LIKELIHOOD ANALYSIS OF NON-GAUSSIAN MEASUREMENT TIME SERIES

NEIL SHEPHARD
Nuffield College, Oxford, OX1 1NF, UK
AND MICHAEL K PITT
Department of Statistics, University of Oxford, OX1 3TG and
Nuffield College, Oxford, OX1 1NF, UK

July 23, 1996

Abstract

In this paper we provide methods for estimating non-Gaussian time series models. These methods rely on Markov chain Monte Carlo to carry out simulation smoothing and Bayesian posterior analysis of parameters, and on importance sampling to estimate the likelihood function for classical inference. The time series structure of the models is used to ensure that our methods are efficient.

Some key words: Blocking; Exponential family; Importance sampling; Markov chain Monte Carlo; Simulation smoother; Stochastic volatility.

1 INTRODUCTION

1·1 *The model and literature*

In this paper we provide a likelihood analysis of an extension of the usual Gaussian state space form (Harvey, 1989, Ch. 3). The Kalman filter and simulation smoother are used to perform efficient signal extraction and parameter estimation.

The univariate observations y_t are assumed to be distributed, conditionally on a univariate θ_t, according to $f(y_t|\theta_t), t = 1, ..., n$. Throughout $\log f$ is assumed to be twice continuously differentiable with respect to θ_t. An example of this is the exponential family, written as

$$\log f(y_t \mid \theta_t) = y_t\theta_t - b(\theta_t) + c(y_t), \tag{1·1}$$

where b and c are flexible but known functions. Here θ_t is the stochastic process

$$\begin{array}{rclcl} h(\theta_t) & = & d_t & = & z_t\alpha_t + x_t\beta \\ & & \alpha_{t+1} & = & W_t\beta + T_t\alpha_t + H_tu_t, \quad u_t \sim NID(0, I), \\ & & \alpha_1 & \sim & N(a_{1|0}, p_{1|0}). \end{array} \tag{1·2}$$

$h(.)$ is a known function which is continuously twice differentiable, while the α_t are called states and NID denotes normally and independently distributed. Typically T_t, z_t, and H_t are selection matrices indexed by a small number of unknown parameters denoted by ψ, while x_t and W_t are sparse regressors.

Only a small number of simple non-Gaussian models possess conjugate filtering recursions. They are studied in Smith (1979, 1981), Smith & Miller (1986), Harvey & Fernandes (1989) and Shephard (1994a). West et al. (1985) propose a dynamic generalized linear class of processes, which is made up of equations (1·1) and (1·2), which they analysed by using approximations to the filtering and smoothing equations. Fahrmeir (1992), Singh & Roberts (1992) and the 1992 unpublished paper by J. Durbin and S. J. Koopman develop algorithms which find the exact mode of the smoothing density, $f(\alpha \mid y)$, and perform approximate likelihood inference on ψ. Here $y = (y_1, ..., y_n)'$ and $\alpha = (\alpha_1', ..., \alpha_n')'$.

General state space models are discussed by Kitagawa (1987) who used numerical integration rules to approximate the required filtering and smoothing recursions. Unfortunately the integrals are of the dimension of α_t and so for many problems these methods are inaccurate and quite slow, although there has been work to overcome these problems (Fruhwirth-Schnatter, 1992). More recently importance sampling has been used to approximate the likelihood by Danielsson & Richard (1993), while Markov chain Monte Carlo (MCMC) methods have been used to perform smoothing and parameter estimation (Carlin et al., 1992a; Carlin et al., 1992b; Jacquier et al., 1994; Chan & Ledolter, 1995).

The merging of the Kalman filter with MCMC methods has been successful, exploiting the time series structure of the model to improve the efficiency of the simulations, where it has been possible to condition on some latent indicator functions or other component to make y_t into a Gaussian state space. Leading references include Carter & Kohn (1994, 1996) and Shephard (1994b). Unfortunately, it has so far been unclear how to adapt these methods to deal with fundamentally non-Gaussian models. We tackle this problem here.

1·2 *Methodology*

This paper presents a new methodology which is based on the estimation by simulation of particular features of the model. This has the advantage that the estimates can be made as accurate as needed and their degree of accuracy can be assessed statistically. These simulation methods replace, in the Gaussian case, the role of the Kalman filter and smoother which usually estimate the states α from the observations y. In a simple example, if it were possible to simulate from $\alpha \mid y, \psi$, then any interesting function of α can be straightforwardly estimated from the draws. Likewise, if a Bayesian paradigm is adopted for ψ, then ψ they can be estimated by recording appropriate summaries of the simulations from $(\alpha', \psi')' \mid y$. If a classical view is taken, then an approximate Gaussian simulation smoother can be used as an input into an importance sample estimator of the likelihood.

It is possible to simulate from $\alpha \mid y, \psi$, which is a highly multivariate non-Gaussian random variable, by employing MCMC methods (Gilks et al., 1996). We use the independence chain 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 states or parameters in the model, given the data. Proposals z are made to possibly replace the current x_i, keeping constant $x_{\backslash i}$ which denotes all the other elements of the x vector. The proposal density is proportional to $q(z, x_{\backslash i})$ while the true density is $f(x_i|x_{\backslash i})$. 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 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).

This scheme suggests taking x_i to stand for individual states or parameters (Carlin et al., 1992b). Then the MCMC algorithm updates a single variable at a time which has the advantage that the conditional densities are usually simple. The main disadvantage is the correlation between variables over successive sweeps of the Metropolis sample, a particularly severe problem in time series models (Carter & Kohn, 1994; Shephard, 1994b).

Multi-move, or block, samplers work by updating several variables at the same time. When carried out sensibly, block samplers tend to be very effective. Liu et al. (1994) produce results which show that "grouping random components in a Gibbs sampler with two or more components usually results in more efficient sampling schemes."

Block samplers raise a number of issues. How should the blocks be constructed so that they are easy to sample? Should the blocks be the same for each sweep (a deterministically updated MCMC) or should they be randomly selected (a random block MCMC)?

The simulations from $f(\alpha|y)$ are correlated. We can estimate $\mathsf{E}g(\alpha)$ by $M^{-1}\sum g(\alpha^k)$, where α^k is the $k-th$ simulated value after reaching equilibrium, while the variance of this estimator is computed using a Parzen window $K()$, see Priestley (1981, p. 443). Let $\widehat{J}_M$ denote the estimation of the $\sqrt{M}$ scaled variance of the sample mean, then

$$\widehat{J}_M = \widehat{\Gamma}(0) + \frac{2M}{M-1}\sum_{i=1}^{B_M} K\left(\frac{i}{B_M}\right)\widehat{\Gamma}(i), \text{ where } \widehat{\Gamma}(i) = \frac{1}{M}\sum_{k=i+1}^{M}(\alpha^k - \overline{\alpha})(\alpha^{k-i} - \overline{\alpha}),$$

and B_M, the bandwidth, is stated in all our calculations.

