STOCHASTIC VOLATILITY: LIKELIHOOD INFERENCE AND COMPARISON WITH ARCH MODELS

By Sangjoon Kim and Neil Shephard
Salomon Brothers Asia Limited and Nuffield College, Oxford

November 5, 1994

Summary

Stochastic volatility models present a natural way of working with time-varying volatility. However the difficulty involved in estimating these types of models has prevented their wide-spread use in empirical applications.

In this paper we exploit Gibbs sampling to provide a likelihood framework for the analysis of stochastic volatility models, demonstrating how to perform either maximum likelihood or Bayesian estimation. The paper includes an extensive Monte Carlo experiment which compares the efficiency of the maximum likelihood estimator with that of quasi-likelihood and Bayesian estimators proposed in the literature. We also compare the fit of the stochastic volatility model to that of ARCH models using the likelihood criterion to illustrate the flexibility of the framework presented.

Some key words: ARCH, Bayes estimation, Gibbs sampler, Heteroscedasticity, Maximum likelihood, Quasi-maximum likelihood, Simulation, Stochastic EM algorithm, Stochastic volatility, Stock returns.

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 important autoregressive conditional heteroscedasticity — ARCH — based models developed by Engle (1982), Bollerslev (1986) and Nelson (1991), which are surveyed in Bollerslev, Chou, and Kroner (1992) and Bollerslev, Engle, and Nelson (1994).

An alternative to using the ARCH framework is to allow the changing variance to follow some latent stochastic process. This is called a stochastic variance or volatility (SV) model. For theoretical work, especially with derivatives, it is convenient to formulate SV models in continuous time. An example of this is the geometric diffusion model used by Hull and White (1987) in their work generalising the Black-Scholes option pricing formula to allow for stochastic volatility. However, for obvious reasons, empirical estimates of the parameters of the model are obtained using a set of discrete time data. Assuming for the simplicity of exposition that the drift term is zero and that the log volatility process has zero mean, a simple Euler discrete time model (by Taylor (1986)) for returns for regularly spaced data is:

$$\begin{array}{rclcl} y_t & = & \epsilon_t \exp(\alpha_t/2), & \epsilon_t & \sim & NID(0,1), \\ \alpha_{t+1} & = & \phi\alpha_t + \eta_t, & \eta_t & \sim & NID(0,\sigma_\eta^2), t = 1, ..., T, \end{array} \tag{1}$$

where y_t is the return on holding the asset. In our analysis we will assume that ϵ_t and η_t are uncorrelated. 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 time series properties are discussed in Taylor (1986) and Taylor (1994). Generalising to irregularly spaced observations raises no new technical issues.

There are now a considerable number of papers which focus on various econometric aspects of the SV model. Most deal with generalized method of moment (GMM) type estimators, examples being Taylor (1994), Duffie and Singleton (1993), Melino and Turnbull (1990), Andersen and Sorensen (1994) and Vetzal (1992). Unfortunately we know that GMM may be inefficient relative to likelihood based procedures. Indeed there is some evidence that even the suboptimal quasi-likelihood procedure put forward by Harvey, Ruiz, and Shephard (1994)[1] may be more efficient than most GMM procedures for high frequency data.[2] The recent work by Andersen and Sorensen (1994) is worth noting,

[1]An extended Kalman filter has been proposed by Watanabe (1992) to deal with some of the non-normality of the quasi-likelihood, but it is difficult to study the properties of this ad hoc suggestion.

[2]Jacquier, Polson, and Rossi (1994) find that the GMM and QML estimators are comparable for

as for their suggested guidelines could improve the performance of the GMM estimator. However it remains the case that likelihood analysis is more efficient than GMM.

In some very recent work, Gallant, Hsieh, and Tauchen (1994) have used an 'efficient method of moments' procedure to fit and check the SV model using the scores generated by the semi non-parameteric (SNP) model of Gallant, Rossi, and Tauchen (1992).

Gourieroux, Monfort, and Renault (1993) have used the quasi-likelihood function as the kernel from which to develop their indirect inference estimator for the continuous time analog[3] of (1), while Ghysels and Jasiak (1994) used the QML estimator to study the relationship between volume and volatility using a time deformation technique. Mahieu and Schotman (1994a) have applied the QML approach in a multivariate analysis of exchange rates, while Harvey (1994) has extended the SV model to allow for long memory effects.

More recently Jacquier, Polson, and Rossi (1994) have proposed a Bayesian approach to the estimation of the SV model using a single-state or single-move Markov Chain Monte Carlo (MCMC) technique (using techniques suggested in Carlin, Polson, and Stoffer (1992)) and Jacquier, Polson, and Rossi (1995) extended these results to some multivariate SV models, while independently Shephard (1993) has used similar methods to implement a stochastic EM algorithm for this problem. Properties of their sampling procedures make these methods computationally demanding for some of these SV models. The Jacquier, Polson, and Rossi (1994) approach has been used by Geweke (1994) in some empirical Bayesian comparisons of SV models with GARCH models. Finally, Danielsson and Richard (1993) and Danielsson (1994) have suggested applying the idea of importance sampling to this problem. Their use of importance sampling seems to allow the computation of an approximation to the likelihood for this model, but their procedure is quite involved and difficult to generalize to more complicated models.

In this paper we develop computationally simple and efficient methods for likelihood-based inference for the model given in (1).[4] The methods we develop rely on simulation-based techniques to solve the very high dimensional integrals which appear in the statistical analysis of the SV model[5]. Our method makes it possible to perform either a Bayesian analysis or classical estimation and inference via a simulated EM algorithm.

weekly stock returns data.

[3]See also Engle and Lee (1994) who use GARCH models as their kernel to estimate the continuous time SV models.

[4]Generalising our procedure to more complicated linear dynamics for the log volatility process α_t, allowing it to be an ARIMA(p,d,q) process for example, raises no new difficulties in the framework we develop. The same is true of adding explanatory models to the volatility process.

[5]The dimensions involved are of the order of the sample size.

The MCMC method we design exploits the linear structure in the SV model so that the integrals can be solved very rapidly. As a result each step of the sampling procedure suggested in this paper can be carried out on a fast PC in a matter of seconds. Furthermore the MCMC procedure, being based on the familiar Kalman filter routines, is reasonably straightforward to program and can be extended in a variety of directions to analyse a wide class of interesting models.

As an application of our framework, we present results obtained using the daily and weekly return data on the S&P 500 index, comparing the performance of SV models to that of ARCH. This type of direct comparison is made possible by the likelihood estimates which are readily available in our new framework. In section 2 we describe our procedure and discuss its sampling properties. We then outline in section 3 how a Bayesian and a maximum likelihood analysis of the problem may be carried out. In section 4 compare the fit of ARCH models to that of stochastic volatility models using the log-likelihood criterion, and conduct diagnostic checks of the fitted models. Finally section 5 concludes.

2 ECONOMETRIC METHODOLOGY

A key feature of the basic SV model, (1), is that it can be transformed into a linear model by taking the logarithm of the squares of the observations. No information is lost by making the transformation: $\log y_t^2 = \alpha_t + \log \epsilon_t^2$. Nelson (1988) and Harvey, Ruiz, and Shephard (1994) have employed a Kalman filter, (21), to estimate the unobservable log volatility, α_t, and a quasi-likelihood:

$$\log L_Q = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log F_t - \frac{1}{2}\sum_{t=1}^{T} v_t^2/F_t, \qquad (2)$$

to perform parameter estimation. This is suboptimal as $\log \epsilon_t^2$ is far away from its normal approximation as shown in Figure 1, which means that the quasi-likelihood estimator based on the normal approximation has poor small sample properties, even though the usual quasi-likelihood asymptotic theory is correct.

The first step of our approach is to use a mixture of normals to obtain an arbitrarily close approximation to the true density of $\log \epsilon_t^2$ without giving up the linear structure obtained by the log-square transformation. The general algorithm was first suggested in Shephard (1994), and subsequently used by Carter and Kohn (1993) who fitted a cubic spline through simulated data from a SV model to illustrate their simulation smoothing methods and by Mahieu and Schotman (1994b) for their analysis of exchange rate news.

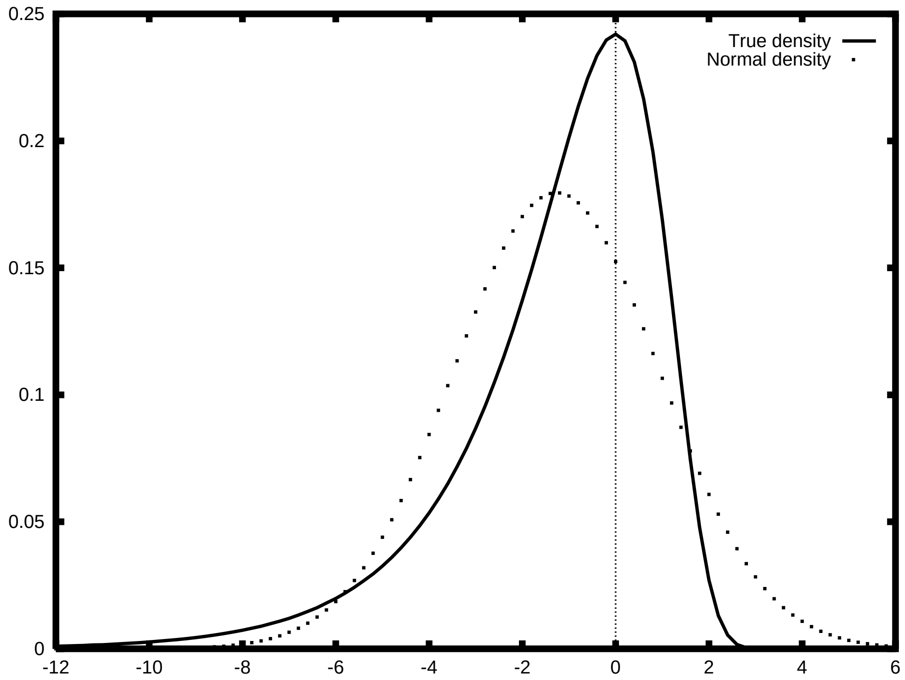

Figure 1: $\log \chi_1^2$ density and normal approximation used in quasi-maximum likelihood estimation.

