JOINT AND ANTITHETIC MCMC FOR NON-GAUSSIAN MEASUREMENTS WITH APPLICATIONS TO STOCHASTIC VOLATILITY

MICHAEL K PITT

Department of Statistics, University of Oxford, OX1 3TG and *Nuffield College, Oxford, OX1 1NF, UK*

AND

NEIL SHEPHARD

Nuffield College, Oxford, OX1 1NF, UK

August 6, 1996

Abstract

In this paper we examine methods for improving the efficiency of sampling the posterior distribution of the parameters of a non-Gaussian measurement models by simultaneously drawing the parameters and the states. We suggest that this is an effective strategy for reducing autocorrelation for MCMC methods. We also investigate the effect of attempting to induce negative correlation in the Metropolis chain by the use of antithetic variables. This is less successful in our experiments. The stochastic volatility model is considered as a motivating example throughout.

KEYWORDS: Antithetic variables; Blocking; Kalman filter; Metropolis sampling; Non-Gaussian measurement models; Simulation smoother; Stochastic volatility.

1 Introduction

1·1 *Inference via simulation*

Markov chain Monte Carlo (MCMC) methods produce simulations from high dimensional non-Gaussian posterior distributions. Early examples of their use in image analysis includes Geman & Geman (1984). The methods are discussed for more widespread statistical implementation by Ripley (1987) and Gelfand & Smith (1990). A thorough outline of these methods is provided in Chib & Greenberg (1995). In addition, many examples of the current use of MCMC techniques are given in the booklength review by Gilks et al. (1996).

In this paper we use the Metropolis algorithm to simulate from the joint distribution of $x_1, x_2, .., x_m$, denoted f, where x_i stands for a group of unobserved states or parameters in the model, given the data. Proposals z are made to possibly replace the current x_i, keeping constant $x_{\setminus i}$ which denotes all the other elements of the x vector. The proposal density is proportional to $q(z, x_{\setminus i})$ while the true density is $f(x_i|x_{\setminus i})$. Both of these densities are assumed to be everywhere positive (Tierney (1994)). If $x^{(k)}$ is the current state of the sampler then the proposal is accepted with probability:

$$\min\left\{\frac{f(z|x_{\backslash i}^{(k)})q(x_i^{(k)},x_{\backslash i}^{(k)})}{f(x_i^{(k)}|x_{\backslash i}^{(k)})q(z,x_{\backslash i}^{(k)})},1\right\}.$$

In our work we would like to select q to be $f(z|x_{\backslash i})$, but this density is generally difficult to sample directly. Instead we sometimes approximate f by $ch(z)$, then we can sample from a density proportional to $\min\left\{f(z|x_{\backslash i}),ch(z)\right\}$ by the scheme: (1) Generate a candidate value z from $h(\cdot)$ and a value u from a standard uniform; (2) If $u \leq f(z|x_{\backslash i})/ch(z)$ return z, otherwise goto (1). Tierney (1994) calls this type of method a pseudo-dominating rejection algorithm. It is discussed in more detail in Chib & Greenberg (1995).

MCMC sampling is now increasingly used as a tool to fit non-Gaussian state space models. The earliest reference to the use of single move Gibbs sampling seems to be Carlin et al. (1992) who noted the conditional independence structure of the state space models. More recently Fruhwirth-Schnatter (1994) used Gibbs sampling to draw from the posterior density of the parameters in Gaussian state space form (GSSF) models. Often the focus of interest is on the parameters of the state space models rather than the unobserved states, $\alpha = (\alpha_1', \alpha_2', ..., \alpha_n')'$. We are therefore interested in drawing samples from $f(\Psi|y)$, the posterior distribution of the parameters, Ψ, given the observations, $y = (y_1', y_2', .., y_n')'$. For GSSF models (see Harvey (1989) and Hannan & Deistler (1988))

$$\begin{array}{lcl} y_t & = & c_t + Z_t\alpha_t + G_t u_t, \quad u_t \sim \mathsf{NID}(0, I), \\ \alpha_{t+1} & = & d_t + T_t\alpha_t + H_t u_t, \quad t = 1, ..., n \\ \alpha_1|Y_0 & \sim & N(a_{1|0}, P_{1|0}), \end{array} \tag{1·1}$$

this density is available up to a constant of proportionality since we know the likelihood $f(y|\Psi)$ through the Kalman filter (KF). When the Gaussianity of the measurement equation (or the state equation) breaks down we cannot in general analytically integrate over the states. We therefore need to apply MCMC methods to sample the states and the parameters ensuring that the resulting samples converge to samples arising from the required posterior $f(\alpha, \Psi|y)$.

1·2 *Non-Gaussian measurement time series*

In this paper we investigate this sampling problem for a specific class of non-Gaussian measurement time series. First define the signal as $s_t = c_t + Z_t\alpha_t$, then we will assume that

$$l(y_t|\alpha_t) = l(\alpha_t) = \log f(y_t|\alpha_t) = \log f(y_t|s_t)$$

is twice differentiable with respect to s_t and is parameter free. This second assumption can be achieved by appropriate parameterisation. We also assume that the transition equation of the states, α_t, remains a linear Gaussian Markov Chain as in (1·1). Often we will employ a concavity assumption on $\log f(y_t|s_t)$, however when this is relaxed the methods outlined in this paper are still valid, although their efficiency may be reduced. In this paper, we shall now assume that y_t and s_t are univariate for compactness of exposition.

To illustrate the type of non-Gaussian model under consideration we shall apply our methods to the stochastic volatility (SV) model on a dataset recording the daily returns on holding the Pound Sterling/US Dollar exchange rate from 1/10/81 to 28/6/85 (946 observations). The univariate SV model, from Taylor (1986), is written as

$$y_t = \epsilon_t \exp(\alpha_t/2), \quad \alpha_{t+1} = \mu + \phi(\alpha_t - \mu) + \eta_t, \tag{1·2}$$

$$\eta_t \sim \mathsf{NID}(0, \sigma_\eta^2), \quad \epsilon_t \sim \mathsf{NID}(0,1),\ t = 1,..,T, \tag{1·3}$$

$$\alpha_1 \sim \mathsf{NID}\left\{\mu, \sigma_\eta^2/(1-\phi^2)\right\}, \tag{1·4}$$

where α_t represents the unobserved log instantaneous volatility or state and $\Psi = \{\mu, \phi, \sigma_\eta^2\}$ represents the parameters. This model has attracted much recent attention in the econometrics literature; see, for instance, Hull & White (1987), Harvey et al. (1994) and Jacquier et al. (1994).