1·3 *Structure of the paper*

The structure of the paper is as follows. In § 2 we look at a simple example of the model and investigate the properties of simple single move MCMC algorithms for this problem. The core of our contribution comes in § 3, which discusses the design of our methods to maximise efficiency for these models. § 4 looks at importance sampling. § 5 concludes, while an Appendix details some key algorithms.

2 EXAMPLE: STOCHASTIC VOLATILITY

2·1 *The model*

The stochastic volatility (SV) model has attracted much recent attention as a way of generalising the Black-Scholes option pricing formula to allow volatility clustering in asset returns (Hull & White, 1987; Harvey et al., 1994; Jacquier et al., 1994). The basic SV model has

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

where ϵ_t and η_t are independent Gaussian processes with variances of 1 and σ_η^2 respectively. MCMC methods have been used on this model by, for instance, Jacquier et al. (1994).

In this paper we analyse the daily returns (difference of the log of the series) on the Pound Sterling/US Dollar exchange rate from 1/10/81 to 28/6/85. This dataset has been previously analysed using quasi-likelihood methods in Harvey et al. (1994).

2·2 *Pseudo-dominating Metropolis sampler*

We exploit a second order expansion of the log-density as the basis of our pseudo-dominating Metropolis sampler. The expansion is, writing $\alpha_t|\alpha_{t-1}, \alpha_{t+1} \sim \mathsf{N}(\mu_t, \sigma_t^2)$, suppressing dependence on ψ, of the form:

$$\begin{aligned} \log f(\alpha_t|y_t, \alpha_{\backslash t}) &= \log f(\alpha_t|\alpha_{t-1}, \alpha_{t+1}) + \log f(y_t|\alpha_t) \\ &= -\tfrac{(\alpha_t-\mu_t)^2}{2\sigma_t^2} - \tfrac{\alpha_t}{2} - \tfrac{y_t^2}{2\beta^2}\exp(-\alpha_t) \\ &\approx -\tfrac{(\alpha_t-\mu_t)^2}{2\sigma_t^2} - \tfrac{\alpha_t}{2} - \tfrac{y_t^2}{2\beta^2}\exp(-\mu_t)\left\{1 - (\alpha_t - \mu_t) + \tfrac{1}{2}(\alpha_t-\mu_t)^2\right\} \\ &= \log g. \end{aligned}$$

A similar quadratic expansion appears in Green & Han (1990) in their analysis of Poisson images with a Gaussian prior. The quadratic term in $\log g$ means that it does not bound $\log f$. This delivers a pseudo-dominating suggestion, with suggested draws z for $\alpha_t|y_t, \alpha_{\backslash t}$ being made from

$$\mathsf{N}\left[\frac{\sigma_t^{*2}}{\sigma_t^2}\mu_t + \frac{\sigma_t^{*2}}{2}\left\{\frac{y_t^2}{\beta^2}\exp(-\mu_t)(1+\mu_t) - 1\right\}, \sigma_t^{*2}\right], \quad \text{where} \quad \sigma_t^{*-2} = \sigma_t^{-2} + \frac{y_t^2}{2\beta^2\exp(\mu_t)}.$$

Notice that if $y_t = 0$, then $\sigma_t^{*2} = \sigma_t^2$ and f is truly normally distributed and equals g. The precision of σ_t^{*-2} increases with y_t^2.

In the accept/reject part of the algorithm, these suggestions are made until acceptance with probability $\min(f/g, 1)$, while the Metropolis probability of accepting the resulting z is

$$\min\left(\frac{f(z \mid y_t, \alpha_{\backslash t}^{(k)}) \min\left[f(\alpha_t^{(k)} \mid y_t, \alpha_{\backslash t}^{(k)}), g(\alpha_t^{(k)})\right]}{f(\alpha_t^{(k)} \mid y_t, \alpha_{\backslash t}^{(k)}) \min\left[f(z \mid y_t, \alpha_{\backslash t}^{(k)}), g(z)\right]}, 1\right).$$

A slightly less direct way of thinking about this analysis is that we are using a Gaussian approximation to the log-likelihood, $\log f(y_t|\alpha_t)$, which is then added to the then-conjugate Gaussian prior to deliver a Gaussian posterior. Notice that as y_t goes to zero this likelihood becomes uninformative, although the posterior is perfectly well-behaved. This way of thinking about this problem is easily extended to updating more than a single state at a time.

Here we pursue a Bayesian analysis of the parameters in the model. Given the states, sampling from the parameters is straightforward for β and σ_η^2. First assuming a flat prior for $\log \beta$ we achieve the posterior $\beta^2|y, \alpha \sim \chi_n^{-2} \sum y_t \exp(-\alpha_t)$, while assuming a prior of $\chi_p^{-2} S_0$ for $\sigma_\eta^2|\phi$ we have

$$\sigma_\eta^2|\phi, \alpha \sim \chi_{n+p}^{-2} \left\{ \sum_{t=2}^{n} (\alpha_t - \phi\alpha_{t-1})^2 + \alpha_1^2(1-\phi^2) + S_0 \right\}.$$

In this work we assume for daily data $p = 10$ and $S_0 = p \times 0.01$, while for weekly data S_0 is taken to be $p \times 0.05$. Our prior on the persistence parameter ϕ is designed to ensure that the log volatility process is stationary. To carry this out we employ a beta prior on $(\phi + 1)/2$, with $\mathsf{E}(\phi) = \{2\delta/(\gamma + \delta)\} - 1$. In the analysis we use below, we set $\delta = 20$ and $\gamma = 1.5$ so the prior mean is 0.86 and standard deviation is 0.11. As a result of our prior choice, the posterior is non-conjugate and we sample from $\phi|\sigma_\eta^2, \alpha$ using an accept/reject algorithm.

2·3 *Empirical effectiveness*

Our MCMC sampler is initialised by setting all the log-volatilities to zero and $\phi = 0.95$, $\sigma_\eta^2 = 0.02$ and $\beta = 1$. We iterated the MCMC sampler on the states for 1,000 iterations and then on the parameters and states for 50,000 more iterations, before recording any answers. The next 1,000,000 iterations are graphed in Figure 1 and summarised in Table 1.

The correlogram shows significant autocorrelations for ϕ at 10,000 lags, for β at 25,000 lags and for σ_η at 5,000 lags.

