, cycles and seasonals, that have a natural interpretation and represent the salient features of the series under investigation.

An explicit link with other approaches such as the ARIMA approach and the regression methodology can usually be made. As far as the former is concerned, linear univariate structural models have a reduced form ARIMA representation, but the latter is subject to restrictions on the parameter space, which play a relevant role for forecasting and signal extraction, providing a sensible way of weighting the available information. Moreover, structural models can easily be extended to handle any frequency of observation (weekly, daily, hourly) and specific features of the series that are difficult to deal with in the ARIMA framework (heteroscedasticity, nonlinearity, non-Gaussianity). From the second standpoint, structural models are set up as regression models in which the explanatory variables are functions of time and the coefficients are allowed to vary over time, thereby encompassing the traditional decomposition of a time series into deterministic components. A thorough presentation of the main ideas and methodological aspects underlying structural time series models is contained in Harvey (1989); other important references are West and Harrison (1997) and Kitagawa and Gersch (1996).

The material presented in this chapter is organised as follows: the next three sections deal with the specifications of time series models respectively for the trend, the cycle, and the seasonal component, and how they can be combined into the main univariate structural models; multivariate extensions are discussed in section 5. The disturbances driving the different components are assumed independent and this can be viewed as an identification restriction. However, models with correlated disturbances can be

specified and they are briefly discussed in section 6. Central to the statistical treatment of structural time series models is the state space representation (sec. 7). The Kalman filter is an essential tool for inferences about the unobserved components and for evaluation of the likelihood function. The algorithm is presented along with the modifications that have to be introduced in the presence of nonstationary components. Section 8 explains how explanatory variables can be included in a structural model. The estimation of the unobserved components based on the full sample of observations is called *smoothing* and is considered in section 9.

Structural time series models are used not only for providing a description of the salient features of the series, but also for forecasting its future values. Forecasting provides the means of projecting the past into the future by attaching suitable weights to the past and current observations of the variable under investigation. It is the topic of section 10, where particular attention is posed on the relation with forecasting with *ad hoc* techniques and ARIMA models. Section 11 presents some non-linear and non-Gaussian extensions. Finally, in section 12 two illustrations of modelling and forecasting with structural time series models are provided.

2 Trend models

The specification of a time series model for the trend component varies according to the features displayed by the series under investigation and any prior knowledge. The most elementary structural model deals with a series whose underlying level changes over time, like in the situation depicted in the first panel of figure 1. The data generating process can be thought of consisting of a trend, μ_t, evolving according to a random walk, with a superimposed irregular component, ϵ_t:

$$\begin{array}{lcll} y_t & = & \mu_t + \epsilon_t, \quad t = 1, 2, \ldots, T, & \epsilon_t \sim \mathsf{NID}(0, \sigma_\epsilon^2) \\ \mu_{t+1} & = & \mu_t + \eta_t, & \eta_t \sim \mathsf{NID}(0, \sigma_\eta^2) \end{array} \tag{1}$$

where NID denotes normally and independently distributed. This is known as the local level model (LLM) and is a straightforward generalisation of the constant level model: $y_t = \mu + \epsilon_t$, arising when $\sigma_\eta^2 = 0$. On the other hand, when $\sigma_\epsilon^2 = 0$, (1) reduces to a pure random walk and the trend coincides with the observations.

The stationary representation of the LLM model is obtained by taking first differences: $\Delta y_t = \eta_{t-1} + \Delta\epsilon_t$. It follows immediately that $\mathsf{E}(\Delta y_t) = 0$, and that the autocovariance function of Δy_t, denoted $c(\tau) = \mathsf{E}(\Delta y_t \Delta y_{t-\tau})$, has non zero values $c(0) = \sigma_\eta^2 + 2\sigma_\epsilon^2$, $c(1) = -\sigma_\epsilon^2$, and $c(\tau) = 0$, $\tau > 1$. Thus, the autocorrelation function, $\varrho(\tau) = c(\tau)/c(0)$, exhibits a cut-off at lag one, with $\varrho(1) = -\sigma_\epsilon^2/(\sigma_\eta^2 + 2\sigma_\epsilon^2)$ taking values in $[0, -1/2]$. The ARIMA, or reduced form, representation corresponding to (1) is $\Delta y_t = (1 + \theta L)\xi_t, \xi_t \sim$

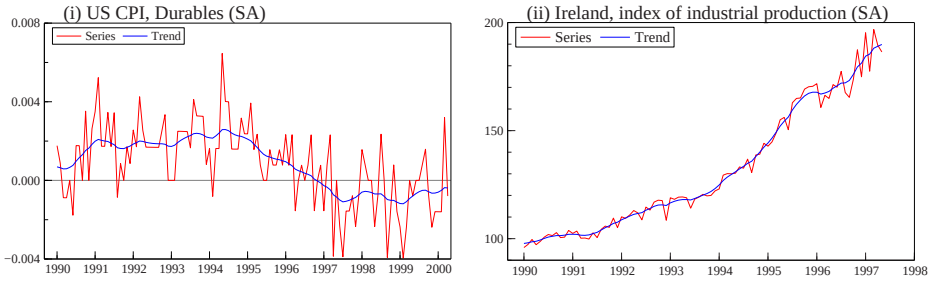

Figure 1: (i) U.S. Consumer Price Index for Durables, Seasonally adjusted, monthly growth rates, January 1990 - April 2000 (Source: Bureau of Labor Statistics). (ii) Index of Industrial Production, Seasonally adjusted, Ireland, January 1990 - July 1997 (Source: OECD Statistical Compedium). The dotted line are the smoothed estimates of the trend, see section 9 for details.

$\mathsf{NID}(0, \sigma^2)$; equating the autocorrelations at lag 1 we can express the MA parameter as a function of the signal to noise ratio, $q = \sigma_\eta^2/\sigma_\epsilon^2$: $\theta = [(q^2 + 4q)^{1/2} - 2 - q]/2$; it can be seen that the MA parameter is constrained to lie in the range [-1,0].

Consider now the situation depicted in panel (ii) of fig. 1, concerning the monthly seasonally adjusted index of industrial production for Ireland. The series displays a steady upward movement, suggesting that we have to bring a slope, or a drift, into the model for the trend. A deterministic linear trend model, $\mu_t = \alpha + \beta t$, is likely to be inadequate, but further flexibility can be introduced. The motivation for the extension is as follows: note that such trend would be generated by the recursive formulae $\mu_{t+1} = \mu_t + \beta_t$ and $\beta_{t+1} = \beta_t$, $t = 1, \ldots, T-1$, respectively for the level and the slope, with starting values $\mu_0 = \alpha$ and $\beta_0 = \beta$; then, we allow for time variation by introducing random disturbances on the right hand side of the recursive formulae.

The resulting model, known as the *local linear trend model* (LLTM), is written as:

$$\begin{array}{rcccccll} y_t & = & \mu_t & + & \epsilon_t, & & \epsilon_t \sim \mathsf{NID}(0, \sigma_\epsilon^2), & t = 1, 2, \ldots, T, \\ \mu_{t+1} & = & \mu_t & + & \beta_t & + & \eta_t, & \eta_t \sim \mathsf{NID}(0, \sigma_\eta^2), \\ \beta_{t+1} & = & & & \beta_t & + & \zeta_t, & \zeta_t \sim \mathsf{NID}(0, \sigma_\zeta^2), \end{array} \tag{2}$$

where η_t, ζ_t, ϵ_t are independent of one another. For $\sigma_\zeta^2 = 0$ the trend reduces to a random walk with constant drift ($\mu_{t+1} = \mu_t + \beta + \eta_t$), whereas for $\sigma_\eta^2 = 0$ the trend is an integrated random walk ($\Delta^2 \mu_{t+1} = \zeta_{t-1}$). The latter is referred to as a *smoothness prior* specification, as the resulting trend varies very smoothly over time (see Kitagawa and Gersch, 1996, ch. 3,4,8, and the

references therein); a special case is the Hodrick and Prescott (1997) filter, which fixes $q_\zeta = \sigma_\zeta^2/\sigma_\epsilon^2$ to predetermined value (eg. $1/1600$ for quarterly data). Finally, when $\sigma_\eta^2 = \sigma_\zeta^2 = 0$ we fall back on the deterministic linear trend.

The time and frequency domain properties of the LLTM can be ascertained from its stationary representation: $\Delta^2 y_t = \Delta\eta_{t-1} + \zeta_{t-2} + \Delta^2\epsilon_t$. The autocovariance function of $\Delta^2 y_t$ takes the values $c(0) = 2\sigma_\eta^2 + \sigma_\zeta^2 + 6\sigma_\epsilon^2$, $c(1) = -\sigma_\eta^2 - 4\sigma_\epsilon^2$, $c(2) = \sigma_\epsilon^2$, and $c(\tau) = 0$ for $\tau > 2$. Hence $\varrho(1)$ and $\varrho(2)$ lie respectively in [-2/3,0] and [0,1/6], and $\varrho(\tau)$ displays the cut-off at lag 2 that is characteristic of an MA(2) process. Therefore, the reduced form is $y_t \sim ARIMA(0,2,2)$, with severe restrictions on the parameter space of the MA parameters.

3 Cyclical models

In economics, the term *business cycle* broadly refers to the recurrent, though not exactly periodic, deviations around the long term path of the series. A model for the cyclical component should be capable of reproducing commonly acknowledged essential features, such as the presence of strong autocorrelation, determining the recurrence and alternation of phases, and the dampening of the fluctuations, or zero long run persistence.

A time series model accounting for these stylised facts can be derived by a stochastic extension of the deterministic cycle model $\psi_t = \alpha\cos\lambda_c t + \alpha^*\sin\lambda_c t$, where λ_c is the angular frequency measured in radians, $\lambda_c \in [0,\pi]$. This defines a perfectly periodic function of time, repeating itself every $\bar{p} = 2\pi/\lambda_c$ time units, where $\bar{p}$ is the period, with constant amplitude $(\alpha^2 + \alpha^{*2})^{1/2}$ and phase $\tan^{-1}(\alpha^*/\alpha)$.

A stochastic cycle can be obtained by letting the coefficients α and α^* follow an AR(1) process with coefficient ρ, $0 \le \rho \le 1$, that is responsible for the dampening of the fluctuations: hence, $\alpha_{t+1} = \rho\alpha_t + \tilde{\kappa}_t$, $\alpha^*_{t+1} = \rho\alpha^*_t + \tilde{\kappa}^*_t$, where $\tilde{\kappa}_t$ and $\tilde{\kappa}^*_t$ are mutually independent NID disturbances with zero mean and common variance σ_κ^2.

Equivalently, recognising that a deterministic cycle can be generated recursively by $\psi_{t+1} = \psi_t\cos\lambda_c + \psi^*_t\sin\lambda_c$ and $\psi^*_{t+1} = -\psi_t\sin\lambda_c + \psi^*_t\cos\lambda_c$, with starting values $\psi_0 = \alpha$ and $\psi^*_0 = \alpha^*$, a stochastic cycle is constructed multiplying the right hand side of these two equations by ρ (*damping factor*), and adding stochastic disturbances in the form of NID sequences, giving:

$$\begin{bmatrix} \psi_{t+1} \\ \psi^*_{t+1} \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_t \\ \psi^*_t \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa^*_t \end{bmatrix}, \tag{3}$$

where again $\kappa_t \sim \mathsf{NID}(0,\sigma_\kappa^2)$ and $\kappa^*_t \sim \mathsf{NID}(0,\sigma_\kappa^2)$, are mutually independent. The reduced form of (3) is the ARMA(2,1) process:

$$(1 - 2\rho\cos\lambda_c L + \rho^2 L^2)\psi_{t+1} = (1 - \rho\cos\lambda_c L)\kappa_t + \rho\sin\lambda_c L\kappa^*_t;$$

when ρ is strictly less than one the cycle is stationary with $\mathsf{E}(\psi_t) = 0$ and $\sigma_\psi^2 = \mathsf{Var}(\psi_t) = \sigma_\kappa^2/(1-\rho^2)$; the autocorrelation at lag τ is $\rho^\tau \cos \lambda_c \tau$. For $\lambda_c \in (0, \pi)$ the roots of the AR polynomial are a pair of complex conjugates with modulus ρ^{-1} and phase λ_c; correspondingly, the spectral density displays a peak at λ_c. When $\lambda_c = 0$, ψ_t collapses to the AR(1) process $\psi_{t+1} = \rho\psi_t + \kappa_t$, whereas in the case $\lambda_c = \pi$, $\psi_{t+1} = -\rho\psi_t + \kappa_t$.

4 Seasonal models

Seasonal fluctuations account for a major part of the variation of a wide spectrum of economic, social and environmental phenomena. Hylleberg (1992, p. 4) defines seasonality as "the systematic, although not necessarily regular, intra-year movement caused by the changes of the weather, the calendar, and timing of decisions". A more operational definition is given by Harvey (1989, p. 301) in terms of prediction, as the "part of the series which, when extrapolated, repeats itself over any one-year time period and averages out to zero over such a time period". There are several specifications of a seasonal component, γ_t, satisfying this requirement; nevertheless, the stochastic process for γ_t complying with the above definition is such that that $S(L)\gamma_t \sim MA(q)$ with $q \leq s-2$, where $S(L) = 1 + L + \cdots + L^{s-1}$ denotes the seasonal summation operator and s the number of seasons in a year (e.g. s is equal to 4 and 12 respectively for quarterly and monthly time series).

In the time domain, a fixed seasonal pattern is modelled as: $\gamma_t = \boldsymbol{z}_t'\boldsymbol{\chi}$, where $\boldsymbol{z}_t' = [D_{1t}, \ldots, D_{st}]$ is a vector containing the values of s seasonal dummies, D_{jt}, taking value 1 in season j and 0 otherwise, and $\boldsymbol{\chi}$ is an $s \times 1$ vector containing the effects associated with the different seasons, which are restricted to sum up to zero in order to enhance identifiability when the level of the series in nonzero. Denoting $\boldsymbol{i}_s = [1, 1, \ldots, 1]'$, an $s \times 1$ vector of ones, the zero sum constraint is expressed as $\boldsymbol{i}_s'\boldsymbol{\chi} = 0$ and ensures that $S(L)\gamma_t = 0$.

