No. 785

Bayesian Analysis of Stochastic Volatility Models with Heavy-Tailed Distributions

by

Manabu Asai

July 1998

1 Introduction

Changes in asset return variance or volatility over time, may be modeled using the GARCH models, developed by Engle (1982) and Bollerslev (1986). In GARCH models, such effects are captured by letting the conditional volatility be a function of squares of previous observations and past volatilities. Since the models are formulated in terms of the conditional distribution, the maximum likelihood estimation may be implemented in a straightforward way. A wide range of GARCH models have now appeared in the econometric literature; see, for example, a survey by Bollerslev *et al.* (1995).

An alternative approach is to use an unobserved volatility component model. The logarithm of an unobserved volatility is modeled as a linear stochastic process, such as an autoregression. Models of this kind are called stochastic volatility (SV) models. A simple SV model is given by $y_t = \sqrt{h_t}\eta_t$ and $\ln h_t = \gamma + \phi \ln h_{t-1} + \sigma_\nu \nu_t$, where $\nu_t \sim \text{NID}(0,1)$ and η_t has a mean of zero and a variance of one, and is generated independently of ν_t. Many researchers assume η_t to have the normal distribution. Compared to the GARCH models, the SV models are more general in several respects. The statistical properties of SV models are obtained easily from the properties of the process generating the volatility component. Their main disadvantage, however, is that they are difficult to estimate by the maximum likelihood method. Taylor (1986), Melino and Turnbull (1990) and Andersen and Sørensen (1996) used the method of moments (MM) to avoid the integration problems associated with evaluating the likelihood directly. Nelson (1988), Harvey, Ruiz, and Shephard (1994) and Ruiz (1994) employed approximate Kalman filtering methods in their quasi maximum likelihood (QML) estimation. The Monte Carlo evidence of Jacquier, Polson, and Rossi (1994), however, implies that MM and Kalman filtering procedures suffer from poor finite sample performance because they do not depend on the exact likelihood.

When researchers can neglect computational costs, there are better alternatives based on the exact likelihood; Danielsson and Richard (1993) and Danielsson (1994a) proposed simulation-based maximum likelihood procedures; Watanabe (1997) developed nonlinear filtering maximum likelihood (NFML) procedures; Jacquier, Polson, and Rossi (1994) suggested a Bayesian inference and used the Markov chain Monte Carlo (MCMC). Although these methods are computationally intensive, experimental results of Jacquier, Polson, and Rossi (1994), Danielsson (1994b) and Watanabe (1997) show that these estimators outperform MM and Kalman filter approaches.

These computationally intensive methods have certain advantages, but they still leave much room for extensions. In the first place, taking account of empirical results that many financial time series are well described as ARMA processes, we need a estimation method for ARMA-SV models. In the second place, SV models may be generalized using heavy-tailed distributions. This is important because the kurtosis in many financial series is greater than the kurtosis which results from incorporating stochastic volatility into a Gaussian process.

In this paper, a Bayesian MCMC technique is developed to estimate SV models, which are more general in two respects: (i) that allow η_t in a SV process to follow a heavy-tailed distribution and (ii) that allow y_t to follow an ARMA-SV process. Regarding the first point, a natural candidate for the distribution of η_t is the Student t distribution. Ruiz (1994) used a scaled t distribution and applied his QML method via the Kalman filter. Shephard and Pitt (1997) also used a scaled t distribution and developed a Bayesian MCMC technique based on the likelihood approach. In this paper, we assume that η_t has a Generalized Error Distribution (hereafter, GED), which is favored by Harvey (1981), Box and Tiao (1992), and Nelson (1991)[1].

Regarding the second point, Asai (1998) proposed a new QML method to estimate ARMA-SV models. The log-GARCH models, proposed by Geweke (1986) and Pantula (1986), are the logarithmic extension of the GARCH models. Their models may be interpreted as a special case of the Exponential GARCH models, originally developed by Nelson (1991). Asai (1998) derived a log-GARCH representation of a class of SV models, including linear regression models with ARMA(p,q)-SV errors. He proposed a new QML method, and conducted Monte Carlo experiments to analyze the finite-sample properties of his method to estimate simple SV models. His log-GARCH approach has an advantage over the NFML procedure with respect to computational

[1] Box and Tiao (1992) call the GED the exponential power distribution

burden, although the finite sample performance of this method is not as good as those of other computer-intensive methods. Thus, it is worth developing a computationally intensive method for ARMA-SV models.

The organization of this paper is as follows. The rest of this section briefly surveys MCMC methods. Section 2 deals with simple heavy-tailed SV models instructively. Section 3 introduces ARMA-SV models with heavy-tailed distributions, and develops a Bayesian MCMC method, based on the results in section 2. Section 4 briefly reports empirical findings for the yen/dollar daily exchange rate. These findings indicate that the heavy-tailed distribution is preferable to the normal distribution. Section 5 concludes the paper.

Before turning to our main results, we briefly introduce an MCMC method. The MCMC method has been used recently to simulate complex multivariate distributions. The Gibbs sampling algorithm is the best known of these methods, and its impact on Bayesian statistics, following the work of Tanner and Wong (1987) and Gelfand and Smith (1990), has been immense as detailed in many articles, *e.g.*, Smith and Roberts (1993), Tanner (1993), and Chib and Greenberg (1994). Many statisticians also use the more general Metropolis-Hastings (MH) algorithm. This technique was originally developed by Metropolis *et al.* (1953) and subsequently generalized by Hastings (1970); see Tierney (1994) and Chib and Greenberg (1996).

Let $\pi(\omega_j|Y)$ be the density of the target distribution from which we want to generate ω_j. If it is easy to generate a sequence from the full conditional posterior density of ω_j given Y and the rest of parameters,$p(\omega_j|Y, \omega_{-j})$ where ω_{-j} denotes all parameters except for ω_j, the Gibbs sampling algorithm may be applied in the following way. For the $(i+1)$th iteration of the Markov chain, $\omega_1^{(i+1)}$ is drawn from $p(\omega_1|Y, \omega_{-1}^{(i)})$, $\omega_2^{(i+1)}$ is drawn from $p(\omega_2|Y, \omega_{-2}^{(i)})$, and so on. Here, $\omega_{-j}^{(i)} = (\omega_1^{(i+1)}, \ldots, \omega_{j-1}^{(i+1)}, \omega_{j+1}^{(i)}, \ldots, \omega_k^{(i)})'$. Then, under weak conditions, $\omega_j^{(i)}$ converges to a random draw from $\pi(\omega_j|Y)$. Furthermore, it can be shown that

$$\frac{1}{n}\sum_{i=1}^{n}\omega_j^{(i)} \to E(\omega_j^{(i)}|Y),$$

where n is the number of Gibbs runs. It can also be shown that the sample mean of any function of $\omega_j^{(i)}$ convergens to its expectation.

The MH algorithm may be helpful in cases where it is not straightforward to sample from $p(\omega_j|Y, \omega_{-j})$. For our setup, the sampling scheme becomes:

1. Initialize $\omega^{(0)}$. Set $i = 0$.

2. Generate a new candidate value ω_j^c based on the current value $\omega_{-j}^{(i)}$ by using a candidate generating density, $q(\omega_j^c|Y, \omega_{-j})$ and write $\omega_j^{(i+1)} = \omega_j^c$ with probability

$$\text{Min}\left(\frac{p(\omega_j^c|Y, \omega_{-j})q(\omega_j^{(i)}|Y, \omega_{-j})}{p(\omega_j^{(i)}|Y, \omega_{-j})q(\omega_j^c|Y, \omega_{-j})}, 1\right),$$

   or reject and keep $\omega_j^{(i+1)} = \omega_j^{(i)}$.

3. Set $i = i + 1$ and goto 2.

As we repeat the MH algorithm, the distribution of $\omega_j^{(i)}$ converges to the target distribution under a regularity condition.