If we denote by $f_Z(z)$ the mixture density, we have

$$f_Z(z) = \sum_{i=1}^{K} \pi_i f_{z|\omega=i}(z|\omega = i), z|\omega = i \sim N(\mu_i - 1.2704, \sigma_i^2). \tag{3}$$

We are now in a position to select K, $\mu_1, ..., \mu_K$, $\pi_1, ..., \pi_K$ and $\sigma_1^2, ..., \sigma_K^2$ to make the approximation "sufficiently good". In our work, following for instance Titterington, Smith, and Makov (1985, p. 133), we matched the first four moments of $f_{\exp(Z)}(r)$ and $f_Z(r)$ 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

ω	$Pr(\omega = i)$	μ_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

Table 1: Selection of the Mixing Distribution to be $\log \chi_1^2$

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 optimisation 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 1, while the approximating and the true density are drawn in Figure 2. It would be easy to improve the fit by increasing the value of K, however further experiments we have conducted suggest that increasing K has little discernible effect on our main results.

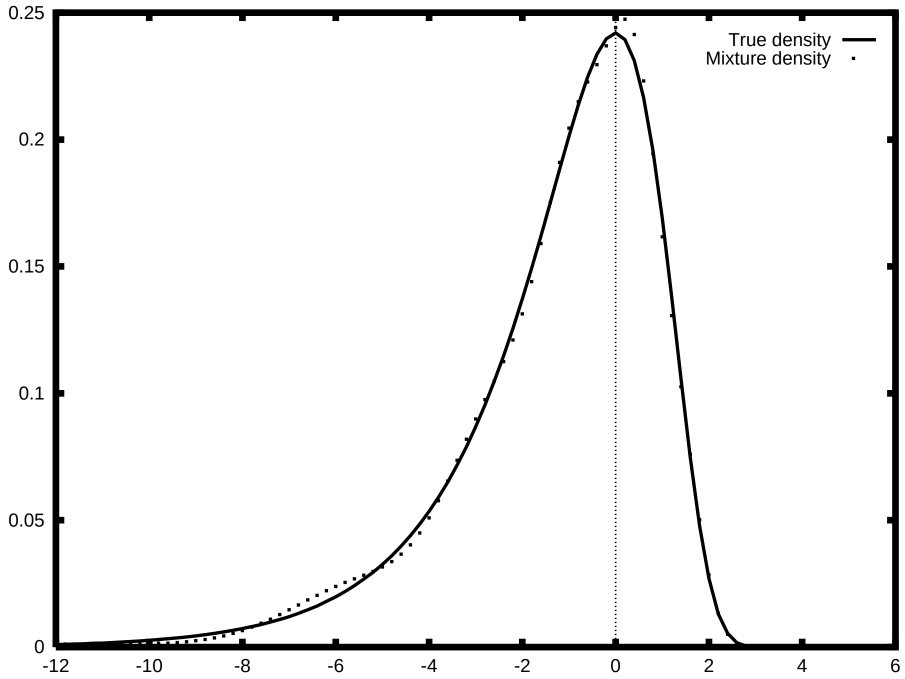

Figure 2: $\log \chi_1^2$ density and mixture approximation.

2.1 Partial Gaussian time series model

When we condition on the random mixture index ω_t, $\log \varepsilon_t^2$ is Gaussian. This implies that $\log y_t^2$ is in a Gaussian state space conditional on a realisation of ω_t. This observation enables us to exploit the structure of partial non-Gaussian time series models recently developed by Shephard (1994). In what follows we shall use the notation $Y_t = (\log y_1^2, ..., \log y_t^2)'$, to denote the information available at time t.

The basic idea of the partial non-Gaussian time series is that conditional on some random vector ω, the model is in a Gaussian state space; see Hamilton (1994, Ch. 13) for a discussion of this class of models. By allowing the elements of ω to have a variety of forms, this framework includes a state space model with switching regimes (generalising the AR model with switches suggested by Hamilton (1989)), heavy tails (solving the signal extraction problem in the presence of outliers), additive-multiplicative components (extending component models to allow cycles and trends to interact), etc.

To be able to use the partial non-Gaussian framework for the SV model we need to be able to simulate from the Gaussian density of $\alpha = (\alpha_1, ..., \alpha_T)'|Y_T, \omega$ and the multinomial $\omega = (\omega_1, ..., \omega_T)'|Y_T, \alpha$. These two components of the simulation are then combined using a rapidly converging multi-move Gibbs sampler,[6] to enable us to draw from the density $f(\alpha|Y_T)$. It has the form:

Multi-move Sampler

1. Generate a candidate point ω.
2. Draw α^* from the density $f(\alpha|\omega, Y_T)$.
3. Draw ω^* from $\Pr(\omega|\alpha^*, Y_T)$.
4. Write $\omega = \omega^*$ and go to 2.

As the sampler iterates, $\alpha^* \xrightarrow{d} \alpha|Y_T$ as long as the joint density of the $\omega|\alpha^*, Y_T$ is everywhere positive (Tierney (1993)). This holds for *all parameter values* in the SV model including when $\sigma_\eta^2 = 0$.

An important observation in the context of the approach we advocate is that $\alpha|Y_T$ and $\omega|Y_T$ are not strongly correlated. As a result this sampler will converge *very quickly*. This contrasts with the Metropolis algorithm studied by Jacquier, Polson, and Rossi (1994) and Shephard (1993) which converge very slowly when the persistence parameter ϕ is close to one, as is often the case in high frequency financial data, e.g., stock returns and exchange rates.

The sampler requires us to be able to simulate from $\alpha|\omega, Y_T$ and $\omega|\alpha, Y_T$. This is straightforward using the simulation smoother of De Jong and Shephard (1995) which is given in the Appendix. It has two parts: the first is to run the Kalman filter, (21), which is started at $P_{1|0} = \sigma_\eta^2/[1-\phi^2]$,and $a_{1|0} = 0$; the second comprises the simulations given in (22) and (24).

[6] Early work in this area include Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) and Hastings (1970). The special case of the Gibbs sampler was put forward by Geman and Geman (1984). Its use is particularly popular in Bayesian statistics, see Smith and Roberts (1993), and in spatial statistics, see Ripley (1987). Early uses of this MCMC technology in econometrics include Jacquier, Polson, and Rossi (1994), Shephard (1993), Albert and Chib (1993), McCulloch and Rossi (1994), Chib (1992) and Chib (1993).

The simulation from $\Pr(\omega|Y_T, \alpha)$ is also straightforward. It is sufficient to note that

$$\Pr(\omega|\alpha^*, Y_T) = \prod_{t=1}^{T} \Pr(\omega_t|\alpha_t^*, \log y_t^2) \propto \prod_{t=1}^{T} f(\log y_t^2|\omega_t, \alpha_t^*) \Pr(\omega_t). \tag{4}$$

The constant of proportionality can be found by evaluating $f(\log y_t^2|\omega_t = i, \alpha_t^*) \Pr(\omega_t = i)$ for all $i = 1, \ldots, 7$, so the simulation boils down to sampling from a multinomial distribution. This only requires a single draw from the uniform distribution for each t.

The ideas described above can be extended to cover the case where α is allowed to follow an arbitrary $ARIMA(p, d, q)$ process. The approach also allows a simple analysis of random level shifts and outliers as discussed by McCulloch and Tsay (1994), Albert and Chib (1993), and McCulloch and Tsay (1993), which may be applicable to the analysis of crashes in financial markets.

2.2 Comparison with single-move simulators

To provide some impression of the quality of the performance of our suggested procedure, we conducted two small experiments. The first uses $\phi = 0.9$ and $\sigma_\eta^2 = 0.05$ (suggestive of weekly financial data); the second sets $\phi = 0.99$ and $\sigma_\eta^2 = 0.01$ (daily data). To study the speed of our multi-move sampler, we use two techniques. In the first we run in parallel 1000 algorithms, all started off by setting $\omega_t, t = 1, \ldots, n$, to values associated with the QML approximation.[7] We average the 1000 answers to obtain estimates of the time varying log of volatility α_t after specific numbers of iterations and graph them against time in Figure 3. This gives some indication of the memory inherent in the algorithm, i.e., how long the effects of the initial values last in the process and hence how much serial correlation there is in the sequence once the algorithm has reached equilibrium. We can see that the algorithm does converge, refining quickly the linear Kalman filter solution. It reaches its equilibrium in a matter of a few iterations.

Figure 3 also gives the results from the second approach[8]. This runs a single-move algorithm for $100,000$ iterations recording the draw α_{50}^j from $\alpha_{50}|Y_n$ after each iteration j. We discard the first $10,000$ iterations and draw the correlogram of the resulting sequence of $N = 90,000$ observations. These confirm the first results, showing how the sampler possess only a moderate degree of correlation. This sequence of observations can also be used to estimate the Monte Carlo variance of the estimate, $\widehat{\alpha_{50}}$ (the average of the α_{50}^j) of the the log volatility. The estimates of $Nvar\{\widehat{\alpha_{50}}\}$, reported in Table 2, is

[7]This is done by introducing an eighth value for ω_t which is used only for the first iteration of the sampling procedure.

[8]We are grateful to Peter Rossi for convincing us of the added value of the second approach.

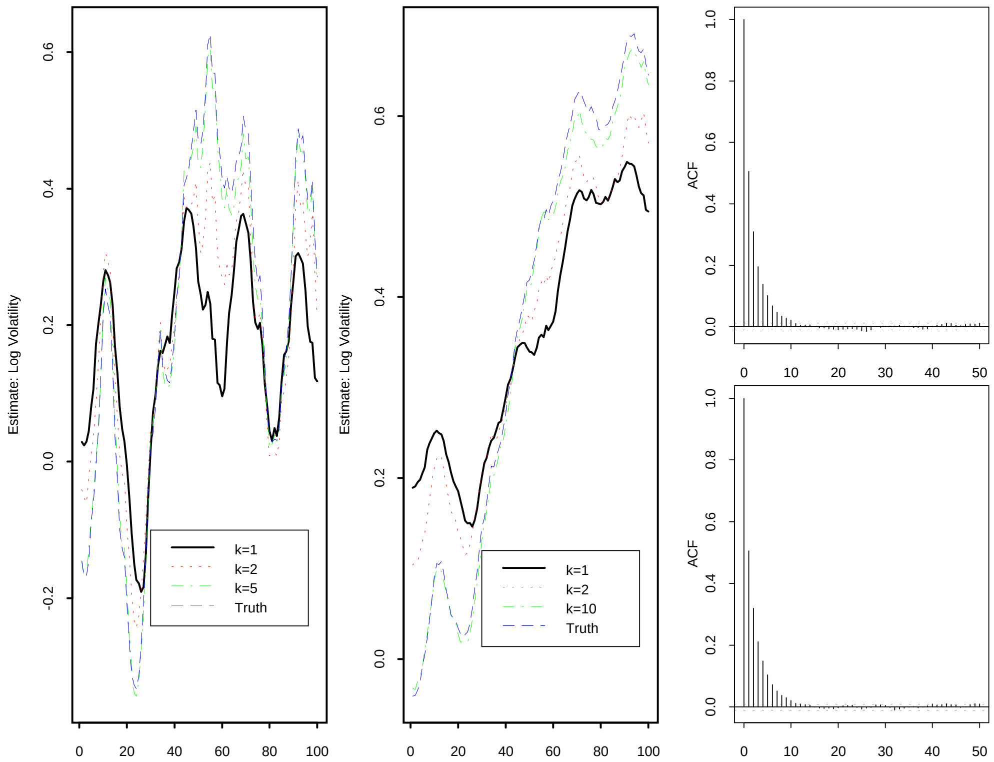

Figure 3: Multi-move algorithms. Weekly parameters ($\phi = 0.9, \sigma_\eta^2 = 0.05$), Daily parameters ($\phi = 0.99, \sigma_\eta^2 = 0.01$) and Corresponding Correlograms. Y-Axis: Log Volatility, X-Axis: Time.

