J. Japan Statist. Soc.
Vol. 36 No. 2 2006 199–212

COMPARISON OF MCMC METHODS FOR ESTIMATING GARCH MODELS*

Manabu Asai**

This paper reviews several MCMC methods for estimating the class of ARCH models, and compare performances of them. With respect to the mixing, efficiency and computational requirement of the MCMC, this paper found the best method is the tailored approach based on the acceptance-rejection Metropolis-Hastings algorithm.

Key words and phrases: Bayesian inference, GARCH, Gibbs sampler, Markov chain Monte Carlo, Metropolis-Hastings algorithm.

1. Introduction

Autoregressive conditional heteroskedasticity (ARCH) models pioneered by Engle (1982) and their extended version have been proven to be very successful in modeling the volatility of financial time series; see Bollerslev *et al.* (1994). Bayesian inference on ARCH models has been implemented using the importance sampling technique proposed by Geweke (1989) and more recently using Markov chain Monte Carlo (MCMC) methods including Bauwens and Lubrano (1998), Kim *et al.* (1998), Nakatsuma (2000), Vrontos *et al.* (2000) and Mitsui and Watanabe (2003).

For each integer t, let ε_t be a model's prediction error and σ_t^2 the variance of ε_t given information at time $t-1$. The most useful ARCH parameterization is the generalized ARCH (GARCH) model introduced by Bollerslev (1986, 1987). The GARCH(p,q) model is given by

$$\varepsilon_t = \sigma_t z_t, \tag{1.1}$$

$$z_t \sim \text{i.i.d. with } E(z_t) = 0, \qquad V(z_t) = 1, \tag{1.2}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{1.3}$$

where ω, the α_i, and the β_j are nonnegative. Stationary conditions impose that $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$; for details, see Bollerslev (1986). Bollerslev (1987) assumes that z_t follows a standardized Student-t distribution.

It has long been recognized that the returns of financial assets are negatively correlated with changes in the volatilities of returns; see Black (1976) and Christie

Received November 21, 2004. Revised June 14, 2005. Accepted December 8, 2005.

*An early version of this paper is Asai and Watanabe (2004). After moving from Tokyo Metropolitan University, the work was continued by the first author.

**Faculty of Economics, Soka University, 1-236 Tangi-cho, Hachioji-shi, Tokyo 192-8577, Japan. Email: m-asai@soka.ac.jp

(1982). In the class of ARCH models, for instance, Nelson (1991) proposed the exponential GARCH (EGARCH) model, while Glosten *et al.* (1992) developed a threshold indicator function GARCH model, which is commonly called GJR model. The volatility equation of the GJR(p, q) model is given by

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} [\alpha_i^+ I(\varepsilon_{t-i} > 0) + \alpha_i^- I(\varepsilon_{t-i} \leq 0)] \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{1.4}$$

where $I(\cdot)$ denotes indicator function.

In this paper, we review the main MCMC estimator proposed for estimating ARCH models, summarize the main advantages and limitations of each of the method considered, and compare the properties of these methods, focusing on Student-t errors and asymmetric behavior of volatility by using simulated and real data. In the Bayesian analysis, various estimation methods produce the same posterior if the same prior and likelihood are used. In other words, neglecting numerical differences, all MCMC methods here yield the same posterior distribution. Based on the fact, conducting Monte Carlo experiments will produce almost the same properties, with respect to bias and mean squared errors, among all MCMC methods. Hence, we need other criterions to measure efficiency of estimation methods, which will be discussed in the next section.

The paper is organized as follows. Section 2 describes the various MCMC methods for estimating parameters of ARCH class, and their main advantages and limitations. This section is completed with an illustration with simulated data. Section 3 illustrates the results with an empirical application that compares the MCMC methods by fitting the GJR-t model to daily Tokyo stock price index. Section 5 concludes the paper.

2. MCMC estimation

As stated above, a comparison of Bayesian MCMC methods needs alternative measures of efficiency since the finite sample bias and mean squared error used by frequentists have little sense in this situation.

Desirable properties for sampling methods in MCMC are efficiency and well mixing, which yield fast convergence. In addition to these properties, computational requirements and model flexibility are important for applied econometrics.

In the following we review the main MCMC methods with respect to each sampling algorithm, and summarize the main advantages and limitations of each method. Some MCMC methods introduced here lack flexibility in model specifications, and some are inconvenient to impose a prior restriction such as $\alpha_1 + \beta_1 < 1$ in the GARCH(1, 1).

2.1. Griddy Gibbs sampler

Bauwens and Lubrano (1998) applied the griddy Gibbs sampler (GGS) proposed by Ritter and Tanner (1992) in order to conduct Bayesian inference on GJR-t models.

