Bayesian analysis of stochastic volatility models with fat-tails and correlated errors

Eric Jacquier[a,*], Nicholas G. Polson[b], Peter E. Rossi[b]

[a] *CIRANO, HEC Montréal, 3000 Cote Sainte-Catherine, Montréal, Canada PQ H4A 3L4*
[b] *The University of Chicago Graduate School of Business, 1101 East 58th Street, Chicago, IL 60637, USA*

Abstract

The basic univariate stochastic volatility model specifies that conditional volatility follows a log-normal auto-regressive model with innovations assumed to be independent of the innovations in the conditional mean equation. Since the introduction of practical methods for inference in the basic volatility model (J. Business Econom. Statist. 12(4) 371), it has been observed that the basic model is too restrictive for many financial series. We extend the basic SVOL to allow for a so-called "leverage effect" via correlation between the volatility and mean innovations, and for fat-tails in the mean equation innovation. A Bayesian Markov Chain Monte Carlo algorithm is developed for the extended volatility model. Thus far, likelihood-based inference for the correlated SVOL model has not appeared in the literature. We develop Bayes Factors to assess the importance of the leverage and fat-tail extensions. Sampling experiments reveal little loss in precision from adding the model extensions but a large loss from using the basic model in the presence of mis-specification. There is overwhelming evidence of a leverage effect for weekly and daily equity indices. The evidence in favor of fat-tails is very strong for daily exchange rate and equity indices, but less so for weekly data. We also find that volatility estimates from the extended model are markedly different from those produced by the basic SVOL.
© 2003 Elsevier B.V. All rights reserved.

---

* Corresponding author.
  E-mail address: eric.jacquier@hec.ca (E. Jacquier).

1. Introduction

Stochastic volatility (hereafter, SVOL) models offer a natural alternative to the GARCH family of time-varying volatility models. SVOL models allow for separate error processes for the conditional mean and conditional variance. The basic SVOL model specifies a log-normal auto-regressive process for the conditional variance with independent innovations in the conditional mean and conditional variance equations. There is evidence that SVOL models offer increased flexibility over the GARCH family, e.g. Geweke (1994b) and Fridman and Harris (1998).

The basic SVOL model can be extended in two natural ways: (1) with a fat-tailed distribution of the conditional mean innovations and (2) with a "leverage" effect, in which the conditional mean and conditional variance innovations are correlated. Evidence in favor of fat-tails has been uncovered by Gallant et al. (1997) and Geweke (1994c). Negative correlations between mean and variance errors can produce a "leverage" effect in which negative (positive) shocks to the mean are associated with increases (decreases) in volatility. This effect has been documented by Black (1976). It is captured in the EGARCH approach of Nelson (1991) and the modified GARCH of Glosten et al. (1993). Dumas et al. (1998) report that volatility estimates implied by index option prices are negatively correlated with the underlying index return.

Jacquier et al. (1994) (hereafter, JPR) develop a Bayesian Markov Chain Monte Carlo method for the basic SVOL model. This allows for exact finite sample inference, prediction and smoothing. Here we extend their method to handle SVOL models with both fat-tailed and correlated errors, providing the first likelihood-based procedure for SVOL models with correlated errors. We develop Bayes factors to assess the extent of sample evidence in favor of these extensions. The direct evaluation of Bayes factors as ratios of marginal likelihoods is computationally intensive and can be numerically unstable for latent variable models. Rather, we show how to rewrite the marginal likelihood ratios as simple functions of posterior quantities. Therefore we only require the already produced MCMC posterior output to compute Bayes factors.

To document evidence in support of the extended SVOL model as well as to investigate performance issues, both simulated and equity index/exchange rate series are analyzed. Using simulated data, we show that assuming the basic SVOL specification in the presence of fat-tails has severe consequences for both parameter and volatility estimation. All but one of the financial series show strong evidence of fat-tailed conditional mean errors, albeit weaker for the weekly series. All equity indices, weekly and daily, display a leverage effect. The Bayes factors display overwhelming evidence in favor of the extensions of the basic SVOL model. The extended SVOL model produces estimates and forecasts of volatility that are markedly different from the basic model, showing that the extensions are of practical relevance.

The paper is organized as follows. Section 2 outlines the basic SVOL model and extensions with details of the MCMC algorithm and Bayes Factor computations. Section 3 documents the performance of the method in parameter estimation and smoothing or volatility inference. Section 4 presents the results of application of the model to financial time series. Section 5 provides evidence on prior sensitivity analysis and convergence of the MCMC method. Section 6 offers concluding remarks.

2. Extensions of the basic SVOL model

2.1. The basic model

2.1.1. Model

This paper presents extensions of the basic log-normal autoregressive SVOL below:

$$y_t = \sqrt{h_t}\varepsilon_t,$$

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T,$$

$$(\varepsilon_t, v_t) \sim N(0, I_2). \tag{1}$$

Here I_2 is the two-dimensional identity matrix. Let ω denote the vector of parameters of the basic SVOL, $(\alpha, \delta, \sigma_v)$, where α is the intercept, δ is the volatility persistence and σ_v is the standard deviation of the shock to $\log h_t$. We can view the model as a hierarchical structure of three conditional distributions, $p(\mathbf{y}|\mathbf{h})$, $p(\mathbf{h}|\omega)$, $p(\omega)$, where $\mathbf{y}$ and $\mathbf{h}$ are the vectors of data and volatilities. The distribution of the data given the volatilities is $p(\mathbf{y}|\mathbf{h})$. Beliefs about the evolution of volatility, here a log-normal AR(1) are modeled by $p(\mathbf{h}|\omega)$ models. $p(\omega)$ reflects beliefs about the parameters of the volatility process. In both the basic and the extended SVOL, the support of the prior $p(\omega)$ restricts $\log h_t$ to the region of stationarity.

2.1.2. Priors

We use a Normal–Gamma prior for $p(\omega)$ as in standard Bayesian analysis of regression models.

For σ_v, we use $p(\sigma_v) \propto e^{-v_0 s_0^2 / 2\sigma^2} / \sigma^{v_0+1}$, an inverse gamma with $v_0 = 1$ degrees of freedom and a very small sum of squares of $s = 0.005$. This is a very flat prior over the relevant posterior range.

We use $\alpha \sim N(0, 100)$ and $\delta \sim N(0, 10)$. α and δ are a priori independent. The prior on δ is essentially flat over $[0, 1]$. We impose stationarity for $\log h_t$ by truncating the prior of δ. Christoffersen and Diebold (1997) find evidence in favor of stationarity in a model free context. Here we believe for substantive reasons that a non-stationary stochastic volatility is unrealistic. For example, it would mean that portfolio managers should permanently re-balance their portfolios, or traders permanently adjust long-term option values, after a volatility shock. Other priors for δ are possible. Geweke (1994a) proposes alternate priors to allow the formulation of odds ratios for non-stationarity. Kim et al. (1998) center an informative Beta prior around 0.9. We use a flat prior as we have no prior view on δ, other than stationarity.

2.1.3. Algorithm

The MCMC algorithm JPR developed for the basic SVOL model exploits the hierarchical structure of the model by augmenting the parameter space to include $\mathbf{h}$. Using a proper prior for $(\mathbf{h}, \omega)$, the JPR MCMC algorithm provides inference about the joint posterior $p(\mathbf{h}, \omega|\mathbf{y})$. As noted in JPR (1994), the transition kernel is irreducible and aperiodic, and, due to the time reversibility of the general Metropolis chain, the resulting posterior is the unique invariant distribution of the chain, see also Tierney (1994).

This will hold for the extensions considered here as the additional parameters have proper priors. Thus the chain for the extensions has the same convergence properties as in JPR (1994).

MCMC algorithms such as *Gibbs* and *Metropolis* construct Markov chains with equilibrium distribution equal to the joint posterior distribution of the parameters given the data. Consider a partition of a parameter vector θ into r blocks. The Gibbs algorithm draws the $n+1$st θ_i from the conditional $p(\theta_i^{(n+1)}|\theta_1^{(n+1)},\ldots,\theta_{i-1}^{(n+1)},\theta_{i+1}^{(n)},\ldots,\theta_r^{(n)})$ or $p(\theta_i^{(n+1)}|\theta_{-i}^{(n)})$, for $i=1,\ldots,r$. For the basic SVOL in (1), JPR break the joint posterior density $p(\omega,\mathbf{h}|\mathbf{y})$ into two Gibbs blocks: $p(\omega|\mathbf{h},\mathbf{y})$ and $p(\mathbf{h}|\omega,\mathbf{y})$.

- $p(\omega|\mathbf{y},\mathbf{h})$: $p(\omega|\mathbf{y},\mathbf{h})=p(\omega|\mathbf{h})$ is the posterior from a linear regression. Using standard analytical results, direct draws can be made. Stationarity is imposed via our prior on δ. We simply reject MCMC draws larger than one. Such draws are truly exceptional, occurring at a rate of less than one in several hundred thousand.
- $p(\mathbf{h}|\omega,\mathbf{y})$: JPR break $p(\mathbf{h}|\omega,\mathbf{y})$ into T univariate conditional distributions

$$p(h_t|h_{t-1},h_{t+1},\alpha,\delta,\sigma_v,\mathbf{y}) \propto \frac{1}{h_t^{0.5}} \exp\frac{-y_t^2}{2h_t} \times \frac{1}{h_t} \exp\frac{-(\log h_t - \mu_t)^2}{2\sigma^2},$$

$$t=1,\ldots,T, \tag{2}$$

where $\mu_t=(\alpha(1-\delta)+\delta(\log h_{t+1}+\log h_{t-1}))/(1+\delta^2)$ and $\sigma^2=\sigma_v^2/(1+\delta^2)$. Let p denote the un-normalized kernel in (2). Here we augment $\mathbf{h}$ with h_0,h_{T+1}. We integrate out h_0 and h_{T+1} by drawing from the AR(1) in (1), see Geweke (1994c). Given the time reversibility of the AR(1), we draw $\log h_0$ by drawing $v_0 \sim \mathrm{N}(0,1)$ and computing $\alpha + \delta\log h_1 + \sigma_v v_0$. We use this method for the extended model as well.

