Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 22 (2013) 1056 – 1064

Procedia Computer Science

17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013

Markov Chain Monte Carlo versus Importance Sampling in Bayesian Inference of the GARCH model

Tetsuya Takaishi[a,*]

[a]Hiroshima University of Economics, 5-37-1, Gion, Asaminami-ku, Hiroshima 731-0192, Japan

Abstract

Usually, the Bayesian inference of the GARCH model is preferably performed by the Markov Chain Monte Carlo (MCMC) method. In this study, we also take an alternative approach to the Bayesian inference by the importance sampling. Using a multivariate Student's t-distribution that approximates the posterior density of the Bayesian inference, we compare the performance of the MCMC and importance sampling methods. The overall performance can be measured in terms of statistical errors obtained for the same size of Monte Carlo data. The Bayesian inference of the GARCH model is performed by the MCMC method implemented by the Metropolis-Hastings algorithm and the importance sampling method for artificial return data and stock return data. We find that the statistical errors of the GARCH parameters from the importance sampling are smaller than or comparable to those obtained from the MCMC method. Therefore we conclude that the importance sampling method can also be applied effectively for the Bayesian inference of the GARCH model as an alternative method to the MCMC method.

Keywords: GARCH model; Markov Chain Monte Carlo; Importance Sampling; Bayesian Inference

1. Introduction

In empirical finance it is of particular importance to estimate and forecast volatility of asset returns which is often utilized for option pricing and portfolio allocation etc. Since volatility is not directly observed in the financial markets one needs to use parametric models which could reproduce properties of asset returns. Empirical properties of asset returns are now classified as the stylized facts[1] which include fat-tailed distributions and volatility clustering etc. The most widely recognized models which capture some of the stylized facts are the GARCH (generalized autoregressive conditional hetroskedasticity) model[2] and its extended versions. The original model to the GARCH model was first created by Engle[3], which is called the ARCH model. Then later the ARCH model is generalized by Bollerslev[2] and the generalized version is called the GARCH model.

To utilize the GARCH model one needs to infer the model parameters so that the model matches the time series of asset returns we consider. In this study we employ the Bayesian inference for the GARCH model. In the Bayesian estimation one assumes that the probability density of the model parameters, called the posterior density, is given through the Bayes's rule. Then the model parameters are inferred as the expectation values under

*Corresponding author. Tel.: +81-82-871-3900.
E-mail address: tt-taka@hue.ac.jp.

doi:10.1016/j.procs.2013.09.191

the probability density of the model parameters. Namely the expectation values are given by the integrals of the parameter function constructed by a multiplication of the posterior density and a parameter variable. In general the integrals are not calculable analytically. The popular approach to perform the integrals is the Markov Chain Monte Carlo (MCMC) method which replace the expectation values of the parameters with the average values over Monte Carlo samples obtained through the Markov Chain. A drawback of the MCMC method is that the time series of the Monte Carlo samples obtained through the Markov Chain are usually correlated. The magnitude of the correlation can be measured by the autocorrelation time and an overall statistical error increases with the autocorrelation time. Thus it is preferable to use a method that can effectively generate uncorrelated Monte Carlo samples.

The simplest MCMC method might be the Metropolis algorithm[4] which creates the Markov Chain through the Metropolis accept/reject test. The Metropolis algorithm is algorithmically so simple that it can be applied for many cases including the inference of the GARCH model. However the Metropolis algorithm is generally not effective, i.e. its autocorrelation time is large.

There exist a variety of algorithms which might implement the MCMC method effectively. For the Bayesian inference of the GARCH model several algorithms have been proposed and applied[5]-[11]. By recent studies[9, 10, 13] it is shown that the Metropolis-Hastings algorithm[12] with a multivariate Student's t-proposal density generates MCMC time series very effectively, i.e. the autocorrelation time of the Monte Carlo data turns out to be small[9, 10, 13, 14, 15, 16, 17]. This effectiveness of the Metropolis-Hastings algorithm implies that the multivariate Student's t-proposal density used in the algorithm is already very similar to the posterior density of the GARCH model.