In Bayesian inference, we evaluate the posterior statistics of a parameter of interest such as its mean, standard deviation, median, and quantiles. For example, when we want to obtain the posterior mean of a parameter ω_j in $\omega : k \times 1$,

$$E(\omega_j|Y) = \int \omega_j \pi(\omega|Y) d\omega,$$

we need to calculate multiple integration. Since it is difficult to perform multiple integration analytically and/or to evaluate the likelihood function in the SV model, Jacquier, Polson, and Rossi (1994) applied an MCMC method to obtain the posterior statistics or to marginalize the joint posterior density.

2 SV models with heavy-tailed distributions

In this section, we focus on simple SV models to show our MCMC method for heavy-tailed distributions instructively. ARMA-SV models with heavy-tailed distributions are discussed in section 3.

We consider the following model:

$$y_t = \sqrt{h_t}\eta_t, \tag{1}$$
$$\ln h_t = \gamma + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \quad \nu_t \sim \text{NID}(0,1), \tag{2}$$

where η_t is a white noise process with unit variance, generated independently of ν_t. We assume that η_t has the GED instead of the standard normal distribution. The density of a GED random variable normalized to have a mean of zero and a variance of one is given by

$$p(\eta|\lambda) = a(\lambda)\ \exp\left[-b(\lambda)|\eta|^{2/(1+\lambda)}\right] \tag{3}$$

where

$$a(\lambda) = \frac{\left[\Gamma\left(\frac{3}{2}(1+\lambda)\right)\right]^{1/2}}{(1+\lambda)\left[\Gamma\left(\frac{1}{2}(1+\lambda)\right)\right]^{3/2}}, \quad b(\lambda) = \left[\frac{\Gamma\left(\frac{3}{2}(1+\lambda)\right)}{\Gamma\left(\frac{1}{2}(1+\lambda)\right)}\right]^{1/(1+\lambda)},$$

and $-1 < \lambda \leq 1$. When $\lambda = 0$, η has the standard normal distribution. For $\lambda > 0$, the distribution of η has thicker tails than normal and for $\lambda < 0$, the distribution of η has thinner tails than normal. In particular, when $\lambda = 1$, the distribution is the double exponential. When λ tends to -1, it can be shown that the distribution tends to the rectangular distribution. The GED is sometimes used in ARCH models, as well as the t distribution, to describe the tail-thickness of the distributin of asset returns. While a t-variate has moments depending on the degrees of freedom, a GED-variate has arbitrary finite moments; see Nelson (1991). Thus, if we assume that η_t has the GED, then y_t in equations (1) and (2) has arbitrary finite moments.

Let $Y \equiv (y_1, \cdots, y_T)' : T\times 1$, and $h \equiv (h_1, \cdots, h_T)' : T\times 1$. We assume the latent variable vector h is generated by (2), and that the data Y are generated by (1). Let the prior distributions for the parameter vector $\omega \equiv (\gamma, \phi, \sigma_\nu^2, \lambda)'$ be $\check{\phi}|\sigma_\nu^2 \sim N_2(\phi_0, \sigma_\nu^2 \Phi_0^{-1}) I_{S_\phi}$, $\sigma_\nu^2 \sim IG(\nu_0/2, s_0/2)$, and $\lambda \sim \text{Uni}(-1,1)$, where $\check{\phi} = (\gamma, \phi)' \sim 2 \times 1$, N_i is the i-variate Gaussian distribution, IG is the inverted gamma distribution, Uni is the uniform distribution, I_S is the indicator function of the set S, S_ϕ is the set of ϕ that satisfies $|\phi| < 1$, and the hyperparameters ϕ_0, Φ_0, ν_0, and s_0 are known.

The joint posterior of (h, ω) is given by the Bayes theorem,

$$\pi(h, \omega|Y) \propto p(Y|h, \lambda)p(h|\omega_{-\lambda})p(\omega)$$

where $\omega_{-\lambda}$ denotes all the parameters in ω other than λ.

$$\begin{aligned} p(h|\omega_{-\lambda}) &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma_\nu} \exp\left[-\frac{1}{2\sigma_\nu^2}\left(\ln h_t - \gamma - \phi \ln h_{t-1}\right)^2\right], \\ p(Y|h, \lambda) &= \prod_{t=1}^{T} a(\lambda) \exp\left[-b(\lambda)\left(y_t^2/h_t\right)^{1/(1+\lambda)}\right]. \end{aligned}$$

The key to constructing the appropriate Markov chain sampler is found by breaking the joint posterior into various conditional distributions. We simulate ω and h from the following conditional densities: $\pi(\check{\phi}|Y, \omega_{-\check{\phi}}, h)$, $\pi(\sigma_\nu^2|Y, \omega_{-\sigma_\nu^2}, h)$ and $\pi(\lambda|Y, \omega_{-\lambda}, h)$, $\pi(h_t|y_t, \omega, h_{t-1}, h_{t+1})$.

Let $LH = (\ln h_1, \cdots, \ln h_T)' : T \times 1$, $LX = \begin{bmatrix} 1 & \cdots & 1 \\ \ln h_0 & \cdots & \ln h_{T-1} \end{bmatrix}' : T \times 2$, $\mu_t = \left((1-\phi)\gamma + \phi(\ln h_{t+1} + \ln h_{t-1})\right)/(1+\phi^2)$ and $\sigma_h^2 = \sigma_\nu^2/(1+\phi^2)$.

We now present the full conditional distributions that are used in the simulation of the simple SV model. By assuming the normal-gamma conjugate prior above[2], the full conditional distributions for $(\check{\phi}, \sigma_\nu^2, \lambda)$ and h_t are given by

(i) $\quad \check{\phi}|Y, \omega_{-\check{\phi}}, h \sim N_2\left((LX'LX + \Phi_0)^{-1}(LX'LH + \Phi_0\phi_0), \sigma_\nu^2(LX'LX + \Phi_0)^{-1}\right) I_{S_\phi},$

(ii) $\quad \sigma_\nu^2|Y, \omega_{-\sigma_\nu^2}, h \sim IG\left((T + \nu_0 + 1)/2, ((\check{\phi} - \phi_0)'\Phi_0(\check{\phi} - \phi_0) + (LH - LX\check{\phi})'(LH - LX\check{\phi}) + s_0)/2\right),$

(iii) $\quad \pi(\lambda|Y, \omega_{-\lambda}, h) \propto \prod_{t=1}^{T} a(\lambda) \exp\left[-b(\lambda)\left(y_t^2/h_t\right)^{1/(1+\lambda)}\right], \quad -1 < \lambda \le 1,$

(iv) $\quad \pi(h_t|y_t, \omega, h_{t-1}, h_{t+1}) \propto h_t^{-1/2} \exp\left[-b(\lambda)\left(y_t^2/h_t\right)^{1/(1+\lambda)} - \frac{(\ln h_t - \mu_t)^2}{2\sigma_h^2}\right].$

While sampling from the full conditional distributions of $\check{\phi}, \sigma_\nu^2$ can be done by the Gibbs sampling algorithm, sampling from λ and h_t requires the MH accept/reject algorithm.

---

[2] It is straightforward to modify our full conditional distributions to the flat prior case.