Efficient draws from p in (2) are critical to the performance of the MCMC algorithm. One possible method is rejection sampling. It requires a *blanket* density q and a finite constant c such that, $p(h) \leqslant cq(h)\ \forall h$. Consider a draw h from q. Accept the draw with probability $p(h)/cq(h)$. Otherwise, draw again. This may be slow within the context of a MCMC sampler either due to the computation of the integration constant in (2) at every iteration, or a high rejection rate. Another approach is the independence Metropolis–Hastings algorithm. Make a candidate $n+1$st draw of h_t, using a transition kernel $f(h)$. Accept the candidate draw with probability

$$\min\left(\frac{p(h_t^{(n+1)})/f(h_t^{(n+1)})}{p(h_t^{(n)})/f(h_t^{(n)})},1\right). \tag{3}$$

Otherwise, repeat $h_t^{(n)}$, and move to $h_{t+1}^{(n+1)}$. The repeat rule adjusts for the difference between f and p. Since (3) uses ratios, the integration constant of p is not needed. See Hastings (1970), Tierney (1994), and Chib and Greenberg (1995) for discussion. JPR combine accept/reject and Metropolis–Hastings. First, draw from q and accept the draw with probability $\min(p(h)/cq(h),1)$. Repeat this until a draw is retained. The candidate draw then enters a Metropolis–Hastings step. If $h_t^{(n+1)}$ is not accepted, repeat $h_t^{(n)}$ and move to $h_{t+1}^{(n+1)}$. Tierney (1994) points out that this amounts to using $f(h) \propto \min(p(h),cq(h))$ in (3).

The choice of q is critical to the efficiency of the algorithm. Fewer rejections and repeats result if $q(h)$ better approximates $p(h)$. p in (2) is the product of an inverse gamma and log-normal kernels. To build q, JPR approximate the log-normal kernel by an inverse gamma, $\propto h_t^{-(\phi_{LN,t}+1)} e^{-\theta_{LN,t}/h_t}$, with the same mean and variance. This requires $\phi_{LN} = (1 - 2e^{\sigma^2})/(1 - e^{\sigma^2})$ and $\theta_{LN,t} = (\phi_{LN} - 0.5) \exp(\mu_t + 0.5\sigma^2)$. The inverse gamma kernel in (2) has parameters $\phi_1 = -0.5$ and $\theta_{1,t} = y_t^2/2$. The product of inverse gammas is an inverse gamma. The resulting blanket q is

$$q(h_t|.) \propto h_t^{-(\phi+1)} e^{-\theta_t/h_t},$$

where $\phi = \phi_1 + \phi_{LN} - 1$ and $\theta_t = \theta_{1,t} + \theta_{LN,t}$. Alternative distributions for q such as log-normals or mixtures of log-normal and inverse gamma were found to be less efficient in practice. Theoretical convergence does not require q to have a fatter tail than p, as it does for a pure accept/reject. However, the less efficient blankets all had thinner tails than the one used here.

Pure and accept/reject Metropolis have the same theoretical convergence properties, but they differ slightly in practice. Rejections require more draws from q to obtain a given number of retained draws. Repeats lead to a less informative sequence of draws. Our algorithm is fast and we are not constrained by computing time. Thus, we favor sequences of retained draws with fewer repeats, i.e. more information content. A larger c tilts the algorithm toward rejections and away from repeats. This increases the acceptance probability (3). We choose c as 1.1 times the value of p/q at the mode of q. This results in about 20% rejection rate and less than 1% repeat rate.

Other strategies have been proposed for $p(\mathbf{h}|\omega, \mathbf{y})$. Geweke (1994c) notes that p in (2) is log-concave, and therefore the algorithm in Wild and Gilks (1993) can be used. Carter and Kohn (1994) propose another approach for the basic SVOL, see also Mahieu and Schotman (1998), and Shephard and Kim (1994). They approximate the distribution of $\log \varepsilon_t^2$ with a discrete mixture of normals. This allows a joint draw of the vector $\mathbf{h}$. Our strategy does not use an approximation. While Chib et al. (2002) note that the approximation can be corrected by re-weighting, the discrete mixture approach does not extend to the correlated case introduced below.

We now allow for fat-tails in ε_t and a correlation between ε_t and v_t. We describe the fat-tail and correlation extensions separately and then observe that they are easy to combine.

2.2. Fat-tailed departures from normality

2.2.1. Model

In the context of stochastic volatility models, Geweke (1994c) and Gallant et al. (1997) provide empirical evidence suggesting fat-tails in the distribution of the conditional mean ε_t. This has also been noted in the GARCH/EGARCH literature (cf. Bollerslev, 1987; Nelson, 1991; Bollerslev et al., 1992). A fat-tailed distribution for ε_t is easily obtained by a scale mixture. In our approach, the realization of the scale mixture variable is a latent variable. The fat-tailed model is as follows:

$$y_t = \sqrt{h_t} \varepsilon_t = \sqrt{h_t} \sqrt{\lambda_t} z_t,$$

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T,$$

$$(z_t, v_t) \sim \mathrm{N}(0, I_2),$$

$$\lambda_t \sim p(\lambda_t | v). \tag{4}$$

We assume that λ_t is distributed i.i.d. inverse gamma, or that $v/\lambda_t \sim \chi_v^2$. This implies that the marginal distribution of $\varepsilon_t = \sqrt{\lambda_t} z_t$ is Student-$t(v)$. This approach has been used by Carlin and Polson (1991) with fixed v, and Geweke (1993) who estimates v for regression models. The Student-t allows for a wide range of kurtosis and approximates the tail behavior of the normal model for very large values of v. Allowing for fat-tails in ε_t extends the SVOL model in a manner similar to fat-tails extensions of GARCH models. Fat tails could also be introduced in the distribution of v_t, but in order for y_t to have finite moments, this would preclude the use of a Student-t distribution—no moments of e^v exist where $v \sim t$.

The basic and fat-tailed SVOLs differ markedly in their treatment of extreme observations or *outliers*. In the basic SVOL, large $|y_t|$'s are evidence that h_t is high. In the fat-tailed model, λ_t provides an additional source of flexibility. The fat-tailed model can deal with outliers by introducing a large λ_t. Therefore, it will take a series of large $|y_t|$'s before h_t is increased. In fact, some have suggested that a posterior analysis of λ_t might serve as an outlier diagnostic. In any case, the fat-tailed SVOL can be viewed as more outlier *resistant* than the basic SVOL.

These differences in the treatment of large $|y_t|$'s can have an important effect on volatility estimates and forecasts. The intuition is that the basic SVOL model will result in a more variable sequence of estimated h_t's than the fat-tailed model. Observations with a large $|y_t|$ will have much greater effects on future volatility forecasts for the basic than for the fat-tailed SVOL. Fig. 1 illustrates this using the UK £/$ exchange rate. Panel (a) shows the time series of log-differences in exchange rates in 1985. Panel (b) shows the smoothed estimates of h_t computed with the MCMC algorithms and priors outlined in Sections 2.1, 2.2.2 and 2.2.3, for both the basic and fat-tailed models. As might be expected, the h_t series resulting from the basic model is much more variable than the series inferred from the fat-tailed model. This is because the fat-tailed model has the ability to "*resist*" outliers by using a large value of λ_t. The high variability in the time series plot of $s_t = \sqrt{h_t \lambda_t}$ in Panel (b) further highlights the flexibility of the fat-tailed model. The observation for September 23rd also illustrates this point well. In the fat-tailed model, a large λ_t accommodates this observation, which reduces its influence on the subsequent smoothed h_t's.

The resistance to outliers has an influence on the future volatility predictions. Panel (c) of Fig. 1 shows the out-of-sample forecasts of s_t, assuming that the time series stops on September 23rd. During the sample period, panel (c) plots the smoothed values of $\sqrt{h_t}$ for the basic and fat-tailed models. The dotted lines after 9/23/85 show the forecasts of s_t. For each draw of $(\mathbf{h}, \omega, v)$, we draw future h's from the AR(1), see Geweke (1989a, 1994c), and future λ's from v/χ^2 if needed. A draw of s_{T+k} follows by computation of $\sqrt{h_{T+k} \lambda_{T+k}}$. The forecasts differ markedly for the two models with the basic SVOL model giving greater influence to the 9/23/85 observation. Note how the estimation of h_t on that 9/23 is affected by information on the upcoming low-volatility
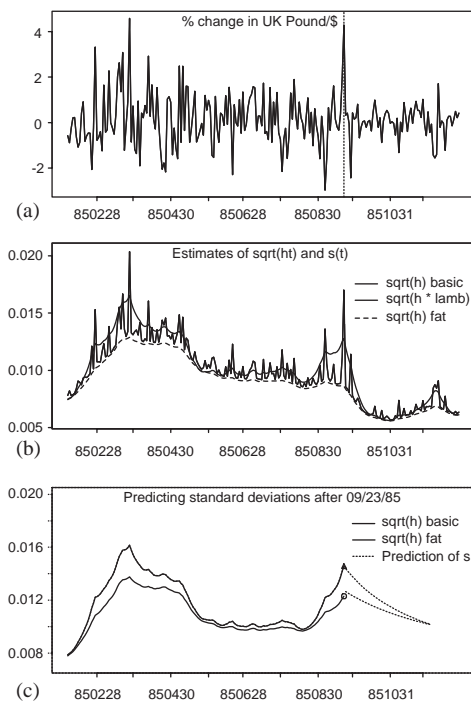
Fig. 1. Differences between fat-tail and basic SVOL volatility forecasts.

period. The fat-tailed SVOL with that information, panel b, attributes most of the shock on 9/23 to λ. Panel c, where the model does not have future information, shows a higher estimate of h_t on 9/23.

Estimates of h_t differ when the fat-tailed model generates λ_t's different from 1. This is especially true if v is low, as is the case for the £/\$ where the posterior mean of v is 10. However, even larger v's can allow some large λ's. Thus, the fat-tail extension has important practical implications on volatility estimation and forecasting.

