Biometrika (1995), **82**, 2, *pp.* 339–50
Printed in Great Britain

The simulation smoother for time series models

BY PIET DE JONG
Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, B.C., V6T 1Z2, Canada

AND NEIL SHEPHARD
Nuffield College, Oxford, OX1 1NF, U.K.

SUMMARY

Recently suggested procedures for simulating from the posterior density of states given a Gaussian state space time series are refined and extended. We introduce and study the simulation smoother, which draws from the multivariate posterior distribution of the disturbances of the model, so avoiding the degeneracies inherent in state samplers. The technique is important in Gibbs sampling with non-Gaussian time series models, and for performing Bayesian analysis of Gaussian time series.

Some key words: Gibbs sampling; Kalman filter; Simulation smoother; Smoothing; State space model.

1. INTRODUCTION

This paper introduces a simple and efficient method, the simulation smoother, for sampling from the smoothing density associated with time series models. The simulation smoother exploits the common structure which most time series models possess.

Simulation smoothing comes up in a number of papers, e.g. Frühwirth-Schnatter (1994, 1995), Carter & Kohn (1994), Shephard (1994), Chib & Greenberg (1994), and may be of importance in future time series work. One area of application is the Markov chain Monte Carlo method or Gibbs sampling, described by Smith & Roberts (1993) and Ripley (1987, pp. 113–6). General time series papers using these methods are by Carlin, Polson & Stoffer (1992), Carter & Kohn (1994) and Shephard (1994). Papers tackling specific cases of models using this framework include Chib (1993), Jacquier, Polson & Rossi (1994), Albert & Chib (1993), Shumway & Stoffer (1991) and Frühwirth-Schnatter (1994). Its application to robust nonparametric regression, which exploits the state space form for computational convenience, has been suggested by Carter & Kohn (1994).

The objective of Gibbs sampling (Tierney, 1995; Tanner, 1991, pp. 89–106) is to produce draws from the 'smoothing' density $p(\alpha, \omega | y)$, where y is the vector of observed data. The vectors α and ω contain quantities of interest or facilitate the estimation of the same. They are viewed as unobserved 'latent' data. There is flexibility in choosing α and ω, depending on the structure of the model, objectives of the analysis and estimation method. With Gibbs sampling, draws from $p(\alpha, \omega | y)$ are produced by cycling over the steps $\alpha \sim p(\alpha | y, \omega)$ and $\omega \sim p(\omega | y, \alpha)$, where each draw serves to redefine the conditioning variable on the next draw.

In time series applications α is usually defined as the stack of state vectors with respect to a state space form and $p(\alpha | y, \omega)$ is typically Gaussian. Carlin et al. (1992) propose that

the draw from $p(\alpha|y,\omega)$ be replaced by the 'single-state' Gibbs sampler which samples from $p(\alpha_t|y,\alpha^t,\omega)$, where α^t is α excluding α_t and cycling over t and likewise for the individual elements of ω. With the single-state sampler each draw serves to redefine a single state α_t in contrast to the 'multi-state' Gibbs sampler where the entire α is drawn at once from $p(\alpha|y,\omega)$.

Recently there has been a series of papers which suggest that the single state Gibbs sampler may be extremely inefficient for a class of time series models which includes many of the most interesting non-Gaussian and nonlinear models. In response, Carter & Kohn (1994) and Frühwirth-Schnatter (1994) have independently suggested a direct implementation of multi-state sampling for time series models. The implementation is based on the identity:

$$p(\alpha|y,\omega)=p(\alpha_n|y,\omega)p(\alpha_{n-1}|y,\alpha_n,\omega)\dots p(\alpha_0|y,\alpha_1,\dots,\alpha_n,\omega). \tag{1}$$

A draw from $p(\alpha|y,\omega)$ can thus be constructed recursively provided the subdraws from the densities in the right of (1) are practical. Carter & Kohn (1994) and Frühwirth-Schnatter (1994) show how to implement these subdraws if $p(\alpha|y,\omega)$ is Gaussian.

The multi-state sampler is simpler to implement than the single-state sampler and more efficient in that there is less correlation between successive α or ω draws and so it converges more quickly: see Liu, Kong & Wong (1994), who show that generating variables from reduced conditionals produces faster convergence. Shortcomings of the multi-state sampler as implemented by Carter & Kohn (1994) and Frühwirth-Schnatter (1994) flow from the fact that a draw from $p(\alpha|y,\omega)$ is constructed recursively in terms of α_t. Typically α is of very high dimension with many identities linking the state variables. The identities are a consequence of forcing the model into state space form. These identities must be kept track of in the direct recursive construction of the draw, necessitating mechanisms for dealing with degeneracies and imposing a large computational overhead.

In this paper we develop an alternative multi-state Gibbs sampler for time series models, in which it is disturbances rather than states which are drawn. Drawing disturbances is not subject to automatic identities, and is typically simpler. All random variables in the state space model are linear combinations of the disturbances, and hence can be constructed from simulated disturbances as required.

Section 2 illustrates the advantages of multi-state over single-state sampling. The multi-state sampler is constructed in terms of the disturbance draws. Section 3 gives the general algorithm for the disturbance sampler. Section 4 discusses some examples, and § 5 deals with regression effects. An appendix proves the correctness of the algorithm.

2. SINGLE VERSUS MULTI-STATE SAMPLING

2·1. *Illustration*

To illustrate the importance of multi-state sampling, consider the stochastic volatility model:

$$y_t=\varepsilon_t\exp(\tfrac{1}{2}\alpha_t),\quad \alpha_{t+1}=\phi\alpha_t+\eta_t,$$

where ε_t and η_t are mutually and serially independent Gaussian random variables with zero means and variances 1 and σ^2 respectively. This non-Gaussian state space model has been used to generalise the Black–Scholes option pricing equation to allow the

variance to change over time (Hull & White, 1987; Chesney & Scott, 1989; Melino & Turnbull, 1990).