1·3 *Block sampling*

This procedure has traditionally been split into two separate sampling problems. We draw the parameters from their distribution conditional upon the states $f(\Psi|y,\alpha)$ and then the states conditional upon the new parameters $f(\alpha|y,\Psi)$. The problem of drawing from $f(\alpha|y,\Psi)$ is by no means trivial since, for state space models, α is typically of high dimension. In addition, the correlations between elements α_t and α_{t+k} arising from $f(\alpha|y,\Psi)$ may be very high, particularly for persistent models. This high correlation causes single move methods (which move one state at a time), such as Jacquier et al. (1994), to converge slowly. Blocking components in this case provides a much more effective strategy and this approach is exploited for special cases of non-Gaussianity in Shephard (1994) and Carter & Kohn (1994). A more general method for sampling blocks of states is given in Shephard & Pitt (1997), and this method will be explored in more detail in Section 2 of this paper.

The problem of sampling the parameters from their conditional posterior distribution $f(\Psi|y,\alpha)$ is less problematic. The dimension of Ψ is low in comparison to α and if we break Ψ into its various components, sampling each conditional upon the others, we often obtain conditional distributions which are easy to sample from.

1·4 *Outline of this paper*

The disadvantage of the separate sampling scheme outlined above is that the posterior for the parameters can be highly correlated with the states as well as with the other parameters. This presents potential problems as it indicates that the MCMC chain may mix slowly leading to an inefficient sampler and difficult statistical diagnostics. In this paper we investigate various strategies for improving the performance of the algorithms.

In Section 3 we exploit antithetic variables which induce negative correlation between successive sweeps of the MCMC chain. We closely follow the stochastic relaxation approach of Green & Han (1990) and Barone & Frigessi (1989). Section 4 presents the joint sampling method, directly drawing from $f(\alpha, \Psi|y)$. Section 5 concludes, while the conclusion outlines some algorithm used in this paper.

2 Separate sampling

The states, conditional upon the parameters, are drawn from their target density using the procedure suggested by Shephard & Pitt (1997). Firstly a small number of states $\alpha_{\backslash}$, widely spaced, are randomly picked to retain their values from the previous MCMC sweep. These states are called "knots". We now update each block of states between two knots using the MCMC method. Suppose that $\alpha_{t,k} = \{\alpha_t, ..., \alpha_{t+k}\}$ represents a block of states between two knots, α_{t-1} and α_{t+k+1}. We then have to sample the states $\alpha_{t,k}$ from the target density $f(\alpha_{t,k}|\Psi, y, \alpha_{t-1}, \alpha_{t+k+1})$. We note that

$$\begin{aligned} f(\alpha_{t,k}|y, \Psi, \alpha_{t-1}, \alpha_{t+k+1}) &\propto \exp\{l(y|\alpha_{t,k})\} f(\alpha_{t,k}|\Psi, \alpha_{t-1}, \alpha_{t+k+1}) \\ &\simeq \exp\{\tilde{l}(y|\alpha_{t,k})\} f(\alpha_{t,k}|\Psi, \alpha_{t-1}, \alpha_{t+k+1}) \\ &\propto \tilde{f}(\alpha_{t,k}|y, \Psi, \alpha_{t-1}, \alpha_{t+k+1}). \end{aligned}$$

where $\tilde{l}(y|\alpha_{t,k})$ represents a second order Taylor expansion of $l(y|\alpha_{t,k})$ in $\alpha_{t,k}$ around the vector $\widehat{\alpha}_{t,k}$. We generally choose $\widehat{\alpha}_{t,k}$ to be the mode of $f(\alpha_{t,k}|y, \Psi, \alpha_{t-1}, \alpha_{t+k+1})$. By construction the approximation $\tilde{f}(\alpha_{t,k}|y, \Psi, \alpha_{t-1}, \alpha_{t+k+1})$ is of Gaussian form since $\tilde{l}(y|\alpha_{t,k})$ is quadratic and $f(\alpha_{t,k}|\Psi, \alpha_{t-1}, \alpha_{t+k+1})$ is Gaussian. This approximate model for $\alpha_{t,k}$ is now in Gaussian state space form and so $\alpha_{t,k}$ (or the corresponding disturbances of the state space) can be sampled using the simulation smoother of de Jong & Shephard (1995) .

If we use only a Metropolis algorithm for deciding whether to accept the new proposal $\alpha_{t,k}^n$ or retain the old values $\alpha_{t,k}^o$ (in Shephard & Pitt (1997) this is combined with an A-R step), we have

$$\Pr(\alpha_{t,k}^o - > \alpha_{t,k}^n) = \min\left\{1, \frac{w(\alpha_{t,k}^n)}{w(\alpha_{t,k}^o)}\right\},$$

where $w(\alpha_{t,k}) = \exp\{l(y|\alpha_{t,k}) - \tilde{l}(y|\alpha_{t,k})\} = \exp\left[\sum_{i=t}^{t+k}\left\{l(\alpha_i) - \tilde{l}(\alpha_i)\right\}\right]$.

Typically $w(\alpha_{t,k})$ is close to 1, yielding a high acceptance rate in the Metropolis algorithm. We proceed in this fashion, sampling all the blocks between the fixed knots. This defines a complete MCMC sweep through the states. The parameters are then sampled from their

