Report Series

Adaptive Information Systems and Modelling in Economics and Management Science

A Comparison of Bayesian Model Selection based on MCMC with an application to GARCH-Type Models

Tatiana Miazhynskaia

Sylvia Frühwirth-Schnatter

Georg Dorffner

Report No. 83

November, 2003

Report Series

November, 2003

SFB
'Adaptive Information Systems and Modelling in Economics and Management Science'

Vienna University of Economics
and Business Administration
Augasse 2–6, 1090 Wien, Austria

in cooperation with
University of Vienna
Vienna University of Technology

http://www.wu-wien.ac.at/am

Papers published in this report series
are preliminary versions of journal articles
and not for quotations.

This piece of research was supported by the Austrian Science Foundation (FWF) under grant SFB#010 ('Adaptive Information Systems and Modelling in Economics and Management Science').

A Comparison of Bayesian Model Selection based on MCMC with an application to GARCH-Type Models

Tatiana Miazhynskaia
Austrian Research Institute for
Artificial Intelligence,
Freyung 6/6, A-1010 Vienna, Austria,
phone:+431.5336112-25,
fax: +431.5336112-77,
tatiana@oefai.at

Sylvia Frühwirth-Schnatter*
Institute for Applied Statistics,
Johannes Kepler University Linz
Altenberger Strae 69, A-4040 Linz,
phone: +43732.2468-8295,
fax: +43732.2468-9846,
sylvia.fruehwirth-schnatter@jku.at

Georg Dorffner
Austrian Research Institute for Artificial Intelligence and
Department of Medical Cybernetics and Artificial Intelligence,
University of Vienna, Freyung 6/2, A-1010 Vienna, Austria,
phone: +431.4277-63116, fax: +431.4277-9631,
georg@ai.univie.ac.at

Abstract

This paper presents a comprehensive review and comparison of five computational methods for Bayesian model selection, based on MCMC simulations from posterior model parameter distributions. We apply these methods to a well-known and important class of models in financial time series analysis, namely GARCH and GARCH-t models for conditional return distributions (assuming normal and t-distributions). We compare their performance vis--vis the more common maximum likelihood-based model selection on both simulated and real market data. All five MCMC methods proved feasible in both cases, although differing in their computational demands. Results on simulated data show that for large degrees of freedom (where the t-distribution becomes more similar to a normal one), Bayesian model selection results in better decisions in favour of the true model than maximum likelihood. Results on market data show the feasibility of all model selection methods, mainly because the distributions appear to be decisively non-Gaussian.

*Corresponding author

Keywords: Bayesian inference, Bayesian model selection, GARCH models, Markov Chain Monte Carlo (MCMC), model likelihood

1 Introduction

The Bayesian inference became very popular in the last decade of the previous century. Scientists got both: methodology and computer power to perform realistic Bayesian data analysis. The difference between the classical and the Bayesian approach to statistics is a fundamental change of paradigm. In the classical analysis, it is assumed that we know the correct structure of the model, and want to estimate the ”true value” of the model parameters which are also assumed to exist. From a Bayesian point of view, there is no such thing as a true parameter value. All we have is a fixed collection of data. Furthermore, we may have some prior idea about what value of the parameters could be expected, summarized in a (possibly non- or little informative) prior density of the parameters. Combining the prior density with the likelihood of observing the data a posterior density of the parameters is constructed. This posterior density is a quantitative, probabilistic description of the knowledge about the parameters in the model. The strong advantage of a Bayesian analysis is that the complete posterior distribution of the parameters can be used for further analysis: to create a prediction or to make a model selection.

To explore the posterior distribution of parameters, the most powerful technique is to use Markov chain Monte Carlo (MCMC) computing methods such as the Gibbs sampler (Gelfand and Smith, 1990) and the Metropolis-Hastings (MH) algorithm (Hastings, 1970). While these algorithms enable direct estimation of posterior and predictive quantities of interest, they do not readily lend themselves to estimation aspects of the model probabilities. As a result, many different approaches have been suggested in the literature. The most widely used is the group of direct methods: harmonic mean estimator of Newton and Raftery (1994), importance sampling (Frühwirth-Schnatter, 1995), reciprocal importance estimator (Gelfand and Dey, 1994), bridge sampling ((Meng and Wong, 1996), (Frühwirth-Schnatter, 2002)). A nice review of some these methods together with background concepts of Bayesian model selection can be found in Kass and Raftery (1995). In 1995 Chib proposed also an indirect methods for estimating model likelihoods from Gibbs sampling output, an idea that recently has been extended to output from MH algorithm (Chib and Jeliazhov, 2001). A slightly more direct approach to compute posterior model probabilities using MCMC is to include the model indicator as a parameter in the sampling algorithm itself. Green (1995) introduced a reversible jump MCMC strategy for generating from the joint posterior distribution based on the standard MH approach.

The purpose of this article is to give the computationally complete review of these methods and to demonstrate the differences and difficulties of the model likelihood estimators discussed above by a simulation study and empirical data. For illustration, we focus on the most important practical models for financial time series - GARCH-type models, introduced for the first time in the original paper of Bollerslev (1986) and then extensively generalized in many aspects, including conditional variance equation and the density specification. We limited ourselves to the case of two models: AR(1)-GARCH(1,1) with gaussian errors and AR(1)-GARCH(1,1)-t with Student t-density.

Our motivation for choosing the GARCH-type models for the illustration is explained by the fact that these models represent a very wide class of heteroskedastic econometric models and, therefore, the Bayesian analysis presented in this paper can be generalized for the whole this class. The Bayesian inference on GARCH-type models has been first implemented using importance sampling - see Kleibergen and van Dijk (1993). More recent approaches include Griddy-Gibbs sampler by Bauwens and Lubrano (1998) and Metropolis-Hastings algorithm with some specific choice of the proposal distribution ((Geweke, 1995), (Kim, Shephard and Chib, 1998), (Müller and Pole, 1998) and (Nakatsuma, 2000)), the model selection based on the reversible jump MCMC in Vrontos, Dellaportas and Politis (2000) and on the bridge sampling algorithm (Frühwirth-Schnatter, 2002). Our approach differs from the above that in this paper we apply different methods for making Bayesian model choice and compare their performance with the results from the classical analysis. Moreover, we compare the MCMC posterior output with the maximum likelihood estimates of the chosen GARCH models.

The empirical analysis was based on return series of stock indices from different financial markets. We used return series of the Dow Jones Industrial Average (USA), FTSE 100 (Great Britain) and NIKKEI 225 (Japan) over a period of 10 years and performed the complete Bayesian inference of GARCH models on these data.

The paper is organized as follows. In Section 2 we review the Bayesian approach to model selection and discuss some computational algorithm. The GARCH-type models chosen for the numerical illustration are presented in Section 3. Full MCMC methodology including the detailed implementation guidelines of the computational aspects is given in Section 4. Section 5 is dedicated to the simulation study and the application to market data is included in Section 6. Finally, Section 7 concludes the paper.

2 Bayesian Model Selection

All the complex models may be viewed as the specification of a joint distribution of observables (data) which we denote by Y and unobservables (model parameters) which we denote by θ. The traditional approach to Bayesian model selection is concerned with the following situation. Suppose the observed data Y are generated by a model M_i, one of a set **M** of competing models. Each model specifies the distribution of Y, $f(Y|\theta_i, M_i)$ apart from an unknown parameter vector θ_i of dimension n_i. Under prior densities $\pi(\theta_i|M_i)$ the marginal distribution of Y are found by integrating out the parameters

$$p(Y \mid M_i) = \int f(Y \mid \theta_i, M_i)\, \pi(\theta_i \mid M_i)\, d\theta_i. \tag{2.1}$$

By analogy with data likelihood function, the quantity $p(Y|M_i)$ is called also *model likelihood*.

Suppose also that we have some preliminary knowledge about model probabilities $\pi(M_i)$. Interest lies in obtaining the posterior probabilities $p(M_i|Y)$ for every model M_i in consideration either to arrive at a single "best" model, or to determine the posterior distribution of some quantity of interest which is common to all models via model averaging. Due to possible differences among the priors $\pi(M_i)$ a choice between two models (say models M_1 and M_2) is often based not on the posterior odds but on the *Bayes factor*

$$\mathrm{B}_{12} = \frac{p(M_1 \mid Y) \,/\, p(M_2 \mid Y)}{\pi(M_1) \,/\, \pi(M_2)},$$

i.e., the ratio of posterior odds of the model M_1 to prior odds of M_1. The Bayes factor is often thought of as the weight of evidence in favour of model M_1 provided "by the data", though Lavine and Scherrish (1999) show that a more accurate interpretation is that B_{12} captures the change in the odds in favour of model M_1 as we move from prior to posterior. Jeffreys (1961) recommended a scale of evidence for interpreting Bayes factors. In Table 1 we consider a slight modification of Jeffreys' scale taken from Wasserman (1997).