Although Jacquier, Polson, and Rossi (1994) argued that it was difficult to find the bounding function of $\pi(h_t|y_t, \omega, h_{t-1}, h_{t+1})$, we can derive it, using the idea of Shephard and Pitt (1997). Note that although Shephard and Pitt (1997) applied their idea to SV models with Gaussian distributions, their essential procedure may be extended without change. The logarithm of the density of $h_t|y_t, \omega, h_{t-1}, h_{t+1}$ is by (iv),

$$\ln(\pi(h_t|y_t, \omega, h_{t-1}, h_{t+1})) = \text{const} + \ln(p_t^*),$$

where

$$\ln(p_t^*) = -\frac{1}{2}\ln\, h_t - b(\lambda)\left(y_t^2/h_t\right)^{1/(1+\lambda)} - \frac{1}{2\sigma_h^2}\left(\ln h_t - \mu_t\right)^2. \tag{4}$$

Let $h_t = \exp\, v_t$ and using the first-order Taylor series expansion of $\exp(-v_t)$ around $v_t^* (= \ln\, h_t^*)$, we obtain

$$\begin{aligned}\left(\frac{1}{h_t}\right)^{1/(1+\lambda)} = \exp\left(-\frac{v_t}{1+\lambda}\right) &\geq \exp\left(-\frac{v_t^*}{1+\lambda}\right) - \frac{(v_t - v_t^*)}{1+\lambda}\exp\left(-\frac{v_t^*}{1+\lambda}\right) \\ &= \left(\frac{1}{h_t^*}\right)^{1/(1+\lambda)}\left(1 + \frac{(\ln h_t^* - \ln h_t)}{1+\lambda}\right).\end{aligned}$$

Rewriting (4) using the above inequality, we have a bounding function,

$$\begin{aligned}\ln(p_t^*) &\leq -\frac{1}{2}\ln\, h_t - b(\lambda)\left(\frac{y_t^2}{h_t^*}\right)^{1/(1+\lambda)}\left(1 + \frac{(\ln h_t^* - \ln h_t)}{1+\lambda}\right) - \frac{(\ln\, h_t - \mu_t)^2}{2\sigma_h^2} \\ &= \ln(g_t^*).\end{aligned}$$

The normalized version of g_t^* is a Gaussian density, which has the mean and the variance

$$m_t^* = \mu_t + \frac{\sigma_h^2 b(\lambda)}{1+\lambda}\left(\frac{y_t^2}{h_t^*}\right)^{1/(1+\lambda)} - \frac{\sigma_h^2}{2}, \quad \text{and} \quad \sigma_h^2.$$

Hence, we can sample from $p(h_t|\cdot)$ by proposing $h_t \sim LN(m_t^*, \sigma_h^2)$ and accepting with probability p_t^*/g_t^*. Following Pitt and Shephard (1995), we choose $h_t^* = \exp(\mu_t)$ as the point at which a Taylor series expansion may be carried out[3].

[3]Watanabe (1996) proposed the following procedure: we first derive the value of h_t, which corresponds to the peak of p_t^* very roughly by iterating the Newton method for a few times; then we choose h_t^* so that the peaks of p_t^* and g_t^* coincide. Applying this method may produce slightly higher acceptance rates for the MH algorithm.

We derive the approximate function of the full conditional distribution for λ. We define

$$M(\lambda) \equiv \ln(\pi(\lambda|Y, \omega_{-\lambda}, h)) = T \ln a(\lambda) - b(\lambda) \sum_{i=1}^{T} (y_t^2/h_t)^{1/(1+\lambda)}.$$

Using a second-order Taylor series expansion[4] of $M(\lambda)$ around λ^*, we obtain

$$M(\lambda) \simeq M(\lambda^*) + M'(\lambda^*)(\lambda - \lambda^*) + \frac{M''(\lambda^*)}{2}(\lambda - \lambda^*)^2 \quad (5)$$

$$= \text{const} + \frac{M''(\lambda^*)}{2}(\lambda - \mu^*)^2 \quad (6)$$

where

$$\mu^* = \lambda^* - \frac{M'(\lambda^*)}{M''(\lambda^*)}.$$

If λ^* is chosen so that $M'(\lambda) = 0$ using Newton's iteration procedure, then $M'(\lambda^*) = 0$ and the candidate of λ, say, λ^c may be generated by $N(\lambda^*, -1/M''(\lambda^*))$. The acceptance probability of the MH algorithm for the kth sampling is

$$\text{Min}\left(\exp\left(M(\lambda^c) - M(\lambda^{(k)}) + \frac{M''(\lambda^*)}{2}(\lambda^{(k)} - \lambda^*)^2 - \frac{M''(\lambda^*)}{2}(\lambda^c - \lambda^*)^2\right), 1\right). \quad (7)$$

The derivatives of $M(\lambda)$ are given in Appendix.

When $\lambda = 0$, the above results may be reduced to the case of SV models with Gaussian distributions treated in Shephard and Pitt (1997).

3 MCMC method for ARMA-SV models

3.1 The Model and Prior Assumptions

In this section, we extend the results of SV models with the GED presented in the previous section to models in a more general framework. We consider

[4] $M'(\lambda)$ denotes the partial derivative of $M(\lambda)$ by λ.

a linear regression model with ARMA(p,q)-SV errors, or simply ARMA-SV model:

$$\begin{aligned} y_t &= X_t\delta + u_t, && (8)\\ A(L)u_t &= B(L)e_t, && (9)\\ e_t &= \sqrt{h_t}\eta_t, && (10)\\ \ln h_t &= W_t\gamma + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \quad \nu_t \sim \text{NID}(0,1), && (11) \end{aligned}$$

where η_t has the normalized version of GED, generated independently of ν_t. X_t and W_t are $1\times k$ vector and $1\times l$ vector, respectively. δ is a $k\times 1$ parameter vector and γ is a $l\times 1$ parameter vector.

$$A(L) = 1 - \alpha_1 L - \cdots - \alpha_p L^p,$$
$$B(L) = 1 - \beta_1 L - \cdots - \beta_q L^q,$$

and L is the lag operator. This ARMA-SV model is a straightforward extension if we take account of the empirical result that many asset return series may be expressed as ARMA processes.

Let $\alpha \equiv (\alpha_1,\cdots,\alpha_p)' : p\times 1$, $\beta \equiv (\beta_1,\cdots,\beta_q)' : q\times 1$, $\omega_1 \equiv (\delta',\alpha',\beta',\lambda)' : (k+p+q+1)\times 1$, $\omega_2 \equiv (\gamma',\phi,\sigma_\nu^2)' : (l+2)\times 1$, $\omega \equiv (\omega_1',\omega_2')' : (k+l+p+q+3)\times 1$, $Y \equiv (y_1,\cdots,y_T)' : T\times 1$, $X \equiv (X_1',\cdots,X_T')' : T\times k$, $W \equiv (W_1',\cdots,W_T')' : T\times l$, and $h \equiv (h_1,\cdots,h_T)' : T\times 1$.

We consider a state space expression of the ARMA model (8)-(10), given the latent variable vector h:

$$\begin{aligned} y_t &= X_t\delta + z a_t + G_t\eta_t, && (12)\\ a_{t+1} &= T a_t + H_t\eta_t, && (13) \end{aligned}$$

where $z \equiv [1,0,\cdots,0] : 1\times m$, $G_t \equiv h_t^{1/2}$,

$$T \equiv \begin{bmatrix} \alpha_1 & & & \\ \alpha_2 & & I_{m-1} & \\ \vdots & & & \\ \alpha_m & 0 & \cdots & 0 \end{bmatrix} : m\times m, \quad H_t \equiv h_t^{1/2} \begin{bmatrix} \alpha_1-\beta_1 \\ \alpha_2-\beta_2 \\ \vdots \\ \alpha_m-\beta_m \end{bmatrix} : m\times 1,$$

$m = \max\{p,q\}$, $\alpha_j = 0$ for $j > p$, $\beta_j = 0$ for $j > q$, and η_t has the GED with mean zero and variance one. We assume $a_0 = 0$. Obviously, $e_t = \sqrt{h_t}\eta_t$. This state space expression[5] is not the same as the typical expression of the ARMA model, given in textbooks such as Harvey (1993) or Hamilton (1994). We can, however, easily verify that this expression reduces to equations (8)–(10).

We now assume that the latent variable vector h is generated by (11), and the data Y are generated by (8)-(10), with p,q,known. We consider the following constraints;

C1: All roots of $A(L) = 0$ lie outside of the unit circle.

C2: All roots of $B(L) = 0$ lie outside of the unit circle.

C3: $|\phi| < 1$.

C1 and C2 are related to the stationarity and invertibility of the error term. C3 guarantees the stationarity of the latent variable $\ln h_t$.

