Review of Economic Studies (1998) 65, 361–393
0034-6527/98/00170361$02.00

Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models

SANGJOON KIM
Salomon Brothers Asia Limited

NEIL SHEPHARD
Nuffield College, Oxford

and

SIDDHARTHA CHIB
Washington University, St. Louis

First version received December 1994; *final version accepted August* 1997 (*Eds.*)

In this paper, Markov chain Monte Carlo sampling methods are exploited to provide a unified, practical likelihood-based framework for the analysis of stochastic volatility models. A highly effective method is developed that samples all the unobserved volatilities at once using an approximating offset mixture model, followed by an importance reweighting procedure. This approach is compared with several alternative methods using real data. The paper also develops simulation-based methods for filtering, likelihood evaluation and model failure diagnostics. The issue of model choice using non-nested likelihood ratios and Bayes factors is also investigated. These methods are used to compare the fit of stochastic volatility and GARCH models. All the procedures are illustrated in detail.

1. INTRODUCTION

The variance of returns on assets tends to change over time. One way of modelling this feature of the data is to let the conditional variance be a function of the squares of previous observations and past variances. This leads to the autoregressive conditional heteroscedasticity (ARCH) based models developed by Engle (1982) and surveyed in Bollerslev, Engle and Nelson (1994).

An alternative to the ARCH framework is a model in which the variance is specified to follow some latent stochastic process. Such models, referred to as stochastic volatility (SV) models, appear in the theoretical finance literature on option pricing (see, for example, Hull and White (1987) in their work generalizing the Black–Scholes option pricing formula to allow for stochastic volatility). Empirical versions of the SV model are typically formulated in discrete time. The canonical model in this class for regularly spaced data is

$$\begin{aligned} y_t &= \beta e^{h_t/2}\varepsilon_t, && t \geqq 1, \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \\ h_1 &\sim \mathcal{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right), \end{aligned} \tag{1}$$

where y_t is the mean corrected return on holding the asset at time t, h_t is the log volatility at time t which is assumed to follow a stationary process ($|\phi| < 1$) with h_1 drawn from the stationary distribution, ε_t and η_t are uncorrelated standard normal white noise shocks and $\mathcal{N}(\cdot, \cdot)$ is the normal distribution. The parameter β or $\exp(\mu/2)$ plays the role of the constant scaling factor and can be thought of as the modal instantaneous volatility, ϕ as the persistence in the volatility, and σ_η the volatility of the log-volatility. For indentifiability reasons either β must be set to one or μ to zero. We show later that the parameterization with β equal to one in preferable and so we shall leave μ unrestricted when we estimate the model but report results for $\beta = \exp(\mu/2)$ as this parameter has more economic interpretation.

This model has been used as an approximation to the stochastic volatility diffusion by Hull and White (1987) and Chesney and Scott (1989). Its basic econometric properties are discussed in Taylor (1986), the review papers by Taylor (1994), Shephard (1996) and Ghysels, Harvey and Renault (1996) and the paper by Jacquier, Polson and Rossi (1994). These papers also review the existing literature on the estimation of SV models.

In this paper we make advances in a number of different directions and provide the first complete Markov chain Monte Carlo simulation-based analysis of the SV model (1) that covers efficient methods for Bayesian inference, likelihood evaluation, computation of filtered volatility estimates, diagnostics for model failure, and computation of statistics for comparing non-nested volatility models. Our study reports on several interesting findings. We consider a very simple Bayesian method for estimating the SV model (based on one-at-a-time updating of the volatilities). This sampler is shown to be quite inefficient from a simulation perspective. An improved (multi-move) method that relies on an offset mixture of normals approximation to a log-chi-square distribution coupled with an importance reweighting procedure is shown to be strikingly more effective. Additional refinements of the latter method are developed to reduce the number of blocks in the Markov chain sampling. We report on useful plots and diagnostics for detecting model failure in a dynamic (filtering) context. The paper also develops formal tools for comparing the basic SV and Gaussian and t-GARCH models. We find that the simple SV model typically fits the data as well as more heavily parameterized GARCH models. Finally, we consider a number of extensions of the SV model that can be fitted using our methodology.

The outline of this paper is as follows. Section 2 contains preliminaries. Section 3 details the new algorithms for fitting the SV model. Section 4 contains methods for simulation-based filtering, diagnostics and likelihood evaluations. The issue of comparing the SV and GARCH models is considered in Section 5. Section 6 provides extensions while Section 7 concludes. A description of software for fitting these models that is available through the internet is provided in Section 8. Two algorithms used in the paper are provided in the Appendix.

2. PRELIMINARIES

2.1. *Quasi-likelihood method*

A key feature of the basic SV model in (1) is that it can be transformed into a linear model by taking the logarithm of the squares of observations

$$\log y_t^2 = h_t + \log \varepsilon_t^2, \tag{2}$$

where $E(\log \varepsilon_t^2) = -1{\cdot}2704$ and $\text{Var}(\log \varepsilon_t^2) = 4{\cdot}93$. Harvey, Ruiz, and Shephard (1994) have employed Kalman filtering to estimate the parameters $\theta = (\phi, \sigma_\eta^2, \mu) \in (-1, 1) \times \Re_+ \times \Re$ by maximizing the quasi likelihood

$$\log L_Q(y \mid \theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{n} \log F_t - \frac{1}{2}\sum_{t=1}^{n} v_t^2 / F_t,$$

where $y = (y_1, \ldots, y_n)$, v_t is the one-step-ahead prediction error for the best linear estimator of $\log y_t^2$ and F_t is the corresponding mean square error.[1] It turns out that this quasi-likelihood estimator is consistent and asymptotically normally distributed but is sub-optimal in finite samples because $\log \varepsilon_t^2$ is poorly approximated by the normal distribution, as shown in Figure 1. As a consequence, the quasi-likelihood estimator under the assumption that $\log \varepsilon_t^2$ is normal has poor small sample properties, even though the usual quasi-likelihood asymptotic theory is correct.

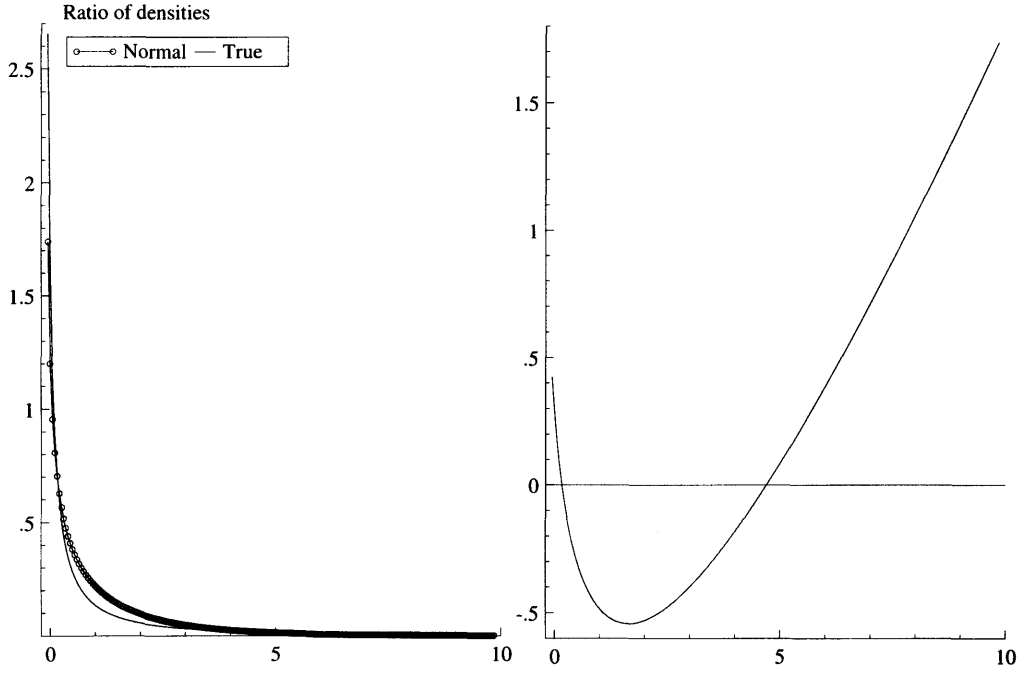

FIGURE 1

Log-normal approximation to χ_1^2 density. Left is the χ_1^2 density and the log-normal approximation which is used in the quasi-likelihood approach. Right is the log of the ratio of the χ_1^2 density to the approximation

2.2. *Markov chain Monte Carlo*

An alternative, exact approach to inference in the SV model is based on Markov chain Monte Carlo (MCMC) methods, namely the Metropolis–Hastings and Gibbs sampling algorithms. These methods have had a widespread influence on the theory and practice of Bayesian inference. Early work on these methods appears in Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953), Hastings (1970), Ripley (1977) and Geman and

1. The Kalman filter algorithms for computing v_t and F_t are given in the Appendix.

Geman (1984) while some of the more recent developments, spurred by Tanner and Wong (1987) and Gelfand and Smith (1990), are included in Chib and Greenberg (1996), Gilks, Richardson and Spiegelhalter (1996) and Tanner (1996, Ch. 6). Chib and Greenberg (1995) provide a detailed exposition of the Metropolis–Hastings algorithm and include a derivation of the algorithm from the logic of reversibility.

The idea behind MCMC methods is to produce variates from a given multivariate density (the posterior density in Bayesian applications) by repeatedly sampling a Markov chain whose invariant distribution is the target density of interest. There are typically many different ways of constructing a Markov chain with this property and one goal of this paper is to isolate those that are simulation-efficient in the context of SV models. In our problem, one key issue is that the likelihood function $f(y|\theta)=\int f(y|h,\theta)f(h|\theta)dh$ is intractable. This precludes the direct analysis of the posterior density $\pi(\theta|y)$ by MCMC methods. This problem can be overcome by focusing instead on the density $\pi(\theta,h|y)$, where $h=(h_1,\ldots,h_n)$ is the vector of n latent volatilities. Markov chain Monte Carlo procedures can be developed to sample this density without computation of the likelihood function $f(y|\theta)$. It should be kept in mind that sample variates from a MCMC algorithm are a high-dimensional (correlated) sample from the target density of interest. These draws can be used as the basis for making inferences by appealing to suitable ergodic theorems for Markov chains. For example, posterior moments and marginal densities can be estimated (simulation consistently) by averaging the relevant function of interest over the sampled variates. The posterior mean of θ is simply estimated by the sample mean of the simulated θ values. These estimates can be made arbitrarily accurate by increasing the simulation sample size. The accuracy of the resulting estimates (the so called numerical standard error) can be assessed by standard time series methods that correct for the serial correlation in the draws. The serial correlation can be quite high for badly behaved algorithms.

2.2.1. *An initial Gibbs sampling algorithm for the* SV *model*

For the problem of simulating a multivariate density $\pi(\psi|y)$, the Gibbs sampler is defined by a blocking scheme $\psi=(\psi_1,\ldots,\psi_d)$ and the associated full conditional distributions $\psi_i|\,\psi_{\setminus i}$, where $\psi_{\setminus i}$ denotes ψ excluding the block ψ_i. The algorithm proceeds by sampling each block from the full conditional distributions where the most recent values of the conditioning blocks are used in the simulation. One cycle of the algorithm is called a sweep or a scan. Under regularity conditions, as the sampler is repeatedly swept, the draws from the sampler converge to draws from the target density at a geometric rate. For the SV model the ψ vector becomes (h,θ). To sample ψ from the posterior density, one possibility (suggested by Jacquier, Polson and Rossi (1994) and Shephard (1993)) is to update each of the elements of the ψ vector one at a time.

1. Initialize h and θ.
2. Sample h_t from $h_t|\,h_{\setminus t},y,\theta,t=1,\ldots,n$.
3. Sample $\sigma_\eta^2|\,y,h,\phi,\mu,\beta$.
4. Sample $\phi|\,h,\mu,\beta,\sigma_\eta^2$.
5. Sample $\mu|\,h,\phi,\sigma_\eta^2$.
6. Goto 2.

Cycling through 2 to 5 is a complete sweep of this (single move) sampler. The Gibbs sampler will require us to perform many thousands of sweeps to generate samples from $\theta,h|\,y$.

The most difficult part of this sampler is to effectively sample from $h_t|h_{\setminus t}, y_t, \theta$ as this operation has to be carried out n times for each sweep. However,

$$f(h_t|h_{\setminus t}, \theta, y) \propto f(h_t|h_{\setminus t}, \theta) f(y_t|h_t, \theta), \qquad t=1, \ldots, n.$$

We sample this density by developing a simple accept/reject procedure.[2] Let $f_N(t|a, b)$ denote the normal density function with mean a and variance b. It can be shown (ignoring end conditions to save space) that

$$f(h_t|h_{\setminus t}, \theta) = f(h_t|h_{t-1}, h_{t+1}, \theta) = f_N(h_t|h_t^*, v^2),$$

where

$$h_t^* = \mu + \frac{\phi\{(h_{t-1}-\mu)+(h_{t+1}-\mu)\}}{(1+\phi^2)}$$

and

$$v^2 = \frac{\sigma_\eta^2}{1+\phi^2}.$$

Next we note that $\exp(-h_t)$ is a convex function and can be bounded by a function linear in h_t. Let $\log f(y_t|h_t, \theta) = \text{const} + \log f^*(y_t, h_t, \theta)$. Then

$$\begin{aligned}\log f^*(y_t, h_t, \theta) &= -\frac{1}{2}h_t - \frac{y_t^2}{2}\{\exp(-h_t)\} \\ &\leqq -\frac{1}{2}h_t - \frac{y_t^2}{2}\{\exp(-h_t^*)(1+h_t^*) - h_t \exp(-h_t^*)\} \\ &= \log g^*(y_t, h_t, \theta, h_t^*).\end{aligned}$$

Hence

$$f(h_t|h_{\setminus t}, \theta) f^*(y_t, h_t, \theta) \leqq f_N(h_t|h_t^*, v^2) g^*(y_t, h_t, \theta, h_t^*).$$

The terms on the right-hand side can be combined and shown to be proportional to $f_N(h_t|\mu_t, v^2)$ where

$$\mu_t = h_t^* + \frac{v^2}{2}[y_t^2 \exp(-h_t^*) - 1]. \tag{3}$$

2. Other MCMC algorithms for simulating from $h_t|h_{t-1}, h_{t+1}, y_t; \theta$ have been given in the literature by Shephard (1993), Jacquier, Polson and Rossi (1994), Shephard and Kim (1994), Geweke (1994) and Shephard and Pitt (1997). The closest to our suggestion is Geweke (1994) who also bounded $\log f^*$, but by $-0{\cdot}5h_t$. This suffers from the property of having a high rejection rate for slightly unusual observations (for example, 0·9 for $|y_t|/\beta \exp(h_t/2) > 3$). Shephard and Pitt (1997), on the other hand, used a quadratic expansion of $\log f^*$ about h_t^*. This increases the generality of the procedure but it involves a Metropolis rejection step and so is more involved. Shephard (1993) approximated f^* by a normal distribution with the same moments as $\log \varepsilon_t^2$.

