

Structural Time Series Models

Andrew C. Harvey and Neil Shephard

1. Introduction

A structural time series model is one which is set up in terms of components which have a direct interpretation. Thus, for example, we may consider the classical decomposition in which a series is seen as the sum of trend, seasonal and irregular components. A model could be formulated as a regression with explanatory variables consisting of a time trend and a set of seasonal dummies. Typically, this would be inadequate. The necessary flexibility may be achieved by letting the regression coefficients change over time. A similar treatment may be accorded to other components such as cycles. *The principal univariate structural time series models are therefore nothing more than regression models in which the explanatory variables are functions of time and the parameters are time-varying.* Given this interpretation, the addition of observable explanatory variables is a natural extension as is the construction of multivariate models. Furthermore, the use of a regression framework opens the way to a unified model selection methodology for econometric and time series models.

The key to handling structural time series models is the *state space form* with the state of the system representing the various unobserved components such as trends and seasonals. The estimate of the unobservable state can be updated by means of a *filtering* procedure as new observations become available. Predictions are made by extrapolating these estimated components into the future, while *smoothing* algorithms give the best estimate of the state at any point within the sample. A structural model can therefore not only provide forecasts, but can also, through estimates of the components, present a set of stylised facts; see the discussion in Harvey and Jaeger (1991).

A thorough discussion of the methodological and technical ideas underlying structural time series models is contained in the monographs by Harvey (1989) and West and Harrison (1989), the latter adopting a Bayesian perspective. Since then there have been a number of technical developments and applications to new situations. One of the purposes of the present article is to describe these new results.

1.1. Statistical formulation

A structural time series model for quarterly observations might consist of trend, cycle, seasonal and irregular components. Thus

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (1.1)$$

where μ_t is the trend, ψ_t is the cycle, γ_t is the seasonal and ε_t is the irregular. All four components are stochastic and the disturbances driving them are mutually uncorrelated. The trend, seasonal and cycle are all derived from deterministic functions of time, and reduce to these functions as limiting cases. The irregular is white noise.

The deterministic linear trend is

$$\mu_t = \alpha + \beta t. \quad (1.2)$$

Since μ_t may be obtained recursively from

$$\mu_t = \mu_{t-1} + \beta, \quad (1.3)$$

with $\mu_0 = \alpha$, continuity may be preserved by introducing stochastic terms as follows:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad (1.4a)$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad (1.4b)$$

where η_t and ζ_t are mutually uncorrelated white noise disturbances with zero means and variances, σ_η^2 and σ_ζ^2 respectively. The effect of η_t is to allow the level of the trend to shift up and down, while ζ_t allows the slope to change. The larger the variances, the greater the stochastic movements in the trend. If $\sigma_\eta^2 = \sigma_\zeta^2 = 0$, (1.4) collapses to (1.2) showing that the deterministic trend is a limiting case.

Let ψ_t be a cyclical function of time with frequency λ_c , which is measured in radians. The period of the cycle, which is the time taken to go through its complete sequence of values, is $2\pi/\lambda_c$. A cycle can be expressed as a mixture of sine and cosine waves, depending on two parameters, α and β . Thus

$$\psi_t = \alpha \cos \lambda_c t + \beta \sin \lambda_c t, \quad (1.5)$$

where $(\alpha^2 + \beta^2)^{1/2}$ is the amplitude and $\tan^{-1}(\beta/\alpha)$ is the phase. Like the linear trend, the cycle can be built up recursively, leading to the stochastic model

$$\begin{pmatrix} \psi_t \\ \psi_t^* \end{pmatrix} = \rho \begin{pmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^* \end{pmatrix}, \quad (1.6)$$

where κ_t and κ_t^* are mutually uncorrelated with a common variance, σ_κ^2 , and ρ is a damping factor, such that $0 \leq \rho \leq 1$. The model is stationary if ρ is strictly

less than one, and if λ_c is equal to 0 or π it reduces to a first-order autoregressive process.

A model of deterministic seasonality has the seasonal effects summing to zero over a year. The seasonal effects can be allowed to change over time by letting their sum over the previous year be equal to a random disturbance term ω_t , with mean zero and variance σ_ω^2 . Thus, if s is the number of seasons in the year,

$$\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t \quad \text{or} \quad \gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t. \quad (1.7)$$

An alternative way of allowing seasonal dummy variables to change over time is to suppose that each season evolves as a random walk but that, at any particular point in time, the seasonal components, and hence the disturbances, sum to zero. This model was introduced by Harrison and Stevens (1976, p. 217–218).

A seasonal pattern can also be modelled by a set of trigonometric terms at the seasonal frequencies, $\lambda_j = 2\pi j/s$, $j = 1, \dots, [s/2]$, where $[s/2]$ is $s/2$ if s is even and $(s-1)/2$ if s is odd. The seasonal effect at time t is then

$$\gamma_t = \sum_{j=1}^{[s/2]} (\gamma_j \cos \lambda_j t + \gamma_j^* \sin \lambda_j t). \quad (1.8)$$

When s is even, the sine term disappears for $j = s/2$ and so the number of trigonometric parameters, the γ_j and γ_j^* , is always $(s-1)/2$, which is the same as the number of coefficients in the seasonal dummy formulation. A seasonal pattern based on (1.8) is the sum of $[s/2]$ cyclical components, each with $\rho = 1$, and it may be allowed to evolve over time in exactly the same way as a cycle was allowed to move. The model is

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t}, \quad (1.9)$$

where, following (1.6),

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} = \begin{pmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix}, \quad (1.10)$$

where ω_{jt} and ω_{jt}^* , $j = 1, \dots, [s/2]$, are zero mean white noise processes which are uncorrelated with each other with a common variance σ_ω^2 . As in the cycles (1.6) $\gamma_{j,t}^*$ appears as a matter of construction, and its interpretation is not particularly important. Note that when s is even, the component at $j = s/2$ collapses to

$$\gamma_{j,t} = \gamma_{j,t-1} \cos \lambda_j + \omega_{jt}. \quad (1.11)$$

If the disturbances in the model are assumed to be normally distributed, the

Table 1
Principal structural time series components and models

Model	Component	Specification	Stationarity operator, $\Delta(L)$	Reduced form	Comments
(A) Local level/ random walk plus noise model	(1a) Random walk	$\mu_t = \mu_{t-1} + \eta_t$	$\Delta = 1 - L$	—	
	(1b) Random walk with drift	$\mu_t = \mu_{t-1} + \beta + \eta_t$	Δ		
(B) Local linear trend	(2) Stochastic trend	$y_t = \mu_t + \varepsilon_t$, with μ_t as in (1a)	Δ	ARIMA(0, 1, 1)	Forecast function is EWMA.
		$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$, $\beta_t = \beta_{t-1} + \zeta_t$	Δ^2	Random walk with drift if $\sigma_\zeta^2 = 0$. Doubly integrated random walk if $\sigma_\eta^2 = 0$.	
(C) Cycle plus noise model	(3) Stochastic cycle	$y_t = \mu_t + \varepsilon_t$, with μ_t as in (2)	Δ^2	ARIMA(0, 2, 2)	Forecast function in nonseasonal Hold-Winters.
		$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}$, where ψ is cycle, $0 \leq \rho < 1$, and $0 \leq \lambda_c \leq \pi$	1	ARMA(2, 1) ARMA(2, 2)	Collapses to AR(1) if $\lambda_c = 0$ or π .
(D) Trend plus cycle		$y_t = \mu + \psi_t + \varepsilon_t$, where $0 \leq \rho < 1$	1	ARIMA(2, 2, 4)	Constant + ARMA(2, 2)
		$y_t = \mu_t + \psi_t + \varepsilon_t$, with μ_t as in (2)	Δ^2	ARIMA(2, 2, 4)	ARIMA(2, 2, 4)
(E) Cyclical trend		$y_t = \mu_t + \varepsilon_t$, $\mu_t = \mu_{t-1} + \psi_{t-1} + \beta_{t-1} + \eta_t$, with β_t as in (2)	Δ^2	ARIMA(2, 2, 4)	Gives the damped trend model if $\lambda_c = 0$ and β_t is removed.
		As (3) but $\rho = 1$	$1 - 2 \cos \lambda_c L + L^2$	—	—
(F) Basic structural model	(4) Nonstationary cycle	$y_t = \sum_{j=1}^{[s/2]} \gamma_{t-j} + \omega_j$	$S(L) = 1 + L + L^2 + \dots + L^{s-1}$	$S(L)y_t \sim WN$	Evolves more smoothly than (5a).
	(5a) Dummy variable seasonality	$y_t = \sum_{j=1}^{[s/2]} \gamma_{t-j}^\dagger$	$S(L)y_t \sim MA(s-2)$		
(5b) Trigonometric seasonality		$y_t = \sum_{j=1}^{[s/2]} \gamma_{t,j}^\dagger$	$S(L)y_t \sim MA(s-2)$		
		where $\gamma_{t,j}^\dagger$ is a nonstationary cycle, (4), with $\lambda_c = \lambda_j = 2\pi j/s$, $j = 1, 2, \dots, [s/2]$			
(F) Basic structural model	(5a) or (5b)	$y_t = \mu_t + \gamma_t + \varepsilon_t$, where μ_t is as in (2) and γ_t is as in (5a) or (5b)	$\Delta_s = (1 - L)(1 - L^s)$	$\Delta \Delta_s y_t \sim MA(s+1)$	(a) Forecasts from Holt-Winters are similar. (b) Close to 'airline' model for some series.

hyperparameters ($\sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2, \sigma_\kappa^2, \rho, \lambda_c, \sigma_\varepsilon^2$) may be estimated by maximum likelihood. This may be done in the time domain using the Kalman filter as described in Section 2, or in the frequency domain as described in Harvey (1989, Chapter 4). Harvey and Peters (1990) present simulation evidence on the performance of different estimators. Once the hyperparameters have been estimated, the state space form may be used to make predictions and construct estimators of the unobserved components.

EXAMPLE. A model of the form (1.1), but without the seasonal component, was fitted to quarterly, seasonally adjusted data on US GNP from 1947Q1 to 1988Q2. The estimated variances of $\eta_t, \zeta_t, \kappa_t$, and ε_t were 0, 0.0015, 0.0664 and 0 respectively, while the estimate of ρ was 0.92. The estimate of λ_c was 0.30, corresponding to a period of 20.82 quarters. Thus the length of business cycles is roughly five years.

A summary of the main structural models and their properties may be found in Table 1. Structural time series models which are linear and time invariant, all have a corresponding *reduced form* autoregressive integrated moving average (ARIMA) representation which is equivalent in the sense that it will give identical forecasts to the structural form. For example in the local level model,

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \\ \mu_t &= \mu_{t-1} + \eta_t, \end{aligned} \tag{1.12}$$

where ε_t and η_t are mutually uncorrelated white noise disturbances, taking first differences yields

$$\Delta y_t = \eta_t + \varepsilon_t - \varepsilon_{t-1}, \tag{1.13}$$

which in view of its autocorrelation structure is equivalent to an MA(1) process with a nonpositive autocorrelation at lag one. Thus y_t is ARIMA(0, 1, 1). By equating autocorrelations at lag one it is possible to derive the relationship between the moving average parameter and q , the ratio of the variance of η_t to that of ε_t . In more complex models, there may not be a simple correspondence between the structural and reduced form parameters. For example in (1.1), $\Delta\Delta_s y_t$ is ARMA(2, $s+3$), where Δ_s is the seasonal difference operator. Note that the terminology of reduced and structural form is used in a parallel fashion to the way it is used in econometrics, except that in structural time series models the restrictions come not from economic theory, but from a desire to ensure that the forecasts reflect features such as cycles and seasonals which are felt to be present in the data.

In addition to the main structural models found in Table 1 many more structural models may be constructed. Additional components may be introduced and the components defined above may be modified. For example, quadratic trends may replace linear ones, and the irregular component may be

formulated so as to reflect the sampling scheme used to collect the data. If observations are collected on a daily basis, a slowly changing day of the week effect may be incorporated in the model, while for hourly observations an intra-day pattern may be modelled in a similar way to seasonality. A more parsimonious way of modelling an intra-day pattern, based on time-varying splines, is proposed in Harvey and Koopman (1993).

1.2. Model selection

The most difficult aspect of working with time series data is model selection. The attraction of the structural framework is that it enables the researcher to formulate, at the outset, a model which is explicitly designed to pick up the salient characteristics of the data. Once the model has been estimated, its suitability can be assessed, not only by carrying out diagnostic tests, but also by checking whether the estimated components are consistent with any prior knowledge which might be available. Thus if a cyclical component is used to model the trade cycle, a knowledge of the economic history of the period should enable one to judge whether the estimated parameters are reasonable. This is in the same spirit as assessing the plausibility of a regression model by reference to the sign and magnitude of its estimated coefficients.

Classical time series analysis is based on the theory of stationary stochastic processes, and this is the starting point for conventional time series model building. Nonstationarity is handled by differencing, leading to the ARIMA class of models. The fact that the simpler structural time series models can be made stationary by differencing provides an important link with classical time series analysis. However, the analysis of series which are thought to be stationary does not play a fundamental role in structural modelling methodology. Few economic and social time series are stationary and there is no overwhelming reason to suppose that they can necessarily be made stationary by differencing, which is the assumption underling the ARIMA methodology of Box and Jenkins (1976). If a univariate structural model fails to give a good fit to a set of data, other univariate models may be considered, but there will be an increased willingness to look at more radical alternatives. For example, a search for outliers might be initiated or it may be necessary to concede that a structurally stable model can only be obtained by conditioning on an observed explanatory variable.

Introducing explanatory variables into a model requires access to a larger information set. Some prior knowledge of which variables should potentially enter into the model is necessary, and data on these variables is needed. In a structural time series model the explanatory variables enter into the model side by side with the unobserved components. In the absence of these unobserved components the model reverts to a regression, and this perhaps makes it clear as to why the model selection methodology which has been developed for dynamic regression is appropriate in the wider context with which we are concerned. Distributed lags can be fitted in much the same way as in

econometric modelling, and even ideas such as the error-correction mechanism can be employed. The inclusion of the unobserved time series components does not affect the model selection methodology to be applied to the explanatory variables in any fundamental way. What it does is to add an extra dimension to the interpretation and specification of certain aspects of the dynamics. For example, it provides a key insight into the vexed question of whether to work with the variables in levels or first differences, and solves the problem by setting up a general framework within which the two formulations emerge as special cases.