A simple way of allowing the seasonal pattern to evolve over time is letting the sum of the seasonal effect be equal to a random disturbance term, that is $S(L)\gamma_t = \omega_t$, $\omega_t \sim \mathsf{NID}(0, \sigma_\omega^2)$; this is referred to as the the *dummy seasonal* model. A richer class of models of stochastic seasonality is derived letting the coefficients $\boldsymbol{\chi}$ change over time according to a multivariate random walk:

$$\gamma_t = \boldsymbol{z}_t'\boldsymbol{\chi}_t, \quad \boldsymbol{\chi}_{t+1} = \boldsymbol{\chi}_t + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathsf{NID}(\boldsymbol{0}, \boldsymbol{\Omega}) \tag{4}$$

The stochastic counterpart of the zero sum constraint is enforced by $\boldsymbol{i}_s'\boldsymbol{\Omega} = \boldsymbol{0}_s'$. This implies $\boldsymbol{i}_s'\boldsymbol{\chi}_t = \boldsymbol{i}_s'\boldsymbol{\chi}_{t-1}$, which for $\boldsymbol{i}_s'\boldsymbol{\chi}_0 = 0$ implies in turn that $S(L)\gamma_t$ is a stationary zero mean process. As a matter of fact, by repeated substitution from (4) it is possible to show that $S(L)\gamma_t \sim MA(q)$, with $q \leq$

$s-2$. A special case of (4) is the Harrison and Stevens (1976) model, which arises for the specification $\boldsymbol{\Omega} = \sigma^2_{\omega^*}[\boldsymbol{I}_s - (1/s)\boldsymbol{i}_s\boldsymbol{i}_s']$, and is such that $S(L)\gamma_t \sim MA(s-2)$. As $\boldsymbol{\Omega}$ has rank $s-1$, one of the elements of $\boldsymbol{\chi}_t$ is redundant and can be dropped.

Representation (4) is amenable since the disturbances $\boldsymbol{\omega}_t$ are season-specific. In Proietti (1998) it is argued that it can account for seasonal heteroscedasticity in a simple fashion. Moreover, the time-varying periodic cubic spline specification, adopted by Harvey and Koopman (1993) to parsimoniously model daily and weekly seasonality, arises imposing an appropriate reduced rank structure on $\boldsymbol{\Omega}$.

In the frequency domain, a fixed seasonal pattern is modelled by the sum of $[s/2]$ cycles defined at the seasonal frequencies $\lambda_j = 2\pi j/s$, $j = 1, 2, \ldots, [s/2]$, where $[s/2] = s/2$ for s even and $(s-1)/2$ if s is odd:

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{jt}, \qquad \gamma_{jt} = \alpha_j \cos\lambda_j + \alpha_j^* \sin\lambda_j.$$

When s is even the sine term disappears for $j = s/2$, so the number of trigonometric terms is always $s-1$.

A possible stochastic extension of the trigonometric seasonal model is such that the seasonal effect at time t arises from the combination of $[s/2]$ stochastic cycles formulated as in (3), setting $\rho = 1$ to allow for a persistent pattern:

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{jt}, \qquad \begin{bmatrix} \gamma_{j,t+1} \\ \gamma^*_{j,t+1} \end{bmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t} \\ \gamma^*_{j,t} \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega^*_{j,t} \end{bmatrix}, \quad (5)$$

For s even, the last component, defined at $\lambda_{s/2} = \pi$, reduces to $\gamma_{\frac{s}{2},t+1} = -\gamma_{\frac{s}{2},t} + \omega_{\frac{s}{2},t}$. The disturbances ω_{jt} and ω^*_{jt} are assumed to be normally and independently distributed with common variance $\sigma^2_{\omega_j}$; it is often assumed that the latter is constant across j: $\sigma^2_{\omega_j} = \sigma^2_\omega$. In the latter case the trigonometric model can be shown to be equivalent to the Harrison and Stevens model with $\sigma^2_{\omega^*} = (s/2)\sigma^2_\omega$; however, when s is even the equivalence holds if $\sigma^2_{\omega_j} = \sigma^2_\omega$, for $j = 1, \ldots, (s-1)/2$, and $\sigma^2_{\omega_{s/2}} = \sigma^2_\omega/2$. The reduced form is $S(L)\gamma_t \sim$MA$(s-2)$, as can be established by aggregating the $[s/2]$ nonstationary ARMA(2,1) reduced forms for each γ_{jt}.

A comparison of the various representations of a seasonal component and a discussion of the implications for forecasting are given in Proietti (2000). The model $y_t = \mu_t + \gamma_t + \epsilon_t$, where μ_t is the local linear trend in (2) and γ_t has one of the specifications above, is referred to as the *basic structural model* (BSM). The terminology alludes to the fact that it is successfully fitted to economic time series for which the *airline model*, $\Delta\Delta_s y_t = (1+\theta L)(1+\Theta L^s)\xi_t$, with negative MA parameters, is appropriate.

5 Multivariate models

Let us suppose that time series observations are available on a cross section of N units and are gathered in the vector $\boldsymbol{y}_t = [y_{1t}, \ldots, y_{it}, \ldots, y_{Nt}]'$. A multivariate structural time series model for $\boldsymbol{y}_t$ is formulated so that each of the individual time series follows a univariate model, say $y_{it} = \mu_{it} + \psi_{it} + \gamma_{it} + \epsilon_{it}, i = 1, \ldots, N$, and is linked to the other series via the contemporaneous correlation among the disturbances driving the components. This specification is called a *Seemingly Unrelated Time Series Equation* (SUTSE) system; it is tailored for variables that are subject to the same overall environment and are not causally related, although cause and effect relationships can be introduced, for instance, by modelling the short run dynamics as a stationary vector autoregression.

The most relevant specification issues raised by the multivariate framework are discussed with reference to the multivariate LLM:

$$\begin{array}{lcll} \boldsymbol{y}_t & = & \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \ldots, T, & \boldsymbol{\epsilon}_t \sim \mathsf{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon) \\ \boldsymbol{\mu}_{t+1} & = & \boldsymbol{\mu}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t \sim \mathsf{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \end{array} \tag{6}$$

where $\boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{\Sigma}_\eta$ are $N \times N$ non-negative definite matrices, and $\mathsf{E}(\boldsymbol{\epsilon}_t \boldsymbol{\eta}_{t-\tau}) = \mathbf{0}, \forall \tau$.

A first special case arises when the covariance matrices of the disturbances are proportional, that is there exists a scalar q such that $\boldsymbol{\Sigma}_\eta = q\boldsymbol{\Sigma}_\epsilon$. This restriction is relevant for it leads to a parsimonius model and can be tested as in Fernandez and Harvey (1990); it implies that each component series and any linear combination thereof follow the same time series process.

Common components arise when the covariance matrices of the relevant disturbances have reduced rank. When $\text{rank}(\boldsymbol{\Sigma}_\eta) = K < N$ in (6) we can write $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Theta}\boldsymbol{D}_\eta\boldsymbol{\Theta}'$, where $\boldsymbol{\Theta}$ is $N \times K$ with elements $\Theta_{ij} = 0, j > i$, $\Theta_{ii} = 1$, and $\boldsymbol{D}_\eta$ is a diagonal $K \times K$ matrix. The trend component can be then reformulated in terms of a set of K common trends, giving

$$\begin{array}{lcll} \boldsymbol{y}_t & = & \boldsymbol{\Theta}\boldsymbol{\mu}_t^\dagger + \boldsymbol{\mu}_\theta + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \ldots, T, & \boldsymbol{\epsilon}_t \sim \mathsf{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon) \\ \boldsymbol{\mu}_{t+1}^\dagger & = & \boldsymbol{\mu}_t^\dagger + \boldsymbol{\eta}_t^\dagger, & \boldsymbol{\eta}_t^\dagger \sim \mathsf{NID}(\mathbf{0}, \boldsymbol{D}_\eta) \end{array} \tag{7}$$

where $\boldsymbol{\mu}_\theta = [\mathbf{0}_K' \;\; \bar{\boldsymbol{\mu}}']'$ and $\bar{\boldsymbol{\mu}}$ is an $(N-K) \times 1$ vector. The matrix $\boldsymbol{\Theta}$ contains the standardised factor loadings, which can be rotated to enhance interpretation. The important feature of the common trends model is that although each component of $\boldsymbol{y}_t$ is integrated of order 1, $N-K$ linear combinations of $\boldsymbol{y}_t$ are stationary, so the system is cointegrated (Engle and Granger, 1987). In other words, there exist an $(N-K) \times N$ matrix $\boldsymbol{C}$, with the property $\boldsymbol{C}\boldsymbol{\Theta} = \mathbf{0}$, so that $\boldsymbol{C}\boldsymbol{y}_t$ is stationary. Further details and extensions to common cycles and common seasonals, along with economic applications, can be found in Harvey and Koopman (1997).

Finally, a *dynamic error component model* is obtained specifying $\mathbf{\Sigma}_\epsilon = \sigma^2_\epsilon \boldsymbol{ii}' + \sigma^2_{\epsilon^*} \boldsymbol{I}$, $\mathbf{\Sigma}_\eta = \sigma^2_\eta \boldsymbol{ii}' + \sigma^2_{\eta^*} \boldsymbol{I}$, where $\boldsymbol{i}$ is an $N \times 1$ vector of 1s. Hence, for each component, the correlation between the disturbances in any two units is the same and is accounted for by a time specific common effect, whereas unit specific effects are mutually independent. Marshall (1992) shows that the model can be trasformed to a system of time series equations consisting of a LLM for the cross sectional mean, $\bar{y}_t$, and a homogeneous LLM for the $N-1$ series in deviation form $y_{it} - \bar{y}_t, i = 1, \ldots, N-1$.

6 Correlated disturbances

The reader may wonder why the specification of structural models assumes independent disturbances. Actually, this is a restriction which has to imposed to achieve identifiability; take for instance the LLM (1) and assume $\mathsf{E}(\epsilon_t \eta_{t-\tau}) = \sigma_{\epsilon\eta}$ if $\tau = 0$, and zero otherwise; matching the nonzero autocovariances of Δy_t yields a (nonlinear) system of two equations in three unknowns $(\sigma^2_\epsilon, \sigma^2_\eta, \sigma_{\epsilon\eta})$ which has infinite solutions. Imposing $\sigma_{\epsilon\eta} = 0$ gives a unique solution, provided $c(1)$ is negative. An identifiable model can also be obtained by assuming perfectly correlated disturbances as in Snyder (1985, p. 273), who poses $\eta_t = \alpha \epsilon_t$.

Thus, structural time series models usually achieve identification assuming that the disturbances driving the components are independent; this is known to place severe constrains on the ARIMA reduced form parameter space (for instance, in the LLM example above, θ is restricted between -1 and 0, so that half of the MA parameter range is admissible). It is a matter of debate whether they are overly restrictive and whether they can be meaningfully relaxed assuming correlated disturbances.

Models with correlated components have been considered by Godolphin (1976) with the explicit intent of extending the parameter range yielding decomposable models. Snyder (1985) and Ord, Koehler and Snyder (1997) propose unobserved components models with only one source of random disturbances, arguing that models with multiple disturbances are unnecessarily complex. Another very popular result, the Beveridge and Nelson (1981) decomposition, is formulated with perfectly correlated disturbances, and is commonly viewed as providing a structural interpretation to *any* ARIMA model (see also section 10).

On the contrary, West and Harrison (1997, sec. 7.3.4) argue against correlated components on the grounds of parsimony in parameterising a forecasting model, whereas Harvey and Koopman (2000), looking at the implications on the weighting patterns for trend extraction by a local level model, conclude that the scope for such extensions is very limited.

Notice also that the unobserved components have been specified in future form; often (Harvey, 1989, West and Harrison, 1996) the contempo-

raneous form of the model is used, eg. in the LLM case $y_t = \mu_t^* + \epsilon_t^*$, $\mu_t^* = \mu_{t-1}^* + \eta_t^*$, with ϵ_t^* and η_t^* mutually and serially independent. Setting $\mu_t = \mu_{t-1}^*$, we can rewrite the model in future form as $y_t = \mu_t + \epsilon_t$, $\mu_{t+1} = \mu_t + \eta_t$, where the disturbances $\epsilon_t = \eta_t^* + \epsilon_t^*$ and $\eta_t = \eta_t^*$ are now correlated.

7 Statistical treatment

The structural time series models considered in the previous sections are special cases of state space models (SSM). A SSM consists of a *measurement equation* and a *transition equation*: the former relates the $N \times 1$ vector time series $\boldsymbol{y}_t$ to an $m \times 1$ vector of unobservable components or state vector, $\boldsymbol{\alpha}_t$:

$$\boldsymbol{y}_t = \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{G}_t \boldsymbol{\varepsilon}_t, \qquad t = 1, 2, \ldots, T, \tag{8}$$

where $\boldsymbol{Z}_t$ is an $N \times m$ matrix, $\boldsymbol{G}_t$ is $N \times g$ and $\boldsymbol{\varepsilon}_t$ is a $g \times 1$ vector of random disturbances that we assume $\mathsf{NID}(\boldsymbol{0}, \boldsymbol{I}_g)$. The *transition equation* is a dynamic linear model for the states $\boldsymbol{\alpha}_t$, taking the form of a first order vector autoregression:

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{T}_t \boldsymbol{\alpha}_t + \boldsymbol{H}_t \boldsymbol{\varepsilon}_t, \tag{9}$$

where $\boldsymbol{T}_t$ and $\boldsymbol{H}_t$ are $m \times m$ and $m \times g$ matrices, respectively.