The idea of GGS is to form a simple approximation to the inverse cdf based on the evaluation of $\pi(\theta_i \mid \theta_{\setminus i}, y)$ on a grid of points, where $\pi(\theta \mid y)$ is the posterior distribution, and $\pi(\theta_i \mid \theta_{\setminus i} y)$ is the full conditional distribution for θ_i. When θ_i is univariate, the kernel of $\pi(\theta_i \mid \theta_{\setminus i}, y)$, conditionally on a previous draw of the conditioning parameters, can be evaluated over a grid of points. One can then compute the corresponding distribution function using deterministic integration rule. Afterwards, one can generate a draw of θ_i by inversion of the distribution at a random value sampled uniformly in $[0, 1]$. For more details of GGS and applications to GARCH models, see chapters 3 and 7 of Bauwens *et al.* (1999). Bauwens and Lubrano (1998) use trapezoidal rule of integration.

A merit of the GGS is that it is successful in dealing with the shape of posterior, such as skewness, by using smaller MCMC outputs compared to other methods. This is due to the fact that integration is done on a grid so that every direction can be explored in detail. Another advantage of the GGS is that the conditioning, which is a variance reduction technique, is easy. Conditioning means that to estimate $E(\theta_i \mid y)$, one uses $\sum_{j=1}^{M} E(\theta_i^{(j)} \mid \theta_{\setminus i}, y)/M$ instead of $\sum_{j=1}^{M} \theta_i^{(j)}/M$. This can be done by the same integration rule used in Gibbs sampling.

The main cost of the method is the evaluation of posterior density kernel. For example, if 33 point grids are used for each parameter of GARCH-t model to sample 1000 draws, then the algorithm requires 132000 functional evaluations. So it can be greedy in computational time.

Bauwens and Lubrano (1998) states that the choice of the grid of points has to be made carefully and constitutes the main difficulty in applying the GGS. Even if parameter space is bounded, such as $0 < \alpha_1 < 1$ in GARCH$(1, 1)$ models, they recommend to restrict the integration to the subset of the parameter space where the value of the posterior density is large enough to contribute the integrals.

As a general rule in MCMC, sets of parameters that are highly correlated should be treated as one block. This is related to another drawback of GGS. For this reason GGS may be inefficient in sampling, say, $(\omega, \alpha_1, \beta_1)$ in GARCH$(1, 1)$ or parameters of higher order models. Although the GGS is flexible with respect to parametric specification of a model, applying GGS to complicated model may be inefficient in sampling and take much numbers of iterations before the burn-in.

2.2. Metropolis-Hastings algorithm

Nakatsuma (2000) developed an MCMC method to estimate GARCH(p, q) models. Based on an ARMA$(\max\{p, q\}, q)$ representation of ε_t^2, Nakatsuma (2000) used Metropolis-Hastings (MH) algorithm for estimating ARMA models proposed by Chib and Greenberg (1994). The method of Nakatsuma (2000) is also an extension of Müller and Pole (1995).

Although this method, originally, deals with normal distribution for conditional density, it is easy to incorporate Student-t distribution by using the sampling method of the degree-of-freedom parameter proposed by Geweke (1993) or by Watanabe (2001). A short explanation about Watanabe (2001) is given in

Section 3.

Stationary conditions, such $\alpha_1+\beta_1 < 1$, can be taken into account in the MH algorithm, by just rejecting all pairs of (α_1, β_1) that did not obey the preceding restriction; see Gelfand *et al.* (1992).

One of the drawbacks of this method is that it is not applicable to GJR and EGARCH, since they have no usual ARMA representations. Another draw back is concerned with computational time in the sense that it needs an optimization for sampling parameters of MA part, β_i's, in each MCMC iteration.

As for the MH algorithm of Müller and Pole (1995), it can be applied to GJR and EGARCH. This method, however, needs a huge number of iterations before the MH chain converges, due to low acceptance rates of the samplers.

It is helpful to describe the MCMC method used in Vrontos *et al.* (2000), since their technique is based on Müller and Pole (1995). Vrontos *et al.* (2000) propose modeling each GARCH and EGARCH model jointly, by using reversible-jump MCMC introduced by Green (1995). In other words, the reversible-jump method can be applied where the dimension of parameter space itself one of the parameters to be simulated, corresponding to the model choice problem; for details, see, for example, Robert and Casella (1999). In this framework, sampling parameters of GARCH/EGARCH is not restrict to the MH of Müller and Pole (1995), and can be conducted by any method except for the approach of Nakatsuma (2000). But we shall make no further inquiry into this matter; to do so would involve us in a discussion of the Bayesian model comparison that is of no immediate relevance.