Kass and Raftery (1995) provide a complete review of Bayes factors, including their interpretation, computation or approximation, robustness to the model-specific prior distributions and applications in a variety of scientific problems.

Bayes Factor	Interpretation
$B_{12} < 1/10$	Strong evidence for M_2
$1/10 < B_{12} < 1/3$	Moderate evidence for M_2
$1/3 < B_{12} < 1$	Weak evidence for M_2
$1 < B_{12} < 3$	Weak evidence for M_1
$3 < B_{12} < 10$	Moderate evidence for M_1
$B_{12} > 10$	Strong evidence for M_1

Table 1: Jeffreys' scale of evidence for Bayes Factors.

Applying Bayes' theorem, we get that

$$B_{12} = \frac{\left[\frac{p(Y|M_1)\pi(M_1)}{p(Y)}\right] \,/\, \left[\frac{p(Y|M_2)\pi(M_2)}{p(Y)}\right]}{\pi(M_1) \,/\, \pi(M_2)} = \frac{p(Y|M_1)}{p(Y|M_2)},$$

i.e., the Bayes factor is equal to the ratio of the model likelihoods for the two models. In such a way, the collection of model likelihoods $p(Y|M_i)$ is equivalent to the model probabilities themselves (since the prior probabilities $\pi(M_i)$ are known in advance) and hence could also be considered as the key quantities needed for Bayesian model choice.

The model likelihoods yield posterior probabilities of all the models as

$$p(M_i|Y) = \frac{p(Y|M_i) \cdot \pi(M_i)}{\sum_{k=1}^{n} p(Y|M_k)\pi(M_k)}$$

The integral (2.1) is analytically tractable in only certain restricted problems and sampling based methods must be used to obtain estimates of the model likelihoods.

2.1 Direct Methods

In what follows, we suppress the dependence on the model indicator M_i in our notations, since all the calculations below must be repeated for all models under considerations.

By $\hat{p}(Y)$ we denote the estimate of the model likelihood $p(Y)$. Directly from (2.1) *simple Monte Carlo* approach follows:

$$\hat{p}(Y) = \frac{1}{G}\sum_{g=1}^{G} f(Y|\theta^{(g)}), \tag{2.2}$$

where $\{\theta^{(g)}\}_{g=1}^{G}$ is the sample from the priors $\pi(\theta)$. The crucial drawback of this direct method is that the likelihood $f(Y \mid \theta^{(g)})$ is typically very peaked compared to the prior $\pi(\theta)$, so that (2.2) will be a very inefficient estimator (most of the terms in the sum will be near 0).

2.1.1 Harmonic Mean Estimator

To increase the efficiency of the model likelihood estimator it's more profitable to use samples from the posterior distribution. Newton and Raftery (1994) suggested the *harmonic mean* of the posterior sample likelihood as the estimator, i.e.

$$\hat{p}_{HM}(Y) = \left[\frac{1}{M} \sum_{m=1}^{M} \frac{1}{f(Y|\theta^{(m)})} \right]^{-1}, \tag{2.3}$$

where $\{\theta^{(m)}\}_{m=1}^{M} \sim p(\theta|Y)$.

The possible problem with this approach is that it can be quite unstable because the inverse likelihood does not have finite variance (some of the likelihood terms in the sum will be near 0).

2.1.2 Reciprocal Importance Estimator

A useful generalization of (2.3) was provided by Gelfand and Dey (1994). Given samples $\{\theta^{(m)}\}_{m=1}^{M}$ from the posterior distribution, they suggested the *reciprocal importance sampling* estimator

$$\hat{p}_{RI}(Y) = \left[\frac{1}{M} \sum_{m=1}^{M} \frac{h(\theta^{(m)})}{f(Y|\theta^{(m)})\,\pi(\theta^{(m)})} \right]^{-1}, \tag{2.4}$$

This estimator appears to be rather sensitive to the choice of h function. High efficiency is most likely to result if h roughly matches the posterior density. Gelfand and Dey (1994) propose a multivariate normal or t-density with mean and covariance estimated from the posterior sample.

2.1.3 Bridge Sampling

The bridge sampling procedure was first described by Meng and Wong (1996). Kaufmann and Frühwirth-Schnatter (2002) applied this technique to the switching ARCH models.

Let $h(\theta)$ be again a density with known normalizing constant, which is some simple approximation to the posterior density. Let $\alpha(\theta)$ be an arbitrary function satisfying rather general regularity condition. The *bridge sampling* method is based on the following key identity

$$p(Y) = \frac{\int \alpha(\theta) f(Y|\theta)\pi(\theta)h(\theta)d\theta}{\int \alpha(\theta)h(\theta)p(\theta|Y)d\theta} = \frac{\mathbf{E}_h(\alpha(\theta)f(Y|\theta)\pi(\theta))}{\mathbf{E}_p(\alpha(\theta)h(\theta))},$$

where $\mathbf{E}_g$ is the expectation with respect to a density $g(\cdot)$.

If samples $\{\theta^{(m)}\}_{m=1}^M$ and $\{\tilde{\theta}^{(l)}\}_{l=1}^L$ from the posterior $p(\theta|Y)$ and the approximate density $h(\theta)$ respectively are available, then we get the bridge sampling estimator

$$\hat{p}_{BS}(Y) = \frac{L^{-1}\sum_{l=1}^L \alpha(\tilde{\theta}^{(l)})f(Y \mid \tilde{\theta}^{(l)})\pi(\tilde{\theta}^{(l)})}{M^{-1}\sum_{m=1}^M \alpha(\theta^{(m)})h(\theta^{(m)})}. \tag{2.5}$$

Reciprocal importance sampling above is a special case of (2.5) if $\alpha(\theta) = \frac{1}{f(Y|\theta)\pi(\theta)}$.

Meng and Wong (1996) proposed an asymptotically optimal choice of $\alpha(\theta)$ minimizing the expected relative error of the estimator $\hat{p}_{BS}(Y)$ as

$$\alpha(\theta) = \frac{1}{Lh(\theta) + Mp(\theta|Y)}.$$

Since this optimal $\alpha(\theta)$ depends on the normalized posterior $p(\theta|Y)$, the following iterative procedure can be applied: using a previous estimate $\hat{p}_{BS}^{(t-1)}$ we normalize the posterior as $\hat{p}(\theta \mid Y) = \frac{1}{\hat{p}_{BS}^{(t-1)}}f(Y|\theta)\pi(\theta)$ and a new estimate $\hat{p}_{BS}^{(t)}$ is calculated accordingly (2.5), i.e.

$$\hat{p}_{BS}^{(t)}(Y) = \hat{p}_{BS}^{(t-1)}(Y)\frac{\frac{1}{L}\sum_{l=1}^L \frac{\hat{p}(\tilde{\theta}^{(l)}|Y)}{Lh(\tilde{\theta}^{(l)})+M\hat{p}(\tilde{\theta}^{(l)}|Y)}}{\frac{1}{M}\sum_{m=1}^M \frac{\hat{h}(\theta^{(m)})}{Lh(\theta^{(m)})+M\hat{p}(\theta^{(m)}|Y)}}.$$

As initial value for this recursion we can use, e.g., $\hat{p}_{RI}(Y)$.

2.2 Chibs' Candidate's Estimator

To avoid the specification of the approximate function h, Chib (1995) proposed an indirect method for estimating model likelihoods from Gibbs sampling output, recently extended to output from MH algorithm (Chib and Jeliazhov, 2001).

