Markov Chain Monte Carlo methods for Generalized Stochastic Volatility Models

Siddhartha Chib*
Olin School of Business, Washington University, St Louis, USA

Federico Nardari
Olin School of Business, Washington University, St Louis, USA

Neil Shephard
Nuffield College, Oxford OX1 1NF, UK

November 9, 1998

Abstract

This paper is concerned with Markov chain Monte Carlo based Bayesian inference in generalized models of stochastic volatility defined by heavy-tailed student-t distributions (with unknown degrees of freedom) and covariate effects in the observation and volatility equations. A simple, fast and highly efficient algorithm, that builds on Kim, Shephard and Chib (1998), is developed for estimating these models. Computation of the likelihood function by a particle filter is considered as are methods for constructing diagnostic measures and the model marginal likelihood. The techniques are applied in detail to daily returns on the S&P 500 index and to weekly changes in the short-term interest rate.

Keywords: Bayes factor, Markov chain monte carlo, marginal likelihood, mixture models, particle filters, simulation based inference, stochastic volatility.

1 Introduction

The efficient fitting of models with stochastic volatility is one of the more challenging problems in modern time series analysis. These models can be interpreted as non-linear state space models in which the unobserved state variable is the stochastically evolving log volatility of the process. From a statistical perspective, therefore, methods for fitting such models have the potential to be applied more broadly although the analysis of the stochastic volatility models is important in its own right due to its significance for applications, especially

*Address for correspondence: John M. Olin School of Business, Washington University, Campus Box 1133, 1 Brookings Dr., St. Louis, MO 63130. Email addresses: chib@simon.wustl.edu.

in the area of high-frequency finance [see for example, Taylor (1994), Shephard (1996), and Ghysels, Harvey and Renault (1996)].

We shall be concerned in this paper with the following generalized stochastic volatility model

$$\begin{aligned} y_t &= x_{1t}'\beta_1 + w_t^{\gamma} \exp(h_t/2) u_t \\ h_t &= \mu + x_{2t}'\beta_2 + \phi(h_{t-1} - \mu) + \sigma\eta_t,\ t \leq n, \end{aligned} \tag{1}$$

where y_t is the response variable, x_{1t}, x_{2t} and w_t are covariates, $\{h_t\}$ is the latent log-volatility process with $\{\eta_t\}$ a white noise Gaussian sequence. Assume that the density of u_t is independent *student-t* with mean zero, dispersion one and *unknown* degrees of freedom ν. Then, the density of y_t conditioned on the parameters and h_t is a t-density with mean $x_{1t}'\beta_1$ and dispersion $w_t^{2\gamma} \exp(h_t)$ given by

$$f(y_t|h_t, \theta) = \frac{, \left((\nu+1)/2\right), \left(\nu/2\right)}{\left(\nu\pi w_t^{2\gamma} \exp(h_t)\right)^{1/2}} \left(1 + \frac{1}{\nu w_t^{2\gamma} \exp(h_t)} (y_t - x_{1t}'\beta_1)^2\right)^{-(1+\nu)/2}, \tag{2}$$

where $\psi \equiv (\beta_1, \beta_2, \gamma, \mu, \phi, \sigma, \nu)$ are the model parameters. Thus, in this model, the distribution of the responses is heavy-tailed and covariate effects are permitted in both the evolution and observation equations. If we assume that $|\phi| < 1$ and $\{x_{1t}, x_{2t}, w_t\}$ are a strictly stationary sequence, then the y_t process is strictly stationary.

The generalized stochastic volatility specified above has not been analyzed in the literature before from either the classical or Bayesian perspectives. Classical fitting of SV models is generally quite difficult because of the problems in constructing the likelihood function. A procedure to deal with this issue is discussed by Sandmann and Koopmans (1998) but the method (which is based on results in Shephard and Pitt (1997) and Durbin and Koopman (1997)), requires a number of tuning steps and is only tested on models that are much simpler than the model above. The existing Bayesian methods are also concentrated on the basic Gaussian SV models without covariates or a level effect [Jacquier, Polson and Rossi (1994) and Kim, Shephard and Chib (1998)].

The main point of the current paper is to develop a highly tuned and efficient MCMC based method for sampling the posterior distribution of ψ in the generalized stochastic volatility model. The method of Jacquier, Polson and Rossi (1994), which produces high

autocorrelation times even in the basic SV model, cannot be generalized to this model. The approach suggested in this paper builds on the framework of Kim, Shephard and Chib (1998) although except for the sampling of the latent volatilities the method developed here is entirely different. The proposed MCMC sampler is extremely efficient and fast and represents in our view the easiest and most effective way of fitting stochastic volatility models.

If we split the observation and evolution equation parameters as $\psi = (\beta_1, \theta, \nu)$, where $\theta = (\gamma, \mu, \beta_2, \phi, \sigma)$, then our proposed strategy rests on the *joint sampling* of the parameters θ from one reduced conditional distribution and ν (the degrees of freedom parameter) from another reduced conditional distribution.

The rest of the paper is organized as follows. Section 2 presents an alternative representation of the generalized stochastic volatility model that forms the basis of the proposed approach. The approach is outlined with details of each step deferred to the Appendix. Section 3 is concerned with posterior inferences based on the MCMC output. Procedures for computing the likelihood function, diagnostic measures and the model marginal likelihood are provided. Two real data examples are discussed in Section 4 and concluding remarks and possible applications of the approach to other models of topical interest are given in the last section.

2 The Proposed Approach

2.1 Alternative representation of the model

We begin by recalling the procedure introduced by Kim, Shephard and Chib (1998) for converting the stochastic volatility model into a conditionally Gaussian state space model. In order to achieve the same simplification, we first represent the student-t errors as scale mixture of normals and write $u_t = \lambda_t^{-1/2}\varepsilon_t$, where ε_t is a standard normal random variable and λ_t is i.i.d. $Gamma(\frac{\nu}{2}, \frac{\nu}{2})$. Now, conditioned on ψ and $\{\lambda_t\}$ the model may be reexpressed as

$$y_t^* = \gamma \log(w_t^2) + h_t + z_t \tag{3}$$

$$h_t = \mu + x_{2t}'\beta_2 + \phi(h_{t-1} - \mu) + \sigma\eta_t \tag{4}$$

where

$$y_t^* = \log\left(\frac{(y_t - x_{1t}'\beta_1)^2}{\lambda_t^{-1}}\right),$$

and $z_t = \log \varepsilon_t^2$.

It is easy to see that the observation equation of the model could also be represented as $y_t^* = h_t + z_t$, with y_t^* now defined y_t^* as $\log\left((y_t - x_{1t}'\beta_1)^2/(\lambda_t^{-1}w_t^{2\gamma})\right)$. We do not adopt this transformation, however, because the term $w_t^{2\gamma}$ in the denominator tends to make the ratio too small in general, leading to numerical problems in the evaluation of the logarithm.