3 DESIGNING BLOCKS

3·1 *Background*

Liu et al. (1994) suggest blocking to improve the speed of convergence for simulators. This idea has motivated a considerable time series statistics literature on this topic in models built out of Gaussianity, but with some non-normality mixed in (Carter & Kohn, 1994; Shephard,

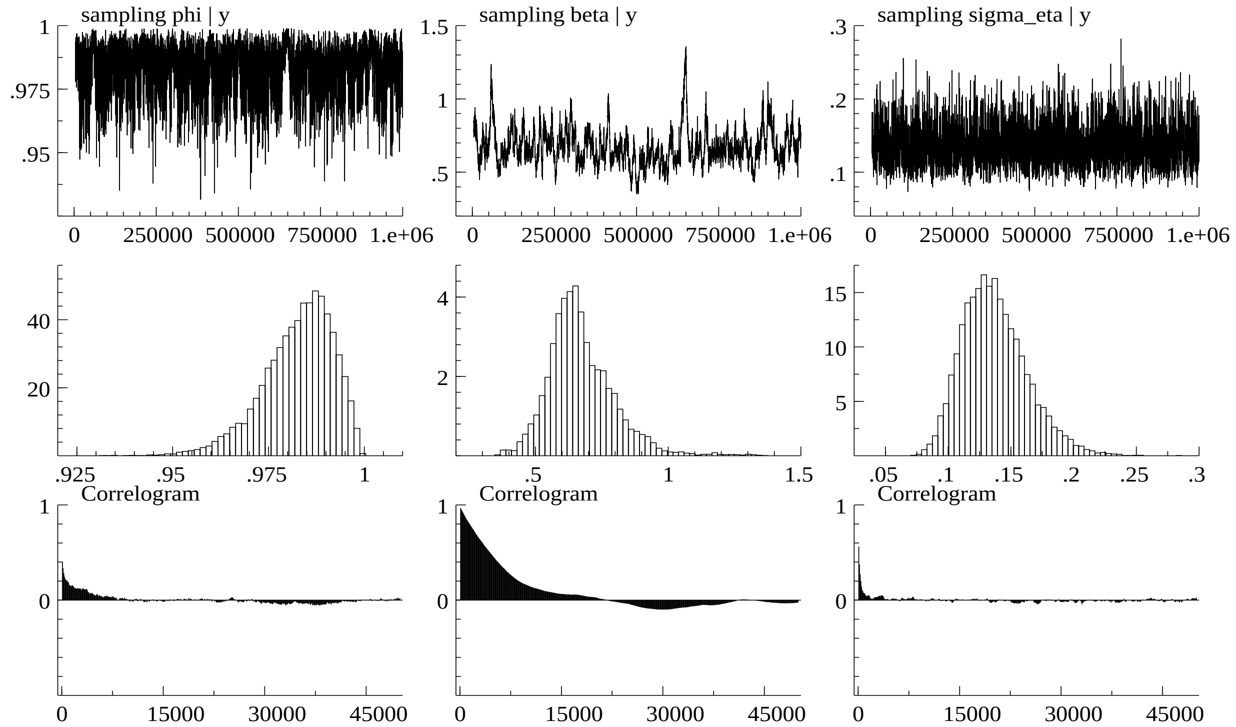

Figure 1: *Returns for Pound/Dollar. Top graphs: simulation against iteration. Middle: histograms of marginal distribution. Bottom: corresponding correlogram for simulation.*

	Mean	Monte Carlo S.E.	Covariance & *Correlation* of Posterior		
$\phi\|y$	0.9821	0.000277	8.343×10^{-5}	*-0.629*	*0.217*
$\sigma_\eta\|y$	0.1382	0.000562	-0.0001479	0.0006631	*-0.156*
$\beta\|y$	0.6594	0.0121	0.0002089	-0.0004228	0.01209
Computer time	233,303				

Table 1: *Returns for the Pound/Dollar. Summaries of Figure 1. The S.E. of simulation computed using B_M = 100,000. Italics are correlations. Computer time: seconds on a P5/133.*

1994b). However, the non-Gaussian measurement state space models are not of this form: they are fundamentally non-Gaussian. Hence there is a substantial need for a new strategy.

Before we discuss our suggestion, a comment on our notation is in order. The states α in (1·2) usually have many identities linking the state variables over time, consequently when we simulate from $\alpha \mid y$ we simulate from a possibly highly degenerate density. To avoid this complication we can focus on the fundamental disturbances in the model u_t, ensuring that $u \mid y$ is non-degenerate. To simplify the exposition we remove the regressors, W_t and x_t, from the analysis in the remainder of this section.

Sampling directly from $u \mid y$ may be too ambitious as this is highly multivariate and so if n is very large it is likely that we will run into very large rejection frequencies, counteracting the effectiveness of the blocking. Hence we employ an intermediate strategy. Our sampling method is based around sampling blocks of disturbances, say $u_{t,k} = (u_{t-1}, ..., u_{t+k-1})'$, given beginning and end conditions, α_{t-1} and α_{t+k+1}, and the observations. We call these end conditions "stochastic knots", the nomenclature being selected by analogy with their role in splines. At the beginning and end of the data set, there is no need to use two-sided knots. In these cases only single knots are required, simulating, for example, from $u_{1,k}|\alpha_{t+k+1}$, $y_1, ..., y_{t+k}$.

3·2 *Proposal density*

The only viable way of carrying out sampling from $u_{t,k}|\alpha_{t-1},\alpha_{t+k+1}, y_t, ..., y_{t+k}$ seems to be to place it within a Metropolis algorithm, employing multivariate Gaussian proposals. We now show how to do this in an effective way. The basis of our work is the use of a Taylor type expansion of

$$\log f = \log f(u_{t-1}, ..., u_{t+k-1} \mid \alpha_{t-1}, \alpha_{t+k+1}, y_t, ..., y_{t+k})$$

around some preliminary estimate of $u_{t,k}$ made using the conditioning arguments $\alpha_{t-1}, \alpha_{t+k+1}$ and $y_t, ..., y_{t+k}$. These estimates, and the corresponding $\alpha_t, ..., \alpha_{t+k}$ and $\theta_t, ..., \theta_{t+k}$, are denoted by hats. How they are formed will be discussed in a moment. We need some notation. We write $l(\theta_t)$ to denote $\log f(y_t|\theta_t)$ and the derivative of this log-density with respect to d_t as l' and l'' respectively. The expansion is then

$$\begin{aligned}
\log f &= -\tfrac{1}{2}u'_{t,k}u_{t,k} + \textstyle\sum_{s=t}^{t+k} l(\theta_s), \quad \alpha_{s+1} = T_s\alpha_s + H_s u_s, \\
&\simeq -\tfrac{1}{2}u'_{t,k}u_{t,k} + \textstyle\sum_{s=t}^{t+k} l(\widehat{\theta}_s) + z_s(\alpha_s - \widehat{\alpha}_s)l'(\widehat{\theta}_s) + \tfrac{1}{2}\left\{z_s(\alpha_s - \widehat{\alpha}_s)\right\}^2 D_s(\widehat{\theta}_s), \\
&= \log g.
\end{aligned}$$