For notational convenience, we simply call the MH method of Nakatsuma (2000) as the 'MH' hereafter.

2.3. Adaptive rejection Metropolis sampling

Kim *et al.* (1998) used the adaptive rejection Metropolis sampling (ARMS) method proposed by Gilks *et al.* (1995) for estimating GARCH-t models.

Adaptive rejection sampling works by constructing an envelope function of the log of the target density (provided this is concave), which is then used in rejection sampling. Whenever a point is rejected, the envelope is updated to correspond more closely to the true log density, thereby reducing the chance of rejecting subsequent points. Fewer rejection steps imply fewer point-evaluations of the log density. For some models the log-concavity constraint in the density does not hold. The ARMS of Gilks *et al.* (1995) deals with this situation by performing a Metropolis step on each point accepted at a rejection step in the adaptive rejection sampling.

To conduct this method of Kim *et al.* (1998), it only requires the coding of the log-likelihood function and the prior, if the C code of ARMS written by Walley Gilks is available. Similarly to the GGS method, it is applicable when the target density is unbounded, but it is better to bound it since the log-likelihood may deliver overflow/underflow problems.

The main drawback of this method is computational requirements, which is due to functional evaluations like the GGS and acceptance rate of metropolis step like the MH.

2.4. Acceptance-rejection/Metropolis-Hastings algorithm

Mitsui and Watanabe (2003) developed a Taylored approach based on the acceptance-rejection Metropolis-Hastings (ARMH) algorithm proposed by Tierney (1994); see also Chib and Greenberg (1995) for details of ARMH algorithm. The method of Mitsui and Watanabe (2003) can work with any kind of parametric ARCH-type models. In the first step, the method maximizes the sum of the log-prior plus the log-likelihood with respect to all parameters. The second step is sampling all parameters by ARMH algorithm from a multivariate Student-t distribution with the mean that maximizes the objective function and the dispersion that is the inverse of negative hessian matrix at the mode. In this step, the proportional constant of the candidate generating density to the full conditional distribution is specified so that the former evaluated at θ^* is equal to the latter at θ^*. It repeats the second step and retains outputs after the Markov chains converge.

The advantage of this method is that it is free to functional form as the GGS and the ARMS. In addition to this, the computational requirement of the ARMH is very small since it is irrelevent to the shape of the full conditional density. The disadvantage of this method is that it may be inefficient when θ^* does not take the global maxima but the local one. This is due to the irrelevance of the full conditional distribution.

2.5. Illustration with simulated data

To illustrate the differences between the estimates of the four MCMC methods previously described, we fit the GARCH(1, 1) model with a normal conditional distribution to simulated series with $(\omega, \alpha_1, \beta_1) = (0.1, 0.1, 0.85)$ and $T = 500$. All computations reported in this article were carried out using the Ox language of Doornik (1998).

Since it is important to restrict the parameter space for using the GGS method, we use uniform priors as in Bauwens and Lubrano (1998):

$$\omega \sim U(0, 0.2), \qquad \alpha_1 \sim U(0, 0.5), \qquad \beta_1 \sim U(0.35, 0.95).$$

These priors are also used in other MCMC methods. In the four methods, common random numbers are used to keep conditions fair. We used 33 grids following Bauwens and Lubrano (1998). In connection to the endpoint restriction of the GGS, the same restrictions are imposed on the ARMS.

For each method, M draws reported above are used for calculating the posterior means, 95% intervals, and the convergence diagnostic statistics and the inverse of the efficiency factors proposed by Geweke (1992). The posterior means are computed by averaging the simulated draws. The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. Geweke (1992) suggests assessing the convergence of the MCMC by comparing values early in the sequence with those late in the sequence. Let $\theta^{(i)}$ be the i-th draw of a parameter in the recorded 10000 draws, and let $\bar{\theta}_A = \frac{1}{n_A}\sum_{i=1}^{n_A}\theta^{(i)}$ and $\bar{\theta}_B = \frac{1}{n_B}\sum_{i=10001-n_B}^{10000}\theta^{(i)}$. Using these values, Geweke (1992) proposes the following statistic called *convergence diagnostics* (CD).

$$\text{CD} = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\hat{\sigma}_A^2/n_A + \hat{\sigma}_B^2/n_B}}, \tag{2.1}$$

where $\sqrt{\hat{\sigma}_A^2/n_A}$ and $\sqrt{\hat{\sigma}_B^2/n_B}$ are standard errors of $\bar{\theta}_A$ and $\bar{\theta}_B$. If the sequence of $\theta^{(i)}$ is stationary, it converges in distribution to the standard normal. We set $n_A = 0.1M$ and $n_B = 0.5M$ and compute $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ using a Parzen window with bandwidths of $0.01M$ and $0.05M$, respectively.