The method is based on the marginal likelihood identity

$$p(Y) = \frac{f(Y|\theta) \cdot \pi(\theta)}{p(\theta|Y)}.$$

Only the denominator on the right hand side is unknown. But since this identity holds for any θ value, we only require a posterior density estimate $\hat{p}(\theta^*|Y)$ at a single point θ^*. Then the proposed estimate of the model likelihood, on the computationally convenient logarithm scale is

$$\log[\hat{p}_{CE}(Y)] = \log f(Y \mid \theta^*) + \log \pi(\theta^*) - \log \hat{p}(\theta^* \mid Y). \tag{2.6}$$

The choice of θ^* is in principal arbitrary, however, the most precise estimate of $\hat{p}(\theta|Y)$ would usually result from using the posterior mode or value close to it. Now our purpose is to estimate the ordinate $\hat{p}(\theta^*|Y)$ given $\{\theta^{(g)}\}_{g=1}^G \sim p(\theta|Y)$.

Suppose that parameter space is split into B blocks, i.e $\theta = (\theta_1, \theta_2, \dots, \theta_B)$
Then we can write

$$p(\theta^*|Y) = p(\theta_1^*|Y) \cdot p(\theta_2^*|Y, \theta_1^*) \cdot \ldots \cdot p(\theta_B^*|Y, \theta_1^*, \dots, \theta_{B-1}^*). \tag{2.7}$$

Suppose also that full conditionals $p(\theta_i|Y, \theta_{-i}) \propto \pi(\theta)\, p(Y|\theta), \quad i = \overline{1, B}$,
where $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_B)$, are sampled by MH algorithm with proposal density $q(\theta_i, \theta_i'|Y, \Theta_{i-1}, \Theta_{i+1})$, where $\Theta_{i-1} = (\theta_1, \dots, \theta_{i-1})$, $\Theta_{i+1} = (\theta_{i+1}, \dots, \theta_B)$, and with probability of move

$$\alpha(\theta_i, \theta_i'|Y, \Theta_{i-1}, \Theta_{i+1}) = \min\{1, \frac{f(Y|\theta_i', \Theta_{i-1}, \Theta_{i+1})\, \pi(\theta_i', \theta_{-i})\, q(\theta_i', \theta_i|Y, \Theta_{i-1}, \Theta_{i+1})}{f(Y|\theta_i, \Theta_{i-1}, \Theta_{i+1})\, \pi(\theta_i, \theta_{-i})\, q(\theta_i, \theta_i'|Y, \Theta_{i-1}, \Theta_{i+1})}\}.$$

Then the reduced ordinate $p(\theta_i^*|Y, \theta_1^*, \dots, \theta_{i-1}^*)$ in (2.7) can be estimated from the output of the reduced MCMC runs as follows:

Step 1. Set $\Theta_{i-1} = \Theta_{i-1}^*$ and draw $\{\theta_i^{(g)}, \dots, \theta_B^{(g)}\}_{g=1}^G$, where $\theta_k^{(g)} \sim p(\theta_k|Y, \theta_{-k})$, $k = i, \dots, B$

Step 2. Include θ_i^* in the condition set. Let $\Theta_i^* = (\Theta_{i-1}^*, \theta_i^*)$ and produce the samples $\{\theta_{i+1}^{(j)}, \dots, \theta_B^{(j)}\}_{j=1}^J$, where $\theta_k^{(j)} \sim p(\theta_k|Y, \theta_{-k})$, $k = i+1, \dots, B$.
At each step also draw $\theta_i^{(j)} \sim q(\theta_i^*, \theta_i|Y, \Theta_{i-1}^*, \Theta_{i+1}^{(j)})$.

Step 3. Estimate the reduced ordinate by the ratio of Monte Carlo averages

$$\hat{p}(\theta_i^*|Y, \theta_1^*, \dots, \theta_{i-1}^*) = \frac{\frac{1}{G}\sum_{g=1}^G \alpha(\theta_i^{(g)}, \theta_i^*|Y, \Theta_{i-1}^*, \Theta_{i+1}^{(g)})\, q(\theta_i^{(g)}, \theta_i^*|Y, \Theta_{i-1}^*, \Theta_{i+1}^{(g)})}{\frac{1}{J}\sum_{j=1}^J \alpha(\theta_i^*, \theta_i^{(j)}|Y, \Theta_{i-1}^*, \Theta_{i+1}^{(j)})}.$$

The reduced runs are obtained by fixing an appropriate set of parameters and continuing the MCMC simulations with a smaller set of distributions.

2.3 Reversible Jump MCMC

The methods described so far all seek to estimate the model likelihood $p(Y)$ for each model separately. They also operate on a posterior sample that has already been produced by some

MCMC method. An alternative approach favoured by many authors is to include the model indicator M_i as a parameter in the sampling algorithm itself.

Green (1995) introduced a reversible jump MCMC strategy for generating from the joint posterior distribution $p(M_i, \theta_i \mid Y)$ based on the standard MH approach. The method is based on creating a Markov chain which operates on the space $\mathbf{M}\times \bigcup_{j\in M} \Theta_j$ ($\mathbf{M}$ - finite set of models, $\theta_j \in \Theta_j \subset \mathbf{R}^{n_j}$) and which can 'jump' between models with parameter spaces of different dimension, while retaining detailed balance ensures the correct limiting distribution $p(M_i, \theta_i \mid Y)$. The condition of detailed balance requires that the equilibrium probability of moving from a state (M_i, θ_i) to (M_j, θ_j) equal to that of moving from (M_j, θ_j) to (M_i, θ_i). For other details see Green (1995).

Suppose the current state of Markov chain be (M_i, θ_i), where θ_i has dimension $\dim(\theta_i) = n_i$. Then the procedure is as follows.

1. Propose a new model M_j with probability $m(M_i, M_j)$;

2. Generate u from a specified proposal density $q(u|\theta_i, M_i, M_j)$;

3. Set $(\theta_j', u') = g_{i,j}(\theta_i, u)$, where $g_{i,j}(\theta_i, u)$ is dimension-matching function, deterministic, invertible. Note that $g_{i,j} = g_{j,i}^{-1}, \quad n_i + \dim(u) = n_j + \dim(u')$;

4. Accept the proposed move to model M_j with probability

   $$\alpha_{i\to j} =$$

   $$\min\left\{1,\ \frac{f(Y|M_j, \theta_j')\,\pi(\theta_j'|M_j)\,\pi(M_j)\,m(M_j, M_i)\,q(u'|\theta_j, M_j, M_i)}{f(Y|M_i, \theta_i)\,\pi(\theta_i|M_i)\,\pi(M_i)\,m(M_i, M_j)\,q(u|\theta_i, M_i, M_j)} \times |J|\right\} \quad (2.8)$$

   with

   $$J = \left|\frac{\partial g_{i,j}(\theta_i, u)}{\partial(\theta_i, u)}\right|.$$

If $M_i = M_j$, then the move is a standard MH step.

To implement the reversible jump MCMC we need to specify the probability $m(M_i, M_j)$ for every proposed move, the proposal distributions $q(u \mid \theta_i, M_i, M_j)$ and the functions $g_{i,j}$ for all i,j. These choices do not affect the results but may be crucial for the convergence rate of the Markov chain. There are many variations or simpler versions of reversible jump that can be applied in specific model selection problems. In particular, if all parameters of the proposed model are generated from a proposal distribution, then $(\theta_j', u') = (u, \theta_i)$ with $\dim(\theta_i) = \dim(u')$ and $\dim(\theta_j') = \dim(u)$ and the Jacobian term in (2.8) is one. If models

M_i and M_j may be described as nested, then there is an extremely natural proposal distribution and transformation function $g_{i,j}$ such that $\dim(u') = 0$ and $\theta'_j = g_{i,j}(\theta_i, u)$. But it should be noted that sometimes the parameters that are "common" to both models M_i and M_j may change dramatically and therefore, the associated parameter values of θ'_j correspond to a region of low posterior probability, which results in reducing of the convergence rate.

Provided that the sampling chain for the model indicator mixes sufficiently well, the posterior probability of model M_i can be estimated by

$$\hat{p}_{RJ}(M_i \,|\, Y) = \frac{1}{G}\sum_{g=1}^{G} I(M^{(g)} = M_i), \quad I(\cdot) \text{ - indicator function,} \tag{2.9}$$

which can in turn be used to estimate the Bayes factor.

3 Models

For the numerical illustration of the strategies above we chose the most important practical models for financial time series - GARCH-type models, introduced for the first time in the original paper of Bollerslev (1986) and then extensively generalized in many aspects, including conditional variance equation and the density specification. Good review of GARCH model formulations and properties was provided by Bollerslev, Engle and Nelson (1994). We refer also to Shephard (1996) for statistical properties of this class of models.