The model is completed by specifying the distribution of z_t. The exact distribution is actually quite complicated but Kim, Shephard and Chib (1998) have determined that the distribution of z_t may be approximated rather closely by a seven-component mixture of normal densities f_N given by

$$\begin{aligned} z_t|s_t &\sim \mathcal{N}(m_{s_t}, v_{s_t}^2) \\ \Pr(s_t &= i) = q_i\,,\ i \leq 7,\ t \leq n, \end{aligned} \tag{5}$$

where $s_t \in (1, 2, ..., 7)$ is an unobserved component indicator with probability mass function $q = \{q_i\}$ and the parameters $\{q, m_{s_t}, v_{s_t}^2\}$ are as reported in Table 1.

s_t	q	m_{s_t}	$v_{s_t}^2$
1	0.00730	-11.40039	5.79596
2	0.10556	-5.24321	2.61369
3	0.00002	-9.83726	5.17950
4	0.04395	1.50746	0.16735
5	0.34001	-0.65098	0.64009
6	0.24566	0.52478	0.34023
7	0.25750	-2.35859	1.26261

Table 1: *Parameters of seven-component Gaussian mixture to approximate the distribution of* $\log \chi_1^2$.

In the MCMC context, the use of this approximation is highly recommended because the minor approximation error can be removed easily (at the conclusion of the posterior sampling) by a reweighting procedure, as discussed by Kim, Shephard and Chib (1998). This strategy of working with a highly efficient approximating model, and then reweighting

the posterior sample ex-post, is a powerful and general method of dealing with complicated models in the simulation context.

If we let $s = \{s_1, ..., s_n\}$, $\lambda = \{\lambda_1, ..., \lambda_n\}$ and $\mathcal{F}_t^* = (y_1^*, ..., y_t^*)$, then it should be noted that given this simplification, the density of y_t^* conditioned on (s, θ) but marginalized over $h = \{h_t\}$ can be expressed

$$f(y_1^*, ..., y_n^* | s, \theta) = \prod_{t=1}^{n} f(y_t^* | \mathcal{F}_{t-1}^*, s, \theta) \tag{6}$$

where each one-step ahead density $f(y_t^* | \mathcal{F}_{t-1}^*, s, \theta)$ can be derived from the output of the Kalman filter recursions (adapted to the differing components, as indicated by the component vector s).

2.2 Prior distributions

The generalized SV model provided above is completed through the specification of the distribution $\pi(\psi)$ on the parameters. In the context of our computational MCMC estimation scheme the nature of the prior is not important and virtually any distributional form may be adopted in the data analysis. Nonetheless, in the examples below we assume that the parameters are mutually independent and that prior information can be represented by the following distributions. For ϕ, we follow Kim, Shephard and Chib (1998) and formulate our prior in terms of $\phi = 2\phi^* - 1$, where ϕ^* is distributed as Beta with parameters $(\phi^{(1)}, \phi^{(2)})$. This implies that the prior on $\phi \in (-1, 1)$ is

$$\pi(\phi) = c\,(0.5(1+\phi))^{\phi^{(1)}-1}\,(0.5(1-\phi))^{\phi^{(2)}-1}, \quad \phi^{(1)}, \phi^{(2)} > 0.5\,, \tag{7}$$

where

$$c = 0.5\frac{,\,(\phi^{(1)} + \phi^{(2)})}{,\,(\phi^{(1)}),\,(\phi^{(2)})}$$

is the normalizing constant. Under this prior, the prior mean is

$$2\phi^{(1)}/(\phi^{(1)} + \phi^{(2)} - 1)\,.$$

This prior distribution is sufficient for our purposes although it is possible to proceed with alternative distributional forms, such as those considered by Chib and Greenberg (1994) and Marriott and Smith(1992). Each of these priors imposes stationarity which in our view

is important in this context. For γ we assume a uniform distribution on the range $(0, 2)$, which covers the values that have been considered in the literature. For σ we assume that our prior information can be represented by an inverse gamma density

$$\pi(\sigma) = \frac{\delta_0^{\nu_0}}{, (\nu_0)} \left(\frac{1}{\sigma}\right)^{\nu_0+1} \exp(-\delta_0/\sigma) \ , \sigma > 0$$

and for μ, β_1 and β_2 we assume independent normal priors $N(\mu|\mu_0, M_0)$, $N(\beta_1|\beta_{10}, B_{10}^{-1})$, and $N(\beta_2|\beta_{20}, B_{20}^{-1})$, respectively, where the hyperparameters $(\delta_0, \nu_0, \mu_0, M_0, \beta_{10}, B_{10}, \beta_{20}, B_{20})$ are assumed known. Finally, for the degrees of freedom ν we assume that the prior is uniform over the range $(2, 128)$.

2.3 Markov chain Monte Carlo algorithm

The model in the mixture representation has a straightforward conditional structure that lends itself to Markov Chain Monte Carlo simulations. The idea behind MCMC sampling is to construct a Markov chain whose limiting distribution is the target posterior density of interest. The Markov chain is then iterated a large number of times and the sampled draws, beyond a burn-in period, are treated as variates from the target distribution. For discussions of MCMC methods, see Chib and Greenberg (1995) and Gilks, Richardson and Spiegelhalter (1996). In the present context, the key idea is to construct the Markov chain based on the blocks $\{\beta_1, [\theta, h], s \text{ and } [\nu, \lambda]\}$, where the notation $[\theta, h]$ means that the parameters $\theta = (\gamma, \mu, \beta_2, \phi, \sigma)$ and h are sampled in one block, conditioned on the remaining blocks. In this algorithm, the parameters ν and λ are also sampled in one block conditioned on the other blocks. Extensive experimentation has shown that these step are the key to minimizing the serial correlation in the MCMC output.

We summarize the algorithm through the following steps.

MCMC algorithm for the generalized SV model

1. *Initialize* β_1, s, λ *and* ν

2. *Sample* θ *and* h *from* $\theta, h|y, s, \lambda, \beta_1$ *by drawing*

   (a) θ *from* $\theta|y^*, s, \beta_1$ *and*

(b) *h from* $h|y^*, s, \beta_1, \theta$

3. *Sample* β_1 *from* $\beta_1|y, h, s, \lambda, \gamma$

4. *Sample* s *from* $s|y^*, h$

5. *Sample* ν *and* λ_t *from* $\nu, \lambda|y, h, \psi$ *by drawing*

   (a) ν *from* $\nu|y, h, \psi$ *and*

   (b) λ_t *from* $\lambda_t|y_t, h_t, \psi, \nu$.

Steps 2a and Step 5a as mentioned above are the key to this algorithm. We implement Step 2a by using the Metropolis-Hastings algorithm [see for example, Chib and Greenberg (1995)] by making a proposal draw $\theta^i = (\gamma^i, \mu^i, \beta_2^i, \phi^i, \sigma^i)$ from a tailored multivariate-t density $q(\theta|m, V, \xi)$ with ξ degrees of freedom (the proposal density may alternatively be specified as multivariate normal). Specifically, the density $\log f(y^*|s, \theta)$ in (6) is numerically maximized over θ within each MCMC iteration; then, the maximizing value, say $\hat{\theta}$, is taken as the mean m of $q(.)$ and the inverse of minus the Hessian matrix (evaluated at $\hat{\theta}$) is taken as the scale matrix V of $q(.)$. The proposal value is accepted or rejected according to the Metropolis-Hastings algorithm.