Geweke (1994) and Shephard and Kim (1994) independently suggested the use of the Gilks and Wild (1992) procedure for sampling from log concave densities such as $\log f(h_t|h_{\setminus t}, \theta, y)$. This is generalizable to non-log-concave densities using the Gilks, Best and Tan (1995) sampler. Typically these routines need about 10 to 12 evaluations of $\log f(h_t|h_{\setminus t}, \theta, y)$ to draw a single random variable. Hence they are about 10 times less efficient than the simple accept/reject algorithm given above.

Jacquier, Polson and Rossi (1994)'s Metropolis algorithm uses a very different approach. They approximate the density of $h_t|h_{\setminus t}$ and so use a non-Gaussian proposal based on f^*. Typically this procedure is considerably slower than the use of the Gilks and Wild (1992) methods suggested above.

With these results, the accept–reject procedure (Ripley (1987)) to sample h_t from $f(h_t|h_{\setminus t}, \theta, y)$ can now be implemented. First, propose a value h_t from $f_N(h_t|\mu_t, v^2)$. Second, accept this value with probability f^*/g^*; if rejected return to the first step and make a new proposal.[3]

Sampling σ_η^2 and ϕ. Sampling the σ_η^2 and ϕ one at a time is straightforward. If we assume a conjugate prior $\sigma_\eta^2|\phi, \mu \sim \mathcal{IG}(\sigma_r/2, S_\sigma/2)$, then σ_η^2 is sampled from

$$\sigma_\eta^2|y, h, \phi, \mu \sim \mathcal{IG}\left\{\frac{n+\sigma_r}{2}, \frac{S_\sigma+(h_1-\mu)^2(1-\phi^2)+\sum_{t=1}^{n-1}((h_{t+1}-\mu)-\phi(h_t-\mu))^2}{2}\right\}, \quad (4)$$

where $\mathcal{IG}$ denotes the inverse-gamma distribution. Throughout we set $\sigma_r=5$ and $S_\sigma=0{\cdot}01\times\sigma_r$.

For ϕ, sampling from the full conditional density is also easy. Let $\phi=2\phi^*-1$ where ϕ^* is distributed as Beta with parameters $(\phi^{(1)}, \phi^{(2)})$. Hence, our prior on ϕ is

$$\pi(\phi)\propto\left\{\frac{(1+\phi)}{2}\right\}^{\phi^{(1)}-1}\left\{\frac{(1-\phi)}{2}\right\}^{\phi^{(2)}-1}, \qquad \phi^{(1)}, \phi^{(2)}>\frac{1}{2}, \quad (5)$$

and has support on the interval $(-1, 1)$ with a prior mean of $\{2\phi^{(1)}/(\phi^{(1)}+\phi^{(2)})-1\}$. In our work we will select $\phi^{(1)}=20$ and $\phi^{(2)}=1{\cdot}5$, implying a prior mean of 0·86. Alternative priors could also be used. For example, the flat prior $\pi(\phi)\propto 1$ is attractive in that it leads to an analytically tractable full conditional density. But this prior can cause problems when the data are close to being non-stationary (Phillips (1991) and Schotman and Van Dijk (1991)). Chib and Greenberg (1994) and Marriott and Smith (1992) discuss other priors (restricted to the stationary region) for autoregressive models. We feel that it is important from a data-analytic view to impose stationarity in the SV model. Further, if $\phi=1$ then the μ terms cancel in (1) and so μ becomes unidentified from the data. The prior we select avoids these two problems rather well.

Under the specified prior, the full conditional density of ϕ is proportional to

$$\pi(\phi)f(h|\mu, \phi, \sigma_\eta^2),$$

where

$$\log f(h|\mu, \phi, \sigma_\eta^2)\propto -\frac{(h_1-\mu)^2(1-\phi^2)}{2\sigma_\eta^2}+\frac{1}{2}\log(1-\phi^2) - \frac{\sum_{t=1}^{n-1}\{(h_{t+1}-\mu)-\phi(h_t-\mu)\}^2}{2\sigma_\eta^2}, \quad (6)$$

This function is concave in ϕ for all values of $\phi^{(1)}$, $\phi^{(2)}$. This means that ϕ can be sampled using an acceptance algorithm. Employ a first order Taylor expansion of the prior about

$$\hat{\phi}=\sum_{t=1}^{n-1}(h_{t+1}-\mu)(h_t-\mu)/\sum_{t=1}^{n-1}(h_t-\mu)^2,$$

and combine with $f(h|\mu, \phi, \sigma^2)$. The resulting density provides a good suggestion density. Alternatively, one can specialize the method of Chib and Greenberg (1994) (which is based

3. This proposal has an average acceptance rate of approximately $1-y_t^2\exp(-h_t^*)v_t^2/(4\beta^2)$. A typical situation is where $v_t^2=0{\cdot}01$. Usually $y_t^2\exp(-h_t^*)v_t^2/\beta^2$ will not be very large as h_t^* is the smoothed log-volatility of y_t and so reflects the variation in y_t. An extreme case is where $y_t^2\exp(-h_t^*)\sigma_t^2/\beta^2=100$, which leads to an average acceptance rate of approximately 0·75. In our experience an average acceptance rate of over 0·995 seems usual for real financial datasets.

on the Metropolis–Hastings algorithm). Given the current value $\phi^{(i-1)}$ at the $(i-1)$-st iteration, sample a proposal value ϕ^* from $N(\hat{\phi}, V_\phi)$ where $V_\phi = \sigma_\eta^2 \{\sum_{t=1}^{n-1} (h_t - \mu)^2\}^{-1}$. Then, provided ϕ^* is in the stationary region, accept this proposal value as $\phi^{(i)}$ with probability $\exp\{g(\phi^*) - g(\phi^{(i-1)})\}$ where

$$g(\phi) = \log \pi(\phi) - \frac{(h_1 - \mu)^2(1-\phi^2)}{2\sigma_\eta^2} + \frac{1}{2}\log(1-\phi^2).$$

If the proposal value is rejected, set $\phi^{(i)}$ to equal $\phi^{(i-1)}$. Both these approaches can be used with alternative priors on ϕ.

Sampling μ. Suppose we work with a diffuse prior[4] on μ, then μ is sampled from the full conditional density

$$\mu \mid h, \phi, \sigma_\eta^2 \sim \mathcal{N}(\hat{\mu}, \sigma_\mu^2), \tag{7}$$

where

$$\hat{\mu} = \sigma_\mu^2 \left\{ \frac{(1-\phi^2)}{\sigma_\eta^2} h_1 + \frac{(1-\phi)}{\sigma_\eta^2} \sum_{t=1}^{n-1} (h_{t+1} - \phi h_t) \right\},$$

and

$$\sigma_\mu^2 = \sigma_\eta^2 \{(n-1)(1-\phi)^2 + (1-\phi^2)\}^{-1}.$$

In our work we sample μ and record the value $\beta = \exp(\mu/2)$.

Illustration. To illustrate this algorithm we analyse the daily observations of weekday close exchange rates for the U.K. Sterling/U.S. Dollar exchange rate from 1/10/81 to 28/6/85. The sample size is $n = 946$. Later in the paper we will also use the corresponding series for the German Deutschemark (DM), Japanese Yen and Swiss Franc (SwizF), all against the U.S. Dollar. This data set has been previously analysed using quasi-likelihood methods in Harvey, Ruiz and Shephard (1994). The mean-corrected returns will be computed as

$$y_t = 100 \times \left\{ (\log r_t - \log r_{t-1}) - \frac{1}{n} \sum_{i=1}^{n} (\log r_i - \log r_{i-1}) \right\}, \tag{8}$$

where r_t denotes the exchange rate at time t. The MCMC sampler was initialized by setting all the $h_t = 0$ and $\phi = 0{\cdot}95$, $\sigma_\eta^2 = 0{\cdot}02$ and $\mu = 0$. We iterated the algorithm on the log-volatilities for 1000 iterations and then the parameters and log-volatilities for 50,000 more iterations, before recording the draws from a subsequent 1,000,000 sweeps. The burn-in period is thus much larger than what is customary in the literature and is intended to ensure that the effect of the starting values becomes insignificant. As a result, there is likely to be no additional information from running multiple chains from dispersed starting

4. Occasionally, for technical reasons, we take a slightly informative prior such as $\mu \sim N(0, 10)$. In this paper, this prior was used for the computation of Bayes factors.

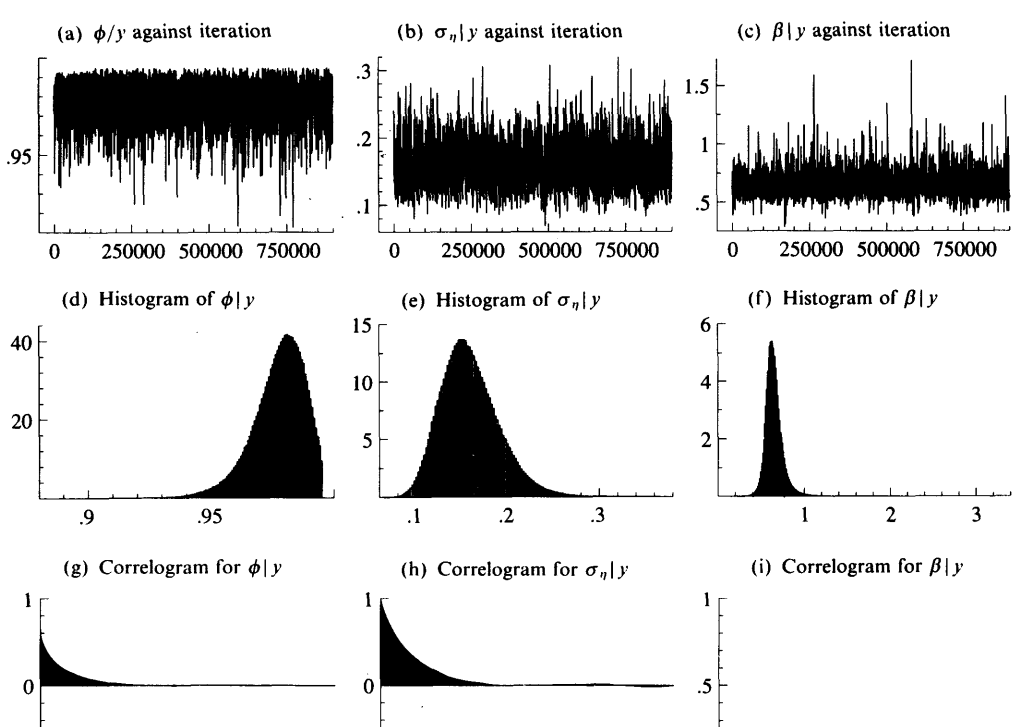

FIGURE 2

Single move Gibbs sampler for the Sterling series. Graphs (a)–(c): simulations against iteration. Graphs (d)–(f): histograms of marginal distribution. Graphs (g)–(i): corresponding correlograms for simulation. In total 1,000,000 iterations were drawn, discarding the first 50,000

values. The complete 1,000,000 iterations[5] are graphed in Figure 2 and summarized in Table 1.[6]

The summary statistics of Table 1 report the *simulation inefficiency* factors of the sampler. These are estimated as the variance of the sample mean from the MCMC sampling scheme (the square of the numerical standard error) divided by the variance of the sample mean from a hypothetical sampler which draws independent random variables from the posterior (the variance divided by the number of iterations). We think that the simulation inefficiency statistic is a useful diagnostic (but by no means the only one) for measuring how well the chain mixes. The numerical standard error of the sample mean is estimated

5. We have employed a 32 bit version of the modified Park and Miller (1988) uniform random number as the basis of all our random numbers. This has a period of $2^{31}-1$, which allows us to draw around 2.1 billion random numbers. In these experiments we are drawing approximately $n \times 2 \times 1{\cdot}05$ random numbers per sweep of the sampler, where 5% is a very conservative estimate of the overall rejection rate. For this dataset this is 1984 draws per sweep. Given that we employ 1,000,000 sweeps, we are close, but not beyond, the period of our random number generator.

6. Timings will be given for all the computations given in this paper. These are made using the authors C^{++} code which has been linked to Ox. The single move algorithm is optimized to this special case and so is about as fast as it is possible to make it. The latter algorithms are much more general and so it is not completely fair to compare the computed time reported here to their times.

TABLE 1.

Daily returns for Sterling: summaries of Figure 2. The Monte Carlo S.E. of simulation is computed using a bandwidth, B_M, of 2000, 4000 and 2000 respectively. Italics are correlations rather than covariances of the posterior. Computer time is seconds on a Pentium Pro/200. The other time is the number of seconds to perform 100 sweeps of the sampler

	Mean	MC S.E.	Inefficiency	Covariance and correlation		
$\phi \mid y$	0·97762	0·00013754	163·55	0·00011062	*−0·684*	*0·203*
$\sigma_\eta \mid y$	0·15820	0·00063273	386·80	−0·00022570	0·00098303	*−0·129*
$\beta \mid y$	0·64884	0·00036464	12·764	0·00021196	−0·00040183	0·0098569
Time	3829·5	0·58295				

by time series methods (to account for the serial correlation in the draws) as

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

where $\hat{\rho}(i)$ is an estimate of the autocorrelation at lag i of the MCMC sampler, B_M represents the bandwidth and K the Parzen kernel (see, for example, Priestley (1981, Ch. 6)) given by

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

The correlogram (autocorrelation function) indicates important autocorrelations for ϕ and σ_η at large lag lengths. If we require the Monte Carlo error in estimating the mean of the posterior to be no more than one percentage of the variation of the error due to the data, then this Gibb sampler would have to be run for around 40,000 iterations. This seems a reasonably typical result: see Table 2.

TABLE 2

Bandwidth B_M was 2000, 4000 and 2000, respectively for the parameters, for all series. In all cases 1,000,000 sweeps were used

	$\phi \mid y$		$\sigma_\eta \mid y$		$\beta \mid y$	
Series	Mean	Inefficiency	Mean	Inefficiency	Mean	Inefficiency
DM	0·96496	122·77	1·15906	292·81	0·65041	15·762
Yen	0·98010	313·03	0·12412	676·35	0·53597	14·192
SwizF	0·95294	145·48	0·20728	231·15	0·70693	13·700

Parameterization. An alternative to this sampler is to replace the draw for $\mu \mid h, \phi, \sigma_\eta^2$ with that resulting from the alternative parameterization $\beta \mid y, h$. Such a move would be a mistake. Table 3 reports the inefficiency factor for this sampler using 1,000,000

TABLE 3

Bandwidth B_M was 4000, 4000 and 15,000, respectively for the parameters. 1,000,000 sweeps were used

	$\phi \mid y$		$\sigma_\eta \mid y$		$\beta \mid y$	
Series	Mean	Inefficiency	Mean	Inefficiency	Mean	Inefficiency
Sterling	0·97793	465·30	0·15744	439·73	0·64280	5079·6

draws of this sampler. There is a small deterioration in the sampler for $\phi | y$ and a very significant reduction in efficiency for $\beta | y$. The theoretical explanation for the inadequacies of the β parameterization is provided by Pitt and Shephard (1998).