The fact that structural time series models are set up in terms of components which have a direct interpretation means that it is possible to employ a model selection methodology which is similar to that proposed in the econometrics literature by writers such as Hendry and Richard (1983). Thus one can adopt the following criteria for a good model: parsimony, data coherence, consistency with prior knowledge, data admissibility, structural stability and encompassing.

2. Linear state space models and the Kalman filter

The linear state space form has been demonstrated to an extremely powerful tool in handling all linear and many classes of nonlinear time series models; see Harvey (1989, Chapters 3 and 4). In this section we introduce the state space form and the associated Kalman filter. We show how the filter can be used to deliver the likelihood. Recent work on smoothing is also discussed.

2.1. The linear state space form

Suppose a multivariate time series y_t possesses N elements. This series is related to a $p \times 1$ vector α_t , which labelled the state, via the measurement equation

$$y_t = Z_t \alpha_t + X_t \beta + \varepsilon_t, \quad t = 1, \dots, T. \quad (2.1)$$

Here Z_t and X_t are nonstochastic matrices of dimensions $N \times p$ and $N \times k$ respectively, β is a fixed k -dimensional vector and ε_t is a zero mean, $N \times 1$ vector of white noise, with variance H_t .

The measurement equation is reminiscent of a classical regression model, with the state vector representing some of the regression coefficients. However, in the state space form, the state vector is allowed to evolve over time. This is achieved by introducing a transition equation, which is given by

$$\alpha_t = T_t \alpha_{t-1} + W_t \beta + R_t \eta_t, \quad t = 1, \dots, T, \quad (2.2)$$

where T_t , W_t and R_t are fixed matrices of size $(p \times p)$, $(p \times k)$ and $(p \times g)$ respectively, η_t is a zero mean and g -dimensional vector of white noise, with

variance Q_t . In the literature η_t and ε_t have always been assumed to be uncorrelated for all $s \neq t$. In this paper we will also assume that η_t and ε_t are uncorrelated, although Anderson and Moore (1979) and more recently De Jong (1991) and Koopman (1993) relax this assumption.

The inclusion of the R_t matrix is somewhat arbitrary, for the disturbance term can always be redefined to have a variance $R_t Q_t R_t'$. However, the transition equation above is often regarded as being more natural. The transition equation involves the state at time zero and so to complete the state space form we need to tie down its behaviour. We assume that α_0 has a mean a_0 and variance P_0 . Further, α_0 is assumed to be uncorrelated with the noise in the transition and measurement equations. This completed state space form is said to be time invariant if Z_t , X_t , H_t , W_t , R_t and Q_t do not change over time.

To illustrate these general points we will put the univariate structural model (1.1) of trends, seasonals and cycles discussed in Section 1 into time invariant state space form by writing $\alpha_t = (\mu_t, \beta_t, \delta_t, \delta_{t-1}, \dots, \delta_{t-s+2}, \psi_t, \psi_t^*)'$, where

$$y_t = (1 \ 0 \ 1 \ 0 \ 0 \ \cdots \ 0 \ 1 \ 0) \alpha_t + \varepsilon_t, \quad (2.3a)$$

$$\alpha_t = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 & \cdots & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \rho \cos \lambda_c & \rho \sin \lambda_c \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -\rho \sin \lambda_c & \rho \cos \lambda_c \end{pmatrix} \alpha_{t-1}$$

$$+ \begin{pmatrix} \eta_t \\ \zeta_t \\ \omega_t \\ 0 \\ 0 \\ \vdots \\ 0 \\ \kappa_t \\ \kappa_t^* \end{pmatrix}. \quad (2.3b)$$

2.2. The Kalman filter

In most structural time series models the individual elements of α_t are unobservable, either because of the presence of some masking irregular term ε_t or because of the way α_t is constructed. However, the observations do carry some information which can be harnessed to provide estimates of the unknownable α_t . This estimation can be carried out using a variety of information sets. We will write Y_s to denote this information, which will be composed of all the observations recorded up to time s and our initial knowledge of α_0 . The two

most common forms of estimation are smoothing, where we estimate α_t using Y_T , and filtering, where we estimate using only Y_t . We will focus on various aspects of smoothing in Section 2.4, but here we look at filtering.

Filtering allows the tracking of the state using contemporaneously available information. The optimal, that is minimum mean square error, filter is given by the mean of the conditional density of α_t , given Y_t , which is written as $\alpha_t|Y_t$. The Kalman filter delivers this quantity if the observations are Gaussian. If they are non-Gaussian the Kalman filter provides the optimal estimator amongst the class of linear estimators. Here we develop the filter under Gaussianity; see Duncan and Horn (1972) for an alternative derivation.

We start at time zero with the knowledge that $\alpha_0 \sim N(a_0, P_0)$. If we combine the transition and measurement equations with this prior and write Y_0 to express the information in it, then

$$\begin{pmatrix} \alpha_1 \\ y_1 \end{pmatrix} | Y_0 \sim N \left(\begin{pmatrix} a_{1|0} \\ Z_1 a_{1|0} + X_1 \beta \end{pmatrix}, \begin{pmatrix} P_{1|0} & P_{1|0} Z'_1 \\ Z_1 P_{1|0} & F_1 \end{pmatrix} \right), \quad (2.4)$$

where

$$\begin{aligned} P_{1|0} &= T_1 P_0 T'_1 + R_1 Q_1 R'_1, & F_1 &= Z_1 P_{1|0} Z'_1 + H_1, \\ a_{1|0} &= T_1 a_0 + W_1 \beta. \end{aligned} \quad (2.5)$$

Usually, we write $v_1 = y_1 - Z_1 \alpha_{1|0} - X_1 \beta$ as the one-step ahead forecast error. It is constructed so that $v_1 | Y_0 \sim N(0, F_1)$. Consequently, using the usual conditioning result for multivariate normal distributions as given, for example, in Rao (1973)

$$\alpha_1 | Y_1 \sim N(a_1, P_1), \quad (2.6)$$

where

$$a_1 = a_{1|0} + P_{1|0} Z'_1 F_1^{-1} v_1, \quad P_1 = P_{1|0} - P_{1|0} Z'_1 F_1^{-1} Z_1 P_{1|0}. \quad (2.7)$$

This result means that the filter is recursive. We will use the following notation throughout the paper to describe the general results: $\alpha_{t-1} | Y_{t-1} \sim N(a_{t-1}, P_{t-1})$, $\alpha_t | Y_{t-1} \sim N(a_{t|t-1}, P_{t|t-1})$ and $v_t | Y_{t-1} \sim N(0, F_t)$. The precise definition of these densities is given in the following three sets of equations. First the prediction equations

$$a_{t|t-1} = T_t a_{t-1} + W_t \beta, \quad P_{t|t-1} = T_t P_{t-1} T'_t + R_t Q_t R'_t, \quad (2.8)$$

then the one-step ahead forecast equations

$$v_t = y_t - Z_t a_{t|t-1} - X_t \beta, \quad F_t = Z_t P_{t|t-1} Z'_t + H_t, \quad (2.9)$$

and finally the updating equations

$$\begin{aligned} a_t &= a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} v_t \quad \text{and} \quad P_t = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}. \end{aligned} \quad (2.10)$$

One immediate result which follows from the Kalman filter is that we can write the conditional joint density of the observations as

$$f(y_1, \dots, y_T | Y_0) = \prod_{t=1}^T f(y_t | Y_{t-1}) = \prod_{t=1}^T f(v_t | Y_{t-1}). \quad (2.11)$$

This fracturing of the conditional joint density into the product of conditionals is called the prediction error decomposition. If α_t is stationary, an unconditional joint density can be constructed since the initial conditions, a_0 and P_0 , are known. The case where we do not have stationarity has been the subject of some interesting research in recent years.

2.3. Initialization for non-stationary models¹

We will derive the likelihood for a model in state space form using the argument in De Jong (1988a). A slightly different approach can be found in Ansley and Kohn (1985). We present a simplified derivation based partly on the results in Marshall (1992a). For ease of exposition β will be assumed to be zero and all the elements in α_0 can be taken to be nonstationary. We start by noting that if we write $y = (y'_1, y'_2, \dots, y'_T)'$, then

$$f(y) = \frac{f(\alpha_0 = 0)f(y | \alpha_0 = 0)}{f(\alpha_0 = 0 | y)}. \quad (2.12)$$

The density $f(y | \alpha_0 = 0)$ can be evaluated by applying the Kalman filter and the prediction error decomposition if we initialize the filter at $a_0 = 0$ and $P_0 = 0$. We denote this filter by KF(0, 0), and the corresponding output as a_t^* , $a_{t|t-1}^*$ and v_t^* . The density $f(\alpha_0 = 0)$ has a simple form, which leaves us with the problem of $f(\alpha_0 = 0 | y)$. If we write $v^* = (v_1^*, v_2^*, \dots, v_T^*)'$, then we can use the result that v^* is a linear combination of y in order to write $f(\alpha_0 = 0 | y) = f(\alpha_0 = 0 | v^*)$. To be able to evaluate $f(\alpha_0 = 0 | v^*)$ we will need to define F as a block diagonal matrix, with blocks being F_t and A as a matrix with row blocks $Z_t G_{t-1}$, where $G_t = T_{t+1}(I - K_t Z_t)G_{t-1}$, $G_0 = T_1$, and $K_t = P_{t|t-1} Z_t' F_t^{-1}$ (the so-called Kalman gain). In all cases the quantities are evaluated by the Kalman filter under the startup condition $P_0 = 0$. Then as

$$v_t(\alpha_0) = y_t - E(y_t | Y_{t-1}, \alpha_0) = y_t - Z_t a_{t|t-1}(\alpha_0), \quad (2.13)$$

¹ The rest of Section 2 is more technical and can be omitted on first reading without loss of continuity.

and

$$\begin{aligned} a_{t+1|t}(\alpha_0) &= T_{t+1}a_{t|t-1}(\alpha_0) + T_{t+1}K_t v_t(\alpha_0) \\ &= T_{t+1}(I - K_t Z_t)a_{t|t-1}(\alpha_0) + T_{t+1}K_t y_t \\ &= G_t \alpha_0 + a_{t+1|t}^*, \end{aligned} \quad (2.14)$$

we have that

$$\begin{aligned} v_t(\alpha_0) &= y_t - Z_t G_{t-1} \alpha_0 - Z_t a_{t|t-1}^* \\ &= v_t^* - Z_t G_{t-1} \alpha_0, \end{aligned} \quad (2.15)$$

and so

$$v^* | \alpha_0 \sim N(A\alpha_0, F). \quad (2.16)$$

Thus we can use Bayes' theorem to deliver the result that

$$\alpha_0 | y \sim N((P_0^{-1} + S_T)^{-1}(P_0^{-1}a_0 + s_T), (P_0^{-1} + S_T)^{-1}), \quad (2.17)$$

where

$$S_T = A'F^{-1}A \quad \text{and} \quad s_T = A'F^{-1}v^*. \quad (2.18)$$

S_T and s_T can be computed recursively, in parallel with KF(0, 0), by

$$S_t = S_{t-1} + G_{t-1}'Z_t'F_t^{-1}Z_tG_{t-1} \quad \text{and} \quad s_t = s_{t-1} + G_{t-1}'Z_t'F_t^{-1}v_t^*, \quad (2.19)$$

with $s_0 = 0$ and $S_0 = 0$; see De Jong (1991). The log-likelihood is then constructed as

$$\begin{aligned} l(y) &= -\frac{1}{2} \log |P_0| - \frac{1}{2} a_0' P_0^{-1} a_0 - \frac{1}{2} \sum_{t=1}^T \log |F_t| - \frac{1}{2} \sum_{t=1}^T v_t'^* F_t^{-1} v_t^* \\ &\quad - \frac{1}{2} \log |P_0^{-1} + S_T| + \frac{1}{2} (s_T + P_0^{-1}a_0)'(P_0^{-1} + S_T)^{-1}(s_T + P_0^{-1}a_0). \end{aligned} \quad (2.20)$$

Traditionally, nonstationary state space models have been initialised into two ways. The first is to use a diffuse prior on $\alpha_0 | Y_0$; this is to allow the diagonal elements of P_0 to go to infinity. We can see that in the limit the result from this is that

$$\begin{aligned} l(y) + \frac{1}{2} \log |P_0| &\rightarrow -\frac{1}{2} \sum_{t=1}^T \log |F_t| - \frac{1}{2} \sum_{t=1}^T v_t'^* F_t^{-1} v_t^* \\ &\quad - \frac{1}{2} \log |S_T| - \frac{1}{2} s_T' S_T^{-1} s_T \\ &= -\frac{1}{2} \log |S_T| - \frac{1}{2} \log |F| \\ &\quad - \frac{1}{2} v^*'(F^{-1} - F^{-1}A(A'F^{-1}A)^{-1}A'F^{-1})v^*. \end{aligned} \quad (2.21)$$

An approximation to (2.21) can be obtained for many models by running

KF(a_0, P_0) with the diagonal elements of P_0 set equal to large, but finite, values. The likelihood is then constructed from the prediction errors once enough observations have been processed to give a finite variance. However, the likelihood obtained from (2.21) is preferable as it is exact and numerically stable.

The other main way nonstationary models are initialised is by taking α_0 to be an unknown constant; see Rosenberg (1973). Thus a_0 becomes a nuisance parameter and P_0 is set to zero. In this case, in the limit, the likelihood becomes

$$l(y) \rightarrow -\frac{1}{2} \sum_{t=1}^T \log |F_t| - \frac{1}{2} \sum_{t=1}^T v_t^{*'} F_t^{-1} v_t^* + a_0' S_T - \frac{1}{2} a_0' S_T a_0, \quad (2.22)$$

$$= -\frac{1}{2} \log |F| - \frac{1}{2} (v^* - Aa_0)' F^{-1} (v^* - Aa_0), \quad (2.23)$$

the term $a_0' S_T a_0$ in (2.22) appearing when $(P_0^{-1} + S_T)^{-1}$ is expanded out. We can concentrate a_0 out at its maximum likelihood value $\hat{a}_0 = (A' F^{-1} A)^{-1} A' F^{-1} v^*$, to deliver the profile or concentrated likelihood function

$$c(y) = -\frac{1}{2} \log |F| - \frac{1}{2} v^{*'} (F^{-1} - F^{-1} A (A' F^{-1} A)^{-1} A' F^{-1}) v^*. \quad (2.24)$$

The difference between the profile likelihood and the likelihood given in (2.21) is simply the $\log |S_T|$ term. The latter is called a marginal or restricted likelihood in the statistics literature; cf. McCullagh and Nelder (1989, Chapter 7). It is based on a linear transformation of y making the data invariant to α_0 .