$$\sum_{t=0}^{1000}\left(1-\frac{|t|}{1000}\right)\left\{\frac{1}{N}\sum_{j=1}^{N}(\alpha_{50}^{j}-\widehat{\alpha_{50}})\alpha_{50}^{j+t}\right\}.$$

The Table also reports the corresponding Rao-Blackwellised estimate (see Gelfand and Smith (1990) and Liu, Wong, and Kong (1994)) which averages the smoothed estimate of α_{50} found by running an analytic smoother (see De Jong (1989)) conditioning on ω^j. We call this the mixing estimate of the log-volatility (see Carter and Kohn (1994)).

In Kim and Shephard (1994) we compared this multi-move algorithm with the Jacquier, Polson, and Rossi (1994) algorithm. Here we will use a simpler single-state algorithm. This is based on the fact that

$$\log f(\alpha_t|\alpha_{t-1},\alpha_{t+1},y_t) = const - \frac{y_t^2}{2\exp(\alpha_t)} - \frac{(\alpha_t-\phi\alpha_{t-1})^2}{2\sigma_\eta^2} - \frac{(\alpha_{t+1}-\phi\alpha_t)^2}{2\sigma_\eta^2} \tag{5}$$

is concave in α_t, which means we can directly simulate from it using the Gilks and Wild (1992) and Wild and Gilks (1993) routine[9]. This gives us the following simple Gibbs

[9] The Wild and Gilks (1993) algorithm requires the user to supply only two subroutines which calculate the log density and its derivative (ignoring constants).

ϕ	σ_η^2	MM	SM	Mixing MM
0.9	0.1	.486	2.181	.302
0.99	0.01	.185	6.732	.110
0.995	0.005	.136	7.726	.082

Table 2: Information content per observation: estimate of $NVar(\widehat{\alpha_{50}})$. Based on 100,000 draws discarding 10,000. MM denotes multi-move sampler, SM is the single-move sample. Mixing MM is the MM sampler using an analytic smoother conditional on ω. The comparative information content per observation is found by calculating the ratio of these estimates. So in the $\phi = 0.9$ case MM is four times more informative than SM, while in the $\phi = 0.995$ case this becomes 57.

sampler:

Single-move Sampler

1. Generate a candidate point α.
2. Draw α_t^* from the density $f(\alpha_t|\alpha_{t-1}^*, \alpha_{t+1}, y_t)$, for $t = 1, ..., n$.
3. Write $\alpha = \alpha^*$ and go to 2.

Figure 4, repeats the above experiment using the single-move algorithm. The convergence speed of Metropolis type algorithms is broadly controlled by the correlation amongst the elements which are being moved. If the elements are nearly orthogonal then the convergence will be almost instantaneous, if they are nearly singular then it will be very slow. In the case of the SV model the relevant correlation matrix is $Corr(\alpha|Y_T)$, which becomes singular as $\phi \to 1$ and $\sigma_\eta^2 \to 0$. For typical high frequency financial data we know that volatility is highly persistent. We may well expect realistic parameter values to yield a $Corr(\alpha|Y_T)$ matrix which is nearly singular. This contributes to the relatively slow convergence, high correlation and high values of $Nvar\{\widehat{\alpha_{50}}\}$ reported in Table 2, of the single-move Gibbs sampler and other single-move algorithms like that advocated by Jacquier, Polson, and Rossi (1994) for SV models and Carlin, Polson, and Stoffer (1992) for more general state-space models. On the other hand, in our new sampling procedure the moves are between the blocks $\alpha|Y_T, \omega$ and $\omega|Y_T, \alpha$, which are only moderately correlated. This results in the very rapid convergence we see in Figure 3.

Some readers will not have been convinced by this example, perhaps arguing that although the single-move sampler is slower, the difference in speed is unimportant given the rapidly increasing speed of computers. We disagree with this sentiment: Figures 5

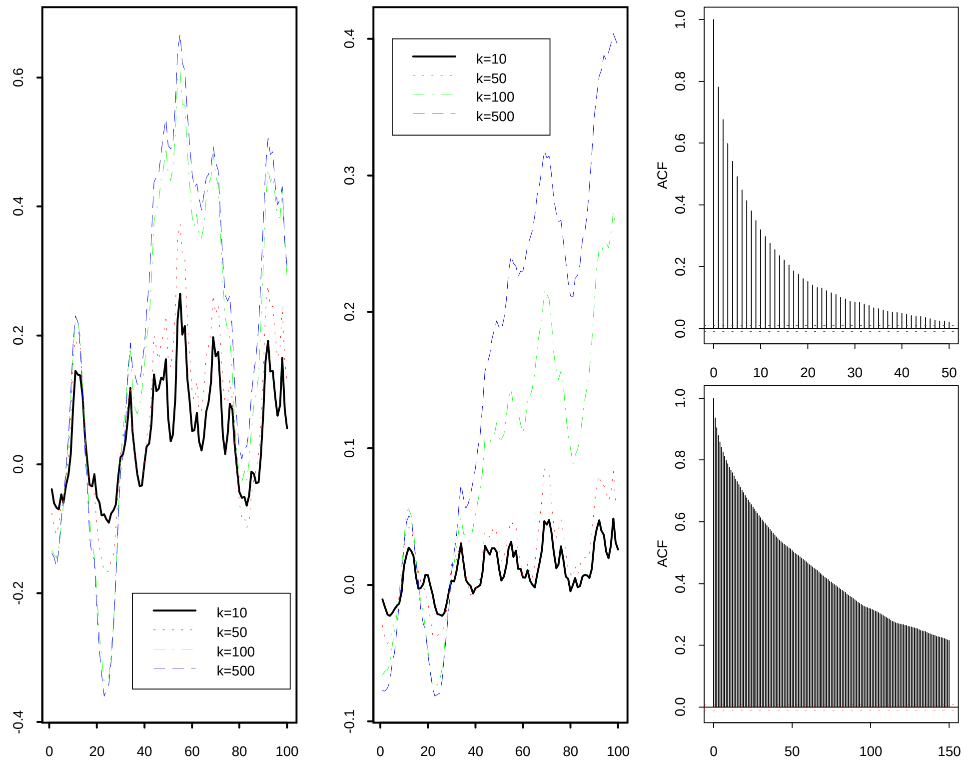

Figure 4: Single-Move algorithms. Weekly parameters ($\phi = 0.9, \sigma_\eta^2 = 0.05$), Daily parameters ($\phi = 0.99, \sigma_\eta^2 = 0.01$) and Corresponding Correlograms. Y-Axis: Log Volatility, X-Axis: Time.

provides evidence for our arguments. These report results obtained from repeating the above experiments, but now generalized to the SV model:

$$\begin{array}{rcllcl} y_t & = & \epsilon_t \exp\{(\alpha_t + \beta_t)/2\}, & \epsilon_t & \sim & NID(0,1), \\ \alpha_{t+1} & = & \phi\alpha_t + \eta_t, & \eta_t & \sim & NID(0,\sigma_\eta^2), \\ \beta_{t+1} & = & \beta_t + \zeta_t, & \zeta_t & \sim & NID(0,\sigma_\zeta^2), \end{array} \tag{6}$$

which allows the 'permanent' level of volatility to evolve according to a random walk. It analogs the Engle and Lee (1992) model of permanent and transitory volatility shocks for ARCH models. In our experiment we allow $\phi = 0.9$,$\sigma_\eta^2 = 0.05, \sigma_\zeta^2 = 0.01$. The value of σ_ζ^2 is perhaps too high, but it is chosen so that the single-move algorithm shows signs of converging: more realistic values would slow it down further. The first picture in Figure 5 shows the estimates of the β_t components for the multi-move algorithm; the single-move's results are in the middle graph. They suggest that single-move algorithms take a great deal of time to converge.

These observations are confirmed by the correlograms in Figure 5 which resulted from recording the draw from β_{50} for both samplers. This time the samplers were allowed

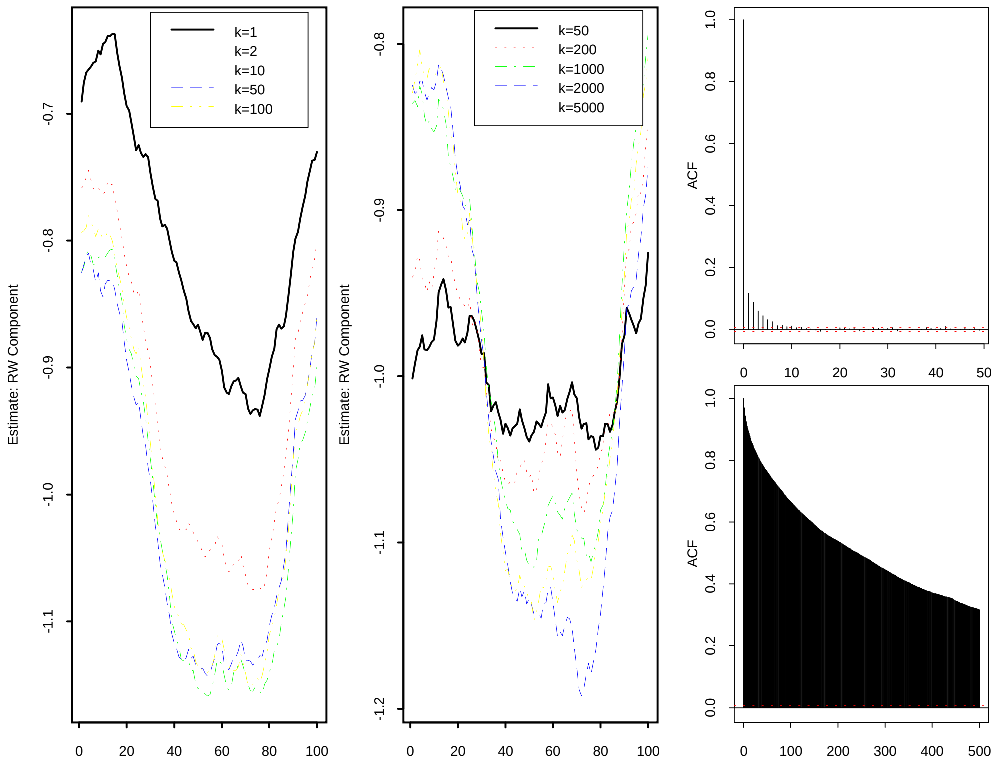

Figure 5: Multi and single-move algorithms for the random walk component. Top correlogram: the multi-move algorithm. Bottom: the single-move algorithm. Y-Axis: Log Volatility, X-Axis: Time.

to run for 100,000 iterations, discarding the first 10,000 draws. The correlograms show that the multi-move algorithm continues to exhibit low correlation, while the single-move algorithm's correlation has increased. The results of this show up in the estimate of the variance of the estimate of $E(\beta_{50}|Y_n)$ for the two samplers. It becomes 39.8 for the single-move algorithm, while it is 0.359 for the multi-move sampler. As σ_ζ^2 goes to 0 this ratio will go to infinity.