2.2.2. Priors

For ω, we employ the same priors as in Section 2.1.2. A conjugate inverse gamma prior is used for $\lambda_t | v$.

$$p(\lambda_t | v) \propto \lambda_t^{-[(v/2)+1]} e^{-v/2\lambda_t} \sim \mathscr{IG}(v/2, 2/v) \sim v/\chi^2(v). \tag{5}$$

We use a uniform discrete prior on $[3, 40]$ for v. A lower bound at 3 enforces the existence of a conditional variance. In practice, draws below 5 are very rare. We note that $p(\lambda | v)$ does not change much for $v \in [41, 50]$. Therefore, excluding this range does not restrict the flexibility of the fat-tailed model. We conduct a sensitivity analysis with respect to the upper bound of the v prior in Section 5.1.

Geweke (1993) proposes an alternative prior, $p(v) \propto e^{-gv}$. Our uniform discrete prior is convenient for two reasons. First, the posterior $p(v | \mathbf{y})$ is proportional to the

marginal likelihood $p(\mathbf{y}|v)$. This implies that the posterior can be interpreted as a marginal likelihood, and Bayes factors for different values of v can be directly computed. Second, the discreteness allows direct draws of v rather than possibly requiring a Metropolis step. It does mean that we can not study intervals shorter than 1, but the data never have enough information on v for this to be relevant. Typical posterior standard deviations are between 2 and 6.

2.2.3. Algorithm

We now describe an MCMC algorithm for the fat-tailed model. The parameters now include v and the T latent variables λ. Consider simulating from the joint posterior distribution $p(\mathbf{h}, \omega, \lambda, v|\mathbf{y})$ by cycling through the conditionals $p(\mathbf{h}, \omega|\lambda, v, \mathbf{y})$ and $p(\lambda, v|\mathbf{h}, \omega, \mathbf{y})$. We draw $p(\lambda, v|.)$ as a block, $p(\lambda|v,.)p(v|.)$, rather than with a Gibbs cycle $[p(\lambda|v,.), p(v|\lambda,.)]$, to avoid the problems documented by Eraker et al. (1998), in which v can get absorbed into the lower bound. The strategy for drawing each set of conditionals is given by

- $p(\mathbf{h}, \omega|\lambda, v, \mathbf{y})$: If we rewrite the model as $y_t^* = y_t/\sqrt{\lambda_t} = \sqrt{h_t}z_t$, $z_t \sim \mathrm{N}(0,1)$, then the algorithm in Section 2.1.3 applies directly.
- $p(\lambda|\mathbf{h}, \omega, v, \mathbf{y})$: note that $p(\lambda|\mathbf{h}, \omega, v, \mathbf{y}) = \prod_{t=1}^{T} p(\lambda_t|y_t, h_t, v)$. Then

$$p(\lambda_t|y_t, h_t, v) \equiv p\left(\lambda_t \left|\frac{y_t}{\sqrt{h_t}}, v\right.\right) \propto p\left(\frac{y_t}{\sqrt{h_t}}\left|\lambda_t, v\right.\right) p(\lambda_t|v).$$

Using our natural conjugate priors in (5), the conditional posterior of λ_t is inverse gamma:

$$p(\lambda_t|y_t, h_t, v) \propto \lambda_t^{-[(v+1)/2+1]}e^{-[(y_t^2/h_t)+v]/2\lambda_t} \sim \mathcal{IG}\left(\frac{v+1}{2}, \frac{2}{(y_t^2/h_t)+v}\right). \quad (6)$$

- $p(v|\mathbf{h}, \omega, \mathbf{y})$: note that $(y_t|h_t, v) \sim t(v)$. v is discrete with probability mass proportional to the product of t distribution ordinates:

$$p(v|\mathbf{h}, \omega, \mathbf{y}) \propto p(v)p(\mathbf{y}|\mathbf{h}, v) = p(v)\prod_{t=1}^{T}\frac{v^{v/2}\Gamma(v+1/2)}{\Gamma(v/2)\Gamma(\frac{1}{2})}(v+y_t^2/h_t)^{-(v+1)/2}. \quad (7)$$

2.3. Correlated errors and leverage effect

2.3.1. Model

The basic SVOL specifies zero correlation between ε_t and v_t, the errors of the mean and variance equations. The correlated errors model introduces a correlation parameter, ρ.

$$y_t = \sqrt{h_t}\varepsilon_t$$

$$\log h_t = \alpha + \delta \log h_{t-1} + u_t, \quad \text{where } u_t = \sigma_v v_t. \quad (8)$$

The covariance matrix of $\mathbf{r}_t \equiv (\varepsilon_t, u_t)'$, is Σ^*

$$\Sigma^* = \begin{pmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{pmatrix}. \tag{9}$$

If the correlation ρ is negative, then a negative innovation in the levels, ε_t, will be associated with higher contemporaneous and subsequent volatilities. On the other hand, a positive innovation ε_t is associated with a decrease in volatility. This asymmetry has been dubbed the leverage effect. The correlated SVOL can exhibit a large leverage effect even for moderate values of ρ. For example, if $\rho = -0.6$ and $|\varepsilon_t| = 1.5$, then $E[h_t]$ will be 60% higher for negative versus positive shocks.

The returns on common equity indices exhibit leverage effects. For example, volatility is 22% higher in the 10 days following returns more than 2 standard deviations below the mean than after returns more than 2 standard deviations above the mean, using the daily equal-weighted CRSP index (62–87).

The asymmetry in the correlated SVOL model also induces skewness in the marginal distribution of the series. With $\rho = -0.6$ and $\delta = 0.95$, data simulated from the correlated SVOL model exhibit significant left-skewness. Returns more than 2.5 standard deviations below and above the mean occur 1.85% and 0.85% of the time, respectively. This skewness is consistent with the non-parametric evidence of Gallant et al. (1997).

2.3.2. Priors

The major challenge in the correlated model is to formulate a prior for Σ^*. Σ^* has its $(1,1)$ element fixed to 1. This means that standard inverted Wishart priors cannot be used. However, we can transform the elements of Σ^* so that conjugate priors can be used. We transform (ρ, σ_v) to (ψ, Ω) as follows:

$$\Sigma^* = \begin{pmatrix} 1 & \psi \\ \psi & \Omega + \psi^2 \end{pmatrix}.$$

As a referee suggested, our transformation is motivated by observing that the volatility innovation u_t can be written as

$$u_t = \log h_t - \alpha - \delta \log h_{t-1} = \sigma_v \rho \varepsilon_t + \sigma_v \sqrt{1 - \rho^2} \eta_t \quad \text{with } (\eta_t, \varepsilon_t) \sim \mathrm{N}(0, I_2). \tag{10}$$

This is the regression of u_t on ε_t, with slope $\psi = \sigma_v \rho$ and error variance $\Omega = \sigma_v^2(1 - \rho^2)$. As observed by McCulloch et al. (2000), one can use a normal prior for ψ and an inverse gamma for Ω. In this paper, $\Omega \sim \mathscr{IG}(v_0 = 1, v_0 t_0^2 = 0.005)$ and $\psi | \Omega \sim \mathscr{N}(\psi_0 = 0, \Omega/p_0)$, where $p_0 = 2$.

Our prior on (ψ, Ω) induces a prior distribution over (ρ, σ_v) as shown in Fig. 2. This distribution is diffuse on ρ, while ruling out very large correlations. The marginal prior on σ_v is very similar to that used in the basic model.
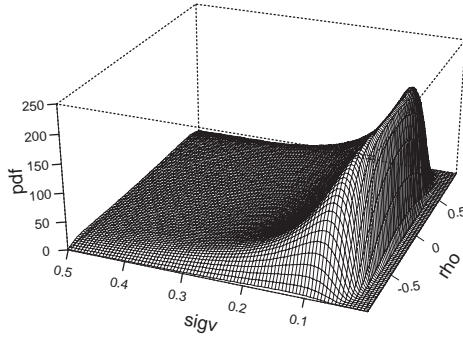
Fig. 2. Joint prior distribution of ρ and σ_v.

2.3.3. Algorithm

The joint distribution of data and volatilities is

$$p(\mathbf{y}, \mathbf{h} | \alpha, \delta, \Sigma^*) \propto \prod_{t=1}^{T} h_t^{-3/2} \, p\left(\frac{y_t}{\sqrt{h_t}}, \log h_t | h_{t-1}, \alpha, \delta, \Sigma^*\right)$$

$$= \prod_{t=1}^{T} h_t^{-3/2} |\Sigma^*|^{-1/2} \exp\left(-\frac{1}{2} \operatorname{tr}(\Sigma^{*-1} A)\right). \tag{11}$$

Here $A = \sum_t \mathbf{r}_t \mathbf{r}_t'$, is the residual matrix. The joint posterior is then

$$p(\mathbf{h}, \Sigma^*, \alpha, \delta | \mathbf{y}) \propto p(\Sigma^*) p(\alpha, \delta) \prod_{t=1}^{T} h_t^{-3/2} |\Sigma^*|^{-1/2} \exp\left(-\frac{1}{2} \operatorname{tr}(\Sigma^{*-1} A)\right). \tag{12}$$

As discussed above, we transform (ρ, σ_v) to (ψ, Ω). Given this re-parameterization and the posterior in (12), we can extend the JPR algorithm as follows.

- $p(\psi, \Omega | \alpha, \delta, \mathbf{h}, \mathbf{y})$: Note that $|\Sigma^*| = \Omega$, and rewrite Σ^{*-1} as

$$\Sigma^{*-1} = \frac{1}{\Omega} \begin{pmatrix} \psi^2 & -\psi \\ -\psi & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \equiv \frac{C}{\Omega} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

where a_{ij} is the ij element of A. Then $\operatorname{tr}(\Sigma^{*-1} A) = \operatorname{tr}(CA)/\Omega + a_{11}$. Since $a_{11} = \sum_t \varepsilon_t^2$, it does not involve ψ or Ω. So it follows from (12) that

$$p(\psi, \Omega | \alpha, \delta, \mathbf{h}, \mathbf{y}) \propto \frac{1}{\Omega^{T/2}} \exp\left(-\frac{\operatorname{tr}(CA)}{2\Omega}\right) p(\psi, \Omega).$$

Let $a_{22.1} = a_{22} - a_{12}^2/a_{11}$ and $\hat{\psi} = a_{12}/a_{11}$. Both are functions of $\alpha, \delta, \mathbf{h}$. We can write $\operatorname{tr}(CA) = a_{22.1} + (\psi - \hat{\psi})^2 a_{11}$. The joint posterior follows by conjugacy of the prior:

$$p(\psi | \Omega, \alpha, \delta, \mathbf{h}, \mathbf{y}) \sim \mathcal{N}(\tilde{\psi}, \Omega/(a_{11} + p_0)),$$

$$p(\Omega | \alpha, \delta, \mathbf{h}, \mathbf{y}) \sim \mathcal{IG}(v_0 + T - 1, v_0 t_0^2 + a_{22.1}), \tag{13}$$

where $\tilde{\psi} = (a_{11}\hat{\psi} + p_0\psi_0)/(a_{11} + p_0)$. A draw of (ψ, Ω) yields a draw of (ρ, σ_v) by computation of $\sigma_v^2 = \psi^2 + \Omega$ and $\rho = \psi/\sigma_v$.