We select size of the retained and discarded sample (M and N) as follows. First of all, we consider $M = \{2000, 3000, \ldots, 10000\}$ and $N = \{1000, 2000\}$ except for the GGS. For the GGS method, we also consider $M = 1000$ since it is enough to deal with the shape of posteriors. Secondly, we choose the pair M and N such that $M + N$ is the smallest among the pairs on which the MCMC converge.

As stated above, desirable properties for sampling methods in MCMC are efficiency and well mixing, which yield fast convergence. While mixing is measured by the autocorrelation time, the efficiency is compared by a ratio of the efficiency factors introduced by Geweke (1992). The idea of Geweke's (1992) efficiency factor is based on the fact that a true posterior density has its unique variance whichever methods are used to estimate them. The efficiency factor is estimated as the variance of the posterior divided by the variance of the sample mean from the MCMC sampling scheme (M times the square of the numerical standard error). The numerical standard errors are computed using a Parzen window; see Geweke (1992) for details. The inverse of the efficiency factor is called inefficiency factor, which is more used in the literature. It should be noted that the inefficiency factor is not affected by M, while the numerical standard error is $O_p(M^{-1/2})$.

Figures 1 and 2 show correlograms of the MH and ARMH simulations, respectively. Those of the GGS and ARMS are omitted, since the MH and ARMH provide two extreme cases. The correlograms for the MH indicate that important autocorrelations for all parameters at large lag lengths, while those of the ARMH vanish at sixth lag. This result implies that, compared to the ARMH, the length of the MH chain will be very large to estimate parameters as accurately as the ARMH.

Table 1 shows the MCMC estimation results along with maximum likelihood (ML) estimates and their asymptotic 95% intervals. The second and third columns of Table 1(a) report the number of iterations. The smallest number for M is specified as 2000 for the MH, ARMS, and ARMH since $M \geq 2000$ is desirable to draw histograms. But for the GGS, it is not the case. The MH method requires larger M than others due to large number of lags of autocorrelation. The fourth column presents geometric averages of inefficiency factors of each method

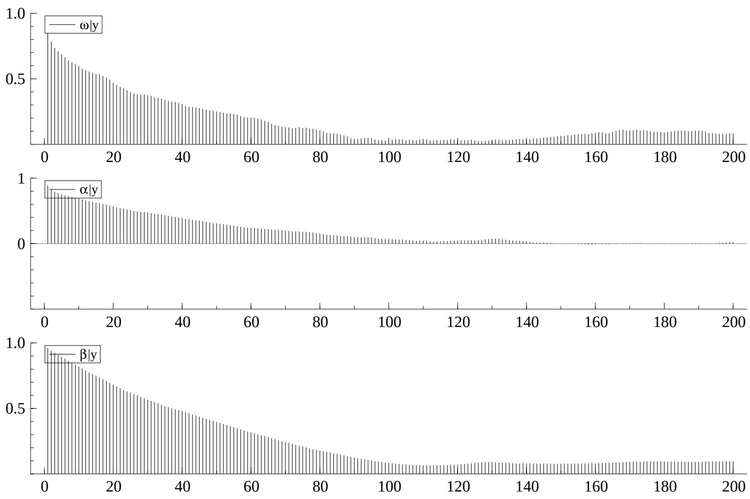

Figure 1. Correlograms of MH simulation for GARCH(1, 1).

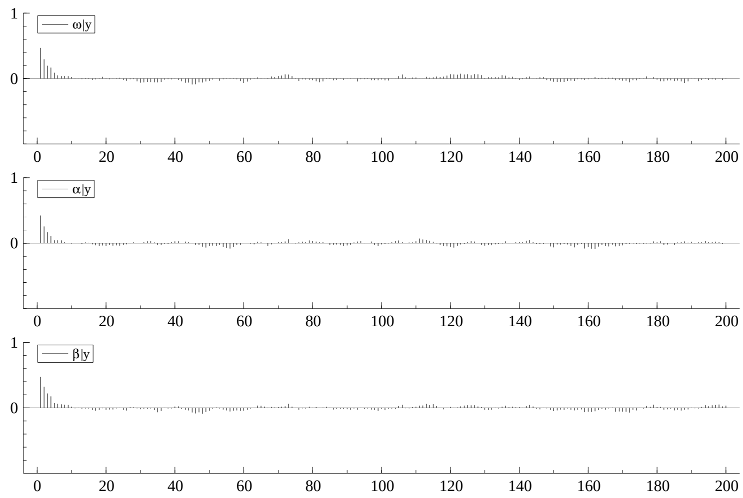

Figure 2. Correlograms of ARMH simulation for GARCH(1, 1).