Step 2b is implemented using the simulation smoother algorithm as in Kim, Shephard and Chib (1998). Step 3 follows from the update of a regression model with heteroskedastic errors. Step 4 corresponds to the sampling of s_t from a seven point discrete distribution in which the prior weights $\Pr(s_t)$ are updated to $\Pr(s_t) f_N(y_t^*|\gamma \log(w_t^2) + h_t + m_{s_t}, v_{s_t}^2)$ and then normalized. Finally, Step 5a involves the sampling of the degrees of freedom by a Metropolis-Hastings step from the reduced conditional density of ν (given by the product of student-t densities in equation (2)) and Step 5b is a drawing from updated gamma distributions. Full details of this algorithm are given in the Appendix.

We mention that while it may appear to be reasonable to sample γ and the remaining parameters of θ as separate blocks, the resulting sampler is ill-behaved due to the strong correlation between γ and ϕ. It is well known that strongly correlated components should be simulated as one block to minimize the serial correlation of the MCMC output. A similar

consideration led to our treating the parameters $(\mu, \beta_2, \phi, \sigma)$ and h as one block. The value of this blocking scheme is demonstrated in the applications below.

3 Posterior Inferences

In this section we turn to methods that provide the likelihood function of the model, diagnostic measures for evaluating model adequacy and the model marginal likelihood for comparing alternative models. It should be noted that we need the likelihood function (not for the sampling of the posterior distribution) but for computing the marginal likelihood. In the latter calculation, the likelihood function is required at just a single point in the parameter space.

3.1 Likelihood estimation

Consider the estimation of the likelihood function $f(y_1, ..., y_n|\mathcal{F}_0, \psi) = \prod_{t=1}^{n} f(y_t|\mathcal{F}_{t-1}, \psi)$ where ψ is some known value (say the estimated posterior mean from the MCMC simulations), $f(y_t|\mathcal{F}_{t-1}, \psi)$ is the student-t density in (2) marginalized over the distribution of $h_t|\mathcal{F}_{t-1}, \psi$ and $\mathcal{F}_t = (y_1, ..., y_t)$ is the data up to time t. Suppose that a sample of h_t^j $(j \leq M)$ is available from the filtered distribution $h_t|\mathcal{F}_t, \psi$ for each t. Then, the sequence of one-step ahead densities $f(y_t|\mathcal{F}_{t-1}, \psi)$ can be obtained by the following procedure.

Estimation of the likelihood function

1. *Set $t = 1$, initialize ψ and obtain a sample of h_{t-1}^j $(j \leq M)$.*

2. *For each value of h_{t-1}^j, sample h_t^j from the Gaussian evolution equation $h_t^j|h_{t-1}^j, \psi \sim \mathcal{N}\left(\mu + \phi(h_{t-1}^j - \mu) + x_{2t-1}'\beta_2, \sigma^2\right)$.*

3. *Estimate the one-step ahead density as*

$$\widehat{f}(y_t|\mathcal{F}_{t-1}, \psi) = \frac{1}{M}\sum_{j=1}^{M} f_T\left(y_t|x_{1t}'\beta_1, w_t^{2\gamma}\exp(h_t^j), \nu\right).$$

4. *Apply the filtering procedure described below to obtain a sample $h_t^1, ..., h_t^M$ from the filtered distribution $h_t|\mathcal{F}_t, \psi$.*

5. *Increment t to $t+1$ and goto Step 2.*

In this procedure Steps 2 and 3 are both straightforward. The key question is the sampling of $h_t^1, ..., h_t^M$ from the filtered distribution. This may be done using what is called the *particle filter* in the literature [see, for example, Gordon, Salmond and Smith (1993), Kitagawa (1996), Berzuini, Best, Gilks and Larizza (1997), Isard and Blake (1996), Pitt and Shephard (1997) and Carpenter, Clifford and Fearnhead (1998)]. From Bayes theorem,

$$f(h_t|\mathcal{F}_t, \psi) \propto f(y_t|h_t, \psi) f(h_t|\mathcal{F}_{t-1}, \psi) \tag{8}$$

where $f(y_t|h_t, \psi) = f_T\left(y_t|x_{1t}'\beta_1, w_t^{2\gamma}\exp(h_t), \nu\right)$ and

$$f(h_t|\mathcal{F}_{t-1}, \psi) = \int f(h_t|h_{t-1}, \psi) f(h_{t-1}|\mathcal{F}_{t-1}, \psi) dh_{t-1}.$$

In this case $f(h_t|h_{t-1}, \psi) = f_N\left(h_t|\mu + \phi(h_{t-1} - \mu) + x_{2t-1}'\beta_2, \sigma^2\right)$, a Gaussian evolution density. The latter integral can be estimated from the sample $h_{t-1}^1, ..., h_{t-1}^M$ leading to the approximations

$$\begin{aligned} f(h_t|\mathcal{F}_{t-1}, \psi) &\simeq \frac{1}{M}\sum_{j=1}^{M} f(h_t|h_{t-1}^j, \psi) \\ f(h_t|\mathcal{F}_t, \psi) &\dot{\propto} f(y_t|h_t, \psi)\frac{1}{M}\sum_{j=1}^{M} f(h_t|h_{t-1}^j, \psi). \end{aligned}$$

The question now is to sample h_t from the latter density. We work with a generic suggestion of Pitt and Shephard (1997) called an *auxiliary particle filter*. This filter requires a first stage in which proposal values $h_t^{*1}, ..., h_t^{*R}$ are created. These values are then reweighted to produce draws $\{h_t^1, ..., h_t^M\}$ that correspond to draws from $f(y_t|h_t, \psi)\sum_{j=1}^{M} f(h_t|h_{t-1}^j, \psi)$. Typically one may take R to be five or ten times larger than M. In all the examples below, we let $M = 2000$ and $R = 10,000$ though one may select even higher values if computer resources are available. We now summarize the steps involved for the filter in period t.

Auxiliary particle filter for generalized SV model

1. *Given values* $\{h_{t-1}^1, ..., h_{t-1}^M\}$ from $(h_{t-1}|\mathcal{F}_{t-1}, \psi)$ *calculate*

$$\hat{h}_t^{*j} = \mu + \phi(h_{t-1}^j - \mu) + x_{2t-1}'\beta_2,$$

$$w_j = f(y_t|\hat{h}_t^{*j}, \psi)\ ,\ \ j = 1, ..., M\,,$$

and sample R *times the integers* $1, 2, ..., M$ *with probability proportional to* $\{w_j\}$. *Let the sampled indexes be* $k_1, ..., k_R$ *and associate these with* $\hat{h}_t^{*k_1}, ..., \hat{h}_t^{*k_R}$.

2. *For each value of k_j from Step 1, simulate the values $\{h_t^{*1}, ..., h_t^{*R}\}$ from the volatility process as*

$$h_t^{*j} = \mu + \phi(h_{t-1}^{k_j} - \mu) + x'_{2t-1}\beta_2 + \sigma\eta_t^j\,,\ j = 1, ..., R,$$

where $\eta_t^j \sim N(0,1)$.