- $p(\alpha, \delta | \psi, \Omega, \mathbf{h}, \mathbf{y})$ is a linear regression as in the basic model (see Section 2.1.3).
- $p(\mathbf{h} | \psi, \Omega, \alpha, \delta, \mathbf{y})$: We break $\mathbf{h}$ into T components $h_t | h_{t-1}, h_{t+1}$. Likelihood (11) implies

$$p(h_t | h_{t-1}, h_{t+1}, \Sigma^*, \alpha, \delta, \mathbf{y}) \propto h_t^{-3/2} \exp\left(-\frac{\operatorname{tr}(\Sigma^{*-1}\mathbf{r}_t\mathbf{r}_t') + \operatorname{tr}(\Sigma^{*-1}\mathbf{r}_{t+1}\mathbf{r}_{t+1}')}{2} \right).$$

After introducing (ψ, Ω), we obtain

$$p(h_t | h_{t-1}, h_{t+1}, \psi, \Omega, \mathbf{y}) \propto \frac{1}{h_t^{\frac{3}{2} + \frac{\delta\psi y_{t+1}}{\Omega\sqrt{h_{t+1}}}}} \exp\left(\frac{-y_t^2}{2h_t}\left(1 + \frac{\psi^2}{\Omega}\right)\right.$$

$$\left. - \frac{(\log h_t - \mu_t)^2}{2\Omega/(1+\delta^2)} + \frac{\psi y_t u_t}{\Omega\sqrt{h_t}} \right), \tag{14}$$

where $\mu_t = (\alpha(1-\delta) + \delta(\log h_{t+1} + \log h_{t-1}))/(1+\delta^2)$, $u_t = v_t\sigma_v$. The density p in (14) has inverse gamma and log-normal kernels. As before, we approximate the log-normal kernel by an inverse gamma and combine with the inverse gamma kernel. The parameters of the resulting inverse gamma are

$$\phi_t = \phi_1 + \phi_{\mathrm{LN}} + 1 = \frac{\psi\delta y_{t+1}}{\Omega\sqrt{h_{t+1}}} - 0.5 + \frac{1 - 2e^{\Omega/(1+\delta^2)}}{1 - e^{\Omega/(1+\delta^2)}} + 1,$$

$$\theta_t = \theta_1 + \theta_{\mathrm{LN}} = \frac{y_t^2}{2}\left(1 + \frac{\psi^2}{\Omega}\right) + (\phi_{\mathrm{LN}} - 1)e^{\mu_t + 0.5\Omega/(1+\delta^2)}. \tag{15}$$

This would suggest the following inverse gamma blanket $q_1(h_t)$ for the accept/reject Metropolis step:

$$q_1(h_t | h_{t-1}, h_{t+1}, \psi, \Omega, y_t) \sim \mathcal{IG}(\phi_t, \theta_t) \propto \frac{1}{h_t^{\phi_t+1}} e^{-\theta_t/h_t}. \tag{16}$$

q_1 omits the third term in the exponent of (14). This should not affect the theoretical capability of the algorithm to produce draws with invariant distribution p. However, in practice q_1 produces an inefficient algorithm. For example, with simulated data of sample size encountered in practice, the sampler based on q_1 produces a posterior with diffusion not much different from the prior. We now discuss a simple modification of q_1 that dramatically improves performance.

To illustrate the problem with q_1, consider a model with $\omega = (-0.368, 0.95, 0.26)$, and $\rho = -0.6$. For simulated samples, we plotted p and q_1 for many observations and draws. Fig. 3 shows one such observation. Fig. 3(a) shows a normalized p (solid line) and the approximating blanket q_1 (dotted line). The vertical bars are the 5th and 95th quantiles of p. q_1 is clearly located far to the right of p. The key to performance of accept/reject and Metropolis methods is the ratio $p(h)/q(h)$ which drives acceptance and repeat probabilities. The flatter p/q is, the more efficient the blanket, yet p/q_1 is far from flat as shown in Fig. 3(b).

To improve upon q_1, we must explore the third term, $u_t/\sqrt{h_t}$ in (14). It involves $\log h_t/\sqrt{h_t}$, which is not present in the kernel of any standard density. We solve this
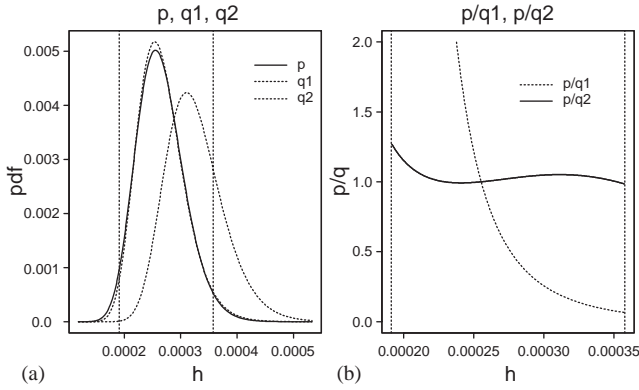
Fig. 3. Improving the blanket density for the correlated model.

problem by approximating $u_t/\sqrt{h_t}$ as a linear function of $1/h_t$. To do this, we compute $u_t/\sqrt{h_t}$ at two points around the mode, and use the slope s between these two points. Using this linearization of the third term, we can combine the third term in the inverse gamma kernel. This results in a new blanket q_2

$$q_2(h_t|h_{t-1}, h_{t+1}, \psi, \Omega, y_t) \sim \mathscr{IG}(\phi_t, \theta_t^*) \propto \frac{1}{h_t^{\phi_t+1}} e^{-\theta_t^*/h_t}. \tag{17}$$

Here $\theta_t^* = \theta_t - s\psi\, y_t/\Omega$. Fig. 3 shows that q_2 provides a much improved approximation to p. We also computed a measure of p/q variability using the squared relative differences of p/q at the mode versus one point on each side, averaged across draws and observations. Using q_2 instead of q_1 results in a much less variable p/q ratio.

2.4. Algorithm for fat-tails and correlation

The full model with both fat-tails and correlated errors is

$$y_t = \sqrt{h_t \lambda_t} z_t,$$

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T,$$

$$v/\lambda_t \sim \chi_v^2,$$

$$(z_t, v_t) \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right). \tag{18}$$

The hierarchical structure allows us to use the correlated errors algorithm in Section 2.3.3 with only minor modifications. First, given λ, we replace y_t with $y_t/\sqrt{\lambda_t}$ and apply the algorithm described in Section 2.3.3. Second, given **h**, we draw from the posteriors in (6) and (7). The iteration of these two steps is the algorithm of the full model.

2.5. Odds ratios

Bayesian comparison of alternative models are often made using posterior odds ratios (see Kass and Raftery (1995) for a survey of the posterior odds literature). A posterior odds ratio is the product of the ratio of the marginal densities of the data, called the Bayes factor, times the prior odds ratio. The Bayes factor requires integrating out the model parameters. For example, if $O_{1|2}$ and $BF_{1|2}$ denote the posterior odds ratio and Bayes factor of model M_1 over M_2, then

$$O_{1|2} = \frac{p(M_1)}{p(M_2)} \times \frac{p(y|M_1)}{p(y|M_2)} = \text{prior odds} \times BF_{1|2} \quad \text{and}$$

$$BF_{1|2} = \frac{\int p(\theta_1|M_1)p(\mathbf{y}|\theta_1, M_1)\,\mathrm{d}\theta_1}{\int p(\theta_2|M_2)p(\mathbf{y}|\theta_2, M_2)\,\mathrm{d}\theta_2}. \qquad (19)$$

The direct computation of these marginal likelihood is intensive, possibly unstable for latent variables models. Instead, a variety of methods have been proposed to use MCMC draws to approximate Bayes Factors (cf. Newton and Raftery, 1994).

In this section, we show how to use the special structure of our models and priors to compute Bayes factors for comparing the basic SVOL model to the models with fat-tails and correlated errors. Let $B, F, C,$ and FC denote the basic, fat-tailed, correlated errors, and full SVOL models respectively. We observe that $BF_{B|FC} = BF_{B|F} \times BF_{F|FC}$. In Section 2.5.1, we show how to compute $BF_{B|F}$. Section 2.5.2 discusses the computation of $BF_{F|FC}$.