Reason for slow convergence. The intuition for the slow convergence reported in Table 1 is that the components of $h | y, \theta$ are highly correlated and in such cases sampling each component from the full conditional distribution produces little movement in the draws, and hence slowly decaying autocorrelations (Chib and Greenberg (1996)). For analytical results, one can think of the Gaussian equivalent of this problem. Under the Gaussian assumption and the linear approximation (2) and (1), the sampler in the simulation of h from $h | y, \theta$ has an analytic convergence rate of (Pitt and Shephard (1998, Theorem 1))

$$4\phi^2/\{1+\phi^2+\sigma_\eta^2/\text{Var}\,(\log \varepsilon_t^2)\}^2,$$

where θ is taken as fixed at the expected values given in the results for the Sterling series. If Var $(\log \varepsilon_t^2)$ is set equal to 4·93, then this result implies a geometric convergence rate of $\rho_A = 0{\cdot}9943$ and an inefficiency factor of $(1+\rho_A)/(1-\rho_A) = 350$ which is in the range reported in Table 1.

In order to improve the above sampler it is necessary to try to sample the log-volatilities in a different way. One method is to sample groups of consecutive log volatilities using a Metropolis algorithm. This is investigated in Shephard and Pitt (1997). In this paper we detail a more ambitious model specific approach. This approach is described next.

3. IMPROVED MCMC ALGORITHMS

In this section we design an offset mixture of normals distribution (defined below) to accurately approximate the exact likelihood. This approximation helps in the production of an efficient (adapted Gibbs sampler) Monte Carlo procedure that allows us to sample all the log-volatilities at once. We then show how one can make the analysis exact by correcting for the (minor) approximation error by reweighting the posterior output.

3.1. *Offset mixture representation*

Our approximating parametric model for the linear approximation (2) will be an offset mixture time series model

$$y_t^* = h_t + z_t, \tag{9}$$

where $y_t^* = \log (y_t^2 + c)$ and

$$f(z_t) = \sum_{i=1}^{K} q_i f_N(z_t | m_i - 1{\cdot}2704, v_i^2),$$

is a mixture of K normal densities f_N with component probabilities q_i, means $m_i - 1{\cdot}2704$, and variances v_i^2. The constants $\{q_i, m_i, v_i^2\}$ are selected to closely approximate the exact density of $\log \varepsilon_t^2$. The "offset" c was introduced into the SV literature by Fuller (1996, pp. 494–497) in order to robustify the QML estimator of the SV model to y_t^2 being very small. Throughout we will set $c = 0.001$ (although it is possible to let c depend on the actual value taken by y_t^2). It should be noted that the mixture density can also be written

in terms of a component indicator variable s_t such that

$$z_t | s_t = i \sim \mathcal{N}(m_i - 1{\cdot}2704, v_i^2), \qquad \Pr(s_t = i) = q_i. \tag{10}$$

This representation will be used below in the MCMC formulation.

We are now in a position to select K and $\{m_i, q_i, v_i^2\}$ $(i \leqq K)$ to make the mixture approximation "sufficiently good". In our work, following for instance Titterington, Smith, and Makov (1985, p. 133), we matched the first four moments of $f_{\exp(z)}(r)$ (the implied log-normal distribution) and $f(z_t)$ to those of a χ_1^2 and $\log \chi_1^2$ random variable respectively, and required that the approximating densities lie within a small distance of the true density. This was carried out by using a non-linear least squares program to move the weights, means and variances around until the answers were satisfactory. It is worth noting that this nonlinear optimization incurs only a one-time cost, as there are no model-dependent parameters involved. We found what we judged to be satisfactory answers by setting $K=$ 7. The implied weights, means and variances are given in Table 4, while the approximating and the true density are drawn in Figure 3. It would be easy to improve the fit by increasing the value of K, however further experiments that we have conducted suggest that increasing K has little discernible effect on our main results.

TABLE 4

Selection of the mixing distribution to be $\log \chi_1^2$

ω	$\Pr(\omega = i)$	m_i	σ_i^2
1	0·00730	−10·12999	5·79596
2	0·10556	−3·97281	2·61369
3	0·00002	−8·56686	5·17950
4	0·04395	2·77786	0·16735
5	0·34001	0·61942	0·64009
6	0·24566	1·79518	0·34023
7	0·25750	−1·08819	1·26261

3.2. *Mixture simulator*

In the MCMC context, mixture models are best estimated by exploiting the representation in (10). The general algorithm for state space models was suggested independently by Shephard (1994) and Carter and Kohn (1994). The posterior density of interest is $\pi(s, h, \phi, \sigma_\eta^2, \mu | y^*)$, where $s = (s_1, \ldots, s_n)$. In this case, both h and s can be sampled separately in one block and the sampler takes the form:

1. Initialize s, ϕ, σ_η^2 and μ.
2. Sample h from $h | y^*, s, \phi, \sigma_\eta^2, \mu$.
3. Sample s from $s | y^*, h$.
4. Update ϕ, σ_η^2, μ according to (6), (4) and (7).
5. Goto 2.

Note that we are using $y^* = \{\log(y_1^2 + c), \ldots, \log(y_T^2 + c)\}$ in the conditioning set above as a *pointer* to the mixture model. The vectors y^* and y, of course, contain the same information.

The important improvement over the methods in Section 2 is that it is now possible to efficiently sample from the highly multivariate Gaussian distribution $h | y^*, s, \phi, \sigma_\eta, \mu$ because $y^* | s, \phi, \sigma_\eta, \mu$ is a Gaussian time series which can be placed into the state-space

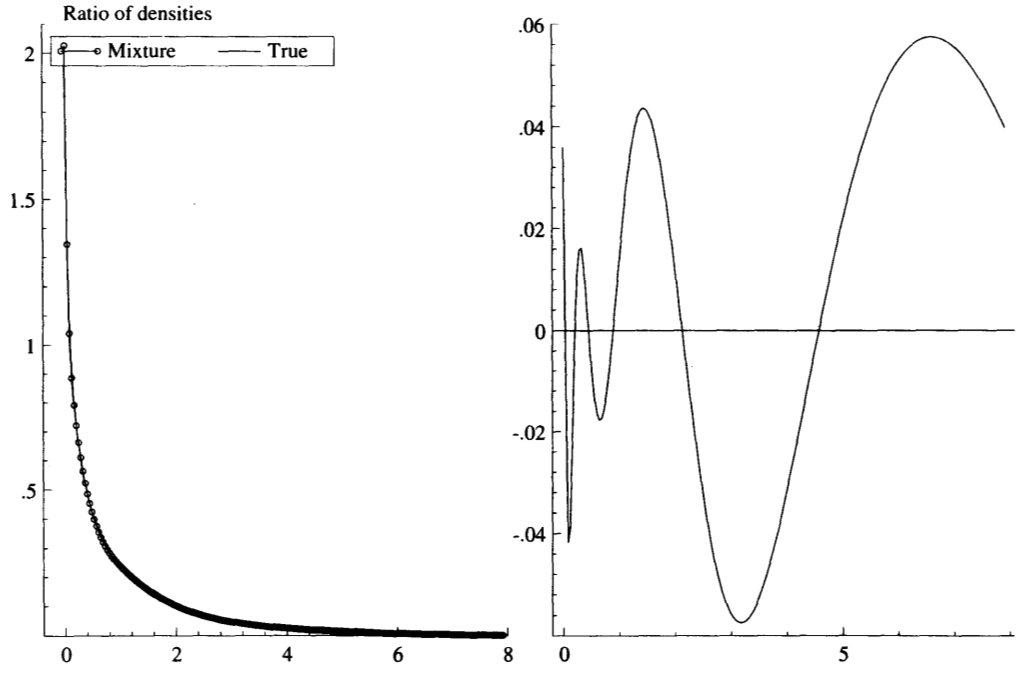

FIGURE 3

Mixture approximation to χ^2_1 density. Left: χ^2_1 density and mixture approximation. Right: the log of the ratio of the χ^2_1 density to the mixture approximation

form associated with the Kalman filter. The time series literature calls such models partially non-Gaussian or conditionally Gaussian. This particular model structure means we can sample from the entire $h \mid y^*, s, \phi, \sigma_\eta, \mu$ using the Gaussian simulation signal smoother detailed in the Appendix. As for the sampling of s from $s \mid y^*, h$, this is done by independently sampling each s_t using the probability mass function

$$\Pr(s_t = i \mid y_t^*, h_t) \propto q_i f_N(y_t^* \mid h_t + m_i - 1{\cdot}2704, v_i^2), \qquad i \geqq K.$$

The results from 750,000 sweeps of this mixture sampler are given in Table 5 and Figure 4. This sampler has less correlation than the single move sampler and suggests that generating 20,000 simulations from this sampler would probably be sufficient for inferential purposes.

TABLE 5

Daily returns for Sterling against Dollar. Summaries of Figure 2. The Monte Carlo S.E. of simulation is computed using a bandwidth, B_M, of 2000, 2000 and 100 respectively. Italics are correlations rather than covariances of the posterior. Computer time is seconds on a Pentium Pro/200. The other time is the number of seconds to perform 100 complete passes of the sampler

	Mean	MC S.E.	Inefficiency	Covariance and correlation		
$\phi \mid y$	0·97779	6·6811e−005	29·776	0·00011093	*−0·690*	*0·203*
$\sigma_\eta \mid y$	0·15850	0·00046128	155·42	−0·00023141	0·0010131	*−0·127*
$\beta \mid y$	0·64733	0·00024217	4·3264	0·00021441	−0·00040659	0·010031
Time	15,374	2·0498				

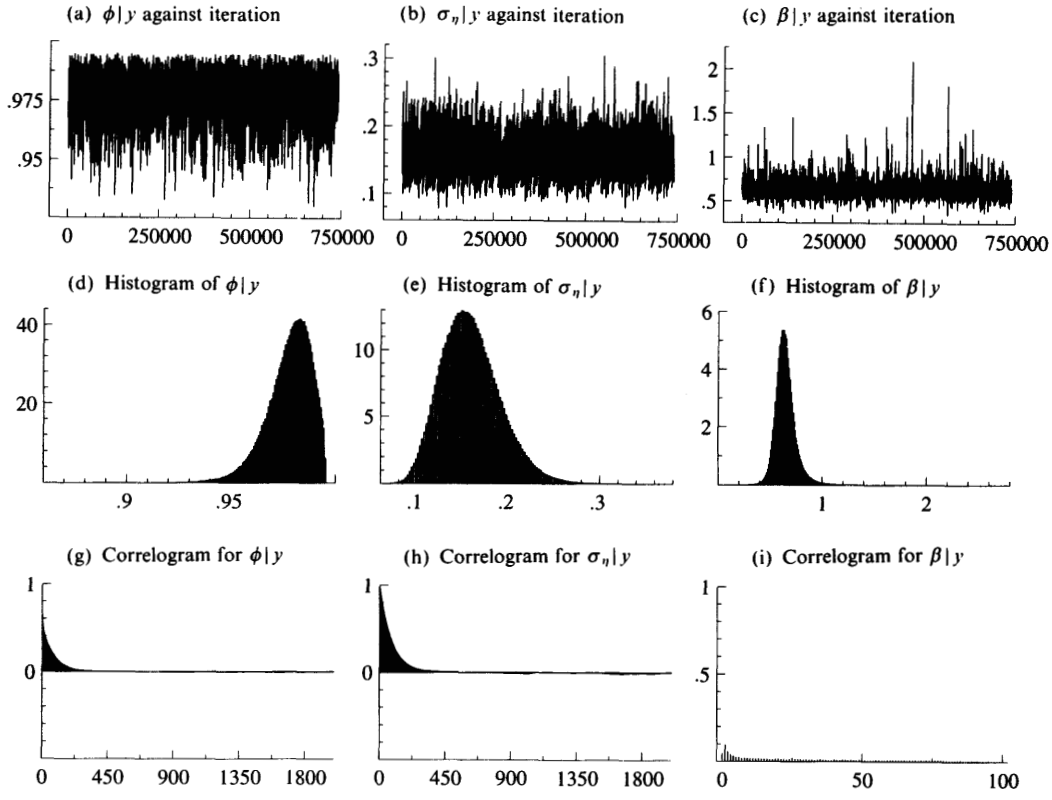

FIGURE 4

Mixture sampler for Sterling series. Graphs (a)–(c): simulations against iteration. Graphs (d)–(f): histograms of marginal distribution. Graphs (g)–(i): corresponding correlograms for simulation. In total 750,000 iterations were drawn, discarding the first 10,000

3.3. *Integrating out the log-volatilities*

Although this mixture sampler improves the correlation behaviour of the simulations, the gain is not very big as there is a great deal of correlation between the volatilities and parameters. However, we can use the Gaussian structure of $y^*|s, \phi, \sigma_\eta^2$ to overcome this. We can sample the joint distribution $\pi(\phi, \sigma_\eta^2, h, \mu|y^*, s)$ by sampling (ϕ, σ_η^2) from $\pi(\phi, \sigma_\eta^2|y^*, s) \propto f(y^*|s, \phi, \sigma_\eta^2)\pi(\phi, \sigma_\eta^2)$, and then sampling (h, μ) from $\pi(h, \mu|y^*, s, \phi, \sigma_\eta^2)$ We are able to sample the former distribution because the density $f(y^*|s, \phi, \sigma_\eta^2)$ can be evaluated using an augmented version of the Kalman filter (analytically integrating out μ and h).[7] Then, writing $\mu|y^*, s, \phi, \sigma_\eta^2 \sim \mathcal{N}(\tilde{\mu}, \sigma_{\tilde{\mu}}^2)$ we have that

$$\pi(\phi, \sigma_\eta^2|y^*, s) \propto \pi(\phi)\pi(\sigma_\eta^2) f(y^*|s, \phi, \sigma_\eta^2) = \pi(\phi)\pi(\sigma_\eta^2)\frac{f(y^*|s, \phi, \sigma_\eta^2, \mu=0)\pi(\mu=0)}{\pi(\mu=0|y^*, s, \phi, \sigma_\eta^2)}$$

$$\propto \pi(\phi)\pi(\sigma_\eta^2) \prod_{t=1}^n F_t^{-1/2} \exp\left(-\frac{1}{2}\sum_{t=1}^n v_t^2/F_t\right) \exp\left(\frac{1}{2\sigma_{\tilde{\mu}}^2}\tilde{\mu}^2\right)\sigma_{\tilde{\mu}},$$

where v_t is the one-step-ahead prediction error for the best mean square estimator of y_t^*, and F_t is the corresponding mean square error. The quantities $v_t, F_t, \tilde{\mu}, \sigma_{\tilde{\mu}}^2$ are

7. Augmented Kalman filters and simulation smoothers are discussed in the Appendix.

computed from the augmented Kalman filter provided in the Appendix, conditional on s.