3. *Resample the values $\{h_t^{*1}, ..., h_t^{*R}\}$ M times with replacement using probabilities proportional to*

$$\frac{f_T\left(y_t|x'_{1t}\beta_1, w_t^{2\gamma}\exp(h_t^{*j}), \nu\right)}{f_T\left(y_t|x'_{1t}\beta_1, w_t^{2\gamma}\exp(\hat{h}_t^{*k_j}), \nu\right)},\ j = 1, ..., R\,,$$

to produce the desired filtered sample $\{h_t^1, ..., h_t^M\}$ from $(h_t|\mathcal{F}_t, \psi)$.

This particle filter is extremely simple to code. The only non-straightforward aspect is the sampling of indexes with replacement from populations with unequal probabilities. Methods for carrying this out are discussed in Pitt and Shephard (1997) and Carpenter, Clifford and Fearnhead (1998).

3.2 Diagnostics

Based on these M draws on h_{t+1} from the prediction density, one can estimate the conditional probability that the random Y_{t+1} will be less than the observed y_{t+1}, $\Pr(Y_{t+1} \leq y_{t+1}|\mathcal{F}_t)$ by

$$u_{t+1}^M = \frac{1}{M}\sum_{j=1}^{M}\Pr(Y_{t+1} \leq y_{t+1}|h_{t+1}^j, \psi)\,. \tag{9}$$

For each $t = 1, \ldots, n$, under the null of a correctly specified model u_t^M converges in distribution to independent and identically distributed uniform random variables as $M \to \infty$ (Rosenblatt (1952)). This provides a valid basis for diagnostic checking. These variables can be mapped into the normal distribution, by using the inverse of the normal distribution function $n_t^M = F^{-1}(u_t^M)$ to give a standard sequence of independent and identically distributed normal variables, which are then transformed one-step-ahead forecasts normed by their correct standard errors. These can be used to carry out formal Box-Ljung, normality, and heteroscedasticity tests, among others. Likewise diagnostic checks which focus on the modelling of the dispersion of returns can be obtained by working with the so-called reflected residuals $2\left|u_t^M - \frac{1}{2}\right|$ introduced by Kim, Shephard and Chib (1998). These random

variables also converge in distribution to independent and identically distributed uniform random variables as $M \to \infty$ and so again can be transformed via the inverse of the normal distribution function $r_t^M = F^{-1}(2\left|u_t^M - \frac{1}{2}\right|)$.

3.3 Marginal likelihood

The likelihood ordinate and the MCMC output can also be used to compute the model marginal likelihood. The marginal likelihood is defined as the integral of the likelihood function with respect to the prior density and is useful for comparing non-nested models. We describe a method for computing the marginal likelihood based on Chib (1995).

Let ψ denote the parameters of a given generalized stochastic volatility model with likelihood function $f(y|\psi)$ and prior density $\pi(\psi)$, where the likelihood function is computed using the particle filtering algorithm given above. Then, the marginal likelihood (with all normalizing constants included) can be written as

$$m(y) = \frac{f(y|\psi)\pi(\psi)}{\pi(\psi|y)},$$

an expression that follows from Bayes theorem. This expression is an identity in ψ and can be evaluated at any appropriately selected point ψ^* (say). If ψ^* denotes a high density point and $\hat{\pi}(\psi^*|y)$ the estimate of the posterior ordinate at ψ^*, then the marginal likelihood on the log scale is estimated as

$$\ln m(y) = \ln f(y|\psi^*) + \ln \pi(\psi^*) - \hat{\pi}(\psi^*|y) \tag{10}$$

The posterior density ordinate is estimated from the MCMC output by either kernel smoothing (if the dimension of ψ is small) or by a marginal/conditional decomposition of the posterior ordinate followed by reduced MCMC runs to generate the draws necessary to estimate each of the marginal/conditional ordinates (see Chib (1995) for further details).

4 Applications

In this section we consider applications of the proposed method to two datasets that have been extensively analyzed in the finance literature. The purpose of these examples is to illustrate the efficacy of the algorithm on two dimensions: the observed serial correlation in

the sampled output and the inefficiency factors in the estimation of the posterior mean. If we let G denote the Monte Carlo sample size, then the *inefficiency factor* is defined as

$$[1 + 2 \sum_{k=1}^{\infty} \rho(k)] ,$$

where $\rho(k)$ is the autocorrelation at lag k for the parameter of interest and the terms in the summation are cut off according to (say) the Parzen window. This is a useful quantity and may be interpreted as the ratio of the *numerical variance* of the posterior mean from the MCMC chain to the variance of the posterior mean from hypothetical independent draws. It serves to quantify the relative loss from using correlated draws, in comparison with hypothetical uncorrelated draws, for computing the posterior mean.

4.1 Stock Market Data

The data series comes from the Center for Research on Security Prices (CRSP) files and consists of daily continuously compounded returns, y_t, on the S&P 500 index (computed without considering dividends) from January 2, 1980 through December 29, 1987, for a total of 2022 observations. For ease of comparison with the existing papers, the sample period that we have selected is the same as in, among others, Gallant et al. (1992), Jacquier, Polson and Rossi (1994) and Sandmann and Koopman (1998). The model specification adopted here is, however, different from the latter papers and is given by

$$\begin{aligned} y_t &= x_{1t}'\beta_1 + \exp(h_t/2)u_t \\ h_t &= \mu + \beta_2 x_{2t} + \phi(h_{t-1} - \mu) + \sigma v_t \end{aligned}$$

where $x_{it} = (1, y_{t-1})$ and $\beta_1 = (a, b)$. Within this setup, four models are fit to the data:

- Model $\mathcal{M}_1$: Gaussian errors in the measurement equation and no covariates in the evolution equation;
- Model $\mathcal{M}_2$: Gaussian errors in the measurement equation and the lagged return, $x_{2t} = y_{t-1}$, in the evolution equation;
- Model $\mathcal{M}_3$: student t-errors with unknown degrees of freedom parameter ν in the measurement equation and no covariate in the evolution equation.

- Model $\mathcal{M}_4$: student t-errors with unknown degrees of freedom parameter ν in the measurement equation and the lagged return, $x_{2t} = y_{t-1}$, in the evolution equation;

4.1.1 Prior distributions

Before turning to the data analysis, we complete the model by specifying the parameters of the prior distributions given in Section 2.2. For ϕ we choose the values $\phi^{(1)} = 20$ and $\phi^{(2)} = 1.5$ which implies that our prior guess of ϕ is 0.86. This reflect the relatively high degree of persistence in volatility commonly found in high frequency financial series. In the prior of σ we let $\nu_0 = 2.25$ and $\delta_0 = 0.25$ which implies that the prior mean of σ is 0.20 and the prior standard deviation is 0.40. For μ, β_1 and β_2 we assume independent normal priors with means and standard deviations that are reported in Table 2. Finally, for the degrees of freedom ν we assume that the prior is uniform over the range $(2, 128)$. These are reasonable priors as they impose some structure but are not particularly informative.