2.5.1. Fat-tails

To compare the basic to the fat-tailed model, we can write the Bayes factor as the expectation of the ratio of un-normalized posteriors with respect to the posterior under the fat-tailed model. We are grateful to a referee for pointing this out.

$$BF_{B|F} = \frac{p(\mathbf{y}|M_B)}{p(\mathbf{y}|M_F)} = \mathrm{E}\left[\frac{p_B(\theta|M_B)p_B(\mathbf{y}|\theta, M_B)}{p_F(\theta|v, M_F)p_F(\mathbf{y}|\theta, v, M_F)}\right]. \qquad (20)$$

Here $\theta = (\alpha, \delta, \sigma_v, \mathbf{h})$, and E[.] refers to the expectation over the joint posterior of (θ, v). Eq. (20) is derived by first observing that $p(\mathbf{y}|M_F) = p_F(\mathbf{y}|\theta, v)p_F(\theta, v)/p_F(\theta, v|\mathbf{y})$. It then follows that:

$$BF_{B|F} = \frac{\int p_B(\theta)p_B(\mathbf{y}|\theta)\,\mathrm{d}\theta}{p(\mathbf{y}|M_F)} = \int_v \int_\theta \frac{p_B(\theta)p_B(\mathbf{y}|\theta)}{p(\mathbf{y}|M_F)}\,\mathrm{d}\theta\, p_F(v)\,\mathrm{d}v$$

$$= \iint \frac{p_B(\theta)p_B(\mathbf{y}|\theta)p_F(\theta, v|\mathbf{y})p_F(v)}{p_F(\mathbf{y}|\theta, v)p_F(\theta, v)}\,\mathrm{d}\theta\,\mathrm{d}v$$

$$= \iint \frac{p_B(\theta)p_B(\mathbf{y}|\theta)}{p_F(\mathbf{y}|\theta, v)p_F(\theta|v)}p_F(\theta, v|\mathbf{y})\,\mathrm{d}\theta\,\mathrm{d}v.$$

Eq. (20) simplifies further. Given our priors, $p_F(\theta|v) = p_F(\theta) = p_B(\theta)$, and $p(v)$ is constant. So the Bayes factor is only the ratio

$$BF_{B|F} = E\left[\frac{p(\mathbf{y}|\theta, M_B)}{p(\mathbf{y}|\theta, v, M_F)}\right]. \qquad (21)$$

This expectation can be approximated by averaging over the MCMC draws.

Note that (21) extends to the computation of $BF_{C|FC}$. With non-zero ρ, a draw of (ω, h_t) implies a draw of v_t, and $\varepsilon_t|v_t \sim N(\rho v_t, 1 - \rho^2)$. So, in (21), we use $y_t^* = (y_t - \rho v_t)/\sqrt{1 - \rho^2}$ instead of y_t.

2.5.2. Correlated errors

To simplify the discussion, we first consider $BF_{B|C}$ and then show how to extend this approach to compute $BF_{F|FC}$. We use the Savage density ratio approach (Dickey, 1971) to compute $BF_{B|C}$. Let (ψ, ϕ) denote the parameters of the correlated model, where $\phi = (\Omega, \alpha, \delta, \mathbf{h})$. The basic SVOL is nested within the correlated model and sets $\psi = 0$. Let p_B and p_C be the priors under the basic and correlated models. Then, if $p_B(\phi) = p_C(\phi|\psi = 0)$, the Bayes factor is simply the ratio of posterior over prior densities:

$$BF_{B|C} = \left.\frac{p_C(\psi|\mathbf{y})}{p_C(\psi)}\right|_{\psi=0}.$$

To compute the marginal posterior of ψ, we simply average the conditional posterior $p_C(\psi|\phi)$ over the draws of ϕ. Jacquier and Polson (2000) provide the details of the Savage density approach. They analytically integrate out Ω. The resulting conditional posterior, $p_C(\psi|\mathbf{h}, \alpha, \delta, \mathbf{y})$ is a Student-t. Given G draws of the MCMC sampler, the Bayes factor can be approximated by

$$\widehat{\mathscr{BF}}_{B/C} = \frac{\Gamma(\frac{v_0+T}{2})\Gamma(\frac{v_0}{2})}{\Gamma(\frac{v_0+T-1}{2})\Gamma(\frac{v_0+1}{2})} \frac{1}{G}\sum_{g=1}^{G} \sqrt{\frac{1 + a_{11}^{(g)}/p_0}{1 + a_{22.1}^{(g)}/v_0 t_0^2}} \left[1 + \frac{\tilde{\psi}^2}{v_1 t_1^2/p_1}\right]^{(g)^{-(v_0+T)/2}}, \qquad (22)$$

where $\tilde{\psi}$ is defined for Eq. (13) and (g) refers to the gth draw of the sampler.

In the full model algorithm described in Section 2.4, fat-tails do not modify the conditional posteriors of ψ and the other parameters. Replace $\mathbf{y}$ by $\mathbf{y}/\sqrt{\lambda}$, and (22) applies directly for the computation of $BF_{F/FC}$, the Bayes factor of a pure fat-tail model F versus a full model FC.

3. Sampling performance of Bayes parameter and volatility estimates

Since likelihood-based inference has not been conducted for the extended model, we do not know the sampling properties of Bayes estimators. In particular, we do not know how informative the sample data can be about the new parameters, v and ρ. In addition, the consequences of using a mis-specified model, such as using the basic

Table 1
Sampling properties of Bayes estimators: full model

True value	δ	σ_v	$E_h \times 10^3$	V_h/E_h^2	v	ρ
	0.95	0.26	0.9	1	10	−0.6
Average	0.94	0.27	0.93	1.02	14	−0.42
RMSE	(0.025)	(0.039)	(0.19)	(0.34)		(0.19)
$Q1$, $Q3$					10 18	

We simulate 500 samples of 1000 observations from the full model. For each sample, the posteriors are based on 25 000 draws of the sampler, after discarding 5000 draws. The rows entitled "Average" and "RMSE" report the average and the Root mean squared errors of the 500 posterior means. For v, however, we report the average of the 500 first and third posterior quartiles.

model in the presence of fat-tails, are unknown. In this section, sampling experiments are conducted to gauge the performance of Bayes volatility and parameters estimators.

3.1. Parameter estimation

Table 1 documents the sampling performance of the posterior mean for data generated from $\omega = (-0.368, 0.95, 0.26)$. These values imply $E_h = 0.0009$ and the squared coefficient of variation, $\text{Var}(h)/E_h^2 = 1$ and have been used by JPR and others. We also set $v = 10$ and $\rho = -0.6$; which induces fat-tails and a reasonably large leverage effect. Five hundred samples of 1000 observations each are simulated. For each sample, we generate 30,000 draws from the MCMC algorithm. Bayes estimators are approximated by the average of the last 25,000 draws. It should be noted that samples of size 1000 are small by the standards of high-frequency financial time series analysis so that the results presented here represent a "stress-test" of our method.

Table 1 presents the sample average of the posterior means as well as their RMSE across the 500 samples. The regression parameters ω, and functions thereof, are recovered very precisely. The estimators of ρ and v appear to have moderate upward biases. Of course, Bayes estimators do not have to be unbiased even with locally uniform priors. However, as the sample size increases this bias must disappear unless the priors are dogmatic. Indeed, we experimented with samples of size 3000 (the approximate size of the data sets used in our empirical analysis below) and found that the bias in estimation of ρ decreases dramatically.

The evidence proposed in Table 1 shows that Bayes estimators have good sampling properties for the extended model. In addition, the sampling performance provides indirect evidence that our algorithm performs well with a modest number of draws. More direct evidence of the algorithm's convergence properties is given in Section 5 below.

We now turn to the performance of estimators under mis-specification. Table 2 presents evidence of the consequences of mis-specification. Panel (a) reports the performance of the Bayes estimator on data simulated from the basic SVOL model with normal errors. Panel (a) shows that virtually no loss of precision is incurred when we fit the fat-tailed model to data generated from the basic model. However, as is shown

Table 2
Estimating the wrong model: effect on parameter estimates

True value	δ	σ_v	$E_h \times 10^3$	V_h/E_h^2	v
	0.95	0.26	0.9	1	∞

Panel a: Data simulated from the basic SVOL with normal errors
Basic model fitted:

Average	0.93	0.29	0.94	1.09	
RMSE	(0.03)	(0.05)	(0.18)	(0.42)	

Fat-tail model fitted:

Average	0.94	0.27	0.88	1.03	28
RMSE	(0.02)	(0.04)	(0.17)	(0.40)	
$Q1$, $Q3$					23 35

True value	δ	σ_v	$E_h \times 10^3$	V_h/E_h^2	v
	0.95	0.26	0.9	1	10

Panel b: Data simulated from the fat-tailed SVOL
Basic model fitted:

Average	0.91	0.37	1.16	1.35	
RMSE	(0.05)	(0.12)	(0.34)	(0.58)	

Fat-tail model fitted:

Average	0.93	0.30	1.00	1.15	18
RMSE	(0.03)	(0.06)	(0.22)	(0.48)	
$Q1$, $Q3$					12 23

We simulate 500 samples of 1000 observations each from the basic model (Table 2(a)), and from the fat-tailed model (Table 2(b)). For each sample, the posteriors are based on 25 000 draws of the sampler, after discarding 5000 draws. The rows entitled "Average" and "RMSE" report the average and the Root mean squared errors of the 500 posterior means. For v, we report the average of the 500 first and third posterior quartiles.

in panel (b), if the data has fat-tails, use of the basic SVOL model imparts substantial biases to the estimation of model parameters. The parameters most affected are σ_v, E_h, and V_h/E_h^2 as the basic SVOL model raises both the level and the variability of volatility to accommodate the observations which are "outlying" relative to a normal distribution.

3.2. Volatility estimation

Volatility estimation and prediction is one of the most important uses of the model. Table 3 documents the performance of the algorithm and the consequence of using the wrong model on the estimates of volatility. For each observation of a simulated sample, we can examine the performance of Bayes estimates of $\sqrt{h_t}$ and of the *standard deviation* $s_t = \sqrt{h_t \lambda_t}$. RMSE and the percentage mean absolute error, %MAE, are computed by averaging over the 1000 observations of the 500 samples. The percentage absolute error is defined as $100|\hat{s}_t - s_t|/s_t$.

Table 3
Estimating the wrong model: effect on smoothing performance

| | Data generated from the | | |
| | Basic model | Fat-tailed model | |
	All obs.	All obs.	$\lambda_t > \lambda_{.9}$
Fat-tailed model fit:			
RMSE(s_t)	0.00659	0.0104	0.0198
RMSE($\sqrt{h_t}$)	0.00656	0.0071	0.0076
%MAE($\sqrt{h_t}$)	18.3	21.4	23.7
Basic model fit:			
RMSE(s_t)	0.00658	0.0107	0.0198
RMSE($\sqrt{h_t}$)	0.00658	0.0082	0.0098
%MAE($\sqrt{h_t}$)	19.2	25.9	30.6

The sampling distributions are based on the 1000 samples of Table 2. For each observation, we compute the estimation error of the posterior mean of $\sqrt{h_t}$ and $s_t = \sqrt{h_t \times \lambda_t}$. We report the root mean squared error, RMSE, and the average of the absolute values of % errors, %MAE. Under each column titled "All obs.", the averages are computed over all the 500,000 observations. When the data are generated by the fat-tailed SVOL with $v = 10$, we also report RMSE and %MAE for the subsets of observations with true λ_t larger than the 90th percentile of $p(\lambda|v = 10)$.