This implies that we can sample from $\phi, \sigma_\eta^2 | y^*, s$ directly by making the proposal $\{\phi^{(i)}, \sigma_\eta^{2(i)}\}$, given the current value $\{\phi^{(i-1)}, \sigma_\eta^{2(i-1)}\}$, by drawing from some density $g(\phi, \sigma_\eta^2)$ and then accepting them using the Metropolis–Hastings probability of move

$$\min\left\{1, \frac{\pi(\phi^{(i)}, \sigma_\eta^{2(i)} | y^*, s)}{\pi(\phi^{(i-1)}, \sigma_\eta^{2(i-1)} | y^*, s)} \frac{g((\phi^{(i-1)}, \sigma_\eta^{2(i-1)})}{g(\phi^{(i)}, \sigma_\eta^{2(i)})}\right\}. \tag{11}$$

If the proposal value is rejected, we then set $\{\phi^{(i)}, \sigma_\eta^{2(i)}\} = \{\phi^{(i-1)}, \sigma_\eta^{2(i-1)}\}$. We call this an "integration sampler" as it integrates out the log-volatilities.

The structure of the integration sampler is then generically:

1. Initialize $(s, \phi, \sigma_\eta, \mu)$.
2. Sample (ϕ, σ_η^2) from $\pi(\phi, \sigma_\eta^2 | y^*, s)$ using a Metropolis–Hastings suggestion based on $g(\sigma_\eta^2, \phi)$, accepting with probability (11).
3. Sample $h, \mu | y^*, s, \phi, \sigma_\eta^2$ using the augmented simulation smoother given in the Appendix.
4. Sample $s | y^*, h$ as in the previous algorithm.
5. Goto 2.

An important characteristic of this sampler is that the simulation smoother can jointly draw h and μ. The scheme allows a free choice of the proposal density $g(\phi, \sigma_\eta^2)$. We have employed a composite method which first draws 200 samples (discarding the first ten samples) from the posterior density $\pi(\phi, \sigma_\eta^2 | y)$ using a Metropolis–Hastings sampler based on Gilks, Best and Tan (1995) which only requires the coding of the function $y^* | s, \phi, \sigma_\eta^2$ and the prior. These 200 draws are used to estimate the posterior mean and covariance. The mean and twice the covariance are then used to form a Gaussian proposal density $g(\phi, \sigma_\eta^2)$ for the Metropolis–Hastings algorithm in (11). As an alternative, one could also use a multivariate Student t proposal distribution instead of the Gaussian. See Chib and Greenberg (1995) for further discussion on the issues involved in choosing a proposal density for the Metropolis–Hastings algorithm.

The output from the resulting sampler is reported in Figure 5 and Table 6. These suggest that 2000 samples from this generator would be sufficient for this problem. This result seems reasonably robust to the data set.

3.4. *Reweighting*

The approach based on our (very accurate) offset mixture approximation provides a neat connection to conditionally Gaussian state space models and leads to elegant and efficient sampling procedures, as shown above. We now show that it is possible to correct for the minor approximation error by appending a straightforward reweighting step at the conclusion of the above procedures. This step then provides a sample from the exact posterior density of the parameters and volatilities. The principle we describe is quite general and may be used in other simulation problems as well.

First write the mixture approximation as making draws from $k(\theta, h | y^*)$, and then define

$$w(\theta, h) = \log f(\theta, h | y) - \log k(\theta, h | y) = \text{const} + \log f(y | h) - \log k(y^* | h),$$

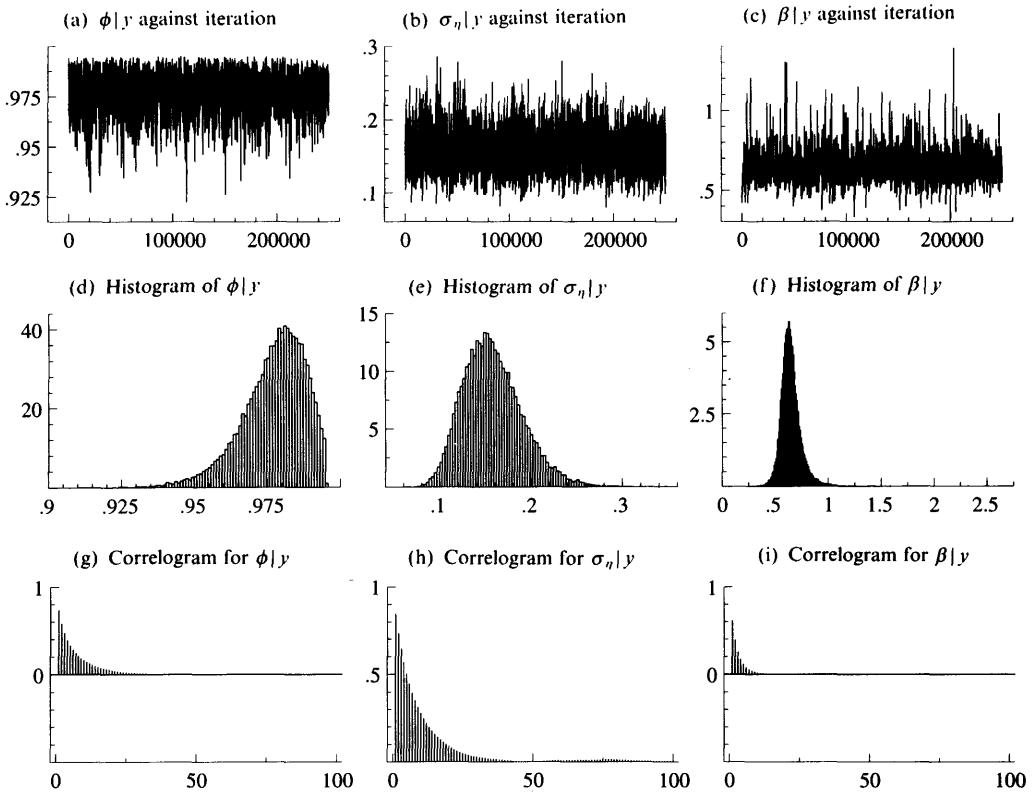

FIGURE 5

The integration sampler for Sterling series. Graphs (a)–(c): simulations against iteration. Graphs (d)–(f): histograms of marginal distribution. Graphs (g)–(i): corresponding correlograms for simulation. In total 250,000 iterations were drawn, discarding the first 250

TABLE 6

Daily returns for Sterling against Dollar. Summaries of Figure 5. The Monte Carlo S.E. of simulation is computed using a bandwidth, B_M, of 100, 100 *and* 100 *respectively. Italics are correlations rather than covariances of the posterior. Computer time is seconds on a Pentium Pro/200. The other time is the number of seconds to perform* 100 *complete passes of the sampler*

	Mean	MC S.E.	Inefficiency	Covariance and correlation		
$\phi\|y$	0·97780	6·7031e−005	9·9396	0·00011297	*−0·699*	*0·205*
$\sigma_\eta\|y$	0·15832	0·00025965	16·160	−0·00023990	0·0010426	*−0·131*
$\beta\|y$	0·64767	0·00023753	1·4072	0·00021840	−0·00042465	0·010020
Time	8635·2	3·4541				

where

$$f(y|h)=\prod_{t=1}^{n} f_N\{y_t|0,\exp(h_t)\},$$

and

$$k(y^*|h)=\prod_{t=1}^{n}\sum_{i=1}^{K} q_i f_N(y_t^*|h_t+m_i-1{\cdot}2704,v_i^2).$$

Both these functions involve Gaussian densities and are straightforward to evaluate for any value of h. Then,

$$\mathrm{E}g(\theta)\,|\,y = \int g(\theta) f(\theta\,|\,y) d\theta$$

$$= \int g(\theta) \exp\{w(\theta, h)\} k(\theta, h\,|\,y^*) d\theta dh \Big/ \int \exp\{w(\theta, h)\} k(\theta, h\,|\,y^*) d\theta dh.$$

Thus we can estimate functionals of the posterior by reweighting the MCMC draws according to

$$\mathrm{E}g(\hat{\theta})|\,y = \sum_j g(\theta^j) c^j,$$

where the weights are

$$c^j = \exp\{w(\theta^j, h^j)\} / \sum_i \exp\{w(\theta^i, h^i)\}. \tag{12}$$

As the mixture approximation is very good, we would expect that the weights c^j would have a small variance.

To see the dispersion of the weights, we recorded the weights from the sampler which generated Figure 5 and plotted the resulting log-weights in Figure 6. The log-weights are close to being normally distributed with a standard deviation of around one.

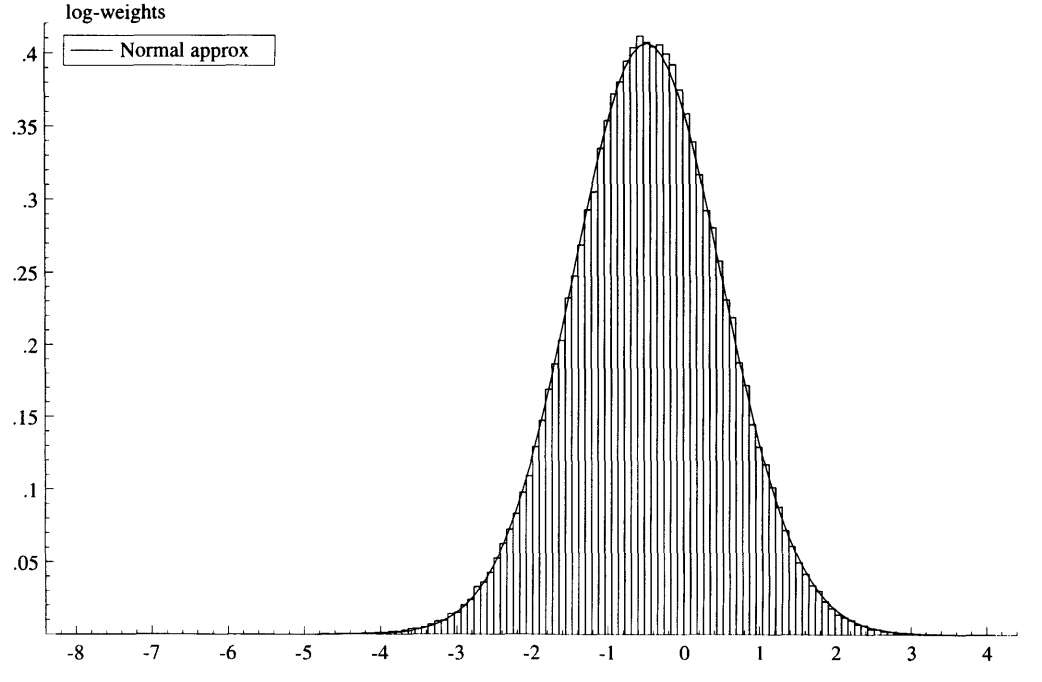

FIGURE 6

Histogram of the log of the $M \times c^j$ for 250,000 sweeps for the integration sampler and a corresponding approximating normal density with fitted mean and standard deviation. All the weights around zero would indicate a perfect sampler

TABLE 7

Daily returns for Sterling against Dollar. Summaries of reweighted sample of 250,000 *sweeps of the integration sampler. The Monte Carlo S.E. of simulation is computed using a block one tenth of the size of the simulation. Italics are correlations rather than covariances of the posterior. Computer time is seconds on a Pentium Pro/200. The other time is the number of seconds to perform* 100 *complete passes of the sampler*

	Mean	MC S.E.	Inefficiency	Covariance and correlation		
$\phi \mid y$	0·97752	7·0324e−005	11·20	0·00010973	*−0·685*	*0·204*
$\sigma_\eta \mid y$	0·15815	0·00024573	14·81	−0·00022232	0·00096037	*−0·129*
$\beta \mid y$	0·64909	0·00025713	1·64	0·00021181	−0·00039768	0·0098312
Time	10,105	4·0423				

To see the effect of the weights on the parameters estimates, we reweighted the 250,000 samples displayed in Figure 5. This produced the estimates which are given in Table 7. These Monte Carlo estimates of the posterior means are statistically insignificantly different from Monte Carlo estimated values given in Table 1. However, the Monte Carlo precision has improved dramatically. Further, the Monte Carlo standard errors indicate that this data set could be routinely analysed using around 1500 sweeps.

This conclusion seems to hold up for some other exchange rate series. Table 8 reports the estimates of the parameters and simulation inefficiency measures for the DM, Yen and Swiss Franc series. This table is the exact analog of Table 2 for the single move algorithm.

4. FILTERING, DIAGNOSTICS AND LIKELIHOOD EVALUATION

4.1. *Introduction*

There has been considerable recent work on the development of simulation based methods to perform filtering, that is computing features of $h_t \mid Y_t, \theta$, for each value of $Y_t = (y_1, \ldots, y_t)$. Leading papers in this field include Gordon, Salmond and Smith (1993), Kitagawa (1996), Isard and Blake (1996), Berzuini, Best, Gilks and Larizza (1997), West (1993) and Muller (1991). We work with a simple approach which is a special case of a suggestion made by Pitt and Shephard (1997). Throughout we will assume θ is known. In practice θ will be set to some estimated value, such as the maximum likelihood estimator or the Monte Carlo estimator of the posterior mean.