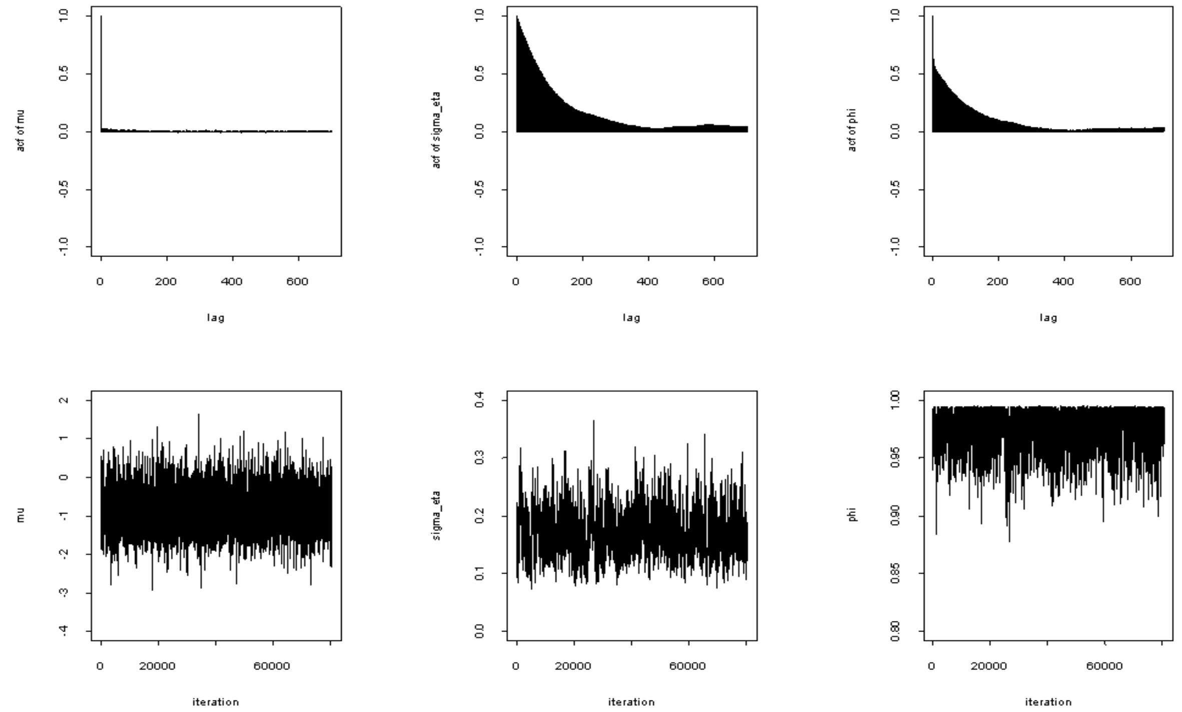

Figure 1: *Correlograms for 80,500 samples of* μ, σ_η^2 *and* ϕ *going from left to right, using the seperate sampling approach with 10 knots. Plots of the samples are shown below the correlograms*

conditional density $f(\Psi|\alpha)$. For the next sweep a fixed number of knots is again randomly chosen from the states and so the process continues.

The parameters Ψ are drawn directly from $f(\Psi|\alpha)$. Indeed they are conditioned on the other parameters so for the SV model we draw from the densities $[\sigma_\eta|\phi,\alpha,\mu]$, $[\mu|\phi,\alpha,\sigma_\eta]$ and $[\phi|\sigma_\eta,\alpha,\mu]$. The conditional distributions of these parameters are described in more detail in Shephard & Pitt (1997).

To illustrate these methods we analyse the exchange rate data discussed in the introduction. Figure 1 shows $M = 80,500$ samples from the separate sampler using 10 knots (an efficient choice) together with their correlograms. The accept-reject method together with the Metropolis step is used for this application. It is clear that at lags of 500 there is some degree of autocorrelation in the parameters ϕ and σ_η. This suggests that these two parameters are highly correlated with the states and with each other.

The summary statistics are given in Table 1. The reported inefficiency factor estimates the variance of the sample mean from the MCMC sampling scheme relative to a hyperthetical sampler which draws independent random variables from the posterior. The ratio is estimated using a Parzen window (see, for example, Priestley (1981, Ch. 6)) with

$$\widehat{R}_M = 1 + \frac{2M}{M-1}\sum_{i=1}^{B_M} K\left(\frac{i}{B_M}\right)\widehat{\rho}(i),$$

Parameter	Mean	Variance	Inefficiency factor
μ	-0.90523	0.077368	3.69 (300)
σ_η	0.16907	0.0014429	191 (800)
ϕ	0.97425	0.00017069	112 (800)
CPU time (s)	4,363		

Table 1: *Summaries of Figure 1. Results for the separate block sampling scheme with* 10 *knots. Time is in seconds on a pentium 133 to carry out 80,500 iterations. Figures in brackets denote the bandwidth for Parzen window.*

where B_M represents bandwidth, while the Parzen kernel is

$$\begin{aligned} K(x) &= 1-6x^2+6x^3, & x\in[0,\tfrac{1}{2}], \\ &= 2(1-x)^3, & x\in[\tfrac{1}{2},1], \\ &= 0, & \text{elsewhere.} \end{aligned}$$

Here $\widehat{\rho}(i)$ is an estimate of the autocorrelation at lag i of the MCMC sampler.

3 Antithetic variables

In this section the approach of Barone & Frigessi (1989), which they call the $\omega-$stochastic relaxation approach, is considered. Barone & Frigessi (1989) use this method in image analysis to reduce autocorrelation in their Metropolis chain when the conditional target distribution is Gaussian. Part of the appeal of this method is that the Metropolis proposals are accepted with probability 1. Green & Han (1990) modify the method for use in a Poisson spatial statistics model. For this single-move method the conditional densities are non-Gaussian although the proposal densities are Gaussian. The proposal densities are constructed in the following manner. Suppose we define the following transition density with x denoting the proposed new state and x' the current state, with $-1<\theta<1$

$$q(x|x') = \mathsf{N}\left\{(1+\theta)\mu-\theta x', (1-\theta^2)\Sigma\right\},$$

To ease notation the conditioning variables have been dropped. Then,

$$\begin{aligned} \log\left\{\frac{q(x'|x)}{q(x|x')}\right\} &= -\frac{1}{2(1-\theta^2)}\left[\begin{array}{c} \{x'-\mu-\theta(\mu-x)\}^T\Sigma^{-1}\{x'-\mu-\theta(\mu-x)\} \\ -\{x-\mu-\theta(\mu-x')\}^T\Sigma^{-1}\{x-\mu-\theta(\mu-x')\} \end{array}\right] \\ &= -\frac{1}{2}\left[(x'-\mu)^T\Sigma^{-1}(x'-\mu)-(x-\mu)^T\Sigma^{-1}(x-\mu)\right] = \log\left\{\frac{q(x')}{q(x)}\right\}, \end{aligned}$$

where $q(x) = \mathsf{N}(\mu,\Sigma)$, the independent proposal density corresponding to $\theta=0$. This expression is independent of θ. Hence the Metropolis expression is the same as the independent sampling ($\theta=0$) case. However, it should be noted that the Metropolis acceptance rate will not be the same as that for the independent sampling case since the proposal values are sampled differently.