The first column of Table 3 reports on data generated by the basic SVOL model. The first three rows document performance with estimates from the fat-tailed model, the next three rows present results from the estimation of the basic model. Note that when estimating the basic model, $s_t = \sqrt{h_t}$, so the RMSEs are equal. It is not so when v is estimated. The two approaches have about the same performance, e.g., an 18% relative MAE for $\sqrt{h_t}$. Thus, the unnecessary estimation of the fat-tailed model does not affect smoothing performance.

The last two columns document performance when the data are generated by the fat-tailed model. The first of these columns uses all the observations. The RMSE for s_t is about the same whether or not fat-tails are incorporated. Recall from the discussion of Fig. 1 in Section 2.2.1 that the key in forecasting performance is the accuracy of the estimation of h_t. If both models fit s_t equally well, the basic model may not fit $\sqrt{h_t}$ as well since λ_t is incorrectly set to 1. The second and third rows of results show the extent of this problem. In absolute and relative terms, the basic model performs markedly worse than the fat-tailed model for $\sqrt{h_t}$, e.g., a %MAE of 26% versus 21%, or a 25% higher error. The last columns shows how performance is related to the size of the true λ_t. The worst fit occurs for the observations with the larger λ_t's. For the observations with λ_t in the top decile, the %MAE for $\sqrt{h_t}$ is 31% for the basic SVOL versus 24% for the fat-tailed model, a 30% higher error.

These differences have significant economic implications. For example, at the money, a stochastic volatility based option pricing model is approximately linear in the standard deviation when volatility is not priced, e.g., Hull and White (1987). Therefore, a reduction in RMSE of volatility may lead to a commensurate reduction of option pricing RMSE. Note, however, that our results apply to one-day ahead volatility forecasts

while most option pricing studies look at longer horizons, over which the reduction in RMSE may be less pronounced.

4. Empirical applications: equities and exchange rates

4.1. Data

We study both weekly and daily financial series. The weekly series used are the equal and value weighted CRSP indices and the first, fifth and tenth deciles of size sorted portfolios, representing small, medium and large firms. All the series are pre-filtered for AR(1) and monthly seasonals as in JPR. The samples include 1539 weekly returns from 1962 to 1991.

The daily series include two stock indices and three exchange rates. One index is the S&P500, 2023 daily returns from 1980 to 1987, filtered to remove calendar effects as in JPR. The second is the CRSP value weighted index, 6409 daily returns from 1962 to 1987, used by Nelson (1991) in his EGARCH. The S&P500 contains only large firms while the CRSP contains all the firms on the exchange. While there may be good reasons for a leverage effect in equities, it is less clear for exchange rates. We study the exchange rates of the UK £ in US dollars per £, and the Deutsche Mark (DM), and Canadian dollar (CAD) relative to the US dollar. The CAD includes 3010 daily observations from January 1975 to December 1986. The £ and DM include 2613 daily observations from January 1980 to June 1990. We study the first difference of the logarithm of the exchange rate series, which removes any autocorrelation.

4.2. Posterior analysis

4.2.1. Weekly series

Table 4 summarizes the posterior distribution of the parameters. First, larger firms appear to have a more persistent pattern of volatility than smaller firms; the posterior of δ shifts upward from $d1$ to $d10$. Second, larger firms exhibit a lower variability of volatility as shown by σ_v and V_h/E_h^2. Both these posteriors shift downward from $d1$ to $d10$. Finally, the posterior unconditional variance, E_h, centers on 0.45 for $d5$, $d10$, EW and VW, about a 15% annualized standard deviation. It is markedly higher, 0.6, for the smallest firm index $d1$. These patterns appear, but much weaker, when comparing EW to VW. Recall that small firms are represented more heavily in EW than VW due to the equal-weighting.

Now consider the posterior evidence regarding v and ρ. Large firms, VW and $d10$, have posterior means of v around 25 while the small firms, $d1$, $d5$, and EW, have means around 20. The extent to which this supports fat-tailed error distributions is difficult to assess directly from v. Section 4.4 provides more formal evidence using posterior odds. The posterior mean of ρ is large, below -0.4 for all but the small firm index $d1$ for which it is -0.15. ρ is estimated precisely with a standard deviation around 0.06.

Table 4
Posterior analysis for weekly series

	EW	VW	$d1$	$d5$	$d10$
δ	0.943	0.945	0.906	0.932	0.950
	(0.015)	(0.014)	(0.023)	(0.017)	(0.014)
	0.917, 0.965	0.920, 0.967	0.868, 0.942	0.902, 0.958	0.925, 0.971
σ_v	0.27	0.252	0.37	0.28	0.23
	(0.03)	(0.03)	(0.05)	(0.03)	(0.03)
	0.216, 0.327	0.203, 0.304	0.28, 0.44	0.23, 0.34	0.185, 0.29
$\frac{V_h}{E_h^2}$	0.98	0.86	1.17	0.86	0.81
	(0.23)	(0.20)	(0.24)	(0.18)	(0.19)
	0.64, 1.52	0.56, 1.33	0.78, 1.70	0.58, 1.28	0.52, 1.26
$10^3 E_h$	0.45	0.43	0.60	0.46	0.44
	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)
	0.35, 0.58	0.34, 0.56	0.47, 0.74	0.37, 0.58	0.35, 0.58
v	21	25	23	20	26
	(8.5)	(8.2)	(9.1)	(8.5)	(8.1)
	9, 18, 37	13, 25, 39	10, 22, 38	9, 18, 37	13, 26, 39
ρ	−0.46	−0.47	−0.15	−0.44	−0.41
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
	−0.55, −0.36	−0.57, −0.36	−0.25, −0.05	−0.53, −0.34	−0.53, −0.28

The first number is the posterior mean. The number between parentheses is the posterior standard deviation. The two numbers below are the posterior 5th and 95th quantiles, and the median for v. The prior domain of v is [3,40]. EW, Equal-weighted NYSE; VW, Value-weighted NYSE; $d1$, small, $d5$, medium, $d10$, large firms; weekly returns, 7/62–12/91. $T = 1539$. Returns have been prefiltered to remove AR(1) and monthly seasonals from the mean equation.

4.2.2. Daily series

Table 5 shows the results for the daily series. The posteriors of δ are located higher than for the weekly series. This is consistent with temporal aggregation, as suggested by Meddahi and Renault (2000). The highest mean is 0.988, for the full sample CRSP. All other posteriors have means below 0.98 and often no appreciable mass above 0.99, despite a locally uniform prior around 1.0. As with the basic SVOL, there is no apparent evidence of unit root in volatility. This is consistent with the model free results in Christoffersen and Diebold (1997). But this differs from the typical persistence reported in the GARCH literature. For example, the EGARCH estimates for these series are on the edge of non-stationarity, see Nelson (1991) for the CRSP series.

The posteriors for v are centered between 10 and 15, at most 20 for the CRSP. The CAD is the exception with a posterior mean of 32 and 95% of the draws above 20. Despite the clear evidence of fat-tails, there is no support for very low degrees of freedom of the Student-t distribution. In contrast, GARCH estimates of v for these series are much lower, 6.3, 9.8, 8.1, 6.3, 8.1 for EW, CRSP, S& P500, £, CAD. GARCH

Table 5
Posterior analysis for daily series

	CRSP	CRSP	SP500	UK £	DM	CAD$
From:	07/2/62	1/2/80	1/2/80	1/2/80	1/2/80	1/2/75
To:	12/31/87	12/31/87	12/31/87	5/31/90	5/31/90	12/10/86
No. Obs.	6409	2023	2023	2613	2613	3010
δ	0.988	0.977	0.980	0.980	0.966	0.956
	(0.002)	(0.01)	(0.008)	(0.007)	(0.009)	(0.009)
	0.984 0.992	0.960 0.990	0.967 0.991	0.967 0.991	0.950 0.979	0.940 0.970
σ_v	0.131	0.129	0.118	0.116	0.167	0.246
	(0.01)	(0.024)	(0.021)	(0.022)	(0.023)	(0.025)
	0.116 0.147	0.09 0.17	0.09 0.154	0.084 0.156	0.132 0.206	0.21 0.29
$\frac{V_h}{E_h^2}$	1.18	0.48	0.48	0.44	0.53	1.07
	(0.28)	(0.14)	(0.15)	(0.12)	(0.11)	(0.19)
	0.79 1.86	0.29 0.82	0.28 0.86	0.27 0.73	0.36 0.79	0.76 1.51
E_h	$\times 10^4$	$\times 10^4$	$\times 1$	$\times 10^4$	$\times 10^4$	$\times 10^5$
	0.62	0.69	0.99	0.39	0.42	0.52
	(0.094)	(0.10)	(0.15)	(0.05)	(0.04)	(0.06)
	0.48 0.84	0.54 0.92	0.75 1.33	0.31 0.49	0.34 0.51	0.42 0.64
v	21	10	15	10	12	32
	(5.9)	(3.3)	(5.4)	(2.3)	(3.9)	(6.1)
	14 20 33	7 10 16	9 13 26	7 9 14	8 11 19	21 33 40
ρ	−0.48	−0.22	−0.20	−0.15	−0.023	−0.29
	(0.04)	(0.09)	(0.09)	(0.08)	(0.07)	(0.05)
	−0.54 −0.42	−0.36 −0.08	−0.35 −0.04	−0.28 −0.01	−0.14 0.09	−0.38 −0.19

The results shown are as in Table 4. For E_h, as the series have different scales, the first row is the scale factor used on the actual posterior. For example, the posterior mean of E_h for the first series is 0.62×10^4. The first two series are the CRSP daily VW returns. The third series is the daily change in log of the S&P500-index, filtered to remove calendar effects, see Gallant et al. (1997). The UK £ and DM/$ daily noon spot rates (log change) from the board of Governors of the Federal Reserve System are supplied by David Hsieh. The CAD$ daily noon interbank market spot rates from the Bank of Canada is supplied by Angelo Melino.

models do not provide as much mixing as the basic SVOL and require lower degrees of freedom in the resulting conditional distribution. The posteriors of v are lower than for weekly series for two possible reasons. First, temporal aggregation induces normality through a central limit theorem effect. Second, the larger sample sizes of the daily series produce posteriors more informative about a low v.