Let the prior distribution of the vector ω, $p(\omega)$, be given by

$$
\begin{aligned}
p(\omega) &= p(\omega_1)p(\omega_2), \\
p(\omega_1) &= p(\delta)p(\alpha)p(\beta)p(\lambda), \\
&\quad \delta \sim N_k(\delta_0, D_0^{-1}), \quad \alpha \sim N_p(\alpha_0, A_0^{-1})I_{S_\alpha}, \\
&\quad \beta \sim N_q(\beta_0, B_0^{-1})I_{S_\beta}, \quad \lambda \sim \text{Uni}(-1,1), \\
p(\omega_2) &= p(\check{\phi}|\sigma_\nu^2)p(\sigma_\nu^2), \\
&\quad \check{\phi}|\sigma_\nu^2 \sim N_{l+1}(\phi_0, \sigma_\nu^2\Phi_0^{-1})I_{S_\phi}, \quad \sigma_\nu^2 \sim IG(\nu_0/2, s_0/2),
\end{aligned}
$$

where $\check{\phi} = (\gamma', \phi)' : (l+1) \times 1$, I_S is the indicator function of the set S, S_α is the set of a that satisfies C1, S_β is the set of b that satisfies C2, S_ϕ is the set of ϕ that satisfies C3, and the hyperparameters δ_0, D_0, α_0, A_0, β_0, B_0, ϕ_0, Φ_0, ν_0, and s_0 are known.

Nakatsuma (1996) shows that y_t is given by

$$
y_t = X_t\delta + \sum_{j=1}^{p} \alpha_j(y_{t-j} - X_{t-j}\delta) + e_t - \sum_{j=1}^{q} \beta_j e_{t-j} + (\alpha_t - \beta_t)e_0, \tag{14}
$$

[5]State space representations of this type are found in De Jong (1991), Koopman (1993), De Jong and Shephard (1995), and Nakatsuma (1996), among others.

and y_t does not depend on e_0 for $t \geq m$ $(= \max\{p, q\})$. One advantage of our state space expression is that we only need to obtain e_0, which is a scalar, in order to start the recursion of the state space model. Chib and Greenberg (1994) derived a similar expression for a regression model with an ARMA(p, q) error. Their expression, however, is based on the initial state variable α_0, instead of the initial error term e_0. Another advantage of our state space model is that we can evaluate e_0 given the data, unobserved volatilities and the rest of the parameters without using any smoothing. Using equation (14), we obtain the following equations:

$$\begin{aligned}
y_1 - X_1\delta &= (\alpha_1 - \beta_1)e_0 + e_1, \\
y_2 - X_2\delta - \alpha_1(y_1 - X_1\delta) &= (\alpha_2 - \beta_2)e_0 + e_2 - \beta_1 e_1, \\
y_3 - X_3\delta - \alpha_1(y_2 - X_2\delta) - \alpha_2(y_1 - X_1\delta) &= (\alpha_3 - \beta_3)e_0 + e_3 - \beta_1 e_2 - \beta_2 e_1, \\
&\vdots \\
y_T - X_T\delta - \sum_{j=1}^{p}\alpha_j(y_{T-j} - X_{T-j}\delta) &= (\alpha_T - \beta_T)e_0 + e_T - \sum_{j=1}^{q}\beta_j e_{T-j},
\end{aligned}$$

or

$$Pu = Ze_0 + Q\Sigma^{1/2}\eta,$$

where $u \equiv (y_1 - X_1\delta, y_2 - X_2\delta, \cdots, y_T - X_T\delta)' : T \times 1$, $Z \equiv (\alpha_1 - \beta_1, \alpha_2 - \beta_2, \cdots, \alpha_T - \beta_T)' : T \times 1$, $\eta \equiv (\eta_1, \eta_2, \cdots, \eta_T) : T \times 1$, $\Sigma \equiv \text{diag}\{h_1, h_2, \cdots, h_T\} : T \times T$,

$$P \equiv \begin{bmatrix}
1 & & & & & & \\
-\alpha_1 & 1 & & & & & \\
-\alpha_2 & -\alpha_1 & 1 & & & & \\
\vdots & & & \ddots & & & \\
-\alpha_p & \cdots & -\alpha_2 & -\alpha_1 & 1 & & \\
\vdots & & & & & \ddots & \\
0 & \cdots & -\alpha_p & \cdots & -\alpha_2 & -\alpha_1 & 1
\end{bmatrix} : T \times T,$$

$$Q \equiv \begin{bmatrix} 1 & & & & & & \\ -\beta_1 & 1 & & & & & \\ -\beta_2 & -\beta_1 & 1 & & & & \\ \vdots & & & \ddots & & & \\ -\beta_q & \cdots & -\beta_2 & -\beta_1 & 1 & & \\ \vdots & & & & & \ddots & \\ 0 & \cdots & -\beta_q & \cdots & -\beta_2 & -\beta_1 & 1 \end{bmatrix} : T \times T,$$

and η_t has the GED with mean zero and variance one. The estimate of e_0 is given by the GLS[6]:

$$\hat{e}_0 = [Z'(Q\Sigma Q')^{-1}Z]^{-1}Z'(Q\Sigma Q')^{-1}Pu. \tag{15}$$

As noted above, we can start the recursion of (12) and (13) once we obtain e_0 by (15).

3.2 Full Conditional Distribution

We will derive full conditional distributions to implement the MCMC in this subsection. The joint posterior of (h, ω) is given by the Bayes theorem,

$$\pi(h, \omega | Y, X, W) \propto p(Y|X, h, \omega_1)p(h|W, \omega_2)p(\omega)$$

where

$$\begin{aligned} p(h|W, \omega_2) &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma_\nu} \exp\left[-\frac{1}{2\sigma_\nu^2}(\ln h_t - W_t\gamma - \phi \ln h_{t-1})^2\right], \\ p(Y|X, h, \omega_1) &= \prod_{t=1}^{T} a(\lambda) \exp\left[-b(\lambda)\left((y_t - y_{t|t-1})^2/h_t\right)^{1/(1+\lambda)}\right], \\ y_{t|t-1} &\equiv X_t\delta + (1 - A(L))(y_t - X_t\delta) + (B(L) - 1)e_t, \end{aligned}$$

and $y_{t|t-1}$ is the one-step-ahead prediction of y_t given X_t and information up to the period $t-1$. Breaking the joint posterior into various conditional distributions is the key to constructing the appropriate Markov chain sampler. We simulate ω and h from the following conditional densities: $\pi(\delta|Y, X, W, \omega_{-\delta}, h)$, $\pi(\alpha|Y, X, W, \omega_{-\alpha}, h)$, $\pi(\beta|Y, X, W, \omega_{-\beta}, h)$, $\pi(\lambda|Y, X, W, \omega_{-\lambda}, h)$,

[6] Of course, e_0 can be estimated by the maximum likelihood method given ω and h.

$\pi(\breve{\phi}|Y, X, W, \omega_{-\breve{\phi}}, h)$, $\pi(\sigma_\nu^2|Y, X, W, \omega_{-\sigma_\nu^2}, h)$ and $\pi(h_t|y_t, X_t, W_t, \omega, h_{t-1}, h_{t+1})$, where, e.g., $\omega_{-\delta}$ denotes all the parameters in ω other than δ.

We use the data transformation derived by Chib and Greenberg (1994). Let the scalars $y_0^* = u_0$, $y_s = y_s^* = 0$ for $s < 0$ and the vectors $X_s = X_s^* = 0$, $s \leq 0$. For t=1,...,T, define

$$\begin{aligned} y_t^* &= y_t - \sum_{i=1}^{p} \alpha_i y_{t-i} - \sum_{j=1}^{q} \beta_j y_{t-j}^*, \\ X_t^* &= X_t - \sum_{i=1}^{p} \alpha_i X_{t-i} - \sum_{j=1}^{q} \beta_j X_{t-j}^*. \end{aligned}$$

Using this definition, we can easily verify that $y_1^* - X_t^* \delta = e_1$ and proceed by induction, making use of

$$y_t^* - X_t^* \delta = y_t - X_t \delta - \sum_{i=1}^{p} \alpha_i (y_{t-i} - X_{t-i}\delta) - \sum_{j=1}^{q} \beta_j (y_{y-j}^* - X_{y-j}^* \delta).$$

As a result, we obtain the regression relationship $y_t^* = X_t^* \delta + e_t$, where e_t has the GED with mean zero and variance h_t. By repeatedly applying transformations given in the following definition, we arrive at the regression relationship about α.