We require the assumption that the as yet unspecified $D_s(\theta)$ is everywhere strictly negative as a function of θ. Typically we take $D_s(\widehat{\theta}_s) = l''(\widehat{\theta}_s)$ so that the approximation is a second order Taylor expansion. This is convenient, for in the vast majority of cases l'' is everywhere strictly negative. However, it is useful in covering unusual cases to have the possibility of not setting D_s equal to the second derivative. Of course for those cases, we have to provide sensible rules for the selection of D_s.

A crucial, and attractive feature of this expansion is that the error in the approximation involves only the difference between $l(\theta_s)$ and $l(\widehat{\theta}_s)+z_s(\alpha_s-\widehat{\alpha}_s)l'(\widehat{\theta}_s)+\frac{1}{2}\left\{z_s(\alpha_s - \widehat{\alpha}_s)\right\}^2 D_s(\widehat{\theta}_s)$, not the transition equation $u'_{t,k}u_{t,k}$. The implication is that the algorithm should not become significantly less effective as the dimension of u_t increases. This can be contrasted with the numerical integration routines used in Kitagawa (1987), which usually deteriorates in its effectiveness as the dimension of α_t increases.

Some interesting relations can be found for the crucial l' function. First, in general

$$l'(\widehat{\theta}_s) = \left.\frac{\partial\theta_t}{\partial d_t}\right|_{\theta_t=\widehat{\theta}_t} \times \left.\frac{\partial l(\theta_t)}{\partial\theta_t}\right|_{\theta_t=\widehat{\theta}_s},$$

which for the exponential family becomes $l'(\widehat{\theta}_s) = \frac{\partial \theta_t}{\partial d_t}\Big|_{\theta_t=\widehat{\theta}_t} \times \left\{ y_t - \dot{b}\,(\widehat{\theta}_t) \right\}$. Similar manipulations are possible for $l''(\widehat{\theta}_t)$. The most interesting case of this is the canonical link, where $l''(\widehat{\theta}_t) = -\ddot{b}\,(\widehat{\theta}_t)$.

The density of g is highly multivariate Gaussian. It is not a dominating density for $\log f$, but it is usually a good approximation. If we write $v_s^{-1} = -D_s(\widehat{\theta}_s)$, then the joint density can be calculated by defining the artificial variables

$$\widehat{y}_s = z_s \widehat{\alpha}_s + v_s l'(\widehat{\theta}_s), \quad s = t, ..., t+k$$

and then modelling them as

$$\begin{array}{rcl} \widehat{y}_s & = & z_s \alpha_s + \varepsilon_s, \quad \varepsilon_s \sim \mathsf{N}\left(0, v_s\right), \\ \alpha_{s+1} & = & T_s \alpha_s + H_s u_s, \quad u_s \sim NID(0, I). \end{array} \tag{3·1}$$

Consequently, it is possible to simulate from the approximating posterior density of $u_{t,k}$, given knots, using the de Jong & Shephard (1995) simulation smoother on the artificial $\widehat{y}_t, ..., \widehat{y}_{t+k}$. As g does not bound f these draws from the simulation smoother provide suggestions for the pseudo-dominating Metropolis algorithm discussed in § 1.

A similar expansion to (3·1), but without the knots, is used in Fahrmeir (1992), Singh & Roberts (1992) and the unpublished work of Durbin and Koopman. These authors use an moment smoother (7·4) to provide an estimate of $u|y$ using this Taylor expanded approximation. By iterating around the latest smoothed values, this algorithm converges to the mode of the density of $u|y$ in cases where $\partial^2 \log l / \partial \alpha_t \partial \alpha_t'$ is negative semi-definite; the same condition is needed for generalized linear regression models to have a unique maximum (see McCullagh & Nelder (1989, p. 117)). It is typically stated as a requirement that the link function be log-concave. In the case being addressed here, the iterations are an efficient way of carrying out a Newton-Raphson hill climbing algorithm using analytic first and second derivatives on a concave objective function. Thus the approximations suggested by previous authors can be interpreted as Laplace approximations to a very high dimensional density function. Our algorithm, by contrast, converges to a sequence of simulations from $u|y$.

It is informative to look at some special cases of the setup given in (3·1), to see the effect of non-Gaussianity on the approximating model.

Example 1: canonical link If a canonical link is assumed, that is $\theta_t = d_t$ and $\log f = y_t\theta_t - b(\theta_t) + c(y_t)$, then

$$v_s^{-1} = \ddot{b}\,(\widehat{\theta}_s) \text{ and } \widehat{y}_s = d_s + v_s \left\{ y_s - \dot{b}\,(\widehat{\theta}_s) \right\}.$$

A special case of this is the poisson model, where $b(\theta_s) = \exp(\theta_s)$ and so

$$v_s^{-1} = \exp(d_s), \;\; \widehat{y}_s = d_s + \exp(-d_s)\left\{ y_s - \exp(d_s) \right\}.$$

Another important example is the binomial, where $b(\theta_s) = n\log\{1+\exp(\theta_t)\}$. For this model

$$v_s^{-1} = np_t(1-p_t),\ \widehat{y}_s = d_s + (y_s - np_s)\,/\,\{np_t(1-p_t)\}\,, \text{where } p_t = \exp(d_t)/\,\{1+\exp(d_t)\}\,.$$

Example 2: SV model This model has $\log f(y_t|\alpha_t) = -\alpha_t/2 - y_t^2\exp(-\alpha_t)/2\beta^2$. Thus

$$v_s^{-1} = \frac{y_s^2}{2\beta^2}\exp(-d_s),\ \widehat{y}_s = d_s + \frac{v_s}{2}\left\{y_s^2\exp(-d_s)/\beta^2 - 1\right\}.$$

This case is particularly interesting as v_s depends on y_s, unlike with canonical links. As $y_s \to 0$ so $v_s^{-1} \to 0$ and $\widehat{y}_s \to -\infty$, suggesting that the draws from the simulation smoother might always be rejected. However, this observation ignores the fact that $v_s \to \infty$, which effectively treats such observations as missing. Of course there is a numerical overflow problem here, but that can be dealt with in a number of ways without resulting in any approximation.