The posteriors for the full sample CRSP and the S& P500, (columns 1 and 3), are different, with v and V_h/E_h^2 markedly higher for the CRSP. The leverage effect is also much stronger; ρ centered at −0.48 versus −0.2. With far more securities in the CRSP, one expects E_h to be smaller than for the S& P500. The differences in the other parameters are a bit surprising. To match the S& P500 sample period, column 2
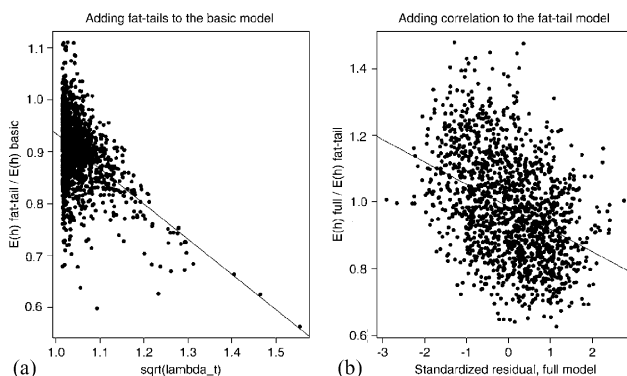
Fig. 4. Effect of fat tails and correlation on $E(h_t)$, weekly EW index.

reports estimates for the CRSP subperiod 80–87. The CRSP posteriors of $\delta, \rho, V_h/E_h^2$ are now consistent with the S&P500. This indicates shifts in the parameters over the 38 year period.

Overall, the leverage effect is strong for both daily and weekly equity indices. The leverage effect is significantly lower for exchange rates, -0.02 for the DM, -0.15 for the £. The CAD again differs markedly with ρ around -0.29. A leverage effect in an exchange rate is conceivable. For example, if Canada, small relative to the US, has a lot of US$ denominated debt, a decrease in the CAD may be viewed as an increase in leverage. In any case, these results show leverage effects in exchange rate volatility cannot be ruled out.

4.3. Effect of the extensions on volatility estimates

Section 3 demonstrates that model mis-specification can induce substantial parameter bias and error in inference about h_t. We now examine the sensitivity of volatility estimates to model specification for the weekly EW index. The results are similar for the other series. We study the fat-tails and the leverage effect separately. In what follows, a "∧" over a quantity denotes the posterior mean. Consider the effect on $\hat{h}_t$ of adding fat-tails to the basic model. The ratio of the two $\hat{h}_t$'s produced by the fat-tailed and basic models, denoted F and B, is a good metric for this effect. The higher λ_t in the fat-tailed model, the smaller the ratio is likely to be.

Fig. 4(a) plots the ratio $\hat{h}_{t,F}/\hat{h}_{t,B}$ versus $\sqrt{\hat{\lambda}_t}$ for each observation. The higher $\hat{\lambda}_t$ the more the basic model overestimates h_t relative to the fat-tailed model. For the observations with the 5% highest $\hat{\lambda}_t$'s, this relative overestimation is above 15%. The basic model overestimates h_t relative to the fat-tailed model for a vast majority of the observations.

We now add the leverage effect. Consider the ratio of $\hat{h}_t$'s under the fat-tailed and full model with both fat-tails and correlation, denoted FC. A negative ρ introduces an asymmetry between observations with negative and positive ε_t. To illustrate this, we

Table 6
Bayes factors

	v	ρ	$BF_{B/F}$	$BF_{C/FC}$	$BF_{F/FC}$	$BF_{B/FC}$
EW	20	−0.46	0.12	0.17	1E-19	1E-20
VW	25	−0.47	0.37	0.33	2E-17	7E-18
$d1$	23	−0.15	0.29	0.20	0.13	0.04
$d5$	20	−0.44	0.13	0.07	1E-19	1E-20
$d10$	26	−0.41	0.47	0.91	2E-05	1E-05
CRSP 62-87	21	−0.48	7E-06	2E-04	1E-144	7E-150
CRSP 80-87	10	−0.22	1E-06	2E-05	0.23	2E-07
SP500	15	−0.20	1E-03	1E-03	0.72	7E-04
CAD	32	−0.29	3.9	1.7	4E-07	2E-06
UK	10	−0.15	3E-08	1E-08	1.25	4E-08
DM	12	−0.02	2E-05	1E-05	4.6	1E-04

B refers to the basic model, F to the model with fat tails only, C to the model with leverage only, and FC to the full model. The columns v and ρ report the posterior means of v and ρ for the full model. Fat-tails are estimated with $v \in [3, 40]$.

plot $\hat{h}_{t,FC}/\hat{h}_{t,F}$ versus $\hat{\varepsilon}_{t,FC}$ in Fig. 4(b). The smaller $\hat{\varepsilon}_t$ is, the more the pure fat-tailed model underestimates h_t relative to the full model. The correlation between the ratio and $\hat{\varepsilon}_t$ in the plot is −0.41. The average ratio for observations in the first decile of ε_t (left tail) is 1.09 while the average ratio in the 10th decile (right tail) is 0.9—a 20% difference. This shows that the leverage model results in markedly different estimates of volatility for a large fraction of the observations.

Nelson (1991) notes that the large outliers in the data are still very unlikely even after extending an EGARCH with fat-tailed conditionals. Geweke (1994c) shows that the basic SVOL has the same problem with the largest outlier, October 87. For the daily CRSP index, the standardized return on October 19, 87 is −23.7. The basic and fat-tailed SVOL residuals $\hat{\varepsilon}_t$ are −4.83 and −3.89. One expects values larger than this once in 113 and once in 1.6 samples respectively. So the SVOL with fat-tails can accommodate large outliers. However, a criterion for a successful model is not necessarily how well it deals with the one or two worst outliers. We must also consider the relative performance of various SVOL models for the entire data set. For this reason, we now compute posterior odds.

4.4. Odds ratio analysis

Table 6 reports the estimated Bayes factors. The first and second columns recall the posterior means of v and ρ in the full model FC. The second column shows the Bayes factors, $BF_{B|F}$, for comparison of the basic to fat-tailed SVOL models. With the exception of CAD, the daily series overwhelmingly favor the fat-tails. The weekly series all favor the fat-tails though not as strongly. The next column reports $BF_{C|FC}$, the
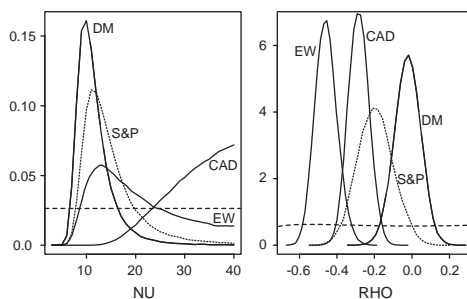
Fig. 5. Posteriors and priors for v and ρ.

Bayes factor comparing the correlated model without and with fat-tails. The odds are nearly identical, which shows that the evidence on fat-tails is robust to the existence of a leverage effect. The next column shows the Bayes factors, $BF_{F|FC}$, for the fat-tailed versus the full model, testing the leverage effect. The evidence very strongly favors the leverage effect for all equity indices and the CAD. It is against the leverage effect for the other two exchange rates, £ and DM. Finally, the last column shows the Bayes factors for the basic versus the full model, $BF_{B/FC}$. For all series, the Bayes factors overwhelmingly support the full model. This is because all series exhibit either fat-tails or correlation, or both. The Bayes factors $BF_{F|FC}$ for the CRSP, 1962–1987, and the S&P500 are very different. We also report these odds for the CRSP from 1980 to 1987, the same sample period as for the S&P500. As was the case for the posterior distribution of ρ, the odds for the CRSP and the SP500 are similar when computed over the same period.

5. Prior sensitivity and convergence

5.1. Sensitivity to prior specification

In all reported results, we use a uniform prior for v on $[3, 40]$. To study the sensitivity of the results to the upper bound on v, we estimate the models with upper bounds at 30 and 60. For the weekly series, an upper bound at 60 increases the posterior means by 5–7. For the daily series, a higher bound hardly affects the posterior distribution (the only exception is CAD with a posterior mean of 47). The posteriors of ρ and ω are largely unaffected. We also computed the odds ratios for bounds at 30 and 60. They were largely unaffected and none of the conclusions from Table 6 changed.