3 LIKELIHOOD BASED INFERENCE

3.1 Posterior distributions

It is straightforward to state the Bayesian approach to the problem of learning about the unknown parameters. We will write[10] $\theta = (\sigma_\eta^2, \phi)'$ and assume that we have a prior for θ

[10]The inclusion of explanatory variables raises no new issues, except that it is now wise to use the diffuse version of the De Jong and Shephard (1995) simulation smoother, which draws from the joint posterior of the states α and the regression parameters. In these calculations a diffuse prior is used for the regression parameters.

that we write $f(\theta)$. Then we can extend the multi-step sampler to cope with the unknown parameters in the following way:

Bayes Samplers

1. Generate a candidate point ω, θ.
2. Draw α^* from the density $f(\alpha|\omega, \theta, Y_T)$.
3. Draw ω^* from the density $\Pr(\omega|\alpha^*, \theta, Y_T)$.
4. Draw θ^* from the density $f(\theta|\alpha^*, \omega^*, Y_T)$.
5. Write $\omega = \omega^*, \theta = \theta^*$ and go to 2.

So long as the prior is always positive, this procedure will converge, providing replications from $\theta|Y_T$. The only new feature of the analysis is the posterior on the θ. The likelihood part of this is easily expressed as the model is constructed to have $\alpha_0 = 0$, which means

$$\begin{aligned} f(\theta|\alpha^*, \omega^*, Y_T) &= f(\theta|\alpha^*, \alpha_0 = 0) \\ &\propto f(\theta) f(\alpha^*, \alpha_0 = 0|\theta) \\ &= f(\theta) \textstyle\prod_{t=0}^{T} f(\alpha_{t+1}^*|\alpha_t^*, \theta). \end{aligned} \tag{7}$$

This posterior will take on particularly simple forms if the standard normal-inverse gamma conjugate priors are assumed for (ϕ, σ_η^2). Although constructing a prior for σ_η^2 is straightforward, the prior for ϕ is more difficult. Although the flat prior $f(\phi) \propto 1$ has some attractive properties, in particular allowing a conjugate analysis (Sims and Uhlig (1991) and Jacquier, Polson, and Rossi (1994)), it is now well known that this prior has some difficulties when the data is close to being non-stationary. Much econometric work has been carried out on this issue in recent years: prominent references to this literature include Phillips (1991) and Schotman and Van Dijk (1991). There is also a burgeoning literature on how priors can be formulated for stationary versions of these models and how posteriors can be computed using the Hastings-Metropolis algorithm. Important references include Chib and Greenberg (1994), Marriott, Ravishanker, Gelfand, and Pai (1992) and Marriott and Smith (1992).

In our work we use a proper prior, constrained so that the log volatility is stationary. A natural distribution to use for ϕ is a transformed beta, which has the log-density

$$\log f(\phi) = (\alpha - 1) \log \left\{ \frac{(1+\phi)}{2} \right\} + (\beta - 1) \log \left\{ \frac{(1-\phi)}{2} \right\}, \phi \in (-1, 1), \alpha > 1, \beta > 1. \tag{8}$$

This has the mean $E(\phi) = \{2\alpha/(\alpha+\beta)\} - 1$. Our analysis will be based on setting $\alpha = 20$ and $\beta = 1.5$, giving a prior mean of 0.86. As the prior is log concave we can sample from the posterior using the Wild and Gilks (1993) algorithm.

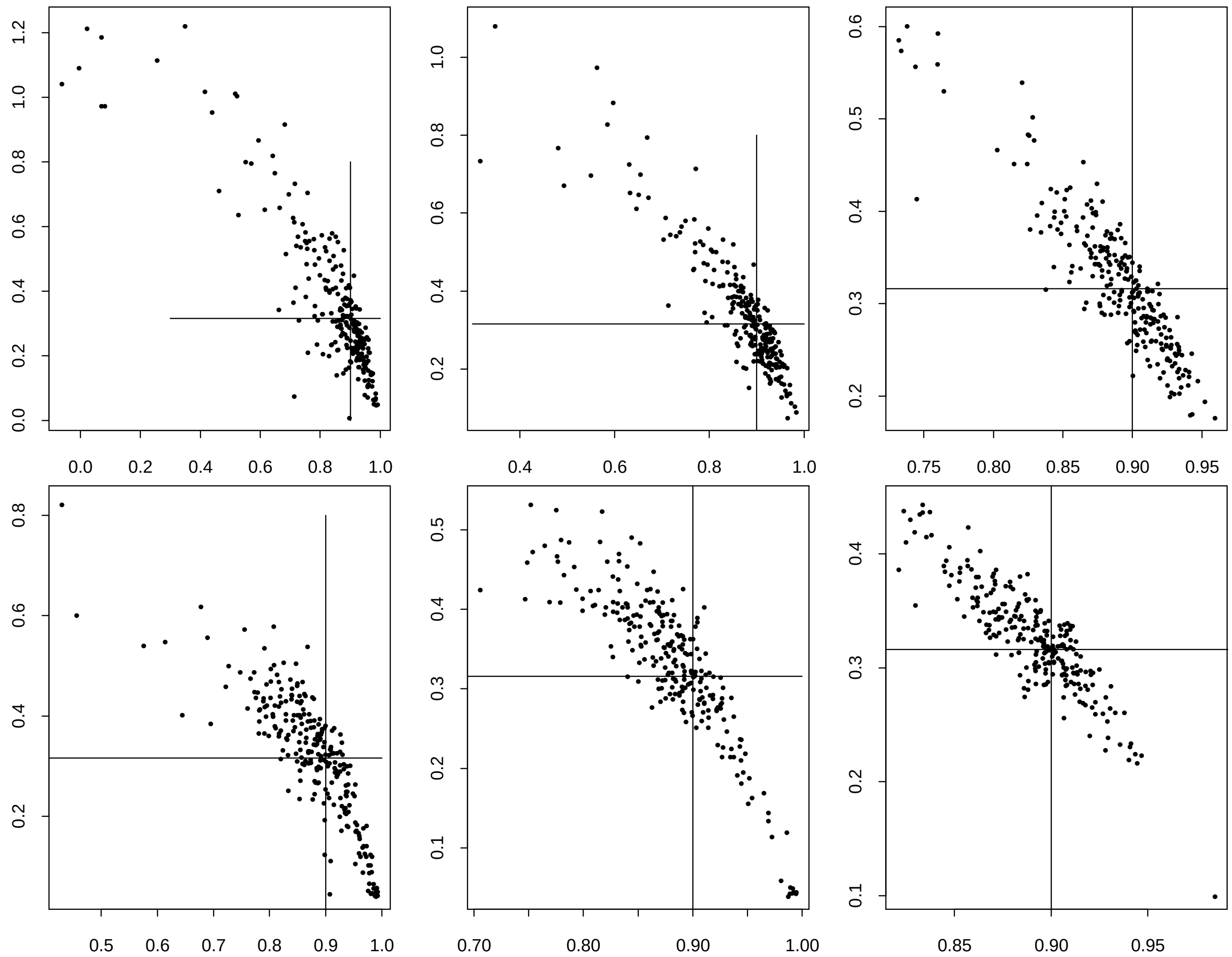

Figure 6: QML and Bayes estimate of SV model. QML on top, Bayes in bottom row. Uses $T = 500$, $T = 1000$, $T = 2000$. T goes from right to left. Y-Axis is $\widehat{\sigma_\eta}$ and on the X-axis is $\widehat{\phi}$.

Figure 6 reports some 250 simulation results which compare this Bayesian solution[11] to the QML estimator of Harvey, Ruiz, and Shephard (1994), using $\phi = 0.9$, $\beta = -0.5$ (β is the time invariant constant in the model – see (6)) and $\sigma_\eta^2 = 0.1$ and allowing the sample size to be $500, 1000$ and 2000. The results confirm the impression given in Jacquier, Polson, and Rossi (1994) of large efficiency gains for the likelihood method over the quasi-likelihood approximation. Further, the results seem not to have been too disturbed by taking a prior which has quite a different mean from the data, a natural result given the large sample sizes of the experiments.

[11]The Bayesian solution uses initial values $\phi = 0.9$ and $\sigma_\eta^2 = 0.1$, a burn in period of 100 iterations and averages the next 400 iterations to get the empirical Bayes estimate. QML uses the same initial values.

3.2 Maximum likelihood estimation

Although the Bayesian approach is straightforward to state, it requires the elicitation of a prior, which is sometimes regarded by some econometricans as being difficult in dynamic models. Although this is perhaps not an insurmountable problem, an alternative is available which allows us to perform maximum likelihood estimation. This involves using an EM algorithm, more specifically a simulated EM technique. The theory behind this approach is straightforward and is discussed in detail in Qian and Titterington (1991) (also Bresnahan (1981), Celeux and Diebolt (1985), Wei and Tanner (1990) and Ruud (1991)).