shown in Table 1(b), and indicates that the ARMH is the most efficient and the MH is the worst, in the sense that the method needs short Markov chains to derive precise results. Although the GGS is less efficient than the ARMS, the former only needs $M = 1000$ to guarantee the convergence of Markov chains. This is an effect of conditioning which reduces variability of Markov chains. The far right column shows CPU time on a Pentium(R) 4 (with Ox 3.30), including discarded and rejected draws. The ARMH method (15 seconds) overwhelms others, and the GGS needs shorter time than MH and ARMS. The reason for the short time requirement of the GGS is that the dimension of parameter vector

Table 1. MCMC estimates of GARCH(1, 1) for simulated data ($T = 500$).

(a) Speeds of convergence

Method	N	M	Ave. INEF	Time
GGS	2000	1000	19.9	0:38.70
MH	2000	4000	73.3	7:44.76
ARMS	2000	3000	33.9	6:06.17
ARMH	1000	2000	2.0	0:14.81

Note. The MCMC simulation is conducted with $M + N$ iterations. The first N samples are discarded and next M draws are recorded. 'Ave. INEF' denotes a geometric average of ineffícacy factors (reported below) for each method. 'Time' denotes CPU time on a Pentium(R) 4 (with Ox 3.30), including discarded and rejected draws.

(b) Parameter estimates

Method	Parameter	Mean	95% Interval	CD	INEF
MLE	ω	0.087	$[-0.007, 0.180]$		
	α	0.075	$[0.022, 0.129]$		
	β	0.878	$[0.793, 0.963]$		
GGS	ω	0.104	$[0.014, 0.184]$	−0.37	18.9
	α	0.090	$[0.037, 0.166]$	0.28	18.4
	β	0.858	$[0.760, 0.933]$	0.16	22.5
MH	ω	0.104	$[0.032, 0.193]$	0.92	60.8
	α	0.085	$[0.038, 0.149]$	0.80	71.8
	β	0.861	$[0.770, 0.927]$	−0.88	90.1
ARMS	ω	0.120	$[0.040, 0.195]$	−0.96	34.5
	α	0.096	$[0.046, 0.170]$	−0.34	26.6
	β	0.843	$[0.754, 0.917]$	0.60	42.5
ARMH	ω	0.118	$[0.039, 0.193]$	−0.70	1.9
	α	0.091	$[0.046, 0.142]$	1.00	2.1
	β	0.847	$[0.771, 0.919]$	−0.62	2.0

Note. True parameter vector is $(0.10, 0.10, 0.85)$. MLE presents maximum likelihood estimates with asymptotic 95% intervals. For each method without MLE, M draws reported above are used for calculating the posterior means, 95% intervals, and the convergence diagnostic (CD) statistics and the inverse of the efficiency factors (inefficiency factor, INEF) proposed by Geweke (1992). The posterior means are computed by averaging the simulated draws. The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. The CD is computed using equation (2.1), where we set $n_A = 0.1M$ and $n_B = 0.5M$ and compute $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ using a Parzen window with bandwidths of $0.01M$ and $0.05M$, respectively.

is only 3. Table 1(b) shows detailed results, indicating that the four methods produce similar results except for inefficiency factors.

Posterior means and 95% intervals are different from the ML results, since the posterior densities are skewed and the ML evaluation of the standard error does not take the restriction into account.

Figure 3 shows the posterior densities for the GGS and ARMH. Even though both GGS and ARMH chains need to set larger M to draw more smoothed

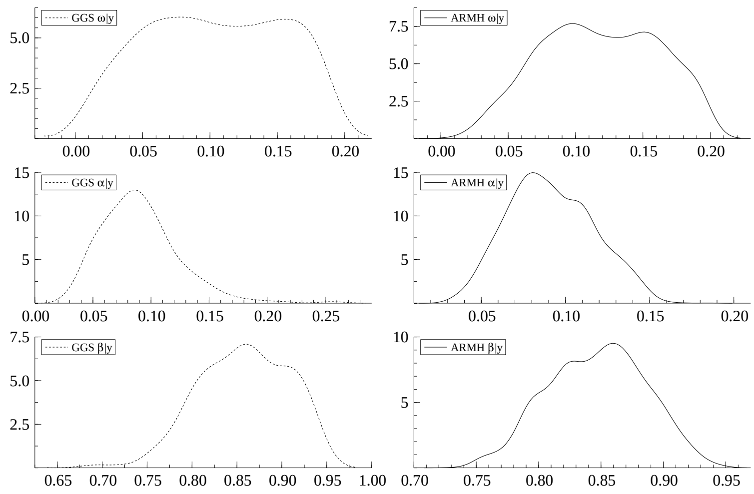

Figure 3. Posterior densities for GARCH(1, 1).