This observation may suggest an alternative approach to implement the Bayesian inference, e.g. the importance sampling. The importance sampling method introduces an importance sampling density which should be handled easily and can generate Monte Carlo data randomly. The Monte Carlo data generated randomly by the importance sampling method can be autocorrelation-free. At first sight the autocorrelation-free nature of the importance sampling could be considered to be an advantage over the MCMC method. However in order to obtain the correct results the Monte Carlo data generated by the importance sampling method should be recalculated by the reweighting factor which corrects the difference between the posterior density and the sampling density. Even if the Monte Carlo data from the importance sampling is autocorrelation-free the statistical errors of the Monte Carlo data could be enhanced by the introduction of such a reweighting factor. In this study we compare performance of the MCMC and importance methods for the GARCH model by the statistical errors estimated from the same size of Monte Carlo data.

2. GARCH model

The general form of the GARCH model by Bollerslev[2], called the GARCH(p,q) model is given by

$$y_t = \sigma_t \epsilon_t, \tag{1}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2, \tag{2}$$

where the GARCH parameters are restricted to $\omega > 0$, $\alpha_i > 0$ and $\beta_i > 0$ to ensure a positive volatility, and the stationary condition $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$ is also required. y_t stands for the return time series of the GARCH(p,q) model and ϵ_t is an independent normal error $\sim N(0,1)$.

The parameters of q and p are lag parameters. Since empirical studies often show that small numbers are selected for q and p by the information criterion such as AIC[18], in this study we focus on the GARCH(1,1) model where the volatility σ_t^2 is given by

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{3}$$

Hereafter the GARCH model represents the GARCH(1,1) model.

3. Bayesian Inference of GARCH model

From Bayes' rule the posterior density $\pi(\theta|y)$ with n observations denoted by $y = (y_1, y_2, \dots, y_n)$ is given by

$$\pi(\theta|y) \propto L(y|\theta)\pi(\theta), \tag{4}$$

where $L(y|\theta)$ is the likelihood function and $\theta \equiv (\theta_1, \theta_2, \theta_3) = (\alpha, \beta, \omega)$ for the GARCH model. $\pi(\theta)$ stands for the prior density which we have to specify depending on θ. In this study we assume that the prior density $\pi(\theta)$ is constant. The likelihood function of the GARCH model with n observations is given by

$$L(y|\theta) = \Pi_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{y_t^2}{\sigma_t^2}\right). \tag{5}$$

Using the posterior density $\pi(\theta|y)$ as a probability density of θ we infer values of θ as expectation values of θ. The expectation values are given by

$$E[\theta_i] = \frac{1}{Z}\int \theta_i \pi(\theta|y) d\theta, \tag{6}$$

where $Z = \int \pi(\theta|y)d\theta$ is the normalization constant. Since the posterior density $\pi(\theta|y)$ is not always normalized, Z is not unity in general. However the value of Z is irrelevant to the MCMC method.

4. Markov Chain Monte Carlo

Since in general eq.(6) is not analytically calculable, we estimate it numerically by the MCMC method. Suppose we obtain N data sampled by an MCMC method. Then we evaluate the expectation value as an average value over the sampled data $\theta_i^{(j)}$,

$$\langle\theta_i\rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \theta_i^{(j)}. \tag{7}$$

The statistical error for N independent (uncorrelated) data is proportional to $\frac{1}{\sqrt{N}}$. When the sampled data are correlated the statistical error increases in proportion to $\sqrt{\frac{2\tau}{N}}$ where τ is the autocorrelation time between the sampled data. In order to have smaller statistical errors it is desirable to take an MCMC method which can generate data with a small τ.