Example 7.1: The state space representation of the LLTM (2) has $\boldsymbol{\alpha}_t = [\mu_t, \ \beta_t]'$, $\boldsymbol{\varepsilon}_t = [\epsilon_t/\sigma_\epsilon \ \ \eta_t/\sigma_\eta \ \ \zeta_t/\sigma_\zeta]'$ (hence $\boldsymbol{\varepsilon}_t$ contains the standardised disturbances), and

$$\boldsymbol{Z}_t = [1 \ \ 0], \ \ \boldsymbol{G}_t = [\sigma_\epsilon, \ \ 0 \ \ 0] \ \ \boldsymbol{T}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \boldsymbol{H}_t = \begin{bmatrix} 0 & \sigma_\eta & 0 \\ 0 & 0 & \sigma_\zeta \end{bmatrix}.$$

The system matrices, $\boldsymbol{Z}_t, \boldsymbol{G}_t, \boldsymbol{T}_t$, and $\boldsymbol{H}_t$, are non stochastic, i.e. they are allowed to vary over time, but in a deterministic fashion, and are functionally related to a set of parameters, $\boldsymbol{\theta}$, which usually will have to be estimated. The SSM is *time invariant* if the system matrices are constant, that is $\boldsymbol{Z}_t = \boldsymbol{Z}, \boldsymbol{G}_t = \boldsymbol{G}, \boldsymbol{T}_t = \boldsymbol{T}$ and $\boldsymbol{H}_t = \boldsymbol{H}$, as occurs in Example 7.1. Moreover, structural time series models are usually specified in a way that the measurement and transition equation disturbances are independent, i.e. $\boldsymbol{H}_t \boldsymbol{G}_t' = \boldsymbol{0}$; nevertheless, in the derivation of the Kalman filter and smoother we will use the general representation with $\boldsymbol{H}_t \boldsymbol{G}_t'$ not necessarily equal to a zero matrix, since this representation encompasses other structural models with correlated disturbances (see section 6).

The SSM is completed by the specification of initial conditions, concerning the distribution of $\boldsymbol{\alpha}_1$: this turns out to be a relevant issue when nonstationary components are present (see section 7.2). When the system is time-invariant and $\boldsymbol{\alpha}_t$ is stationary (an instance is provided by the cycle

plus irregular model, $y_t = \psi_t + \epsilon_t$), and the stochastic process governing $\boldsymbol{\alpha}_t$ started in the indefinite past, then $\mathsf{E}(\boldsymbol{\alpha}_1) = \mathbf{0}$ and $\mathsf{Var}(\boldsymbol{\alpha}_1) = \boldsymbol{P}$, satisfying the matrix equation $\boldsymbol{P} = \boldsymbol{T}\boldsymbol{P}\boldsymbol{T}' + \boldsymbol{H}\boldsymbol{H}'$. Hence, the initial conditions are provided by the unconditional mean and covariance matrix of the state vector. A time invariant SSM is stationary if the eigenvalues of the transition matrix, $\boldsymbol{T}$, are inside the unit circle.

7.1 The Kalman filter and the prediction error decomposition

The Kalman filter (KF) is a fundamental algorithm for the statistical treatment of a SSM. Under the Gaussian assumption it produces the minimum mean square estimator (MMSE) of the state vector along with its mean square error (MSE) matrix, conditional on past information; this is used to build the one-step-ahead predictor of $\boldsymbol{y}_t$ and its mean square error matrix. Due to the independence of the one-step-ahead prediction errors, the likelihood can be evaluated via the prediction error decomposition.

Denoting $\boldsymbol{Y}_t = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$ the information set available at time t, $\tilde{\boldsymbol{\alpha}}_{t|t-1} = \mathsf{E}(\boldsymbol{\alpha}_t|\boldsymbol{Y}_{t-1})$ and $\boldsymbol{P}_{t|t-1} = \mathsf{E}[(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{\alpha}}_{t|t-1})(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{\alpha}}_{t|t-1})'|\boldsymbol{Y}_{t-1}]$, the KF for a standard state space model with initial conditions $\boldsymbol{\alpha}_1 \sim \mathsf{N}(\tilde{\boldsymbol{\alpha}}_{1|0}, \boldsymbol{P}_{1|0})$, where $\tilde{\boldsymbol{\alpha}}_{1|0}$ and $\boldsymbol{P}_{1|0}$ are known and finite, consists of the following recursive formulae for $t = 1, \ldots, T$:

$$\begin{array}{rclrcl} \boldsymbol{\nu}_t & = & \boldsymbol{y}_t - \boldsymbol{Z}_t\tilde{\boldsymbol{\alpha}}_{t|t-1}, & \boldsymbol{F}_t & = & \boldsymbol{Z}_t\boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{G}_t\boldsymbol{G}_t', \\ & & & \boldsymbol{K}_t & = & (\boldsymbol{T}_t\boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{H}_t\boldsymbol{G}_t')\boldsymbol{F}_t^{-1}, \\ \tilde{\boldsymbol{\alpha}}_{t+1|t} & = & \boldsymbol{T}_t\tilde{\boldsymbol{\alpha}}_{t|t-1} + \boldsymbol{K}_t\boldsymbol{\nu}_t, & \boldsymbol{P}_{t+1|t} & = & \boldsymbol{T}_t\boldsymbol{P}_{t|t-1}\boldsymbol{T}_t' + \boldsymbol{H}_t\boldsymbol{H}_t' - \boldsymbol{K}_t\boldsymbol{F}_t\boldsymbol{K}_t'; \end{array} \tag{10}$$

as $\tilde{\boldsymbol{y}}_{t|t-1} = \mathsf{E}(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1}) = \boldsymbol{Z}_t\tilde{\boldsymbol{\alpha}}_{t|t-1}$, $\boldsymbol{\nu}_t$ represent the one-step-ahead prediction error, also known as the *innovation* at time t because it represents the part of $\boldsymbol{y}_t$ that cannot be predicted from the past, and $\boldsymbol{F}_t$ is its variance matrix.

A proof of the KF is found in Anderson and Moore (1979) and in the appendix. When the Gaussianity assumption is removed the KF still yields the minimum mean square linear estimator (MMSLE) of the state vector.

For a time-invariant model, the recursions for $\boldsymbol{F}_t$, $\boldsymbol{K}_t$ and $\boldsymbol{P}_{t+1|t}$ become redundant when the KF has reached a steady state, which occurs if, for some t, $\boldsymbol{P}_{t|t-1} = \boldsymbol{P}$. The conditions under which $\lim_{t\to\infty} \boldsymbol{P}_{t|t-1} = \boldsymbol{P}$ are given in Harvey (1989, sec. 3.3.3 and 3.3.4), and, when they are met, the matrix $\boldsymbol{P}$ is the solution of the Riccati equation

$$\boldsymbol{P} = \boldsymbol{T}\boldsymbol{P}\boldsymbol{T}' + \boldsymbol{H}\boldsymbol{H}' - \boldsymbol{K}\boldsymbol{F}\boldsymbol{K}', \tag{11}$$

with $\boldsymbol{K} = (\boldsymbol{T}\boldsymbol{P}\boldsymbol{Z}' + \boldsymbol{H}\boldsymbol{G}')\boldsymbol{F}^{-1}$ and $\boldsymbol{F} = \boldsymbol{Z}\boldsymbol{P}\boldsymbol{Z}' + \boldsymbol{G}\boldsymbol{G}'$.

The updated estimates of the state vector, $\tilde{\boldsymbol{\alpha}}_{t|t} = \mathsf{E}(\boldsymbol{\alpha}_t|\boldsymbol{Y}_t)$, and their

MSE matrix are:

$$\tilde{\boldsymbol{\alpha}}_{t|t} = \tilde{\boldsymbol{\alpha}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t, \quad \boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{Z}_t\boldsymbol{P}_{t|t-1}. \tag{12}$$

Also, when $\boldsymbol{H}_t\boldsymbol{G}_t' = \boldsymbol{0}$, the KF recursions for the states can be broken up into an updating step, (12), followed by a prediction step:

$$\tilde{\boldsymbol{\alpha}}_{t+1|t} = \boldsymbol{T}_t\tilde{\boldsymbol{\alpha}}_{t|t}, \quad \boldsymbol{P}_{t+1|t} = \boldsymbol{T}_t\boldsymbol{P}_{t|t}\boldsymbol{T}_t' + \boldsymbol{H}_t\boldsymbol{H}_t' \tag{13}$$

As the KF filter provides $\boldsymbol{y}_t|\boldsymbol{Y}_{t-1} \sim \mathsf{NID}(\tilde{\boldsymbol{y}}_{t|t-1}, \boldsymbol{F}_t)$, it enables the likelihood function to be written in prediction error decomposition form. Apart from a constant term the log-likelihood of the observations is computed as follows:

$$\mathsf{L}(\boldsymbol{Y}_T;\boldsymbol{\theta}) = \sum_{t=1}^{T}\mathsf{L}(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1};\boldsymbol{\theta}) = -\frac{1}{2}\left(\sum_{t=1}^{T}\ln|\boldsymbol{F}_t| + \sum_{t=1}^{T}\boldsymbol{\nu}_t'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t\right). \tag{14}$$

The likelihood function can be maximised numerically by a quasi-Newton optimisation routine. Analytical expressions for the score vector, with respect to the parameters in $\boldsymbol{G}_t$ and $\boldsymbol{H}_t$ (Koopman and Shepard, 1992), and for the information matrix (Harvey, 1989, pp. 140-143) are available. The dimension of the problem can be reduced by concentrating one of the variance parameters out of the likelihood function: if we write $\boldsymbol{\theta} = [\boldsymbol{\theta}^{*'}, \sigma^2]'$, and $\boldsymbol{G}_t = \sigma\boldsymbol{G}_t^*, \boldsymbol{H}_t = \sigma\boldsymbol{H}_t^*, \boldsymbol{\alpha}_1 \sim \mathsf{N}(\tilde{\boldsymbol{\alpha}}_{1|0}, \sigma^2\boldsymbol{P}_{1|0})$ the recursions (10), run with $\boldsymbol{G}_t$ and $\boldsymbol{H}_t$ replaced by $\boldsymbol{G}_t^*$ and $\boldsymbol{H}_t^*$, yield $\boldsymbol{y}_t|\boldsymbol{Y}_{t-1} \sim \mathsf{NID}(\tilde{\boldsymbol{y}}_{t|t-1}, \sigma^2\boldsymbol{F}_t)$. Maximising the corresponding likelihood with respect to σ^2 gives $\hat{\sigma}^2 = \sum\boldsymbol{\nu}_t'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t/(NT)$ and the concentrated likelihood $\mathsf{L}_{\sigma^2}(\boldsymbol{Y}_T;\boldsymbol{\theta}^*) = -0.5[NT(\ln\hat{\sigma}^2 + 1) + \sum\ln|\boldsymbol{F}_t|]$.

7.2 Initial conditions and nonstationary models*

When there are d nonstationary elements in the state vector, we write in general $\boldsymbol{\alpha}_1 = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{D}\boldsymbol{\delta}$, where $\boldsymbol{a}$ and $\boldsymbol{B}$ are respectively an $m \times 1$ known vector and an $m \times (m-d)$ known matrix that may depend on $\boldsymbol{\theta}$, and $\boldsymbol{D}$ is a $m \times d$ selection matrix assigning the appropriate elements of $\boldsymbol{\delta}$ to the states. The vectors $\boldsymbol{\eta} \sim \mathsf{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ and $\boldsymbol{\delta}$ are used to initialise the stationary and nonstationary elements of $\boldsymbol{\alpha}_t$ and are mutually independent and independent of $\boldsymbol{\varepsilon}_t, \forall t$. Two assumptions can be made: (i) $\boldsymbol{\delta}$ is considered as a fixed, but unknown vector; (ii) $\boldsymbol{\delta}$ is a diffuse random vector, i.e. it has an improper distribution with a mean of zero and an arbitrarily large variance matrix: $\boldsymbol{\delta} \sim \mathsf{N}(\boldsymbol{0}, \kappa\boldsymbol{I}_d)$, $\kappa \to \infty$. The first assumption is suitable if it is deemed that the transition process (9) governing the states has started at time $t = 1$; the second if the process has started in the remote past. For instance, if one of the state components is a random walk, as for LLM model,

*This section is more technical and can be omitted on first reading.

then its starting value is the cumulative sum of infinite NID disturbances, so its variance will go to infinity. The use of a diffuse prior amounts to leaving the distribution of initial conditions unspecified.

Example 7.2: The model $y_t = \mu_t + \gamma_t + \psi_t + \epsilon_t$, for quarterly data, with the components specified as in (2), (3) and (5), has $\boldsymbol{\alpha}_t = [\mu_t \;\; \beta_t \;\; \gamma_{1t} \;\; \gamma^*_{1t} \;\; \gamma_{2t} \;\; \psi_t \;\; \psi^*_t]'$, $\boldsymbol{Z} = [1,0,1,0,1,1,0]$, $\boldsymbol{G} = [\sigma_\epsilon, 0, 0, 0, 0]$, $\boldsymbol{T} = \text{diag}(\boldsymbol{T}_\mu, \boldsymbol{T}_\gamma, \boldsymbol{T}_\psi)$, where

$$\boldsymbol{T}_\mu = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{T}_\gamma = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \boldsymbol{T}_\psi = \rho \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix},$$

$\boldsymbol{H} = [\boldsymbol{0} \quad \tilde{\boldsymbol{H}}], \tilde{\boldsymbol{H}} = \text{diag}(\sigma_\eta, \sigma_\zeta, \sigma_\omega, \sigma_\omega, \sigma_\omega, \sigma_\kappa, \sigma_\kappa)$, and, as regards initial conditions,

$$\boldsymbol{a} = \boldsymbol{0}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{0} \\ \sigma_\psi \boldsymbol{I}_2 \end{bmatrix}, \boldsymbol{D} = \begin{bmatrix} \boldsymbol{I}_5 \\ \boldsymbol{0} \end{bmatrix}.$$