Fig. 5 shows the posterior distributions for v and ρ, overlaid with the priors (dashed lines). Fig. 5(a) shows posteriors and the prior for v. The posteriors are peaked relative to the prior. Except for CAD, all posteriors put mass in a region well within the support of the prior. Fig. 5(b) shows the posteriors for ρ. Again, the posteriors are very peaked over subsets of the parameter domain, showing that their shape is not affected by the prior.
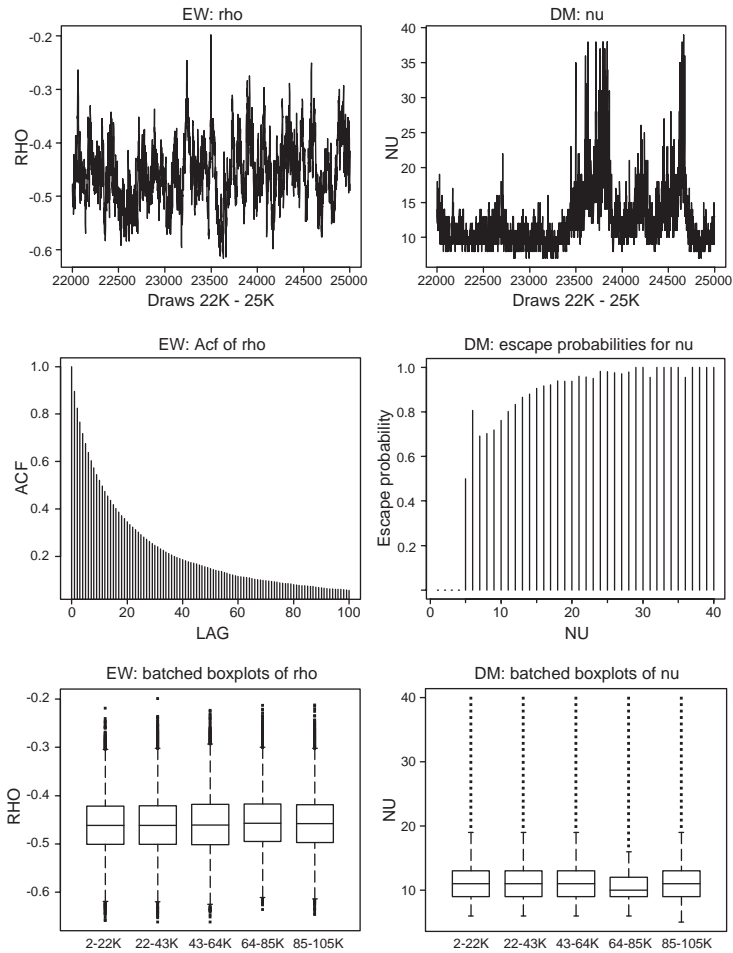
Fig. 6. Diagnostics from the samples of draws of v and ρ.

5.2. Convergence and efficiency

Experiments with different starting points show that initial conditions always dissipate extremely fast. A few hundred draws are more than enough for all parameters. The 5000 burn-in period used appears more than adequate. Standard tests of convergence, such as the autocorrelation of the draw sequences, confirm the high speed of convergence.

Fig. 6 shows some of these diagnostics for two series. Results for other series are similar. The left plots in Fig. 6 show diagnostics for ρ for the EW series. The top plot shows a sequence of 4000 draws for EW. The results in the paper are based on 100,000 draws. The sampler clearly makes large moves and navigates the parameter space. The middle plot shows the ACF of ρ. The first order is 0.9 but decays quickly.

For other series the first order is often lower. The bottom plot shows batched boxplots of the draws, each with 20,000 draws, confirming the stability of the distribution.

The right plots of Fig. 6 shows diagnostics for v for the DM series. The top plot, a time series plots of the draws, shows that even though the distribution is clearly centered around small values, the sampler can make large moves over the entire space. The first-order autocorrelation of v is about 0.6. For a discrete parameter, such as v here, one can easily compute an escape probability. For $v^{(n)}$ the nth draw of v, the escape probability is $\Pr(v^{(n+1)} \neq v^{(n)})$. Escape probabilities are inversely related to the probability of getting *stuck* in a subset of the parameter space, e.g., Diaconis and Strook (1991), Eraker et al. (1998). The middle plot shows that the escape probabilities are always high. This confirms the intuition of the time series plots that v moves well through the parameter space. Last, the batched boxplots in the bottom show that the posterior distribution of v is stable.

The ACF of a sample of draws allows to compute standard errors for the Monte Carlo means. Given the positive autocorrelation of the samplers, these standard errors are larger than those assuming zero autocorrelation. The ratio of these two quantities is the relative numerical efficiency, see Geweke (1989b, 1992). Consider estimating the variance with 100 lags of the ACF and a triangular window. For example for the VW series, the standard errors for the MCMC estimates of the posterior means of $\delta, \sigma_v, \rho, v$ are 0.0003, 0.0008, 0.0012, 0.14, with relative numerical efficiencies of $\frac{1}{7}$, $\frac{1}{9}$, $\frac{1}{5.5}$ and $\frac{1}{6}$.

The results for both simulated and actual data reported here show that our algorithm is fast and converges rapidly with acceptable levels of numerical efficiency. As a practical matter, we can obtain accurate estimates of posterior quantities even with a modest number of draws. It should be emphasized that our sampling experiments provide strong evidence of convergence of the chain. The autocorrelation function and associated numerical efficiencies of a sequence of draws is useful to assess the accuracy of the Monte Carlo estimates of posterior quantities. However, it is not a substitute for the evidence obtained from sampling experiments. For example, a chain that gets stuck in a sub-space of the parameter space can exhibit low autocorrelation but may not navigate rapidly enough for practical use, see Eraker et al. (1998) for an example with the block sampler. As researchers develop other algorithms in the future, we encourage them to perform careful sampling experiments to establish convergence across a wide range of empirically relevant parameter values.

6. Conclusion

We develop a MCMC algorithm to conduct inference in an extended SVOL model, featuring fat-tails and a leverage effect. Since basic model parameters are augmented with both the conditional volatilities h_t and the scale-mixture parameters λ_t, exact finite sample smoothing is possible. In order to assess the weight of the sample evidence in favor of the extensions, methods for computation of Bayes Factors are introduced. Our Bayes factor computations are designed to use the MCMC draw sequence without additional simulation.

The algorithm is shown to be reliable and fast. The extensions impose very little additional computational burden over the basic SVOL model. For typical sample sizes, 100,000 draws of the full model are generated in about 30 min with a 400 Mhz workstation.

The performance of our method is illustrated with simulated data. The parameters of the extended model are estimated efficiently. We also illustrate the consequences of model mis-specification. If the data is generated from the basic SVOL model, using the extended model imposes little cost in terms of reduced precision. However, if the basic model is used in the presence of mis-specification significant biases occur in both volatility and parameter estimation.

Application of the model to equity index return and exchange rate time series provides ample evidence in support of the extensions. All but one of the series require fat-tailed errors, although the evidence is not as strong for the weekly as for the daily series. The equity indices and the Canadian dollar exchange rate exhibit a strong leverage effect. Overall, the Bayes factors overwhelmingly support the extended model. The fat-tailed model is more resistant to outliers than the basic model and provides smaller conditional variance estimates in periods of high volatility. The leverage model leads to higher conditional variance estimates when the shock to the conditional mean is negative. These differences are often large enough to matter when variance forecasts are needed, as in asset allocation, risk management, and option pricing.

References

Black, F., 1976. Studies of stock market volatility changes. Proceedings of the American Statistical Association, Business and Economic Statistics Section, pp. 177–181.

Bollerslev, T., 1987. A conditionally heteroskedastic time-series model for security prices and rates of return. Review of Economics and Statistics 69, 542–547.

Bollerslev, T., Chou, R., Kroner, K., 1992. ARCH modelling in finance: a review of the theory and empirical evidence. Journal of Econometrics 52, 5–59.

Carlin, B.P., Polson, N.G., 1991. Inference for non-conjugate Bayesian models using the Gibbs sampler. Canadian Journal of Statistics 19, 399–405.

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

Chib, S., Greenberg, 1995. Understanding the Metropolis–Hastings algorithm. The American Statistician 49, 327–335.

Chib, S., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics 108, 281–316.

Christoffersen, P., Diebold, F., 1997. How relevant is volatility forecasting for risk management. Working Paper, Pennsylvania Economics Department.

Diaconis, P., Strook, D., 1991. Geometric bounds for eigen values of Markov chains. Annals of Applied Probability 1, 36–61.

Dickey, J., 1971. The weighted likelihood ratio, linear hypotheses on normal location parameters. Annals of Mathematical Statistics 42, 204–224.

Dumas, B., Fleming, J., Whaley, R., 1998. Implied volatility functions: empirical tests. Journal of Finance 53, 2059–2106.

Eraker, B., Jacquier, E., Polson, N., 1998. Pitfalls in MCMC algorithms. Department of Econometrics and Statistics. Graduate School of Business, University of Chicago Technical Report, 98-04.

Fridman, M., Harris, L., 1998. A maximum likelihood approach for non-Gaussian stochastic volatility models. Journal of Business and Economics Statistics 16 (3), 284–291.

Gallant, A.R., Hsieh, D., Tauchen, G., 1997. Estimation of stochastic volatility models with diagnostics. Journal of Econometrics 81 (1), 159–192.

Geweke, J., 1989a. Exact predictive densities for linear models with ARCH disturbances. Journal of Econometrics 40, 63–86.

Geweke, J., 1989b. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317–1339.

Geweke, J., 1992. Evaluating the accuracy of sampling-base approaches to the calculation of posterior moments. In: Berger, J.O., Bernardo, J.M., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 4. Oxford University Press, Oxford, pp. 169–194.

Geweke, J., 1993. Bayesian treatment of the independent student-t linear model. Journal of Applied Econometrics 8, S19–S40.

Geweke, J., 1994a. Priors for macroeconomic time series and their application. Econometric Theory 10, 609–632.

Geweke, J., 1994b. Bayesian comparison of econometric models. Working Paper, Federal Reserve Bank of Minneapolis Research Department.

Geweke, J., 1994c. Comment on Bayesian analysis of stochastic volatility. Journal of Business and Economics Statistics 12 (4), 371–417.

Glosten, L., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatility. Journal of Finance 3, 281–300.

Jacquier, E., Polson, N., 2000. Odds ratios for non-nested models: application to stochastic volatility models. Boston College Working Paper.

Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models (with discussion). Journal of Business and Economic Statistics 12 (4), 371–417.

Kass, R., Raftery, A., 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.

Mahieu, R., Schotman, P., 1998. An empirical application of stochastic volatility models. Journal of Applied Econometrics 13 (4), 333–360.

McCulloch, R., Polson, N., Rossi, P., 2000. Bayesian analysis of the multinomial probit with fully identified parameters. Journal of Econometrics 99, 173–193.

Meddahi, N., Renault, E., 2000. Temporal aggregation of volatility models. CIRANO Discussion Paper 2000s-22.

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

Newton, M., Raftery, A., 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society Series B 57, 3–48.

Shephard, N., Kim, S., 1994. Comment of Bayesian analysis of stochastic volatility by Jacquier, Polson, and Rossi. Journal of Business and Economics Statistics 12 (4), 371–417.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701–1762.

Wild, P., Gilks, W., 1993. Adaptive rejection sampling from log-concave densities. Journal of the Royal Statistical Society Series C 42, 701–708.