Let $\tilde{y}_0 = e_0$, and let the scalars $y_s = \tilde{y}_s = 0$, $s < 0$. For $s \leq 0$, let the scalars $y_s = \tilde{y}_s = 0$. For $t = 1, \ldots, T$, define

$$\begin{aligned} \tilde{y}_t &= y_t - X_t \delta - \sum_{j=1}^{q} \beta_j \tilde{y}_{t-j}, \\ \tilde{X}_t &= (\tilde{y}_{t-1}, \cdots, \tilde{y}_{t-p}) : 1 \times p. \end{aligned}$$

As mentioned above, the relationship of $\tilde{y}_1 - \tilde{X}_1 \alpha = e_1$ and

$$\tilde{y}_t - \tilde{X}_t \alpha = y_t - X_t \delta - \sum_{j=1}^{p} \alpha_j (y_{t-j} - X_{t-j}\delta) - \sum_{j=1}^{q} \beta_j (\tilde{y}_{t-j} - \tilde{X}_{t-j}\alpha),$$

can be rewritten compactly as $\tilde{y}_t = \tilde{X}_t \alpha + e_t$.

We must extend the definition of μ_t, and LX for the ARMA-SV model as $\mu_t = ((1-\phi) W_t \gamma + \phi(\ln h_{t+1} + \ln h_{t-1})) / (1 + \phi^2)$, and

$$LX = \begin{bmatrix} W_1' & \cdots & W_T' \\ \ln h_0 & \cdots & \ln h_{T-1} \end{bmatrix}' : T \times (l+1).$$

We now present the full conditional distributions that are used in the simulation of the regression model with ARMA(p,q)-SV errors. Under prior assumptions given in subsection 3.1, the full conditional distributions for $(\delta, \alpha, \beta, \check{\phi}, \sigma_\nu^2)$ and h_t are given by

$$\text{1-(i)}\quad \pi(\delta|Y,X,\omega_{-\delta},h) \propto \prod_{t=1}^{T} a(\lambda)\, \exp\left[-b(\lambda)\left((y_t^* - X_t^*\delta)^2/h_t\right)^{1/(1+\lambda)}\right] \times (2\pi)^{-k/2}|D_0|^{1/2}\exp\left[-\frac{1}{2}(\delta-\delta_0)'D_0(\delta-\delta_0)\right],$$

$$\text{1-(ii)}\quad \pi(\alpha|Y,X,\omega_{-\alpha},h) \propto \prod_{t=1}^{T} a(\lambda)\, \exp\left[-b(\lambda)\left((\tilde{y}_t - \tilde{X}_t\alpha)^2/h_t\right)^{1/(1+\lambda)}\right] \times (2\pi)^{-p/2}|A_0|^{1/2}\exp\left[-\frac{1}{2}(\alpha-\alpha_0)'A_0(\alpha-\alpha_0)\right] I_{S_\alpha},$$

$$\text{1-(iii)}\quad \pi(\beta|Y,X,\omega_{-\beta},h) \propto \prod_{t=1}^{T} a(\lambda)\, \exp\left[-b(\lambda)\left(e_t(\beta)^2/h_t\right)^{1/(1+\lambda)}\right] \times (2\pi)^{-q/2}|B_0|^{1/2}\exp\left[-\frac{1}{2}(\beta-\beta_0)'B_0(\beta-\beta_0)\right] I_{S_\beta},$$

$$\text{1-(iv)}\quad \check{\phi}|Y,X,\omega_{-\check{\phi}},h \sim N_{l+1}\left((LX'LX+\Phi_0)^{-1}(LX'LH+\Phi_0\phi_0), \sigma_\nu^2(LX'LX+\Phi_0)^{-1}\right) I_{S_\phi},$$

$$\text{1-(v)}\quad \sigma_\nu^2|Y,X,\omega_{-\sigma_\nu^2},h \sim IG\left((T+\nu_0+1)/2, ((\check{\phi}-\phi_0)'\Phi_0(\check{\phi}-\phi_0) + (LH-LX\check{\phi})'(LH-LX\check{\phi}) + s_0)/2\right),$$

$$\text{1-(vi)}\quad \pi(\lambda|Y,\omega_{-\lambda},h) \propto \prod_{t=1}^{T} a(\lambda)\, \exp\left[-b(\lambda)\left(\frac{(y_t - y_{t|t-1})^2}{h_t}\right)^{1/(1+\lambda)}\right] I_{S_\lambda},$$

$$\text{1-(vii)}\quad \pi(h_t|y_t,\omega,h_{t-1},h_{t+1}) \propto h_t^{-1/2}\exp\left[-b(\lambda)\left(\frac{(y_t - y_{t|t-1})^2}{h_t}\right)^{1/(1+\lambda)} - \frac{(\ln h_t - \mu_t)^2}{2\sigma_h^2}\right].$$

Note that our full conditional distributions can be modified easily to the flat prior case.

If we use the Gaussian distribution instead of the GED, that is, we have the prior information of $\lambda = 0$, our full conditional distribution may be reduced to

$$\text{2-(i)}\quad \delta|Y,X,\omega_{-\delta},h \sim N_k\left((D_0 + X^{*\prime}\Sigma^{-1}X^*)^{-1}(D_0\delta_0 + X^{*\prime}\Sigma^{-1}Y^*), (D_0 + X^{*\prime}\Sigma^{-1}X^*)^{-1}\right),$$

2-(ii) $\alpha|Y, X, \omega_{-\alpha}, h$

$$\sim N_p\left((A_0 + \tilde{X}'\Sigma^{-1}\tilde{X})^{-1}(A_0\alpha_0 + \tilde{X}\Sigma^{-1}\tilde{y}), (A_0 + \tilde{X}'\Sigma^{-1}\tilde{X})^{-1}\right) I_{S_\alpha},$$

2-(iii) $\pi(\beta|Y, X, \omega_{-\beta}, h) \propto \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi h_t}} \exp\left[-\frac{e_t(\beta)^2}{2h_t}\right]$

$$\times (2\pi)^{-q/2}|B_0|^{1/2}\exp\left[-\frac{1}{2}(\beta - \beta_0)'B_0(\beta - \beta_0)\right] I_{S_\beta},$$

2-(iv) $\check{\phi}|Y, X, \omega_{-\check{\phi}}, h$

$$\sim N_{l+1}\left((LX'LX + \Phi_0)^{-1}(LX'LH + \Phi_0\phi_0), \sigma_\nu^2(LX'LX + \Phi_0)^{-1}\right) I_{S_\phi},$$

2-(v) $\sigma_\nu^2|Y, X, \omega_{-\sigma_\nu^2}, h$

$$\sim IG\left((T + \nu_0 + 1)/2, ((\check{\phi} - \phi_0)'\Phi_0(\check{\phi} - \phi_0) + (LH - LX\check{\phi})'(LH - LX\check{\phi}) + s_0)/2\right),$$

2-(vi) $\pi(h_t|y_t, X_t, \omega, h_{t-1}, h_{t+1}) \propto h_t^{-1/2}\exp\left[-\frac{(y_t - y_{t|t-1})^2}{2h_t}\right] \exp\left[-\frac{(\ln h_t - \mu_t)^2}{2\sigma_h^2}\right],$

where Y^* is the $T \times 1$ column vector of the y_t^*, $X^* \equiv (X_1^{*\prime}, \cdots, X_T^{*\prime})' : T \times k$, $\tilde{Y}$ is the $T \times 1$ column vector of the $\tilde{y}_t$, $\tilde{X} \equiv (\tilde{X}_1', \cdots, \tilde{X}_T')' : T \times p$, and $\Sigma \equiv \text{diag}\{h_1, \cdots, h_T\} : T \times T$. In this case, most of the parameters may be generated by Gibbs sampling; see Asai (1997) for more detailed discussion and implementation.