The method may be applied to the states, α, of our model since the proposal for the states is Gaussian. If we denote the current states by α' and the proposed states by α, suppressing the conditioning parameters and knots, then the target density is given by $p(\alpha)$ where

$$\log p(\alpha) = c + l(\alpha) + \pi(\alpha).$$

The log-prior $\pi(\alpha)$, the state density, is quadratic in α and, as before $l(\alpha)$ represents the non-quadratic, but concave, log-likelihood (the measurement density). The unnormalised proposal density is given by $q^*(\alpha)$, where

$$\log q^*(\alpha) = c + \tilde{l}(\alpha) + \pi(\alpha),$$

where, as before, $\tilde{l}(\alpha)$ represents the second order expansion of the log-likelihood. Hence we draw $\alpha \sim \mathsf{N}(\mu, \Sigma)$, using the simulation smoother, then the log of the Metropolis ratio is

$$\log \left\{ \frac{p(\alpha)q(\alpha')}{p(\alpha')q(\alpha)} \right\} = l(\alpha) - \tilde{l}(\alpha) - l(\alpha') + \tilde{l}(\alpha').$$

If we instead draw $\alpha \sim \mathsf{N}\left\{(1+\theta)\mu - \theta\alpha', (1-\theta^2)\Sigma\right\}$ then the construction of the Metropolis ratio remains the same since we have,

$$\log \left\{ \frac{q(\alpha|\alpha')}{q(\alpha'|\alpha)} \right\} = \log \left\{ \frac{q(\alpha)}{q(\alpha')} \right\}. \tag{3·1}$$

So again this expression does not depend on θ.

The fact that this expression still does not involve the density arising from the state equation, but only the likelihood arising from the measurement equation and its approximation, is computationally attractive.

We may also require to place the state sampler within an accept-reject method (A-R), see Tierney (1994), before performing the calculation of the Metropolis ratio. The reason for this being that the Metropolis rejection probability is reduced at the expense of some computer time. The equality of (3·1), arising from the results of Barone & Frigessi (1989), implies that

$$\tilde{l}(\beta) + \pi(\beta) - \tilde{l}(\beta') - \pi(\beta') = \tilde{l}(\alpha) + \pi(\alpha) - \tilde{l}(\alpha') - \pi(\alpha') \tag{3·2}$$

where, regarding them as functions of α we define

$$\beta(\alpha) = \frac{\{\alpha + \theta\alpha' - (1+\theta)\mu\}}{\sqrt{1-\theta^2}} + \mu \qquad \text{and} \qquad \beta'(\alpha') = \frac{\{\alpha' + \theta\alpha - (1+\theta)\mu\}}{\sqrt{1-\theta^2}} + \mu,$$

If we perform A-R sampling by sampling $\beta \sim \mathsf{N}(\mu, \Sigma)$ until acceptance by inspecting the ratio, the log of which is

$$\log r(\beta) = l(\beta) - \tilde{l}(\beta),$$

then the density of the resulting β is

$$f(\beta) \propto \min\left\{p(\beta), q^*(\beta)\right\}.$$

If we then transform the β value by setting $\alpha = \sqrt{1-\theta^2}(\beta-\mu) - \theta\alpha' + (1+\theta)\mu$ then the distribution of α is (since the Jacobian of the transformation is constant),

$$f(\alpha) \propto \min\left[p\left\{\beta(\alpha)\right\}, q^*\left\{\beta(\alpha)\right\}\right],$$

where β is defined in terms of α as above.

The log of the contribution to the Metropolis ratio of (4·2) now becomes,

$$\begin{aligned} \log\left\{\tfrac{p(\alpha)f(\alpha')}{p(\alpha')f(\alpha)}\right\} &= l(\alpha) + \pi(\alpha) + \min\{l(\beta') + \pi(\beta'), \tilde{l}(\beta') + \pi(\beta')\} \\ &+ \; -l(\alpha') - \pi(\alpha') - \min\{l(\beta) + \pi(\beta), \tilde{l}(\beta) + \pi(\beta)\}. \end{aligned}$$

Noting (3·2) we obtain,

$$\begin{aligned} \log\left\{\frac{p(\alpha)f(\alpha')}{p(\alpha')f(\alpha)}\right\} &= l(\alpha) - l(\alpha') + \tilde{l}(\alpha') - \tilde{l}(\alpha) \\ &\quad + \min\{l(\beta') - \tilde{l}(\beta'), 0\} - \min\{l(\beta) - \tilde{l}(\beta), 0\}. \end{aligned}$$

The Metropolis step is therefore simple to calculate and still only involves the likelihood of the states and their second order approximation. The justification for using the A-R method in the context of non-independent proposals is less compelling than its use for independent proposals since, even if we have coverage, we do not accept at the Metropolis stage with probability 1. However, in practice, we have found that for this form of proposal the Metropolis acceptance probability was increased by the addition of the A-R stage.

The advantage of this antithetic variable approach is that it is possible to induce negative correlation in the states conditional upon the previous draw. This is achieved by choosing $0 < \theta < 1$. However, it is apparent that the conditional variance of the proposal, given by $(1-\theta^2)\Sigma$, is reduced as θ becomes closer to 1. This is not an attractive feature since as $\theta \to 1$ we get no movement in the resulting MCMC chain.

Figure 2 and Table 2 presents the results of the separate sampling scheme using stochastic relaxation on the states ($\theta = 0.65$). In this case the Metropolis method with A/R was implemented. Clearly, the efficiency is less than that for the standard separate sampling scheme. This may be due to the fact that the relaxation method leads to higher rates of Metropolis rejection.

4 Joint sampling

The proposed joint algorithm aims at sampling all the states and parameters given some known randomly selected knots $\alpha_\backslash$. We shall denote the states to be sampled by $\underline{\alpha}$ and the whole state space by $\alpha = \{\underline{\alpha}, \alpha_\backslash\}$. The joint distribution of the parameters and the states, given the observations and the knots, can be written as $f(\underline{\alpha}, \Psi | y, \alpha_\backslash)$. Now

$$f(\underline{\alpha}, \Psi | y, \alpha_\backslash) \quad \propto \quad f(y|\underline{\alpha}) f(\underline{\alpha}|\Psi, \alpha_\backslash) f(\Psi) = \exp\{l(y|\underline{\alpha})\} f(\underline{\alpha}|\Psi, \alpha_\backslash) f(\Psi)$$