graphs, the histograms of the GGS seem to be preferable to the ARMH in the sense that the latter does not explore enough the tails of the distribution. In addition, the number of M of the GGS is half that of the ARMH. It should be noted that the computational requirement of the ARMH is still smaller than that of the GGS, even if $M = 10000$ is used for the former. One of the drawbacks of the GGS is that increasing the number of parameters leads to large increase in the computational burden, which will be shown in the next section.

3. Empirical example

We illustrate comparison of MCMC methods using daily data of the Tokyo stock price index (TOPIX) from January 4, 2000 to June 28, 2002 ($T = 615$). We define the return as $r_t = 100 \times \{\log P_t - \log P_{t-1}\}$ where P_t is the closing value on day t.

Table 2 shows estimation results for the GARCH(1, 1) model. Although their parameter estimates resemble each other, the ARMH is most efficient and least time-demanding of the four methods.

Table 3 presents estimation results for the GARCH-t(1, 1) model. As noted before, the MH methods of Nakatsuma (2000) is extended for the GARCH-t, by incorporating the MCMC technique of Watanabe (2001). The prior distribution for ν is specified as a truncated exponential with probability density function,

$$f(\nu) = \begin{cases} c\lambda \exp(-\lambda\nu), & \nu > 4, \\ 0, & \text{otherwise}, \end{cases}$$

where $c = \exp(4\lambda)$. Specifically, we set λ equal to 0.1. The extreme increase of computational requirements for the GGS and ARMS shown in Table 3(a) indicates that both methods need much time for an additional parameter. Table 3

Table 2. MCMC estimates of GARCH$(1,1)$ for TOPIX ($T = 614$).

(a) Speeds of convergence

Method	N	M	Ave. INEF	Time
GGS	2000	1000	17.8	0:47.35
MH	2000	5000	97.9	12:22.25
ARMS	2000	3000	11.3	7:55.43
ARMH	1000	2000	1.8	0:14.73

Note. The MCMC simulation is conducted with $M + N$ iterations. The first N samples are discarded and next M draws are recorded. 'Ave. INEF' denotes a geometric average of ineffícacy factors (reported below) for each method. 'Time' denotes CPU time on a Pentium(R) 4 (with Ox 3.30), including discarded and rejected draws.

(b) Parameter estimates

Method	Parameter	Mean	95% Interval	CD	INEF
MLE	ω	0.116	$[0.016, 0.215]$		
	α	0.078	$[0.031, 0.125]$		
	β	0.867	$[0.800, 0.934]$		
GGS	ω	0.111	$[0.019, 0.187]$	0.54	16.4
	α	0.083	$[0.034, 0.149]$	1.66	16.6
	β	0.868	$[0.791, 0.931]$	-1.18	20.8
MH	ω	0.138	$[0.066, 0.196]$	0.46	74.9
	α	0.082	$[0.041, 0.135]$	0.41	101.9
	β	0.852	$[0.786, 0.914]$	-0.35	123.1
ARMS	ω	0.127	$[0.053, 0.194]$	1.68	14.4
	α	0.091	$[0.046, 0.147]$	1.15	7.1
	β	0.852	$[0.794, 0.909]$	-1.62	14.1
ARMH	ω	0.130	$[0.055, 0.194]$	-0.13	1.5
	α	0.087	$[0.045, 0.132]$	0.16	1.6
	β	0.854	$[0.798, 0.905]$	-0.15	2.4

Note. MLE presents maximum likelihood estimates with asymptotic 95% intervals. For each method without MLE, M draws reported above are used for calculating the posterior means, 95% intervals, and the convergence diagnostic (CD) statistics and the inverse of the efficiency factors (inefficiency factor, INEF) proposed by Geweke (1992). The posterior means are computed by averaging the simulated draws. The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. The CD is computed using equation (2.1), where we set $n_A = 0.1M$ and $n_B = 0.5M$ and compute $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ using a Parzen window with bandwidths of $0.01M$ and $0.05M$, respectively.

also shows that the ARMH is the most efficient and the least time-demanding of the four methods.

Table 4 reports estimation results for the GJR-$t(1,1)$ model. The MH method of Nakatsuma (2000) does not apply to this model. Table 4(a) shows that differences of averaged inefficiencies of three methods are smaller than those for GARCH and GARCH-t, although the ARMH is the most efficient. Table 4(a) also reports that the effects of an additional parameter yields large increases in

Table 3. MCMC estimates of GARCH-$t(1,1)$ for TOPIX ($T = 614$).