The case when $\boldsymbol{\delta}$ is fixed and unknown has been considered by Rosenberg (1973), who shows that it can be concentrated out of the likelihood function. As a matter of fact, the innovations and the conditional mean of the state vector delivered by the the KF with initial conditions $\tilde{\boldsymbol{\alpha}}_{1|0} = \boldsymbol{a} + \boldsymbol{D}\boldsymbol{\delta}$ and $\boldsymbol{P}_{1|0} = \boldsymbol{B}\boldsymbol{B}'$, denoted KF($\boldsymbol{\delta}$), can be written respectively as $\boldsymbol{\nu}_t = \boldsymbol{\nu}^*_t - \boldsymbol{V}_t\boldsymbol{\delta}$ and $\tilde{\boldsymbol{\alpha}}_{t+1|t} = \tilde{\boldsymbol{\alpha}}^*_{t+1|t} - \boldsymbol{A}_{t+1|t}\boldsymbol{\delta}$. The starred quantities, $\boldsymbol{\nu}^*_t$ and $\tilde{\boldsymbol{\alpha}}^*_{t+1|t}$, are produced by the KF run with $\boldsymbol{\delta} = \boldsymbol{0}$, i.e. with initial conditions $\tilde{\boldsymbol{\alpha}}^*_{1|0} = \boldsymbol{a}$ and $\boldsymbol{P}^*_{1|0} = \boldsymbol{B}\boldsymbol{B}'$; we denote this filter by KF($\boldsymbol{0}$). Note that it produces the matrices $\boldsymbol{F}^*_t$ and $\boldsymbol{P}^*_{t+1|t}, t = 1, \ldots T$, which, being invariant to $\boldsymbol{\delta}$, equal the corresponding matrices $\boldsymbol{F}_t$ and $\boldsymbol{P}_{t+1|t}$ produced by KF($\boldsymbol{\delta}$). The matrices $\boldsymbol{V}_t$ and $\boldsymbol{A}_{t+1|t}$ are generated by the following recursions, that are run in parallel to KF($\boldsymbol{0}$):

$$\boldsymbol{V}_t = -\boldsymbol{Z}_t\boldsymbol{A}_{t|t-1}, \quad \boldsymbol{A}_{t+1|t} = \boldsymbol{T}_t\boldsymbol{A}_{t|t-1} + \boldsymbol{K}_t\boldsymbol{V}_t, \quad t = 1, \ldots, T, \tag{15}$$

with initial value $\boldsymbol{A}_{1|0} = -\boldsymbol{D}$. Then, replacing $\boldsymbol{\nu}_t = \boldsymbol{\nu}^*_t - \boldsymbol{V}_t\boldsymbol{\delta}$ and $\boldsymbol{F}_t = \boldsymbol{F}^*_t$ into (14), yields:

$$\mathsf{L}(\boldsymbol{Y}_T; \boldsymbol{\theta}, \boldsymbol{\delta}) = -\frac{1}{2}\left(\sum_{t=1}^{T} \ln|\boldsymbol{F}^*_t| + \sum_{t=1}^{T} \boldsymbol{\nu}^{*\prime}_t \boldsymbol{F}^{*-1}_t \boldsymbol{\nu}^*_t - 2\boldsymbol{\delta}'\boldsymbol{s}_T + \boldsymbol{\delta}'\boldsymbol{S}_T\boldsymbol{\delta}\right) \tag{16}$$

where $\boldsymbol{s}_T = \sum_1^T \boldsymbol{V}'_t \boldsymbol{F}^{*-1}_t \boldsymbol{\nu}^*_t$ and $\boldsymbol{S}_T = \sum_1^T \boldsymbol{V}'_t \boldsymbol{F}^{*-1}_t \boldsymbol{V}_t$. Hence, the maximum likelihood estimate of $\boldsymbol{\delta}$ is $\hat{\boldsymbol{\delta}} = \boldsymbol{S}_T^{-1}\boldsymbol{s}_T$ and the concentrated likelihood is

$$\mathsf{L}_\delta(\boldsymbol{Y}_T; \boldsymbol{\theta}) = -0.5(\sum_{t=1}^{T} \ln|\boldsymbol{F}^*_t| + \sum_{t=1}^{T} \boldsymbol{\nu}^{*\prime}_t \boldsymbol{F}^{*-1}_t \boldsymbol{\nu}^*_t - \boldsymbol{s}'_T \boldsymbol{S}_T^{-1} \boldsymbol{s}_T) \tag{17}$$

When $\boldsymbol{\delta}$ is diffuse, $\boldsymbol{\delta} \sim \mathsf{N}(\boldsymbol{0}, \kappa \boldsymbol{I}_d)$, $\kappa \to \infty$, the definition of the likelihood needs to be amended; in particular, de Jong (1991) shows that only

$\mathsf{L}(\boldsymbol{Y}_T;\boldsymbol{\theta},\boldsymbol{\delta}) + \frac{d}{2}\ln\kappa$ is a proper likelihood and that its limiting expression for $\kappa \to \infty$ is

$$\mathsf{L}_\infty(\boldsymbol{Y}_T;\boldsymbol{\theta}) = -\frac{1}{2}\left(\sum \ln|\boldsymbol{F}_t^*| + \ln|\boldsymbol{S}_T| + \sum \boldsymbol{\nu}_t^{*\prime}\boldsymbol{F}_t^{*-1}\boldsymbol{\nu}_t^* - \boldsymbol{s}_T'\boldsymbol{S}_T^{-1}\boldsymbol{s}_T\right). \tag{18}$$

Also, the limiting expressions for the mean and variance of $\boldsymbol{\delta}$, conditional on $\boldsymbol{Y}_t$ are $\boldsymbol{S}_t^{-1}\boldsymbol{s}_t$ and $\boldsymbol{S}_t^{-1}$, $\boldsymbol{s}_t = \sum_{i=1}^t \boldsymbol{V}_i'\boldsymbol{F}_i^{*-1}\boldsymbol{\nu}_i^*$ and $\boldsymbol{S}_t = \sum_{i=1}^t \boldsymbol{V}_i'\boldsymbol{F}_i^{*-1}\boldsymbol{V}_i$, provided the latter matrix is invertible. All the relevant quantities are available by the run of KF($\mathbf{0}$), augmented by the recursions (15). Furthermore, de Jong shows that the limiting expressions for the innovations, the one-step-ahead prediction of the state vector and the corresponding covariance matrices are

$$\begin{array}{lcllcl} \boldsymbol{\nu}_t & = & \boldsymbol{\nu}_t^* - \boldsymbol{V}_t\boldsymbol{S}_{t-1}^{-1}\boldsymbol{s}_{t-1}, & \boldsymbol{F}_t & = & \boldsymbol{F}_t^* + \boldsymbol{V}_t\boldsymbol{S}_{t-1}^{-1}\boldsymbol{V}_t', \\ \tilde{\boldsymbol{\alpha}}_{t|t-1} & = & \tilde{\boldsymbol{\alpha}}_{t|t-1}^* - \boldsymbol{A}_{t|t-1}\boldsymbol{S}_{t-1}^{-1}\boldsymbol{s}_{t-1}, & \boldsymbol{P}_{t|t-1} & = & \boldsymbol{P}_{t|t-1}^* + \boldsymbol{A}_{t|t-1}\boldsymbol{S}_{t-1}^{-1}\boldsymbol{A}_{t|t-1}'. \end{array} \tag{19}$$

Usually, the augmented recursions can be dropped after processing d observations provided $\boldsymbol{S}_d$ is invertible, in which case $\mathsf{E}(\boldsymbol{\delta}|\boldsymbol{Y}_d) = \boldsymbol{S}_d^{-1}\boldsymbol{s}_d$, and a collapse can be made to the standard KF.

The notion of a diffuse likelihood is close to that of a marginal likelihood, being based on a rank $T-d$ linear transformation of the series that eliminates dependence on δ. Comparing (18) with (17), it turns out that $\mathsf{L}_\infty(\boldsymbol{Y}_T;\boldsymbol{\theta})$ differs from $\mathsf{L}_\delta(\boldsymbol{Y}_T;\boldsymbol{\theta})$ because of the presence of the term $\ln|\boldsymbol{S}_T|$, which could bear relevant effects on the estimation of $\boldsymbol{\theta}$, especially when the data generating process is close to nonstationarity and noninvertibility. In this situation, as shown by Tunnicliffe-Wilson (1986) and Shephard and Harvey (1990), the latter referring to the estimation of the signal to noise ratio for the LLM when the true value is close to zero, the estimators based on (18) exhibit better small sample properties.

Recently, Koopman (1997) has provided an exact analytic solution to the initialisation problem that is computationally more efficient than augmenting the KF by the matrix recursions (15). His approach entails the derivation of a modified KF, hinging on the fundamental idea, adopted by Ansley and Kohn (1985, 1989), of expressing the KF quantities explicitly in terms of κ and letting $\kappa \to \infty$ to get the exact solution.

A simple illustration is provided with reference to the LLM (1), whose state space space representation has $\alpha_t = \mu_t$, $\boldsymbol{\varepsilon}_t = [\epsilon_t/\sigma_\epsilon \quad \eta_t/\sigma_\eta]'$, $Z = T = 1$, $\boldsymbol{G} = [\sigma_\epsilon \quad 0]$, $\boldsymbol{H} = [0 \quad \sigma_\eta]$, and initial conditions $\tilde{\mu}_{1|0} = 0$, $P_{1|0} = \kappa$ (i.e. $a = 0, B = 0$, and $D = 1$). The run of the KF at time $t = 1$ gives: $\nu_1 = y_1$, $F_1 = \kappa + \sigma_\epsilon^2$, $K_1 = \kappa/(\kappa + \sigma_\epsilon^2)$, $\tilde{\mu}_{2|1} = y_1\kappa/(\kappa + \sigma_\epsilon^2)$ and $P_{2|1} = \sigma_\epsilon^2\kappa/(\kappa + \sigma_\epsilon^2) + \sigma_\eta^2$. Letting $\kappa \to \infty$, we get the limiting expressions $K_1 = 1$, $\tilde{\mu}_{2|1} = y_1$ and $P_{2|1} = \sigma_\epsilon^2 + \sigma_\eta^2$. Note that $P_{2|1}$ does not depend on κ and $\nu_2 = y_2 - y_1$ has a proper distribution, $\nu_2 \sim \mathsf{N}(0, F_2)$, with $F_2 = \sigma_\eta^2 + 2\sigma_\epsilon^2$. In general the innovations at subsequent times can be written as a linear

combination of past and current changes of the series, and therefore have a proper distribution.

Suppressing dependence on $\boldsymbol{\theta} = [\sigma_\epsilon^2 \quad \sigma_\eta^2]'$, the diffuse log-likelihood is $\mathsf{L}(Y_T) + 0.5\ln\kappa = \mathsf{L}(\nu_1) + 0.5\ln\kappa + \sum_{t=2}^{T}\mathsf{L}(\nu_t)$, but $\mathsf{L}(\nu_1) + 0.5\ln\kappa = -0.5\ln(\kappa/(\kappa+\sigma_\epsilon^2)) - 0.5y_1^2/(\kappa+\sigma_\epsilon^2)$ vanishes as $\kappa \to \infty$ Hence the usual KF can be started at $t = 2$ with $\tilde{\mu}_{2|1}$ and $P_{2|1}$ as given. The log likelihood is computed as in (14) with the summation starting at $t = 2$. These calculations confirm the fact that the diffuse likelihood is based on a rank $T-1$ linear transformation of the series that makes it invariant to δ.

In closing this section, we hint that a simple expedient to obtain an approximation to the diffuse likelihood is to run the standard KF with κ replaced by a large number, such as 10^7. Although this solution is practical, it is theoretically unsatisfactory and prone to numerical inaccuracies.

8 Explanatory variables

Explanatory variables can be brought into the model so as to capture exogenous effects and various types of interventions. If we let $\boldsymbol{X}_t$ and $\boldsymbol{W}_t$ denote fixed and known matrices of dimension $N \times k$ and $m \times k$, respectively, the state space model is written

$$\boldsymbol{y}_t = \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{G}_t\boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\alpha}_{t+1} = \boldsymbol{T}_t\boldsymbol{\alpha}_t + \boldsymbol{W}_t\boldsymbol{\beta} + \boldsymbol{H}_t\boldsymbol{\varepsilon}_t. \tag{20}$$

When the SSM is stationary, $\boldsymbol{\beta}$ can be estimated by generalised least squares, which amounts to apply the same KF to $\boldsymbol{y}_t$ and each of the explanatory variables in the columns of $\boldsymbol{X}_t$, and perform a weighted least squares regression (with weights provided by $\boldsymbol{F}_t^{-1}$); see Harvey (1989), pp. 130-133 for details.

When the SSM is nonstationary (see section 7.2), the vector $\boldsymbol{\beta}$ can be incorporated in the vector $\boldsymbol{\delta}$ that is redefined as the $d+k$ vector $\boldsymbol{\delta} = [\boldsymbol{\delta}_\alpha' \quad \boldsymbol{\beta}']'$, with $\boldsymbol{\delta}_\alpha$ being associated to the nonstationary elements in the state vector. The matrix $\boldsymbol{D}$ in this context is the $m \times (k+d)$ matrix $[\boldsymbol{D}_\alpha \quad \boldsymbol{0}]$ where the first block is associated with $\boldsymbol{\delta}_\alpha$. The treatment under fixed and diffuse assumption is the same as in the previous subsection, with the matrix recursions (15) replaced by

$$\boldsymbol{V}_t = [\boldsymbol{0} \quad \boldsymbol{X}_t] - \boldsymbol{Z}_t\boldsymbol{A}_{t|t-1}, \quad \boldsymbol{A}_{t+1|t} = \boldsymbol{T}_t\boldsymbol{A}_{t|t-1} - [\boldsymbol{0} \quad \boldsymbol{W}_t] + \boldsymbol{K}_t\boldsymbol{V}_t$$

and initialised with $\boldsymbol{A}_{1|0} = -\boldsymbol{D}$.

An alternative equivalent approach consists in inserting $\boldsymbol{\beta}$ in the state vector and redefining the SSM accordingly:

$$\boldsymbol{y}_t = \boldsymbol{Z}_t^\dagger\boldsymbol{\alpha}_t^\dagger + \boldsymbol{G}_t\boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\alpha}_{t+1}^\dagger = \boldsymbol{T}_t^\dagger\boldsymbol{\alpha}_t^\dagger + \boldsymbol{H}_t^\dagger\boldsymbol{\varepsilon}_t$$

where

$$\boldsymbol{\alpha}_t^\dagger = \begin{bmatrix} \boldsymbol{\alpha}_t \\ \boldsymbol{\beta}_t \end{bmatrix}, \quad \boldsymbol{Z}_t^\dagger = [\boldsymbol{Z}_t \;\; \boldsymbol{X}_t], \quad \boldsymbol{T}_t^\dagger = \begin{bmatrix} \boldsymbol{T}_t & \boldsymbol{W}_t \\ \mathbf{0} & \boldsymbol{I}_k \end{bmatrix}, \boldsymbol{H}_t^\dagger = \begin{bmatrix} \boldsymbol{H}_t \\ \mathbf{0} \end{bmatrix}.$$

9 Smoothing

Once the parameters of a structural model have been estimated, interest lies on estimation of the components based on the full sample of observations, $\boldsymbol{Y}_T$. This operation is referred to as *smoothing* (see Anderson and Moore, 1979, ch. 7, for a comprehensive account of the classic algorithms). The time series plot of the smoothed components against time is also a valuable diagnostic tool to check if the components extracted provide a suitable representation of the stylised facts concerning the series.