We consider two GARCH models differing only in the type of distribution of the random shocks. To account for possible autocorrelation in financial data, we chose the autoregressive process of the order 1 for the mean filtering. Thus, the model 1 is the AR(1)-GARCH(1,1) model with gaussian errors

$$(M_1) \qquad \begin{cases} y_t = a_0 + a_1 y_{t-1} + e_t, & t = 1, 2, \ldots, T \\ e_t \,|\, I_{t-1} \sim \mathbf{N}(0, \sigma_t^2), & \\ \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad , & \end{cases}$$

with the restrictions $\alpha_0, \alpha_1, \beta_1 \geq 0$ to ensure $\sigma_t^2 > 0$. Stationarity in variance impose that $\alpha_1 + \beta_1 < 1$. I_{t-1} denotes time series history up to time $t-1$.

The model 2 is the AR(1)-GARCH(1,1)-t model in the same parameterization but with

Student errors, i.e.

$$(M_2) \qquad \begin{cases} y_t = a_0 + a_1 y_{t-1} + e_t, & t = 1, 2, \ldots, T \\ e_t \mid I_{t-1} \sim \mathbf{T}_\nu(0, \sigma_t^2), & \\ \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & . \end{cases}$$

As reported in Bollerslev et al. (1994), the Student t-density with its fat tails is more able than the normal distribution to account for the excess kurtosis present in financial data.

Thus, the parameter vector to be estimated in (M_1) is $\theta_1 = (a_0, a_1, \alpha_0, \alpha_1, \beta_1)$ and the likelihood for a sample of T observations $Y = (y_1, y_2, \ldots, y_T)$ can be written as

$$(L1) \qquad f(Y \mid \theta_1) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{e_t^2}{2\sigma_t^2}\right\}.$$

Under the assumption of Student t-distribution, the likelihood for the sample Y is

$$(L2) \qquad f(Y \mid \theta_2) = \prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi(\nu-2)\sigma_t^2}} \left(1 + \frac{e_t^2}{(\nu-2)\sigma_t^2}\right)^{-(\nu+1)/2}$$

where $\nu > 2$ denotes the degrees of freedom of the Student t-distribution and the parameter vector to be estimated is $\theta_2 = (a_0, a_1, \alpha_0, \alpha_1, \beta_1, \nu)$.

The starting point for the Bayesian inference is a prior distribution over the model parameters. Choice of suitable priors is generally a contentious issue. One wants the priors to reflect one's believes about parameter values and at the same time to use a non-informative (flat) priors that does not favor particular values of the parameter over other values.

Note that for all parameters we have to use the proper priors in order to avoid possible non-integrability of the posterior parameter distribution that would make the Bayesian model selection rather questionable (Kass and Raftery, 1995). The choice of the prior on the Student-t degrees of freedom parameter ν is of special relevance with this respect. As shown in Bauwens and Lubrano (1998) the posterior density is not proper when one chooses a flat prior on ν on $(0, \infty)$. The conclusion is that sufficient prior information is needed on ν to force the posterior to tend to zero quickly enough at the tail, in order to be integrable. A possible choices for priors on ν is, e.g., an exponential density $\lambda e^{-\lambda\nu}$ as in Geweke (1993) with parameter λ chosen subjectively to fix the prior mean and variance of ν.

In such a way, we used the following proper priors: $a_0 \sim \mathbf{N}(0, 3)$, $a_1 \sim \mathbf{N}(0, 3)$, $\alpha_0 \sim \mathbf{logN}(-2.3, 5)$, $\alpha_1 \sim \mathbf{logN}(-2.0, 5)$, $\beta_1 \sim \mathbf{logN}(-0.2, 5)$ and $\nu \sim \mathbf{Exp}(0.1)$. The parameters of the priors were approximated based on the maximum likelihood results for these models on various financial data and then the variances were multiplied by factor 50. In this way, such priors turned out to be practically noninformative because their effective range is from 10 to 50 times larger that the effective range of the resulting posterior density. For the prior on degree of freedom parameter, we fix $\lambda = 0.1$ after tuning on synthetic data.

4 MCMC Implementation

The starting point for all model selection strategies presented above is Bayesian inference about the parameter vector θ conditional on data Y via the posterior density $p(\theta \mid Y)$. Using the Bayes theorem, this density takes the form $p(\theta \mid Y) = c \cdot f(Y \mid \theta)\pi(\theta)$ for some normalizing constant c, likelihood function $f(Y \mid \theta)$ and prior density $\pi(\theta)$.

For many realistic problems, evaluation of $p(\theta \mid Y)$ is analytically intractable, so numerical or asymptotic methods are necessary. In this article we adopt the MCMC sampling strategies as the tool to obtain posterior summaries of interest. The idea is based on the construction of an irreducible and aperiodic Markov chain with realizations $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(t)}, \ldots$ in the parameter space, equilibrium distribution $p(\theta|Y)$, and a transition probability $K(\theta'', \theta') = \pi(\theta^{(t+1)} = \theta'' \mid \theta^{(t)} = \theta')$, where θ' and θ'' are the realized states at time t and $t+1$, respectively. Under appropriate regularity conditions, asymptotic results guarantee that as $t \to \infty$ then $\theta^{(t)}$ tends in distribution to a random variable with density $p(\theta|Y)$, and the ergodic average of an integrable function of θ is a consistent estimator of the (posterior) mean of the function. For the underlying statistical theory of MCMC see Tierney (1994).

One popular MCMC procedure is Gibbs sampling but because of the recurrent structure of variance equation in the GARCH model no one of full conditional distributions (i.e., densities of each element or subvector of θ given all other elements) is of known form from which random numbers could be easily generated. There is no property of conjugacy for GARCH model parameters. In this way, we use the Metropolis-Hastings algorithm which gives the easiest black-box sampling strategy yielding the required realization of $p(\theta|Y)$ (see, e.g., Geweke (1995), Kim et al. (1998), Müller and Pole (1998) and Nakatsuma (2000)).

After initial exploratory runs of the Markov chain it was checked for the correlation between the parameters and the blocking update of highly correlated parameters was implemented to increase the efficiency and improve the convergence of Markov chain. Moreover, it appeared that it is more computationally convenient to work with the logarithmic transformation of the variance parameters $(\alpha_0, \alpha_1, \beta_1)$ to a subvector taking values on $(-\infty, +\infty)$. In such a way, we applied the random walk Metropolis algorithm with Student t-distribution as a proposal for the mean parameters and degrees of freedom and with Gaussian proposal for the transformed variance parameters, where the variances of these proposal distributions were tuned to come near ”optimal” acceptance rate in the range 25-40% (see Carlin and Louis (1996) for details).

The output sample of every MCMC run was constructed as follows. First, starting from different initial parameter values, two large samples were simulated, each 40000 iterations. An initial (burn-in) parts of them were discarded after a visual inspection of the series plots for each parameter (usually 10-20% of the run size). Second, the inefficiency factors[1] were calculated and a decision was made about the lag interval with which the sample should be collected to achieve a nearly independent sample (”thinning”). Third, the resulting samples were checked for convergence by using the scale reduction factor[2] $\hat{R}$. And finally, the chains were combined in one output sample used for further analysis.

5 Simulation Study

To be able to compare the model likelihood estimators, we created 210 synthetic data sets, 60 of model 1 and 150 of model 2, generating the model parameters from the priors above. In order to investigate the dependence of the results below on the ”true” parameter ranges,

[1]Inefficiency factor κ is the factor by which we have to increase the run length of the MCMC sampler compared to iid sampling and it accounts for the whole serial dependence in the sampled values (Geweke, 1992) in the following way:

$$\kappa = 1 + 2\sum_{l=1}^{L}\left(1 - \frac{l}{L+1}\right)\rho(l),$$

where $\rho(l)$ represents the autocorrelation at lag l of the sampled parameter values. The bandwidth L is chosen such that $\rho(l),\ l = \overline{1, L}$, significantly contributes to the serial dependence of the sampled values.