(a) Speeds of convergence

Method	N	M	Ave. INEF	Time
GGS	2000	1000	11.2	2:48.62
MH	2000	6000	80.7	16:09.62
ARMS	2000	3000	7.2	13:31.64
ARMH	1000	2000	2.0	0:31.26

Note. The MCMC simulation is conducted with $M + N$ iterations. The first N samples are discarded and next M draws are recorded. 'Ave. INEF' denotes a geometric average of inefficacy factors (reported below) for each method. 'Time' denotes CPU time on a Pentium(R) 4 (with Ox 3.30), including discarded and rejected draws.

(b) Parameter estimates

Method	Parameter	Mean	95% Interval	CD	INEF
MLE	ω	0.127	$[0.008, 0.245]$		
	α	0.068	$[0.019, 0.117]$		
	β	0.870	$[0.790, 0.950]$		
	ν	10.20	$[3.424, 16.98]$		
GGS	ω	0.126	$[0.038, 0.188]$	−1.95	19.9
	α	0.081	$[0.029, 0.148]$	−1.73	18.8
	β	0.862	$[0.788, 0.932]$	1.92	24.6
	ν	11.12	$[6.742, 13.19]$	−0.74	1.7
MH	ω	0.126	$[0.047, 0.195]$	−0.42	69.9
	α	0.068	$[0.029, 0.117]$	0.42	55.4
	β	0.870	$[0.809, 0.934]$	0.05	97.2
	ν	11.83	$[5.971, 23.41]$	−1.11	112.6
ARMS	ω	0.134	$[0.060, 0.196]$	1.11	12.5
	α	0.080	$[0.035, 0.142]$	1.21	9.2
	β	0.858	$[0.795, 0.915]$	−1.67	14.2
	ν	11.36	$[6.020, 21.48]$	1.62	1.6
ARMH	ω	0.136	$[0.056, 0.196]$	−1.50	1.9
	α	0.079	$[0.033, 0.129]$	1.80	1.7
	β	0.856	$[0.785, 0.914]$	0.03	1.9
	ν	10.43	$[6.068, 15.87]$	0.43	2.5

Note. MLE presents maximum likelihood estimates with asymptotic 95% intervals. For each method without MLE, M draws reported above are used for calculating the posterior means, 95% intervals, and the convergence diagnostic (CD) statistics and the inverse of the efficiency factors (inefficiency factor, INEF) proposed by Geweke (1992). The posterior means are computed by averaging the simulated draws. The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. The CD is computed using equation (2.1), where we set $n_A = 0.1M$ and $n_B = 0.5M$ and compute $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ using a Parzen window with bandwidths of $0.01M$ and $0.05M$, respectively.

Table 4. MCMC estimates of GJR-GARCH-$t(1,1)$ for TOPIX ($T = 614$).

(a) Speeds of convergence

Method	N	M	Ave. INEF	Time
GGS	2000	1000	5.0	3:27.73
ARMS	2500	3000	7.1	19:43.28
ARMH	1000	2000	3.6	0:34.71

Note. The MCMC simulation is conducted with $M + N$ iterations. The first N samples are discarded and next M draws are recorded. 'Ave. INEF' denotes a geometric average of inefficacy factors (reported below) for each method. 'Time' denotes CPU time on a Pentium(R) 4 (with Ox 3.30), including discarded and rejected draws.

(b) Parameter estimates

Method	Parameter	Mean	95% Interval	CD	INEF
MLE	ω	0.097	$[-0.001, 0.195]$		
	α^+	0.024	$[-0.016, 0.065]$		
	α^-	0.123	$[0.046, 0.200]$		
	β	0.881	$[0.810, 0.951]$		
	ν	10.91	$[3.143, 18.67]$		
GGS	ω	0.114	$[0.031, 0.185]$	0.94	7.9
	α^+	0.038	$[0.011, 0.080]$	-1.01	5.4
	α^-	0.145	$[0.072, 0.235]$	-1.58	7.6
	β	0.861	$[0.796, 0.930]$	0.47	9.4
	ν	11.69	$[7.135, 13.98]$	-1.18	1.0
ARMS	ω	0.127	$[0.048, 0.195]$	-0.02	10.6
	α^+	0.040	$[0.004, 0.098]$	0.01	5.4
	α^-	0.150	$[0.070, 0.247]$	-1.43	12.6
	β	0.850	$[0.786, 0.912]$	0.46	18.0
	ν	11.84	$[6.103, 23.11]$	0.10	1.4
ARMH	ω	0.123	$[0.044, 0.192]$	1.14	2.0
	α^+	0.034	$[0.004, 0.069]$	0.29	4.2
	α^-	0.142	$[0.073, 0.236]$	-0.25	4.8
	β	0.857	$[0.807, 0.916]$	-0.30	2.9
	ν	11.32	$[6.206, 18.79]$	0.97	5.7