In the Gaussian case smoothing provides the MMSE of $\boldsymbol{\alpha}_t$ using $\boldsymbol{Y}_T$, $\tilde{\boldsymbol{a}}_{t|T} = \mathsf{E}(\boldsymbol{\alpha}_t|\boldsymbol{Y}_T)$, along with its MSE matrix $\boldsymbol{P}_{t|T} = \mathsf{E}[(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{a}}_{t|T})(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{a}}_{t|T})'|\boldsymbol{Y}_T]$. It can be performed efficiently by the algorithm proposed by de Jong (1989), consisting of the following backwards recursions starting at $t = T$, with initial values $\boldsymbol{r}_T = \mathbf{0}$ and $\boldsymbol{N}_T = \mathbf{0}$:

$$\begin{array}{rclrcl} \tilde{\boldsymbol{\alpha}}_{t|T} & = & \tilde{\boldsymbol{\alpha}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{r}_{t-1}, & \boldsymbol{P}_{t|T} & = & \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{N}_{t-1}\boldsymbol{P}_{t|t-1} \\ \boldsymbol{r}_{t-1} & = & \boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t + \boldsymbol{L}_t'\boldsymbol{r}_t, & \boldsymbol{N}_{t-1} & = & \boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{N}_t\boldsymbol{L}_t \end{array} \tag{21}$$

where $\boldsymbol{L}_t = \boldsymbol{T}_t - \boldsymbol{K}_t\boldsymbol{Z}_t$. A preliminary forward KF pass is required to store the quantities $\tilde{\boldsymbol{\alpha}}_{t|t-1}$, $\boldsymbol{P}_{t|t-1}$, $\boldsymbol{\nu}_t$, $\boldsymbol{F}_t$ and $\boldsymbol{K}_t$. The proof of (21) is found in De Jong (1989), who also deals with the modifications to introduce under diffuse initial conditions. These involve storage of $\boldsymbol{V}_t$ (15) and extra matrix recursions. Alternatively, using the exact initial KF of Koopman (1997), (21) still apply for $t = T, \ldots, d+1$ and a straightforward adjustment can be made for the initial stretch $t = d, \ldots, 1$. When Gaussianity does not hold, the smoother still delivers the MMSLE of $\boldsymbol{\alpha}_t$.

The estimation of the disturbances $\boldsymbol{\varepsilon}_t$ associated with the the various components in a SSM, referred to as *disturbance smoothing*, is built upon the *smoothing errors* (De Jong, 1988) $\boldsymbol{u}_t = \boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t - \boldsymbol{K}_t'\boldsymbol{r}_t$, with variance $\boldsymbol{D}_t = \boldsymbol{F}_t^{-1} + \boldsymbol{K}_t'\boldsymbol{N}_t\boldsymbol{K}_t$. Koopman (1992) shows that $\tilde{\boldsymbol{\varepsilon}}_{t|T} = \mathsf{E}(\boldsymbol{\varepsilon}_t|\boldsymbol{Y}_t) = \boldsymbol{G}_t'\boldsymbol{u}_t + \boldsymbol{H}_t'\boldsymbol{r}_t$ and $\mathsf{Var}(\tilde{\boldsymbol{\varepsilon}}_{t|T}) = \boldsymbol{G}_t'\boldsymbol{D}_t\boldsymbol{G}_t + \boldsymbol{H}_t'\boldsymbol{N}_t\boldsymbol{H}_t$. The standardised smoothed estimates of the disturbances are known referred to as *auxiliary residuals*. When $\boldsymbol{H}_t\boldsymbol{G}_t' = \mathbf{0}$, the irregular auxiliary residual is $\boldsymbol{G}_t\boldsymbol{G}_t'\boldsymbol{u}_t$, standardised by the square root of the diagonal elements in $\boldsymbol{G}_t\boldsymbol{G}_t'\boldsymbol{D}_t\boldsymbol{G}_t\boldsymbol{G}_t'$, whereas the auxiliary residuals associated with the unobserved components in $\boldsymbol{\alpha}_t$ are the elements of $\boldsymbol{H}_t\boldsymbol{H}_t'\boldsymbol{r}_t$, scaled by the square root of the diagonal elements of $\boldsymbol{H}_t\boldsymbol{H}_t'\boldsymbol{N}_t\boldsymbol{H}_t\boldsymbol{H}_t'$. They provide test statistics for outliers and structural change in the state components (Harvey and Koopman, 1992, De Jong and Penzer, 1998). Unlike the innovations, the auxiliary residuals are serially

correlated; Harvey and Koopman (1992) derive their autocorrelation structure and show how they can be employed to form appropriate tests of normality.

10 Forecasting

For a SSM without regression effects, the update of the KF at time t produces $\tilde{\boldsymbol{\alpha}}_{t+1|t}$ and $\boldsymbol{P}_{t+1|t}$, which are used to yield the one-step-ahead forecast $\tilde{\boldsymbol{y}}_{t+1|t} = \boldsymbol{Z}_{t+1}\tilde{\boldsymbol{\alpha}}_{t+1|t}$, along with its MSE matrix $\boldsymbol{F}_{t+1} = \boldsymbol{Z}_{t+1}\boldsymbol{P}_{t+1|t}\boldsymbol{Z}'_{t+1} + \boldsymbol{G}_{t+1}\boldsymbol{G}'_{t+1}$.

For multistep prediction, taking the expectation of both sides of the measurement equation (8) conditional on $\boldsymbol{Y}_t$, gives $\tilde{\boldsymbol{y}}_{t+l|t} = \boldsymbol{Z}_{t+l}\tilde{\boldsymbol{\alpha}}_{t+l|t}$; moreover, $\mathsf{MSE}(\tilde{\boldsymbol{y}}_{t+l|t}) = \boldsymbol{F}_{t+l|t} = \boldsymbol{Z}_{t+l}\boldsymbol{P}_{t+l|t}\boldsymbol{Z}'_{t+l} + \boldsymbol{G}_{t+l}\boldsymbol{G}'_{t+l}$, where the l-step-ahead forecast of the state vector, $\tilde{\boldsymbol{\alpha}}_{t+l|t} = \mathsf{E}(\boldsymbol{\alpha}_{t+l}|\boldsymbol{Y}_t)$, and its MSE, $\boldsymbol{P}_{t+l|t} = \mathsf{E}[(\boldsymbol{\alpha}_{t+l} - \tilde{\boldsymbol{\alpha}}_{t+l|t})(\boldsymbol{\alpha}_{t+l} - \tilde{\boldsymbol{\alpha}}_{t+l|t})'|\boldsymbol{Y}_t]$, are built up recursively by the chain rule:

$$\tilde{\boldsymbol{\alpha}}_{t+l|t} = \boldsymbol{T}_{t+l-1}\tilde{\boldsymbol{\alpha}}_{t+l-1|t}, \quad \boldsymbol{P}_{t+l|t} = \boldsymbol{T}_{t+l-1}\boldsymbol{P}_{t+l-1|t}\boldsymbol{T}'_{t+l-1} + \boldsymbol{H}_{t+l-1}\boldsymbol{H}'_{t+l-1}.$$

These expressions are initialised by $\tilde{\boldsymbol{\alpha}}_{t+1|t}$ and $\boldsymbol{P}_{t+1|t}$ delivered by the KF at time t. Two things should be noticed: first, multi-step-ahead prediction errors are correlated and, secondly, the MSE matrices do not take into account the uncertainty arising from estimation of the parameters in $\boldsymbol{\theta}$. When regression and diffuse initial effects are present, $\tilde{\boldsymbol{y}}_{t+l|t} = \boldsymbol{Z}_{t+l}\tilde{\boldsymbol{\alpha}}_{t+l|t} + \boldsymbol{X}_{t+l}\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\alpha}}_{t+l|t} = \boldsymbol{T}_{t+l-1}\tilde{\boldsymbol{\alpha}}_{t+l-1|t} + \boldsymbol{W}_{t+l-1}\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}} = \mathsf{E}(\boldsymbol{\beta}|\boldsymbol{Y}_t) = \boldsymbol{S}_t^{-1}\boldsymbol{s}_t$ and the initialisation for $\tilde{\boldsymbol{\alpha}}_{t+l|t}$ and $\boldsymbol{P}_{t+l|t}$ is provided by the second row of (19). In such case the MSE matrices will reflect the uncertainty associated with the regression and the initial diffuse effects, the same holding when $\boldsymbol{\beta}$ is incorporated into the state vector.

Writing $\tilde{\boldsymbol{y}}_{t+l|t} = \boldsymbol{Z}_{t+l}\boldsymbol{T}_{t+l,t+1}\tilde{\boldsymbol{\alpha}}_{t+1|t}$, where $\boldsymbol{T}_{j,s} = \boldsymbol{T}_{j-1}\boldsymbol{T}_{j-2}\cdots\boldsymbol{T}_s$ for $j > s$ and $\boldsymbol{T}_{j,s} = \boldsymbol{I}$ for $j = s$, the forecasts can be expressed as a weighted linear combination of past and current innovations:

$$\tilde{\boldsymbol{y}}_{t+l|t} = \boldsymbol{Z}_{t+l}\boldsymbol{T}_{t+l,t+1}\sum_{j=0}^{t-1}\left[\boldsymbol{T}_{t+1,t+1-j}\boldsymbol{K}_{t-j}\boldsymbol{\nu}_{t-j}\right].$$

Also, writing $\tilde{\boldsymbol{\alpha}}_{t+1|t} = \boldsymbol{L}_t\tilde{\boldsymbol{\alpha}}_{t|t-1} + \boldsymbol{K}_t\boldsymbol{y}_t$, where $\boldsymbol{L}_t = \boldsymbol{T}_t - \boldsymbol{K}_t\boldsymbol{Z}_t$, repeated substitution allows the forecast function to be written as a weighted average of past and current observations:

$$\tilde{\boldsymbol{y}}_{t+l|t} = \boldsymbol{Z}_{t+l}\boldsymbol{T}_{t+l,t+1}\sum_{j=0}^{t-1}\left[\boldsymbol{L}_{t+1,t+1-j}\boldsymbol{K}_{t-j}\boldsymbol{y}_{t-j}\right],$$

where we have set $\boldsymbol{L}_{j,s} = \boldsymbol{L}_{j-1}\boldsymbol{L}_{j-2}\cdots\boldsymbol{L}_s$ for $j > s$ and $\boldsymbol{L}_{j,s} = \boldsymbol{I}$ for $j = s$. Once new information, $\boldsymbol{y}_{t+1}$, becomes available the forecast of $\boldsymbol{y}_{t+l}$ can be

updated by the formula $\tilde{\boldsymbol{y}}_{t+l|t+1} = \tilde{\boldsymbol{y}}_{t+l|t} + \boldsymbol{Z}_{t+l}\boldsymbol{T}_{t+l,t+2}\boldsymbol{K}_{t+1}\boldsymbol{\nu}_{t+1}$, which stresses that the forecast revision depends on the innovation at time $t+1$.

Example 10.1: The forecast function of the LLM (1) is a horizontal straight line, being given by $\tilde{y}_{t+l|t} = \tilde{\mu}_{t+1|t}, l = 1, 2, \ldots$, where $\tilde{\mu}_{t+1|t} = \tilde{\mu}_{t|t-1} + K_t\nu_t$; moreover, $\mathsf{MSE}(\tilde{y}_{t+l|t}) = P_{t+1|t} + (l-1)\sigma_\eta^2 + \sigma_\epsilon^2$. When the KF has reached a steady state, $K_t = K = P/(P+\sigma_\epsilon^2)$ so that the previous forecast is revised according to a fraction of the one-step-ahead forecast error. This fraction measures the *persistence* of the innovations and is always between 0 and 1. The constant P can be obtained as the only admissible solution of (11): $P = \sigma_\epsilon^2(q + \sqrt{q^2+4q})/2$, where $q = \sigma_\eta^2/\sigma_\epsilon^2$ is the signal to noise ratio. Furthermore, the forecasts are generated as an *exponentially weighted moving average* of the available observation, since $\tilde{y}_{t+l|t} = Ky_t + (1-K)Ky_{t-1} + (1-K)^2Ky_{t-2} + \cdots + (1-K)^jKy_{t-j} + \cdots$. The discounting of past observations varies with q: the larger q is, the greater the weight placed on the most recent observations.

For a time invariant model with uncorrelated measurement and transition disturbances ($\boldsymbol{HG}' = \boldsymbol{0}$), it is also useful to write the multistep forecasts in terms of $\tilde{\boldsymbol{\alpha}}_{t|t}$, as $\tilde{\boldsymbol{y}}_{t+l|t} = \boldsymbol{Z}\boldsymbol{T}^l\tilde{\boldsymbol{\alpha}}_{t|t}$. This formulation can be employed to evaluate the shape of the forecast function: for instance, for the structural model $y_t = \mu_t + \psi_t + \gamma_t + \epsilon_t$, with μ_t as in (2) and trigonometric seasonality,

$$\tilde{y}_{t+l|t} = \tilde{\mu}_{t|t} + l\tilde{\beta}_{t|t} + \rho^l[\tilde{\psi}_{t|t}\cos(l\lambda_c) + \tilde{\psi}^*_{t|t}\sin(l\lambda_c)] + \sum_{j=1}^{[s/2]}[\tilde{\gamma}_{t|t}\cos(l\lambda_j) + \tilde{\gamma}^*_{t|t}\sin(l\lambda_j)];$$

it should be noted that the trend forecasts are linear in the forecast horizon, l, with intercept and slope that are adaptive in the forecast origin, the cycle and the seasonal contribute via sine and cosine waves, the former vanishing as $l \to \infty$. Moreover, writing $\tilde{\boldsymbol{\alpha}}_{t|t} = \boldsymbol{T}\tilde{\boldsymbol{\alpha}}_{t-1|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{Z}'\boldsymbol{F}_t^{-1}\boldsymbol{\nu}_t$, we can view forecasting with structural time series models as an "error learning" process or generalised *exponential smoothing*. The weights attached to the observations by the components can be obtained recursively from $\tilde{\boldsymbol{\alpha}}_{t|t} = (\boldsymbol{I} - \boldsymbol{P}_{t|t-1}\boldsymbol{Z}'\boldsymbol{F}_t^{-1}\boldsymbol{Z})\boldsymbol{T}\tilde{\boldsymbol{\alpha}}_{t-1|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{Z}'\boldsymbol{F}_t^{-1}\boldsymbol{y}_t$.