The single-state sampler involves drawing α_t, ideally from its conditional distribution

$$p(\alpha_t|\alpha^t, y) = p(\alpha_t|\alpha_{t-1}, \alpha_{t+1}, y_t) = cp(y_t|\alpha_t)p(\alpha_{t+1}|\alpha_t)p(\alpha_t|\alpha_{t-1}),$$

where c, the constant of proportionality, is unknown. A good approach to overcoming the problem is suggested by Jacquier et al. (1994), building on the accept/reject Metropolis work of Carlin et al. (1992). However, we are going to avoid using this by noting that $\log\{p(\alpha_t|\alpha^t, y)\}$ is concave in α_t. This means we can directly sample from $p(\alpha_t|\alpha_{t-1}, \alpha_{t+1}, y_t)$ using the routine of Wild & Gilks (1993), and so use the single-state Gibbs sampler.

This Gibbs sampler will converge to drawings from $p(\alpha|y)$ so long as $\sigma^2 > 0$. However, the speed of convergence may be slow, in the sense of taking a large number of draws. To illustrate these features, Fig. 1 reports some results using two sets of parameter values: $\phi = 0{\cdot}9$, $\sigma^2 = 0{\cdot}05$ and $\phi = 0{\cdot}99$, $\sigma^2 = 0{\cdot}01$. These lines report the average over 1000 replications of α_t for $t = 0, 1, \ldots, 100$ after k iterations, all initialised at zero. The experiment shows how long the initial conditions last in the single-state sampler and so reflect the memory or correlation in the sampler. Figure 2 shows the results from running a single sampler for 100 000 iterations, discarding the first 10 000. The resulting correlogram of that series, recording the draw for α_{50}, compactly presents the correlation in the sampler in equilibrium.

The results of Fig. 1 show that as ϕ increases, and similarly as $\sigma^2 \to 0$, the sampler slows up, reflecting the increased correlation amongst the states drawn. This unfortunate characteristic of the single-state Gibbs sampler is common in state space models. If a component, such as α_t, changes slowly and persistently, the single-state sampler will be slow.

2·2. *Multi-state sampling*

To use the simulation smoother in this context we transform the Stochastic Volatility model (Harvey, Ruiz & Shephard, 1994; Shephard, 1994):

$$\log(y_t^2) = \alpha_t + \log(\varepsilon_t^2).$$

Carter & Kohn (1993) and Shephard (1994) have suggested using the multi-state samplers on this model by approximating the distribution of $\log(\varepsilon_t^2)$ by a mixture of normals, so that $p\{\log(\varepsilon_t^2)|\omega_t\}$ is Gaussian with mean $\mu(\omega_t)$ and variance $\sigma^2(\omega_t)$, where the ω_t are assumed independent and identically distributed integer random variables. The advantage of this representation of the model is that, conditionally on the ω_t, the state space model is Gaussian. It is possible to directly draw from $p(\alpha|y, \omega)$ using the simulation smoother developed in this paper. The smoother, specialised to the current model, is based on running forwards, for $t = 1, 2, \ldots, n$, the Kalman filter

$$e_t = \log(y_t^2) - \mu(\omega_t) - a_t, \quad d_t = p_t + \frac{\sigma^2(\omega_t)}{\sigma^2}, \quad k_t = \frac{p_t}{\phi d_t},$$

$$a_{t+1} = \phi a_t + k_t e_t, \quad p_{t+1} = 1 + \phi p_t(1 - k_t),$$

with starting conditions $a_1 = 0$, $p_1 = 1/(1-\phi^2)$. On the forward pass, only the scalars e_t, d_t and k_t are stored. These are respectively the innovation, innovation variance, and

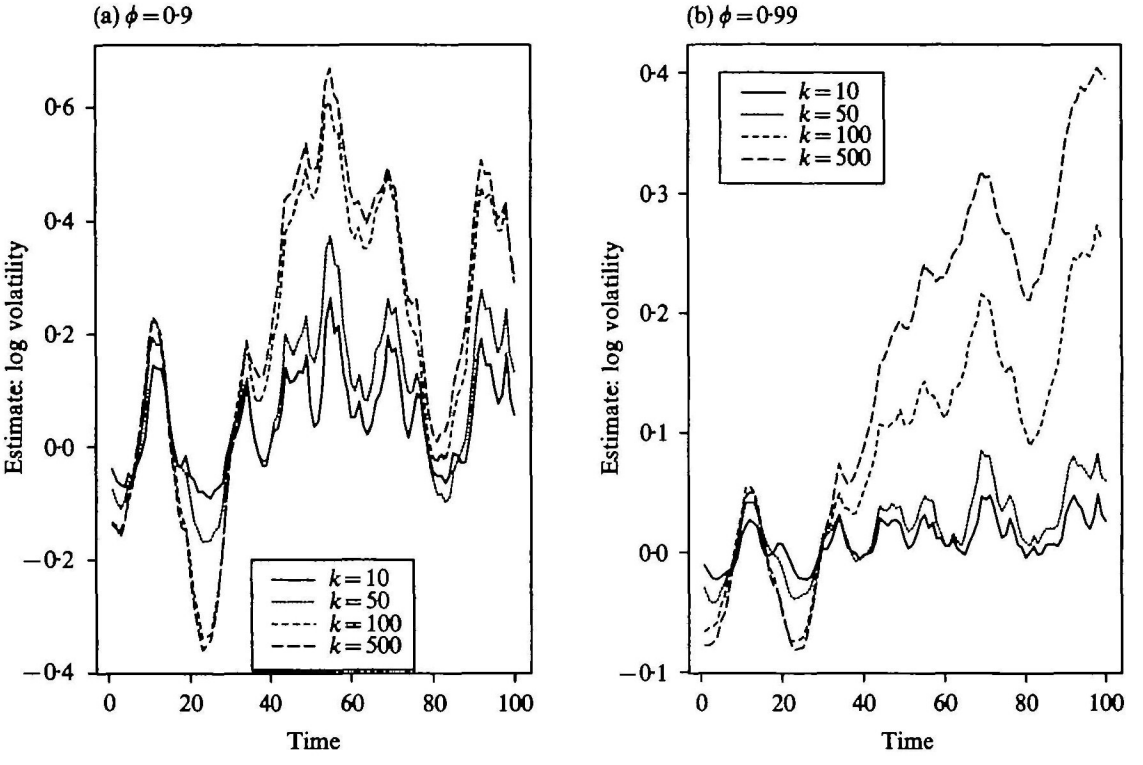

Fig. 1. Signal extraction of stochastic volatility model using single-move Gibbs, indicating rate of convergence.