Note. MLE presents maximum likelihood estimates with asymptotic 95% intervals. For each method without MLE, M draws reported above are used for calculating the posterior means, 95% intervals, and the convergence diagnostic (CD) statistics and the inverse of the efficiency factors (inefficiency factor, INEF) proposed by Geweke (1992). The posterior means are computed by averaging the simulated draws. The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. The CD is computed using equation (2.1), where we set $n_A = 0.1M$ and $n_B = 0.5M$ and compute $\hat{\sigma}^2_A$ and $\hat{\sigma}^2_B$ using a Parzen window with bandwidths of $0.01M$ and $0.05M$, respectively.

the computational burden of the GGS and ARMS.

The time requirement of the ARMS can be reduced, for instance, by using the following strategy. By using ARMS method to draw first, say, 400 samples (discarding the first 100 samples), posterior mean and covariance are calculated.

The mean and covariance is used to form a multivariate Student-t proposal distribution for the MH algorithm. This strategy, however, will blur the feature of the ARMS method.

4. Concluding remarks

In this paper, the MCMC estimation procedures of the class of ARCH models have been reviewed. For each method, we describe the main advantages and drawbacks. As for the flexibility to model extensions, the GGS, ARMS and ARMH give easy black-box samplings. With respect to the mixing, efficiency and computational requirement of the MCMC, this paper found the ARMH is the best method.

Acknowledgements

This research is partially supported by the Grant-in-Aid for the 21st Century COE program "Microstructure and Mechanism Design in Financial Markets" from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Asai, M. and Watanabe, T. (2004). Comparison of MCMC methods for estimating GARCH models, COE discussion paper series, No. 18, Tokyo Metropolitan University.

Bauwens, L. and Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler, *Econometrics Journal*, **1**, c23–c46.

Bauwens, L., Lubrano, M. and Richard, J.-F. (1999). *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press.

Black, F. (1976). Studies of stock market volatility changes, *1976 Proceedings of the American Statistical Association Bisiness and Economic Statistics Section*, American Statistical Association, 177–181.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.

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

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

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

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hasting algorithm, *The American Statistician*, **49**, 327–335.

Christie, A. A. (1982). The stochastic behavior of common stock variances: value, leverage and interest rate effects, *Journal of Financial Economics*, **10**, 407–432.

Doornik, J. A. (1998). *Object-Oriented Matrix Programming using Ox 2.0*, London, Timberlake Consultants Press.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Ecnometrica*, **50**, 987–1008.

Gelfand, A. E., Smith, A. F. M. and Lee, T. M. (1992). Bayesian analysis of constrained parameter and truncated data problem using Gibbs sampling, *Journal of the American Statistical Association*, **87**, 523–532.

Geweke, J. (1989). Exact predictive density for linear models with ARCH disturbances, *Journal of Econometrics*, **40**, 63–86.

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

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

Geweke, J. (2001). Getting it right: Checking for errors in likelihood based inference, unpublished manuscript, University of Iowa.

Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling, *Applied Statistics*, **44**, 455–473.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1992). On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, **48**, 1779–1801.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**, 711–732.

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

Mitsui, H. and Watanabe, T. (2003). Bayesian analysis of GARCH option pricing models, *the Journal of the Japan Statistical Society* (Japanese Issue), **33**, 307–324 (in Japanese).

Müller, P. and Pole, A. (1995). Monte Carlo posterior integration in GARCH models, unpublished manuscript, Duke University.

Nakatsuma, T. (2000). Bayesian analysis of ARMA-GARCH models: A Markov chain sampling approach, *Journal of Econometrics*, **95**, 57–69.

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

Ritter, C. and Tanner, M. A. (1992). The Gibbs stopper and the Griddy Gibbs sampler, *Journal of the American Statistical Association*, **87**, 861–868.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*, Springer-Verlag.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion), *Annals of Statistics*, **21**, 1701–1762.

Vrontos, I. D., Dellaportas, P. and Politis, D. (2000). Full Bayesian inference for GARCH and EGARCH models, *Journal of Business and Economic Statistics*, **18**, 187–198.

Watanabe, T. (2001). On sampling the degree-of-freedom of student-t disturbances, *Statistics and Probability Letters*, **52**, 177–181.