Example 10.2: For the LLTM of example 7.1 the forecast function is $\tilde{y}_{t+l|t} = \tilde{\mu}_{t|t} + l\tilde{\beta}_{t|t}$; the steady state recursions for $\tilde{\mu}_{t|t}$ and $\tilde{\beta}_{t|t}$ are equivalent to those of the Holt-Winters' forecasting technique:

$$\begin{array}{rcccccc} \tilde{\mu}_{t|t} & = & \tilde{\mu}_{t-1|t-1} & + & \tilde{\beta}_{t-1|t-1} & + & \lambda_0\nu_t \\ \tilde{\beta}_{t|t} & = & & + & \tilde{\beta}_{t-1|t-1} & + & \lambda_0\lambda_1\nu_t \end{array}$$

with $\lambda_0 = p_{11}/(p_{11}+\sigma_\epsilon^2)$ and $\lambda_1 = p_{12}/p_{11}$, where $\boldsymbol{P} = \{p_{ij}\}$ satisfies (11). The smoothing constants λ_0 and λ_1, both in the range (0,1), as $\sigma_\eta^2 \geq 0$

implies $0 < p_{12} < p_{11}$, are functionally related to the signal to noise ratios $q_\eta = \sigma_\eta^2/\sigma_\epsilon^2$ and $q_\zeta = \sigma_\zeta^2/\sigma_\epsilon^2$ (see Harvey, 1989, p. 175-177). The forecast MSE is a quadratic function of the forecast horizon: $\mathsf{MSE}(\tilde{y}_{t+l|t}) = p_{11} + 2(l-1)p_{12} + (l-1)^2 p_{22} + (l-1)\sigma_\eta^2 + l(l-1)\sigma_\zeta^2/2 + \sigma_\epsilon^2$. The steady state weights attributed to the observations can be derived from replacing $\boldsymbol{PZ'F}^{-1} = [\lambda_0 \quad \lambda_0\lambda_1]'$ in $\tilde{\boldsymbol{\alpha}}_{t|t} = [\boldsymbol{I} - (\boldsymbol{I} - \boldsymbol{PZ'F}^{-1}\boldsymbol{Z})\boldsymbol{T}L]^{-1}\boldsymbol{PZ'F}^{-1}y_t$. This gives $\tilde{\mu}_{t|t} = \theta(L)^{-1}\lambda_0[1-(1-\lambda_1)L]y_t$ and $\tilde{\beta}_{t|t} = \theta(L)^{-1}\lambda_0\lambda_1\Delta y_t$, where $\theta(L) = 1 - [2 - \lambda_0(1+\lambda_1)]L + (1-\lambda_0)L^2$. Note that the weights are less than 1 in modulus and sum up to 1 and to 0 respectively for the level and the slope.

In the ARIMA framework the forecast function can be decomposed into components associated to the roots of the autoregressive polynomial (Box *et al.*, 1994, ch. 5). It would appear that for *any* ARIMA model a structural interpretation can be provided (i.e. in terms of trends, cycles, and so forth), but this is unwarranted. This point can be illustrated with respect to the ARIMA(1,1,1) model $(1-\rho L)\Delta y_t = (1+\theta L)\xi_t$, with $\rho, \theta \in (-1, 1)$ and $\xi_t \sim \mathsf{NID}(0, \sigma^2)$: the forecast function can be written $\tilde{y}_{t+l|t} = m_t + \rho^l c_t$, where the updating equations for the components are:

$$m_t = m_{t-1} + [(1+\theta)/(1-\rho)]\xi_t, \quad c_t = \rho c_{t-1} - [(\rho+\theta)/(1-\rho)]\xi_t, \quad (22)$$

The sequence $\{m_t, c_t\}$ is also known as the Beveridge and Nelson (1981) decomposition of y_t into a permanent and a transitory component, and the constant $[(1+\theta)/(1-\rho)]$ is referred to as persistence, since it represents the scalar multiple of the innovation determining the amount of revision in the long-run forecast of the series.

It is easily shown that the trend is related to the observations via $m_t = [(1+\theta)/(1-\rho)](1-\rho L)(1+\theta L)^{-1}y_t$, so that the weights attributed to y_{t-j} are geometrically decreasing from time $t-1$ on and sum up to one. However, they can be greater than 1 if persistence is greater than 1. For instance, when $\theta = .8, \rho = 0.5$, the weights attributed to $y_{t-j}, j = 0, 1, 2, 3$, are 3.60, -4.68, 3.74, -3.00, respectively. In this situation the trend is, loosely speaking, more volatile than the series itself, since the innovations are not discounted, but their effect is amplified (in the example above by a factor of 3.60), and its time series plot will have a very uneven appearance. It is questionable whether this can be called a trend.

The structural time series model with ARIMA(1,1,1) reduced form is the trend plus AR(1) model $y_t = \mu_t + \psi_t$, $\mu_{t+1} = \mu_t + \eta_t$, $\psi_{t+1} = \rho\psi_t + \kappa_t$, with independent disturbances $\eta_t \sim \mathsf{NID}(0, \sigma_\eta^2)$ and $\kappa_t \sim \mathsf{NID}(0, \sigma_\kappa^2)$, and is such that persistence is constrained to be less than unity. As a matter of fact, the stationary representation of the structural model is $(1-\rho L)\Delta y_t = (1-\rho L)\eta_{t-1} + \Delta\kappa_{t-1}$, and equating the autocovariances of $(1-\rho L)\Delta y_t$ at lags 0 and 1 yields $\sigma^2(1+\theta^2) = (1+\rho^2)\sigma_\eta^2 + 2\sigma_\kappa^2$ and $\sigma^2\theta = -\rho\sigma_\eta^2 - \sigma_\kappa^2$, which can be solved for σ_η^2 and σ_κ^2 to give $\sigma_\eta^2 = \sigma^2[(1+\theta)/(1-\rho)]^2$ and

$\sigma_\kappa^2 = -\sigma^2[\rho(1+\theta)^2/(1-\rho)^2 + \theta]$. Now, $\sigma_\kappa^2 \geq 0$ requires $\theta \leq -\rho$, and this amounts to constraining persistence to be not greater than one. The state space representation has $\boldsymbol{Z} = [1 \quad 1]$, $\boldsymbol{G} = \boldsymbol{0}'$, $\boldsymbol{T} = \text{diag}(1, \rho)$, $\boldsymbol{H} = \text{diag}(\sigma_\eta, \sigma_\kappa)$, so that $\boldsymbol{P}\boldsymbol{Z}'\boldsymbol{F}^{-1} = [(1+\theta)/(1-\rho),\ -(\rho+\theta)/(1-\rho)]'$; therefore, the steady state recursions for the components are exactly as in (22), with m_t, c_t replaced by $\tilde{\mu}_{t|t}$, $\tilde{\psi}_{t|t}$, and $\xi_t = \nu_t$.

10.1 Post-sample predictive testing and model evaluation

Diagnostic checking is usually carried out using the standardised innovations $\boldsymbol{v}_t = \boldsymbol{F}_t^{-1/2}\boldsymbol{\nu}_t$, which play a role in detecting various types of misspecifications, such as serial correlation, heteroscedasticity, nonnormality and structural change (see Harvey, 1989, sec. 5.4. and 8.4.2).

Assessing the goodness of fit of a structural model is closely bound up with forecasting: a basic measure is the prediction error variance (*pev*), defined as the variance of the one step-ahead prediction errors in the steady state. For a time invariant SSM, the *pev* is the steady state matrix $\boldsymbol{F} = \lim_{t\to\infty} \boldsymbol{F}_t$, which can be approximated by $\boldsymbol{F}_T$ (*finite pev*), and corresponds to the variance of the ARIMA reduced form disturbances.

The definition of a scale free measure of goodness of fit, analogous to the coefficient of determination in regression, varies according to the nature of the series under investigation. A comparison is made with a corresponding näive forecasting model: for instance, if the BSM is fitted to a univariate seasonal time series, the sum of squares of the one-step-ahead prediction errors (SSE) is compared to the sum of squares of the first differences of the series around the seasonal means (SSDSM); this is the sum of the prediction errors arising from a random walk with seasonal drift model, and the coefficient of determination is correspondingly defined as R_s^2 = 1-SSE/SSDSM.

The availability of post-sample observations, $y_{T+1}, \ldots, y_{T+l}$ (for simplicity we refer to the univariate case), can be exploited for assessing forecasting performance. Two types of prediction errors emerge: the one-step-ahead prediction errors, $\nu_{T+j}, j = 1, \ldots, l$, and the extrapolative residuals or j-steps-ahead prediction errors $\nu_{T+j|T} = y_{T+j} - \tilde{y}_{T+j|T}$; various measures of forecast accuracy can be built upon them, such as the sum of their absolute values and of their squares, in order to compare rival models.

A test of predictive failure aims at comparing the model performance in the future relative to its past performance. To assess whether the prediction errors in the post-sample period are significantly greater than those within the sample period the following *post-sample predictive failure* statistic is used:

$$\xi(l) = \frac{T-d}{l}\left(\sum_{h=1}^{l} v_{T+h}^2\right)\left(\sum_{t=d+1}^{T} v_t^2\right)^{-1},$$

where $v_t = \nu_t/\sqrt{F_t}$ are the standardised one-step-ahead prediction errors and d the number of nonstationary elements. If the model is correctly specified, $v_t \sim \mathsf{NID}(0,1)$ and the distribution of $\xi(l)$ is $F(l, T-d)$. The cumulative sum (CUSUM) of the standardised prediction errors is useful for detecting if the model is systematically over or underpredicting.

11 Non-linear and non-Gaussian models

The literature on non-linear and non-Gaussian structural models has been growing very rapidly during the last decade, paralleling the advances in computational inference using stochastic simulation techniques. This section provides only an incomplete and non technical account, placing more emphasis on applications in economics and finance; introductory material can be found in Harvey (1989), sec. 3.7, 6.5 and 6.6, Kitagawa and Gersh (1996), ch. 6, West and Harrison (1997), ch. 12-15.

An important class of non-linear models arises when the system matrices are functionally related to the information available at time $t-1$, that is $\boldsymbol{Z}_t = \boldsymbol{Z}_t(\boldsymbol{Y}_{t-1}), \boldsymbol{G}_t = \boldsymbol{G}_t(\boldsymbol{Y}_{t-1}), \boldsymbol{T}_t = \boldsymbol{T}_t(\boldsymbol{Y}_{t-1}), \boldsymbol{H}_t = \boldsymbol{H}_t(\boldsymbol{Y}_{t-1})$. The resulting SSM is *conditionally Gaussian*, as given $\boldsymbol{Y}_{t-1}$ the system matrices can be regarded as fixed. The KF still delivers the MMSLE of the state vector, but $\tilde{\boldsymbol{\alpha}}_{t|t-1}$ is no longer coincident with $\mathsf{E}(\boldsymbol{\alpha}_t|\boldsymbol{Y}_{t-1})$, latter being non-linear in $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$; similarly, $\boldsymbol{P}_{t|t-1}$ represents only the conditional MSE matrix of $\tilde{\boldsymbol{\alpha}}_{t|t-1}$. The attractive feature is that the likelihood function can be still obtained via the prediction error decomposition. Forecasting is discussed in Harvey (1989, p. 159). A conditionally Gaussian setup is used in Harvey *et al.* (1992) in order to provide approximate filtering and quasi-maximum likelihood estimation for structural models with ARCH disturbances (STARCH models). Another example is provided by the smooth transition structural models used by Proietti (1999) to model business cycle asymmetries, like that occurring when contractions are steeper, but shorter, than expansions.

The general framework for handling non-linear and non-Gaussian models is such that the measurement equation is replaced by the observation conditional density

$$f(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T|\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T; \boldsymbol{\theta}) = \prod_{t=1}^{T} f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t; \boldsymbol{\theta}),$$

which specifies that $\boldsymbol{\alpha}_t$ is sufficient for $\boldsymbol{y}_t$, and the transition equation is replaced by the Markovian transition density, $f(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T; \boldsymbol{\theta}) = f(\boldsymbol{\alpha}_0; \boldsymbol{\theta}) \prod_{t=1}^{T-1} f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t; \boldsymbol{\theta})$. A unified treatment of statistical inference via simulation in this framework is provided in Shephard and Pitt (1997) and Durbin and Koopman (1997, 2000). Leaving aside further details, we highlight the following applications:

Structural models with Markov switching This class of models, introduced by Harrison and Stevens (1976) as a particular case of the *multiprocess* class, postulates that the system matrices vary according to the states of a latent first order Markov chain. It is adopted in Kim (1993), who proposes an approximate filter and smoother, to decompose the US inflation rate into a random walk trend plus stationary AR(1) components with heteroscedastic disturbances, whose variances vary according to two independent two-states (low and high uncertainty) Markov chain. Simulated inference is considered in Shephard (1994) and the monograph by Kim and Nelson (1999) provides a comprehensive treatment and illustrations.

Multiplicative models Multiplicative models arise in nonlinear seasonal adjustment, when the trend and the seasonal component combine multiplicatively and the irregular variance depends on the underlying trend; see Harvey (1989, p. 174) and Shephard (1994).

Dynamic generalised linear models This class of models arises for time series observations originating from the exponential family, such as count data with Poisson and binomial distribution and continuous data with skewed distributions such as the exponential and the gamma distributions. A binomial application concerning advertising awareness is illustrated in West and Harrison (1997, sec 14.4), whereas Durbin and Koopman (2000) present a Poisson application with respect to monthly series of van drivers killed in road accidents in the UK.