The objective is to obtain a sample of draws from $h_t \mid Y_t, \theta$ given a sample of draws $h_{t-1}^1, \ldots, h_{t-1}^M$ from $h_{t-1} \mid Y_{t-1}, \theta$. Such an algorithm is called a particle filter in the literature. We now show how this may be done. From Bayes theorem,

$$f(h_t \mid Y_t, \theta) \propto f(y_t \mid h_t, \theta) f(h_t \mid Y_{t-1}, \theta), \tag{13}$$

TABLE 8

Bandwidth, B_M, *for each parameter was* 100 *on all series. In all cases* 250,000 *sweepes were used*

	$\phi \mid y$		$\sigma_\eta \mid y$		$\beta \mid y$	
Series	Mean	Inefficiency	Mean	Inefficiency	Mean	Inefficiency
DM	0·96529	8·31	0·15812	11·99	0·65071	9·73
Yen	0·97998	23·10	0·12503	35·66	0·53534	2·71
SwizF	0·95276	13·52	0·20738	15·33	0·70675	8·38

where

$$f(h_t|Y_{t-1},\theta)=\int f(h_t|h_{t-1},\theta)f(h_{t-1}|Y_{t-1},\theta)dh_{t-1},$$

and $f(h_t|h_{t-1},\theta)=f_N(h_t|\mu+\phi(h_{t-1}-\mu),\sigma_\eta^2)$ is the normal evolution density. The latter integral can be estimated from the sample $h_{t-1}^1,\ldots,h_{t-1}^M$ leading to the approximations

$$f(h_t|Y_{t-1},\theta)\simeq\frac{1}{M}\sum_{j=1}^M f(h_t|h_{t-1}^j,\theta),$$

and

$$f(h_t|Y_t,\theta)\propto f(y_t|h_t,\theta)\frac{1}{M}\sum_{j=1}^M f(h_t|h_{t-1}^j,\theta). \tag{14}$$

The question now is to sample h_t from the latter density. The obvious importance sampling procedure of producing a sample $\{h_t^j\}$ from $f(h_t|h_{t-1}^j,\theta)$ and then resampling these draws with weights proportional to $\{f(y_t|h_t^j,\theta)\}$ is not efficient. An improved procedure runs as follows. Let $h_{t|t-1}=\mu+\phi(M^{-1}\sum h_{t-1}^j-\mu)$ and $\log f(y_t|h_t,\theta)=$ const $+\log f^*(y_t,h_t,\theta)$. Now expand $\log f^*(y_t,h_t,\theta)$ in a Taylor series around the point $h_{t|t-1}$ as

$$\begin{aligned}\log f^*(y_t,h_t,\theta)&=-\frac{1}{2}h_t-\frac{y_t^2}{2}\{\exp(-h_t)\}\\&\leqq-\frac{1}{2}h_t-\frac{y_t^2}{2}\{\exp(-h_{t|t-1})(1+h_{t|t-1})-h_t\exp(-h_{t|t-1})\}\\&=\log g^*(h_t,h_{t|t-1},\theta).\end{aligned}$$

Also, after some algebra it can be shown that

$$g^*(h_t,h_{t|t-1},\theta)\,f(h_t|h_{t-1}^j,\theta)\propto\pi_j f_N(h_t|h_{t|t-1}^j,\sigma_\eta^2), \tag{15}$$

where

$$\pi_j=\exp\left\langle-\frac{1}{2\sigma_\eta^2}[\{\mu+\phi(h_{t-1}^j-\mu)\}^2-h_{t|t-1}^{j2}]\right\rangle$$

and

$$h_{t|t-1}^j=\mu+\phi(h_{t-1}^j-\mu)+\frac{\sigma_\eta^2}{2}\{y_t^2\exp(-h_{t|t-1})-1\}.$$

Hence, the kernel of the target density in (14) can be bounded as

$$f^*(y_t,h_t,\theta)\frac{1}{M}\sum_{j=1}^M f(h_t|h_{t-1}^j,\theta)\leqq g^*(h_t,h_{t|t-1},\theta)\frac{1}{M}\sum_{j=1}^M f(h_t|h_{t-1}^j,\theta),$$

where the right-hand side terms are proportional to $1/M\sum_{j=1}^M\pi_j f_N(h_t|h_{t|t-1}^j,\sigma_\eta^2)$ due to (15).

These results suggest a simple accept–reject procedure for drawing h_t. First, we draw a proposal value h_t from the mixture density $\sum_{j=1}^M\pi_j^*f_N(h_t|h_{t|t-1}^j,\sigma_\eta^2)$, where $\pi_j^*=\pi_j/\sum_j\pi_j$. Second, we accept this value with probability $f^*(y_t,h_t,\theta)/g^*(h_t,h_{t|t-1},\theta)$. If the value is rejected, we return to the first step and draw a new proposal.

TABLE 9

Daily returns for Sterling series. Summaries of reweighted sample of 2500 *sweeps of the integration sampler. The Monte Carlo S.E. of simulation is computed using a block one tenth of the size of the simulation. Italics are correlations rather than covariances of the posterior. Computer time is seconds on a Pentium Pro/*200*. The other time is the number of seconds to perform* 100 *complete passes of the sampler*

	Mean	MC S.E.	Inefficiency	Covariance and correlation		
$\phi \mid y$	0·97611	0·0018015	11·636	0·00014783	*−0·765*	*0·277*
$\sigma_\eta \mid y$	0·16571	0·0065029	17·657	−0·00033148	0·0012693	*−0·232*
$\beta \mid y$	0·64979	0·0047495	1·4563	0·00030503	−0·00074971	0·008209
Time	97·230	3·8892				

By selecting a large M this filtering sampler will become arbitrarily accurate.

4.1.1. *Application*

To illustrate this, we apply these methods to the Sterling/Dollar series, filtering the volatility. Throughout we will employ $M = 2500$. Similar results were obtained when M fell to 1000, although reducing M below that figure created important biases. The results are made conditional of the estimated parameters, which are taken from Table 9 and based on 2500 sweeps of the integration sampler.

The resulting filtered and smoothed estimates of the volatility are given in Figure 7, together with a graph of the absolute values of the returns. The graph shows the expected feature of the filtered volatility lagging the smoothed volatility. Throughout the sample,

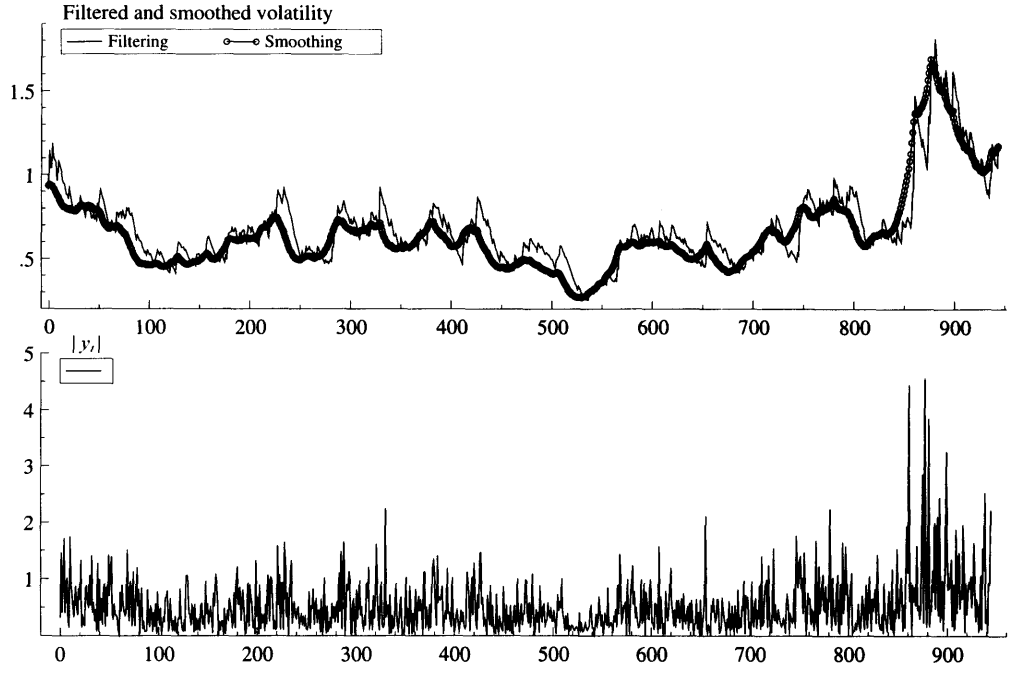

FIGURE 7

Top: filtered and smoothed estimate of the volatility $\exp(h_t/2)$, computed using $M = 2000$. Bottom: $|y_t|$, the absolute values of the returns

the filtered volatility is slightly higher than the smoothed values due to the gradual fall in volatility observed for these series during this period.

4.2. *Diagnostics*

Having designed a filtering algorithm it is a simple matter to sample from the one-step-ahead prediction density and distribution function. By definition the prediction density is

$$f(y_{t+1}|Y_t, \theta) = \int f(y_{t+1}|Y_t, h_{t+1}, \theta) f(h_{t+1}|Y_t, h_t, \theta) f(h_t|Y_t, \theta) dh_{t+1} dh_t,$$

which can be sampled by the method of composition as follows. For each value h_t^j $(j=1, 2, \ldots, M)$ from the filtering algorithm, one samples h_{t+1}^j from

$$h_{t+1}^j | h_t^j \sim \mathcal{N}\{\mu + \phi(h_t^j - \mu), \sigma_\eta^2\}.$$

Based on these M draws on h_{t+1} from the prediction density, we can estimate the probability that y_{t+1}^2 will be less than the observed y_{t+1}^{o2}

$$\Pr(y_{t+1}^2 \leqq y_{t+1}^{o2} | Y_t, \theta) \cong u_{t+1}^M = \frac{1}{M} \sum_{j=1}^M \Pr(y_{t+1}^2 \leqq y_{t+1}^{o2} | h_{t+1}^j, \theta). \tag{16}$$

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 cary out Box–Ljung, normality, and heteroscedasticity tests, among others.

The computed forecast uniforms and resulting correlograms and *QQ* plots are given in Figure 8. The results suggest that the model performs quite well, although it reveals some outliers. However, closer inspection shows that the outliers correspond to small values of y_t^2. This suggests that the SV model fails to accommodate some of the data values that have limited daily movements. On the other hand it appears to perform well when the movements in the data are large. This will be made more formal in the next subsection.

4.2.1. *Likelihood estimation*

The one-step-ahead predictions can also be used to estimate the likelihood function since the one-step-ahead prediction density, $f(y_{t+1}|Y_t)$, can be estimated as

$$\frac{1}{M} \sum_{j=1}^M f(y_{t+1}|h_{t+1}^j), \qquad h_{t+1}^j | h_t^j \sim \mathcal{N}\{\mu + \phi(h_t^j - \mu), \sigma_\eta^2\}, \tag{17}$$

using drawings from the filtering simulator. The same argument gives a filtered estimate of h_{t+1} using the information up to time t.

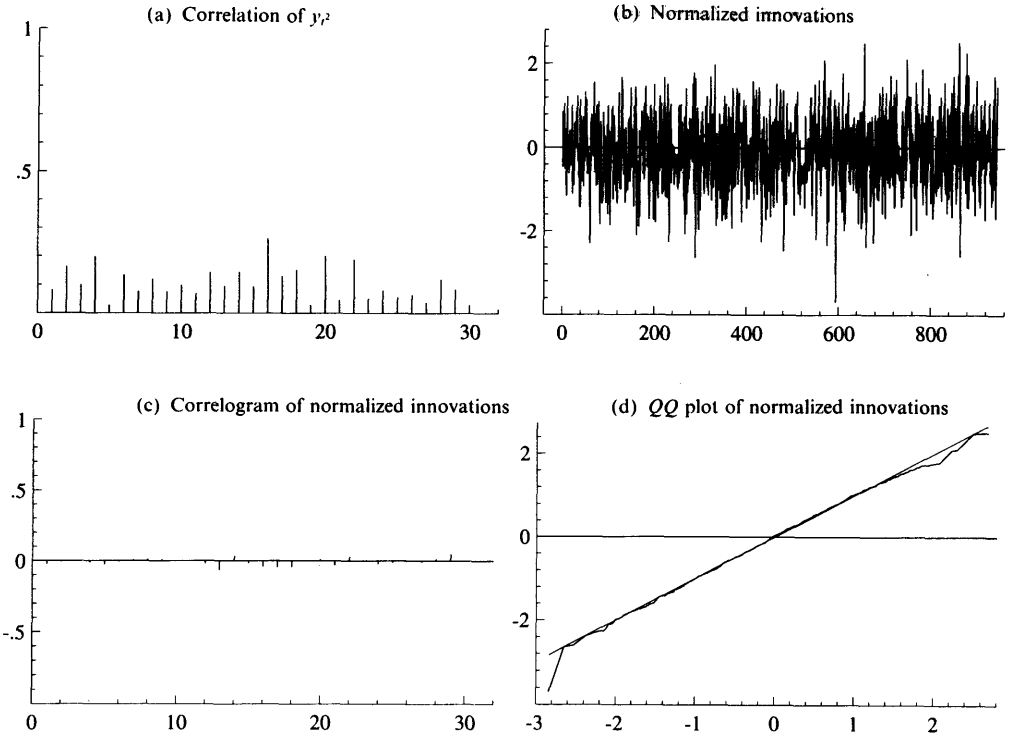

FIGURE 8

Diagnostic checks. Graph (a): correlogram of y_t^2. Graph (b): normalized innovations. Graph (c): the corresponding correlogram. Graph (d): associated QQ-plot

Table 10 shows the results from some standard diagnostic checks on the $n_1^M, \ldots, n_n^M$ produced by the fitted model. Under the correctness of the model, the diagnostics should indicate that the variables are Gaussian white noise. We report the skewness and kurtosis coefficients,

$$\text{Skew} = \frac{nb_3}{6}, \qquad \text{Kurtosis} = \frac{n(b_4 - 3)^2}{24},$$

where b_3 and b_4 denote the standardized estimators of the third and fourth moment of $\{n_t^M\}$ about the mean, an overall Bowman and Shenton (1975) normality statistic which combines these two measures and the Box–Ljung statistic using 30 lags. The table also gives the simulation standard error for these statistics, based on repeating the simulation

TABLE 10

Diagnostics of the SV model using $M = 2500$. BL(l) denotes a Box–Ljung statistic on l lags. The figures in brackets are simulation standard errors using 10 replications. The two other models are fitted using ML. The estimated degrees of the Student t model is given in brackets.

	Skew	Kurtosis	Normality	BL (30)	Log-lik
SV	1·4509	0·54221	2·3992	18·555	−918·56
	(0·057)	(0·083)	(0·295)	(0·120)	(0·558)
NID	11·505	21·640	600·65	401·20	−1018·2
tID(4·87)	1·2537	1·2156	3·0494	700·62	−964·56

ten times with different random draws but with the data fixed. Finally, for comparison the table gives the same diagnostics for the $\mathcal{N}(0, \sigma^2)$ and scaled Student t *iid* models. The results suggest that there are no straightforward failures in the way the model has been fitted.

5. COMPARISON OF NON-NESTED MODELS VIA SIMULATION

5.1. *GARCH model*

In this section we compare the fit of basic SV models with the GARCH models commonly used in the literature. Two approaches are used in this non-nested model comparison—one based on likelihood ratios and another based on ratios of marginal likelihoods resulting in what are called Bayes factors.

The notation we use for the Gaussian GARCH(1, 1) model is

$$y_t | Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2), \qquad \text{where } \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2. \tag{18}$$

while the equivalent Student-t model introduced by Bollerslev (1987) is denoted as t-GARCH with v as the notation for the positive degrees of freedom.