The term $\log |S_T|$ can have a significant effect on small sample properties of maximum likelihood (ML) estimators in certain circumstances. This can be seen by looking at some results from the paper by [Shephard and Harvey \(1990\)](#) which analyses the sampling behaviour of the ML estimator of q , the ratio of the variances of η_t and ε_t , in the local level model (1.12). When q is zero the reduced form of the local level model is strictly noninvertible. Evaluating the probability that q is estimated to be exactly zero for various true values of q

Table 2
Probability that ML estimator of signal–noise ratio q is exactly equal to zero

<i>Marginal likelihood</i>					
<i>T</i> – 1	$q = 0$	$q = 0.01$	$q = 0.1$	$q = 1$	$q = 10$
10	0.64	0.61	0.47	0.21	0.12
30	0.65	0.49	0.18	0.03	0.01
50	0.65	0.35	0.07	0.01	0.00

<i>Profile likelihood</i>					
<i>T</i> – 1	$q = 0$	$q = 0.01$	$q = 0.1$	$q = 1$	$q = 10$
10	0.96	0.95	0.88	0.60	0.44
30	0.96	0.87	0.49	0.20	0.13
50	0.96	0.72	0.28	0.08	0.05

and sample sizes, gives the results summarised in Table 2. It can be seen that using a profile likelihood instead of a marginal results in a much higher probability of estimating q to be zero. Unless q is actually zero, this is undesirable from a forecasting point of view since there is no discounting of past observations. This provides a practical justification for the use of diffuse initial conditions and marginal likelihoods.

2.4. Smoothing

Estimating α_t using the full set of observations Y_T is called smoothing. The minimum mean square estimator of α_t using Y_T is $E \alpha_t | Y_T$. An extensive review of smoothing is given in Anderson and Moore (1979, Chapter 7).

Recently there have been some important developments in the way $E \alpha_t | Y_T$ is obtained; see, for example, De Jong (1988b, 1989), Kohn and Ansley (1989) and Koopman (1993). These breakthroughs have dramatically improved the speed of the smoothers. The new results will be introduced by using the framework of Whittle (1991). For ease of exposition, R_t will be assumed to be an identity matrix and β will be assumed to be zero.

Under Gaussianity, $E \alpha_t | Y_T$ is also the mode of the density of $\alpha_t | Y_T$. Thus we can use the general result that under weak regularity, if $f(\cdot)$ is a generic density function and m denotes the mode, then

$$\frac{\partial f(x|z)}{\partial x} \Big|_{x=m} = 0 \text{ if and only if } \frac{\partial f(x,z)}{\partial x} \Big|_{x=m} = 0. \quad (2.25)$$

The smoother can therefore be found by searching for turning points in the joint density of $\alpha'_1, \alpha'_2, \dots, \alpha'_T, y'_1, \dots, y'_T$, the logarithm of which is

$$\begin{aligned} D = & \text{constant} - \frac{1}{2} (\alpha_0 - a_0)' P_0^{-1} (\alpha_0 - a_0) \\ & - \frac{1}{2} \sum_{t=1}^T (y_t - Z_t \alpha_t)' H_t^{-1} (y_t - Z_t \alpha_t) \\ & - \frac{1}{2} \sum_{t=1}^T (\alpha_t - T_t \alpha_{t-1})' Q_t^{-1} (\alpha_t - T_t \alpha_{t-1}). \end{aligned} \quad (2.26)$$

Thus

$$\frac{\partial D}{\partial \alpha_t} = Z_t' H_t^{-1} \varepsilon_t - Q_t^{-1} \eta_t + T_{t+1}' Q_{t+1}^{-1} \eta_{t+1} \quad \text{for } t = 1, \dots, T. \quad (2.27)$$

Equating to zero, writing the solutions as $\hat{\alpha}_t$ and $\hat{\varepsilon}_t = y_t - Z_t \hat{\alpha}_t$ and $\hat{\eta}_t = \hat{\alpha}_t - T_t \hat{\alpha}_{t-1}$ results in the backward recursion

$$\begin{aligned} \hat{\alpha}_{t-1} &= T_t^{-1} (\hat{\alpha}_t - Q_t (Z_t' H_t^{-1} \hat{\varepsilon}_t + T_{t+1}' Q_{t+1}^{-1} \hat{\eta}_{t+1})) \\ &= T_t^{-1} (\hat{\alpha}_t - \hat{\eta}_t), \quad t = 1, \dots, T, \end{aligned} \quad (2.28)$$

as

$$Z_t' H_t^{-1} \hat{\varepsilon}_t - Q_t^{-1} \hat{\eta}_t + T_{t+1}' Q_{t+1}^{-1} \hat{\eta}_{t+1} = 0. \quad (2.29)$$

The starting point $\hat{\alpha}_T = a_T$ is given by the Kalman filter. Unfortunately, using

$$\hat{\eta}_t = Q_t (T_{t+1}' Q_{t+1}^{-1} \hat{\eta}_{t+1} + Z_t' H_t^{-1} \hat{\varepsilon}_t), \quad (2.30)$$

will lead to a numerically unstable filter even though mathematically this result holds exactly. Koopman's (1993) shows that it can be stabilised by computing $\hat{\varepsilon}_t$ not by $y_t - Z_t \hat{\alpha}_t$, but by

$$\hat{\varepsilon}_t = H_t (F_t^{-1} v_t - K_t' T_{t+1}' Q_{t+1}^{-1} \hat{\eta}_{t+1}), \quad (2.31)$$

where F_t and K_t are computed using KF(0, 0) and $v_t = v_t^* - Z_t G_{t-1} S_T^{-1} s_T$. Thus the efficient smoother uses (2.28), (2.30) and (2.31).

Recently, Harvey and Koopman (1992) have proposed using the smoothed estimates of ε_t and η_t to check for outliers and structural breaks, while Koopman (1993) uses them to implement a rapid EM algorithm and [Koopman and Shephard \(1992\)](#) show how to construct the exact score by smoothing.

3. Explanatory variables

Stochastic trend components are introduced into dynamic regression models when the underlying level of a nonstationary dependent variable cannot be completely explained by observable explanatory variables. The presence of a stochastic trend can often be rationalised by the fact that a variable has been excluded from the equation because it is difficult, or even impossible, to measure. Thus in Harvey et al. (1986) and [Slade \(1989\)](#), a stochastic trend is used as a proxy for technical progress, while in the demand equation for UK spirits estimated by [Ansley and Kohn \(1989\)](#) the stochastic trend can be thought of as picking up changes in tastes. Such rationalisation not only lends support to the specification of the model, but it also means that the estimated stochastic trend can be analysed and interpreted.

If stochastic trends are appropriate, but are not explicitly modelled, their effects will be picked up indirectly by time trends and lags on the variables. This can lead to a proliferation of lags which have no economic meaning, and which are subject to common factors and problems of inference associated with unit roots. An illustration of the type of problems which can arise with such an approach in a single equation context can be found in Harvey et al. (1986), where a stochastic trend is used to model productivity effects in an employment output equation and is compared with a more traditional autoregressive distributed lag (ADL) regression model with a time trend. Such problems may become even more acute in multivariate systems, such as vector autoregressions and simultaneous equation models; see Section 5.

Other stochastic components, such as time-varying seasonals or cycles, can

also be included in a model with explanatory variables. Since this raises no new issues of principle, we will concentrate on stochastic trends.

3.1. Formulation and estimation

A regression model with a stochastic trend component may be written

$$y_t = \mu_t + x_t' \delta + \varepsilon_t, \quad t = 1, \dots, T, \quad (3.1)$$

where μ_t is a stochastic trend (1.4), x_t is a $k \times 1$ vector of exogenous explanatory variables, δ is a corresponding vector of unknown parameters, ε_t is a normally distributed, white noise disturbance term with mean zero and variance σ_ε^2 . A standard regression model with a deterministic time trend emerges as a special case, as does a model which could be estimated efficiently by OLS regression in first differences; in the latter case $\sigma_\varepsilon^2 = \sigma_\zeta^2 = 0$.

In the reduced form of (3.1), the stochastic part, $\mu_t + \varepsilon_t$, is replaced by an ARIMA(0, 2, 2) process. If the slope is deterministic, that is $\sigma_\zeta^2 = 0$ in (1.3), it is ARIMA(0, 1, 1). Box and Jenkins (1976, pp. 409–412) report a distributed lag model fitted to first differences with an MA(1) disturbance term. This model can perhaps be interpreted more usefully as a relationship in levels with a stochastic trend component of the form

$$\mu_t = \mu_{t-1} + \beta + \eta_t. \quad (3.2)$$

Maximum likelihood estimators of the parameters in (3.1) can be constructed in the time domain *via* the prediction error decomposition. This is done by putting the model in state space form and applying the Kalman filter. The parameters δ and β can be removed from the likelihood function either by concentrating them out of form of a profile likelihood function as in Kohn and Ansley (1985) or by forming a marginal likelihood function; see the discussion in Section 2.3. The marginal likelihood can be computed by extending the state so as to include β and δ , even though they are time-invariant, and then initializing with a diffuse prior.

The difference between the profile and marginal likelihood is in the determinantal term of the likelihood. There are a number of arguments which favour the use of marginal likelihoods for inference in small samples or when the process is close to nonstationarity or noninvertibility; see Tunnicliffe-Wilson (1989). In the present context, the difference in behaviour shows up most noticeably in the tendency of the trend to be estimated as being deterministic. To be more specific, suppose the trend is as in (3.2). The signal-noise ratio is $q = \sigma_n^2 / \sigma_\varepsilon^2$ and if this is zero the trend is deterministic. The probability that q is estimated to be zero has been computed by Shephard (1993a). Using a profile likelihood by concentrating out β leads to this probability being relatively high when q is small but nonzero. The properties of the estimator obtained from the marginal likelihood are much better in this respect.

3.2. Intervention analysis

Intervention analysis is concerned with making inferences about the effects of known events. These effects are measured by including intervention, or dummy, variables in a dynamic regression model. In pure intervention analysis no other explanatory variables are present.

Model (3.1) may be generalized to yield the intervention model

$$y_t = \mu_t + x_t' \delta + \lambda w_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (3.3)$$

where w_t is the intervention variable and λ is its coefficient. The definition of w_t depends on the form which the intervention effect is assumed to take. If the intervention is transitory and has an effect only at time t , $t = \tau$, and is zero otherwise. More generally the intervention may have a transitory effect which dies away gradually, for example, we may have $w_t = \varphi^{t-\tau}$, when $|\varphi| < 1$, for $t \geq \tau$. A permanent shift in the level of the series can be captured by a step variable which is zero up to the time of the intervention and unity thereafter. An effect of this kind can also be interpreted as a transitory shock to the level equation in the trend, in which case it appears as a pulse variable in (1.4a). Other types of intervention variable may be included, for example variables giving rise to changes in the slope of the trend or the seasonal pattern. The advantage of the structural time series model framework over the ARIMA framework proposed by Box and Tiao (1975) is that it is much easier to formulate intervention variables having the desired effect on the series.

Estimation of a model of the form (3.3) can be carried out in both the time and frequency domains by treating the intervention variable just like any other explanatory variable. In the time domain, various tests can be constructed to check on the specification of the intervention; see the study by [Harvey and Durbin \(1986\)](#) on the effect of the UK seat belt law.

4. Multivariate time series models

4.1. Seemingly unrelated time series equations (SUTSE)

The structural time series models introduced in Section 1 have straightforward multivariate generalisations. For instance, the local level with drift becomes, for an N -dimensional series $y_t = (y_{1t}, \dots, y_{Nt})'$,

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \Sigma_\varepsilon), \\ \mu_t &= \mu_{t-1} + \beta + \eta_t, \quad \eta_t \sim \text{NID}(0, \Sigma_\eta), \end{aligned} \quad (4.1)$$

where Σ_ε and Σ_η are nonnegative definite $N \times N$ matrices. Such models are called seemingly unrelated time series equations (SUTSE) reflecting the fact that the individual series are connected only via the correlated disturbances in

the measurement and transition equations. Estimation is discussed in Fernández (1990).

The maximisation of the likelihood function for this model can be computationally demanding if N is large. The evaluation of the likelihood function requires $O(N^3)$ floating point operations and, although β can be concentrated, there are still $N \times (N + 1)$ parameters to be estimated by numerical optimisation. However, for many applications there are specific structures on Σ_ϵ and Σ_η that can be exploited to make the computations easier. One example is where Σ_ϵ and Σ_η are proportional, that is $\Sigma_\eta = q\Sigma_\epsilon$. Such a system is said to be homogeneous. This structure allows each of the series in y_t to be handled by the same Kalman filter and so the likelihood can be evaluated in $O(N)$ operations. Further, Σ_ϵ can be concentrated out of the likelihood, leaving a single parameter q to be found by numerical maximisation. The validity of the homogeneity assumption can be assessed by using the Lagrange multiplier test of Fernández and Harvey (1990).

4.2. Error components models

Consider the classical error components model

$$y_{it} = \mu + \lambda_i + v_t + \omega_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (4.2)$$

where μ represents the overall mean and λ_i , v_t and ω_{it} are unit specific and time specific effects respectively, assumed to be serially and mutually independent, Gaussian and with expected values equal to zero. The dynamic versions of this model studied in the literature usually include lagged dependent variables and autoregressive processes for the components v_t and ω_{it} ; see Anderson and Hsiao (1982).