Outlier models Outlying observations and structural breaks in the components can be handled as in section (8) by the inclusion of appropriate dummy variables on the right hand side of the measurement and transition equations. This strategy has several drawbacks: for instance, when a dummy is used to model an additive outlier, this amounts to considering the observation as missing, so that a weight of zero is assigned to it in signal extraction and forecasting; on the contrary, the observation may still contain some information, which could be elicited by downweighting it suitably. Moreover, in some empirical applications it is not infrequent to find out that the quest for a specification compatible with the Gaussian assumption does lead to the detection of a relevant number of occurrences of outliers and structural breaks, which should rather be taken as evidence for departure from Gaussianity. The alternative strategy consists in allowing the disturbances of a structural model to possess a heavy tailed density, such as Students' t-distribution, the general error distribution (Durbin and Koopman, 1997), or a mixture of Gaussian (Harrison and Stevens, 1976).

Stochastic variance models This class of models allows the variability of the series to change over time. The basic univariate specification for a time series of stock returns is $y_t = \epsilon_t \exp(h_t/2)$, $h_{t+1} = \beta + \rho h_t + \eta_t$, where ϵ_t and η_t are independent Gaussian processes with variances 1 and σ_η^2 respectively. This specification captures the empirical regularities found in financial time series, such as leptokurtosis, volatility clustering, and the fact that returns exhibit little or no serial correlation, whereas their squares show pronounced serial dependence. A comprehensive review of the various approaches to inference for stochastic variance models is provided by Shephard (1996); a freeware package (SVpack) linked to the Ox programming language (Doornik, 1998) is also made available by the same author (`http://www.nuff.ox.ac.uk/users/shephard/ox/`). A recent addition implementing the approach of Durbin and Koopman (1997) is Sandmann and Koopman (1998).

12 Illustrations

12.1 Italian Gross Domestic Product

The first illustration deals with the quarterly series of the Italian Gross Domestic Product at 1995 prices (source: ISTAT, *National Economic Accounts*), shown in panel (i) of fig. 2. The series displays a clear upward trend, with a changing slope: in effect, average yearly growth declines from about 3.5% at the beginning of the sample period to about 1.0% at the end. Moreover, some there is some graphical evidence for the presence of some cyclical behaviour especially in the period between the two oil crises and from 1993 onwards.

When the local linear trend model (2) is fitted to the logarithms of the series using STAMP 6.0, the interactive menu-driven programme for fitting and forecasting structural time series models documented in Koopman *et al.* (2000), the maximum likelihood estimates are $\hat{\sigma}_\epsilon^2 = 0$ (the irregular is absent), $\hat{\sigma}_\eta^2 = 183 \times 10^{-7}$, $\hat{\sigma}_\zeta^2 = 347 \times 10^{-7}$. Elaborating results from example 10.2 the forecast function implied by the model is $\tilde{\ln y}_{t+l|t} = \tilde{\mu}_{t|t} + l\tilde{\beta}_{t|t}$ with $\tilde{\mu}_{t|t} = y_t$ (notice that $\hat{\sigma}_\epsilon^2 = 0$ implies $\lambda_0 = 1$), so that the trend is coincident with the observations, and $\tilde{\beta}_{t|t} = [1 - (1 - \lambda_1)L]^{-1}\lambda_1 \Delta \ln y_t$ - the current estimate of the slope is an *exponentially weighted moving average* of current and past growth rates. As far as goodness of fit is concerned, the pev is 0.000065 and the coefficient of determination is 0.08. However, the model is misspecified as the Ljung-Box test of residual autocorrelation, $Q(P) = T^*(T^* + 2)\sum_{\tau=1}^{P}(T^* - \tau)r_v^2(\tau)$, where $T^* = T - d$ and $r_v(\tau)$ is the autocorrelation coefficient of v_t at lag τ, is significant at the 5% level for all $\tau < 13$.

Diagnostic checking and a priori considerations suggest to fit the trend

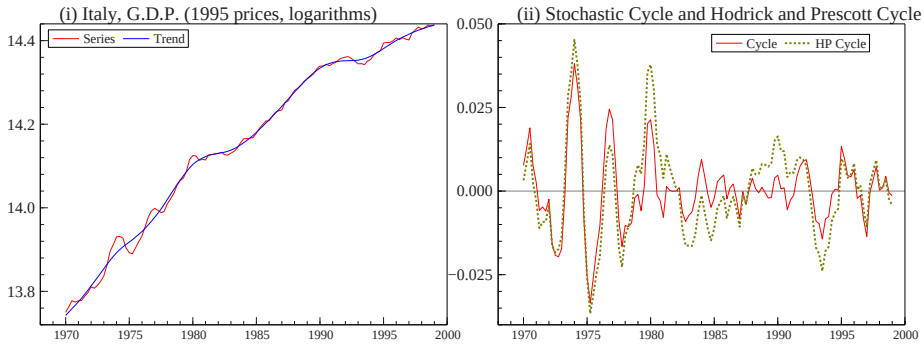

Figure 2: Italy, Gross Domestic Product at constant 1995 prices (logarithms), 1970.1-1999.1. (i) Series with trend, $\tilde{\mu}_{t|T}$. (ii) Comparison of $\tilde{\psi}_{t|T}$ from the model $\ln y_t = \mu_t + \psi_t$, and the cycle resulting from the Hodrick and Prescott (1997) detrending procedure.

plus cycle model $\ln y_t = \mu_t + \psi_t$: the estimated parameters for the trend are $\hat{\sigma}_\eta^2 = 2 \times 10^{-7}$, $\hat{\sigma}_\zeta^2 = 25 \times 10^{-7}$, $\hat{\sigma}_\psi^2 = 1402 \times 10^{-7}$, $\hat{\rho} = 0.92$, $\lambda_c = 0.52$, which implies a period of 12 quarters (3 years). As a result the trend extracted, $\tilde{\mu}_{t|T}$, has a fairly smooth appearance (see fig. 2, panel (i)), whereas $\tilde{\psi}_{t|T}$ provides a good representation of the Italian business cycle. The latter is is compared with the deviations from the Hodrick and Prescott trend, which is the trend extracted by the model (2) with the following restrictions on the parameters: $\sigma_\eta^2 = 0$ and $\sigma_\zeta^2/\sigma_\epsilon^2 = 0.000625$. This is clearly a misspecified model and the resulting cycle, although highly coherent with that extracted by the trend plus cycle model, overemphasises the short run dynamics, especially during the second decade which is commonly acknowledged as a period of steady growth with little or no business cycle fluctuations.

The absence of residual autocorrelation ($Q(P)$ is never significant at the 5% level) and of departures from the normality assumption, coupled with the better within sample performance ($pev = 0.000052$; the coefficient of determination is 0.27), suggest that the trend plus cycle model is to be preferred.

12.2 BEA Auto Unit Sales

We now provide an example of univariate modelling and forecasting with structural time series models with respect to the monthly domestic auto unit sales made available by the US *Bureau of Economic Analysis* (BEA) at the URL `http://www.bea.doc.gov/bea/pn/ndn0207.exe`. The series covers the sample period 1967.1-1998.6, and is an extension of that studied in Findley *et al.* (1998), to compare a subjective pre-adjustment made by an expert analyst and an objective one, based upon five user de-

fined regressors and automatic outlier identification. These adjustments are preliminary to the seasonal adjustment of the series and aim at removing the effects of short-duration sales incentive programmes used by automobile manufacturers, causing "a large increase in the month or two in which they occur, followed by a substantial decrease in the subsequent month or two" (Findley *et al.*, 1998, p. 146).

The series, shown in panel (i) of figure 3, along with time-varying level and seasonal components, contains also a relevant calendar component in the form of trading days effects, since the level of sales varies with the day of the week, eg. being higher than average in Thursday and Monday. This effect is modelled including six trading days regressors measuring respectively the number of Mondays, Tuesdays, ..., Saturdays minus the number of Sundays in every month. Actually, the treatment of this component is less trivial than it appears at first sight, due to the reporting habits of the manufacturers; see the comment by Cleveland on the paper by Findley *et al.* (1998, p. 153).

The initial specification of a structural time series model for the logarithmic transformation of the series is the following: $\ln y_t = \mu_t + \gamma_t + \boldsymbol{x}_t'\boldsymbol{\beta} + \epsilon_t$, where μ_t is a random walk, $\mu_{t+1} = \mu_t + \eta_t$, γ_t has the trigonometric specification (5), $\epsilon_t \sim \mathsf{NID}(0, \sigma_\epsilon^2)$, and $\boldsymbol{x}_t$ contains 6 trading days regressors and the 5 user-defined regressors employed by Findley *et al.* (1998).

This model is estimated using the package STAMP 6.0; diagnostic checking highlights that the standardised innovations are affected by excess kurtosis and significant autocorrelation (as revealed by the Ljung-Box statistic). Moreover, the shape of the autocorrelation function and the estimated spectral density, which shows a peak around the frequency 0.9 corresponding to a period of 6-7 months, seems to suggest the inclusion of a cyclical component.

When the component ψ_t in (3) is added there is a considerable improvement in the fit: the prediction error variance is *pev*=0.006394 and the coefficient of determination $R_s^2 = 0.44$ (for the previous specification we had *pev*=0.006954 and $R_s^2 = 0.39$). Furthermore, the Ljung-Box statistic is never significant and excess kurtosis is reduced. The maximum likelihood estimates of the parameters are $\hat{\sigma}_\eta^2 = 13279 \times 10^{-7}$, $\hat{\sigma}_\omega^2 = 36 \times 10^{-7}$, $\hat{\sigma}_\epsilon^2 = 4361 \times 10^{-7}$, $\hat{\sigma}_\kappa^2 = 16898 \times 10^{-7}$, $\hat{\rho} = 0.6785$, $\hat{\lambda}_c = 0.8964$.

The stochastic cycle, with variance $\hat{\sigma}_\psi^2 = 31316 \times 10^{-7}$ and period equal to 7 months, captures the quasi-seasonal effect of short duration sales programmes. As a matter of fact, panel (ii) of figure 3 shows that the smoothed estimates of the cycle are highly coherent with the subjective pre-adjustment factors. Of course, the cycle extracted is no substitute for genuine external information on the timing and extent of sales programmes, but structural modelling appears to capture properly this feature of the series.

Forecasts up to 12 steps ahead are displayed in panel (iv) in the original

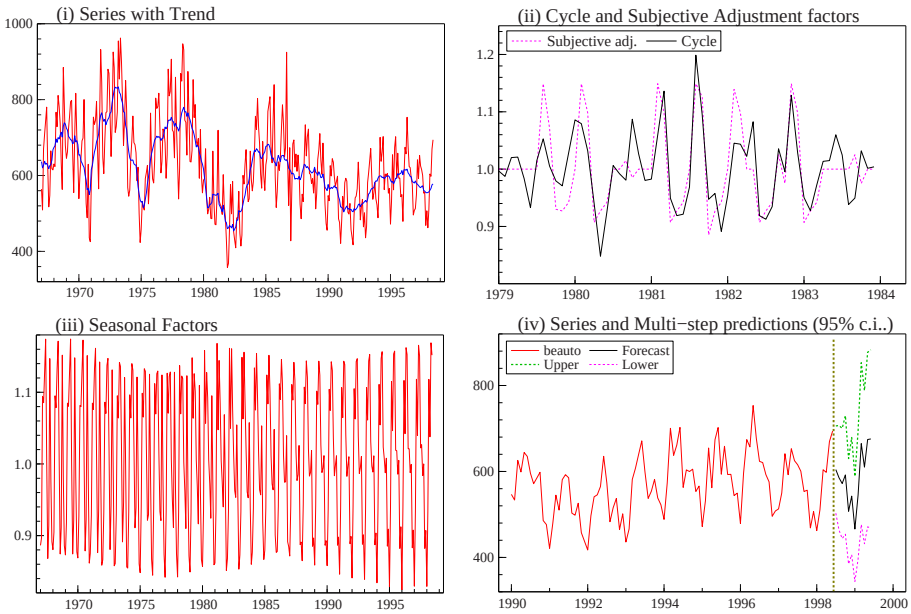

Figure 3: BEA Domestic Unit Auto Sales, 1967.1-1999.6. (i) Series with trend, $\exp(\tilde{\mu}_{t|T})$. (ii) Comparison of subjective pre-adjustment factors and the cycle extracted, $\exp(\tilde{\psi}_{t|T})$, for the period 1979.1-1983.12. (iii) Plot of seasonal factors, $\exp(\tilde{\gamma}_{t|T})$. (iv) l-step-ahead forecasts with upper and lower 95% confidence limits.

scale of the observations: if we denote $z_t = \ln y_t$, and we write $z_{T+l}|z_1, \ldots, z_T \sim \mathsf{N}(\tilde{z}_{T+l|T}, F_{T+l|t})$, then by properties of the lognormal distribution, the forecast in the original scale is obtained as $\tilde{y}_{T+l|T} = \exp(\tilde{z}_{T+l|T} + F_{T+l|T}/2)$.

For comparison, a regression model with ARIMA errors was fitted to the same series using the RegARIMA module included in X-12-ARIMA package with GiveWin interface (see the URL `http://www.nuff.ox.ac.uk/users/doornik/` for downloads and information about GiveWin. Documentation and downloads for X-12-ARIMA are available from `http://www.census.gov/-pub/ts/x12a/final/`). The model selected according to various information criteria is ARIMA (2,0,1) $\times$ (0,1,1):

$$(1-1.4007L+0.4382L^2)\Delta_{12}(\ln y_t - C_t) = (1-0.7180L)(1-0.7786L^{12})\xi_t,\ \xi_t \sim \mathsf{NID}(0, 0.006623)$$

where C_t represents the regression kernel, including six trading days regressors and five user-defined regressors.