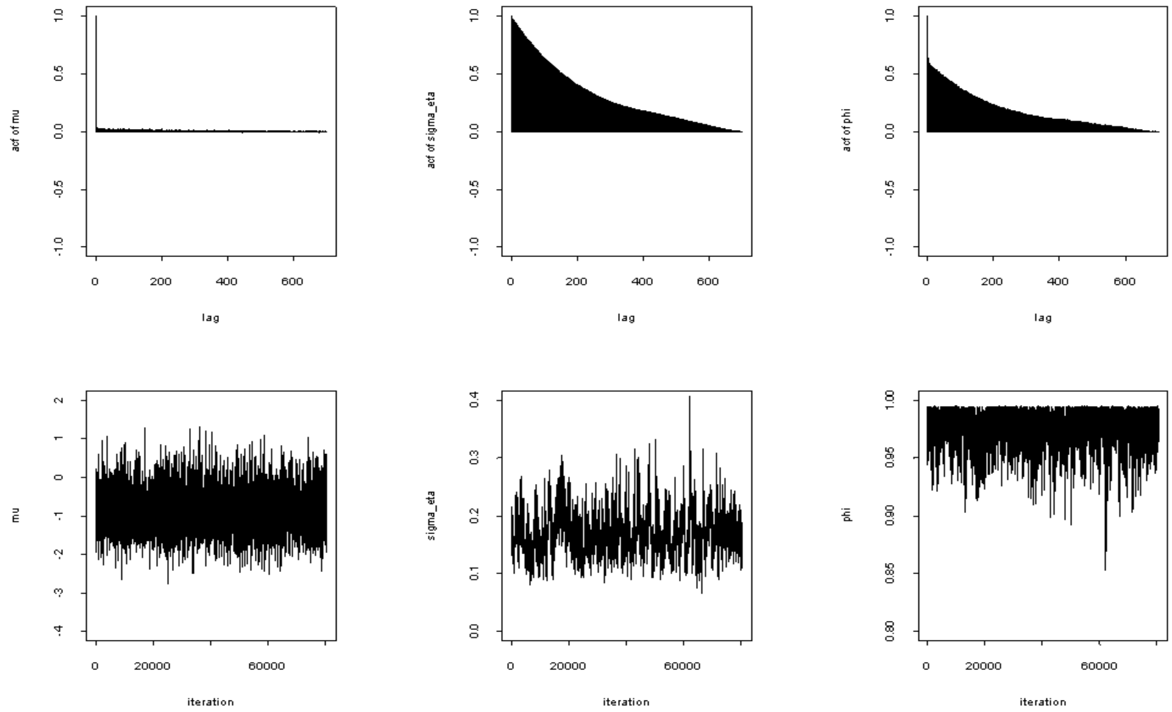

Figure 2: *Correlograms for 80,500 samples of μ, σ_η^2 and ϕ going from left to right, using the seperate sampling approach with 10 knots. The relaxation method ($\theta = 0.65$) is applied to the states. Plots of the samples are shown below the correlograms*

$$\begin{aligned} &\simeq \exp\{\tilde{l}(y|\underline{\alpha})\} f(\underline{\alpha}|\Psi, \alpha_{\backslash}) f(\Psi) \\ &= \tilde{f}(y|\underline{\alpha}) f(\underline{\alpha}|\Psi, \alpha_{\backslash}) f(\Psi), \end{aligned} \tag{4·1}$$

where again $\tilde{l}(y|\alpha)$ represents the second order expansion of $l(y|\alpha)$ around a suitable vector $\widehat{\alpha}$, the choice of which will be outlined presently. Now we have to sample from the approximate joint density of (4·1) to provide proposals for the Metropolis method. Again it can be seen, as in Shephard & Pitt (1997), that (4·1) represents a Gaussian state space form (GSSF) in $\underline{\alpha}$. Indeed $\tilde{f}(y|\underline{\alpha}) \propto f(z|\underline{\alpha})$ where z represents a vector of pseudo-observations arising from a GSSF. We can now see how to sample from (4·1) since,

$$\tilde{f}(y|\underline{\alpha}) f(\underline{\alpha}|\Psi, \alpha_{\backslash}) f(\Psi) \propto f(z|\underline{\alpha}) f(\underline{\alpha}|\Psi, \alpha_{\backslash}) f(\Psi) \propto f(\underline{\alpha}|\Psi, y, \alpha_{\backslash}) f(\Psi|z, \alpha_{\backslash}).$$

Parameter	Mean	Variance	Inefficiency factor
μ	-0.9051572	0.07813187	4.780 (300)
σ_η	0.1680184	0.0017689	302 (800)
ϕ	0.9745084	0.0001751277	179 (800)
CPU time (s)	7,981		

Table 2: *Summaries of Figure 2. Results for the separate block sampling scheme with* 10 *knots and using relaxation on the states. Time is in seconds on a pentium 133 to carry out 80,500 iterations. Figures in brackets denote the bandwidth for Parzen window.*

Here $f(\underline{\alpha}|\Psi, y, \alpha_\backslash)$ represents the smoothing density for $\underline{\alpha}$, which can be simulated by using the method of de Jong & Shephard (1995). The sampling of $f(\Psi|z, \alpha_\backslash)$ is less straightforward. The log of $f(\Psi|z, \alpha_\backslash)$ is given by the Kalman filter (together with the parameter prior). If it were possible to sample directly from this parameter density then the overall Metropolis switching probability would be

$$\Pr(\{\underline{\alpha}^o, \Psi^o\} -> \{\underline{\alpha}^n, \Psi^n\}) = \min\left\{1, \frac{w(\underline{\alpha}^n)}{w(\underline{\alpha}^o)}\right\}.$$

This is exactly the same as the construction of the Metropolis step were we only sampling the states. In practice, sampling the parameters from $f(\Psi|z, \alpha_\backslash)$, is not possible exactly. Instead we generate proposal samples for $f(\Psi|z, \alpha_\backslash)$ from an approximation to this density $\tilde{f}(\Psi|z, \alpha_\backslash)$. The approximation is constructed as follows. We use the BFGS method for obtaining the mode of $f(\Psi|z, \alpha_\backslash)$, or a point near the mode, $\widehat{\Psi}$. The BFGS method is described in detail in Fletcher (1987, p55–57). The method is a quasi-Newton optimisation algorithm. Let us express $f(\Psi|z, \alpha_\backslash) = \exp\{l(\Psi|z, \alpha_\backslash)\}$. Our approximation to this density (up to a constant of proportionality) is $\tilde{f}(\Psi|z, \alpha_\backslash) = \exp\{\tilde{l}(\Psi|z, \alpha_\backslash)\}$ where,

$$\tilde{l}(\Psi|z, \alpha_\backslash) = -\frac{1}{2}(\Psi - \widehat{\Psi})' H^{-1} (\Psi - \widehat{\Psi}).$$