The autocorrelation time is calculated through the autocorrelation function (ACF). The ACF of θ_i with N Monte Carlo data is defined as

$$ACF(t) = \frac{\frac{1}{N}\sum_{j=1}^{N}(\theta_i^{(j)} - \langle\theta_i\rangle)(\theta_i^{(j+t)} - \langle\theta_i\rangle)}{\sigma_{\theta_i}^2}, \tag{8}$$

where $\langle\theta_i\rangle$ and $\sigma_{\theta_i}^2$ are the average value and the variance of θ_i respectively. The autocorrelation time τ is defined by

$$\tau = \frac{1}{2} + \sum_{t=1}^{\infty} ACF(t). \tag{9}$$

There exist a variety of methods to implement MCMC algorithm. In this study for the Bayesian inference of the GARCH model we employ the Metropolis-Hastings algorithm with a multivariate Student's t-proposal density which is shown to be effective[9, 10, 13, 14], (i.e. the autocorrelation time is small).

5. Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm[12] is an extension of the original Metropolis algorithm[4]. Let $P(x)$ be a probability distribution from which we would like to sample data x. The MH algorithm generates x sequentially as follows.

(1) First we set an initial value x_0 and $i = 1$.

(2) Then we generate a new value x_i from a certain probability distribution $g(x_i|x_{i-1})$ which we call proposal density.

(3) We accept the candidate x_i with a probability of $P_{MH}(x_{i-1}, x_i)$ where

$$P_{MH}(x_{i-1}, x_i) = \min\left[1, \frac{P(x_i)}{P(x_{i-1})}\frac{g(x_i|x_{i-1})}{g(x_{i-1}|x_i)}\right]. \tag{10}$$

When x_i is rejected we keep x_{i-1}, i.e. $x_i = x_{i-1}$.

(4) Go back to (2) with an increment of $i = i + 1$.

From the studies of Refs.[9, 10, 13, 14], it is shown that a multivariate Student's t-distribution as the proposal density is very effective for the Bayesian inference of the GARCH model. In this study we also use a multivariate Student's t-distribution for the MH algorithm. Let $g(\theta)$ be a (p-dimensional) multivariate Student's t-distribution given by

$$g(\theta) = \frac{\Gamma((\nu+p)/2)/\Gamma(\nu/2)}{\det\Sigma^{1/2}(\nu\pi)^{p/2}}\left[1 + \frac{(\theta-M)^t\Sigma^{-1}(\theta-M)}{\nu}\right]^{-(\nu+p)/2}, \tag{11}$$

where θ and M are column vectors,

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix}, M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_p \end{bmatrix}, \tag{12}$$

and $M_i = E(\theta_i)$. Σ is the covariance matrix defined as

$$\frac{\nu\Sigma}{\nu-2} = E[(\theta-M)(\theta-M)^t]. \tag{13}$$

ν is a parameter to tune the shape of Student's t-distribution. For $\nu \to \infty$ the Student's t-distribution goes to a Gaussian distribution. In this study we use $\nu = 10$[13].

The random number generation for the multivariate Student's t-distribution can be done easily as follows. First we decompose the symmetric covariance matrix Σ by the Cholesky decomposition as $\Sigma = LL^t$. Then substituting this result to eq.(11) we obtain

$$g(X) \sim \left[1 + \frac{X^tX}{\nu}\right]^{-(\nu+p)/2}, \tag{14}$$

where $X = L^{-1}(\theta - M)$. The random numbers X are given by $X = Y\sqrt{\frac{\nu}{w}}$, where Y follows $N(0, I)$ and w is taken from the chi-square distribution with ν degrees of freedom, χ^2_ν. Finally we obtain the random number θ by $\theta = LX + M$.

6. Importance Sampling

Let us recall eq.(6),

$$\langle\theta_i\rangle = \frac{1}{Z}\int \theta_i\pi(\theta|y)d\theta. \tag{15}$$