A more natural approach to the specification of dynamic error components models, can be based on the ideas of structural time series models. This is suggested by Marshall (1992b), who allowed both time specific and time-unit specific effects to evolve over time according to random walk plus noise processes. The error components model becomes

$$\begin{aligned} y_{it} &= \mu_{it} + \varepsilon_t + \varepsilon_{it}^*, \\ \mu_{it} &= \mu_{i,t-1} + \eta_t + \eta_{it}^*, \end{aligned} \quad (4.3)$$

where μ_{it} is the mean for unit i at time t and ε_t , ε_{it}^* , η_t and η_{it}^* are assumed to be independent, zero mean, Gaussian random variables, with variances σ_ε^2 , $\sigma_{\varepsilon^*}^2$, σ_η^2 and $\sigma_{\eta^*}^2$ respectively. This model is a multivariate local level model, with the irregular and level random shocks decomposed as common effects, ε_t and η_t , and specific effects, ε_{it}^* and η_{it}^* . This means that

$$\Sigma_\varepsilon = \sigma_\varepsilon^2 I + \sigma_{\varepsilon^*}^2 \boldsymbol{\iota} \boldsymbol{\iota}' \quad \text{and} \quad \Sigma_\eta = \sigma_\eta^2 I + \sigma_{\eta^*}^2 \boldsymbol{\iota} \boldsymbol{\iota}', \quad (4.4)$$

where $\boldsymbol{\iota}$ is the N -dimensional unit vector and I the identity matrix.

If σ_η^2 and $\sigma_{\eta^*}^2$ are equal to zero, the model reduces to the static error components model discussed in (4.2). On the other hand if σ_η^2 is greater than zero, but $\sigma_{\eta^*}^2$ is equal to zero, the N time series have, apart from a time invariant effect, the same time-dependent mean. In this situation, the time series are cointegrated in the sense of Engle and Granger (1987).

Optimal estimates of the components μ_{it} can be obtained by means of the Kalman filter. That requires the manipulation of $N \times N$ matrices and so it becomes cumbersome if N is large. However, the idea of homogeneity can be used to reduce these calculations dramatically. Take for each time t the average of the observations across units and the first $N - 1$ deviations from this average. Thus, in an obvious notation, (4.3) becomes

$$\bar{y}_t = \bar{\mu}_t + \varepsilon_t + \bar{\varepsilon}_t^*, \quad (4.5a)$$

$$\begin{aligned} \bar{\mu}_t &= \bar{\mu}_{t-1} + \eta_t + \bar{\eta}_t^* \\ t &= 1, \dots, T, \end{aligned} \quad (4.5b)$$

$$(y_{it} - \bar{y}_t) = (\mu_{it} - \bar{\mu}_t) + (\varepsilon_{it}^* - \bar{\varepsilon}_t^*), \quad (4.6a)$$

$$\begin{aligned} (\mu_{it} - \bar{\mu}_t) &= (\mu_{i,t-1} - \bar{\mu}_{t-1}) + (\eta_{it}^* - \bar{\eta}_t^*), \\ i &= 1, \dots, N-1, \quad t = 1, \dots, T, \end{aligned} \quad (4.6b)$$

with the equations in (4.5) and (4.6) being statistically independent of one another. As the transformation to this model is nonsingular, the estimation of the trends μ_{it} can be obtained from this model instead of from the original error components model. The estimation of the average level can be carried out by running a univariate Kalman filter over the average values of the observations \bar{y}_t . The remaining $N - 1$ equations can be dealt with straightforwardly as they are a homogeneous system, with variances proportional to $(I - \boldsymbol{\mu}'\boldsymbol{\mu}/N)$, where I and $\boldsymbol{\mu}$ are now $N - 1$ -dimensional unit matrices and vectors.

The Kalman filter which provides the estimator of $\bar{\mu}_t$ using the information available up to time t is

$$\bar{m}_t = \bar{m}_{t-1} + \frac{\bar{p}_{t-1} + \sigma_\eta^2 + \sigma_{\eta^*}^2/N}{\bar{p}_{t-1} + \sigma_\eta^2 + \sigma_{\eta^*}^2/N + \sigma_\varepsilon^2 + \sigma_{\varepsilon^*}^2/N} (\bar{y}_t - \bar{m}_{t-1}), \quad (4.7)$$

where \bar{p}_t is the MSE of \bar{m}_t , given by

$$\bar{p}_t = (\bar{p}_{t-1} + \sigma_\eta^2 + \sigma_{\eta^*}^2/N) - \frac{(\bar{p}_{t-1} + \sigma_\eta^2 + \sigma_{\eta^*}^2)^2}{(\bar{p}_{t-1} + \sigma_\eta^2 + \sigma_{\eta^*}^2/N + \sigma_\varepsilon^2 + \sigma_{\varepsilon^*}^2/N)}. \quad (4.8)$$

These recursions are run from $t = 2$ and with initial values $\bar{m}_1 = y_1$ and $\bar{p}_1 = (\sigma_\varepsilon^2 + \sigma_{\varepsilon^*}^2/N)$. With respect to the formulae to obtain the estimators of the components $(\mu_{it} - \bar{\mu}_t)$ using the information up to time t , m_{it}^* and their MSEs, p_{it}^* , these have exactly the same form as (4.7) and (4.8) but with $(\sigma_\eta^2 + \sigma_{\eta^*}^2/N)$ and $(\sigma_\eta^2 + \sigma_{\eta^*}^2/N)$ replaced by $((N-1)\sigma_{\eta^*}^2/N)$ and $((N-1)\sigma_{\varepsilon^*}^2/N)$ respectively.

ly and with initial values $m_{i1} = (y_{i1} - \bar{y}_1)$ for $i = 1, \dots, N - 1$. The estimators of each μ_{it} , m_{it} , and its MSE, p_{it} , are given by

$$\begin{aligned} m_{it} &= \bar{m}_t + m_{it}^*, \quad i = 1, \dots, N - 1, \quad t = 1, \dots, T, \\ p_{it} &= \bar{p}_t + p_{it}^*, \quad i = 1, \dots, N - 1, \quad t = 1, \dots, T, \end{aligned} \quad (4.9)$$

while m_{Nt} is obtained by differencing.

EXAMPLE. In Marshall (1992b), a error components model of the form given above, but with a fixed slope as in (3.2), is estimated for the logarithm of the quarterly labour costs time series in Austria, Belgium, Luxembourg and The Netherlands. The sample period considered in 1970 to 1987 and so $N = 4$ and $T = 72$. The maximum likelihood estimates of the parameters were

$$\begin{aligned} \sigma_\epsilon^2 &= 0, \quad \sigma_{\epsilon^*}^2 = 0.115 \times 10^{-3}, \\ \sigma_\eta^2 &= 0.249 \times 10^{-3}, \quad \sigma_{\eta^*}^2 = 0.159 \times 10^{-3}. \end{aligned} \quad (4.10)$$

4.3. Explanatory variables in SUTSE models

The introduction of explanatory variables into the SUTSE model opens up the possibility of incorporating ideas from economic theory. This is well illustrated in the paper by [Harvey and Marshall \(1991\)](#) on the demand for energy in the UK. The assumption of a translog cost function leads to the static share equation system

$$s_i = \alpha_i + \sum_j \alpha_{ij} \log(p_j/\tau_j), \quad i = 1, \dots, N, \quad (4.11)$$

where the α_i , $i = 1, \dots, N$ and α_{ij} , $i, j = 1, \dots, N$, are parameters, s_i is the share of the i -th input, p_j is the (exogenous) price of the j -th input and τ_j is an index of relative technical progress for the input j which takes the factor augmenting form; see Jorgenson (1986).

The model can be made dynamic by allowing the $\log \tau_{jt}$, relative technical progress at time t for input j , to follow a random walk plus drift

$$\log \tau_{jt} = \log \tau_{j,t-1} + \bar{\beta}_j + \bar{\eta}_{jt}, \quad i = 1, \dots, N. \quad (4.12)$$

If the random disturbance term ε_{jt} is added to each share equation, this leads to a system of share equations which can be written in matrix form as

$$\begin{aligned} y_t &= \mu_t + Ax_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \Sigma_\varepsilon), \\ \mu_t &= \mu_{t-1} + \beta + \eta_t, \quad \eta_t \sim \text{NID}(0, \Sigma_\eta), \end{aligned} \quad (4.13)$$

where y_t is an $N \times 1$ vector of shares $(s_1, \dots, s_N)'$. Here μ_t is an $N \times 1$ vector depending on the α'_i , α'_{ij} and $\log \tau'_{it}$, so that the i -th element of μ_t is $\alpha_i + \sum \alpha_{ij} \log \tau_{jt}$, while A is an $N \times N$ matrix of α'_{ij} and x_t is the $N \times 1$ vector of the $\log p_{jt}$'s.

Harvey and Marshall (1991) estimated (4.13) under the assumption of statistical homogeneity, that is $\Sigma_\eta = q\Sigma_\epsilon$ and found this to be a reasonable assumption using the LM test referred to in Section 4.1. One equation was dropped to ensure that the shares summed to one. Finally restrictions from economic theory, concerning cost exhaustion, homogeneity and symmetry, were incorporated into the A matrix.

4.4. Common trends

Many economic variables seem to move together, indicating common underlying dynamics. This feature of data has been crystallised in the econometric literature as the concept of cointegration; see, for example Engle and Granger (1987) and Johansen (1988). Within a structural time series framework this feature can be imposed by modifying (4.1) so as to construct a common trends model

$$\begin{aligned} y_t &= \Theta\mu_t^* + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \Sigma_\varepsilon), \\ \mu_t^* &= \mu_{t-1}^* + \beta^* + \eta_t^*, \quad \eta_t^* \sim \text{NID}(0, \Sigma_{\eta^*}), \end{aligned} \quad (4.14)$$

where Θ is a $N \times K$ fixed matrix of factor loadings. The $K \times K$ matrix Σ_{η^*} is constrained to be a diagonal matrix and $\Theta_{ij} = 0$ for $j > i$, while $\Theta_{ii} = 1$ in order to achieve identifiability; see Harvey (1989, pp. 450–451). As Σ_{η^*} is diagonal, the common trends, the elements of μ_t^* , are independent.

The common trends model has $K \leq N$, but if $K = N$, it is equivalent to the SUTSE model, (4.1), with $\beta = \Theta\beta^*$ and $\Sigma_\eta = \Theta\Sigma_{\eta^*}\Theta'$ where Θ and Σ_{η^*} are the Cholesky decomposition of Σ_η . This suggests first estimating a SUTSE model and carrying out a principal components analysis on the estimated Σ_η to see what value of K accounts for a suitably large proportion of the total variation. A formal test can be carried out along the lines suggested by Stock and Watson (1988), but its small sample properties have yet to be investigated in this context. Once K has been determined, the common trends model can be formulated and estimated.

EXAMPLE. Tiao and Tsay (1989) fitted various multivariate models to the logarithms of indices of monthly flour prices in three cities, Buffalo, Minneapolis and Kansas City, over the period from August 1972 to November 1980. In their comment on this paper, Harvey and Marshall fit (4.1) and

Table 3
Principal components analysis of estimated covariance matrix of trend disturbances

Eigenvalues	Cumulative proportion	Eigenvectors		
7.739	0.965		-0.55	-0.59
0.262	0.998		0.35	0.48
0.015	1.00		0.76	-0.65

conduct a principal components analysis on the estimated Σ_η . The results, given in Table 3, indicate that the first principal component dominates the variation in the transition equation and represents the basic underlying price in the three cities. Setting K equal to one or two might be appropriate.

Models with common components have also been used in the construction of leading indicators; see Stock and Watson (1990).

4.5. Modelling and estimation for repeated surveys

Many economic variables are measured by using sample survey techniques. Examples include the labour force surveys which are conducted in each member state of the European Community. It is now quite common practice to analyse the results from repeated surveys using time series methods.

If sample surveys are nonoverlapping, then the survey errors are independent and a simple model for the vector of characteristics at time t , θ_t , might be

$$y_t = \theta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t), \quad t = 1, \dots, T, \quad (4.15)$$

where the sampling errors ε_t are independent over time and are independent of θ_t . A simple estimator of θ_t would then be y_t . However, by imposing a model on the evolution of θ_t , an improvement in the precision of the estimate is possible. This improvement will be very marked if θ_t moves very slowly and H_t is large.

Scott and Smith (1974) suggested fitting ARIMA models to θ_t ; see also Smith (1978) and Jones (1980). A more natural approach is to use structural models. The analysis of repeated, nonoverlapping surveys is based on the same principles as standard time series model building except that constraints are imposed on the measurement error covariance matrix through sampling theory.

EXAMPLE. Consider the repeated sample survey of a set of proportions θ_{1t} , $\theta_{2t}, \dots, \theta_{pt}$, using simple random sampling with sample size n_t for $t = 1, \dots, T$. If $p = 2$, and y_t denotes the sample proportion in group one, the simple model

$$\begin{aligned} y_t &= \theta_t + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \frac{\theta_t(1-\theta_t)}{n_t}\right), \\ \theta_t &= \frac{1}{1 + \exp(-\alpha_t)}, \\ \alpha_t &= \alpha_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2), \end{aligned} \quad (4.16)$$

will allow $\theta_{1t} = \theta_t$ and $\theta_{2t} = 1 - \theta_t$ to evolve over time in the range zero to one. If p is greater than two or the state α_t evolves in a more complicated way, perhaps with seasonals, the model can be modified appropriately. However, the modelling principle is unchanged, sampling theory dictates the measurement error and time series considerations the transition equation. A discussion of the way in which such models can be estimated may be found in Section 6.2.

When the repeated surveys are overlapping the model for the measurement equation can become very involved. A clear discussion of the principles involved is given in Scott and Smith (1974). More recent work in this area includes Hausman and Watson (1985), Binder and Dick (1989), Tam (1987), Pfeffermann and Burck (1990) and Pfeffermann (1991).