TABLE II

Diagnostics of the ML estimators of the Gaussian and Student t distributed GARCH models. BL(l) denotes a Box–Ljung statistic on l lags. Above the line are the answers of the real data, the ones below are the corrected observations. Figures in brackets for the t-GARCH model are the estimated degrees of freedom

	α_0	$\alpha_1+\alpha_2$	Skew	Kurt	Normality	BL (30)	Log-lik
GARCH	0·0086817	0·98878	4·5399	4·3553	39·580	16·183	−928·13
t-GARCH (8·44)	0·00058463	0·99359	0·56281	0·31972	0·41897	22·515	−917·22

The diagnostic statistics given in Table 11 suggest that the Gaussian GARCH model does not fit the data very well, suffering from positive skewness and excess kurtosis. This suggests that the model cannot accommodate the extreme positive observations in the data. The t-GARCH model is better, with much better distributional behaviour. Again its diagnostics for serial dependence are satisfactory. The fitted likelihood is very slightly better than the SV model, although it has one more parameter.

5.2. *Likelihood ratio statistics*

There is an extensive literature on the statistical comparison of non-nested models based on likelihood ratio statistics. Much of the econometric literature on this topic is reviewed in Gourieroux and Monfort (1994). The approach we suggest here relies on simulation and is based on Atkinson (1986). Related ideas appear in, for instance, Pesaran and Pesaran (1993) and Hinde (1992).

Let $\mathcal{M}_1$ denote the SV model and $\mathcal{M}_0$ the GARCH model. Then, the likelihood ratio test statistic for comparative fit that is investigated here is given by

$$LR_y = 2\{\log \hat{f}(y | \mathcal{M}_1, \hat{\theta}_1) - \log f(y | \mathcal{M}_0, \tilde{\theta}_0)\},$$

where $\log \hat{f}(y | \mathcal{M}_1, \hat{\theta}_1)$ and $\log f(y | \mathcal{M}_0, \tilde{\theta}_0)$ denote the respective estimates of the log likelihoods, the former estimated by simulation as described above,[8] $\hat{\theta}_1$ is the estimated

8. The GARCH process has to be initialized by setting σ_0^2. The choice of this term effects the likelihood function. In our calculations we set $\sigma_0^2 = \alpha_0/(1-\alpha_1-\alpha_2)$.

posterior mean of SV model parameters and $\tilde{\theta}_0$ the MLE of the GARCH model parameters. The sampling variation of LR_y under the hypothesis that the SV model is true or under the alternative that the GARCH model is true is approximated by simulation, following Atkinson (1986). Clearly, analytical derivations of the sampling distribution are difficult given the unconventional estimators of the log-likelihood.

Under the assumption that the SV model is true and the true values of its parameters are $\theta_1^{(0)}$, we generate simulations y^i, $i=1,\ldots,M$ from the true model. For each simulated series we estimate the parameters of the GARCH and SV models and record the value of LR_y, which we denote as LR_y^i. The resulting scatter of values $LR_y^1,\ldots,LR_y^M$ are a sample from the exact distribution of LR_y under the SV null. The fact that we estimated the likelihood and the parameters of the SV model for each y^i does not alter this result. Hence we could use these simulations LR_y^i as inputs into a trivial Monte Carlo test (see, for example, Ripley (1987, pp. 171–174)) of the hypothesis that the GARCH model is true. Unfortunately $\theta_1^{(0)}$ is unknown and so it is estimated from the data and chosen to be $\hat{\theta}_1$. This introduces an additional approximation error into the sampling calculation which falls as the sample size $n \to \infty$.

The estimated approximate sampling distributions of LR_y under each hypothesis based on 99 simulations plus the realization from the data are given in Figure 9. This figure shows that if the null of the SV model is true, then LR_y can be expected to be

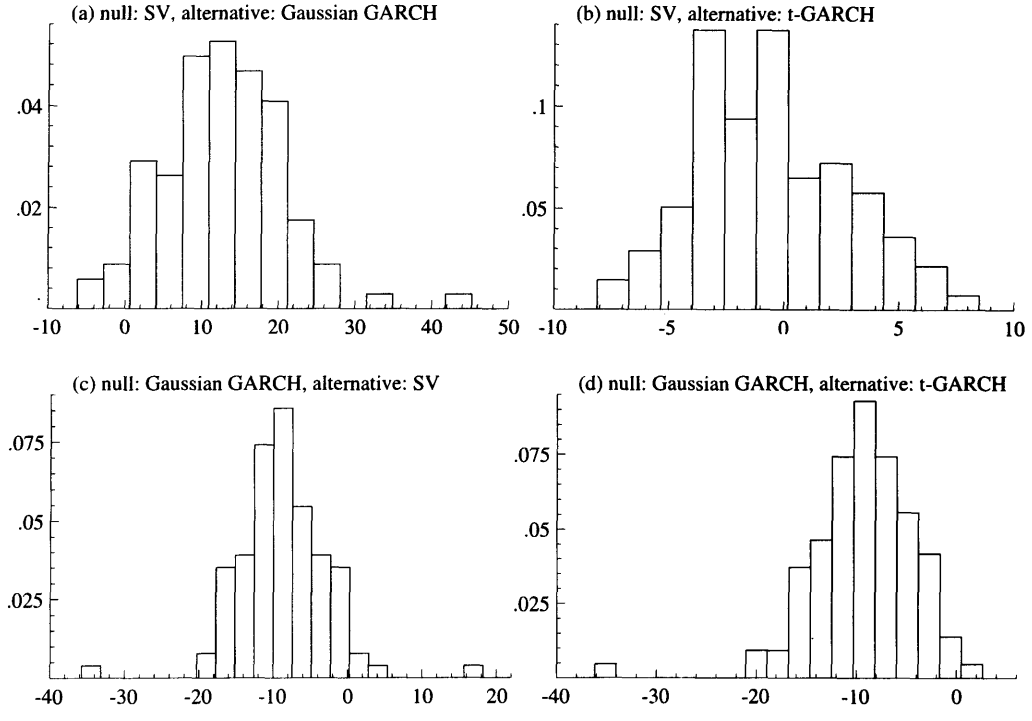

FIGURE 9

Non-nested testing. Graphs (a)–(b) LR_y computed when SV is true. Graph (a): SV against a GARCH model. Graph (b): SV against a t-GARCH. The observed values are 19·14 and −2·68 respectively, which are 80th and 29th out of the 100 samples. Graphs (c)–(d): LR_y computed when GARCH model is true. Graph (c): GARCH against SV. Graph (d): GARCH against t-GARCH. The observed values are 19·14 and −2·68 respectively, which ranks them 100th and 79th out of the 100 samples

positive when the alternative is a Gaussian GARCH, while it is expected to be around zero when the alternative is a t-GARCH.

For the Sterling series the observed LR_y is 19·14 for the SV model against GARCH and −2·68 for the SV model against t-GARCH. This suggests that the SV model fits the data better than the GARCH model but slightly worse than the t-GARCH model (which has one more parameter). These results are confirmed by looking at the simulated LR_y. Table 12 records the ranking of the observed LR_y amongst the 99 simulations conducted under the assumption that the SV model is true. Hence if the observed LR_y is the 96th largest, then it is ranked as being 96th. If the ranking is either close to zero or 100 then this would provide evidence against the SV model.

The recorded rankings under the SV hypothesis are not very extreme, with about 20% of the simulations generating LR tests against the GARCH model which are higher than that observed, while 30% of the simulations were lower than that observed for the t-GARCH LR test. Although suggestive, neither of these tests are formally significant. This implies that they are both consistent with the SV model being true.

A more decisive picture is generated when the Gaussian GARCH model is the null hypothesis. No value is as extreme as the observed LR test against the SV model, rejecting the Gaussian GARCH model for these data. The evidence of the test against the t-GARCH model is less strong.

In summary, the observed non-nested LR_y tests give strong evidence against the use of Gaussian GARCH models. The two remaining models are the t-GARCH and SV models. The statistics show a slight preference for the t-GARCH model, but this model is less parsimonious than the SV model and so it would be fairer to argue for the statement that they fit the data more or less equally well. These results carry over to the other three exchange rates. The results from the non-nested tests are given in Table 12, although there is a considerable evidence that the t-GARCH model is preferable to the SV model for the Yen series.

5.3. *Bayes factors*

An alternative to likelihood ratio statistics is the use of Bayes factors, which are symmetric in the models and extremely easy to interpret. The approach adopted here for the computation of Bayes factors relies on the method developed by Chib (1995). From the basic marginal likelihood identity in Chib (1995), the log of the Bayes factor can be written as

$$\begin{aligned}\log f(y|\mathcal{M}_1) - \log f(y|\mathcal{M}_0) &= \{\log f(y|\mathcal{M}_1, \theta_1^*) + \log f(\theta_1^*|\mathcal{M}_1) - \log f(\theta_1^*|\mathcal{M}_1, y)\} \\ &\quad - \{\log f(y|\mathcal{M}_0, \theta_0^*) + \log f(\theta_0^*) - \log f(\theta_0^*|\mathcal{M}_0, y)\},\end{aligned}$$

for any values of θ_0^* and θ_1^*. Here $f(\theta_0^*)$ is the GARCH prior density, while $f(\theta_1^*|\mathcal{M}_1)$ is the prior for the SV parameters. The likelihood for the GARCH model is known, while that of the SV model is estimated via simulation as described above. Next, the posterior densities $f(\theta_0^*|\mathcal{M}_0, y)$ and $f(\theta_1^*|\mathcal{M}_1, y)$ are estimated at the single points θ_0^* and θ_1^* using a Gaussian kernel applied to the posterior sample of the parameters. We follow the suggestion in Chib (1995) and use the posterior means of the parameters as θ_0^* and θ_1^* since the choice of these points is arbitrary.

TABLE 12

Non-nested LR tests of the SV model against the ARCH models. In each case the 99 simulations were added to the observed LR_y to form the histograms. The reported r-th rankings are the r-th largest of the observed LR test out of the 100 LR_y tests conducted under SV or GARCH model

	SV versus GARCH			SV against t-GARCH		
Series	Observed	Rank SV	Rank GARCH	Observed	Rank SV	Rank GARCH
Sterling	19·14	81st	100th	−2·68	29th	79th
DM	11·00	61st	100th	−3·84	9th	87th
Yen	19·84	99th	100th	−30·50	1st	1st
SwizF	53·12	100th	100th	−3·62	20th	98th

To perform a Bayes estimation of the GARCH model we have to write down some priors for the GARCH parameters. This is most easily done by representing the model in its ARMA(1, 1) form for squared data

$$y_t^2 = \alpha_0 + (\alpha_1 + \alpha_2) y_{t-1}^2 + v_t - \alpha_2 v_{t-1}, \qquad v_t = (\varepsilon_t^2 - 1)\sigma_t^2.$$

Hence $\alpha_1 + \alpha_2$ is the persistence parameter, α_2 (which has to be positive) is the negative of the moving average coefficient, while $\alpha_0/(1-\alpha_1-\alpha_2)$ is the unconditional expected value of y_t^2. We will place the same prior on $\alpha_1 + \alpha_2$ as was placed on the persistence parameter ϕ in the SV model (see (5)). This will force the GARCH process to be covariance stationary. The prior specification is completed by assuming that $\alpha_2/(\alpha_1+\alpha_2) \mid \alpha_1 + \alpha_2 = r_a$ follows a Beta distribution with

$$\log f\{\alpha_2/(\alpha_1+\alpha_2) \mid \alpha_1+\alpha_2 = r_a\}$$
$$= \text{const} + \{\phi^{(1)} - 1\} \log\left\{\frac{\alpha_2}{r_a}\right\} + \{\phi^{(2)} - 1\} \log\left\{\frac{r_a - \alpha_2}{r_a}\right\}. \tag{19}$$

Since we would expect that $\alpha_2/(\alpha_1+\alpha_2)$ to be closer to one than zero, we will take $\phi^{(1)} = 45$ and let $\phi^{(2)} = 2$. This gives a mean of 0·957. The scale parameter $\alpha_0/(1-\alpha_1-\alpha_2) \mid \alpha_1, \alpha_2$ are given a standard diffuse inverse chi-squared prior distribution. Finally, for the t-GARCH model, $v-2$ was given in chi-squared prior with a mean of ten.

In order to carry out the MCMC sampling we used the Gilks, Best, and Tan (1995) procedure which just requires the programming of the priors and the GARCH likelihood.

The results of the calculations are given in Table 13. They are very much in line with the likelihood ratio analysis given in Table 12. Again the SV model dominates the Gaussian GARCH model, while it suffers in comparison with the t-GARCH model, especially for the Yen data series. It should be mentioned, however, that these conclusions are in relation to the simplest possible SV model. The performance of the SV model can be improved by considering other versions of the model, for example, one that relaxes the Gaussian assumption. We discuss this and other extensions next.

6. EXTENSIONS

6.1. *More complicated dynamics*

This paper has suggested three ways of performing Bayesian analysis of the SV model: single move, offset mixture and integration sampling. All three extend to the problem where the volatility follows a more complicated stochastic process than an AR(1). A useful

TABLE 13

Estimated Bayes factors for SV model against GARCH model and t-GARCH. All the densities were evaluated at the estimated posterior mean

	GARCH			t-GARCH			
Series	$\alpha_1+\alpha_2$	$\log f_{GARCH}$	Log Bayes	$\alpha_1+\alpha_2$	υ	$\log f_{GARCH}$	Log Bayes
Sterling	0·9802	−928·64	9·14	0·9822	9·71	−918·13	−3·512
DM	0·9634	−952·88	6·52	0·9712	12·82	−945·63	−2·688
Yen	0·9850	−798·79	13·22	0·9939	6·86	−774·8	−11·28
SwizF	0·9153	−1067·2	27·86	0·9538	7·57	−1039·0	−0·84

framework is

$$h_t = c_t + Z_t\gamma_t, \qquad \text{where } \gamma_{t+1} = d_t + T_t\gamma_t + H_t u_t,$$

where $u_t \overset{iid}{\sim} \mathcal{N}(0, I)$, c_t and d_t are assumed to be strictly exogenous, and Z_t, T_t and H_t are selected to represent the log-volatility appropriately. With this framework the log volatility process can be specified to follow an ARMA process.

In the single move Gibbs algorithm, it is tempting to work with the γ_t as

$$f(\gamma_t | y_t, \gamma_{t-1}, \gamma_{t+1}) \propto f(y_t | c_t + Z_t\gamma_t) f(\gamma_t | \gamma_{t-1}) f(\gamma_{t+1} | \gamma_t), \tag{20}$$

has a simple structure. However, this would suffer from the problems of large MCMC simulation inefficiency documented above especially if γ_t is high dimensional or if the $\{\gamma_t\}$ process displayed considerable memory (akin to the example given in Carter and Kohn (1994)). Alternatively, one could sample h_t using

$$f(h_t | h_{\setminus t}, y) \propto f(y_t | h_t) f(h_t | h_{\setminus t}),$$

as we can evaluate $h_t | h_{\setminus t}$ using the de Jong (1998) scan sampler. This is uniformly superior to the algorithms built using (20). Neither of these choices would be competitive, however, with versions of the multi-move and integration sampler which rely on the state space form and can thus be trivially extended to cover these models.