Example 3: heavy-tailed SV model This argument extends to allow ϵ_t in (2·1) to follow a scaled t-distribution, $\epsilon_t = t_t/\sqrt{(\nu-2)/v}$, where $t_t \sim t_\nu$. Then

$$\log f(y_t|\alpha_t) = -\frac{\alpha_t}{2} - (\nu+1)/2\log\left[1 + \frac{y_t^2\exp(-\alpha_t)}{\beta^2(\nu-2)}\right], \quad \text{so}$$

$$l'(\alpha_t) = \frac{1}{2}\left[\frac{\frac{(\nu+1)}{2\beta^2(\nu-2)}y_t^2\exp(-\alpha_t)}{1+\frac{y_t^2\exp(-\alpha_t)}{\beta^2(\nu-2)}} - 1\right], \quad \text{and}$$

$$l''(\alpha_t) = -\frac{(\nu+1)}{4}\left[\frac{y_t^2\exp(-\alpha_t)/\beta^2(\nu-2)}{(1+\frac{y_t^2\exp(-\alpha_t)}{\beta^2(\nu-2)})^2}\right].$$

The resulting v_s^{-1} and $\widehat{y_s}$ are easy to compute. This approach has some advantages over the generic outlier approaches for Gaussian models suggested in Shephard (1994b) and Carter & Kohn (1994), which explicitly use the mixture representation of a t-distribution.

It is important to select sensible values for the sequence $\widehat{\theta}_t, ..., \widehat{\theta}_{t+k}$. The most straightforward choice would be to take them as the mode of $f(\theta_t, ..., \theta_{t+k}|\alpha_{t-1}, \alpha_{t+k+1}, y_t, ..., y_{t+k})$, the points at which the quadratic expansion is carried out. Although we have just noted that iterating the smoothing algorithm on (3·1) achieves this mode, this will slow our simulation algorithm if we have to iterate this procedure until full convergence (although rapid convergence is typical for these procedures). Instead we suggest using the pseudo-dominating MCMC method and only a fixed number of iterations of the smoothing algorithm to get a reasonably good sequence $\widehat{\theta}_t, ..., \widehat{\theta}_{t+k}$ instead of an "optimal" one. Typically we use two to five iterations.

3·3 *Stochastic knots*

The stochastic knots play a crucial role. They ensure that as the sample size increases our algorithm does not fail due to excessive numbers of rejections. In our discussion so far the knots are regarded as being fixed, however in practice there are advantages in selecting them randomly, allowing the points of conditioning to change over the iterations. Our proposal is to work with a collection of stochastic knots, at times $\underset{\sim}{k} = (k_1, ..., k_K)'$ and corresponding values $\underset{\sim}{\alpha} = (\alpha'_{k_1}, ..., \alpha'_{k_K})'$, which appropriately cover the time span of the sample.

The selection of the knots is carried out randomly and independently of the outcome of the MCMC process, with U_i being independent uniforms and

$$k_i = int\left[n \times \{(i + U_i)/(K+2)\}\right], \quad i = 1, ..., K. \tag{3·2}$$

Thus the selection of knots is indexed by a single parameter K which we control. Notice that in the way we have set this up, it is not possible to select K to recover the single-move algorithm as a special case.

In practice K is a tuning parameter, allowing the lengths of blocks to be selected. If K is too large the sampler will be slow because of rejections; if K is too small it will be correlated because of the structure of the model.

3·4 *Illustration on SV model*

To illustrate the effect of blocking we work with the SV model, analysing the MCMC algorithms on simulated and real data. Our simulated data allow two sets of parameters, designed to reflect typical problems for weekly and daily financial data sets. In the weekly case, $\beta = 1$, $\sigma^2_\eta = 0.1$ and $\phi = 0.9$, while in the daily case $\beta = 1$, $\sigma^2_\eta = 0.01$ and $\phi = 0.99$.

Table 2 reports some results from a simulation using $n = 1,000$. The Table splits into two sections. The first is concerned with estimating the states given the parameters, a pure signal extraction problem. It is clearly difficult to summarise the results for all 1,000 time periods and so we focus on the middle state, α_{500}, in all our calculations. Extensive simulations suggest that our results reported here are representative of these general results.

The second part of the Table looks at the estimation of the states at the same time as estimating the parameters of the model. Hence for that simulation the problem is a four-fold one: estimate the states and three parameters.

The Table reports the ratio of the resulting variance of the single-move sampler to the multi-move sampler. Numbers bigger than one reflect gains from using a multi-move sampler. One interpretation of the Table is that if the ratio is x then the single-move sampler has to be iterated x times more than the multi-move sampler to achieve the same degree of precision in the estimates of interest. So if a sample is 10 times more efficient, then it produces the same degree of accuracy from 1,000 iterations as 10,000 iterations from the inferior simulator.

Weekly parameters	K=0	K=1	K=3	K=5	K=10	K=20	K=50	K=100	K=200
States|parameters	1.7	4.1	7.8	17	45	14	12	4.3	3.0
States	20	22	32	28	39	12	12	21	1.98
σ_η	1.3	1.3	1.1	1.7	2.2	1.7	1.5	1.7	1.5
ϕ	1.5	1.4	1.1	1.6	2.6	1.7	1.5	2.0	1.5
β	16	14	32	14	23	6	9	10	1

Daily parameters	K=0	K=1	K=3	K=5	K=10	K=20	K=50	K=100	K=200
States|parameters	66	98	98	85	103	69	25	8.5	2.5
States	91	40	30	60	47	80	14	27	18
σ_η	2.7	3.1	2.9	2.8	3.4	3.8	2.6	3.6	1.8
ϕ	16	13	18	18	16	18	6.8	11	5.7
β	93	51	51	76	65	106	23	27	26

Table 2: *Relative efficiency of block to single-move MCMC sampler. K denotes number of stochastic knots. Reported are the ratio of the computed variances, which reflect efficiency gains. The variances are computed using B_M = 10,000 and 100,000 iterations in all cases except for the single-move sampler on daily parameters cases. For that problem B_M = 100,000 and 1,000,000 iterations were used. The burn-in period is B_M.*

Table 2 shows that when the number of knots is extremely small, say 0, there is a possibility in low persistence cases that the sampler actually performs worse than a single-move algorithm. This is because of large rejection rates. However, this result is very specialised. In the vast majority of cases the gains are quite significant and most importantly, they are largest in cases where the single-move algorithms do badly. This suggests that the use of block samplers does increase the reliability of MCMC techniques for state space models.