3.3 Implementation Issues

By the first-order Taylor series expansion of $z^{1/(1+\lambda)}$ about z^*, we obtain

$$z^{1/(1+\lambda)} \simeq z^{*1/(1+\lambda)} + \frac{1}{1+\lambda} z^{*-\lambda/(1+\lambda)}(z - z^*).$$

Thus, we have the approximation for a given δ^*,

$$\left(\frac{(y_t^* - X_t^*\delta)^2}{h_t}\right)^{1/(1+\lambda)} \simeq \left(\frac{(y_t^* - X_t^*\delta^*)^2}{h_t}\right)^{1/(1+\lambda)}$$
$$+ \frac{1}{1+\lambda}\left(\frac{(y_t^* - X_t^*\delta^*)^2}{h_t}\right)^{-\lambda/(1+\lambda)} \left[\frac{(y_t^* - X_t^*\delta)^2}{h_t} - \frac{(y_t^* - X_t^*\delta^*)^2}{h_t}\right].$$

Therefore, the approximate function of the full conditional distribution of δ is proportional to

$$\exp\left(-\frac{1}{2}(Y^* - X^*\delta)'\Sigma_\delta(Y^* - X^*\delta) - \frac{1}{2}(\delta - \delta_0)'D_0(\delta - \delta_0)\right)$$

where

$$\Sigma_\delta \equiv \text{diag}\left\{ \frac{2b(\lambda)}{h_t(1+\lambda)} \left(\frac{(y_t^* - X_t^*\delta^*)^2}{h_t} \right)^{-\lambda/(1+\lambda)} \right\} : T \times T.$$

The candidate generating function of δ is given by the normal distribution

$$N_k\left((D_0 + X^{*\prime}\Sigma_\delta X^*)^{-1}(D_0\delta_0 + X^{*\prime}\Sigma_\delta Y^*), (D_0 + X^{*\prime}\Sigma_\delta X^*)^{-1} \right).$$

δ^* may be obtained by applying Newton's method for

$$\frac{\partial}{\partial\delta} \sum_{t=1}^{T} \left(\frac{(y_t^* - X_t^*\delta)^2}{h_t} \right)^{1/(1+\lambda)} = 0.$$

Using the analogous approximation, we have the candidate generating function of α,

$$N_p\left((A_0 + \tilde{X}'\Sigma_\alpha\tilde{X})^{-1}(A_0\alpha_0 + \tilde{X}\Sigma_\alpha\tilde{y}), (A_0 + \tilde{X}'\Sigma_\alpha\tilde{X})^{-1} \right) I_{S_\alpha},$$

where I_S is the indicator function of the set S, S_α is the set of a that satisfies C1, and

$$\Sigma_\alpha \equiv \text{diag}\left\{ \frac{2b(\lambda)}{h_t(1+\lambda)} \left(\frac{(\tilde{y}_t - \tilde{X}_t\alpha^*)^2}{h_t} \right)^{-\lambda/(1+\lambda)} \right\} : T \times T.$$

Again α^* is the solution of

$$\frac{\partial}{\partial\alpha} \sum_{t=1}^{T} \left(\frac{(\tilde{y}_t - \tilde{X}_t\alpha)^2}{h_t} \right)^{1/(1+\lambda)} = 0,$$

by Newton's procedure.

Since $e_t(\beta)$ is the nonlinear function of β, the above approximation may not be applied in a straightforward way. Thus, we use the second-order Taylor series expansion of

$$M_\beta(\beta) \equiv \sum_{t=1}^{T} \left(e_t(\beta)^2/h_t \right)^{1/(1+\lambda)}$$

around β^*,

$$M_\beta(\beta) \simeq M_\beta(\beta^*) + (\beta - \beta^*)' \frac{M_\beta''(\beta)}{2}(\beta - \beta^*),$$

where β^* is the solution of $M_\beta'(\beta) = 0$ by Newton's procedure. In this case, the candidate generating function of β becomes

$$N_q\left((B_0 + b(\lambda)M_\beta''(\beta))^{-1}(B_0\beta_0 + b(\lambda)M_\beta''(\beta)\beta^*), (B_0 + b(\lambda)M_\beta''(\beta))^{-1} \right) I_{S_\beta}.$$

The derivatives of $M_\beta(\lambda)$ are given in the Appendix.

4 Empirical Example: Daily Exchange Rates

The present data set consists of daily observations on the yen/dollar exchange rate over the period January 1, 1987 through December 31, 1997, for a total of 2870 observations[7]. A broad consensus has emerged that nominal exchange rates over the free float period are best described as a non-stationary-type process; see, *e.g.*, Baillie and Bollerslev (1989). We shall therefore concentrate on modeling the nominal returns; *i.e.*, $r_t \equiv \ln s_t - \ln s_{t-1}$, where s_t denotes the spot yen/dollar exchange rate at day t.

Instead of prefiltering the return series to take out AR terms and day-of-the-week effects in the mean returns, we consider the following MA(1)-SV model:

$$\begin{aligned} r_t &= \delta_0 + u_t \\ u_t &= e_t - b_1 e_{t-1}, \\ e_t &= \sqrt{h_t}\eta_t, \qquad \eta_t \sim \text{GED}(\lambda), \\ \ln h_t &= \gamma_0 + \gamma_1 w_t + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \qquad \nu_t \sim \text{NID}(0,1), \end{aligned}$$

where w_t denotes a weekend dummy equal to one following a closure of the market. The MA(1) term is included to take account of the weak serial dependence in the mean. We relax the assumption C3, which requires that ϕ is within $[-1, 1]$, to test the stationarity of $\ln h_t$.

For our implementation of the MCMC algorithm, we use QML estimates as the prior information. QML estimates and standard errors are obtained by the log-GARCH approach proposed by Asai (1998). Asai (1998) derived log-GARCH representation of a class of SV models, including ARMA-SV models, and proposed QML estimation based on the standardized t distribution. Smoothed estimates of h_t may be obtained using the approximate Kalman filtering method.

The iterations are started from QML estimates and smoothed estimates of h_t[8]. The Markov chain sampler is run for 20000 draws such that the first m

[7]The rate data were obtained through Datastream.

[8]Asai's (1998) log-GARCH approach assumes the normality of η_t, *i.e.,*, $\lambda = 0$. Thus, we set the starting value of λ as 0.

draws are discarded and then the next n $(= 20000-m)$ are retained. Posterior means are computed as n sample averages. Geweke (1992) recommended using methods from spectral analysis to assess convergence of the MCMC after $m + n$ iterations. Though created for the Gibbs sampler, Geweke's (1992) method may be applied to the output of any MCMC algorithm; see, for example, Cowles and Carlin (1996). If we let our estimate of a marginal posterior mean $E(\omega_.|y, w)$ be

$$\bar{\omega}_. = \frac{1}{n}\sum_{i=1}^{n}\omega_.^{(i)},$$

where $\omega_.^{(i)}$ is the $m + i$th draw of a parameter, then $\{\omega_.^{(i)}\}$ is a univariate stochastic process. Geweke (1992) shows that $\bar{\omega}_.$ is asymptotically normal with mean $E(\omega_.|y, w)$ and asymptotic variance $n^{-1}S(0)$, where $S(0)$ is the spectral density of $\{\omega_.^{(i)}\}$ evaluated at frequency zero. This finding suggests that $[n^{-1}S(0)]^{1/2}$ can be used as a numerical standard error (NSE) for $\bar{\omega}_.$ and that the calculation of the NSE of the Markov chain is feasible. In all results reported in this paper $\hat{S}(0)$ is formed from the period gram of $\{\omega_.^{(i)}\}$ using a Daniell window of width $2\pi/(.375n^{1/2})$.