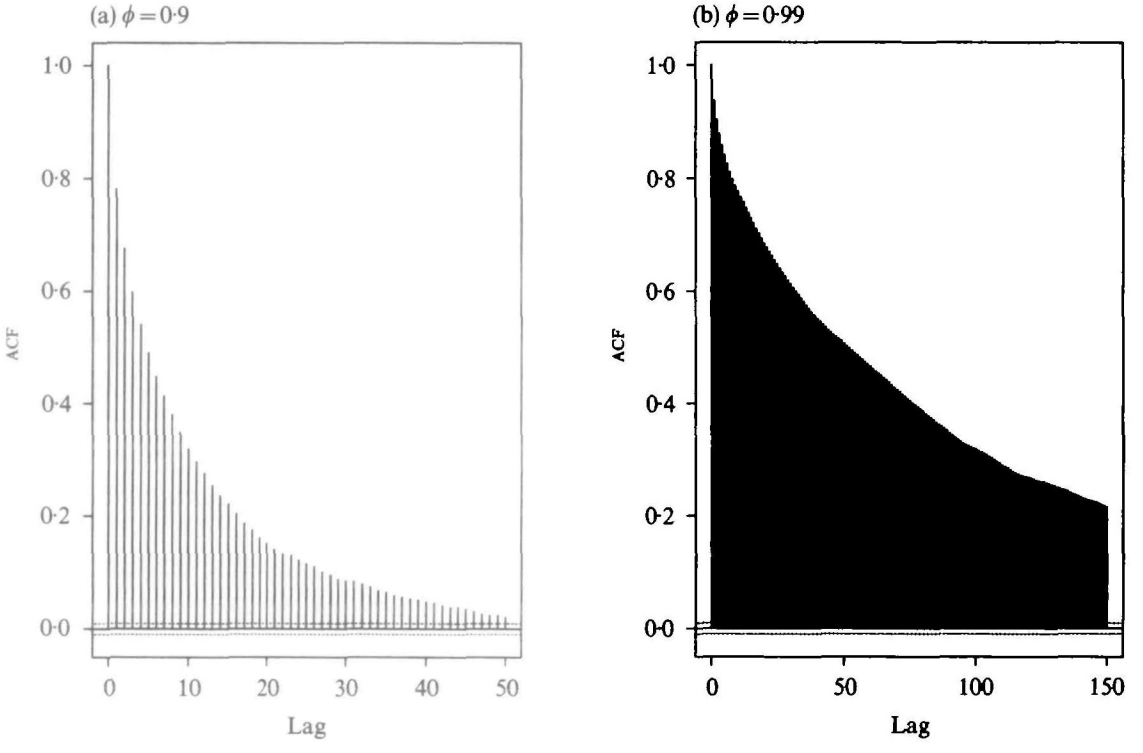

Fig. 2. Correlograms from Fig. 1.

Kalman gain. Then setting $r_n = 0$, $w_n = 0$, we compute, for $t = n, n-1, \ldots, 1$,

$$c_t = 1 - w_t, \quad \eta_t \sim N(r_t, \sigma^2 c_t), \quad v_t = w_t(\phi - k_t),$$

$$r_{t-1} = \frac{e_t}{d_t} + (\phi - k_t) r_t - \frac{v_t(\eta_t - r_t)}{c_t}, \quad w_{t-1} = \frac{1}{d_t} + (\phi - k_t)^2 w_t + \frac{v_t^2}{c_t},$$

where $\eta_t \sim N(r_t, \sigma^2 c_t)$ indicates that η_t is drawn from a Gaussian density with mean r_t and variance $\sigma^2 c_t$. The final η_0 is drawn from $N\{p_1 r_0, \sigma^2 p_1(1 - p_1 w_0)\}$. The simulated α vector is then constructed via the recursion $\alpha_{t+1} = \phi \alpha_t + \eta_t$ $(t = 0, 1, \ldots, n-1)$ starting with $\alpha_0 = 0$.

The simulation smoother is based on simulating from the posterior distribution of the disturbances of the model, η_t, which then allow us, as required, to form the simulation from the states α_t. Simulation smoothing is the analogue of the recent analytic disturbance smoothers, introduced partially by de Jong (1988) and elaborated by Koopman (1993). In fact, putting $v_t = 0$ for all t gives the Koopman algorithm to compute $E(\eta_t | y, \omega)$ and cov $(\eta_t | y, \omega)$.

Likewise, it is easy to draw from $p(\omega | y, \alpha)$ using uniform random numbers. This means it is possible to use a multi-state sampler on this model. Figures 3 and 4 repeat the experiments of Figs 1 and 2, but now using the multi-state sampler.

It is possible to argue that, although there are substantial differences between Figs 1 and 3, and particularly Figs 2 and 4, it is not worth the trouble of blocking and using the multi-state sampler. This is a dangerous view. The inclusion of any really slowly moving component will retard the single-state sampler unacceptably.

3. The simulation smoother

We shall use lower case letters to denote column vectors, and upper case letters for matrices. Dimensions may vary with time. If A and B have the same number of columns, then $(A; B) = (A', B')'$ denotes the matrix formed by placing A above B.

We work with the following general model. Conditional on $\omega = (\omega_0; \omega_1; \ldots; \omega_n)$, it is supposed that y_t is generated by the state space model (de Jong, 1991):

$$y_t = X_t \beta + Z_t \alpha_t + G_t u_t \quad (t = 1, 2, \ldots, n),$$

$$\alpha_{t+1} = W_t \beta + T_t \alpha_t + H_t u_t \quad (t = 0, 1, \ldots, n),$$

where $\alpha_0 = 0$. The coefficients matrices may depend, implicitly, on ω_t. The u_t are independent $N(0, \sigma^2 I)$ variables.