H is the Hessian matrix which is output by the BFGS method. It is guaranteed to be positive definite and approximates the negative of the inverse matrix of second derivatives of $l(\Psi|z)$ evaluated at $\widehat{\Psi}$. Hence we have the following Metropolis updating step

$$\Pr\{(\alpha^o, \Psi^o) -> (\alpha^n, \Psi^n)\} = \min\left\{1, \frac{w(\alpha^n)}{w(\alpha^o)}\frac{\nu(\Psi^n)}{\nu(\Psi^o)}\right\} \tag{4·2}$$

where

$$\nu(\Psi) = \frac{f(\Psi|z, \alpha_\backslash)}{\tilde{f}(\Psi|z, \alpha_\backslash)} = \exp\{l(\Psi|z, \alpha_\backslash) - \tilde{l}(\Psi|z, \alpha_\backslash)\}.$$

It should be noted that again A/R methods for α can be used in conjunction with this Metropolis algorithm although the details are similar to those given Shephard & Pitt (1997) and will not, therefore, be given in detail here. As in the separate sampling scheme it is hoped that the approximations lead to a high Metropolis acceptance probability.

4·1 *Finding points of expansion and starting values*

We know that $l(\Psi|z, \alpha_\backslash)$, the true log-density, consists of the log-likelihood arising from the KF together with a weak prior. The dominating KF log-likelihood is not, in general, concave making an exact Newton method, for instance, unstable. The BFGS method is computationally efficient because it avoids explicitly calculating the matrix of second derivatives, an expensive task when each call of the log-density involves the calculation of the KF (over a potentially large number of time points). However, the performance of the BFGS method relies on the method being initialised at sensible starting values and with a reasonable initial Hessian, H. One way to do this is to simply run the separate MCMC scheme for a fixed short run-in period, recording the means and the variance matrix of the parameters. The proposed joint sampling scheme would then be started at this stage with the starting values for $\widehat{\Psi}$ in the BFGS method always being set to these means and the corresponding Hessian always being set to the sampled variance matrix. Of course, the samples from the run-in period would need to be discarded when making inference from the MCMC chain.

In fact in the applications presented in this paper a different strategy has been employed for obtaining good starting values of $\widehat{\Psi}$ and $\widehat{H}$. Before the MCMC sampling procedure begins we find an approximation to the overall mode of the posterior distribution of Ψ.

The following scheme is implemented:

1. An initial guess Ψ^0 for the parameters is obtained.

2. The mode, $\widehat{\alpha}$, of $f(\alpha|y, \Psi^0)$ is found. This can be done by using the expectation smoother and is detailed in Fahrmeir (1992) and Shephard & Pitt (1997).

3. The mode $\widehat{\alpha}$ is used as an expansion point to yield a GSSF.

4. The log-density of Ψ in the GSSF is maximised, starting at Ψ^0, with respect to Ψ. We then set $\Psi^1 = \Psi$.

This process is continued until we have convergence. We then obtain a rough estimate of the posterior mode $\widehat{\Psi}$. The Hessian, H, is set to the negative of the inverse of the matrix of second derivatives of the log-density of Ψ evaluated at $\widehat{\Psi}$. The important aspect of this method is that it only needs to be run once, before the MCMC simulation, in order to yield starting values for the subsequent BFGS optimisation schemes.

So far we have not addressed the issue of how to choose the expansion points $\widehat{\alpha}$ which are used to obtain the GSSF as the proposal density. The algorithm describing one entire sweep of the MCMC procedure is given below where the states and parameters from the previous sweep are denoted by α^o and Ψ^o (we have obtained $\widehat{\Psi}$ and $\widehat{H}$ as above):

- Randomly allocate a fixed number of the current states to retain their previous sampled values, $\alpha_\setminus$.

- Find the mode, $\underline{\widehat{\alpha}}$, of $f(\underline{\alpha}|\widehat{\Psi}, y, \alpha_\setminus)$ and perform a second order expansion of $f(y|\underline{\alpha}, \widehat{\Psi})$ around $\underline{\widehat{\alpha}}$.

- Construct the resulting GSSF, with pseudo-observations z and find the mode,$\widetilde{\Psi}$, and Hessian, H, of Ψ from $f(\Psi|z, \alpha_\setminus)$, using the BFGS method intialised with $\widehat{\Psi}$ and $\widehat{H}$.

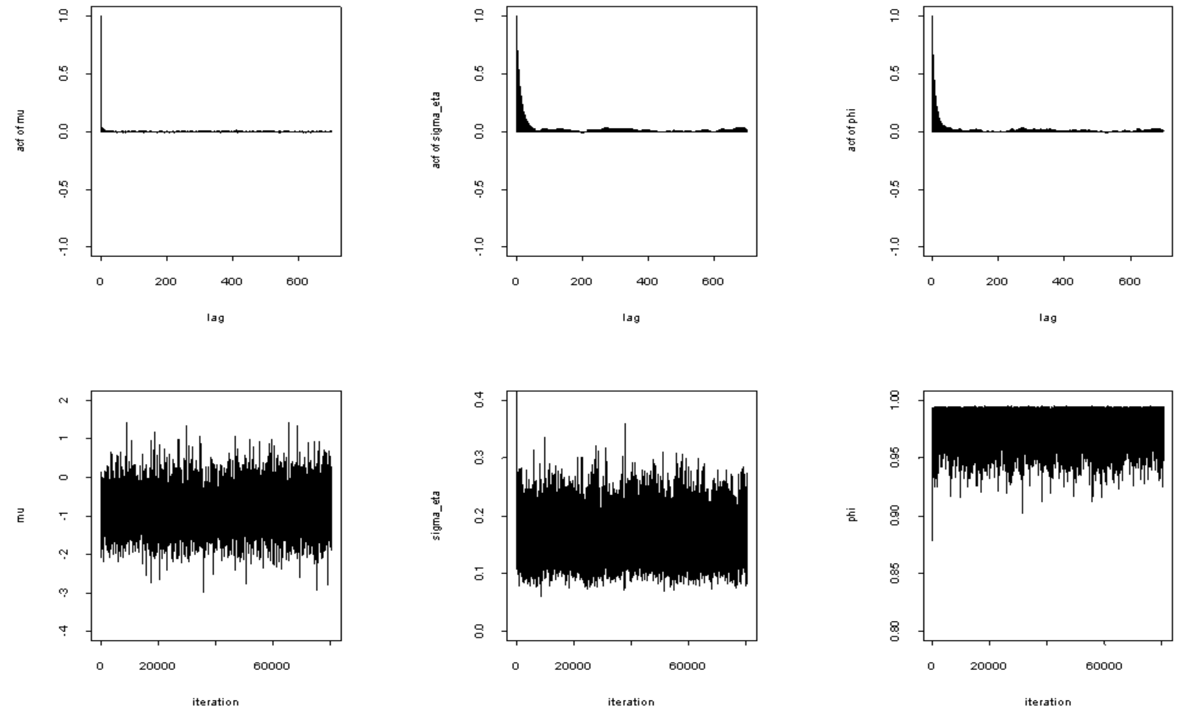

Figure 3: *Correlograms for 80,500 samples of* μ, σ_η^2 *and* ϕ *going from left to right, using the joint sampling approach with 45 knots. Plots of the samples are shown below the correlograms*