To illustrate the effectiveness of this method we return to the application considered in § 2. We use 10 stochastic knots in our block sampler. The results are displayed in Figure 2. This was generated by using 200 iterations using the initial parameter, 300 iterations updating the parameters and states and finally we record 10,000 iterations from the equilibrium path of the sampler. The Figure shares the features of Figure 1, with the same distribution for the parameters. However, the correlations amongst the simulations are now quite manageable.

The precision of the results from the simulation is calculated using the Parzen window with $B_M = 1,000$. The results given in Table 3 are consistent with those given in Table 1 for the single-move algorithm. However, the precision achieved with the 1,000,000 iterations is broadly comparable with that achieved by 10,000 iterations from the multi-move sampler.

One of the most powerful aspects of the approach we are advocating is that the block size controls the efficiency of the methods. It is important to note, however, that the methods

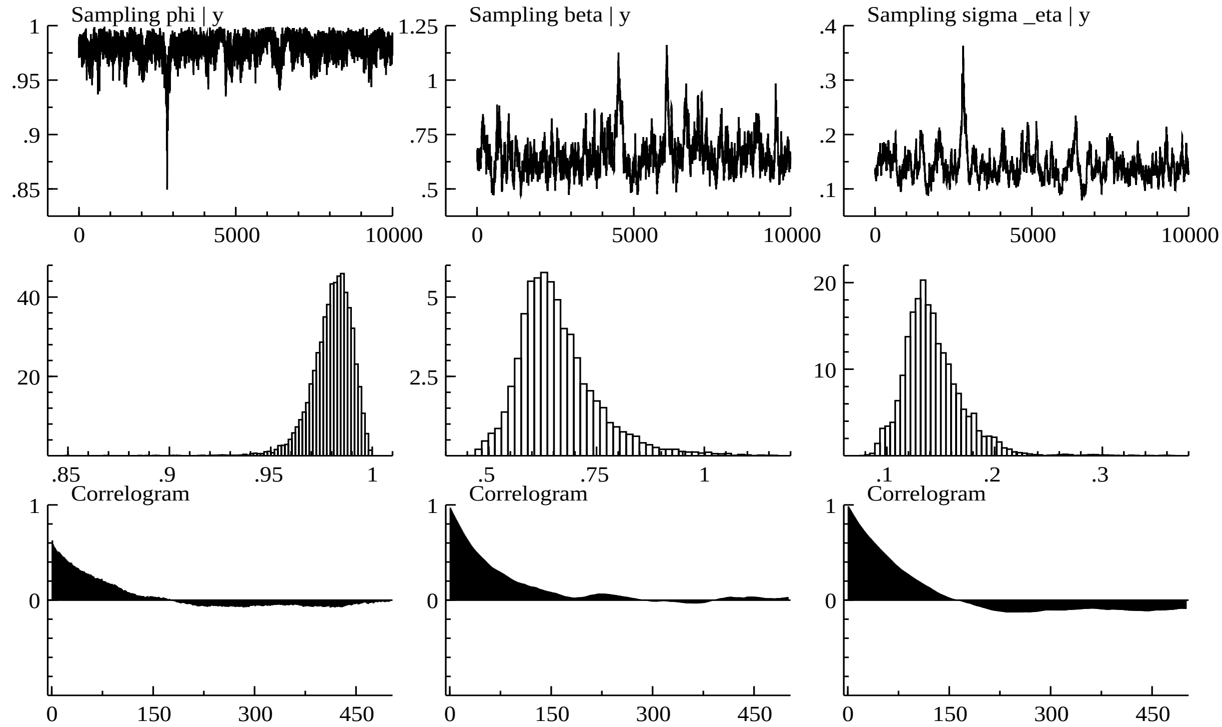

Figure 2: *Returns for Pound/Dollar. Top graphs: simulation against iteration. Middle: histograms of marginal distribution. Bottom: corresponding correlogram for simulation.*

	Mean	Monte Carlo S.E.	Covariance & *Correlation* of Posterior		
$\phi\|y$	0.9802	0.000734	0.000105	*-0.689*	*0.294*
$\sigma_\eta\|y$	0.1431	0.00254	-0.000198	0.000787	*-0.178*
$\beta\|y$	0.6589	0.0100	0.000273	-0.000452	0.00823
Computer time	322				

Table 3: *Returns for the Pound/Dollar. Summaries of Figure 2. The S.E. of simulation computed using $B_M = 1{,}000$. Italics are correlations. Computer time: seconds on a P5/133.*

work whatever the block size.

It would be convenient if we had a method which would automatically find the "optimal" value for K. But it seems likely that this size will be different for different aspects of the problem and in any case, our results suggest extensive robustness of the efficiency gains, with large efficiency gains for blocks between 50 and 500.

4 CLASSICAL ESTIMATION

4·1 *An importance sampler*

So far this paper has focused on Bayesian estimation of the parameters which index the models. We now turn our attention to classical inference. Our suggestion is to construct the likelihood via the use of importance sampling (Ripley, 1987, pp. 122-3) drawing samples from the Gaussian approximation to the smoothing density we developed in § 3. There is no use of stochastic knots, nor of acceptance/rejection sampling. We write this density as $f(u|y;\psi)\tilde{}$, using the disturbances rather than the states, and then use the manipulation:

$$f(y;\psi) = \int \frac{f(y|u;\psi)f(u;\psi)}{f(u|y;\psi)\tilde{}} f(u|y;\psi)\tilde{}\, du.$$

The estimator of this integral takes on the form

$$\widehat{f(y;\psi)}_M = \frac{1}{M}\sum_{j=1}^{M} \frac{f(y|u^j;\psi)f(u^j;\psi)}{f(u^j|y;\psi)\tilde{}}, \quad \text{where} \quad u^j \sim f(u|y;\psi)\tilde{}. \tag{4·1}$$

Crucially, this estimator is differentiable in ψ due to the Gaussianity of the smoothing density. If $f(u|y;\psi)\tilde{} = f(u|y;\psi)$ then this estimate is exact even if $M = 1$. Further, it is straightforward to exploit antithetic variables (Ripley (1987, pp. 129–32)) for this problem by simply switching the signs on the κ_t in the simulation smoother (7·1).

4·2 *Technical issues*

The estimate (4·1) involves simulating from $f(u|y;\psi)\tilde{}$ and evaluating $f(u^j|y;\psi)\tilde{}$, $f(u^j;\psi)$ and $f(y|u^j;\psi)$. The simulation is straightforward using the simulation smoother, while $f(u^j;\psi)$ and $f(y|u^j;\psi)$ are trivial to evaluate. The term $f(u^j|y;\psi)\tilde{}$ can be found using the smoothing density (de Jong & Shephard, 1995).