[2]According to Gelman and Rubin (1992), running the m parallel chains (initially overdispersed with respect to the true posterior) for $2n$ iterations each, we then attempt to check whether the variation within the chains for a given parameter approximately equals the total variation across the chains during the latter n iterations. Specially, we monitor convergence by the scale factor

$$\hat{R} = \left(\frac{n-1}{n} + \frac{m+1}{mn}\frac{B}{W}\right)\frac{df}{df-2},$$

where B/n is the variance between the means from the m parallel chains, W is the average of the m within-chain variances, and df is the degrees of freedom of an approximating t-density to the posterior distribution. As $n \to \infty$ then $\hat{R} \to 1$ and values of $\hat{R}$ close to 1 suggest good convergence.

we considered the groups of data sets with high (≥ 0.99) and low persistence $\omega = \alpha_1 + \beta_1$ as well as three groups according to the "true" value of degrees of freedom: in the first DF group we included the data sets with $2 < \nu < 10$; for the second DF group we have $10 \leq \nu < 20$ and the third group is $\nu \geq 20$.

MCMC simulations were performed as described above. The final output sample from the posterior densities were of size at least 4000 points.

5.1 MCMC estimates versus maximum likelihood estimates

In order to evaluate the accuracy of the MCMC simulations and compare them with the maximum likelihood estimations, we calculated the estimation errors $\hat{\epsilon}_i = \hat{\theta}_i - \theta_i^*$, $i = \overline{1, 210}$, where θ_i^* denotes the "true" parameter values for synthetic data set i. As estimate $\hat{\theta}$ we used the maximum likelihood estimation as well as the mean and the median of the resulting posterior samples.

In Table 2 we collected the average bias in the estimates together with mean square error MSE= $\frac{1}{210}\sum_{i=1}^{210}\hat{\epsilon}_i^2$ separately for every data group with low(high) persistence and over the DF groups. Obviously, for the mean parameters a_0 and a_1 there are no significant differences between MLE and MCMC estimates over the considered data groups.

The estimation errors for the variance parameters $(\alpha_0, \alpha_1, \beta_1)$ are plotted additionally in Fig.1. Because of the skewness of the posterior distributions for these parameters, the median of the MCMC simulation seems to give more accurate estimates comparing with the mean of the resulting chain, but in general, the maximum likelihood estimates show mostly better performance compared with the main posterior statistics. A possible explanation for this can be that the mean and median statistics of the posterior sample are not completely adequate because of non-normal shape of the posterior distribution for these parameters.

If we again consider these differences over our data groups with low(high) persistence and different degrees of freedom from Table 2, we find no significant differences in the error behavior between these groups.

In turn, the estimation accuracy of the degrees of freedom parameter is significantly different over the considered DF groups (see Fig.2 and Table 2). Fat-tailed conditional distribution (with small values of the degrees of freedom) is recognized quite well by either the MLE or MCMC simulations. But when the degrees of freedom increases, the accuracy of the MLEs is going extremely low and the main MC statistics clearly outperform the

Parameter	"True" distribution	Low persistence (<0.99)						High persistence ($\geq$ 0.99)					
		MLE		Bayesian Analysis				MLE		Bayesian Analysis			
				MC mean		MC median				MC mean		MC median	
		bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE
a_0	$\mathbf{N}$	-0.0266	0.0116	-0.0265	0.0115	-0.0266	0.0115	-0.0012	0.0061	-0.0010	0.0059	-0.0011	0.0060
	$\mathbf{T}_{2<\nu<10}$	-0.0303	0.0151	-0.0285	0.0146	-0.0284	0.0145	0.0192	0.0022	0.0188	0.0021	0.0187	0.0021
	$\mathbf{T}_{10\leq\nu<20}$	-0.0052	0.0033	-0.0055	0.0032	-0.0057	0.0032	0.0143	0.0019	0.0128	0.0019	0.0129	0.0019
	$\mathbf{T}_{\nu\geq 20}$	-0.0236	0.0079	-0.0241	0.0080	-0.0242	0.0080	0.0161	0.0049	0.0160	0.0048	0.0159	0.0048
a_1	$\mathbf{N}$	-0.0014	0.0014	-0.0020	0.0014	-0.0020	0.0014	0.0089	0.0024	0.0084	0.0024	0.0086	0.0024
	$\mathbf{T}_{2<\nu<10}$	-0.0133	0.0015	-0.0136	0.0015	-0.0134	0.0015	-0.0235	0.0031	-0.0236	0.0031	-0.0234	0.0031
	$\mathbf{T}_{10\leq\nu<20}$	-0.0113	0.0021	-0.0115	0.0021	-0.0115	0.0021	-0.0039	0.0019	-0.0043	0.0018	-0.0041	0.0018
	$\mathbf{T}_{\nu\geq 20}$	0.0014	0.0025	0.0012	0.0025	0.0014	0.0025	0.0062	0.0010	0.0055	0.0010	0.0056	0.0010
α_0	$\mathbf{N}$	0.0406	0.0581	0.0817	0.0594	0.0506	0.0483	0.0564	0.0089	0.0538	0.0090	0.0412	0.0070
	$\mathbf{T}_{2<\nu<10}$	0.1880	0.5557	0.1788	0.2418	0.0813	0.0952	0.0411	0.0118	0.0693	0.0197	0.0508	0.0144
	$\mathbf{T}_{10\leq\nu<20}$	-0.0652	0.0486	-0.0213	0.0400	-0.0522	0.0497	0.0236	0.0051	0.0308	0.0058	0.0230	0.0048
	$\mathbf{T}_{\nu\geq 20}$	0.0584	0.0755	0.0762	0.0690	0.0450	0.0717	0.0639	0.0150	0.0606	0.0138	0.0447	0.0104
α_1	$\mathbf{N}$	-0.0002	0.0053	0.0082	0.0057	0.0033	0.0056	-0.0094	0.0018	0.0082	0.0026	0.0037	0.0025
	$\mathbf{T}_{2<\nu<10}$	-0.0007	0.0033	0.0312	0.0061	0.0173	0.0036	-0.0284	0.0051	-0.0030	0.0050	-0.0159	0.0048
	$\mathbf{T}_{10\leq\nu<20}$	-0.0044	0.0033	0.0098	0.0033	0.0041	0.0032	0.0102	0.0032	0.0255	0.0035	0.0198	0.0032
	$\mathbf{T}_{\nu\geq 20}$	0.0037	0.0057	0.0078	0.0046	0.0014	0.0045	0.0196	0.0027	0.0316	0.0031	0.0262	0.0028
β_1	$\mathbf{N}$	-0.0163	0.0196	-0.0506	0.0204	-0.0466	0.0218	-0.0049	0.0014	-0.0134	0.0015	-0.0100	0.0014
	$\mathbf{T}_{2<\nu<10}$	-0.0159	0.0502	-0.0749	0.0398	-0.0781	0.0471	-0.0034	0.0028	-0.0278	0.0046	-0.0200	0.0040
	$\mathbf{T}_{10\leq\nu<20}$	0.0024	0.0063	-0.0239	0.0083	-0.0184	0.0078	-0.0191	0.0045	-0.0296	0.0047	-0.0256	0.0046
	$\mathbf{T}_{\nu\geq 20}$	-0.0704	0.0583	-0.0775	0.0405	-0.0827	0.0494	-0.0225	0.0022	-0.0267	0.0023	-0.0232	0.0021
ν	$\mathbf{T}_{2<\nu<10}$	1.3039	18.8982	1.5954	15.4388	1.0101	10.0927	0.6374	2.0644	1.2725	4.9588	0.8259	2.6000
	$\mathbf{T}_{10\leq\nu<20}$	9.6148	760.4358	1.2402	27.9151	-0.6073	21.5133	7.8483	458.6900	2.3369	22.2701	0.2665	13.7749
	$\mathbf{T}_{\nu\geq 20}$	15.0428	1292.9426	-8.9342	138.6652	-11.2424	183.8033	24.7443	1942.6505	-8.1767	112.3830	-10.4228	151.3605