Geweke's (1992) convergence diagnostics are based on the following idea: a Markov chain sampler yields draws from the posterior only as the number of passes becomes large; hence, comparison of early n_A passes with late n_B passes can reveal failures of convergence. Define

$$\bar{\omega}_.^A = \frac{1}{n_A}\sum_{j=1}^{n_A}\omega_.^{(j)},$$
$$\bar{\omega}_.^B = \frac{1}{n_B}\sum_{j=n_C}^{n}\omega_.^{(j)},$$

where $n_C = n - n_B + 1$, and let nse_A and nse_B be the numerical standard errors for the two estimates, $\bar{\omega}_.^A$ and $\bar{\omega}_.^B$, calculated as in the previous paragraph. If the sequence of $\omega_.^{(i)}$ is stationary, the ratios n_A/n and n_B/n are held fixed, and $n_A + n_B < n$, then by the central limit theorem, the distribution of the convergence diagnostic (CD),

$$(\bar{\omega}_.^A - \bar{\omega}_.^B) \Big/ \sqrt{\text{nse}_A^2 + \text{nse}_B^2}\ ,$$

approaches a standard normal as n tends to infinity. Following the suggestion of Geweke (1992), we calculated this statistic by setting $n_A = 0.1n$ and $n_B = 0.5n$.

It has sometimes been suggested that inferences should be based on every kth iteration of each sequence, with k set to some value high enough that successive draws of the parameter vector, ω, are approximately independent. This strategy cannot be used in this situation since the set of simulated values is not so large that reducing the number of simulations by a factor of k gives important savings in storage and computation time.

Table 1 presents the QML estimates via the log-GARCH approach and the Bayes results for the ARMA-SV model. For the log-GARCH approach, the robust QML covariance estimators of White (1982) are used to compute the standard errors. According to convergence diagnostics values, the null hypothesis that the sequence of 20000 samples is stationary cannot be rejected at five percent significance level for all parameters. Therefore, $m = 11000$ and $n = 9000$ are chosen. The marginal posterior means for the tail-thickness parameter, λ is 0.896. This result indicates that η_t in our model has tihcker tails than normal. The marginal posterior means for δ_0 and b_1 are almost zero. Each 95 percent highest posterior density region contains zero. These results support the empirical findings of Hsieh (1988,1989) and Baillie and Bollerslev (1989) for the levels of exchnage rate returns. The marginal posterior means for SV parameters (γ_0,γ_1,ϕ,σ_ν^2) are (-0.993,-0.00655,0.899,0.135). The 95 percent highest posterior density region for ϕ is $[0.868, 0.929]$. Thus the yen/dollar exchange rate data exhibit a high degree of persistence in volatility although the posterior is massed well away from the unit root case.

5 Concluding Remarks

In this paper, a Bayesian MCMC technique is developed to estimate SV models, which are more general in two points than previous models: (i) allowing η_t in (1) to follow a heavy-tailed distribution and (ii) allowing y_t to follow an ARMA-SV process. Using the daily yen/dollar exchange rate, we showed that our Bayesian MCMC technique performs well.

Appendix

A Derivatives of $M(\lambda)$

$$
\begin{aligned}
M'(\lambda) &= \frac{3n}{4}\left[\psi\left(\frac{3}{2}(1+\lambda)\right) - \psi\left(\frac{1}{2}(1+\lambda)\right)\right] - \frac{n}{1+\lambda} \\
&-b(\lambda)\sum_{t=1}^{T}\left(\frac{(y_t - y_{t|t-1})^2}{h_t^*}\right)^{1/(1+\lambda)}\left[\frac{1}{1+\lambda}\left(\frac{3}{2}\psi\left(\frac{3}{2}(1+\lambda)\right) - \frac{1}{2}\psi\left(\frac{1}{2}(1+\lambda)\right)\right)\right. \\
&\left.-\frac{1}{(1+\lambda)^2}\left(\ln\Gamma\left(\frac{3}{2}(1+\lambda)\right) - \ln\Gamma\left(\frac{1}{2}(1+\lambda)\right) + \ln\left(\frac{(y_t - y_{t|t-1})^2}{h_t^*}\right)\right)\right], \\
M''(\lambda) &= \frac{3n}{4}\left[\frac{3}{2}\psi\left(\frac{3}{2}(1+\lambda)\right) - \frac{1}{2}\psi\left(\frac{1}{2}(1+\lambda)\right)\right] + \frac{n}{(1+\lambda)^2} \\
&-b(\lambda)\sum_{t=1}^{T}\left(\frac{(y_t - y_{t|t-1})^2}{h_t^*}\right)^{1/(1+\lambda)}\left[\left[\frac{1}{1+\lambda}\left(\frac{3}{2}\psi\left(\frac{3}{2}(1+\lambda)\right) - \frac{1}{2}\psi\left(\frac{1}{2}(1+\lambda)\right)\right)\right.\right. \\
&\left.-\frac{1}{(1+\lambda)^2}\left(\ln\Gamma\left(\frac{3}{2}(1+\lambda)\right) - \ln\Gamma\left(\frac{1}{2}(1+\lambda)\right) + \ln\left(\frac{(y_t - y_{t|t-1})^2}{h_t^*}\right)\right)\right]^2 \\
&+\frac{1}{1+\lambda}\left(\frac{9}{4}\psi'\left(\frac{3}{2}(1+\lambda)\right) - \frac{1}{4}\psi'\left(\frac{1}{2}(1+\lambda)\right)\right) \\
&-\frac{2}{(1+\lambda)^2}\left(\frac{3}{2}\psi'\left(\frac{3}{2}(1+\lambda)\right) - \frac{1}{2}\psi'\left(\frac{1}{2}(1+\lambda)\right)\right) \\
&\left.+\frac{2}{(1+\lambda)^3}\left(\ln\Gamma\left(\frac{3}{2}(1+\lambda)\right) - \ln\Gamma\left(\frac{1}{2}(1+\lambda)\right) + \ln\left(\frac{(y_t - y_{t|t-1})^2}{h_t^*}\right)\right)\right].
\end{aligned}
$$

B Derivatives of $M_\beta(\beta)$

$$
M_\beta(\beta) \equiv \sum_{t=1}^{T}\left(e_t(\beta)^2/h_t\right)^{1/(1+\lambda)},
$$

$$
M'_\beta(\beta) = \sum_{t=1}^{T}\frac{2}{h_t(1+\lambda)}\left(\frac{e_t(\beta)^2}{h_t}\right)^{-\lambda/(1+\lambda)} e_t(\beta)\frac{\partial e_t(\beta)}{\partial\beta},
$$

$$M''_\beta(\beta) = \sum_{t=1}^{T} \frac{2}{h_t(1+\lambda)} \left(\frac{e_t(\beta)^2}{h_t}\right)^{-\lambda/(1+\lambda)} \left(\frac{1-\lambda}{1+\lambda}\frac{\partial e_t(\beta)}{\partial \beta}\frac{\partial e_t(\beta)}{\partial \beta'} + e_t(\beta)\frac{\partial^2 e_t(\beta)}{\partial\beta\partial\beta'}\right),$$

where for $t > 0$,

$$\begin{aligned}
\frac{\partial e_t(\beta)}{\partial \beta_i} &= e_{t-i}(\beta) + \sum_{s=1}^{q} \beta_s \frac{\partial e_{t-s}(\beta)}{\partial \beta_i}, \quad (i = 1, \ldots, q), \\
\frac{\partial^2 e_t(\beta)}{\partial \beta_i \partial \beta_j} &= \frac{\partial e_{t-i}(\beta)}{\partial \beta_j} + \frac{\partial e_{t-j}(\beta)}{\partial \beta_i} + \sum_{s=1}^{q} \beta_s \frac{\partial^2 e_{t-s}(\beta)}{\partial \beta_i \partial \beta_j}, \quad (i, j = 1, \ldots, q),
\end{aligned}$$

and for $t \leq 0$,

$$\frac{\partial e_t(\beta)}{\partial \beta} = 0, \qquad \frac{\partial^2 e_t(\beta)}{\partial \beta \partial \beta'} = 0.$$

References

[1] Andersen, T.G., and B.E. Sørensen, 1996, GMM estimation of a stochastic volatility model: a Monte Carlo study, Journal of Business and Economic Statistics 14, 328–353.