The EM algorithm exploits the decomposition:

$$\log f(Y_T;\theta) = \log \Pr(\omega) + \log f(Y_T|\omega;\theta) - \log \Pr(\omega|Y_T;\theta). \tag{9}$$

Since $f(Y_T|\omega;\theta)$ is a normal likelihood it can be computed using the Kalman filter given in Appendix and:

$$\log f(Y_T|\omega;\theta) = const - -\frac{1}{2}\sum_{t=1}^{T}\log F_t - \frac{1}{2}\sum_{t=1}^{T} v_t^2/F_t. \tag{10}$$

The EM algorithm iterates:

$$\theta^{(i+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta,\theta^i;Y_T) = \underset{\theta}{\operatorname{argmax}} \sum_{\omega} \log f(Y_T|\omega;\theta)\Pr(\omega|Y_T;\theta^i), \tag{11}$$

which guarantees that the likelihood will not decrease. The objective function for this optimisation problem can be estimated as

$$\frac{1}{M^i}\sum_{j=1}^{M^i}\log f(Y_T|\omega^j;\theta) \tag{12}$$

where ω^j is a simulated value from the MCMC chain $\omega|Y_T;\theta^i$. This is maximized with respect to the parameters θ using a numerical optimisation routine. Each numerical maximization step keeps the ω^j fixed until convergence is obtained, then a new ω vector is obtained through another simulation step.

It is clear that the simulated EM algorithm is not as clean as the Bayesian analysis for it involves two stages of 'approximation' — running the MCMC and then using only M^i replications — rather than one in the Bayesian case — just running the MCMC. This is more of a problem in simulating the estimators for the Monte Carlo experiments than it is for actually carrying out estimation using actual data. As a practical matter the convergence of this algorithm is very easy to monitor. One can easily design a visual, interactive estimation procedure in which one increases M as the process converges to less variable parameter estimates. While this type of approach has its merits, the construction of elegant methods for obtaining a simple maximum likelihood estimator that does not require such monitoring and control is still an open area of research.[12]

Some impression of the performance of the simulated EM algorithm can be found by inspecting Figure 7, which repeats the experiments conducted for the QML and empirical Bayes estimators reported in Figure 6. Figure 7 gives results using $M = 1$, $M = 3$ and $M = 10$, using 10 steps of the EM algorithm before reporting the results. In effect, the simulated EM algorithm takes $10 \times M$ longer on a computer than its QML counterpart, for the simulation takes a small amount of time compared to using a numerical optimisation routine. Applied work may require a more elaborate scheme for setting M, but the results presented should provide a good indication of the performance of the algorithm.

The results suggest that the simulated EM algorithm is at least competitive with the Bayes estimator, although it is always difficult to be very decisive as the Bayes solution is a function of the prior.

Finally more extensive simulation results are available in Table 6 which compares the QML procedure with the simulated EM algorithm for a variety of parameter values. Throughout these simulations we use 1000 replications. The simulation results indicate that most of the gains from using the simulation procedure accrue by taking $M = 1$. Although increasing M beyond this does allow for some gains, particularly in cases where the QML procedure behaves very poorly, sensitivity to M is not very great in general. This may not hold for more complicated models, with more components.

The numbers indicate that the use of the simulated EM algorithm yields large efficiency gains compared to QML as measured by the reduction in the mean squared errors. Gains of the order of 10 to 100 do not seem exceptional when T is small. For large samples QML actually performs very credibly in high persistence cases. This is because the QML's poor performance is due to the occasional occurrence of bad estimates generated by an inlier (a small value of ε_t, which produces a large negative $\log \varepsilon_t^2$) whose importance decreases as sample size increases. The simulated EM algorithm is much less susceptible to this difficulty because it is robustified by using the mixture representation.

[12] A recent paper which discusses the asymptotic theory of using the MCMC method in conjunction with the EM algorithm is Chan and Ledolter (1994). They suggest stopping rules for the procedure. Although this type of approach is in its infancy, it seems to hold some promise.

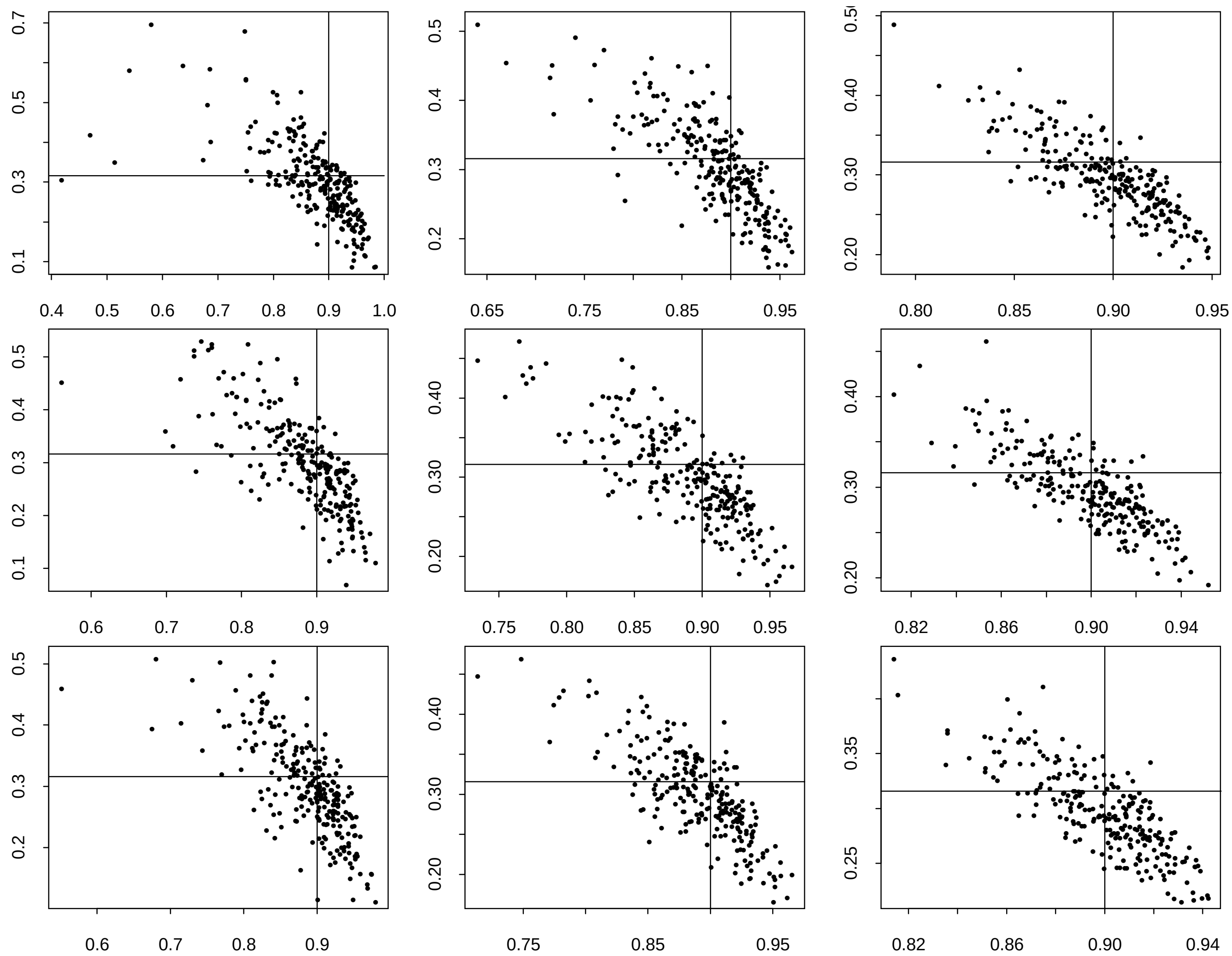

Figure 7: Simulating EM estimate of SV model. Uses $T = 500$, $T = 1000$, $T = 2000$ and $M = 1$, $M = 3$, $M = 10$. T goes from right to left. $M = 1$ in the top row. Y-Axis is $\widehat{\sigma_\eta}$ and on the X-axis is $\widehat{\phi}$.

3.3 Estimating the score and information

The simulated EM algorithm estimates the ML estimates. Likewise the score can be estimated using the EM algorithm results, as in Tanner (1991, pg. 36). That is,

$$\left.\frac{\partial \log L_y(\theta)}{\partial \theta}\right|_{\theta=\theta^*} = \int \left.\frac{\partial \log f(y|\omega;\theta)}{\partial \theta}\right|_{\theta=\theta^*} f(\omega|y;\theta^*)d\omega, \tag{13}$$

which yields the estimator

$$\left.\frac{\partial \log L_y(\theta)}{\partial \theta}\right|_{\theta=\theta^*} \doteq \frac{1}{M^i}\sum_{j=1}^{M^i} \left.\frac{\partial \log f(y|\omega^j;\theta)}{\partial \theta}\right|_{\theta=\theta^*}, \tag{14}$$

where the ω^j are drawn using the multi-move sampler from $\Pr(\omega|y;\theta^*)$. Of some interest is the availability of analytic expressions for $\partial \log f(y|\omega^j;\theta)/\partial\theta$ in the time series literature (Koopman and Shephard (1992)). We can easily work with this expression directly.

Furthermore, by using common initial values and fixing a pool of random numbers, it is possible to numerically differentiate this expression to obtain an estimate of the second derivative matrix — minus the observed information.[13] This provides a valid basis for inference if we take M^i to be large, and offers computational advantages by eliminating the need to compute numerical derivatives for the optimisation problem. The more conventional method of using numerical derivatives remains a valid option for carrying out the optimisation and for computing the scores.

3.4 Diagnostic checking and likelihood evaluation

Although there is now quite a large literature on the estimation of SV models, there is a dearth of work on checking their fit[14]. Diagnostic checks on the model are possible using the innovations from the fitted model, i.e., the one step ahead forecast errors. However they are not immediately available from the fitting exercise. Instead, we use a special algorithm to extract them using a multi-move sampler:

Diagnostic Simulator

1. Generate a candidate point ω.
2. Draw α_t^* from $f(\alpha_t|Y_t,\omega)$.
3. Draw ω^* from $f(\omega|\alpha_t^*, Y_t)$.
4. Simulate α_{t+1} from α_t^*.
5. Record $\Pr(y_{t+1}^* \leq y_{t+1}|\alpha_{t+1}^*)$.
6. Write $\omega = \omega^*$. Go to 2.

where y_{t+1} denotes the observed value of the series and y_{t+1}^* the corresponding random variable. By averaging the recorded probabilities over M iterations for $t = 1, \ldots, T$, we obtain estimates of the probability of exceeding the observed y_{t+1}:

$$u_{t+1}^M = \frac{1}{M}\sum_{i=1}^{M} \Pr(y_{t+1}^* \leq y_{t+1}|\alpha_{t+1}^i). \tag{15}$$

For each $t = 1, \ldots, T$, under the null of a correctly specified model, $u_t^M \xrightarrow{d} UID(0,1)$ variable. This provides a valid basis for diagnostic checking. These variables can be

[13] The Koopman and Shephard (1992) algorithm can be avoided completely by using numerical derivatives to evaluate the score and numerical second derivatives for the information. This is of use for complicated models which may not have simple derivatives when we condition on ω.