The work of Pfeffermann (1991) fits well within the framework of this discussion. He identifies three features of overlapping samples which may effect the way the measurement error is modelled. The first is the way the sample is rotated. For example a survey consisting of four panels which are interviewed quarterly, three of the panels will have been included in past surveys while one is wholly new. Thus each panel will remain in the panel for four quarters. This rotation will interact with the second feature of overlapping surveys, the correlation between individual observations. Pfeffermann, in common with most researchers in this area, relies on Henderson's behavioural model for the i -th individual of the survey made at time t . The model is

$$y_{it} - \theta_t = \rho(y_{i,t-1} - \theta_{t-1}) + \omega_{it}, \quad \omega_{it} \sim \text{NID}(0, \sigma_\omega^2), \quad |\rho| < 1. \quad (4.17)$$

The Pfeffermann model is completed by the third feature, which is that the design of the survey is ignorable, although this assumption can be relaxed at the loss of algebraic simplicity.

With these assumptions it is possible to derive the behaviour of the measurement error in a model. If we use y_{it}^{t-j} to denote the i individual at time t , from a panel established at time $t-j$, then we can write

$$\bar{y}_t^{t-j} = \frac{1}{M} \sum_{i=1}^M y_{it}^{t-j}, \quad j = 0, 1, 2, 3, \quad (4.18)$$

as the aggregate survey estimate of θ_t from the panel established at time $t-j$, then

$$y_t = \begin{pmatrix} \bar{y}_t^t \\ \bar{y}_t^{t-1} \\ \bar{y}_t^{t-2} \\ \bar{y}_t^{t-3} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \theta_t + \varepsilon_t, \quad \varepsilon_t = \begin{pmatrix} \bar{\varepsilon}_t^t \\ \bar{\varepsilon}_t^{t-1} \\ \bar{\varepsilon}_t^{t-2} \\ \bar{\varepsilon}_t^{t-3} \end{pmatrix}, \quad (4.19)$$

where

$$\varepsilon_t = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \rho & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 \\ 0 & 0 & \rho & 0 \end{pmatrix} \varepsilon_{t-1} + \begin{pmatrix} \bar{\varepsilon}_t^t \\ \bar{\omega}_t^{t-1} \\ \bar{\omega}_t^{t-2} \\ \bar{\omega}_t^{t-3} \end{pmatrix} = T\varepsilon_{t-1} + \eta_t. \quad (4.20)$$

The covariance of η_t will be

$$\frac{\sigma_\omega^2}{M} \begin{pmatrix} (1-\rho)^{-1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (4.21)$$

The model can be routinely handled by using the Kalman filter to estimate θ_t , as well as the hyperparameters ρ , and σ_ω^2 . In some cases the individual panel results will not be available, but instead only the aggregate will be recorded. Then the measurement equation becomes

$$\begin{aligned} y_t^* &= (\bar{y}_t^t + \bar{y}_t^{t-1} + \bar{y}_t^{t-2} + \bar{y}_t^{t-3}) \\ &= \theta_t + \frac{1}{4}(\bar{\varepsilon}_t^t + \bar{\varepsilon}_t^{t-1} + \bar{\varepsilon}_t^{t-2} + \bar{\varepsilon}_t^{t-3}). \end{aligned} \quad (4.23)$$

5. Simultaneous equation system

This section considers how simultaneous equation models can be estimated when stochastic trend components of the kind described in Section 4 are specified in some or all of the structural equations. We draw on the paper by [Streibel and Harvey \(1993\)](#), which develops and compares a number of methods for the estimation of single equations using instrumental variable (IV) procedures or limited information maximum likelihood (LIML). The question of identifiability is dealt with in [Harvey and Streibel \(1991\)](#).

5.1. Model formulation

Consider a dynamical simultaneous model in which some or all of the structural equations contain stochastic trend components, which, to simplify matters, will be assumed to follow a multivariate random walk. Thus

$$\begin{aligned} \Gamma y_t &= \Phi_1 y_{t-1} + \cdots + \Phi_r y_{t-r} + B_0 x_t + \cdots + B_s x_{t-s} + S \mu_t + \varepsilon_t, \\ \mu_t &= \mu_{t-1} + \eta_t, \end{aligned} \quad (5.1)$$

where Γ is an $N \times N$ matrix of unknown parameters, Φ_1, \dots, Φ_r are $N \times N$ matrices of autoregressive parameters, B_0, \dots, B_s are $N \times K$ matrices of parameters associated with the $K \times 1$ vector of exogenous variables x_t and its lagged values, μ_t is an $n \times 1$ vector of stochastic trends, S is an $N \times n$ selection matrix of ones and zeros, such that each of the stochastic trends appears in a particular equation, and η_t and ε_t are mutually independent, normally distributed white noise disturbance vectors with positive definite covariance matrices Σ_η and Σ_ε respectively. Equations which do not contain a stochastic trend will usually have a constant term and if the exogenous variables are stochastic, it will be assumed that they are generated independently of μ_t and ε_t .

The model is subject to restrictions which usually take the form of certain variables being excluded from certain equations on the basis of prior economic

knowledge. In a similar way, it will normally be the case that there is some rationale for the appearance of stochastic trend components in particular equations. Indeed many econometric models contain a time trend. For example the wage equation in the textbook Klein model has a time trend which is explained in terms of union pressure. Time trends also appear because of technical progress just as in single equations. The argument here is that such effects are more appropriately modelled by stochastic trends.

Pre-multiplying (5.1) by Γ^{-1} gives the econometric reduced form. Dropping the lags, this may be written as

$$y_t = \theta\mu_t + \Pi x_t + \varepsilon_t, \quad (5.2)$$

where $\Pi = \Gamma^{-1}B$, $\varepsilon_t^* = \Gamma^{-1}\varepsilon_t$ and $\theta = \Gamma^{-1}S$. If stochastic trends only appear in some of the equations, that is $1 \leq n < N$, then (5.2) contains common trends; see Section 4.4.

The presence of stochastic trend components in an econometric model has interesting implications for conventional dynamic simultaneous equation models, for the corresponding reduced form models, and for the associated vector autoregression (VAR) for $(y'_t, x'_t)'$. Some of the points can be illustrated with a simple demand and supply system. Let y_{1t} denote quantity, y_{2t} price and x_t an exogenous variable which is stationary after first differencing, that is integrated of order one, and write

$$\begin{aligned} D: \quad & y_{1t} = \gamma_1 y_{2t} + \mu_t + \varepsilon_{1t}, \\ S: \quad & y_{1t} = \gamma_2 y_{2t} + \beta x_t + \varepsilon_{2t}. \end{aligned} \quad (5.3)$$

The stochastic trend component μ_t may be a proxy for changes in tastes. The first equation could be approximated using lags of y_1 and y_2 , but long lags may be needed and, unless μ_t is constant, a unit root is present; compare the employment-output equation of Harvey et al. (1986). The econometric reduced form is

$$\begin{aligned} y_{1t} &= \theta_1 \mu_t + \pi_1 x_t + \varepsilon_{1t}^*, \\ y_{2t} &= \theta_2 \mu_t + \pi_2 x_t + \varepsilon_{2t}^*, \end{aligned} \quad (5.4)$$

where $\theta_1 = \gamma_2 / (\gamma_2 - \gamma_1)$, $\theta_2 = 1 / (\gamma_2 - \gamma_1)$, and so on. Thus there is a common trend. This can be regarded as a reflection of the fact that there is just a single co-integrating relationship, namely the supply curve; compare a similar, but simpler, example in Engle and Granger (1987, p. 263). Attempting to estimate a reduced form with lagged variables but without the stochastic trends runs into complications; if first differences are taken the stochastic part of the model is strictly noninvertible, so the approximation is not valid, while in levels any inference must take account of the unit root; see Sims, Stock and Watson (1990). The VAR representation of $(y'_t, x'_t)'$ is also subject to constraints because of the common trend, and although estimation can be carried out using the method of Johansen (1988), the point remains that long lags may be

required for a satisfactory approximation and so the number of parameters may be very large for moderate size N and K .

In summary, models which approximate stochastic trends by lags may be highly unparsimonious and uninformative about dynamic relationships. If economic theory does suggest the presence of stochastic trend components, therefore, there are likely to be considerable gains from estimating the implied structural relationships directly. If the complete system of equations can be specified, a full information maximum likelihood (FIML) procedure may be employed. If only a subsystem is specified, but all the predetermined variables are named, a limited information maximum likelihood (LIML) procedure is appropriate. When the rest of the system has not been specified at all, ML methods cannot be applied, but a valid instrumental variable (IV) estimator can be obtained.

5.2. Instrumental variable estimation

Suppose the equation of interest is written in matrix notation as

$$y = Z\delta + u \quad (5.5)$$

where Z is a $T \times m$ matrix with observations on explanatory variables and u is a $T \times 1$ vector of disturbances with mean zero and covariance matrix, $\sigma_e^2 V$. The explanatory variables may include variables which are not exogenous. However, the K exogenous variables in the system provide a set of instrumental variables contained in a $T \times K$ matrix, X .

Multiplying (5.5) through by a $T \times T$ matrix L with the property that $L'L = V^{-1}$ yields

$$Ly = LZ\delta + Lu, \quad (5.6)$$

where $\text{Var}(Lu) = \sigma_e^2 I$. If the same transformation is applied to X , the matrix of optimal instruments is formed over a multivariate regression of LZ on LX . The resulting IV estimator is then

$$\hat{\delta} = (Z'L'P_v LZ)^{-1} Z'L'P_v Ly, \quad (5.7)$$

where P_v is the idempotent projection matrix $P_v = LX(X'V^{-1}X)^{-1}X'L'$. It is known as *generalized two stage least squares* (G2SLS). Under standard regularity conditions, as in Bowden and Turkington (1984, p. 26), $T^{1/2}\hat{\delta}$ has a limiting normal distribution. If V is unknown, but depends on a finite number of parameters which can be estimated consistently, the asymptotic distribution is unaffected. When there are no lagged endogenous variables in (5.5) it can be shown that the G2SLS estimator is at least as efficient as 2SLS in the sense that the determinant of its asymptotic covariance matrix cannot exceed the determinant of the corresponding expression for 2SLS. In a similar way, it can be shown that G2SLS is more efficient than an IV estimator in which instruments are formed from X without first transforming by L .

We now consider the estimation of a model which contains a random walk component as well as explanatory variables, that is

$$y_t = \mu_t + z'_t \delta + \varepsilon_t, \quad t = 1, \dots, T. \quad (5.8)$$

If z_t were exogenous, the GLS estimator of δ could be computed by applying the Kalman filter appropriate for the stochastic part of the model, $\mu_t + \varepsilon_t$, to both y_t and z_t and regressing the innovations from y_t on those from z_t ; see Kohn and Ansley (1985). The same approach may be used with IV estimation. In the notation of (5.5) the Kalman filter makes the transformations Ly , LZ and LX . However, the L matrix is now $(T - 1) \times T$ because the diffuse prior for μ_t means that only $T - 1$ innovations can be formed. The variables in (5.8) may be differenced so as to give a stationary disturbance term. Thus

$$\Delta y_t = \Delta z'_t \delta + u_t, \quad t = 2, \dots, T, \quad (5.9)$$

where $u_t = \eta_t + \Delta \varepsilon_t$. This equation corresponds more directly to (5.5) than does (5.8) since a covariance matrix may be constructed for the disturbance vector and the associated L matrix is square. However, postmultiplying this matrix by the $(T - 1) \times 1$ vector of differenced y_t 's gives exactly the same result as postmultiplying the L matrix for (5.8) by the $T \times 1$ vector of y_t 's.

A number of estimation procedures for (5.8) are considered in Streibel and Harvey (1993). In the preferred method, a consistent estimator of δ is first obtained by applying a suitable IV estimator to (5.9); if there are no lagged dependent variables, 2SLS will suffice. Consistent estimators of the hyperparameters are then obtained from the residuals, and these estimators are used to construct a feasible IV estimator of the form (5.7). There are a number of ways of estimating the hyperparameters. In simple cases, closed form expressions based on the residual autocorrelations are available but, even with δ known, such estimators are not efficient. However, what would be the ML estimator if δ were known can always be computed by an iterative optimisation procedure. Given values of the hyperparameters, an IV estimate is constructed for δ . The hyperparameters are then estimated by ML applied to the residuals. This procedure is then iterated to convergence. Although iterating will not change the asymptotic properties of the estimators of δ or the hyperparameters when there are no lagged dependent variables, it may yield estimators with better small sample properties. When this stepwise estimation procedure is used to estimate an equation in a simultaneous equation system it may be referred to as G2SLS/ML. All the above procedures can be implemented in the frequency domain as well as the time domain.

5.3. Maximum likelihood estimation

It is relatively easy to construct the likelihood function for a model of the form (5.1). Maximising this function then gives the FIML estimators. Of course this may not be straightforward computationally, and the estimators obtained for

any one particular equation may be very sensitive to misspecification in other parts of the system.

If interest centres on a single equation, say the first, and there is not enough information to specify the remaining equations, a limited information estimation procedure is appropriate. In a model of the form (5.1) where u_t is NID(0, Ω), the LIML estimator of the parameters in the first equation can be obtained by applying FIML to a system consisting of the first (structural) equation and the reduced form for the endogenous variables appearing in that equation. Since the Jacobian of this system is unity, the estimator can be computed by iterating a feasible SURE estimator to convergence; see Pagan (1979).

Now consider the application of LIML in a Gaussian system with stochastic trends generated by a multivariate random walk. It will also be assumed that the system contains no lags, although the presence of lags in either the endogenous or exogenous variables does not alter the form of the estimator. Thus

$$\Gamma y_t = \mu_t + Bx_t + \varepsilon_t, \quad \text{Var}(\varepsilon_t) = \Sigma_\varepsilon, \quad (5.10)$$

with Γ being positive definite and μ_t given by (5.1). Hence the reduced form is

$$y_t = \mu_t^* + \Pi x_t + \varepsilon_t^*, \quad \text{Var}(\varepsilon_t^*) = \Sigma_\varepsilon^* = \Gamma^{-1} \Sigma_\varepsilon (\Gamma^{-1})', \quad (5.11a)$$

$$\mu_t^* = \mu_{t-1}^* + \eta_t^*, \quad \text{Var}(\eta_t^*) = \Sigma_\eta^* = \Gamma^{-1} \Sigma_\eta (\Gamma^{-1})', \quad (5.11b)$$

where $\mu_t^* = \Gamma^{-1} \mu_t$. The equation of interest, the first in (5.10) corresponds to (5.8) and may be written as

$$y_{1t} = \mu_{1t} + \gamma' y_{2t} + \beta' x_{1t} + \varepsilon_{1t}, \quad (5.12a)$$

$$\mu_{1t} = \mu_{1,t-1} + \eta_{1t}, \quad (5.12b)$$

where y_{2t} is $g \times 1$, x_{1t} is $k \times 1$, and both ε_{1t} and η_{1t} may be correlated with the corresponding disturbances in the other structural equations. Prior knowledge suggests the presence of a stochastic trend in (5.10). There is no information on whether or not stochastic trends are present in the other structural equations in the system, and so they are included for generality. The reduced form for the endogenous variables included in (5.10) may be written as

$$y_{2t} = \mu_{2t}^* + \Pi_2 x_t + \varepsilon_{2t}^*, \quad (5.13a)$$

$$\mu_{2t}^* = \mu_{2,t-1}^* + \eta_{2t}^*. \quad (5.13b)$$

The LIML estimator is obtained by treating (5.12) and (5.13) as though they were the structural form of a system and applying FIML. The Jacobian is unity and estimation proceeds by making use of the multivariate version of the GLS algorithm described in Harvey (1989, p. 133).