Table 2: Bias and MSE for MLE and MC estimations

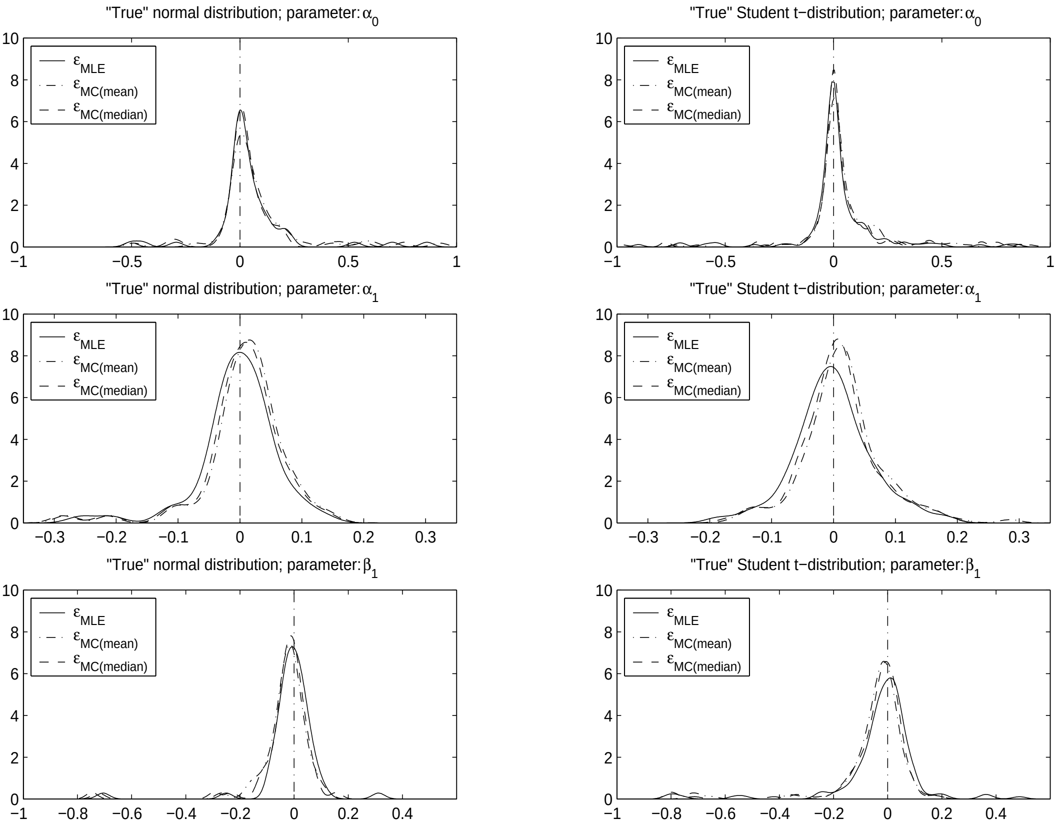

Figure 1: Estimation error plots of variance parameters α_0, α_1 and β_1. ε_{MLE} denotes the error of the maximum likelihood estimates, the notations $\varepsilon_{\text{MC(mean)}}$ ($\varepsilon_{\text{MC(median)}}$) are used for the estimation errors of the mean (median) of the posterior samples.

maximum likelihood estimates.

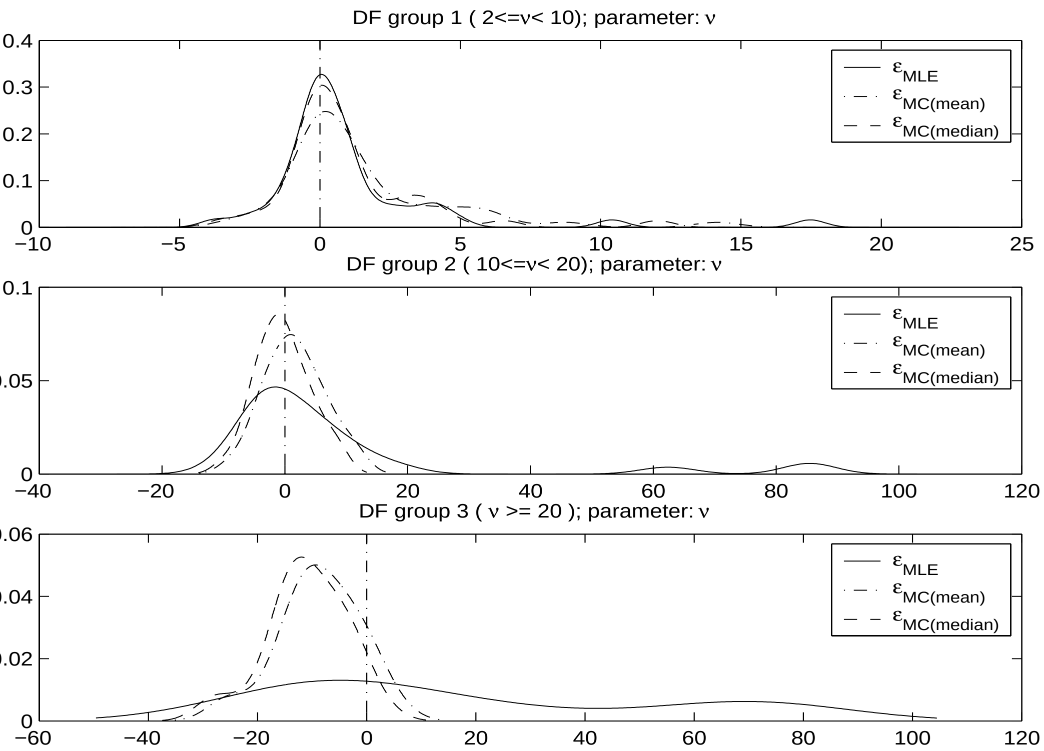

Figure 2: Estimation error plots of the degrees of freedom parameter separately for every DF group. $\varepsilon_{\mathrm{MLE}}$ denotes the error of the maximum likelihood estimates, the notations $\varepsilon_{\mathrm{MC(mean)}}$ ($\varepsilon_{\mathrm{MC(median)}}$) are used for the estimation errors of the mean (median) of the posterior samples.

5.2 Comparison of model likelihood estimators

Taking the resulting chain output from above, we used it as an input to the computational methods for estimating of model likelihoods discussed in section 2. According to formula (2.3) we estimated $\hat{p}_{HM}(Y)$. As importance density in reciprocal importance procedure we used a multivariate normal for the estimates $\hat{p}^{(1)}_{RI}(Y)$ and t-distribution with DF=3 for $\hat{p}^{(2)}_{RI}(Y)$ with parameters estimated from the posterior sample. The same multivariate gaussian approximation of the posterior sample we took as $h(\cdot)$ function to compute the bridge sampling estimate $\hat{p}_{BS}(Y)$. According to (2.6) we calculated the estimation $\hat{p}_{CE}(Y)$.

Following Vrontos et al. (2000), for reversible jump MCMC we implemented the algorithm where all the parameters of the proposed model are generated from a proposal distribution. Consequently, in (2.8) we have that $(\theta'_j, u') = (u, \theta_i)$ with $\dim(\theta_i) = \dim(u')$, $\dim(\theta'_j) = \dim(u)$ and $q(u|\theta_i, M_i, M_j) = q(u|M_j)$, $q(u'|\theta'_j, M_j, M_i) = q(u'|M_i)$ and the Jacobian term is 1. The proposal densities $q(u|M_j)$ and $q(u'|M_i)$ are chosen by investigation of MCMC output as a multivariate normal density with mean and covariance matrix estimated from the chains. To complete the specification of our reversible jump MCMC algorithm we need to specify the probabilities $m(M_i, M_j)$. We have used equally defined probabilities of the ”jump” to the other model and for staying within the model.

In such a way, we run the reversible jump MCMC for 40000 iterations. The behaviour of the MCMC chains was good with rapid convergence to the probabilities $p(M \mid Y)$. The Bayes factor in favour of the ”right” model was calculated based on (2.9).

In order to compare the performance of the discussed model selection algorithms, we calculated the percentage of correct decisions based on the scale of evidence for Bayes factors from Table 1. Thus, we separated the sets where we get strong or moderate or weak evidence in favour of the ”right” model as well as the evidence against the ”right” model. In Tables 3-7 we report these results related to the common number of sets in the corresponding group. The abbreviation ’HM’ is used for the harmonic mean estimator, ’CE’ denotes the Chib candidate estimator, ’RI-N’ and ’RI-T’ are the reciprocal importance estimators with multivariate normal and Student t importance densities, respectively. ’BS’ denotes the bridge sampling estimator and under ’RJMCMC’ we collect the results of the reversible jump MCMC.

There is high correlation between the estimated Bayes factors and the degree of freedom value used to create the corresponding synthetic data set. Whereas for values less than

Distribution	HM	CE	RI-N	RI-T	BS	RJMCMC	$\triangle$BIC
$\mathbf{N}$	14/60	20/60	19/60	18/60	20/60	5/60	50/60
$\mathbf{T}_{2<\nu<10}$	42/48	44/48	45/48	45/48	44/48	42/48	39/48
$\mathbf{T}_{10\leq\nu<20}$	18/49	17/49	17/49	17/49	17/49	14/49	6/49
$\mathbf{T}_{\nu\geq 20}$	9/53	4/53	4/53	4/53	4/53	4/53	2/53

Table 3: Strong evidence for ”right” model (BF$\geq$10).