[14] The notable exception to this comment seems to be Gallant, Hsieh, and Tauchen (1994), who use a SNP model to fit and check the SV model.

mapped into the normal distribution, by using the inverse of the normal distribution function

$$n_t^M = F^{-1}(u_t^M) \xrightarrow{d} NID(0,1) \tag{16}$$

to give a standard sequence of normal variables, which are the transformed one step ahead forecasts normed by their correct standard errors. These can be used to carry out Box-Ljung, normality, and heteroscedasticity tests, among others.

The innovations can also be used to estimate the likelihood for the one-step-ahead density can be estimated as

$$\frac{1}{M}\sum_{i=1}^{M} f(y_{t+1}|\alpha_{t+1}^i), \tag{17}$$

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

4 COMPARISON WITH ARCH MODELS

In this section we will compare the fit of basic SV models with the GARCH models commonly used in the literature. The notation we use for the GARCH(1,1) model is:

$$y_t|Y_{t-1} \sim N(0,\sigma_t^2), \text{ where } \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1\sigma_{t-1}^2. \tag{18}$$

On occasions the normality assumption will be replaced by a student t formulation, following the work of Bollerslev (1987). We will write these models as t-GARCH, and use ν as the notation for the integer degrees of freedom.

The form of SV model estimated in this section will take on the parameterization:

$$\begin{array}{rclcrcl} y_t & = & \epsilon_t \exp[(\alpha_t+\beta)/2], & & \epsilon_t & \sim & NID(0,1), \\ \alpha_{t+1} & = & \phi\alpha_t + \eta_t, & & \eta_t & \sim & NID(0,\sigma_\eta^2). \end{array} \tag{19}$$

Thus β will play the role of a scale parameter. It can be arbitrarily determined by rescaling the observations. The simulated EM algorithm and Bayes procedures will estimate the unknown β, ϕ and σ_η^2 parameters.

Throughout this section, the reported simulated EM algorithm's are obtained by iterating the sampling procedure 50 times using $M = 50$. Our experiments with various starting values indicate that the results presented here are rather insensitive to their choice. On the other hand, we also report Bayes estimators of the model using 1,000 burn in iterations and 20,000 iterations to provide draws from $\theta|Y_n$.

4.1 Data description

We will look at two data sets in this comparison, both taken from the CRSP tape. The first is a weekly series of returns, to be denoted y_t, which we constructed from the daily return data on the S&P 500 index using Wednesday-to-Wednesday returns for the sample period 7/11/1962 to 12/30/1992: yielding 1591 observations.[15] The second uses the daily data reported in Nelson (1991), which runs from 7/03/1962 to 12/31/1987, yielding 6409 observations. Both series will be multiplied by 100, so that they represent percentage returns. This is important, for the techniques we use to handle the SV model can be sensitive to observations which are grossly inappropriately scaled.

The summary statistics for the weekly series are presented in Table 3. As is typical for this type of data, the squared returns exhibit much higher autocorrelations than the returns themselves and these autocorrelations decay slowly, indicating that the volatility process is quite persistent.

	Means	Std. Dev.	BL(32)	Autocorrelations at				
				lag 1	lag 2	lag 3	lag 4	lag 5
y_t	0.128	2.023	25.670	0.036	-0.005	0.029	-0.029	-0.021
y_t^2	4.093	10.00	355.16	0.296	0.142	0.053	0.039	0.046

Table 3: Summary Statistics for Weekly Returns on the S&P 500 Index. y_t denotes 100 times the returns, and y_t^2 denotes squared returns. The sample size is 1591 observations. BL(32) is the Box-Ljung statistic for 32 lags.

Finally, before we fit and check the various volatility models, the series are regressed against a constant and a lagged value of the series to remove the mean and autocorrelation in the levels of the series. This could have been carried out during the estimation of the SV and GARCH models. Carrying out the procedure later did not in fact change the results very much.

4.2 An empirical comparison: weekly data

The results of the fitting and model checking exercise, given in Table 4, suggest a slightly higher likelihood for the SV model than the Gaussian GARCH(1,1) model. It is difficult to be particularly rigorous in this comparison as the models are not nested, which means that

[15] Whenever Wednesday returns are unavailable, data from the next trading day are used. In almost all cases this amounts to using Thursday returns.

basing the usual likelihood ratio test on asymptotic χ^2 distributions is not particularly informative (Cox (1961)).

The Bayes and simulated EM algorithm's parameter estimates are very similar. The posterior densities are reported in Figure 8. The posterior densities, directly estimated from the draws of the multimove Gibbs sampler and using an Splus density estimator, are not particularly tightly peaked: the result of having only a moderate sample size for the weekly data. They are close to being symmetric, which partially explains the closeness of the empirical Bayes estimator and the simulated EM solution.

The estimated parameters for the SV and ARCH models are interesting. The persistence of the GARCH models is much higher than for the SV models — a half life of 74 weeks compared to 32 for the SV model. This seems to be a standard finding (Taylor (1994)) although it has yet to be explained.

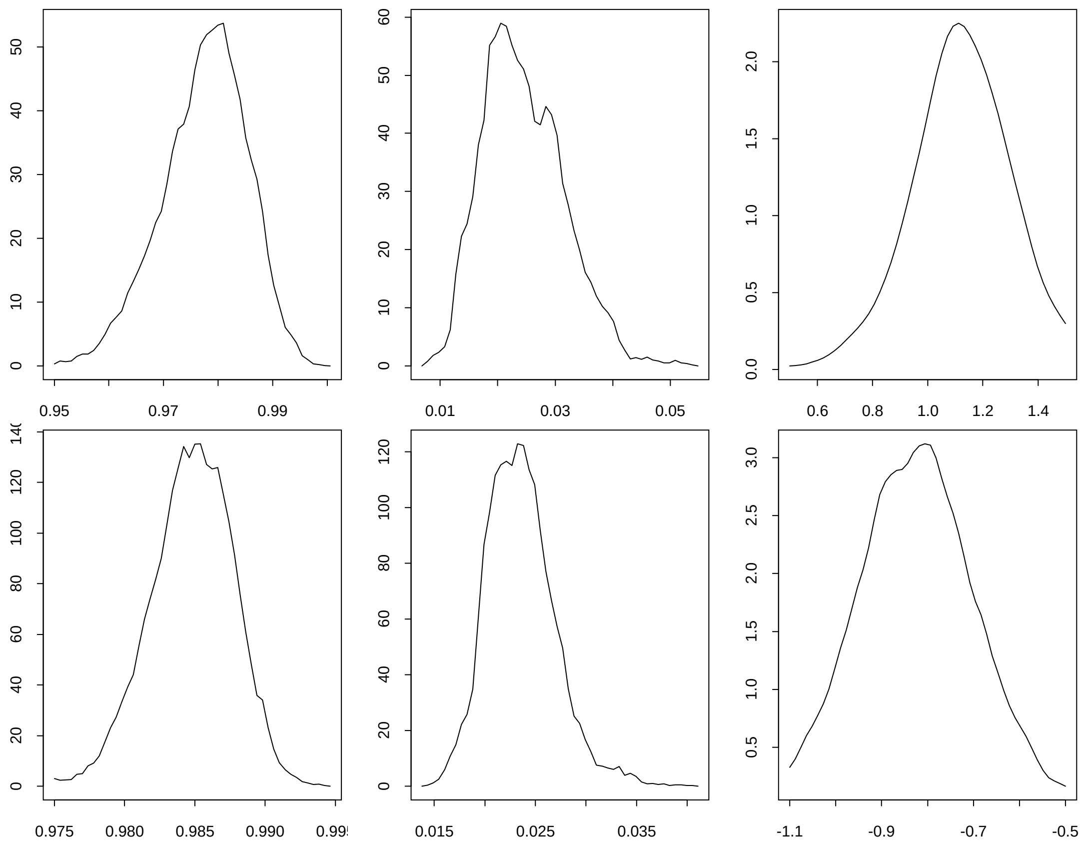

Figure 8: Posterior densities for ϕ, σ_η^2 and β for the weekly data (on top) and daily data (on bottom). Computed using the default options of the Splus density estimation function.

The square root of the estimated variance is given in Figure 9, which shows that the filtered and smoothed estimates from the SV model are almost identical to the results from the GARCH model. The same impression is gained from the Box-Ljung statistics for the squares of the transformed innovations from the model. Both are in the mid 40s, with 30 degrees of freedom — a reasonably satisfactory result. This suggests there is very little to choose between these models of changing variance.