Initially we suppose β is known: the case of unknown β is discussed in § 5. Our simulation smoother draws $\eta \sim p(\eta | y, \omega)$, where $\eta = (\eta_0; \eta_1; \ldots; \eta_n)$ with $\eta_t = F_t u_t$ and the F_t are arbitrary matrices whose choice is discussed below. Initially we run, for $t = 1, 2, \ldots, n$, the Kalman filter

$$\begin{gathered} e_t = y_t - X_t \beta - Z_t a_t, \quad D_t = Z_t P_t Z_t' + G_t G_t', \quad K_t = (T_t P_t Z_t' + H_t G_t') D_t^{-1}, \\ a_{t+1} = W_t \beta + T_t a_t + K_t e_t, \quad P_{t+1} = T_t P_t L_t' + H_t J_t', \end{gathered} \tag{2}$$

where $a_1 = W_0 \beta$, $P_1 = H_0 H_0'$, $L_t = T_t - K_t Z_t$ and $J_t = H_t - K_t G_t$. On this Kalman filter pass, the quantities e_t, D_t and K_t are stored. Then setting $r_n = 0$ and $U_n = 0$, we run, for $t = n, n-1, \ldots, 1$,

$$\begin{gathered} C_t = F_t(I - G_t' D_t^{-1} G_t - J_t' U_t J_t) F_t', \quad \varepsilon_t \sim N(0, \sigma^2 C_t), \quad V_t = F_t(G_t' D_t^{-1} Z_t + J_t' U_t L_t), \\ r_{t-1} = Z_t' D_t^{-1} e_t + L_t' r_t - V_t' C_t^{-1} \varepsilon_t, \quad U_{t-1} = Z_t' D_t^{-1} Z_t + L_t' U_t L_t + V_t' C_t^{-1} V_t, \end{gathered} \tag{3}$$

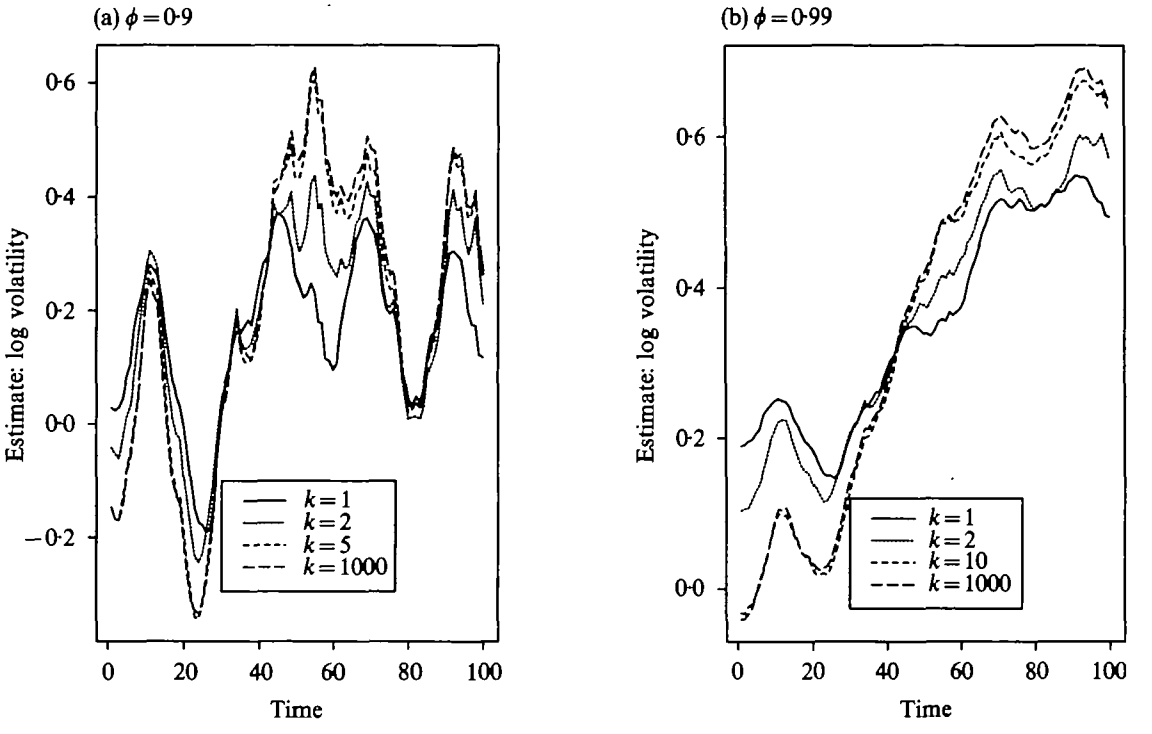

Fig. 3. Signal extraction of stochastic volatility model using multi-move Gibbs, indicating rate of convergence.

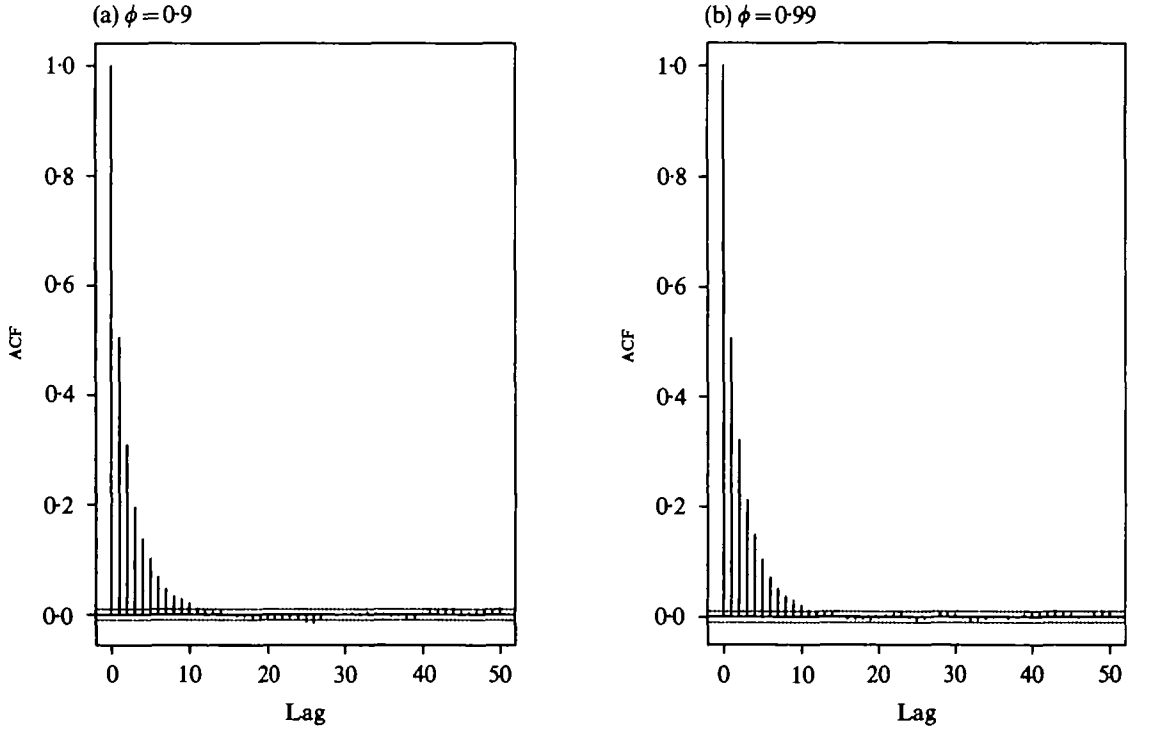

Fig. 4. Correlograms from Fig. 3.