Streibel and Harvey (1993) derive the asymptotic distribution of the LIML

estimator and compare the asymptotic covariance matrix of the estimators of β and γ with the corresponding matrix from the G2SLS/ML estimation procedure for a model without lagged endogenous variables. If $\Sigma_\eta = q\Sigma_\epsilon$ in (5.10), where q is a scalar, the multivariate part of the model is homogenous; see Section 4. In this case G2SLS/ML is as efficient as LIML. Indeed efficiency is achieved with G2SLS without iterating, provided an initial consistent estimator of q is used.

Although G2SLS/ML is not, in general, asymptotically efficient as compared with LIML, the Monte Carlo experiments reported in Streibel and Harvey suggest that in small samples its performance is usually better than that of LIML. Since it is much simpler than LIML, it is the recommended estimator.

6. Nonlinear and non-Gaussian models

Relaxing the requirement that time series models be linear and Gaussian opens up a vast range of possibilities. This section introduces the work in this field which exploits the structural time series framework. It starts with a discussion of conditionally Gaussian nonlinear state space models and then progresses to derive a filter for dynamic generalised linear models. Some recent work on exact filters for nonlinear, non-Gaussian state space models is outlined. Finally, some structural approaches to modelling changing variance is discussed.

6.1. Conditionally Gaussian state space models

The state space form and the Kalman filter provides such a strong foundation for the manipulation of linear models that it is very natural to try to extend their use to deal with nonlinear time series. Some progress can be made by defining a conditionally Gaussian state space model

$$\begin{aligned} y_t &= Z_t(Y_{t-1})\alpha_t + X_t\beta + \varepsilon_t, & \varepsilon_t &\sim N(0, H_t(Y_{t-1})) , \\ \alpha_t &= T_t(Y_{t-1})\alpha_{t-1} + W_t\beta + \eta_t, & \eta_t &\sim N(0, Q_t(Y_{t-1})) . \end{aligned} \quad (6.1)$$

Here ε_t and η_s are assumed to be independent for all values of t and s . In this model the matrices in the state space model are allowed to depend on Y_{t-1} , the available information at time $t - 1$. The Kalman filter still goes through in this case and so the likelihood for the model can be built up from the prediction error decomposition.

The theory behind this type of modelling framework has been studied at considerable length in Liptser and Shirayev (1978). The following examples illustrate its flexibility.

EXAMPLE. The coefficient of a first-order autoregression can be allowed to

follow a random walk, as y_{t-1} is in Y_{t-1} . Thus

$$\begin{aligned} y_t &= y_{t-1}\alpha_{t-1} + \varepsilon_t, & \varepsilon_t &\sim \text{NID}(0, \sigma_\varepsilon^2), \\ \alpha_t &= \alpha_{t-1} + \eta_t, & \eta_t &\sim \text{NID}(0, \sigma_\eta^2). \end{aligned} \quad (6.2)$$

EXAMPLE. Some macro-economic time series appear to exhibit cycles in which the downswing is shorter than the upswing. A simple way of capturing such a feature is to specify a cyclical component which switches from one frequency to another as it moves from downswing into upswing and vice versa. This could be achieved by setting

$$\lambda_c = \begin{cases} \lambda_1, & \text{if } \hat{\psi}_{t|t-1} - \hat{\psi}_{t-1} > 0, \\ \lambda_2, & \text{if } \hat{\psi}_{t|t-1} - \hat{\psi}_{t-1} \leq 0, \end{cases} \quad \lambda_1 \leq \lambda_2, \quad (6.3)$$

where $\hat{\psi}_{t|t-1}$ and $\hat{\psi}_{t-1}$ are estimates of the state of the cycle at times t and $t-1$ respectively, made at time $t-1$. This model, which belongs within the class of threshold models described in Tong (1990), in effect fits two separate linear cycle models to the date, the division taking place and $\hat{\psi}_{t|t-1} - \hat{\psi}_{t-1}$ switches sign.

6.2. Extended Kalman filter

For ease of exposition suppose y_t and α_t are univariate and

$$\begin{aligned} y_t &= z_t(\alpha_t) + \varepsilon_t, & \varepsilon_t &\sim \text{NID}(0, \sigma_\varepsilon^2(\alpha_t)), \\ \alpha_t &= T_t(\alpha_{t-1}) + \eta_t, & \eta_t &\sim \text{NID}(0, \sigma_\eta^2(\alpha_{t-1})). \end{aligned} \quad (6.4)$$

This model cannot be handled exactly by using the Kalman filter. However, for some functions it is possible to expand $z_t(\alpha_t)$ and $T_t(\alpha_{t-1})$ using a Taylor series to give

$$\begin{aligned} z_t(\alpha_t) &\cong z_t(a_{t|t-1}) + \frac{\partial z_t(a_{t|t-1})}{\partial \alpha_t} (\alpha_t - a_{t|t-1}), \\ \alpha_t &\cong T_t(a_{t-1}) + \frac{\partial T_t(a_{t-1})}{\partial \alpha_{t-1}} (\alpha_{t-1} - a_{t-1}). \end{aligned} \quad (6.5)$$

If, in addition, the dependence of the variances on the states is dealt with by replacing them by estimates, made at time $t-1$, then the new approximate model becomes

$$\begin{aligned} y_t &= z_t(a_{t|t-1}) + \frac{\partial z_t(a_{t|t-1})}{\partial \alpha_t} (\alpha_t - a_{t|t-1}) + \varepsilon_t, & \varepsilon_t &\sim \text{N}(0, \sigma_\varepsilon^2(a_{t|t-1})), \\ \alpha_t &= T_t(a_{t-1}) + \frac{\partial T_t(a_{t-1})}{\partial \alpha_{t-1}} (\alpha_{t-1} - a_{t-1}) + \eta_t, & \eta_t &\sim \text{N}(0, \sigma_\eta^2(a_{t-1})). \end{aligned} \quad (6.6)$$

This model is then in the conditionally Gaussian framework and so the Kalman filter can be used to estimate the state. Since the model itself is an approximation, we call the conditionally Gaussian Kalman filter an extended Kalman filter for the original model (6.4); see Anderson and Moore (1979, Chapter 8).

EXAMPLE. Suppose the logistic transformation is being used to keep $z_t(\alpha_t)$ between zero and one as in (4.16). Then

$$z_t(\alpha_t) = \frac{1}{1 + \exp(-\alpha_t)} . \quad (6.7)$$

Then the expanded model becomes

$$y_t = \frac{1}{1 + \exp(-a_{t|t-1})} + \frac{\exp(-a_{t|t-1})}{(1 + \exp(-a_{t|t-1}))^2} (\alpha_t - a_{t|t-1}) + \varepsilon_t . \quad (6.8)$$

This idea can be used to construct a model of opinion polls. Suppose there are just two parties. If the level of support for one party is modelled as a logistic transformation of a Gaussian random walk and the measurement error originates from using a simple random sample size n_t , then

$$y_t = \mu_t + \varepsilon_t , \quad \varepsilon_t \sim N(0, \sigma_t^2) , \quad \sigma_t^2 = \frac{\mu_t(1 - \mu_t)}{n_t} , \quad (6.9a)$$

$$\mu_t = \frac{1}{1 + \exp(-\alpha_t)} , \quad (6.9b)$$

$$\alpha_t = \alpha_{t-1} + \eta_t , \quad \eta_t \sim NID(0, \sigma_\eta^2) . \quad (6.9c)$$

As μ_t is unknown, this model cannot be analysed by using the Kalman filter. Instead, an estimate of α_t can be made at time $t-1$, written $a_{t|t-1}$, and it can be used to replace μ_t in the variance. One of the problems with this approach is that this model does not constrain the observations to lie between zero and one, as ε_t is assumed Gaussian. Although this could be a problem if μ_t were to be close to zero or one, this is unlikely to pose a difficulty for moderate sample sizes.

The Kalman filter can be applied in the standard way once the logistic transformation has been Taylor expanded. The resulting model is

$$y_t = m_{t-1} + \exp(-a_{t-1})m_{t-1}^2(\alpha_t - a_{t-1}) + \varepsilon_t , \\ \varepsilon_t \sim N\left(0, \frac{m_{t-1}(1 - m_{t-1})}{n_t}\right) , \quad (6.10a)$$

$$m_{t-1} = \frac{1}{1 + \exp(-a_{t-1})} . \quad (6.10b)$$

An approach similar to this, but using a multivariate continuous time model to

allow for irregular observations, was followed by Shephard and Harvey (1989) in their analysis of opinion poll data from the British general election campaigns of October 1974, 1979, 1983 and 1987.

6.3. Non-Gaussian state space models

Although the Gaussian state space form provides the basis for the analysis of many time series, it is sometimes not possible to adequately model the data, or a transformation of it, in this way. Some series, such as count data, are intrinsically non-Gaussian and so using a Gaussian model could harm forecasting precision. In this section we outline the methods for directly modelling non-Gaussian series.

The key to modelling non-Gaussian time series is the non-Gaussian state space form. It will be built out of two assumptions. Firstly the measurement equation is such that we can write

$$f(y_1, \dots, y_T | \alpha_1, \dots, \alpha_T) = \prod_{t=1}^T f(y_t | \alpha_t). \quad (6.11)$$

This assumes that given the state α_t , the observation y_t is independent of all the other states and observations. Thus α_t is *sufficient* for y_t . The second assumption is that the transition equation is such that

$$f(\alpha_1, \dots, \alpha_T | Y_0) = f(\alpha_1 | Y_0) \prod_{t=2}^T f(\alpha_t | \alpha_{t-1}), \quad (6.12)$$

that is the state follows a Markov process.

Filtering can be derived for a continuous state by the integrals

$$f(\alpha_t | Y_{t-1}) = \int f(\alpha_t | \alpha_{t-1}) f(\alpha_{t-1} | Y_{t-1}) d\alpha_{t-1}, \quad (6.13a)$$

$$f(\alpha_t | Y_t) = f(y_t | \alpha_t) f(\alpha_t | Y_{t-1}) / \int f(y_t | \alpha_t) f(\alpha_t | Y_{t-1}) d\alpha_t. \quad (6.13b)$$

Thus it is technically possible to carry out filtering, and indeed smoothing, for any state space model if the integrals can be computed. Kitagawa (1987) and Pole and West (1990) have suggested using particular sets of numerical integration rules to evaluate these densities. The main drawback with this general approach is the computational requirement, especially if parameter estimation is required. This is considerable if a reasonable degree of accuracy is to be achieved and the dimension of the state is larger; the dimension of the integral will equal the dimension of the state and so will be 13 for a basic structural model for monthly data. It is well known from the numerical analysis literature that the use of numerical integration rules to evaluate high-dimensional integrals is fraught with difficulty.

The computational explosion associated with the use of these numerical

integration rules has prompted research into alternative methods for dealing with non-Gaussian state space models. Recent work by West and Harrison (1989) and West, Harrison and Mignon (1985) has attempted to extend the use of the Kalman filter to cover cases where the measurement density is a member of the exponential family, which includes the binomial, Poisson and gamma densities, while maintaining the Gaussian transition density. As such this tries to extend the generalised linear model, described in McCullagh and Nelder (1989), to allow for dynamic behaviour.