Distribution	HM	CE	RI-N	RI-T	BS	RJMCMC	$\triangle$BIC
$\mathbf{N}$	19/60	23/60	22/60	23/60	20/60	33/60	8/60
$\mathbf{T}_{2<\nu<10}$	4/48	2/48	1/48	1/48	2/48	3/48	3/48
$\mathbf{T}_{10\leq\nu<20}$	8/49	7/49	7/49	7/49	8/49	8/49	2/49
$\mathbf{T}_{\nu\geq 20}$	3/53	2/53	4/53	4/53	3/53	1/53	2/53

Table 4: Moderate evidence for ”right” model (3$\leq$BF$<$10).

10 the right decision (strong and moderate support for the ”right” model) is taken in 95 % cases (46/48), for the degrees of freedom values more than 20 - only in max 20 % of all cases (12/53). And in this last group with $\nu \geq 20$ the decision against the ”right” model was accepted at least in 26 % (14/53). It seems also that Chibs candidate’s algorithm, reciprocal importance estimators with multivariate normal or fat-tailed t-distributions as importance densities and bridge sampling method are statistically equal in their performance for GARCH-type models. Harmonic mean estimator behaved better for the higher DF group, giving more strong support to the ”right” model, but for the normal distribution it shows significantly worse performance with the large percentage of the undefined and incorrect decisions.

Distribution	HM	CE	RI-N	RI-T	BS	RJMCMC	$\triangle$BIC
$\mathbf{N}$	15/60	8/60	11/60	11/60	11/60	15/60	0/60
$\mathbf{T}_{2<\nu<10}$	1/48	2/48	2/48	2/48	2/48	3/48	2/48
$\mathbf{T}_{10\leq\nu<20}$	9/49	8/49	12/49	12/49	9/49	10/49	8/49
$\mathbf{T}_{\nu\geq 20}$	9/53	12/53	12/53	11/53	13/53	11/53	0/53

Table 5: Weak evidence for ”right” model (1$\leq$BF$<$3).

It is interesting to compare results above to the classical approach to model selection. The most common classical technique is Bayesian Information Criterion (BIC) (Schwarz,

Distribution	HM	CE	RI-N	RI-T	BS	RJMCMC	△BIC
$\mathbf{N}$	7/60	7/60	6/60	6/60	7/60	5/60	1/60
$\mathbf{T}_{2<\nu<10}$	0/48	0/48	0/48	0/48	0/48	0/48	1/48
$\mathbf{T}_{10\leq\nu<20}$	6/49	13/49	9/49	10/49	12/49	14/49	8/49
$\mathbf{T}_{\nu\geq 20}$	15/53	17/53	18/53	17/53	17/53	23/53	2/53

Table 6: Weak evidence against "right" model (1/3≤BF<1).

Distribution	HM	CE	RI-N	RI-T	BS	RJMCMC	△BIC
$\mathbf{N}$	5/60	2/60	2/60	2/60	2/60	2/60	1/60
$\mathbf{T}_{2<\nu<10}$	1/48	0/48	0/48	0/48	0/48	0/48	3/48
$\mathbf{T}_{10\leq\nu<20}$	8/49	4/49	4/49	3/49	3/49	3/49	25/49
$\mathbf{T}_{\nu\geq 20}$	17/53	18/53	15/53	17/53	16/53	14/53	47/53

Table 7: Moderate and strong evidence against "right" model (BF<1/3).

1978) where one chooses the model M_i that maximizes

$$\log f(Y \mid \hat{\theta}_i, M_i) - \frac{n_i}{2} \log T,$$

where $\hat{\theta}_i$ denotes the maximum likelihood estimate of the parameters of the model M_i, n_i is the dimension of θ_i and T is the size of the data Y. It was also shown that for large sample size T an approximation to the Bayes factor $\mathbf{B}_{12}$ is given by

$$\log \mathbf{B}_{12} \approx \triangle\text{BIC} = \log \frac{f(Y \mid \hat{\theta}_1, M_1)}{f(Y \mid \hat{\theta}_2, M_2)} + \frac{n_2 - n_1}{2} \log T. \qquad (5.1)$$

The second term in △BIC acts as a penalty term which corrects for differences in size between the models. For more details on BIC see Carlin and Louis (1996).

The resulting values of △BIC are reported in Tables 3-7 in the last column. It is obvious that BIC clearly favour the "simpler" model with normal distribution. It gives stronger evidence for the "true" AR(1)-GARCH(1,1) model but poorly recognizes the t-distribution with the higher degrees of freedom values.

6 Market Data

In our empirical study we used three data sets related to different financial markets:

1. daily closing values of the American stock index Dow Jones Industrial Average (DJIA);

2. daily closing values of the FTSE 100 traded at the London Stock Exchange;

3. daily closing values of the Japan index NIKKEI 225.

The data were downloaded from http://finance.yahoo.com. The taken time interval for all series was 10 years starting with Januar, 1992 and ending with December, 2001. The data were transformed into continuously compounded returns r_t (in percent) in the standard way by the natural logarithm of the ratio of consecutive daily closing levels and the whole samples of about 2500 observations were used for the Bayesian inference.

For MCMC simulations we applied again the random walk Metropolis algorithm. After initial exploratory runs of MCMC we found high correlation between the variance parameters α_0, α_1 and β_1 with cor$(\alpha_0, \beta_1) = -0.8$ cor$(\alpha_1, \beta_1) = -0.9$ and during further simulations we updated these parameters in one block using multivariate normal proposal with the corresponding correlation matrix. The inefficiency factors varied from 2.8 for the mean parameters till 6.5 for the variance parameter β_1 with the common chain size of 5000 points. The convergence indicator $\hat{R}$ were in the range 1.0002-1.08, proving good convergence of the chains. The resulting marginal posterior densities are presented on Fig.3 and 4. Solid line denotes the posterior parameter plots when estimated on DJIA return series, dotted line denotes the results for FTSE 100 data and dashed line is used for NIKKEI 225 data.

Note that the differences between the markets in the posterior plots for the parameters of AR(1)-GARCH(1,1) model become less noticeable when we apply the conditionally fat-tailed AR(1)-GARCH(1,1)-t model. Moreover, Fig 4 shows that the conditional distribution of FTSE 100 returns has lower kurtosis comparing with the American and the Japan index returns. The 95% posterior confidence interval for degrees of freedom parameter came to $[9.1\ 20.8]$ for this return series against $[5.3\ 9.4]$ for another two.

We want to note that we did not impose stationarity conditions in a Bayesian context, neither in priors, nor in proposal distributions. But after each MCMC run we also estimate the posterior probability of $\alpha_1 + \beta_1 < 1$ as

$$\mathbf{E}(\mathbf{P}(\alpha_1 + \beta_1 < 1)) \approx \frac{1}{G} \sum_{g=1}^{G} I(\alpha_1^{(g)} + \beta_1^{(g)} < 1),$$

where $I(\cdot)$ denotes the indicator function. The estimated posterior probabilities for all data sets are in the range 0.93-1.00, showing that GARCH process has the finite unconditional variance for all return series under consideration.