and store $\eta_t = F_t(G_t' D_t^{-1} e_t + J_t' r_t) + \varepsilon_t$, where we take $G_0 = 0$. The vector $\eta = (\eta_0; \eta_1; \ldots; \eta_n)$ is a draw from $p(\eta | y, \omega)$. The proof is given in the Appendix.

If $F_t = I$ then η_t is drawn from $p(u_t | y, \omega)$. This choice is typically not optimal since, as shown in the next section, it implies that C_t is singular for some t.

If $F_t = G_t$ or $F_t = H_t$ then the sampled η_t correspond to the measurement noise $G_t u_t$ or state noise $H_t u_t$ respectively. The filters (2) and (3) are conceptually easier in this case provided measurement and state noise are uncorrelated, $G_t H_t' = 0$. For example if $F_t = H_t$, a sensible choice for most models, then, putting $\Omega_t = H_t H_t'$, (3) reduces to

$$\begin{gathered} C_t = \Omega_t - \Omega_t U_t \Omega_t, \quad \varepsilon_t \sim N(0, \sigma^2 C_t), \quad V_t = \Omega_t U_t L_t, \\ r_{t-1} = Z_t' D_t^{-1} e_t + L_t' r_t - V_t' C_t^{-1} \varepsilon_t, \quad U_{t-1} = Z_t' D_t^{-1} Z_t + L_t' U_t L_t + V_t' C_t^{-1} V_t, \end{gathered} \tag{4}$$

and $\eta_t = \Omega_t r_t + \varepsilon_t$ is a draw from $p(H_t u_t | y, \omega)$. Similarly if $F_t = G_t$ then, putting $\Gamma_t = G_t G_t'$, $n_t = D_t^{-1} e_t - K_t' r_t$ and $N_t = D_t^{-1} + K_t' U_t K_t$, (3) becomes

$$\begin{gathered} C_t = \Gamma_t - \Gamma_t N_t \Gamma_t, \quad \varepsilon_t \sim N(0, \sigma^2 C_t), \quad V_t = \Gamma_t (N_t Z_t - K_t' U_t T_t), \\ r_{t-1} = Z_t' D_t^{-1} e_t + L_t' r_t - V_t' C_t^{-1} \varepsilon_t, \quad U_{t-1} = Z_t' D_t^{-1} Z_t + L_t' U_t L_t + V_t' C_t^{-1} V_t, \end{gathered} \tag{5}$$

and $\eta_t = \Gamma_t n_t + \varepsilon_t$ is a draw from $p(G_t u_t | y, \omega)$, or, equivalently, $y_t - \eta_t$ is a draw from the conditional 'signal' distribution, $p(X_t \beta + Z_t \alpha_t | y, \omega)$.

The recursion (3) generalises analytic smoothing in the sense that, if V_t and ε_t are set to zero for each t, then (3) reduces to the smoothing recursions given by de Jong (1988, 1989), Kohn & Ansley (1989) and Koopman (1993), and the computed η_t equal $E(F_t u_t | y, \omega)$ while $\sigma^2 C_t = \text{cov}(F_t u_t | y, \omega)$. More particularly if V_t and ε_t are set to zero in (4) or (5) then we obtain the analytic smoothed disturbances of Koopman (1993) and de Jong (1988); that is η_t equals $E(H_t u_t | y, \omega)$ or $E(G_t u_t | y, \omega)$, with associated mean squared error matrix $\sigma^2 C_t$.

The stored Kalman filter quantities e_t, D_t and K_t are, respectively, the innovation vector, scaled innovation covariance matrix and Kalman gain matrix. On the simulation smoothing pass (3), η_t is a draw from $p(F_t u_t | y, \eta_{t+1}, \ldots, \eta_n, \omega)$ since $F_t(G_t' D_t^{-1} e_t + J_t' r_t)$ and $\sigma^2 C_t$ are the conditional mean and covariance matrix of this Gaussian density. Thus the η draw is built up using the decomposition (1), written in terms of η rather than α.

The advantages of the simulation smoother over the state sampler are as follows. First, the storage requirements are typically much less than that required for the state sampler, which stores the one-step-ahead state vector estimates a_t and associated scaled covariance matrices P_t. Secondly, the recursion operates in minimal dimension and does not require inversion of the matrix P_t, which is typically of large dimension and/or singular. Thirdly, the recursion can be operated in square root form, enhancing numerical stability. Fourthly, provided the F_t are of full row rank, there are no automatic degeneracies in $p(\eta | \omega)$. In turn, degeneracies in $p(\eta | y, \omega)$ can always be avoided through a transparent choice of the F_t, as discussed in § 4. Finally, it is often easier to draw from $p(\omega | y, \eta)$ as opposed to $p(\omega | y, \alpha)$.

4. Examples

Example 4·1: MA (q) *model.* This model can be written in state space form with $G_t \equiv 1$ and $H_t \equiv (h_1, \ldots, h_q)'$, where $(h_1, \ldots, h_q)$ is the vector of moving average coefficients with $h_q \neq 0$. If $F_t = I = 1$ then the ε_t and C_t are scalar. However the only nonzero C_t are C_n, $C_{n-1}, \ldots, C_{n-q+1}$ since, given y and $u_n, u_{n-1}, \ldots, u_{n-q+1}$, the remaining u_t are determined

recursively from

$$u_t = (y_{t+q} - u_{t+q} - h_1 u_{t+q-1} - \ldots - h_{q-1} u_{t+1})/h_q.$$

Hence a draw from $p(u|y)$ amounts to drawing q random variables $u_n, u_{n-1}, \ldots, u_{n-q+1}$, which go on to determine the remaining disturbances.