When $\pi(\theta|y)$ is properly normalized, then $Z = 1$. However in general $\pi(\theta|y)$ of the GARCH model can not be easily normalized and in such a case $Z \neq 1$. By introducing an importance sampling density $h(\theta)$ eq.(15) can be rewritten as

$$\langle \theta_i \rangle = \frac{1}{Z'}\frac{Z'}{Z}\int \theta_i \frac{\pi(\theta|y)}{h(\theta)} h(\theta) d\theta, \tag{16}$$

where

$$Z' = \int h(\theta) d\theta. \tag{17}$$

Eq.(16) can be estimated by Monte Carlo integration as

$$\langle \theta_i \rangle = \frac{\langle \theta_i r(\theta) \rangle_{h(\theta)}}{\langle r(\theta) \rangle_{h(\theta)}}, \tag{18}$$

where the reweighting factor $r(\theta)$ is given by

$$r(\theta) = \frac{\pi(\theta|y)}{h(\theta)}, \tag{19}$$

and $\langle O(\theta) \rangle_{h(\theta)}$ stands for the expectation value of $O(\theta)$ with respect to $h(\theta)$. Eq.(18) can be obtained as average values over Monte Carlo data of θ generated by the importance sampling density $h(\theta)$. The importance sampling density $h(\theta)$ should be simple enough to easily generate θ at random. Here we use a multivariate Student's t-distribution as $h(\theta)$ which is also used in the MH algorithm. As seen in the previous section one can easily generate θ randomly from the multivariate Student's t-distribution. Since the Monte Carlo data of θ can be generated randomly they appear to be autocorrelation-free. However what we need is eq.(18) which includes the reweighting factor. The introduction of the reweighting factor may enhance the statistical errors. In order to estimate the statistical errors of eq.(18) we use the jackknife method.

Note that both MH and importance sampling methods use a multivariate Student's t-distribution. The essential difference between two methods comes from how to correct the difference between the posterior density and the multivariate Student's t-distribution. Namely the MH method corrects the difference by the Metropolis test. On the other hand the importance sampling method corrects it by the reweighting factor. In principle two method could give the same final results. However the performance of them could appear differently. We measure the performance of them in terms of the statistical errors and make a comparison of performance between two methods.

7. Simulation Study

In this section we compare the MCMC and importance sampling methods using artificially generated GARCH data. We set the GARCH parameters to $\alpha = 0.05$, $\beta = 0.9$ and $\omega = 0.1$, and generated 3000 data. Fig.1 shows the time series of the data.

We implement the MCMC method as in Ref.[13]. First we make a pilot run by the Metropolis algorithm to estimate M and Σ of a multivariate Student's t-distribution. The first 5000 Monte Carlo data by the Metropolis algorithm are discarded as burn-in process. We then switch from the Metropolis algorithm to the MH algorithm. During simulations we recalculate the values of M and Σ every 1000 Monte Carlo updates by using the accumulated Monte Carlo data so far. Fig.2 shows the acceptance at the Metropolis test of the MH algorithm as a function of every 1000 updates. At the beginning of the simulation the acceptance is low, which indicates that the values of M and Σ are still not accurate enough. As we proceed simulations the acceptance increases rapidly and reaches a plateau around 75%-80%. Final results are obtained by 200000 Monte Carlo data. The autocorrelation times are calculated to be very small, $2\tau \approx 2$, which means that the MH algorithm is a very efficient MCMC method for the GARCH model[13].

The importance sampling algorithm also uses a multivariate Student's t-distribution. We determine the values of M and Σ for the multivariate Student's t-distribution by a pilot run of the MH algorithm. Here we accumulate 5000 Monte Carlo data by the MH algorithm and calculate M and Σ of the multivariate Student's t-distribution.