For ease of exposition we will only look at the extension of the local level model to cover the exponential family measurement density. More specifically, we will assume that

$$\begin{aligned} f(y_t | \mu_t) &= b(y_t, \sigma_{et}^2) \exp\left(\frac{y_t \mu_t - a(\mu_t)}{\sigma_{et}^2}\right), \\ f(\mu_t | \mu_{t-1}) &= \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(\mu_t - \mu_{t-1})^2}{2\sigma_n^2}\right) \end{aligned} \quad (6.14)$$

and follow the development given in West and Harrison (1989). Here σ_{et}^2 will be assumed to be known at time t . By selecting $a(\cdot)$ and $b(\cdot)$ appropriately, a large number of distributions can result. A simple example of this is the binomial distribution

$$f(y_t | \pi_t) = \frac{n_t!}{y_t!(n_t - y_t)!} \pi_t^{y_t} (1 - \pi_t)^{n_t - y_t}, \quad (6.15)$$

which is obtained by writing

$$\begin{aligned} \mu_t &= \log \frac{\pi_t}{1 - \pi_t}, & a(\mu_t) &= \log(1 + \exp(\mu_t)), \\ b(y_t, \sigma_{et}^2) &= \frac{n_t!}{y_t!(n_t - y_t)!}. \end{aligned} \quad (6.16)$$

Although it is relatively straightforward to place densities into their exponential form, the difficulty comes from filtering the unobservable component μ_t as it progresses through time. Suppose we have a distribution for $\mu_{t-1} | Y_{t-1}$. The first two moments of this prior will be written as m_{t-1} and p_{t-1} . The random walk transition means that the first two moments of $\mu_t | Y_{t-1}$ will be

$$m_{t|t-1} = m_{t-1}, \quad p_{t|t-1} = p_{t-1} + \sigma_e^2. \quad (6.17)$$

As the measurement density is in the exponential family, it is always possible to find a conjugate density. Generically it takes the form

$$f(\mu_t | Y_{t-1}) = c(r_{t|t-1}, s_{t|t-1}) \exp(\mu_t r_{t|t-1} - s_{t|t-1} a(\mu_t)). \quad (6.18)$$

For a particular form of this density it is typically possible to select $r_{t|t-1}$ and $s_{t|t-1}$ so that the first two moments of this density match $m_{t|t-1}$ and $p_{t|t-1}$. Thus the actual prior density of $\mu_t | Y_{t-1}$ will be approximated by a density which has

identical first two moments and is conjugate to the measurement density. Having determined r and s , this conjugate prior can be used to construct the one-step ahead density

$$f(y_t | Y_{t-1}) = \int f(y_t | \mu_t) f(\mu_t | Y_{t-1}) d\mu_t \quad (6.19a)$$

$$= \frac{c(r_{t|t-1}, s_{t|t-1}) b(y_t, \sigma_{\epsilon t}^2)}{c(r_{t|t-1} + y_t/\sigma_{\epsilon t}^2, s_{t|t-1} + (1/\sigma_{\epsilon t}^2))}. \quad (6.19b)$$

Further

$$f(\mu_t | Y_t) = c(r_t, s_t) \exp(r_t \mu_t - s_t a(\mu_t)), \quad (6.20)$$

where

$$r_t = r_{t|t-1} + \frac{y_t}{\sigma_{\epsilon t}^2}, \quad s_t = s_{t|t-1} + \frac{1}{\sigma_{\epsilon t}^2}. \quad (6.21)$$

By finding the first two moments of this density, implied values for m_t and p_t can be deduced, so starting the cycle off again. As the approximate density of $y_t | Y_{t-1}$ is known analytically, a maximum quasi-likelihood procedure can be used to estimate the unknown parameters of this model by using a predictive distribution decomposition of the joint density of the observations

$$f(y_1, \dots, y_T | Y_0) = \prod_{t=1}^T f(y_t | Y_{t-1}). \quad (6.22)$$

EXAMPLE. If the measurement equation is normal then

$$r_{t|t-1} = \frac{m_{t|t-1}}{p_{t|t-1}}, \quad s_{t|t-1} = \frac{1}{p_{t|t-1}}, \quad (6.23)$$

so

$$r_t = \frac{m_{t|t-1}}{p_{t|t-1}} + \frac{y_t}{\sigma_{\epsilon t}^2}, \quad s_t = \frac{1}{p_{t|t-1}} + \frac{1}{\sigma_{\epsilon t}^2} \quad (6.24)$$

implying

$$m_t = p_t r_t = \frac{p_t}{p_{t|t-1}} m_{t|t-1} + \frac{p_t}{\sigma_{\epsilon t}^2} y_t. \quad (6.25)$$

As

$$p_t = \frac{\sigma_{\epsilon t}^2 p_{t|t-1}}{p_{t|t-1} + \sigma_{\epsilon t}^2}, \quad (6.26)$$

this is the usual Kalman filter.

EXAMPLE. If the measurement density is binomial then the conjugate prior is beta,

$$f(\pi_t | Y_{t-1}) = \frac{\Gamma(r_{t|t-1} + s_{t|t-1})}{\Gamma(r_{t|t-1})\Gamma(s_{t|t-1})} \pi_t^{r_{t|t-1}-1} (1 - \pi_t)^{s_{t|t-1}-1}. \quad (6.27)$$

But as $\mu_t = \log \pi_t / 1 - \pi_t$ it follows that using our prior knowledge of μ_t ,

$$\begin{aligned} m_{t|t-1} &= E\mu_t | Y_{t-1} = \gamma(r_{t|t-1}) - \gamma(s_{t|t-1}), \\ p_{t|t-1} &= \text{Var } \mu_t | Y_{t-1} = \dot{\gamma}(r_{t|t-1}) + \dot{\gamma}(s_{t|t-1}), \end{aligned} \quad (6.28)$$

where $\gamma(\cdot)$ is the digamma function and $\dot{\gamma}(\cdot)$ is its derivative, we can allow $r_{t|t-1}$ and $s_{t|t-1}$ to be selected numerically. When $r_{t|t-1}$ and $s_{t|t-1}$ are updated to give r_t and s_t , the corresponding m_t and p_t can be deduced from

$$\begin{aligned} m_t &= \gamma(r_t) - \gamma(s_t), \\ p_t &= \dot{\gamma}(r_t) + \dot{\gamma}(s_t). \end{aligned} \quad (6.29)$$

This completes the cycle, since $m_{t+1|t} = m_t$ and $p_{t+1|t} = p_t + \sigma_\eta^2$.

The work on the dynamic generalised linear model and the extended Kalman filter share some important characteristics. The most important of these is that both are approximations, where the degree of approximation is difficult to determine. In neither case does the filtered estimate of the state possess the important property that it is the minimum mean square error estimate.

An alternative approach is to design transition equations which are conjugate to the measurement density so that there exists an exact analytic filter. In the last five years there has been some important work carried out on these exact non-Gaussian filters. Most of this work has been based on a gamma–beta transition equation; see the discussion in Lewis, McKenzie and Hugus (1989). A simple example is

$$\begin{aligned} \alpha_t &= \omega^{-1} \alpha_{t-1} \eta_t, \quad \eta_t \sim \text{Beta}(\omega a_{t-1}, (1 - \omega)a_{t-1}), \\ \alpha_{t-1} | Y_{t-1} &\sim G(a_{t-1}, b_{t-1}), \quad \omega \in (0, 1]. \end{aligned} \quad (6.30)$$

The transition equation is multiplicative. The rather strange constraints on the form of the beta variable are required for conjugacy. They imply $\alpha_t | Y_{t-1} \sim G(a_{t|t-1}, b_{t|t-1})$, where

$$a_{t|t-1} = \omega a_{t-1}, \quad b_{t|t-1} = \omega b_{t-1}. \quad (6.31)$$

This means that

$$\begin{aligned} E\alpha_t | Y_{t-1} &= \frac{a_{t|t-1}}{b_{t|t-1}} = \frac{a_t}{b_t} = E\alpha_{t-1} | Y_{t-1}, \\ \text{Var } \alpha_t | Y_{t-1} &= \frac{a_{t|t-1}}{b_{t|t-1}^2} = \omega^{-1} \text{Var } \alpha_{t-1} | Y_{t-1}. \end{aligned} \quad (6.32)$$

Thus the expectation of the level remains the same, but its variance increases just as it does in a Gaussian local model.

Gamma–beta transition equations have been used by Smith and Miller (1986) in their analysis of extreme value time series to enable them to forecast athletic world records. Harvey and Fernandes (1989a) exploited them to study the goals scored by the England football team, against Scotland in their matches at Hampden Park. A more interesting example from an economic viewpoint is the paper by Harvey and Fernandes (1989b) on insurance claims. Both papers use a Poisson measurement equation

$$f(y_t | \alpha_t) = \frac{e^{-\alpha_t} \alpha_t^{y_t}}{y_t!}. \quad (6.33)$$

As a gamma is the conjugate prior to a Poisson distribution, this model is closed by using a gamma–beta transition equation, for the use of Bayes' theorem shows that

$$\alpha_t | Y_t \sim G(a_t, b_t), \quad a_t = a_{t|t-1} + y_t, \quad b_t = b_{t|t-1} + 1. \quad (6.34)$$

This means that if $a_0 = b_0 = 0$, the filtered estimate of α_t is

$$E\alpha_t | Y_t = \frac{a_t}{b_t} = \frac{\sum_{j=0}^{t-1} \omega^j y_{t-j}}{\sum_{j=0}^{t-1} \omega^j} \quad (6.35)$$

which is an exponentially weighted moving average of the observations. The one-step ahead predictive distribution is

$$\begin{aligned} f(y_t | Y_{t-1}) &= \int f(y_t | \alpha_t) f(\alpha_t | Y_{t-1}) d\alpha_t \\ &= \frac{(a_t - 1)!}{y_t! (a_t - 1 - y_t)!} (b_{t|t-1})^{a_{t|t-1}} b_t^{-a_t}, \end{aligned} \quad (6.36)$$

which is negative binomial and so the likelihood for this model can be computed using the predictive distribution decomposition, as in (6.22).

6.4. Stochastic variance models

One of the most important modelling techniques to emerge in the 1980s was autoregressive conditional heteroskedasticity (ARCH); see Engle (1982) and Bollerslev (1986). These authors suggested modelling the variability of a series by using weights of the squares of the past observations. The important

GARCH(1, 1) model has

$$\begin{aligned} y_t | Y_{t-1} &\sim N(0, h_t), \\ h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 h_{t-1}, \end{aligned} \quad (6.37)$$

that is the one step ahead predictive distribution depends on the variable h_t . Thus the conditional variance of the process is modelled directly, just like in ARMA models the conditional mean is modelled directly.

Although the development of these models has had a strong influence in the econometric literature, a rather different modelling approach has been suggested in the finance literature; see, for example, Hull and White (1987), Chesney and Scott (1989) and Melino and Turnbull (1990). These papers have been motivated by the desire to allow time varying volatility in opinion pricing models, so producing a more dynamic Black–Scholes type pricing equation. This requires that the volatility models be written down in terms of continuous time Brownian motion. In general ARCH models do not tie in with such a formulation, although as Nelson (1991) shows there are links with EGARCH.

The finance models, which are usually called stochastic volatility models, although we prefer to call them stochastic variance models, have some very appealing properties. They directly model the variability of the series, rather than the conditional variability. Thus they are analogous to the structural models discussed in the rest of this paper which are all direct models for the mean of the series at a particular point in time. A simple example is

$$\begin{aligned} y_t &= \varepsilon_t \exp(h_t/2), \quad \varepsilon_t \sim NID(0, 1), \\ h_t &= \alpha_0 + \alpha_1 h_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_h^2). \end{aligned} \quad (6.38)$$

where, for simplicity, ε and η , are assumed to be independent for all t and s . Here the logarithm of the standard deviation of the series follows an AR(1) process, which has an obvious continuous time generalisation. It is not observable, but it can be estimated from the linear state space form

$$\begin{aligned} \log y_t^2 &= h_t + \log \varepsilon_t^2 = h_t + \varepsilon_t^* \\ h_t &= \alpha_0 + \alpha_1 h_{t-1} + \eta_t, \end{aligned} \quad (6.39)$$

where ε_t^* is independent and identically distributed, but not Gaussian. In fact $E\varepsilon_t^* = -1.27$ and $\text{Var } \varepsilon_t^* = 4.93$; see Abramowitz and Stegun (1970, p. 943). The Kalman filter provides the minimum mean square linear estimator of h_t from the $\log y_t^2$ series. Further, the corresponding smoother inherits the same property of being the best linear estimator given the whole data set.

As y_t is the product of two strictly stationary processes, it must also be strictly stationary. Thus for any stochastic variance model, the restrictions needed to ensure the stationarity of y_t are just the standard restrictions to ensure the stationarity of the process generating h_t ; compare the simplicity of this to the GARCH(1, 1) model, as analysed by Nelson (1990). The properties

of this particular autoregressive stochastic variance model can be worked out if $|\alpha_1| < 1$, for then h_t must be strictly stationary, with

$$\gamma_h = Eh_t = \frac{\alpha_0}{1 - \alpha_1}, \quad \sigma_h^2 = \text{Var } h_t = \frac{\sigma_\eta^2}{1 - \alpha_1^2}. \quad (6.40)$$

The fact that y_t is white noise follows almost immediately given the independence of ε_t and η_t . The mean is clearly zero, while

$$Ey_t y_{t-\tau} = E\varepsilon_t \varepsilon_{t-\tau} E \left(\exp \left(\frac{h_t + h_{t-\tau}}{2} \right) \right) = 0 \quad (6.41)$$

as $E\varepsilon_t \varepsilon_{t-\tau} = 0$. The odd moments of y_t are all zero because ε_t is symmetric. The even moments can be obtained by making use of a standard result for the lognormal distribution, which in the present context tells us that since $\exp(h_t)$ is lognormal, its j -th moment about the origin is $\exp(j\gamma_h + j\sigma_h^2/2)$. Therefore

$$\text{Var } y_t = E\varepsilon_t^2 E \exp(h_t) = \exp(\gamma_h + \sigma_h^2/2). \quad (6.42)$$

The fourth moment is

$$Ey_t^4 = E\varepsilon_t^4 E \exp(h_t)^2 = 3 \exp(2\gamma_h + 2\sigma_h^2) \quad (6.43)$$

and so the kurtosis is $3 \exp(\sigma_h^2)$, which is greater than 3 when σ_h^2 is positive. Thus the model exhibits excess kurtosis compared with the normal distribution. The dynamic properties of the model appear in $\log y_t^2$ rather than y_t^2 . In (6.39) h_t is an AR(1) process and ε_t^* is white noise so $\log y_t^2$ is an ARMA(1, 1) process and its autocorrelation function is easy to derive.

The parameter estimation of stochastic variance models is also reasonably simple. Although the linear state space representation of $\log y_t^2$ allows the computation of the innovations and their associated variances, the innovations are not actually Gaussian. If this fact is ignored for a moment and the ‘Gaussian’ likelihood is constructed, then this objective function is called a quasi-likelihood. A valid asymptotic theory is available for the estimator which results from maximising this function; see Dunsmuir (1979, p. 502).

The model can be generalised so that h_t follows any stationary ARMA process, in which case y_t is also stationary and its properties can be deduced from the properties of h_t . Other components could also be brought into the model. For example, the variance could be related to a changing daily or intra daily pattern.

Multivariate generalisations of the stochastic volatility models have been suggested by Harvey, Ruiz and Shephard (1991). These models overcome many of the difficulties associated with multivariate ARCH based models; see Bollerslev, Chou and Kroner (1992) for a survey of these models. The basic

idea is to let the i th element of the N -dimensional vector y_t be

$$\begin{aligned} y_{it} &= \varepsilon_{it} \exp(h_{it}), \\ h_{it} &= \alpha_{0i} + \alpha_{1i} h_{it-1} + \eta_{it}, \end{aligned} \quad (6.44)$$

where ε_t and η_t are N -dimensional multivariate Gaussian white noise processes with covariances Σ_ε and Σ_η . The matrix Σ_ε will be constrained to have ones down its leading diagonal and so can be thought of as being a correlation matrix.