More sophisticated dynamics for the volatility could be modelled by exploiting factor type models. An example of this is

$$h_t = h_{1t} + h_{2t}, \qquad h_{1t+1} = \phi_1 h_{1t} + \eta_{1t}, \qquad h_{2t+1} = \phi_2 h_{2t} + \eta_{2t}.$$

where $\phi_1 > \phi_2$ and η_{1t}, η_{2t} are independent Gaussian white noise processes. Here h_{1t} and h_{2t} would represent the longer-term and shorter-term fluctuations in log-volatility. The introduction of such components, appropriately parameterized, produce volatility versions of the long memory models advocated by Cox (1991).

6.2. *Missing observations*

The framework described above can also be extended to handle missing data. Suppose that the exchange rate r_{34} at time 34 is missing. Then the returns y_{34} and y_{35} would be missing. We could complete the data by adding in r_{34} to the list of unknowns in the sampling. Given r_{34} we could generate y and then sweep $h, \theta | y$. Having carried this out we could update r_{34} by drawing it given h, θ and y. Iterating this procedure gives a valid MCMC algorithm and so would efficiently estimate θ from the non-missing data.

This argument generalizes to any amount of missing data. Hence this argument also generalizes to the experiment where we think of the SV model (1) holding at a much finer

discretization than the observed data. Think of the model holding at intervals of 1/d-th of a day, while suppose that the exchange rate r_t is available daily. Then we can augment the "missing" intra-daily data $\tilde{r}_t = (r_{t_1}, \ldots, r_{t_{d-1}})$ to the volatilities $\tilde{h}_t = (h_{t_1}, \ldots, h_{t_{d-1}}, h_t)$ and design a simple MCMC algorithm to sample from

$$\tilde{r}_1, \ldots, \tilde{r}_n, \tilde{h}_1, \ldots, \tilde{h}_n, \theta \mid r_0, \ldots, r_n.$$

This will again allow efficient estimation of θ from the "coarse" daily data even though the model is true at the intra-daily level. This type of argument is reminiscent of the indirect inference methods which have recently been developed for diffusions by Gourieroux, Monfort and Renault (1993) and Gallant and Tauchen (1996), however our approach has the advantage of not depending on the ad hoc choice of an auxiliary model and is automatically fully efficient.

6.3. *Heavy-tailed* SV *models*

The discrete time SV model can be extended to allow ε_t in (1) to be more heavy-tailed than the normal distribution. This would help in overcoming the comparative lack of fit indicated by Table 12 for the Yen series. One approach, suggested in Harvey, Ruiz and Shephard (1994) amongst others, is to use an *ad hoc* scaled Student t distribution, so that

$$\varepsilon_t = \sqrt{\frac{v-2}{v}}\, \zeta_t / \sqrt{\chi^2_{t,v}/v}, \qquad \text{where} \qquad \zeta_t \overset{iid}{\sim} \mathcal{N}(0,1), \chi^2_{t,v} \overset{iid}{\sim} \chi^2_v,$$

and the ζ_t and $\chi^2_{t,v}$ are independent of one another. The single move and offset mixture algorithms immediately carry over to this problem if we design a Gibbs sampler for $\chi^2_{1,v}, \ldots, \chi^2_{n,v}, h, \theta \mid y$ or $\chi^2_{1,v}, \ldots, \chi^2_{n,v}, h, \theta, \omega \mid y$ respectively.

An alternative to this, which can be carried out in the single move algorithm, would be to directly integrate out the $\chi^2_{t,v}$, which would mean $f(y_t \mid h_t, \theta)$ would be a scaled Student t distribution. This has the advantage of reducing the dimension of the resulting simulation. However, the conditional sampling becomes more difficult. This is because $f(y_t \mid h_t, \theta)$ is no longer log-concave in h_t and the simple accept/reject algorithm will no longer work. However, one could adopt the pseudo-dominating accept/reject procedure that is discussed in Tierney (1994) and Chib and Greenberg (1995). This version of the algorithm incorporates a Metropolis step in the accept/reject method and does not require a bounding function. The same ideas can also be extended for multivariate models and models with correlated ε_t, η_t errors.

6.4. *Semi-parametric* SV

The offset mixture representation of the SV model naturally leads to a semi-parametric version of the SV model. Suppose we select the "parameters" $m_1, \ldots, m_K$, $v_1^2, \ldots, v_K^2, q_1, \ldots, q_K$ freely from the data. Then, this procedure is tantamount to the estimation of the density of the shocks ε_t. The constraint that Var $(\varepsilon_t) = 1$ is automatically imposed if μ is incorporated into these mixture weights.

This generic approach to semi-parametric density estimation along with MCMC type algorithms for the updating of the mixture parameters has been suggested by Escobar and West (1995) and Richardson and Green (1997). Mahieu and Schotman (1997) use a simulated EM approach to estimate a small number of mixtures inside an SV model.

6.5. *Prior sensitivity*

The methods developed above can be easily modified to assess the consequences of changing the prior. Instead of rerunning the entire samplers with the alternative prior, one can reweight the simulation output so that it corresponds to the new prior—in much the same way as the simulation was reweighted to overcome the bias caused by the offset mixture. Since the posterior is

$$f(\theta, h|y) \propto f(y|h, \theta) f(h|\theta) f(\theta) = f(y|h, \theta) f(h/\theta) f^*(\theta) \frac{f(\theta)}{f^*(\theta)},$$

where $f(\theta)$ denotes the new prior and $f^*(\theta)$ the prior used in the simulations, the reweighting follows the form of (12) where $w^j = \log f(\theta^j) - \log f^*(\theta^j)$. This is particularly attractive as the reweighting is a smooth function of the difference between the old prior f^* and the new prior f. Rerunning the sampler will not have this property.

6.6. *Multivariate factor* SV *models*

The basis of the N dimensional factor SV model will be

$$y_t = Bf_t + \varepsilon_t,$$

where

$$\begin{pmatrix} \varepsilon_t \\ f_t \end{pmatrix} \sim \mathcal{N}\langle 0, \text{diag}\{\exp(h_{1t}), \ldots, \exp(h_{Nt}), \exp(h_{N+1t}), \ldots, \exp(h_{N+Kt})\}\rangle,$$

where f_t is K dimensional and

$$(h_{t+1} - \mu) = \begin{pmatrix} \phi_\varepsilon & 0 \\ 0 & \phi_f \end{pmatrix}(h_t - \mu) + \eta_t, \qquad \eta_t \sim \mathcal{N}\left\{0, \begin{pmatrix} \Sigma_{\varepsilon\eta} & 0 \\ 0 & \Sigma_{f\eta} \end{pmatrix}\right\}.$$

As it stands the model is highly overparameterized. This basic structure was suggested in the factor ARCH models analysed[9] by Diebold and Nerlove (1989) and refined by King, Sentana and Wadhwani (1994), but replaces the unobserved ARCH process for f_t by SV processes. It was mentioned as a possible multivariate model by Shephard (1996) and discussed by Jacquier, Polson and Rossi (1995).

Jacquier, Polson and Rossi (1995) discussed using MCMC methods on a simplified version[10] of this model, by exploiting the conditional independence structure of the model to allow the repeated use of univariate MCMC methods to analyse the multivariate model. This method requires the diagonality of ϕ_ε, ϕ_f, $\Sigma_{\varepsilon\eta}$ and $\Sigma_{f\eta}$ to be successful. However, their argument can be generalized in the following way for our offset mixture approach.

Augment the unknown h, θ with the factors f, for then $h|f, y, \theta$ has a very simple structure. In our case we can transform each f_{jt} using

$$\log(f_{jt}^2 + c) = h_{N+jt} + z_{jt}, \qquad z_{jt}|s_{jt} = i \sim \mathcal{N}(m_i - 1{\cdot}2704, v_i^2),$$

noting that given the mixtures the z_{jt} are independent over j as well as t. Hence we can draw from all at once $h|f, s, y, \theta$. This can then be added to routines which draw from $f|y, h, \theta$ and $\theta|y, h, f$ to complete the sampler.

9. Using approximate likelihood methods. Exact likelihood methods are very difficult to construct for factor ARCH models.

10. Their model sets $\Sigma_t \sim \text{NID}(0, \Sigma_\varepsilon)$, rather than allowing the elements to be stochastic volatility models.

7. CONCLUSION

In this paper we have described a variety of new simulation-based strategies for estimating general specifications of stochastic volatility models. The single move accept/reject algorithm is a natural extension of the previous work in the literature. It is very simple to implement, reliable and is easily generalizable. However, it can have poor convergence properties which has prompted us to develop other samplers which exploit the time series structure of the model.

The key element of our preferred sampler is the linearization made possible by a log-square transformation of the measurement equation and the approximation of a $\log \chi^2$ random variable by a mixture of normal variables. This, coupled with the Bayesian reweighting procedure to correct for the linearization error, enables the analysis of complex models using the well-established methods for working with conditionally Gaussian state-space models. The simulation conducted in this paper shows that our proposed methods can achieve significant efficiency gains over previously proposed methods for estimating stochastic volatility models. Furthermore, this approach will continue to perform reliably as we move to models with more complicated dynamics.

The paper also discusses the computation of the likelihood function of SV models which is required in the computation of likelihood ratio statistics and Bayes factors. A formal comparison of the SV model in relation to the popular heavy tailed version of GARCH model is also provided for the first time. An interesting set of methods for filtering the volatilities and obtaining diagnostics for model adequacy are also developed. The question of missing data is also taken up in the analysis. The results in this paper, therefore, provide a unified set of tools for a complete analysis of SV models that includes estimation, likelihood evaluation, filtering, diagnostics for model failure, and computation of statistics for comparing non-nested models. Work continues to refine these results, with the fitting of ever more sophisticated stochastic volatility models.

8. AVAILABLE SOFTWARE

All the software used in this paper can be downloaded from the World Wide Web at the URL:

`http://www.nuff.ox.ac.uk/users/shephard/ox/`

The software is fully documented. We have linked raw `C++` code to the graphics and matrix programming language `Ox` of Doornik (1996) so that these procedures can be easily used by non-experts.

In the case of the single move Gibbs sampler and the diagnostics routines the software is unfortunately specialized to the SV model with AR(1) log-volatility. However, the other procedures for sampling $h \mid y, s, \theta$ and the resampling weights are general.

APPENDIX

This appendix contains various algorithms which allow the efficient computations of some of the quantities required in the paper.

1. *Basic Gaussian state space results*

We discuss general filtering and simulation smoothing results which are useful for a general Gaussian state space model. We analyse the multivariate model

$$\begin{aligned} y_t &= c_t + Z_t\gamma_t + G_t u_t, \\ \gamma_{t+1} &= d_t + T_t\gamma_t + H_t u_t, \qquad u_t \overset{iid}{\sim} \mathcal{N}(0, I). \\ \gamma_1 | Y_0 &\sim N(a_{1|0}, P_{1|0}). \end{aligned} \tag{21}$$

For simplicity we assume that $G_t H_t' = 0$ and we write the non-zero rows of H_t as M_t, $G_t G_t' = \Sigma_t$ and $H_t H_t' = \Sigma_{\eta t}$. Throughout c_t and d_t are assumed known.

In the context of our paper we have mostly worked with the simplest of models where, putting $\beta = 1$ and letting r_t^d *denote* daily returns computed as (8),

$$\log (r_t^{d2} + \text{const}) = h_t + \varepsilon_t,$$

and

$$h_{t+1} = \mu(1-\phi) + \phi h_t + \eta_t,$$

where we condition on the mixture s_t such that

$$\varepsilon_t | s_t = i \sim \mathcal{N}(m_i, v_i^2)$$

and

$$\eta_t \sim \mathcal{N}(0, \sigma_\eta^2).$$

So this puts $y_t = \log (r_t^{d2} + \text{const})$, $c_t = m_i$, $G_t = (\sigma_i, 0)$, $\gamma_t = h_t$ and $Z_t = 1$. Likewise $d_t = \mu(1-\phi)$, $T_t = \phi$ and $H_t = (0, \sigma_\eta)$. Finally, for a stationary initial condition, $a_{1|0} = \mu$ and $P_{1|0} = \sigma_\eta^2/(1-\phi^2)$. This means that $\varepsilon_t = G_t u_t$, $\eta_t = H_t u_t$ and u_t is a bivariate standard normal.

The Kalman filter is run for $t = 1, \ldots, n$,

$$\begin{aligned} &\gamma_{t+1|t} = d_t + T_t\gamma_{t|t-1} + K_t v_t, \qquad P_{t+1|t} = T_t P_{t|t-1} L_t' + \Sigma_{\eta t}, \qquad v_t = y_t - Z_t\gamma_{t|t-1} - c_t, \\ &F_t = Z_t P_{t|t-1} Z_t' + \Sigma_t, \qquad K_t = T_t P_{t|t-1} Z_t' F_t^{-1}, \qquad L_t = T_t - K_t Z_t. \end{aligned} \tag{22}$$

Here $\gamma_{t+1|t} = \mathsf{E}(\gamma_{t+1} | y_1, \ldots, y_t)$ while $P_{t+1|t}$ is the corresponding mean square error. More detailed discussion of the state space form and the Kalman filter is given in Harvey (1989).

The simulation signal smoother (de Jong and Shephard (1995)) draws from the multivariate normal posterior

$$(c_1 + Z_1\gamma_1, \ldots, c_n + Z_n\gamma_n) | y, \theta,$$

where θ denotes the parameters of the model. Setting $r_n = 0$ and $N_n = 0$, and writing $D_t = F_t^{-1} + K_t' N_t K_t$, $n_t = F_t^{-1} v_t - K_t' r_t$, we run for $t = n, \ldots, 1$,

$$\begin{aligned} &C_t = \Sigma_t - \Sigma_t D_t \Sigma_t, && \kappa_t \sim \mathcal{N}(0, C_t), \\ &r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t - V_t' C_t^{-1} \kappa_t, && V_t = \textstyle\sum_t (D_t Z_t - K_t' N_t T_t), \\ &N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t + V_t' C_t^{-1} V_t. \end{aligned} \tag{23}$$

Then $y_t - \Sigma_t n_t - \kappa_t$ is a draw from the signal $c_t + Z_t\gamma_t | y, \theta, c_{t+1} + Z_{t+1}\gamma_{t+1}, \ldots, c_n + Z_n\gamma_n$.

The freeware package SSFPack, due to Koopman, Shephard and Doornik (1996), provides easy to use functions which perform Kalman filtering and the simulation signal smoothing for an arbitrary state space model.

2. *Augmented state space*

Suppose that we write

$$c_t = X_t\beta, \qquad d_t = W_t\beta, \qquad \beta \sim \mathcal{N}(0, \Lambda),$$

where β is independent of the u_t process. Then we can estimate the states and the regression parameter β at the same time using the augmented Kalman filter and simulation smoother. The first of these ideas is due to de Jong (1991), the second to de Jong and Shephard (1995).