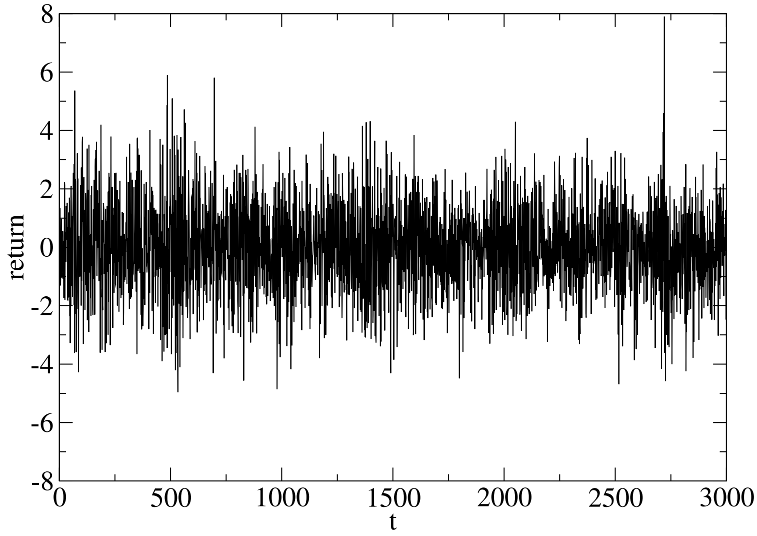

Fig. 1. Return time series generated by the GARCH model with $\alpha = 0.05$, $\beta = 0.9$ and $\omega = 0.1$.

Then using the multivariate Student's t-distribution as the importance sampling density we proceed the importance sampling simulation and generate 200000 Monte Carlo data of θ. Since those Monte Carlo data are generated randomly in the importance sampling the correlation between those Monte Carlo data should be nonexistent. We checked the autocorrelation time of those Monte Carlo data and found that $2\tau \approx 1$, which means there exists no correlation between those Monte Carlo data.

The GARCH parameters obtained by the the MCMC and importance sampling methods are summarized in Table 1. It is confirmed that both methods correctly reproduce the values of θ used for the generation of the GARCH data. The standard deviations are defined by $\sqrt{\langle(\theta_i - \langle\theta_i\rangle)^2\rangle}$. Since we employ the same posterior density for both methods the standard deviations should be the same, as confirmed in Table 1.

The statistical errors are estimated by the jackknife method. It is found that the statistical errors from the importance sampling method are smaller than those of the MCMC method. The gain g of the importance sampling method is defined by "statistical error(MCMC)/statistical error(Importance sampling)". We find that $g > 1$ which means that the greater performance is obtained for the importance sampling method.

Table 1. Results from artificial GARCH data.

		α	β	ω
	true value	0.05	0.9	0.1
MCMC	$\langle\theta_i\rangle$	0.0510	0.913	0.082
	standard deviation	0.0095	0.018	0.026
	statistical error	0.000033	0.000080	0.00013
	2τ	2.1 ± 0.1	2.3 ± 0.2	2.3 ± 0.2
Importance sampling	$\langle\theta_i\rangle$	0.0510	0.912	0.082
	standard deviation	0.0095	0.018	0.026
	statistical error	0.000025	0.000040	0.000042
gain	g	1.32	2.0	3.10

8. Empirical Application

In this section we perform the Bayesian inference for stock data traded on the Tokyo Stock Exchange. The stock data we use is daily stock price data of Toshiba Co. from June 3, 2006 to December 30, 2009. The price

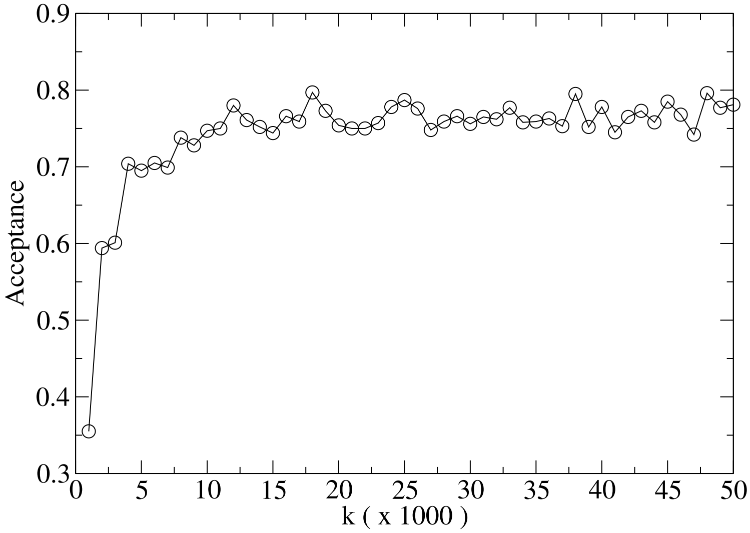

Fig. 2. Acceptance of the Metropolis-Hastings algorithm for artificial GARCH data as a function of every 1000 updates.

data p_i are transformed to return data $r_i = \ln(p_i/p_{i-1})$. The daily return time series is shown in fig.3.