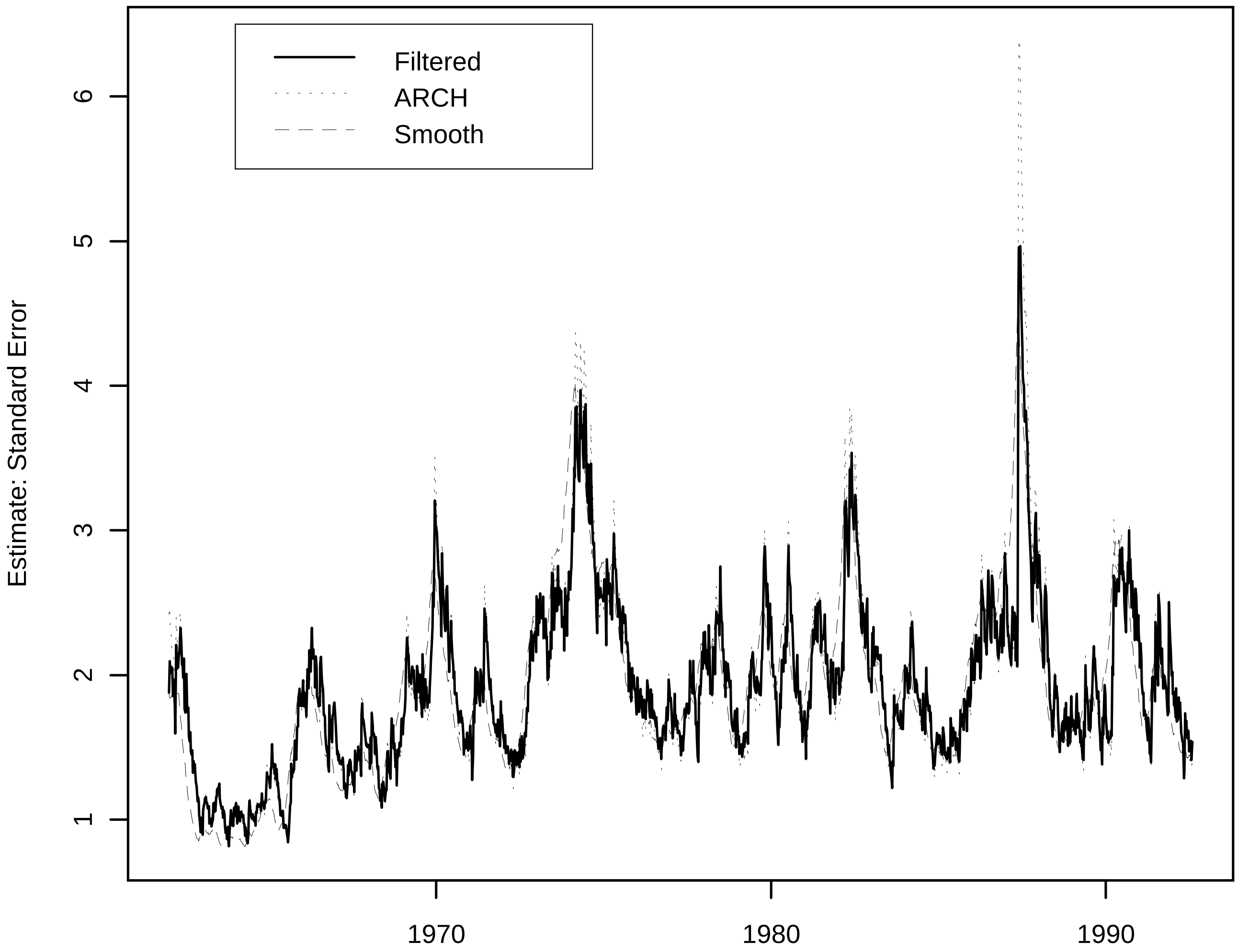

Figure 9: Estimate of the standard errors for the weekly data from 7/11/62 to 12/30/92 using SV and ARCH models.

The kurtosis statistics, reported in Table 4 are 4.4 and 3.0 for the GARCH and SV models respectively. This implies that the generalisation to heavy tailed GARCH and SV models is probably not too important for this data set. This point is reinforced in Table 4, which shows that the estimated t-GARCH model has an almost identical likelihood as the normal based SV model, although the difference is now smaller.

Interestingly, the resulting kurtosis statistic for the t-GARCH model is now satisfactory. It is important to understand how we computed this. As the residuals are t-distributed we mapped them into uniforms using the t-distribution and then mapped them into normals using the inverse distribution function of the standard normal dis-

		α_0	α_1	β_1	ν	BL(30)	Kurtosis	$\log L$
GARCH		0.0539	0.1069	0.8843		34.0	4.41	-3243.9
t-GARCH		0.0558	0.0976	0.8916	14	40.8	3.06	-3231.8
		β	ϕ	σ_η^2		BL(30)	Kurtosis	$\log L$
SV	Sim EM	1.1213	0.9785	0.0230		39.7	3.04	-3231.9
	Bayes	1.1273	0.9775	0.0245				
		[.727,1.522]	[.960,.991]	[.013,.0395]				

Table 4: Estimation Results: 100 times Weekly Returns on the S&P 500 Index. The first row for each model presents the parameter estimates. The second row contains the standard errors. BL(30) is the Box-Ljung statistic (30 lags) for the squared residuals transformed into normality from the fitted model. Kurtosis is the corresponding fourth moment. The figures in [] are Bayesian symmetric 95% intervals.

tribution. The kurtosis statistic is then used on this mapping. Taken together, Table 4 suggests the SV and t-GARCH give almost identical fits to this data. This point is reinforced by Figure 9 which displays the filtered forecasts of the standard error for the SV model and that computed by the ARCH model. It is difficult to find any difference between the two of these lines. The smoothed estimate from the SV model, though, is rather different.

4.3 An empirical comparison: daily data

The above analysis is now repeated for the daily data and the results are given in Table 5. Again it suggests a higher likelihood for the SV model than the Gaussian GARCH(1,1) model, while the Bayes and simulated EM give very similar results (see Figure 8). The half life of the volatility estimates now becomes 168 and 46 days for the ARCH and SV models respectively.

The Box-Ljung statistics for the squares of the transformed innovations from the model are both in the low 50s, with 32 degrees of freedom, a reasonably satisfactory result. The corresponding kurtosis statistics are 5.16 and 3.69, with the GARCH model performing rather poorly in the tails of the distribution. The usual response to this is to use the t-GARCH model. Its estimated parameters are in Table 5. This model has a higher likelihood than the SV model and corrects the kurtosis statistic. This suggests t-GARCH is a better empirical model of the data and indicates a heavy tailed SV model may be important.

		α_0	α_1	β_1	ν	BL(30)	Kurtosis	$\log L$
GARCH		0.0052	0.0980	0.8979		43.6	5.16	-6788.6
t-GARCH		0.0044	0.0823	0.9132	9	48.8	3.04	-6688.8
		β	ϕ	σ_η^2		BL(30)	Kurtosis	$\log L$
SV	Sim EM	-0.8272	0.9849	0.0231		55.3	3.69	-6710.5
	Bayes	-0.8255	0.9848	0.0236				
		[-.563,-1.083]	[.979,.990]	[.0176,.0314]				

Table 5: Estimation Results: 100 times Daily Returns on the S&P 500 Index: Nelson's data set. The first row for each model presents the parameter estimates. The second row contains the standard errors. BL(30) is the Box-Ljung statistic (30 lags) for the squared residuals transformed into normality from the model. The figures in [] are Bayesian symmetric 95% intervals.

5 CONCLUSION

In this paper we have described a new simulation based strategy for estimating general specifications of stochastic volatility models. The key element of the procedure is the linearization made possible by a log-square transformation of the measurement equation and the approximation of a log χ^2 random variable by a mixture of normal variables. This enables the analysis of complex models using the well-established methods for working with Gaussian state-space models.

The simulation conducted in this paper shows that a significant efficiency gain over previously proposed methods for estimating stochastic volatility models is possible, particularly for small sample sizes.

Our empirical work suggests that the simple SV models fit the data as well as the popular ARCH models used in the econometrics literature. Work continues to refine these results, fitting more sophisticated models.

Stochastic volatility models have long attracted the attention of researchers in finance. Our efficient estimation algorithm promises to broaden the scope of the models that can be contemplated for empirical applications.

6 ACKNOWLEDGEMENT

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. We would like to thank the ESRC and Princeton University for their financial support. The comments of Peter Rossi on a previous version of this paper were particularly helpful. The comments of the participants of the 'Stochastic volatilty' conference of October 1994, held at HEC (Université de Montreal), are gratefully acknowledged.

7 APPENDIX

7.1 Gaussian state space model

In this section we discuss general filtering and simulation smoothing results which are useful for a general state space model. We analyse the multivariate model:

$$\begin{array}{lcl} y_t & = & Z_t \alpha_t + G_t u_{1t}, \\ \alpha_{t+1} & = & T_t \alpha_t + H_t u_{2t}, \\ \alpha_1 | Y_0 & \sim & N(a_{1|0}, P_{1|0}), \end{array} \quad \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \sim NID(0, I). \tag{20}$$

For simplicity we assume that $G_t' H_t = 0$ and we write the non-zero rows of H_t as M_t, $G_t G_t' = \Sigma_{\varepsilon t}$ and $H_t H_t' = \Sigma_{\eta t}$. An example of this is an AR(2) with measurement error. Using an obvious notation:

$$z_t = \begin{pmatrix} 1 & 0 \end{pmatrix}, G_t = \begin{pmatrix} \sigma_\varepsilon & 0 \end{pmatrix}, T_t = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix}, H_t = \begin{pmatrix} 0 & \sigma_\eta \\ 0 & 0 \end{pmatrix}, M_t = \begin{pmatrix} 0 & \sigma_\eta \end{pmatrix}.$$

7.2 Kalman filter

The Kalman filter is run for $t = 1, ..., T$,

$$\begin{array}{lcllcllcl} a_{t+1|t} & = & T_t a_{t|t-1} + K_t v_t, & P_{t+1|t} & = & T_t P_{t|t-1} L_t' + \Sigma_{\eta t}, & v_t & = & y_t - Z_t a_{t|t-1}, \\ F_t & = & Z_t P_{t|t-1} Z_t' + \Sigma_{\varepsilon t}, & K_t & = & T_t P_{t|t-1} Z_t' F_t^{-1}, & L_t & = & T_t - K_t Z_t. \end{array} \tag{21}$$

7.3 Simulation smoother

Early sampling methods for drawing from $\alpha | Y_T$ were proposed by Fruhwirth-Schnatter (1994), Chib (1992) and Carter and Kohn (1994). We exploit the approach of De Jong and Shephard (1995), which is computationally simpler, more efficient and avoids singularities. This requires that F_t^{-1}, v_t and K_t be stored from the run of the Kalman filter.

Setting $r_T = 0$ and $N_T = 0$, we run for $t = T, ..., 1$

$$\begin{array}{lcllcl} C_t & = & M_t M_t' - M_t H_t' N_t H_t M_t', & r_{t-1} & = & Z_t' F_t^{-1} v_t + L_t' r_t - L_t' N_t H_t M_t' C_t^{-1} \kappa_t, \\ \kappa_t & \sim & N(0, C_t), & N_{t-1} & = & Z_t' F_t^{-1} Z_t + L_t' N_t L_t + L_t' N_t H_t M_t' C_t^{-1} M_t H_t' N_t L_t \end{array} \tag{22}$$

storing the simulated $M_t u_t$ as $\widehat{M_t u_t} = M_t H_t' r_t + \kappa_t$. It will then be convenient to add to $\widehat{M_t u_t}$ the corresponding zero rows so that we simulate from the whole $H_t u_t$ vector (recall

H_t is M_t plus some rows of zeros). We will write this as $\widehat{\eta_t}$. The end condition $\widehat{\eta_0}$ is calculated by

$$\begin{array}{rcl} C_0 & = & P_{1|0} - P_{1|0} N_0 P_{1|0}, \quad \kappa_0 \quad \sim \quad N(0, C_0), \\ \widehat{\eta_0} & = & P_{1|0} r_0 + \kappa_0. \end{array} \tag{23}$$

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

$$\alpha_{t+1} = T_t \alpha_t + \widehat{\eta_t}, t = 0, ..., n-1. \tag{24}$$

References

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

Andersen, T. and B. Sorensen (1994). GMM estimation of a stochastic volatility model: a Monte Carlo study. Working paper: Northwestern University.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *J. Econometrics 51*, 307–327.