The augmented Kalman filter adds two equations to the Kalman filter (22) which is run with $c_t = 0$, $d_t = 0$, additionally computing

$$V_t^a = -Z_t \gamma_{t|t-1}^a - X_t, \qquad \gamma_{t+1|t}^a = W_t + T_t \gamma_{t|t-1}^a + K_t V_t^a,$$

where $\gamma_{1|0}^a = W_1$. Here V_t^a is a $\dim(y_t) \times \dim(\beta)$ matrix. The augmented innovations V_t^a are the innovations resulting from running $-X_t$ through the Kalman filter (22) with $d_t = W_t$. Hence we can compute the posterior of $\beta | y$ by looking at the weighted least squares regression of v_t on $V_t^a \beta$ with prior information $\beta \sim \mathcal{N}(0, \Lambda)$ and variances F_t. If we set $S_1 = \Lambda^{-1}$ and $s_1 = 0$ (the notation of s_t is local to this discussion) then recursively calculating

$$s_{t+1} = s_t + V_t^{a\prime} F_t^{-1} v_t \quad \text{and} \quad S_{t+1} = S_t + V_t^{a\prime} F_t^{-1} V_t^a,$$

we have that $\beta | y, \theta \sim \mathcal{N}(-S_n^{-1} s_n, S_n^{-1})$, where θ now denotes the remaining parameters in the model.

The random regression effect β can be analytically integrated out of the joint density of y and β (given θ) as

$$f(y|\theta) = \int f(y|\beta, \theta)\pi(\beta)d\beta = \frac{f(y|\beta = 0, \theta)\pi(\beta = 0)}{\pi(\beta = 0|y, \theta}$$

$$\propto \prod_{t=1}^n F_t^{-1/2} \exp\left(-\frac{1}{2}\sum_{t=1}^n v_t^2/F_t\right) |\Lambda|^{-1/2} \exp\left(\frac{1}{2} s_n' S_n^{-1} s_n\right) |S_n^{-1}|^{1/2}, \tag{24}$$

using the terms from the augmented Kalman filter. This result is due to de Jong (1991).

If we draw from $b \sim \beta | y, \theta$ we can calculate a new set of innovations $v_t^s = v_t + V_t^a b$, which are the innovations from running the Kalman filter on a state space with known $\beta = b$. Hence we can use the simulation signal smoother which draws $c_1 + Z_1\gamma_1, \ldots, c_n + Z_n\gamma_n | y, \beta = b, \theta$ using the simulation signal smoother (23) just by plugging in the v_t^s instead of the v_t. By using both of these draws we are actually sampling directly from the distribution of

$$(\beta, c_1 + Z_1\gamma_1, \ldots, c_n + Z_n\gamma_n) | y, \theta.$$

Acknowledgements: This paper is an extensively revised version of a paper with the same title by Sangjoon Kim and Neil Shephard. That version of the paper did not have the section on the use of the reweighting which corrects for the mixture approximation, nor the formal non-nested testing procedures for comparison with GARCH models. Neil Shephard would like to thank the ESRC for their financial support through the project "Estimation via Simulation in Econometrics" and some computational help from Michael K. Pitt and Jurgen A. Doornik. All the authors would like to thank the referees for their comments on the previous version of the paper.

Sangjoon Kim's work was carried out while he was a Ph.D. student at Department of Economics, Princeton University, under the supervision of John Campbell. Sangjoon Kim would like to thank Princeton University for their financial support. The comments of the participants of the "Stochastic volatility" conference of October 1994, held at HEC (Université de Montreal), are gratefully acknowledged. Finally, Neil Shephard would like to thank A. C. Atkinson and D. R. Cox for various helpful conversations on non-nested likelihood ratio testing.

REFERENCES

ATKINSON, A. C. (1986), "Monte-Carlo tests of separate families of hypotheses" (Unpublished paper, Imperial College, London).

BERZUINI, C., BEST, N. G., GILKS, W. R. and LARIZZA, C. (1997), "Dynamic graphical models and Markov chain Monte Carlo methods", *Journal of the American Statistical Association*, **92**, (forthcoming).

BOLLERSLEV, T. (1987), "A conditional heteroskedastic time series model for speculative prices and rates of return", *Review of Economics and Statistics*, **69**, 542–547.

BOLLERSLEY, T., ENGLE, R. F. and NELSON, D. B. (1994), "ARCH models", in R. F. Engle and D. McFadden (eds.), *The Handbook of Econometrics, Volume 4* (Amsterdam: North-Holland), 2959–3038.

BOWMAN, K. O. and SHENTON, L. R. (1975), "Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2", *Biometrika*, **62**, 243–250.

CARTER, C. K. and KOHN, R. (1994), "On Gibbs sampling for state space models", *Biometrika*, **81**, 541–553.

CHESNEY, M. and SCOTT, L. O. (1989), "Pricing European options: a comparison of the modified Black–Scholes model and a random variance model", *Journal of Financial and Qualitative Analysis*, **24**, 267–284.

CHIB, S. (1995), "Marginal likelihood from Gibbs output", *Journal of the American Statistical Association*, **90**, 1313–1321.

CHIB, S. and GREENBERG, E. (1994), "Bayes inference for regression models with ARMA (p, q) errors", *Journal of Econometrics*, **64**, 183–206.
CHIB, S. and GREENBERG, E. (1995), "Understanding the Metropolis–Hastings algorithm", *The American Statistican*, **49**, 327–335.
CHIB, S. and GREENBERG, E. (1996), "Markov chain Monte Carlo simulation methods in econometrics", *Econometric Theory*, **12**, 409–431.
COX, D. R. (1991), "Long-range dependence, non-linearity and time irreversibility", *Journal of Time Series Analysis*, **12**, 329–335.
DE JONG, P. (1991), "The diffuse Kalman filter", *Annals of Statistics*, **19**, 1073–1083.
DE JONG, P. (1998), "The scan sampler", *Biometrika*, **85**, (forthcoming).
DE JONG, P. and SHEPHARD, N. (1995), "The simulation smoother for time series models", *Biometrika*, **82**, 339–350.
DIEBOLD, F. X. and NERLOVE, M. (1989), "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model", *Journal of Applied Econometrics*, **4**, 1–21.
DOORNIK, J. A. (1996), *Ox: Object Oriented Matrix Programming, 1.10* (London: Chapman & Hall).
ENGLE, R. F. (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation", *Econometrica*, **50**, 987–1007.
ESCOBAR, M. and WEST, M. (1995), "Bayesian density estimation and inference using mixtures", *Journal of the American Statistical Association*, **90**, 577–588.
FULLER, W. A. (1996) *Introduction to Time Series* (2nd ed.) (New York: John Wiley).
GALLANT, A. R. and TAUCHEN, G. (1996), "Which moments to match", *Econometric Theory*, **12**, 657–681.
GELFAND, A. and SMITH, A. F. M. (1990), "Sampling-based approaches to calculating marginal densities", *Journal of the American Statistical Association*, **85**, 398–409.
GEMAN, S. and GEMAN, D. (1984). "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images", *IEEE Transactions, PAMI*, **6**, 721–741.
GEWEKE, J. (1994), "Comment on Bayesian analysis of stochastic volatility models", *Journal of Business and Economic Statistics*, **12**, 397–399.
GHYSELS, E., HARVEY, A. C. and RENAULT, E. (1996), "Stochastic volatility", in C. R. Rao and G. S. Maddala (eds.), *Statistical Methods in Finance* (Amsterdam: North-Holland), 119–191.
GILKS, W. K., RICHARDSON, S. and SPIEGELHALTER, D. J. (1996) *Markov Chain Monte Carlo in Practice* (London: Chapman & Hall).
GILKS, W. R., BEST, N. G. and TAN, K. K. C. (1995), "Adaptive rejection Metropolis sampling within Gibbs sampling", *Applied Statistics*, **44**, 155–173.
GILKS, W. R. and WILD, P. (1992), "Adaptive rejection sampling for Gibbs sampling", *Applied Statistics*, **41**, 337–348.
GORDON, N. J., SALMOND, D. J. and SMITH, A. F. M. (1993), "A novel approach to non-linear and non-Gaussian Bayesian state estimation", *IEE-Proceedings F*, **140**, 107–133.
GOURIEROUX, C. and MONFORT, A. (1994), "Testing non-nested hypotheses", in R. F. Engle and D. McFadden (eds.), *The Handbook of Econometrics, Volume 4* (Amsterdam: North-Holland), 2583–2637.
GOURIEROUX, C., MONFORT, A. and RENAULT, E. (1993), "Indirect inference", *Journal of Applied Econometrics*, **8**, S85–S118.
HARVEY, A. C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge: Cambridge University Press).
HARVEY, A. C., RUIZ, E. and SHEPHARD, N. (1994), "Multivariate stochastic variance models", *Review of Economic Studies*, **61**, 247–264.
HASTINGS, W. K. (1970), "Monte-Carlo sampling methods using Markov chains and their applications", *Biometrika*, **57**, 97–109.
HINDE, J. (1992), "Choosing between non-nested models: a simulation approach", in L. Fahrmeir, B. Francis, R. Gilchrist and G. Tutz (eds.), *Advances in GLIM and Statistical Modelling, Proceedings of the GLIM92 Conference and the 7th International Workshop on Statistical Modelling, Munich 13–17 July 1992* (New York: Springer-Verlag), 119–124.
HULL, J. and WHITE, A. (1987), "The pricing of options on assets with stochastic volatilities", *Journal of Finance*, **42**, 281–300.
ISARD, M. and BLAKE, A. (1996), "Contour tracking by stochastic propagation of conditional density", *Proceedings of the European Conference on Computer Vision, Cambridge 1*, 343–356.
JACQUIER, E., POLSON, N. G. and ROSSI, P. E. (1994), "Bayesian analysis of stochastic volatility models (with discussion)", *Journal of Business and Economic Statistics*, **12**, 371–417.
JACQUIER, E., POLSON, N. G. and ROSSI, P. E. (1995), "Models and prior distributions for multivariate stochastic volatility" (Unpublished paper, Graduate School of Business, University of Chicago).
KING, M., SENTANA, E. and WADHWANI, S. (1994), "Volatility and links between national stock markets", *Econometrica*, **62**, 901–933.
KITAGAWA, G. (1996), "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models", *Journal of Computational and Graphical Statistics*, **5**, 1–25.
KOOPMAN, S. J., SHEPHARD, N. and DOORNIK, J. A. (1996), "SSFPack 1.1: filtering, smoothing and simulation algorithms for state space models in Ox" (Unpublished paper and software), available at http://www.nuff.ox.ac.uk/users/shephard/ox/.

MAHIEU, R. and SCHOTMAN, P. (1997), "Stochastic volatility and the distribution of exchange rate news", *Journal of Applied Econometrics*, **15**, (forthcoming).
MARRIOTT, J. M. and SMITH, A. F. M. (1992), "Reparameterization aspects of numerical Bayesian methodology for autoregressive moving-average models", *Journal of Time Series Analysis*, **13**, 327–343.
METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953), "Equations of state calculations by fast computing machines", *Journal of Chemical Physics*, **21**, 1087–1092.
MULLER, P. (1991), "Numerical integration in general dynamic models", *Contemporary Mathematics*, **115**, 145–163.
PARK, S. and MILLER, K. (1988), "Random number generators: good ones are hard to find", *Communications of the ACM*, **31**, 1192–1201.
PESARAN, M. H. and PESARAN, B. (1993), "A simulation approach to the problem of computing Cox's statistic for testing nonnested models", *Journal of Econometrics*, **57**, 377–392.
PHILLIPS, P. C. B. (1991), "To criticise the critics: an objective Bayesian analysis of stochastic trends", *Journal of Applied Econometrics*, **6**, 333–364.
PITT, M. K. and SHEPHARD, N. (1997), "Filtering via simulation: an auxiliary variable approach" (Working paper, Nuffield College, Oxford).
PITT, M. K. and SHEPHARD, N. (1998), "Analytic convergence rates and parameterisation issues for the Gibbs sampler applied to state space models", *Journal of Time Series Analysis*, **19**, (forthcoming).
PRIESTLEY, M. B. (1981) (London: Academic Press). *Spectral Analysis and Time Serie.*
RICHARDSON, S. and GREEN, P. (1997), "On Bayesian analysis of mixtures with unknown number of components", *Journal of the Royal Statistical Society*, **59**, (forthcoming).
RIPLEY, B. D. (1977), "Modelling spatial patterns (with discussion)", *Journal of the Royal Statistical Society B* **39**, 172–212.
RIPLEY, B. D. (1987) *Stochastic Simulation* (New York: Wiley).
ROSENBLATT, M. (1952), "Remarks on a multivariate transformation", *Annals of Mathematical Statistics*, **23**, 470–472.
SCHOTMAN, P. C. and VAN DIJK, H. K. (1991), "A Bayesian analysis of the unit root in real exchange rates, *Journal of Econometrics*, **49**, 195–238.
SHEPHARD, N. (1993), "Fitting non-linear time series models, with applications to stochastic variance models", *Journal of Applied Econometrics*, **8**, S135–S152.
SHEPHARD, N. (1994), "Partial non-Gaussian state space", *Biometrika*, **81**, 115–131.
SHEPHARD, N. (1996), "Statistical aspects of ARCH and stochastic volatility", in D. R. Cox, O. E. Barndorff-Nielson and D. V. Hinkley (eds.), *Time Series Models in Econometrics, Finance and Other Fields* (London: Chapman & Hall), 1–67.
SHEPHARD, N. and KIM, S. (1994), "Comment on 'Bayesian analysis of stochastic volatiality models' by Jacquier, Polson and Rossi", *Journal of Business and Economic Statistics*, **11**, 406–410.
SHEPHARD, N. and PITT, M. K. (1997), "Likelihood analysis of non-Gaussian measurement time series", *Biometrika*, **84**, 653–667.
TANNER, M. A. (1996) *Tools for Statistical Inference: methods for exploration of posterior distributions and likelihood functions* (3 edn.) (New York: Springer-Verlag).
TANNER, M. A. and WONG, W. H. (1987), "The calculation of posterior distributions by data augmentation (with discussion)", *Journal of the American Statistical Association*, **82**, 528–550.
TAYLOR, S. J. (1986) *Modelling Financial Time Series* (Chichester: John Wiley).
TAYLOR, S. J. (1994), "Modelling stochastic volatility", *Mathematical Finance*, **4**, 183–204.
TIERNEY, L. (1994), "Markov Chains for exploring posterior distributions (with discussion)", *Annals of Statistics*, **21**, 1701–1762.
TITTERINGTON, D. M., SMITH, A. F. M. and MAKOV, U. E. (1985) *Statistical Analysis of Finite Mixture Distributions* (Chichester: John Wiley).
WEST, M. (1993), "Approximating posterior distributions by mixtures", *Journal of the Royal Statistical Society B*, **55**, 409–442.