Our implementation scheme is the same as in the previous section. Using the daily return data we perform the Bayesian inference of the GARCH model by the MCMC and importance sampling methods. The final results are calculated from 200000 Monte Carlo data. Fig.4 shows the acceptance of the MH algorithm which is similar to that obtained in fig.2. Table 2 summarizes the results from both methods. It is found that both methods give almost the same values for θ and standard deviations. The gain turns out to be $g \approx 1$, which indicates that the performance of the importance sampling method is comparable to that of the MCMC method.

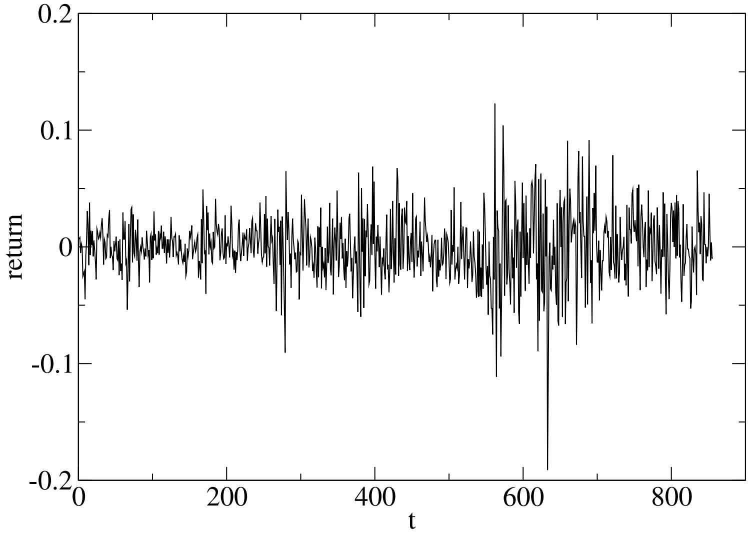

Fig. 3. Daily stock return time series of Toshiba Co on the Tokyo Stock Exchange.

9. Conclusions

We have performed the Bayesian inference of the GARCH model by two methods: MCMC and importance sampling, and compared performance of two methods. Both methods utilize a multivariate Student's distribution.

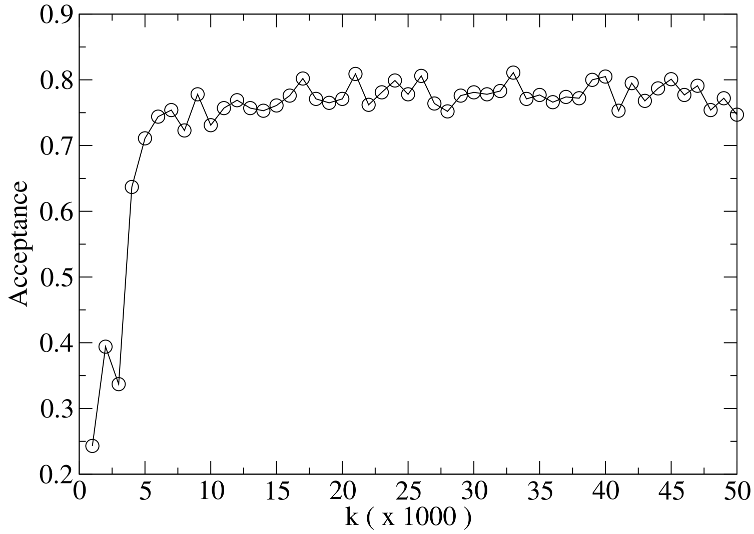

Fig. 4. Acceptance of the Metropolis-Hastings algorithm for daily return data as a function of every 1000 updates.

Table 2. Results from daily stock returns of Toshiba Co.