Proposition 1

Define $X_j = f(u^j|y;\psi)/f(u^j|y;\psi)\tilde{}$ where $u^j \sim f(u^j|y;\psi)\tilde{}$, then X_j is (conditional on y) independent and identically distributed with the moments

$$E(X_j|y) = 1 \quad \text{and} \quad Var(X_j|y;\psi) = \int \left(\frac{f(u|y;\psi)}{f(u|y;\psi)\tilde{}} - 1\right)^2 f(u|y;\psi)\tilde{}\, du = \sigma^2(y;\psi).$$

We define $f(y;\psi)^j = f(y|u^j;\psi)f(u^j;\psi)/f(u^j|y;\psi)\tilde{}$ and so

$$f(y;\psi)^j = f(y;\psi)\frac{f(u^j|y;\psi)}{f(u^j|y;\psi)\tilde{}} = f(y;\psi)X_j.$$

Consequently the estimator $\widehat{f(y;\psi)}_M$ is unbiased with variance of $\sigma^2(y;\psi)/M$.

For classical inference it is the log-likelihood which is the focus of attention. Our estimator of this $\log \widehat{f(y;\psi)}_M$ has the property

$$\log \widehat{f(y;\psi)}_M = \log f(y;\psi) + \log \frac{1}{M}\sum_{j=1}^{M} X_j.$$

The implication of this is that

$$E\left[\log\left\{1 + \frac{1}{M}\sum_{j=1}^{M}(X_j - 1)\right\}\right] = -\frac{\sigma^2(y;\psi)}{2M} + O(M^{-2}),$$

and so the estimator of $\log f\widehat{(y;\psi)}_M$ is biased to $O(M^{-1})$. As the $\sigma^2(y;\psi)$ depends on the parameters of the model, this causes an $O(M^{-1})$ bias in the estimates of the parameters of the model. The magnitude of this bias will depend on the merit of the importance sampling device. However, importantly, it is likely that $\sigma^2(y;\psi) = O(n)$. This means that as the sample size increases, the importance sampling bias will become larger unless M increases at a rate larger than n. This is unfortunate.

The bias is proportional to $\sigma^2(y;\psi)$, which can be unbiasedly estimated by

$$\frac{1}{M-1}\sum_{j=1}^{M}\left(f(y;\psi)^j - f\widehat{(y;\psi)}_M\right)^2, \quad \text{hence}$$

$$\log f\widehat{(y;\psi)}_M + \frac{1}{2M}\frac{1}{M-1}\sum_{j=1}^{M}\left(f(y;\psi)^j - f\widehat{(y;\psi)}_M\right)^2,$$

gives an unbiased estimator to $O(M^{-2})$. Of course this does not mean that the resulting estimator of ψ is superior, for the addition of this term may sufficiently increase the variance of the estimator of the log-likelihood so as to compensate for the reduction in bias.

5 CONCLUSIONS

This paper uses simulation to provide a likelihood basis for non-Gaussians extensions of state space models. We argue that the development of Taylor expansion-based multi-move simulation smoothing algorithms can deliver reliable methods.

The methods have five basic advantages. (1) They integrate the role of the Kalman filter and simulation smoother into the analysis of non-Gaussian models, so exploiting the structure of the model to improve the speed of the methods. (2) They exploit the Taylor expansion which has been previously used in various approximate methods suggested in the literature. (3) They approximate only $\log f(y_t|\theta_t)$ and so as the dimension of the state increases the computational efficiency of the method should not diminish significantly. (4) They extend to many multivariate cases by using a multivariate Taylor expansion of $\log f$ to deliver a multivariate approximating version of the model (3·1). (5) They allow the transition equation (1·2) to become non-Gaussian by Taylor expanding the transition density $\log f(\alpha_{t+1} \mid \alpha_t)$. Then the Metropolis rejection rate will additionally depend on the accuracy of the approximation to transition density.

Although there have been very significant recent advances in signal extraction of non-Gaussian processes, there is much work to be carried out in this area. We have not provided any model checking devices for our fitted models. In principle this is straightforward if based on one-step-ahead prediction densities. Some progress has recently been made in finding methods for computing these densities (Gordon et al., 1993; Muller, 1991; West, 1993).

6 ACKNOWLEDGEMENTS

We would like to thank the ESRC for their financial support under the grant "Estimation via simulation in econometrics." All the computations reported in this paper were generated using `C++` code dynamically linked to the `Ox` Matrix Programming Language (Doornik (1996)). We thank Jurgen Doornik for helpful computational advice and Aurora Manrique, the referee and the Associate Editor for making some useful suggestions.

7 APPENDIX

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

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

We assume that $G_t' H_t = 0$ and write the non-zero rows of H_t as M_t. The Kalman filter (de Jong, 1989) computes $a_{t|t-1} = E\alpha_t|Y_{t-1}, \beta$ and $P_{t|t-1} = MSE(\alpha_t|Y_{t-1}, \beta)$,

$$\begin{array}{lcllcllcl} a_{t+1|t} & = & W_t\beta + T_t a_{t|t-1} + K_t v_t, & P_{t+1|t} & = & T_t P_{t|t-1} L_t' + H_t H_t', & v_t & = & y_t - Z_t a_{t|t-1} - x_t\beta, \\ F_t & = & Z_t P_{t|t-1} Z_t' + G_t G_t', & K_t & = & T_t P_{t|t-1} Z_t' F_t^{-1}, & L_t & = & T_t - K_t Z_t. \end{array}$$

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) samples from $\alpha|Y_n, \beta$. Setting $r_n = 0$ and $N_n = 0$, for $t = n, ..., 1$

$$\begin{array}{lcl} C_t & = & M_t M_t' - M_t H_t' N_t H_t M_t', \qquad \kappa_t \sim \mathsf{N}(0, C_t), \\ r_{t-1} & = & Z_t' F_t^{-1} v_t + L_t' r_t - L_t' N_t H_t M_t' C_t^{-1} \kappa_t, \\ N_{t-1} & = & Z_t' F_t^{-1} Z_t + L_t' N_t L_t + L_t' N_t H_t M_t' C_t^{-1} M_t H_t' N_t L_t, \end{array} \tag{7·1}$$

writing $\widehat{M_t u_t} = M_t H_t' r_t + \kappa_t$. Adding to $\widehat{M_t u_t}$ the corresponding zero rows so that we simulate from the whole $H_t u_t$ vector, written $\widehat{\eta}_t$. The end condition $\widehat{\eta_0}$ is calculated by

$$C_0 = P_{1|0} - P_{1|0} N_0 P_{1|0}, \quad \kappa_0 \sim \mathsf{N}(0, C_0), \quad \eta_0 = P_{1|0} r_0 + \kappa_0. \tag{7·2}$$

The α vector to be simulated via the forward recursion, starting with $\alpha_0 = 0$,

$$\alpha_{t+1} = W_t\beta + T_t\alpha_t + \widehat{\eta}_t, \quad t = 0, ..., n-1. \tag{7·3}$$

The moment smoother (de Jong, 1989; Koopman, 1993) computes $a_{t|n} = E\alpha_t|Y_n, \beta$. Setting $r_n = 0$ it runs backwards:

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \quad t = n, \dots, 1. \tag{7·4}$$

The $\widehat{M_t u_t} = M_t H_t' r_t$ completes, as above, $\widehat{\eta}_t$ allowing a recursion of the form (7·3) for the conditional expectations.