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

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *J. Econometrics 52*, 5–59.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*. North-Holland. Forthcoming.

Bresnahan, T. F. (1981). Departures from marginal-cost pricing in the American automobile industry, estimates from 1977-1978. *J. Econometrics 17*, 201–227.

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

Carter, C. K. and R. Kohn (1993). On the applications of markov chain monte carlo methods to linear state space models. *Proceeding of the American Statistical Association, Business and Economic Statistics Section*. Forthcoming.

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

Celeux, G. and J. Diebolt (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comput. Statist. Quarter. 2*, 73–82.

Chan, K. S. and J. Ledolter (1994). Monte Carlo EM estimation for time series models involving counts. *J. Am. Statist. Assoc. 89*. Forthcoming.

Chesney, M. and L. O. Scott (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. (1992). Bayes inference in the Tobit censored regression model. *J. Econometrics 51*, 79–99.

Chib, S. (1993). Bayes regression with autoregression errors: A Gibbs sampling approach. *J. Econometrics 58*, 275–294.

Chib, S. and E. Greenberg (1994). Bayes inference for regression models with ARMA(p,q) errors. *J. Econometrics*. Forthcoming.

Cox, D. R. (1961). Tests of seperate families of hypotheses. *Proceedings of the Berkeley Symposium 4*, 105–123.

Danielsson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *J. Econometrics 61*, 375–400.

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

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

De Jong, P. and N. Shephard (1995). Efficient sampling from the smoothing density in time series models. *Biometrika 82*. Forthcoming.

Duffie, D. and K. J. Singleton (1993). Simulated moments estimation of markov models of asset proces. *Econometrica 61*, 929–952.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F. and G. G. J. Lee (1992). A permanent and transitory component model of stock return volatility. Working paper: Department of Economics, UCSD.

Engle, R. F. and G. G. J. Lee (1994). Estimating diffusion models of stochastic volatility. Working paper: Department of Economics, UCSD.

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

Gallant, A. R., D. Hsieh, and G. Tauchen (1994). Estimation of stochastic volatility models with diagnostics. Working paper, Dept. of Economics, Duke.

Gallant, A. R., P. E. Rossi, and G. Tauchen (1992). Stock prices and volume. *Rev. Financial Studies 5*, 199–242.

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

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

Geweke, J. (1994). Bayesian comparison of econometric models. Working paper: Federal Reserve Bank of Minneapolis.

Ghysels, E. and J. Jasiak (1994). Stochastic volatility and time deformation: an application of trading volume and leverage effects. Working paper: Universite de Montreal.

Gilks, W. R. and P. Wild (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics 41*, 337–348.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *J. Appl. Econometrics 8*, S85–S118.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

Hamilton, J. (1994). *Time Series Analysis.* Princeton: Princeton University Press.

Harvey, A. C. (1994). Long memory and stochastic volatility. Submitted.

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

Hastings, W. K. (1970). Monte-carlo sampling methods using markov chains and their applications. *Biometrika 57*, 97–109.

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

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. *J. Economics and Business Statist. 12*. Forthcoming.

Jacquier, E., N. G. Polson, and P. E. Rossi (1995). Multivariate and prior distributions for stochastic volatility models. Unpublished paper, Graduate School of Business, University of Chicago.

Kim, S. and N. Shephard (1994). Comment on 'Bayesian analysis of stochastic volatility models' by Jacquier, Polson and Rossi. *J. Economics and Business Statist. 11*.

Koopman, S. J. and N. Shephard (1992). Exact score for time series models in state space form. *Biometrika 79*, 823–6.

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

Mahieu, R. and P. Schotman (1994a). Neglected common features in exchange rate volatility. *J. Empirical Finance 1*, 279–311.

Mahieu, R. and P. Schotman (1994b). Stochastic volatility and the distribution of exchange rate news. Working paper: University of Limburg.

Marriott, J., N. Ravishanker, A. Gelfand, and J. Pai (1992). Bayesian analysis of ARMA processes: complete sampling based inference under full likelihoods. Research Report, Department of Statistics, University of Connecticut.

Marriott, J. M. and A. F. M. Smith (1992). Reparameterization aspects of numerical Bayesian methodology for autoregressivce moving-average models. *J. Time Series Analysis 13*, 327–343.

McCulloch, C. E. and P. E. Rossi (1994). An exact likelihood analysis of the multinomial probit model. *J. Econometrics 64*, 207–240.

McCulloch, R. E. and R. S. Tsay (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *J. Am. Statist. Assoc. 88*, 968–978.

McCulloch, R. E. and R. S. Tsay (1994). Bayesian analysis of autoregressive time series via Gibbs sampling. *J. Time Series Analysis 15*, 235–250.

Melino, A. and S. M. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *J. Econometrics 45*, 239–265.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *J Chemical Physics 21*, 1087–1092.

Nelson, D. B. (1988). The time series behaviour of stock market volatility and returns. Unpublished Ph.D.: MIT.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica 59*, 347–370.

Phillips, P. C. B. (1991). To criticise the critics: an objective Bayesian analysis of stochastic trends. *J. Appl. Econometrics 6*, 333–364.

Qian, W. and D. M. Titterington (1991). Estimation of parameters in hidden Markov Chain models. *Phil. Trans. R. Soc. Lond. A. 337*, 407–428.

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

Ruud, P. (1991). Extensions of estimation methods using the EM algorithm. *J. Econometrics 49*, 305–341.

Schotman, P. C. and H. K. Van Dijk (1991). A Bayesian analysis of the unit root in real exchange rates. *J. Econometrics 49*, 195–238.

Shephard, N. (1993). Fitting non-linear time series models, with applications to stochastic variance models. *J. Appl. Econometrics 8*, S135–S152.

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

Sims, C. and H. Uhlig (1991). Understanding unit rooters: A helicopter tour. *Econometrica 59*, 1591–1599.

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

Tanner, M. A. (1991). *Tools for Statistical Inference, Observed Data and Data Augmentation Methods*, Volume 67. New York: Springer-Verlag.

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. (1993). Markov Chains for exploring posterior distributions. *Ann. Statist. 21.* Forthcoming.

Titterington, D. M., A. F. M. Smith, and U. E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions.* Chichester: Wiley.

Vetzal, K. (1992). Stochastic short rate volatility and the pricing of bonds and bond options. University of Toronto, Ph.D. Thesis.

Watanabe, T. (1992). A nonlinear filtering technique for estimating stochastic volatility models with an application to daily returns in the New York stock exchange. Mimeo: Dept. of Economics, Yale University.

Wei, G. C. G. and M. A. Tanner (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Am. Statist. Assoc. 85*, 699–704.

Wild, P. and W. R. Gilks (1993). AS 287: Adaptive rejection sampling from log-concave density functions. *Applied Statistics 42*, 701–709.

M		T=500	T=1000	T=3000		T=500	T=1000	T=3000
0	.995	.943 (.184)	.983 (.069)	.994 (.004)	.95	.686 (.460)	.767 (.376)	.862 (.240)
1		.972 (.090)	.989 (.019)	.993 (.004)		.834 (.267)	.898 (.145)	.932 (.046)
3		.975 (.088)	.990 (.013)	.993 (.004)		.890 (.165)	.918 (.086)	.937 (.031)
10		.983 (.041)	.990 (.011)	.993 (.004)		.909 (.119)	.924 (.072)	.937 (.029)
0	.005	.043 (.200)	.012 (.055)	.005 (.002)	.01	.326 (.133)	.086 (.212)	.046 (.117)
1		.012 (.028)	.007 (.009)	.006 (.002)		.053 (.033)	.023 (.039)	.015 (.012)
3		.009 (.018)	.006 (.006)	.006 (.002)		.033 (.023)	.019 (.033)	.014 (.009)
10		.008 (.013)	.006 (.005)	.006 (.002)		.028 (.021)	.018 (.030)	.014 (.008)
0	.995	.980 (.063)	.990 (.026)	.994 (.003)	.95	.932 (.069)	.942 (.038)	.947 (.013)
1		.986 (.023)	.991 (.008)	.994 (.003)		.943 (.033)	.948 (.019)	.950 (.010)
3		.987 (.022)	.991 (.008)	.994 (.003)		.944 (.028)	.948 (.018)	.950 (.009)
10		.987 (.021)	.992 (.007)	.994 (.003)		.945 (.027)	.948 (.017)	.950 (.009)
0	.010	.021 (.069)	.013 (.035)	.011 (.003)	.1	.183 (.141)	.117 (.093)	.104 (.029)
1		.013 (.012)	.011 (.007)	.010 (.003)		.048 (.098)	.094 (.032)	.091 (.019)
3		.012 (.010)	.011 (.006)	.010 (.002)		.040 (.095)	.093 (.028)	.091 (.017)
10		.012 (.010)	.011 (.006)	.010 (.002)		.038 (.094)	.093 (.027)	.091 (.017)
0	.975	.839 (.322)	.929 (.160)	.967 (.046)	.9	.830 (.209)	.869 (.121)	.891 (.042)
1		.939 (.111)	.961 (.054)	.970 (.014)		.882 (.079)	.895 (.047)	.902 (.023)
3		.948 (.085)	.964 (.042)	.970 (.012)		.888 (.066)	.898 (.036)	.902 (.019)
10		.951 (.080)	.965 (.040)	.971 (.011)		.891 (.057)	.899 (.034)	.902 (.018)
0	.010	.107 (.314)	.047 (.162)	.015 (.037)	.1	.373 (.223)	.152 (.193)	.113 (.062)
1		.023 (.057)	.016 (.041)	.012 (.007)		.086 (.106)	.096 (.050)	.091 (.025)
3		.019 (.057)	.015 (.039)	.012 (.005)		.078 (.101)	.094 (.037)	.090 (.022)
10		.018 (.049)	.015 (.040)	.012 (.005)		.070 (.098)	.093 (.034)	.090 (.021)

Table 6: Monte Carlo Results: QML and Simulated EM Algorithm. 1000 replications using starting values $\phi = 0.95$ and $\sigma_\eta^2 = 0.02$. Numbers in parentheses denote root mean squared errors.