The reduced form of the structural model with a cyclical component has the same autoregressive structure, but the implied representation of the stationary AR(2) polynomial is $(1-0.8473L+0.4604L^2)$, with a pair of complex conjugates roots with modulus 1.4738 and phase 0.8964, whereas the roots

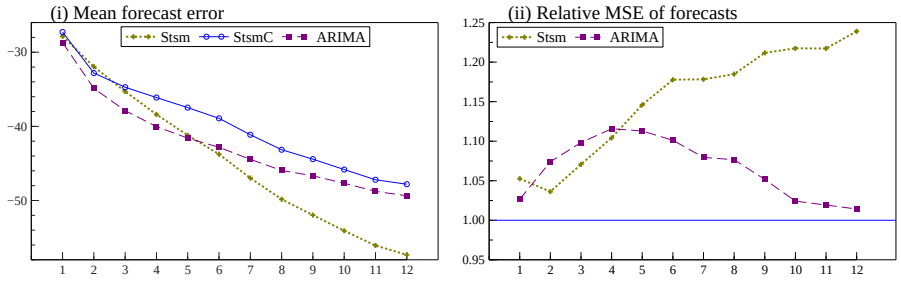

Figure 4: BEA Domestic Unit Auto Sales: comparison of forecast accuracy of structural model without a cyclical component (Stsm), including a cyclical component (StsmC) and the ARIMA(2,0,1)×(0,1,1) model, based on rolling l-step-ahead forecasts ($l = 1, \ldots, 12$) 1989.12-1999.5.

of $(1 - 1.4007L + 0.4382L^2)$ are real and equal 1.0614, 2.1500. Moreover, the MA part is additive (with order 14) rather than multiplicative.

The comparison of the prediction error variance of the structural model (0.006394) with that of the ARIMA model (0.006623) highlights that the performance of the former is superior as far as one-step-ahead forecast errors are concerned. Comparison of MSE of out-of-sample multistep forecasts can be based on the following rolling forecast exercise: starting from 1989.12, three alternative models, namely the structural model without (Stsm) and including (StsmC) the cyclical component and the ARIMA(2,0,1)×(0,1,1) model, are estimated using the observations through a given forecast origin; l-steps-ahead forecasts, $l = 1, \ldots, 12$, are computed. The procedure is repeated shifting the forecast origin by one month until the end of the sample period is reached; this yields a total of 102 one-step-ahead forecast errors and 91 twelve-steps-ahead forecast errors for the three models.

For this computationally demanding task we use the library of state space function SsfPack 2.2 by Koopman *et al.* (1999), linked to the object oriented matrix programming language Ox 2.1 of Doornik (1998). Moreover, a considerable simplification is obtained considering only one trading days regressor (accounting for the number of weekdays in the month minus (5/2) times the number of Saturdays and Sundays in the month), and dropping the 5 user defined regressors.

The main results are summarised in figure 4, which displays for the three models the mean of the forecast errors at forecast horizons ranging from 1 to 12 months (panel (i)) and the ratio of the mean square forecast error of the Stsm and the ARIMA models to that of StsmC (panel (ii)). All models present a negative bias at all forecast horizons, which is less pronounced for StsmC. The best forecast performance is provided by StsmC, which clearly outperforms the ARIMA model at horizons from 2-9, and the

Stsm specification, particularly at longer horizons.

Further empirical evidence on the forecasting performance of structural time series models is reported in Harvey and Todd (1983) and Andrews (1994). The overall conclusion is that the latter is similar and often superior to that of rival specifications.

12.2.1 References

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

Andrews, R.L.: "Forecasting Performance of Structural Time Series Models", Journal of Business and Economic Statistics, 12 (1994), 129-133.

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

Ansley, C.F., and Kohn, R.: "Filtering and smoothing in state space models with partially diffuse initial conditions", Journal of Time Series Analysis, 11 (1990), 275-293.

Beveridge, S., Nelson, C.R.: "A new Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to the Measurement of the 'Business Cycle'", Journal of Monetary Economics, 7 (1981), 151-174.

Box, G.E.P., Jenkins, G.M., and Reinsel, G.C.: Time Series Analysis, Third Edition (Englewood Cliffs NJ: Prentice-Hall, 1994).

de Jong, P.: "A cross-validation filter for time series models", Biometrika, 75 (1988), 594-600.

de Jong, P.: "Smoothing and interpolation with the state space model", Journal of the American Statistical Association, 84 (1989), 1085-1088.

de Jong, P.:, "The diffuse Kalman filter", Annals of Statistics, 19 (1991), 1073-1083.

de Jong, P., and Penzer, J.: "Diagnosing shocks in time series", Journal of the American Statistical Association, 93 (1998), 796-806.

Doornik, J.A.: Object-Oriented Matrix Programming Using Ox 2.0 (London, Timberlake Consultants Press, 1998).

Durbin, J., and Koopman, S.J.: "Monte Carlo maximum likelihood estimation for non-Gaussian state space models", Biometrika, 84 (1997), 669-684.

Durbin, J., and Koopman, S.J.: "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives", Journal of the Royal Statistical Society, Series B, 62 (2000), 3-56.

Engle, R.F., and Granger, C.W.J.: "Co-integration and error correction: representation, estimation and testing", Econometrica, 55 (1987), 251-276.

Fernandez, F.J., and Harvey, A.C.: "Seemingly unrelated time series equations and a test for homogeneity", Journal of Business and Economic Statistics, 8 (1990), 71-81.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., and Chen B.: "New capabilities and methods of the X-12-ARIMA seasonal adjustment program" (with comments), Journal of Business and Economic Statistics, 16 (1998), 127-177.

Godolphin, E.J.: Discussion on the paper by P.J. Harrison and C.F. Stevens "Bayesian Forecasting", Journal of the Royal Statistical Society, Series B, 38 (1976), 238-239.

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

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

Harvey, A.C., and Koopman, S.J.: "Diagnostic Checking of Unobserved Components Time Series Models", Journal of Business and Economic Statistics, 10 (1992), 377-389.

Harvey, A.C., and Koopman, S.J.: "Forecasting Hourly Electricity Demand using Time-Varying Splines", Journal of the American Statistical Association, 88 (1993), 377-389.

Harvey, A.C., and Koopman, S.J.: "Multivariate structural time series models", Systematic Dynamics in Economic and Financial Models, ed., C. Heij, H. Schumaker, B. Hanzon and C. Praagman (Chichester UK, Wiley, 1997).

Harvey, A.C., Koopman, S.J.: "Signal extraction and the formulation of unobserved component models", Econometrics Journal, forthcoming (2000).

Harvey, A.C., and Todd, P.H.J.: "Forecasting economic time series with structural and Box-Jenkins models" (with discussion), Journal of Business and Economic Statistics, 1 (1983), 299-315.

Harvey, A.C., Ruiz E., and Sentana, E.: "Unobserved component time series models with ARCH disturbances", Journal of Econometrics, 52 (1992), 129-157.

Hodrick R.J., and Prescott, E.C.: "Postwar U.S. Business Cycles: an Empirical Investigation", Journal of Money, Credit and Banking, 29 (1997), 1-16.

Hylleberg, S., ed.,: Modelling Seasonality (Oxford UK, Oxford University Press, 1992).

Kim, C.J.: "Unobserved-component time series models with Markov-switching heteroscedasticity: changes in regime and the link between inflation rates and inflation uncertainty", Journal of Business and Economic Statistics, 11 (1993), 341-349.

Kim, C.J., and Nelson, C.R.: State-Space Models with Regime Switching. (Cambridge MA, MIT Press, 1999).

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

Koopman, S.J.: "Disturbance Smoother for State Space Models", Biometrika, 80 (1993), 117-126.

Koopman, S.J.: "Exact initial Kalman filtering and smoothing for non-stationary time series models", Journal of the American Statistical Association, 92 (1997), 1630-1638.

Koopman, S.J., and Shephard, N.: "Exact score for time series models in state form", Biometrika, 79 (1992), 823-826.

Koopman S.J., Harvey, A.C., Doornik, J.A., and Shephard, N.: STAMP 6.0 Structural Time Series Analyser, Modeller and Predictor (London, Timberlake Consultants Ldt., 2000).

Koopman S.J., Shepard, N., and Doornik, J.A.: "Statistical algorithms for models in state space using SsfPack 2.2", Econometrics Journal, 2 (1999), 113-166.

Marshall, P.: "Estimating time-dependent means in dynamic models for cross-sections of time series", Empirical Economics, 17 (1992), 25-33.

Ord J.K., Koehler A.B., Snyder R.D.: "Estimation and prediction for a class of Dynamic nonlinear statistical models", Journal of the American Statistical Association, 92 (1997), 1621-1629.

Proietti, T.: "Seasonal heteroscedasticity and trends", Journal of Forecasting, 17 (1998), 1-17.

Proietti, T.: "Characterising asymmetries in business cycles using smooth transition structural time series models", Studies in Nonlinear Dynamics and Econometrics, 3.3 (1999), 141-156.

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

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

Sandmann G., and Koopman, S.J.: "Estimation of stochastic volatility models via Monte Carlo maximum likelihood", Journal of Econometrics, 87 (1998), 271-301.

Shephard, N.G.: "Partial non-Gaussian state space", Biometrika, 81, (1994), 115-131.

Shephard, N.G.: "Statistical aspects of ARCH and stochastic volatility", Time Series Models in Econometrics, Finance and Other Fields, ed., Cox, D.R., Hinkley, D.V., and Barndorff-Nielsen, O.E. (London, Chapman and Hall, 1996).

Shephard, N.G., and Harvey, A.C.: "On the probability of estimating a deterministic component in the local level model", Journal of Time Series Analysis, 11 (1990), 339-347.

Shephard, N.G., and Pitt, M.K.: "Likelihood analysis on non-Gaussian measurement time series", Biometrika, 84 (1997), 653-667.

Snyder, R.D.: "Recursive estimation of dynamic linear models", Journal of the Royal Statistical Society, Series B, 47 (1985), 272-276.

Tunnicliffe-Wilson, G.: "On the use of marginal likelihood in time series model estimation", Journal of the Royal Statistical Society, Series B, 51 (1989), 15-27.

West, M., and Harrison, J.: Bayesian Forecasting and Dynamic Models, 2nd ed. (New York, Springer-Verlag, 1997).

A Proof of the Kalman filter

The preliminary result which is used to determine the optimality of the KF is that for any two random variables x and z, the MMSE of x given z is the conditional expectation $\mathsf{E}(x|z)$. Assuming that at time t $\tilde{\boldsymbol{\alpha}}_{t|t-1}$ and $\boldsymbol{P}_{t|t-1}$ are given, taking the expectation of both sides of the measurement equation conditional on $\boldsymbol{Y}_{t-1}$ produces $\tilde{\boldsymbol{y}}_{t|t-1} = \mathsf{E}(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1}) = \boldsymbol{Z}_t\tilde{\boldsymbol{\alpha}}_{t|t-1}$. Denoting

the one-step-ahead prediction error, $\boldsymbol{y}_t - \tilde{\boldsymbol{y}}_{t|t-1}$, by $\boldsymbol{\nu}_t$, and substituting from (8), gives $\boldsymbol{\nu}_t = \boldsymbol{Z}_t(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{\alpha}}_{t|t-1}) + \boldsymbol{G}_t\boldsymbol{\varepsilon}_t$. Then, $\mathsf{Var}(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1}) = \boldsymbol{Z}_t\boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{G}_t\boldsymbol{G}_t' = \boldsymbol{F}_t$, since $\boldsymbol{\alpha}_t - \tilde{\boldsymbol{\alpha}}_{t|t-1}$ is uncorrelated with $\boldsymbol{\varepsilon}_t$ (as a matter of fact, repeated substitution from the transition equation shows that $\boldsymbol{\alpha}_t$ is linear in $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{t-1}$).

The updating equations for the state and its covariance matrix are produced as follows: writing $\boldsymbol{Y}_t = \{\boldsymbol{Y}_{t-1}, \boldsymbol{y}_t\}$, by properties of the normal distribution,

$$\begin{aligned}
\mathsf{E}(\boldsymbol{\alpha}_{t+1}|\boldsymbol{Y}_t) &= \mathsf{E}(\boldsymbol{\alpha}_{t+1}|\boldsymbol{Y}_{t-1}) + \mathsf{Cov}(\boldsymbol{\alpha}_{t+1}, \boldsymbol{y}_t|\boldsymbol{Y}_{t-1})[\mathsf{Var}(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1})]^{-1}(\boldsymbol{y}_t - \mathsf{E}(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1})), \\
\mathsf{Var}(\boldsymbol{\alpha}_{t+1}|\boldsymbol{Y}_t) &= \mathsf{Var}(\boldsymbol{\alpha}_{t+1}|\boldsymbol{Y}_{t-1}) - \mathsf{Cov}(\boldsymbol{\alpha}_{t+1}, \boldsymbol{y}_t|\boldsymbol{Y}_{t-1})[\mathsf{Var}(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1})]^{-1}\mathsf{Cov}(\boldsymbol{y}_t, \boldsymbol{\alpha}_{t+1}|\boldsymbol{Y}_{t-1}).
\end{aligned} \tag{23}$$

Now, the expectation and variance of both sides of (9) conditional on $\boldsymbol{Y}_{t-1}$ are $\mathsf{E}(\boldsymbol{\alpha}_{t+1}|\boldsymbol{Y}_{t-1}) = \boldsymbol{T}_t\tilde{\boldsymbol{\alpha}}_{t|t-1}$ and $\mathsf{Var}(\boldsymbol{\alpha}_{t+1}|\boldsymbol{Y}_{t-1}) = \boldsymbol{T}_t\boldsymbol{P}_{t|t-1}\boldsymbol{T}_t + \boldsymbol{H}_t\boldsymbol{H}_t'$, respectively; moreover, $\mathsf{Cov}(\boldsymbol{\alpha}_{t+1}, \boldsymbol{y}_t|\boldsymbol{Y}_{t-1}) = \boldsymbol{T}_t\boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{H}_t\boldsymbol{G}_t'$. Replacing into (23) and writing $\boldsymbol{K}_t = (\boldsymbol{T}_t\boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{H}_t\boldsymbol{G}_t')\boldsymbol{F}_t^{-1}$ yields the last line of (10).

The KF performs a Choleski transformation of the observations: if $\boldsymbol{\nu}$ denotes the stack of the innovations and $\boldsymbol{y}$ that of the observations: then $\boldsymbol{\nu} = \boldsymbol{C}\boldsymbol{y}$, where $\boldsymbol{C}$ is a lower triangular matrix such that $\mathsf{Cov}(\boldsymbol{y}) = \boldsymbol{C}^{-1}\boldsymbol{F}\boldsymbol{C}'^{-1}$ and $\boldsymbol{F} = \mathrm{diag}(\boldsymbol{F}_1, \ldots, \boldsymbol{F}_t, \ldots, \boldsymbol{F}_T)$. Hence, $\boldsymbol{\nu}_t$ is a linear combination of the current and past observations and is orthogonal to the information set $\boldsymbol{Y}_{t-1}$.