- Sample $\Psi^n \sim \mathsf{N}(\widetilde{\Psi}, H)$.
- Sample $\underline{\alpha}^n \sim f(\underline{\alpha}|\widehat{\Psi}, z, \alpha_\backslash)$ using the simulation smoother.
- Change to the new values with probability $\Pr\{(\underline{\alpha}^o, \Psi^o) \to (\underline{\alpha}^n, \Psi^n)\} = \min\left\{1, \frac{w(\underline{\alpha}^n)}{w(\underline{\alpha}^o)} \frac{\nu(\Psi^n)}{\nu(\Psi^o)}\right\}$.

The number of knots has little impact upon the discrepancy between $f(\Psi|z, \alpha_\backslash)$ and its approximation. However, as the number of knots becomes larger the discrepancy between the density of the states and its approximation is reduced and so $w(\underline{\alpha}^n)/w(\underline{\alpha}^o)$ becomes less variable around 1, increasing the Metropolis switching probability. However, the price of this is increased dependency in the sampled chain since we are retaining more states drawn from the previous sweep. In practice, the best choice for the number of knots involves a trade off between these two considerations. In the applications which follow the number of knots is periodically increased, every few iterations, for one sweep to ensure that the MCMC chain moves.

The results of the joint sampling approach with 45 knots are shown in Figure 3 and Table 3. The parameter μ is sampled separately, not jointly, since it is fairly uncorrelated with the states and the other parameters. There is clearly a substantial improvement in efficiency resulting from this approach. The autocorrelations of the parameters die out at about lag 50.

The results of the joint sampling approach when stochastic relaxation ($\theta = 0.7$) is used

Parameter	Mean	Variance	Inefficiency factor
μ	-0.90508	0.077538	1.60 (30)
σ_η	0.16795	0.0012925	22.9 (200)
ϕ	0.97476	0.00015346	17.4 (200)
CPU time (s)	$6,857$		

Table 3: *Summaries of Figure 3. Results for the joint sampling scheme with* 45 *knots. Time is in seconds on a pentium 133 to carry out 80,500 iterations. Figures in brackets denote the bandwidth for Parzen window.*

for μ, are given in Figure 4 and Table 4. There is a slight increase in efficiency resulting from the relaxation approach.

Parameter	Mean	Variance	Inefficiency factor
μ	-0.90561	0.07829744	0.698 (30)
σ_η	0.16735	0.001331371	18.5 (200)
ϕ	0.9747017	0.000161953	14.2 (200)
CPU time (s)	$6,913$		

Table 4: *Summaries of Figure 4. Results for the joint sampling scheme with* 45 *knots and using relaxation on* μ. *Time is in seconds on a pentium 133 to carry out 80,500 iterations. Figures in brackets denote the bandwidth for Parzen window.*

5 Conclusions

The methods outlined in this paper represent a general MCMC approach for non-Gaussian measurement time series. The approach is flexible, allowing the practitioner choice in which parameters to update jointly with the states and which to update separately conditional upon the states. Those parameters which are most highly correlated with the states, and with each other, are the most suitable for the joint updating strategy. In the application presented, the method enables gains in efficiency of around 10 fold as it reduces the autocorrelation associated with the parameters of interest.

The use of the antithetic variables approach introduced by Green & Han (1990) and Barone & Frigessi (1989) for Metropolis sampling requires careful consideration of the antithetic parameter θ. For values of θ close to 1, convergence can be slow. In general, even with judicious choice of θ, the efficiency gains resulting from the use of antithetic variables was slight. This is in part due to the fact that the Metropolis rejection probability increases as θ increases above 0. In addition the conditional variance of the proposal samples is reduced which may cause slow mixing.