The model can be put into state space form, as in (6.39), by writing

$$\log y_{it}^2 = h_{it} + \varepsilon_{it}^*, \quad i = 1, \dots, N. \quad (6.45)$$

The covariance of $\varepsilon_i^* = (\varepsilon_{1i}^*, \dots, \varepsilon_{Ni}^*)'$ can be analytically related to Σ_ε , so allowing straightforward estimation of Σ_ε and Σ_η by using a quasi-likelihood, although the signs of the elements of Σ_ε cannot be identified using this procedure. However, these signs can be estimated directly from the data, for $y_{it}y_{jt} > 0$ if and only if $\varepsilon_{it}\varepsilon_{jt} > 0$ implying the sign of the i,j -th element of Σ_ε should be estimated to be positive if the number of occurrences of $y_{it}y_{jt} > 0$ is greater than $T/2$.

Harvey, Ruiz and Shephard (1991) analyse four daily exchange rates for the US dollar using (6.38) and find that α_1 is approximately equal to unity for all the rates, suggesting that a random walk is appropriate for h_t . This model has very similar properties to IGARCH in which $\alpha_1 + \alpha_2 = 1$ in (6.37). The multivariate generalisation is straightforward and the transformed observations, as in (6.45), are a SUTSE model of the form (4.1). Further investigation of the model indicates that it can be made even more parsimonious by specifying just two common trends, thereby implying co-integration in volatility; compare (4.14). The first common trend affects all four exchange rates, while the second is associated primarily with the Yen.

Although stochastic variance models can be made to fit within the linear space framework and so can be handled by using the Kalman filter, this filter does not deliver the optimal (minimum mean square error) estimate. It is not possible to derive the optimal filter analytically and so it is tempting to change the transition equation in an attempt to allow the derivation of exact results for this problem. This approach has been followed by Shephard (1993b) using the techniques discussed in the previous subsection. He proposed a local scale model

$$y_t | \alpha_t \sim N(0, \alpha_t^{-1}), \quad (6.46)$$

where α_t , the precision of the series at time t , satisfies the gamma-beta transition equation of (6.30). Although α_t is unknown, it can be estimated because

$$\alpha_t | Y_t \sim G(a_t, b_t), \quad a_t = a_{t|t-1} + \frac{1}{2}, \quad b_t = b_{t|t-1} + \frac{1}{2}y_t^2 \quad (6.47)$$

and also

$$\text{E}\alpha_t | Y_t = \frac{a_t}{b_t} = \frac{\sum_{j=0}^{t-1} \omega^j}{\sum_{j=0}^{t-1} \omega^j y_{t-j}^2} \quad (6.48)$$

this being the inverse of the EWMA of the squares of the observations.

When the focus shifts to the one-step ahead forecast density, then

$$y_t | Y_{t-1} \sim t_{2a_{t|t-1}} \left(0, \frac{b_{t|t-1}}{a_{t|t-1}} \right) \quad (6.49)$$

that is $y_t | Y_{t-1}$ is a scaled Student's t variable, with scale which is an exact EWMA of the squares of the past observations. If t is large then the degrees of freedom in the predictive density will approximately equal $\omega/(1-\omega)$. As $\omega \rightarrow 1$, the degrees of freedom increase and so the one-step ahead density becomes like a normal. The parameter ω has to be larger than 0.8 for the fourth moment to exist. Setting ω to 0.5 means that the density is a Cauchy random variable.

Many extensions of this model are possible, allowing, amongst other things, an exponential power measurement density instead of normal, irregularly spaced observations and multistep ahead forecasts. The difficulty with the model is that it is hard to significantly depart from the gamma–beta transition equation. As this is constrained to be a nonstationary process and is technically awkward to generalise to the multivariate case, it is of less practical use than the stochastic variance models. However, for dealing with this very special case it does provide a rather interesting alternative.

References

- Abramowitz, M. and I. A. Stegun (1970). *Handbook of Mathematical Functions*. Dover, New York.
- Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ.
- Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *J. Econometrics* 18, 47–82.
- Ansley, C. F. and R. Kohn (1985). Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. *Ann. Statist.* 13, 1286–1316.
- Ansley, C. F. and R. Kohn (1989). Prediction mean square error for state space models with estimated parameters. *Biometrika* 73, 467–474.
- Binder, D. A. and J. P. Dick (1989). Modelling and estimation for repeated surveys. *Survey Method.* 15, 29–45.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31, 307–327.
- Bollerslev, T., R. Y. Chou and K. F. Kroner (1992). ARCH models in finance: A review of the theory and empirical evidence. *J. Econometrics* 52, 5–59.

- Bowden, R. J. and D. A. Turkington (1984). *Instrumental Variables*. Cambridge Univ. Press, Cambridge.
- Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*. Revised edition, Holden-Day, San Francisco, CA.
- Box, G. E. P. and G. C. Tiao (1975). Intervention analysis with applications to economic and environmental problems. *J. Amer. Statist. Assoc.* **70**, 70–79.
- Chesney, M. and L. O. Scott (1989). Pricing European currency options: A comparison of the modified Black–Scholes model and a random variance model. *J. Financ. Quant. Anal.* **24**, 267–284.
- De Jong, P. (1988a). The likelihood for a state space model. *Biometrika* **75**, 165–169.
- De Jong, P. (1988b). A cross-validation filter for time series models. *Biometrika* **75**, 594–600.
- De Jong, P. (1989). Smoothing and interpolation with the state space model. *J. Amer. Statist. Assoc.* **84**, 1085–1088.
- De Jong, P. (1991). The diffuse Kalman filter. *Ann. Statist.* **19**, 1073–1083.
- Duncan, D. B. and S. D. Horn (1972). Linear dynamic regression from the viewpoint of regression analysis. *J. Amer. Statist. Assoc.* **67**, 815–821.
- Dunsmuir, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its applications to models for a signal observed with noise. *Ann. Statist.* **7**, 490–506.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* **50**, 987–1007.
- Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica* **55**, 251–276.
- Fernández, F. J. (1990). Estimation and testing of a multivariate exponential smoothing model. *J. Time Ser. Anal.* **11**, 89–105.
- Fernández, F. J. and A. C. Harvey (1990). Seemingly unrelated time series equations and a test for homogeneity. *J. Business Econ. Statist.* **8**, 71–82.
- Harrison, P. J. and C. F. Stevens (1976). Bayesian forecasting. *J. Roy. Statist. Soc. Ser. B* **38**, 205–247.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge Univ. Press, Cambridge.
- Harvey, A. C. and A. Jaeger (1991). Detrending, stylized facts and the business cycle. Mimeo, Department of Statistics, London School of Economics.
- Harvey, A. C. and C. Fernandes (1989a). Time series models for count or qualitative observations (with discussion). *J. Business Econ. Statist.* **7**, 407–422.
- Harvey, A. C. and C. Fernandes (1989b). Time series models for insurance claims. *J. Inst. Actuar.* **116**, 513–528.
- Harvey, A. C. and S. J. Koopman (1992). Diagnostic checking of unobserved components time series models. *J. Business Econ. Statist.* **10**, 377–389.
- Harvey, A. C. and S. J. Koopman (1993). Short term forecasting of periodic time series using time-varying splines. *J. Amer. Statist. Assoc.*, to appear.
- Harvey, A. C. and P. Marshall (1991). Inter-fuel substitution, technical change and the demand for energy in the UK economy. *Appl. Econ.* **23**, 1077–1086.
- Harvey, A. C. and S. Peters (1990). Estimation procedures for structural time series models, *J. Forecast.* **9**, 89–108.
- Harvey, A. C., E. Ruiz and N. G. Shephard (1991). Multivariate stochastic variance models. Mimeo, Department of Statistics, London School of Economics.
- Harvey, A. C. and M. Streibel (1991). Stochastic trends in simultaneous equation systems. In: P. Hackl and A. Westlund, eds., *Economic Structural Change*, Springer, Berlin, 169–178.
- Harvey, A. C., B. Henry, S. Peters and S. Wren-Lewis (1986). Stochastic trends in dynamic regression models: An application to the output-employment equation. *Econ. J.* **96**, 975–985.
- Harvey, A. C. and J. Durbin (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *J. Roy. Statist. Soc. Ser. A* **149**, 187–227.
- Hausman, J. A. and M. W. Watson (1985). Errors in variables and seasonal adjustment procedures. *J. Amer. Statist. Assoc.* **80**, 541–552.

- Hendry, D. F. and J.-F. Richard (1983). The econometric analysis of economic time series. *Internat. Statist. Rev.* **51**, 111–164.
- Hull, J. and A. White (1987). Hedging the risks from writing foreign currency options. *J. Internat. Money Finance* **6**, 131–152.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *J. Econ. Dynamics Control* **12**, 231–254.
- Jones, R. H. (1980). Best linear unbiased estimators for repeated survey. *J. Roy. Statist. Soc. Ser. B* **42**, 221–226.
- Jorgenson, D. W. (1986). Econometric methods for modelling producer behaviour. In: Z. Griliches and M. D. Intriligator, eds. *Handbook of Econometrics*, Vol. 3, North-Holland, Amsterdam, 1841–1915.
- Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *J. Amer. Statist. Assoc.* **82**, 1032–1041.
- Kohn, R. and C. F. Ansley (1985). Efficient estimation and prediction in time series regression models. *Biometrika* **72**, 694–697.
- Kohn, R. and C. F. Ansley (1989). A fast algorithm for signal extraction, influence and cross-validation in state space models. *Biometrika* **76**, 65–79.
- Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* **80**, to appear.
- Koopman, S. J. and N. Shephard (1992). Exact score for time series models in state space form. *Biometrika* **79**, 823–826.
- Lewis, P. A. W., E. McKenzie and D. K. Hugus (1989). Gamma processes, *Comm. Statist. Stochast. Models* **5**, 1–30.
- Liptser, R. S. and A. N. Shirayev (1978). *Statistics of Random Processes II: Applications*. Springer, New York.
- Marshall, P. (1992a). State space models with diffuse initial conditions. *J. Time Ser. Anal.* **13**, 411–414.
- Marshall, P. (1992b). Estimating time-dependent means in dynamic models for cross-sections of time series. *Empirical Econ.* **17**, 25–33.
- McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*. 2nd ed., Chapman and Hall, London.
- Melino, A. and S. M. Turnbull (1990). Pricing options with stochastic volatility. *J. Econometrics* **45**, 239–265.
- Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**, 318–334.
- Nelson, D. B. (1991). ARCH models as diffusion approximations. *J. Econometrics* **45**, 7–38.
- Pagan, A. R. (1979). Some consequences of viewing LIML as an iterated Aitken estimator. *Econ. Lett.* **3**, 369–372.
- Pfeffermann, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys (with discussion). *J. Business Econ. Statist.* **9**, 163–177.
- Pfeffermann, D. and L. Burck (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Method.* **16**, 217–338.
- Pole, A. and M. West (1990). Efficient Bayesian learning in non-linear dynamic models. *J. Forecasting* **9**, 119–136.
- Rao, C. R. (1973). *Linear Statistical Inference*. 2nd ed., Wiley, New York.
- Rosenberg, B. (1973). Random coefficient models: The analysis of a cross section of time series by stochastically convergent parameter regression. *Ann. Econ. Social Measurement* **2**, 399–428.
- Scott, A. J. and T. M. F. Smith (1974). Analysis of repeated surveys using time series models. *J. Amer. Statist. Assoc.* **69**, 674–678.
- Shephard, N. (1993a). Maximum likelihood estimation of regression models with stochastic trend components. *J. Amer. Statist. Assoc.* **88**, 590–595.
- Shephard, N. (1993b). Local scale model: State space alternatives to integrated GARCH processes. *J. Econometrics*, to appear.
- Shephard, N. and A. C. Harvey (1989). Tracking to the level of party support during general election campaigns. Mimeo, Department of Statistics, London School of Economics.

- Shephard, N. and A. C. Harvey (1990). On the probability of estimating a deterministic component in the local level model. *J. Time Ser. Anal.* **11**, 339–347.
- Sims, C. A., J. H. Stock and M. W. Watson (1990). Inference in linear time series models with some unit roots. *Econometrica* **58**, 113–144.
- Slade, M. E. (1989). Modelling stochastic and cyclical components of structural change: An application of the Kalman filter. *J. Econometrics* **41**, 363–383.
- Smith, R. L. and J. E. Miller (1986). A non-Gaussian state space model and application to prediction of records. *J. Roy. Statist. Soc. Ser. B* **48**, 79–88.
- Smith, T. M. F. (1978). Principles and problems in the analysis of repeated surveys. In: N. K. Nawoodi, ed., *Survey Sampling and Measurement*, Academic Press, New York, 201–216.
- Stock, J. H. and M. Watson (1988). Testing for common trends. *J. Amer. Statist. Assoc.* **83**, 1097–1107.
- Stock, J. H. and M. Watson (1990). A probability model of the coincident economic indicators. In: K. Lahiri and G. H. Moore, eds., *Leading Economic Indicators*, Cambridge Univ. Press, Cambridge, 63–89.
- Streibel, M. and A. C. Harvey (1993). Estimation of simultaneous equation models with stochastic trend components. *J. Econ. Dynamics Control* **17**, 263–288.
- Tam, S. M. (1987). Analysis of repeated surveys using a dynamic linear model. *Internat. Statist. Rev.* **55**, 63–73.
- Tiao, G. C. and R. S. Tsay (1989). Model specification in multivariate time series (with discussion). *J. Roy. Statist. Soc. Ser. B* **51**, 157–214.
- Tong, H. (1990). *Non-Linear Time Series: A Dynamic System Approach*. Clarendon Press, Oxford.
- Tunnicliffe-Wilson, G. (1989). On the use of marginal likelihood in time series estimation. *J. Roy. Statist. Soc. Ser. B* **51**, 15–27.
- West, M., P. J. Harrison and H. S. Migon (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *J. Amer. Statist. Assoc.* **80**, 73–97.
- West, M. and P. J. Harrison (1989). *Bayesian Forecasting and Dynamic Models*. Springer, New York.
- Whittle, P. (1991). Likelihood and cost as path integrals. *J. Roy. Statist. Soc. Ser. B* **53**, 505–538.