Example 4·2: *Stochastic volatility model.* Consider the stochastic volatility model transformed into the partial Gaussian state space form discussed in § 2, written as

$$\log(y_t^2) = \mu(\omega_t) + \alpha_t + g(\omega_t)u_{1t}, \quad \alpha_{t+1} = \phi\alpha_t + u_{2t},$$

where $g(\omega_t) = \sigma(\omega_t)/\sigma$ and the $u_t = (u_{1t}; u_{2t})$ are independent $N(0, \sigma^2 I)$ vectors. Thus $\log(y_{t+1}^2) - \log(y_t^2)$ is, apart from the mean,

$$\alpha_{t+1} - \alpha_t + g(\omega_{t+1})u_{1,t+1} - g(\omega_t)u_{1t} = u_{2t} + g(\omega_{t+1})u_{1,t+1} - g(\omega_t)u_{1t}.$$

Hence, for $t = n-1, \ldots, 1$, $g(\omega_t)u_{1t}$ is determined from subsequent u_t and data y. Hence if $F_t = ((g(\omega_t), 0); (0, 1))$, then C_n is nonsingular but $C_{n-1}, \ldots, C_1$ are of rank 1. Further $C_0 = 0$ since

$$H_0 u_0 = \alpha_1 = \log(y_1^2) - \mu(\omega_1) - g(\omega_1)u_{11}.$$

Thus drawing from the $2n+1$ dimensional density $p(\eta|y, \omega)$ amounts to drawing $n+1$ Gaussian random variables.

An issue that arises in running (3) is the occurrence of singularities in one or more of the C_t. These singularities will always occur if, for each t, the rows of F_t span the row space of $(G_t; H_t)$. For then, without loss of generality we may assume that $(G_t; H_t)$ is of full column rank, which implies u_t can be determined from $F_t u_t$, and hence we may assume $F_t = I$ implying $\eta = u = (u_0; u_1; \ldots; u_n)$. The vector u satisfies $y = X\beta + Gu$ for some matrices X and G. Thus $Gu = y - X\beta$ and hence $p(u|y, \omega)$ is degenerate. Thus the sampled vector η in (3) is subject to linear constraints. These constraints reflect themselves in the C_t, since

$$p(\eta|y, \omega) = p(\eta_n|y, \omega)p(\eta_{n-1}|y, \eta_n, \omega) \ldots p(\eta_0|y, \eta_1, \ldots, \eta_n, \omega),$$

and $\sigma^2 C_t$ is the covariance matrix associated with the conditional density on the right corresponding to η_t. Thus, if for each t, the rows of F_t span the row space of $(G_t; H_t)$ then C_t will be singular for at least some t.

It is possible to deal with singular C_t in (3) via generalised inversion. However it is more sensible to choose F_t to have rows making up a basis for the row space of H_t, as this often avoids singularities in the C_t. The next two examples provide illustrations.

Example 4·2: *Stochastic volatility model (cont.).* For this model put $F_t = H_t = (0, 1)$. Then $G_t = (g(\omega_t), 0)$ is not in the row space of F_t. In this case all the C_t in (2) are scalar of rank 1. Drawing from $p(\eta|y, \omega)$ reduces to drawing $n+1$ Gaussian random variables. Now C_t is scalar and nonzero for all t. The sampler given in § 2 for the stochastic volatility model is (2) and (3) specialised to this case.

Example 4·3: *Seasonal model.* Consider the stochastic volatility model given above with the addition of a seasonal term

$$s_{t+1} = -\sum_{j=1}^{p-1} s_{t+1-j} + hu_{3t}.$$

In this case $G_t = (g(\omega_t), 0, 0)$, while H_t is $(p+1) \times 3$ with p the number of seasons. The first two rows of H_t are (0, 1, 0) and (0, 0, h) while the rest are zero, reflecting the identities amongst state elements. Form F_t from the first two rows of H. All the C_t are nonsingular, and the drawn ε_t in (2) are nondegenerate random vectors of length 2. The state vector for this model is of length $p+1$, and state sampling, if p is say 12, is highly inefficient compared to disturbance sampling.

In general if, for each t, the rows of F_t form a basis for the row space of H_t then $H_t u_t$ can be determined from η_t. In turn the α_t and $G_t u_t$ can be determined from the $H_t u_t$ via the equations

$$\alpha_{t+1} = W_t\beta + T_t\alpha_t + H_t u_t, \quad G_t u_t = y_t - X_t\beta - Z_t\alpha_t,$$

which can be conveniently computed alongside the next Kalman filter pass. The advantage of this choice for F_t is that if there is measurement noise, that is if for each t the rows of G_t are not in the row space of H_t, then C_t will be nonsingular.

Example 4·4: *Multiplicative model.* Consider the model

$$y_t = \omega_t\alpha_t + gu_{1t}, \quad \alpha_{t+1} = \alpha_t + hu_{2t}, \quad \omega_{t+1} = \rho\omega_t + v_t,$$

where v_t is Gaussian noise with variance σ^2, uncorrelated with the u_t. Conditional on ω this is a Gaussian state space model and hence the scalar $\eta_t = hu_{2t}$ can be drawn as described above. In turn the simulated α_t can be built up in a forward pass from the equation $\alpha_{t+1} = \alpha_t + \eta_t$. Conditional on η and hence the α, the first and third equation define a Gaussian state space model and the above shows how to draw v_t, which in turn can be used to build up the draw of ω_t conditional on y and α.