References

Carlin, B. P., Gelfand, A. E., & Smith, A. F. M. (1992a). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics* **41**, 389–405.

Carlin, B. P., Polson, N. G., & Stoffer, D. (1992b). 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.

Carter, C. K. & Kohn, R. (1996). Markov Chain Monte Carlo in conditionally Gaussian state space models. *Biometrika* **83**. Forthcoming.

Chan, K. S. & Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *J. Am. Statist. Assoc.* **89**, 242–252.

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

Danielsson, J. & Richard, J. F. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *J. Appl. Econometrics* **8**, S153–S174.

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.

Doornik, J. A. (1996). *Ox: Object Oriented Matrix Programming, 1.09.* Chapman & Hall, London.

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

Fruhwirth-Schnatter, S. (1992). Approximate predictive integrals for dynamic generalized linear models. In Fahrmeir, L., Francis, B., Gilchrist, R., & Tutz, G., editors, *Advances in GLIM and Statistical Modelling, Proceedings of the GLIM92 Conference and the 7th International Workshop on Statistical Modelling, Munich, 13-17 July 1992*, pages 101–106. Springer-Verlag, New York.

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

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). A novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F* **140**, 107–133.

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.

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

Harvey, A. C. & Fernandes, C. (1989). Time series models for count data or qualitative observations. *J. Business and Economic Statist.* **7**, 407–417.

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.

Kitagawa, G. (1987). Non-Gaussian statee space modelling of non-stationary time series. *J. Am. Statist. Assoc.* **82**, 503–514.

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

Liu, J., Wong, W. H., & Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes. *Biometrika* **81**, 27–40.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models.* Chapman & Hall, London. 2nd Edition.

Muller, P. (1991). Numerical integration in general dynamic models. *Contemporary Mathematics* **115**, 145–163.

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

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

Shephard, N. (1994a). Local scale model: state space alternative to integrated GARCH processes. *J. Econometrics* **60**, 181–202.

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

Singh, A. C. & Roberts, G. R. (1992). State space modelling cross-classified time series and counts. *International Statistical Review* **60**, 321–335.

Smith, J. Q. (1979). A generalization of the Bayesian steady forecasting model. *J. R. Statist. Soc. B* **41**, 375–387.

Smith, J. Q. (1981). The multiparameter steady model. *J. R. Statist. Soc. B* **43**, 256–260.

Smith, R. L. & Miller, J. E. (1986). A non-Gaussian state space model and application to prediction records. *J. R. Statist. Soc. B* **48**, 79–88.

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

West, M. (1993). Approximating posterior distributions by mxitures. *J. R. Statist. Soc. B*

55, 409–442.

West, M., Harrison, P. J., & Migon, H. S. (1985). Dynamic generalised models and Bayesian forecasting (with discussion). *J. Am. Statist. Assoc.* **80**, 73–97.

	Mean	Monte Carlo S.E.	Covariance & *Correlation* of Posterior		
$\phi \| y$	0.9821	0.000277	8.343×10^{-5}	*-0.629*	*0.217*
$\sigma_\eta \| y$	0.1382	0.000562	-0.0001479	0.0006631	*-0.156*
$\beta \| y$	0.6594	0.0121	0.0002089	-0.0004228	0.01209
Computer time	233,303				

Table 1: *Returns for the Pound/Dollar. Summaries of Figure 1. The S.E. of simulation computed using B_M = 100,000. Italics are correlations. Computer time: seconds on a P5/133.*

Weekly parameters	K=0	K=1	K=3	K=5	K=10	K=20	K=50	K=100	K=200
States|parameters	1.7	4.1	7.8	17	45	14	12	4.3	3.0
States	20	22	32	28	39	12	12	21	1.98
σ_η	1.3	1.3	1.1	1.7	2.2	1.7	1.5	1.7	1.5
ϕ	1.5	1.4	1.1	1.6	2.6	1.7	1.5	2.0	1.5
β	16	14	32	14	23	6	9	10	1

Daily parameters	K=0	K=1	K=3	K=5	K=10	K=20	K=50	K=100	K=200
States|parameters	66	98	98	85	103	69	25	8.5	2.5
States	91	40	30	60	47	80	14	27	18
σ_η	2.7	3.1	2.9	2.8	3.4	3.8	2.6	3.6	1.8
ϕ	16	13	18	18	16	18	6.8	11	5.7
β	93	51	51	76	65	106	23	27	26

Table 2: *Relative efficiency of block to single-move MCMC sampler. K denotes number of stochastic knots. Reported are the ratio of the computed variances, which reflect efficiency gains. The variances are computed using B_M = 10,000 and 100,000 iterations in all cases except for the single-move sampler on daily parameters cases. For that problem B_M = 100,000 and 1,000,000 iterations were used. The burn-in period is B_M.*

	Mean	Monte Carlo S.E.	Covariance & *Correlation* of Posterior		
$\phi\|y$	0.9802	0.000734	0.000105	*-0.689*	*0.294*
$\sigma_\eta\|y$	0.1431	0.00254	-0.000198	0.000787	*-0.178*
$\beta\|y$	0.6589	0.0100	0.000273	-0.000452	0.00823
Computer time	322				

Table 3: *Returns for the Pound/Dollar. Summaries of Figure 2. The S.E. of simulation computed using* $B_M = 1{,}000$. *Italics are correlations. Computer time: seconds on a P5/133.*

Figure 1: *Returns for Pound/Dollar. Top graphs: simulation against iteration. Middle: histograms of marginal distribution. Bottom: corresponding correlogram for simulation.*

Figure 2: *Returns for Pound/Dollar. Top graphs: simulation against iteration. Middle: histograms of marginal distribution. Bottom: corresponding correlogram for simulation.*