4.1.2 Results

Table 2 reports posterior quantities and inefficiency factors for the most general of the four models, ie., $\mathcal{M}_4$ computed from 5000 iterations of the MCMC algorithm. The inefficiency factors are below ten for each parameter (except ν) which indicates that the sampler is mixing well. In addition, the drift parameter b is tightly estimated and indicates that there is some autocorrelation in the S&P 500 return, after controlling for stochastic volatility. The posterior means for the parameters μ, ϕ and σ are similar to the estimates one finds in models without covariates, for example, as reported in Sandmann and Koopman (1998). Furthermore, the distribution of β_2 is concentrated around -0.04 showing that past returns affect not only the level of current returns but also the current volatility of the stock index. Finally, the marginal posterior distribution of ν reveals that the Gaussian assumption is not appropriate for these data. The posterior density is centered around 9: this value is slightly higher than that reported by Sandmann and Koopman (1998) ($\nu = 7.6$) for the SV model without covariates.

Next, we check the adequacy of each of the four models by computing the autocorrelation plots of the n_t^M and r_t^M diagnostics (for convenience these are not reported). The absence of serial correlation in these diagnostics suggests that each model provides a reasonable

	Prior		Posterior				
	Mean	Std dev	Mean	Std dev	Lower	Upper	INEFF
a	0.000	0.400	0.035	0.019	-0.001	0.072	1.773
b	0.000	0.400	0.071	0.022	0.028	0.113	1.860
μ	-5.000	5.000	-0.349	0.161	-0.653	-0.003	5.862
ϕ	0.860	0.100	0.981	0.007	0.965	0.993	7.691
σ	0.200	0.400	0.102	0.018	0.069	0.138	9.702
β_2	0.000	0.400	-0.040	0.012	-0.065	-0.016	9.201
ν	65.00	36.37	8.973	2.178	5.953	14.023	16.89

Table 2: Summary output for the S&P500 data using the generalized SV model. In the table, "Lower" and "Upper" denote the 2.5th percentile and the 97.5th percentile, respectively, and INEFF denotes the inefficiency factor. The results are based on 5000 draws of which the first 1000 are discarded.

explanation for the serial dependence and the dispersion dependence in the data.

Next we compare the four models presented above by computing the model marginal likelihood in (10) and Bayes factors. The likelihood function which is in an input into the marginal likelihood computation is computed using the particle filter with parameters $M = 2000$ and $R = 10,000$. The results are summarized in Table 3. The Bayes factor indicate that there is decisive evidence for including y_{t-1} in the volatility equation. Similarly, the support for the student-t error distribution is found to be decisive. We conclude that the extensions from the simplest SV model in $\mathcal{M}_1$ to richer structures are necessary to adequately capture the volatility dynamics of the S&P500 return.

Model	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$\mathcal{M}_1$	-6.26	-9.91	-11.88
$\mathcal{M}_2$	–	-3.64	-5.61
$\mathcal{M}_3$	–	–	-1.96

Table 3: Models for the S&P500 index return. The entries in the table are Log (base 10) of Bayes factors for row model against column model (see text for a definition of the four models).

4.2 Interest Rate Data

We next analyze the interest rate data set considered, for example, by Gallant and Tauchen (1998). Bank discount rates on three-month Treasury Bills from January 13, 1962 to April 30, 1995 (1737 observations) are collected from the Federal Reserve Bank H.15 database. The raw data are transformed into continuously compounded yields, r_t, prior to the analysis. The response variable y_t is the change in the yield $(r_t - r_{t-1})$ and the model of interest is

$$\begin{aligned} y_t &= x_{1t}'\beta_1 + r_{t-1}^{\gamma} \exp(h_t/2)u_t \\ h_t &= \mu + \beta_2 x_{2t} + \phi(h_{t-1} - \mu) + \sigma\eta_t \end{aligned}$$

where $x_{it} = (1, r_{t-1})$, $\beta_1 = (a, b)$, x_{2t} is the spread between the yields on a 30-year Treasury Bond and a 1-year Treasury Bill and the u_t errors are independent student-t with v degrees of freedom. The term r_{t-1}^{γ} represents the level effect of the interest rate variable on its volatility. Note that in Gallant and Tauchen (1998) the parameter γ is set equal to one but is treated as an unknown parameter in the present analysis.

4.2.1 Prior distributions

As for the parameters $\beta_1, \mu, \beta_2, \phi, \sigma$ and ν, the prior distributions are the same as those used above for the S&P 500 dataset. For γ we assume a uniform distribution on the range $(0, 2)$ as this encompasses the typical values chosen for this parameter in the literature.

4.2.2 Results

The results on the fitted model, based on 5000 draws from the MCMC algorithm, are summarized in Table 4 and Figure 1. Once again we find from the estimated inefficiency factors that our algorithm delivers an extremely efficient sampler with the largest inefficiency factor being around ten. These are quite low given the complicated nature of the model. The parameter γ, which plays a large role in many interest rate models discussed in the finance literature, is quite poorly determined in this dataset. The posterior interval of γ, computed from the 2.5th and 97th percentile of the MCMC output, ranges from 0.3 to around 1.35 and the volatility dynamics are quite persistent as measured by the distribution of ϕ. Interestingly, the parameter ν is quite accurately estimated and its posterior mean of

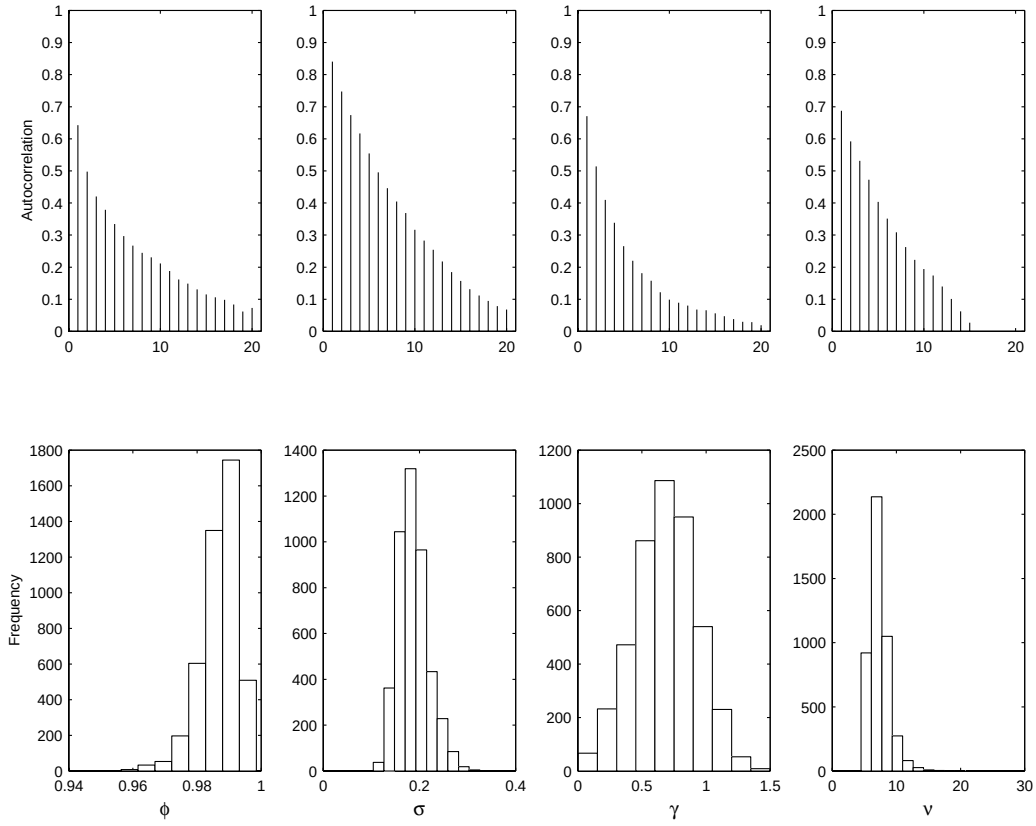

Figure 1: Autocorrelation functions and posterior histograms for ϕ, σ, γ and ν in the generalized SV model with independent student-t observation errors.