5. Regression effects

Situations with unknown regression vector β can be handled by supposing $p(\beta|\omega)$ is Gaussian with $E(\beta|\omega) = b$ and $\text{cov}(\beta|\omega) = \sigma^2 BB'$, which may be singular. In this case the draw $\eta \sim p(\eta|y, \omega)$ is built up using the decomposition

$$p(\beta, \eta|y, \omega) = p(\beta|y, \omega)p(\eta|\beta, y, \omega).$$

Drawing from $p(\eta|\beta, y, \omega)$ was, in effect, considered in §§ 2 and 3. Drawing from $p(\beta|y, \omega)$ is achieved by replacing the equations for e_t and a_{t+1} in (2) by

$$E_t = (0, y_t) - X_t(B, b) - Z_t A_t, \quad A_{t+1} = W_t(B, b) + T_t A_t + K_t E_t$$

and adding the recursion $Q_{t+1} = Q_t + E_t' D_t^{-1} E_t$, where $A_1 = W_0(B, b)$ and $Q_1 = 0$. Thus the Kalman filter (2) is modified in two of its equations, yielding the so-called diffuse Kalman filter. It is shown by de Jong (1991) that $p(\beta|y, \omega)$ is Gaussian with mean and covariance matrix given by

$$E(\beta|y, \omega) = b + B(S+I)^{-1}s, \quad \text{cov}(\beta|y, \omega) = \sigma^2 B(S+I)^{-1}B',$$

where S and s are defined such that the matrix $(S, -s)$ equals the matrix formed from the k top rows of the $k+1$ order matrix Q_{n+1}.

Thus a draw from $p(\beta, \eta|y, \omega)$ is made as follows. First the Kalman filter (4) is run, modified as described above. This is followed by the draw

$$\delta \sim N\{(S+I)^{-1}s, \sigma^2(S+I)^{-1}\}.$$

Given δ, a draw η is made with the simulation smoothing pass (3), where $e_t = E_t(\delta; 1)$.

The above steps assume $p(\beta|\omega)$ is a proper prior for β. In some cases it is reasonable to suppose a vague or improper prior for β: a Gaussian density with $\text{cov}(\beta|\omega) = \sigma^2\kappa BB'$, where $\kappa \to \infty$. In this case, δ is drawn from $N(S^{-1}s, \sigma^2 S^{-1})$ and the draw from $p(\eta|\beta, y, \omega)$ proceeds as above. A similar treatment can be given for the case where α_1 has a vague prior.

Example 5·1: *Stochastic volatility model.* We continue with the stochastic volatility model transformed into the partial Gaussian state space form discussed in §§ 2 and 4. Suppose $\mu(\omega_t)$ as specified in § 2·2 is of the form $x'(\omega_t)\beta$, where β is unknown and $x(\omega_t)$ a vector of zeros except in one position where it contains a one, indicating the state of ω_t. If β is completely unknown we put $b = 0$ and $B = I$ and, at the completion of the modified Kalman filter pass (2), draw β from the Gaussian density $N(S^{-1}s, \sigma^2 S^{-1})$. The simulation smoother (3) is then applied with $e_t = E_t(\beta; 1)$. The resulting vector $(\beta; \eta)$ is a draw from $p(\beta, \eta|y, \omega)$ and a sequence of such draws, cycling over draws from $p(\beta, \eta|y, \omega)$ and $p(\omega|y, \beta, \eta)$, behaves like draws from the posterior $p(\beta, \eta|y)$.

ACKNOWLEDGEMENT

The comments of the referees are gratefully acknowledged. The authors would like to thank NSERC and ESRC for financial support.

APPENDIX

Proof of correctness

Conditioning on ω will be implicit. Put

$$\eta_t = F_t u_t \quad (t = 0, 1, \ldots, n), \quad \varepsilon_n = \eta_n - E(\eta_n|y).$$

Then, since $\sigma^2 D_n = \text{cov}(e_n)$ and $\text{cov}(u_n, e_n) = \sigma^2 G_n'$,

$$E(\eta_n|y) = E(\eta_n|e) = E(\eta_n|e_n) = \sigma^{-2}\,\text{cov}(\eta_n, e_n)D_n^{-1}e_n = F_n G_n' D_n^{-1} e_n,$$

$$\text{cov}(\eta_n|y) = \text{cov}(\varepsilon_n) = \text{cov}(\eta_n, \eta_n - F_n G_n' D_n^{-1} e_n) = \sigma^2 F_n(I - G_n' D_n^{-1} G_n)F_n = \sigma^2 C_n,$$

and hence the assertions of § 3 hold for $t = n$. For $t < n$ put $\varepsilon_t = \eta_t - E(\eta_t|y, \eta_{t+1}, \ldots, \eta_n)$ and suppose the assertions hold for $s > t$. Then

$$\begin{aligned}
E(\eta_t|y, \eta_{t+1}, \ldots, \eta_n) &= E(\eta_t|e_t, e_{t+1}, \ldots, e_n, \eta_{t+1}, \ldots, \eta_n) \\
&= \sum_{s=t}^{n} E(\eta_t|e_s) + \sum_{s=t+1}^{n} E\{\eta_t|\eta_s - E(\eta_s|e_t, \ldots, e_n, \eta_{s+1}, \ldots, \eta_n)\} \\
&= E(\eta_t|e_t) + \sum_{s=t+1}^{n} \{E(\eta_t|e_s) + E(\eta_t|\varepsilon_s)\} \\
&= F_t\left[G_t' D_t^{-1} e_t + \sigma^{-2} \sum_{s=t+1}^{n} \{\text{cov}(u_t, e_s)D_s^{-1}e_s + \text{cov}(u_t, \varepsilon_s)C_s^{-1}\varepsilon_s\}\right].
\end{aligned}$$

The first equality follows on orthogonalising y, since $\eta_t = F_t u_t$ is uncorrelated with $e_1, e_2, \ldots, e_{t-1}$. The second follows from orthogonalising $e_t, e_{t+1}, \ldots, e_n, \eta_{t+1}, \ldots, \eta_n$. It is shown below that, for $s > t$,

$$\text{cov}(e_s, u_t) = \sigma^2 Z_s(L_{s-1} \ldots L_{t+1})J_t, \tag{A1}$$

$$\text{cov}(\varepsilon_s, u_t) = -\sigma^2 V_s(L_{s-1} \ldots L_{t+1})J_t. \tag{A2}$$

Substituting into the above expression yields $E(\eta_t|y, \eta_{t+1}, \ldots, \eta_n)$ as given in § 3. Next note that

$$\begin{aligned}\operatorname{cov}(\eta_t|y, \eta_{t+1}, \ldots, \eta_n) &= \operatorname{cov}(\varepsilon_t) = \operatorname{cov}\{\eta_t, \eta_t - E(\eta_t|y, \eta_{t+1}, \ldots, \eta_n)\} \\ &= \sigma^2 F_t\{I - \operatorname{cov}(u_t, G_t'D_t^{-1}e_t + J_t'r_t)\}F_t' \\ &= \sigma^2 F_t\{I - \operatorname{cov}(u_t, e_t)D_t^{-1}G_t - \operatorname{cov}(u_t, r_t)J_t\}F_t',\end{aligned}$$

which expands to $\sigma^2 C_t$.