		α	β	ω
MCMC	$\langle\theta_i\rangle$	0.101	0.896	7.7×10^{-6}
	standard deviation	0.019	0.019	4.1×10^{-6}
	statistical error	0.000075	0.000079	1.5×10^{-8}
	2τ	2.1 ± 0.1	2.2 ± 0.1	2.5 ± 0.2
Importance sampling	$\langle\theta_i\rangle$	0.101	0.894	7.6×10^{-6}
	standard deviation	0.019	0.019	4.0×10^{-6}
	statistical error	0.000066	0.000084	2.0×10^{-8}
gain	g	1.2	0.94	0.75

From the results obtained for the artificial GARCH data we found that the importance sampling outperforms the MCMC method implemented by the MH algorithm. We also performed the Bayesian inference for the daily stock return data of Toshiba Co. and found that the performance of importance sampling is comparable to that of the MCMC method. Our results might indicate that the performance of two methods depends on the return data we consider. Since there exists a case that the performance of importance sampling is better than that of the MCMC method, it is concluded that the importance sampling method can be used as an alternative method to the MCMC method for the Bayesian inference of the GARCH model.

Acknowledgements

Numerical calculations in this work were carried out at the Yukawa Institute Computer Facility and the facilities of the Institute of Statistical Mathematics. This work was supported by Grant-in-Aid for Scientific Research (C) (No.22500267).

References

[1] Cont R. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance* 2001; **1**: 223–236.
[2] Bollerslev T. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 1986; **31** : 307-327.
[3] Engle RF. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom inflation. *Econometrica* 1982; **50**: 987-1007.

[4] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equations of State Calculations by Fast Computing Machines. *J. of Chem. Phys.* 1953;**21**: 1087–1091.

[5] Bauwens L, Lubrano M. Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal* 1998;**1**: c23-c46.

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

[7] Nakatsuma T. A Markov-Chain Sampling Algorithm for GARCH Models. *Studies in Nonlinear Dynamics and Econometrics* 1998; **3(2)**: 107–117.

[8] Nakatsuma T. Bayesian analysis of ARMA-GARCH models: Markov chain sampling approach. *Journal of Econometrics* 2000; **95**: 57–69.

[9] Mitsui H, Watanabe T. Bayesian analysis of GARCH option pricing models. *J. Japan Statist. Soc. (Japanese Issue)* 2003;**33**: 307–324.

[10] Asai M. Comparison of MCMC Methods for Estimating GARCH Models. *J. Japan Statist. Soc.* 2006; **36**: 199–212.

[11] Takaishi T. Bayesian Estimation of GARCH model by Hybrid Monte Carlo. *Proceedings of the 9th Joint Conference on Information Sciences* 2006; CIEF-214
doi:10.2991/jcis.2006.159

[12] Hastings WK. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* 1970;**57**: 97–109.

[13] Takaishi T. An Adaptive Markov Chain Monte Carlo Method for GARCH Model. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Complex Sciences* 2009; Vol. **5**: 1424.

[14] Takaishi T. Bayesian Estimation of GARCH Model with an Adaptive Proposal Density. *New Advances in Intelligent Decision Technologies, Studies in Computational Intelligence* 2009 ;Vol. **199**: 635–643.

[15] Takaishi T. Bayesian Inference on QGARCH Model Using the Adaptive Construction Scheme *Proceedings of 8th IEEE/ACIS International Conference on Computer and Information Science* 2009; 525. doi:10.1109/ICIS.2009.173

[16] Takaishi T. Bayesian inference with an adaptive proposal density for GARCH models. *J. Phys.: Conf. Ser.* 2010; **221**: 012011.

[17] Takaishi T, Chen TT. Bayesian Inference of the GARCH model with Rational Errors *International Proceedings of Economics Development and Research* 2012;**29**: 303–307

[18] Akaike H. Information theory and an extension of the maximum likelihood principle. *Proceedings of the 2nd International Symposium on Information Theory* Petrov B. N. and Caski F (eds.), Akadimiai Kiado, Budapest 1973: 267-281.