7.35 suggests that the data is non-Gaussian. The regression coefficient β_2 is estimated to be negative but quite small with its upper interval just including zero.

We next compare four different models on the basis of our diagnostic measures and our marginal likelihood/Bayes factor criterion. In addition to the general model given above (which we denote as $\mathcal{M}_1$) we define three other models in order to formally assess the statistical evidence for a simplification of the model. Model $\mathcal{M}_2$ is the same as model $\mathcal{M}_1$ except that the degrees of freedom is fixed at $\nu = 32$. Model $\mathcal{M}_3$ is the same as model $\mathcal{M}_1$ except that the parameter γ is set to one. Finally, model $\mathcal{M}_4$ is defined by replacing the t-assumption of $\mathcal{M}_1$ with the Gaussian and removing the covariate from the volatility equation. The residuals n_t^M and r_t^M which are defined above, and the likelihood function for each of these models, is computed using the particle filter that is run with $M = 2000$

	Prior		Posterior				
	Mean	Std dev	Mean	Std dev	Lower	Upper	INEFF
a	0.000	0.400	0.007	0.006	-0.004	0.019	1.877
b	0.000	0.400	-0.001	0.001	-0.004	0.002	2.285
μ	-5.000	5.000	-5.910	1.113	-8.028	-3.570	7.593
ϕ	0.860	0.100	0.987	0.006	0.973	0.996	9.791
σ	0.200	0.400	0.187	0.031	0.134	0.259	14.09
β_2	0.000	0.400	-0.006	0.003	-0.014	0.001	2.299
γ	1.000	0.333	0.684	0.242	0.192	1.142	8.004
ν	65.00	36.37	7.322	1.558	5.096	11.173	7.895

Table 4: Prior-posterior summary from the generalized SV model based on 5000 MCMC draws. "Lower" and "Upper" denote the 2.5th percentile and the 97.5th percentile, respectively, and INEFF denotes the inefficiency factor.

and $R = 10,000$.

The autocorrelation plots of the n_t^M and r_t^M (which are not reported) indicate no serial dependence in either residual for any of the models. This indicates that there is no evidence for any missing linear dependence or for any missing dispersion dependence. We next consider the QQ plots in Figures 2 and 3 of n_t^M and r_t^M, plotted against the quantiles of the uniform distribution, to check for the correctness of the model specification. A well fitted model should have residual quantiles that are equal to those of the uniform distribution. A visual analysis of the figures shows that both models $\mathcal{M}_1$ and $\mathcal{M}_4$ are suitable according to this criteria.

The model diagnostic given above provides useful information about model adequacy but do not help in the formal choice amongst these models. To do so we compute the marginal likelihoods and the associated Bayes factors on the log base 10 scale for the models under contention. The results appear in Table 5. This table clearly indicates that the data evidence for the general SV model $\mathcal{M}_1$ over the other models is decisive.

5 Concluding Remarks

This paper has considered a class of generalized stochastic volatility models defined by heavy tails, a level effect on the volatility and covariate effects in the observation and evolution equations. A simple, fast and highly efficient MCMC fitting algorithm has been developed

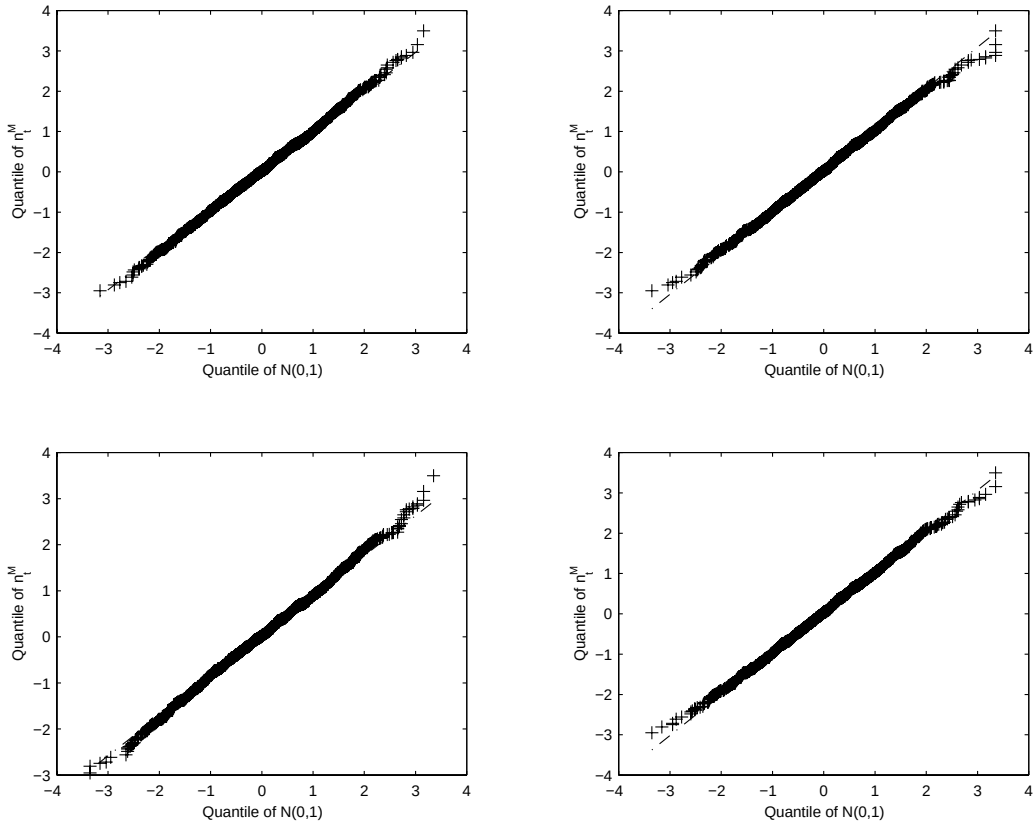

Figure 2: QQ plots of the $n_t^M = F^{-1}(u_t^M)$ across four models (defined in the text).

for such models. The discussion has also considered the construction of diagnostic measures and the estimation of the model marginal likelihood for comparing alternative stochastic volatility models. Taken together the framework and results will be important for the practical analysis of high frequency data.