It remains to prove the relations in (A1) and (A2). For $s > t$,

$$\begin{aligned}e_s &= Z_s(\alpha_s - a_s) + G_s u_s = Z_s\{L_{s-1}(\alpha_{s-1} - a_{s-1}) + J_{s-1}u_{s-1}\} + G_s u_s \\ &= Z_s(L_{s-1}\ldots L_{t+1})\{(\alpha_t - a_t) + J_t u_t\} + (\text{terms linear in } u_{t+1}, \ldots, u_s).\end{aligned}$$

Thus

$$\operatorname{cov}(e_s, u_t) = Z_s(L_{s-1}\ldots L_{t+1})\operatorname{cov}(J_t u_t, u_t) = \sigma^2 Z_s(L_{s-1}\ldots L_{t+1})J_t,$$

which establishes (A1). Next note that

$$\operatorname{cov}(\varepsilon_n, u_t) = \operatorname{cov}(u_n - G_n'D_n^{-1}e_n, u_t) = -\sigma^2 G_n'D_n^{-1}Z_n(L_{n-1}\ldots L_{t+1})J_t,$$

which proves (A2) for $s = n$. For $t < s < n$, using the inductive hypothesis,

$$\begin{aligned}\operatorname{cov}(\varepsilon_s, u_t) &= \operatorname{cov}(u_s - G_s'D_s^{-1}e_s - J_s'r_s, u_t) \\ &= -G_s'D_s^{-1}\operatorname{cov}(e_s, u_t) - J_s'\operatorname{cov}(r_s, u_t) \\ &= -\sigma^2 G_s'D_s^{-1}Z_s(L_{s-1}\ldots L_{t+1})J_t - J_s'\sum_{j=s+1}^{n}(L_{s+1}'\ldots L_{j-1}') \\ &\quad \times \{Z_j'D_j^{-1}\operatorname{cov}(e_j, u_t) - V_j'C_j^{-1}\operatorname{cov}(\varepsilon_j, u_t)\}.\end{aligned}$$

Assuming the truth of the expression for $\operatorname{cov}(\varepsilon_j, u_t)$ for $j > s$ yields

$$\operatorname{cov}(\varepsilon_s, u_t) = -\sigma^2(G_s'D_s^{-1}Z_s + J_s'U_sL_s)(L_{s-1}\ldots L_{t+1})J_t = -\sigma^2 V_s(L_{s-1}\ldots L_{t+1})J_t,$$

which completes the proof.

References

Albert, J. H. & Chib, S. (1993). Bayesian inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *J. Econ. Bus. Statist.* **11**, 1–15.

Carlin, B. P., Polson, N. G. & Stoffer, D. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *J. Am. Statist. Assoc.* **87**, 493–500.

Carter, C. K. & Kohn, R. (1993). On the applications of Markov Chain Monte Carlo methods to linear state space models. In *Proc. Bus. Econ. Statist. Sect.*, pp. 131–6. Washington, D.C.: Am. Statist. Assoc.

Carter, C. K. & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541–53.

Chesney, M. & Scott, L. O. (1989). Pricing European options: a comparison of the modified Black–Scholes model and a random variance model. *J. Financ. Qualitat. Anal.* **24**, 267–84.

Chib, S. (1993). Bayes regression with autoregression errors: A Gibbs sampling approach. *J. Econom.* **58**, 275–94.

Chib, S. & Greenberg, E. (1994). Bayes inference in regression models with ARMA (p, q) errors. *J. Econom.* **64**, 183–206.

de Jong, P. (1988). 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. Am. Statist. Assoc.* **84**, 1085–8.

de Jong, P. (1991). The diffuse Kalman filter. *Ann. Statist.* **2**, 1073–83.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *J. Time Ser. Anal.* **15**, 183–202.

Frühwirth-Schnatter, S. (1995). Bayesian model discrimination and Bayes factors for linear Gaussian state space models. *J. R. Statist. Soc.* B **57**, 237–46.

Harvey, A. C., Ruiz, E. & Shephard, N. (1994). Multivariate stochastic variance models. *Rev. Econ. Studies* **61**, 247–64.

HULL, J. & WHITE, A. (1987). Hedging the risk for writing foreign currency options. *J. Int. Money Finance* **6**, 131–52.

JACQUIER, E., POLSON, N. G. & ROSSI, P. E. (1994). Bayesian analysis of stochastic volatility. *J. Bus. Econ. Statist.* **12**, 371–417.

KOHN, R. & ANSLEY, C. F. (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**, 117–26.

LIU, J., WONG, W. H. & KONG, A. (1994). Correlation structure and convergence rate of the Gibbs sampler (I): applications to the comparison of estimators and augmentation schemes. *Biometrika* **81**, 27–40.

MELINO, A. & TURNBULL, S. M. (1990). Pricing foreign currency options with stochastic volatility. *J. Economet.* **45**, 239–65.

RIPLEY, B. D. (1987). *Stochastic Simulation.* New York: Wiley.

SHEPHARD, N. (1994). Partial non-Gaussian state space. *Biometrika* **81**, 115–31.

SHUMWAY, R. H. & STOFFER, D. S. (1991). Dynamic linear models with switching. *J. Am. Statist. Assoc.* **86**, 763–9.

SMITH, A. F. M. & ROBERTS, G. (1993). Bayesian computations via the Gibbs sampler and related Markov Chain Monte Carlo Methods. *J. R. Statist. Soc.* B **55**, 3–23.

TANNER, M. A. (1991). *Tools for Statistical Inference, Observed Data and Data Augmentation Methods.* New York: Springer-Verlag.

TIERNEY, L. (1995). Markov chain for exploring posterior distributions. *Ann. Statist.* To appear.

WILD, P. & GILKS, W. R. (1993). AS 287: Adaptive rejection sampling from log-concave density functions. *Appl. Statist.* **42**, 701–9.

[*Received December* 1993. *Revised December* 1994]