[2] Asai, M., 1997, A new method to estimate stochastic volatility models: a log-GARCH approach, Discussion Paper Series No.744 (Institute of Policy and Planning Sciences, University of Tsukuba, Ibaraki).

[3] Asai, M., 1998, A new method to estimate stochastic volatility models: a log-GARCH approach, forthcoming in: Journal of the Japan Statistical Society 28.

[4] Baillie, R.T. and T. Bollerslev, 1989, The message in daily exchange rates: a conditional-variance tail, Journal of Business and Economic Statistics 7, 297–305.

[5] Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307–327.

[6] Bollerslev, T., R.F. Engle and D.B. Nelson, 1995, ARCH models, in: R.F. Engle and D. McFadden, eds., The Handbook of Econometrics Vol.4 (North-Holland, Amsterdam).

[7] Box, G.E.P. and G.C. Tiao, 1992, Bayesian Inference in Statistical Analysis (John Wiley & Sons, New York, NY)

[8] Chib, S. and E. Greenberg, 1994, Bayes inference in regression models with ARMA(p,q) errors, Journal of Econometrics 64, 183–286.

[9] Chib, S. and E. Greenberg, 1996, Understanding the Metropolis–Hastings algorithm, American Statistician 49, 327–335.

[10] Cowles, M.K. and B.P. Carlin, 1996, Markov chain Monte Carlo convergence diagnostics: a comparative review, Journal of the American Statistical Association 91, 863–904.

[11] Danielsson, J., 1994a, Stochastic volatility in asset prices: estimation with simulated maximum likelihood, Journal of Econometrics 64, 375–400.

[12] Danielsson, J., 1994b, Bayesian analysis of stochastic volatilities: comment, Journal of Business and Economic Statistics 12, 393–395.

[13] Danielsson, J., and J.F. Richard, 1993, Quadratic acceleration for Monte Carlo likelihood evaluation, Journal of Applied Econometrics 8, 153–174.

[14] De Jong, P., 1991, The diffuse Kalman filter, Annals of Statistics 19, 1073–1083.

[15] De Jong, P. and N. Shephard, 1995, The simulation smoother for time series models, Biometrika 82, 339–350.

[16] Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, Econometrica 50, 987–1008.

[17] Gelfand, A.E. and A.F.M. Smith, 1990, "Sampling based approaches to calculating marginal densities," Journal of the American Statistical Association 85, 398–409.

[18] Geweke, J., 1986, Modeling the persistence of conditional variances: a comment, Econometric Reviews 5, 57–61.

[19] Geweke, J., 1992, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in: J.M. Bernardo, J. Berger, A.P. David, and A.F.M. Smith, eds., Bayesian Statistics 4 (Oxford University Press, Oxford) 169–193.

[20] Hamilton, J.D., 1994, Time Series Analysis (Princeton University Press, NJ).

[21] Harvey, A.C., 1990, The Econometric Analysis of Time Series, 2nd ed. (Philip Allan, London).

[22] Harvey, A.C., 1993, Time Series Models, 2nd ed. (MIT Press, Cambridge, MA).

[23] Harvey, A.C., E. Ruiz and N. Shephard, 1994, Multivariate stochastic variance models, Review of Economic Studies 61, 247–264.

[24] Hastings, W.K., 1970, Monte Calro sampling methods using Mrkov chains and their applications, Biometrika 57, 97–109.

[25] Hsieh, D.A., 1988, The statistical properties of daily foreign exchange rates: 1974–1983, Journal of International Economics 24, 129–145.

[26] Hsieh, D.A., 1989, Modeling heteroskedasticity in daily foreign-exchange rates, Journal of Business and Economic Statistics 7, 307–317.

[27] Jacquier, E., N.G. Polson and P.E. Rossi, 1994, Bayesian analysis of stochastic volatilities, Journal of Business and Economic Statistics 12, 371–389.

[28] Koopman, S.J., 1993, Disturbance smoother for state space models, Biometrika 80, 117–126.

[29] Melino, A. and S.M. Turnbull, 1990, Pricing foreign currency options with stochastic volatility, Journal of Econometrics 45, 239–265.

[30] Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, 1953, Equations of state calculations by fast computing machines, Journal of ChemicalPhysics 21, 1087–1092.

[31] Nakatsuma, T., 1996, Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach, Unpublished paper (Department of Economics, Rutgers University, NJ).

[32] Nelson, D.B., 1988, The time-series behaviour of stock market volatility and returns, Unpublished Ph.D. dissertation (MIT, Cambridge, MA).

[33] Nelson, D.B., 1991, Conditional heteroskedasticity in asset returns: a new approach, Econometrica 59, 347–370.

[34] Pantula, S.G., 1986, Modeling the persistence of conditional variances: a comment, Econometric Reviews 5, 79–97.

[35] Ruiz, E., 1994, Quasi-maximum likelihood estimation of stochastic volatility models, Journal of Econometrics 63, 289–306.

[36] Shephard, N., 1996, Statistical aspects of ARCH and stochastic volatility, in: D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, eds., Time Series Models in econometrics, finance and other fields (Chapman & Hall).

[37] Shephard, N. and M. Pitt, 1997, Likelihood analysis of non-Gaussian measurement time series, Biometrika 84, 653–667.

[38] Smith, A.F.M. and G.O. Roberts, 1993, Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, Journal of the Royal Statistical Society B55, 3–24.

[39] Tanner, M., 1993, Tools for Statistical Inference, 2nd ed. (Springer-Verlag, New York, NY).

[40] Tanner, M. and W.H. Wong, 1987, The calculation of posterior distributions by data augmentation, Journal of the American Statistical Association 82, 528–549.

[41] Taylor, S.J., 1986, Modeling Financial Time Series (John Wiley, Chichester).

[42] Tierney, L., 1994, Exploring posterior distributions using Markov chains, Annals of Statistics 22, 1701–1727.

[43] Watanabe, T., 1996, Permanent and transitory component models of stock return volatility: quasi-maximum likelihood estimation and Bayes inference, Discussion Paper No.434 (Institute of Economic Research, Kyoto University, Kyoto).

[44] Watanabe, T., 1997, A nonlinear filtering approach to stochastic volatility models with an application to daily stock returns, Unpublished paper (Department of Economics, Tokyo Metropolitan University, Tokyo).

Table 1: MA(1)-SV Models

	Posterior distribution					
	log-GARCH		Lower	Upper		
Parameter	estimate	Mean	95% limit	95% limit	Corr.	CD
δ_0	0.000145	6.38×10^{-5}	-3.23×10^{-5}	0.000200	0.62	0.37
	(0.000161)	(1.92×10^{-6})				
b_1	0.0537	-0.000336	-0.00255	0.00194	0.93	0.73
	(0.0271)	(0.000357)				
γ_0	-0.501	-0.993	-1.305	-0.705	0.75	-1.38
	(0.294)	(0.0145)				
γ_1	-0.144	-0.00654	-0.0537	0.0397	0.10	1.94
	(0.076)	(0.000284)				
ϕ	0.945	0.899	0.868	0.929	0.65	-1.37
	(0.032)	(0.00148)				
σ_ν^2	0.0184	0.135	0.0957	0.180	0.97	0.85
	(0.0166)	(0.00255)				
λ	—	0.896	0.863	0.915	0.71	-0.38
		(0.00185)				

Note: Numerical standard error of posterior mean is in parentheses. Correlation denotes the first-order correlation of the Markov chain run. For the log-GARCH approach, standard error is in parentheses. 9000 simulations.