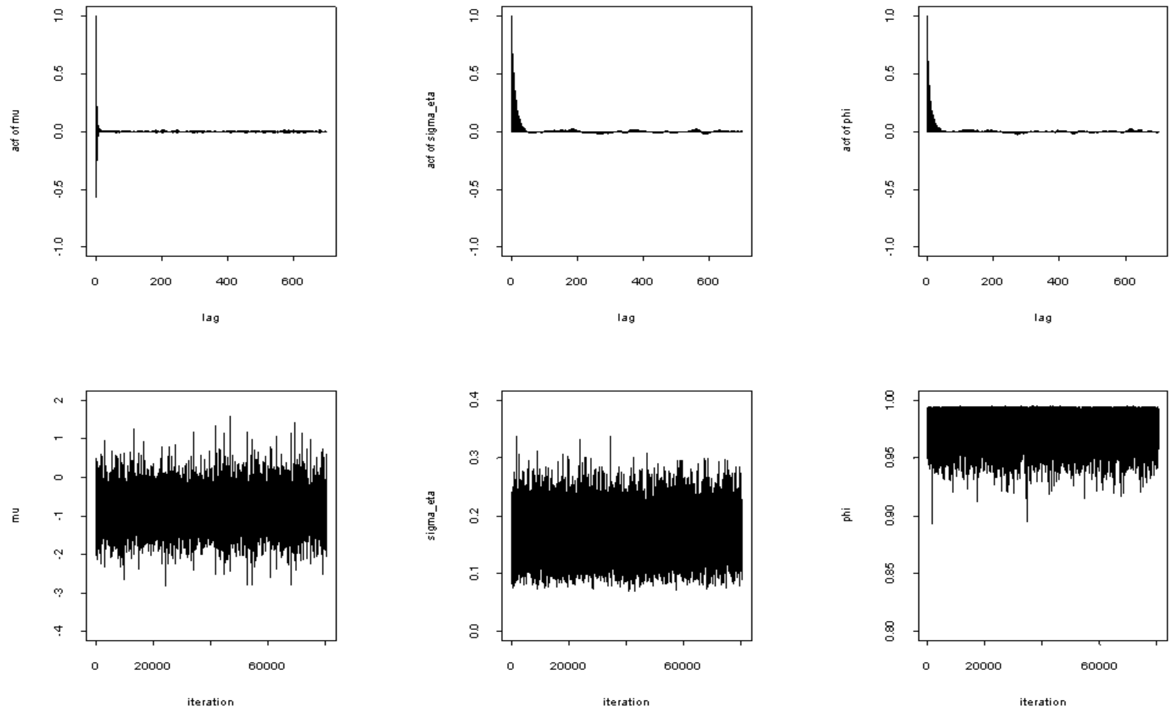

Figure 4: *Correlograms for 80,500 samples of μ, σ_η^2 and ϕ going from left to right, using the joint sampling approach with 45 knots. Relaxation ($\mu = 0.7$) used. Plots of the samples are shown below the correlograms*

In general, it appears that consideration of blocking strategies which exploit the dependency structure of the model of interest as fully as possible is more important than attempting to induce negative correlation between successive sweeps of the MCMC chain. In very highly persistent models, which may arise from intra-daily observations using the SV model, increased efficiency is particularly important.

6 Appendix

This Appendix details Gaussian filtering and smoothing. The Gaussian state space puts

$$\begin{aligned} y_t &= c_t + Z_t\alpha_t + G_t u_t, \quad u_t \sim NID(0, I), \\ \alpha_{t+1} &= d_t + T_t\alpha_t + H_t u_t, \\ \alpha_1 | Y_0 &\sim N(a_{1|0}, P_{1|0}), \end{aligned}$$

We assume that $G_t' H_t = 0$ and write $G_t G_t' = \Sigma_t$ which we assume is full rank. The Kalman filter (de Jong (1989)) computes $a_{t|t-1} = E\alpha_t | Y_{t-1}$ and $P_{t|t-1} = MSE(\alpha_t | Y_{t-1})$,

$$\begin{aligned} a_{t+1|t} &= d_t + T_t a_{t|t-1} + K_t v_t, & P_{t+1|t} &= T_t P_{t|t-1} L_t' + H_t H_t', & v_t &= y_t - Z_t a_{t|t-1} - c_t, \\ F_t &= Z_t P_{t|t-1} Z_t' + G_t G_t', & K_t &= T_t P_{t|t-1} Z_t' F_t^{-1}, & L_t &= T_t - K_t Z_t. \end{aligned}$$

The filter yields forecast errors v_t, MSEs F_t and the Gaussian likelihood

$$\log f(y_1, ..., y_n) = const - \frac{1}{2}\sum \log |F_t| - \frac{1}{2}\sum v_t' F_t^{-1} v_t.$$

The simulation smoother (de Jong & Shephard (1995)) draws from $(c_1 + Z_1\alpha_1, ..., c_n + Z_n\alpha_n)\,|y$. Setting $r_n = 0$ and $N_n = 0$, for $t = n, ..., 1$, and writing $D_t = F_t^{-1} + K_t' N_t K_t$, $n_t = F_t^{-1} v_t - K_t' r_t$

$$\begin{array}{rcl} C_t & = & \Sigma_t - \Sigma_t N_t \Sigma_t, \qquad \kappa_t \sim \mathsf{N}(0, C_t), \\ r_{t-1} & = & Z_t' F_t^{-1} v_t + L_t' r_t - V_t' C_t^{-1} \kappa_t, \qquad V_t = \Sigma_t \left(D_t Z_t - K_t' N_t T_t \right), \\ N_{t-1} & = & Z_t' F_t^{-1} Z_t + L_t' N_t L_t + V_t' C_t^{-1} V_t. \end{array} \tag{6·1}$$

and recording $\eta_t = \Sigma_t n_t + \kappa_t$. Then $y_t - \eta_t$ is a draw from the signal $(c_t + Z_t \alpha_t)\,|y$.

The de Jong (1989) and Koopman (1993) moment smoother is a special case of the simulation smoother as it computes the means of $(c_1 + Z_1\alpha_1, ..., c_n + Z_n\alpha_n)\,|y$. Setting $r_n = 0$ for $t = n, ..., 1$, and writing $n_t = F_t^{-1} v_t - K_t' r_t$

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \tag{6·2}$$

and recording $\eta_t = \Sigma_t n_t$. Then $y_t - \eta_t$ is the mean of the signal $(c_t + Z_t \alpha_t)\,|y$.

References

Barone, P. & Frigessi, A. (1989). Improving stochastic relaxation for Gaussian random fields. *Probability in the Engineering and Informational Sciences* **4**, 369–389.

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. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541–553.

Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistican* **49**, 327–335.

de Jong, P. (1989). Smoothing and interpolation with the state space model. *J. Am. Statist. Assoc.* **84**, 1085–1088.

de Jong, P. & Shephard, N. (1995). The simulation smoother for time series models. *Biometrika* **82**, 339–50.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalised linear models. *J. Am. Statist. Assoc.* **87**, 501–509.

Fletcher, R. (1987). *Practical Methods of Optimisation.* John Wiley, New York, 2 edition.

Fruhwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *J. Time Series Analysis* **15**, 183–202.

Gelfand, A. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* **85**, 398–409.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions, PAMI* **6**, 721–741.

Gilks, W. K., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice.* Chapman & Hall, London.

Green, P. J. & Han, X. L. (1990). Metropolis methods, Gaussian proposals and antithetic variables. In Barone, P., Frigessi, A., & Piccioni, M., editors, *Stochastic Models, Statistical Methods and Algorithms in Image Analysis*, volume 74, pages 142–164. Springer-Verlag, Berlin.

Hannan, E. J. & Deistler, M. (1988). *The Statistical Theory of Linear Systems.* John Wiley, New York.

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

Harvey, A. C., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *Rev. Economic Studies* **61**, 247–264.

Hull, J. & White, A. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* **42**, 281–300.

Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *J. Business and Economic Statist.* **12**, 371–417.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* **80**, 117–126.

Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press, London.

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

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika* **81**, 115–131.

Shephard, N. & Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**. Forthcoming.

Taylor, S. J. (1986). *Modelling Financial Time Series.* John Wiley, Chichester.

Tierney, L. (1994). Markov Chains for exploring posterior distributions (with discussion). *Ann. Statist.* **21**, 1701–1762.