The analysis can be extended in a number of directions. First, one can consider generalized SV models in which the parameters are allowed to switch amongst a given number of states according to a hidden Markov process. The basic SV model under this assumption has been considered recently by So, Lam and Li (1998). The MCMC implementation follows from the procedures developed in Albert and Chib (1993) and Chib (1996). Second, one can fit continuous time analogues of the model discussed in this paper [Andersen and Lund (1997), Gallant and Tauchen (1998) and Gallant, Hsu and Tauchen (1998)]. Such

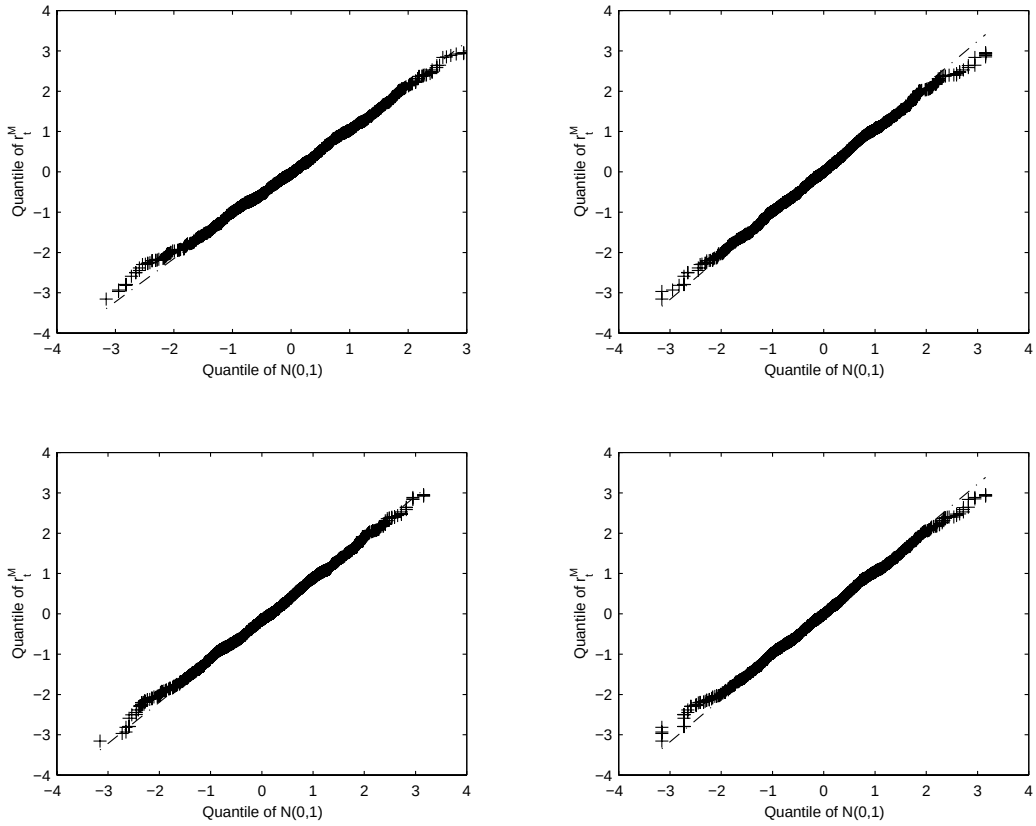

Figure 3: QQ plots of the inverse normally transformed reflected residuals $2\left|u_t^M - \frac{1}{2}\right|$ across four models (defined in the text).

extensions can be handled in the MCMC context by amalgamating the approach of this paper with that of Elerian, Chib and Shephard (1998). Another possible extension is to generalized models of multivariate stochastic volatility of the type recently investigated in detail by Pitt and Shephard (1998). This extension will be reported elsewhere.

Appendix

Algorithm

We include full details of the Metropolis-Hastings algorithm described in the text. This algorithm provides a mechanism for sampling the joint distribution

$$h, s, \lambda, \psi | y$$

Model	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$\mathcal{M}_1$	2.21	6.70	3.18
$\mathcal{M}_2$	–	4.49	0.97
$\mathcal{M}_3$	–	–	-3.52

Table 5: Log (base 10) of Bayes factors for row model against column model (see text for a definition of the four models).

by simulating various blocks of parameters in sequence.

1. Initialize s, λ, β_1, ν.

2. Sample θ and h from $\theta, h|y, s, \lambda, \beta_1$ by drawing

   (a) θ from $\theta|y, s, \lambda, \beta_1 \stackrel{d}{=} \beta_2|y^*, s$ which is proportional to

   $$g(\theta) = \pi(\theta) \prod_{t=1}^{n} f(y_t^*|y_{t-1}^*, s, \theta)\,,$$

   where

   $$\pi(\theta) = \pi(\phi|\phi^{(1)}, \phi^{(2)}) N(\mu|\mu_0, M_0) N(\beta_2|\beta_{20}, B_{20}^{-1}) IG(\sigma|\nu_0, \delta_0) I[0 < \gamma < 2]$$

   and the moments of the Gaussian conditional density $f(y_t^*|y_{t-1}^*, s, \theta)$ can be determined from the Kalman filter recursions. One computes for $t = 1, ..., n$

   $$e_t = y_t^* - \gamma \log(w_t^2) - a_t - m_{s_t}, \qquad d_t = p_t + v_{s_t}^2, \qquad k_t = \phi p_t / d_t, \tag{11}$$

   $$a_{t+1} = \mu + x_{2t}' \beta_2 + \phi\,(a_t - \mu) + k_t e_t, \qquad p_{t+1} = \sigma^2 + \phi p_t\,(\phi - k_t)\,. \tag{12}$$

   From these recursions,

   $$\log f(y_t^*|y_{t-1}^*, s, \theta) = const - \frac{1}{2} \sum_{t=1}^{n} \log d_t - \frac{1}{2} \sum_{t=1}^{n} \frac{e_t^2}{d_t}.$$

   To sample $g(\theta)$, let $q(\theta|m, V, \xi)$ denote a multivariate normal density with ξ degrees of freedom, mean vector m and scale matrix V defined as the mode and inverse of the negative Hessian, respectively, of $\log g(\theta)$. These are found by numerical optimization, typically initializing at the current value of θ and using the previous value of the Hessian. Then

i. Sample a proposal value θ' from the density $q(\theta|\mu, V, \xi)$

ii. Move to θ' given the current point θ with probability of move

$$\min\left\{\frac{\pi(\theta')\prod_{t=1}^{n} f(y_t^*|y_{t-1}^*, s, \theta')}{\pi(\theta)\prod_{t=1}^{n} f(y_t^*|y_{t-1}^*, s, \theta)}\frac{q(\theta|\mu, V, \xi)}{q(\theta'|\mu, V, \xi)}, 1\right\},$$

otherwise stay at θ.

(b) h from $h|y, s, \lambda, \psi \stackrel{d}{=} h|y^*, s, \theta$ in one block using the simulation smoother of de Jong and Shephard (1995). This involves running the Kalman filter (11) and (12) storing $\{e_t, d_t, k_t\}$ followed by backward recursions, defining $n_t = d_t^{-1} + k_t^2 u_t$ and $f_t = \frac{e_t}{d_t} + r_t k_t$, where going from $t = n, ..., 1$ with $r_n = 0$ and $u_n = 0$,

$$c_t = v_{s_t}^2 - v_{s_t}^4 n_t, \qquad \zeta_t \sim N(0, c_t), \qquad b_t = v_{s_t}^2\left(n_t - \phi k_t u_t\right),$$

$$r_{t-1} = \frac{e_t}{d_t} + (\phi - k_t)r_{t-1} - b_t\frac{\zeta_t}{c_t}, \qquad u_{t-1} = d_t^{-1} + (\phi - k_t)^2 u_t + \frac{b_t^2}{c_t}.$$

Then, $h_t = y_t^* - v_{s_t}^2 f_t - \zeta_t$.

3. Sample β_1 from $\beta_1|y, h, s, \lambda, \gamma$. Under the prior $\mathcal{N}(\beta_{10}, B_{10}^{-1})$, the draw is from the distribution $N(\hat{\beta}_1, B^{-1})$ where

$$B = B_{10} + \sum_{t=1}^{n}\frac{x_{1t}x_{1t}'}{\lambda_t^{-1}\exp(h_t)w_t^{2\gamma}}; \hat{\beta}_1 = B^{-1}\left(B_{10}\beta_{10} + \sum_{t=1}^{n}\frac{x_{1t}y_t}{\lambda_t^{-1}\exp(h_t)w_t^{2\gamma}}\right).$$

4. Sample s from $[s|y, h, \lambda, \psi] \stackrel{d}{=} [s|y^*, h, \psi]$, where $[s|y^*, h, \psi] = \prod_{t=1}^{n}\Pr(s_t|y_t^*, h_t, \psi)$ and

$$\Pr(s_t|y_t^*, h_t, \psi) \propto \Pr(s_t) f_N(y_t^*|\gamma\log(w_t^2) + h_t + \mu_{s_t}, v_{s_t}^2).$$

5. Sample ν and λ from $\nu, \lambda|y, h, \psi$ by drawing

(a) ν from $\nu|y, h, \psi$ which is proportional to

$$g(\nu) = \pi(\nu)\prod_{t=1}^{n} f_T\left(y_t|x_{1t}'\beta_1, w_t^{2\gamma}\exp(h_t), \nu\right)$$

This density is not in a known family but it can be sampled by finding a proposal density that is tailored to the target $g(\nu)$ and applying the Metropolis-Hastings algorithm in a manner analogous to the case of θ above.

(b) λ_t independently from $\lambda_t|y_t, h_t, \psi$ $(t \leq n)$ where

$$\lambda_t|y_t, h_t, \psi \sim \text{Gamma}\left(\frac{v+1}{2}, \frac{\nu + (y_t - x_{1t}'\beta_1)^2 / \left(w_t^{2\gamma} \exp(h_t)\right)}{2}\right),$$

6. Repeat Steps 2-6 using the most recent values of the conditioning variables.

A complete loop through steps 2-6 completes one sweep or cycle of the MCMC iterations. These sweeps are fast. On a 400 Mhz Pentium machine with a data set of 1800 observations, 5000 cycles take about twenty minutes for the most general interest rate model investigated in this study. Given the low inefficiency factors of our algorithm, 2000 sweeps may be adequate for typical datasets.

References

Albert, J. and S. Chib (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts, *Journal of Business and Economic Statistics,* 11, 1-15.

Andersen, T. G and J. Lund (1997). Estimating continuous-time stochastic volatility models of the short term interest rate, *Journal of Econometrics*, 343–77.

Berzuini, C., N. G. Best, W. R. Gilks and C. Larizza (1997). Dynamic conditional independence models and Markov chain Monte Carlo methods, *Journal of the American Statistical Association*, 92, 1403-12.

Carpenter, J. R., P. Clifford, P. Fearnhead (1998). An improved particle filter for non-linear problems, Working paper, January 1998, Dept. of Statistics, University of Oxford, OX1 3TG, UK.

Carter, C.K. and R. Kohn (1994). On Gibbs sampling for state space models, *Biometrika,* 81, 541-553.

Chib, S. (1995). Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, 90, 1313–21.

Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models, *Journal of Econometrics*, 75, 79-97.

Chib, S. and E. Greenberg (1994). Bayes inference for regression models with ARMA(p,q) errors, *Journal of Econometrics*, 64, 183–206.

Chib, S. and E. Greenberg (1995). "Understanding the Metropolis-Hastings algorithm, *American Statistician*, 49, 327–35.

de Jong, P. and N. Shephard (1995). The Simulation Smoother for time series models, *Biometrika*, 82, 339-350

Durbin, J. and S.J. Koopman (1997) Monte Carlo maximum likelihood estimation of Gaussian state space model, *Biometrika*, 84, 669-84.

Elerian, O., S. Chib and N. Shephard (1998). Likelihood inference for discretely observed non-linear diffusions, manuscript, Olin School of Business, Washington University, St. Louis.

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

Gallant, A.R., D. Hsieh. and G. Tauchen (1992). Stock prices and volume, *Review of Financial Studies*, 5, 199–242.

Gallant, A.R. and G. Tauchen (1998). Reprojection partially observed systems with applications to interest rate diffusions, *Journal of the American Statistical Association*, 93, 10–24.

Gallant, A. R., C. Hsu and G. Tauchen (1998). Calibrating volatility diffusions and extracting integrated volatility, manuscript.

Ghysels, E., A.C. Harvey and E. Renault (1996). Stochastic volatility, in *Statistical Methods in Finance* (eds. C.R. Rao and G S Maddala, 119-191, Amsterdam: North-Holland.

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

Gordon, N. J., D J. Salmond and A F M Smith (1993). A novel approach to non-linear and Gaussian Bayesian state estimation, *IEE-Proceedings F*, 140, 107–13.

Isard, M. and A. Blake (1996). Contour tracking by stochastic propagation of conditional density, *Proceedings of the European Conference on Computer Vision, Cambridge*, 1, 343-56.

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

Kim, S., N. Shephard and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies*, 65, 361–93.

Kitagawa, G (1996). Monte Carlo filter and smoother for Gaussian nonlinear state space models, *J. Computational and Graphical Statistics*, 5, 1–25.

Marriott, J. M. and A. F. M. Smith (1992). Reparameterization aspects of numerical Bayesian methodology for autoregressivce moving-average models, *Journal of Time Series Analysis*, 13, 327-43.

Nardari F. (1998). Comparing Stochastic Volatility Models of the Short Term Interest Rate: A Bayesian Approach, manuscript, Washington University, St. Louis.

Pitt, M. and N. Shephard (1997). Filtering via simulation based on auxiliary particle filters, manuscript, Nuffield College, Oxford.

Pitt, M. and N. Shephard (1998). Time varying covariances: a factor stochastic volatility approach (with discussion), in *Bayesian Statistics 6* (eds J. Bernardo, J.O. Berger, A. P. Dawid and A. F. M. Smith), Oxford: Oxford University Press, in press.

Rosenblatt, M. (1952). Remarks on a multivariate transformation, *Annals of Mathematical Statistics*, 23, 470-2.

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

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In *Time Series Models with Econometric, Finance and Other Applications*, (eds D.R.Cox, D.V.Hinkley and O.E.Barndorff-Nielson), 1-67, London: Chapman and Hall.

Shephard, N. and M.K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series, *Biometrika*, 84, 653-67.

So, M.K.P., K. Lam and W.K. Lee (1998). A stochastic volatility model with Markov switching, *Journal of Business and Economic Statistics*, 16, 244-253.

Taylor, S. J. (1994). Modelling stochastic volatility, *Mathematical Finance*, 4, 183-204.