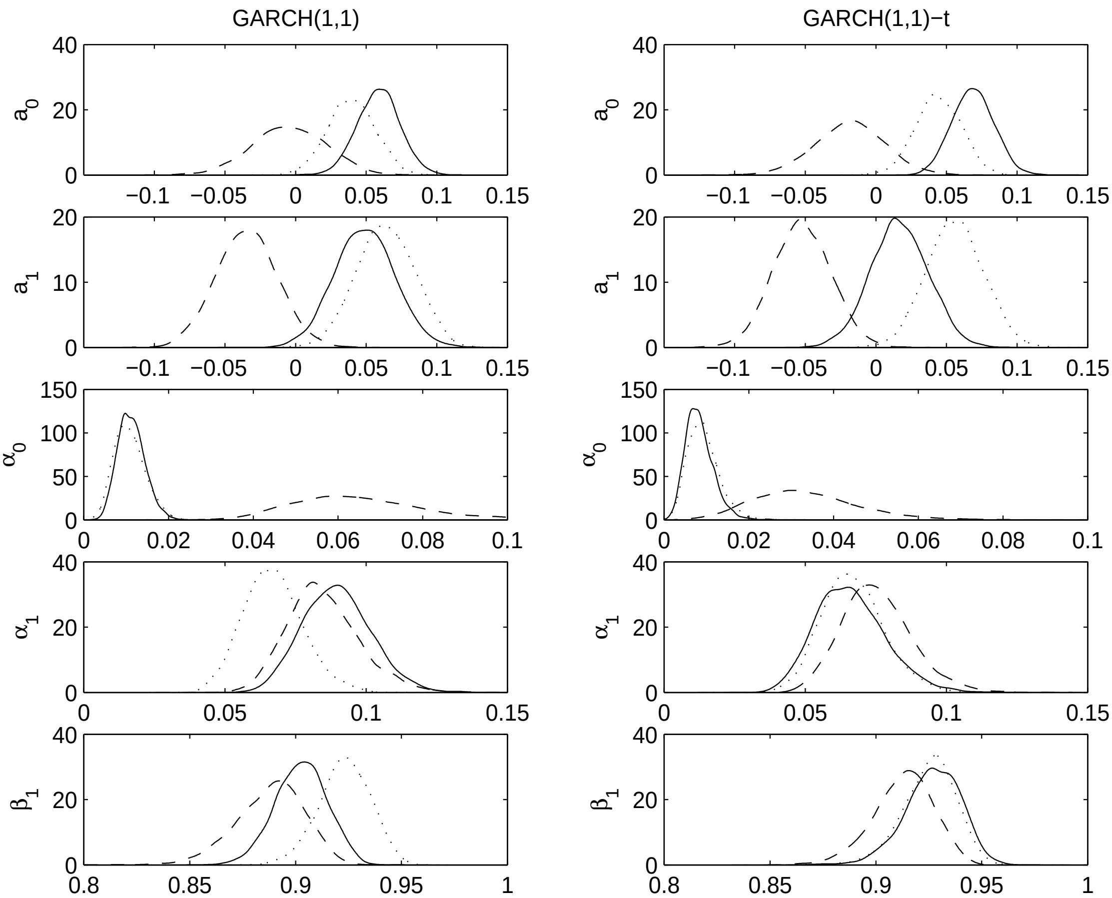

Figure 3: Marginal posterior densities of the parameters a_0, a_1, α_0, α_1 and β_1 for AR(1)-GARCH(1,1) and AR(1)-GARCH(1,1)-t models are depicted on the left- and on the right-hand side, respectively. Solid line corresponds DJIA return series, dotted line denotes the results for FTSE 100 data and dashed line is used for NIKKEI 225 data.

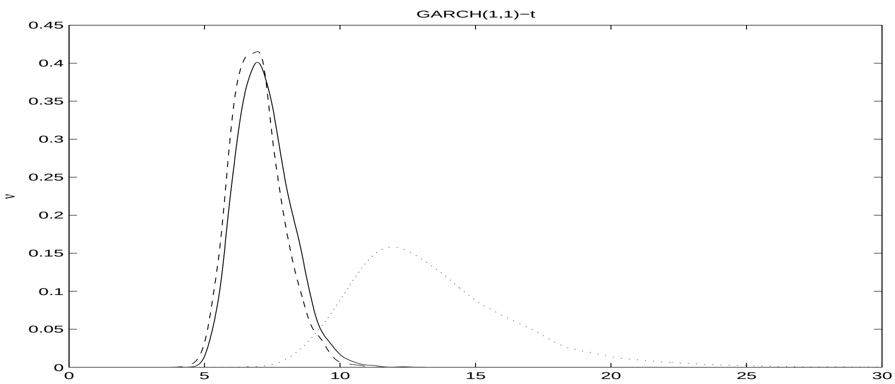

Figure 4: Marginal posterior densities of the degrees of freedom parameter for AR(1)-GARCH(1,1)-t model. Solid line corresponds DJIA return series, dotted line denotes the results for FTSE 100 data and dashed line is used for NIKKEI 225 data.

As the next step we computed all model likelihood estimators discussed above. The resulting values of the logarithm of the model likelihood together with the Bayes factor in favour of model AR(1)-GARCH(1,1) with the gaussian distribution are presented in Table 8. The posterior model probabilities estimated by RJMCMC algorithm were 0 and 1 for model 1 and 2,respectively, and are not included in the table.

Data	HM	CE	RI-N	RI-T	BS	BIC
	-3230.18	-3246.65	-3246.73	-3246.59	-3246.74	-3245.02
DJIA	-3177.42	-3194.24	-3194.28	-3194.09	-3194.30	-3194.48
	1.22e-23	1.73e-23	1.66e-23	1.58e-23	1.68e-23	1.12e-22
	-3355.14	-3371.64	-3371.69	-3371.50	-3371.72	-3370.00
FTSE	-3338.77	-3355.48	-3355.36	-3355.16	-3355.41	-3356.13
100	7.77e-08	9.59e-08	8.09e-08	8.01e-08	8.25e-08	9.47e-07
	-4342.75	-4356.85	-4356.90	-4356.73	-4356.90	-4356.15
NIKKEI	-4284.97	-4301.60	-4301.55	-4301.35	-4301.56	-4302.79
225	8.06e-26	1.01e-24	9.16e-25	8.89e-25	9.25e-25	6.70e-024

Table 8: Model likelihoods for market data. The first and second rows in every data set part contain the estimated model likelihood values (logarithm) for the models GARCH(1,1) and GARCH(1.1)-t, respectively. In the third row the Bayes factors in favour of model 1 are given.

As expected we observe the overwhelming superiority of the fat-tailed conditional distribution for all data considered. Within the discussed strategies, the reciprocal importance (RI-N and RI-T), bridge sampling (BS) and Chib candidate (CE) estimators seem to have equal accuracy with rather small standard errors, while the harmonic mean (HM) estimator leads to the biased model likelihood values.

In the last column of Table 8 we presented the BIC statistics together with the approximation to the BF (5.1). It seems that the BIC again tends to smooth the differences between the models.

7 Conclusions

In this paper we give a review of popular model selection methods in the Bayesian framework, such as harmonic mean, reciprocal sampling, bridge sampling, Chibs' candidate formula and reversible jump MCMC algorithm. For numerical illustration we provide the detailed implementation guidelines for these methods together with MCMC methodology for GARCH-type models from financial econometrics. We ground our method comparison on the simulated data grouped with respect to the persistence level and the degree of kurtosis in the conditional distribution. Based on the simulation study and from our experience on the algorithm implementations we can make the following conclusions:

- the harmonic mean estimator is the simplest in implementation but showed the largest standard errors. Surprisingly, its performance on the synthetic data from the weakly leptokurtic conditional distribution appeared to be good compared to the others methods.

- we found equal accuracy of the Chibs candidate's algorithm; reciprocal importance estimators with both choices of importance densities; and bridge sampling method. But the computational time required for the Chibs candidate's estimates was five times longer compared with the others.

- the reversible jump MCMC algorithm seems to produce significantly weaker evidence for the right models over all data groups considered. Moreover, the coding of this method takes extra time when comparing with the direct methods.

- all the model selection procedures considered above mostly fail to make the right choice when data were generated by the weakly leptokurtic conditional distribution, preferring the simpler gaussian model in these cases. The high level of the variance persistence in the data seems to have no influence on the results.

- The applied classical approach to model selection(BIC) results in clear preferences for the simpler model over all data groups. Without doubt, the Bayesian framework offers advantages compared to the classical analysis, taking into account model complexity in more unbiased way.

In addition, we compared the accuracy of the MCMC simulations with maximum likelihood estimators over the considered data groups. The mean parameters were estimated in both cases with the same accuracy. With respect to the variance parameters, the calculated estimation errors of the maximum likelihood procedure seem to be in general lower compared to the MCMC results. We can explain this effect by the non-symmetrical and non-normal shape of the posterior densities for the variance parameters and, therefore, the taken posterior statistics like mean and median of these distributions are not completely adequate. These results for the mean and variance parameters are not correlated with the kurtosis of the ”true” conditional distribution. As regards the degree of freedom parameter, we got the opposite picture. With the increase of the ”true” degree of freedom value the accuracy of maximum likelihood estimates is getting very low compared to the posterior statistics. It seems that the Bayesian approach by its ”subjectivism” (in the choice of the priors) can manage a situation like this. Again, all results above were not sensitive to the level of the variance persistence.

In the empirical study we performed complete Bayesian analysis of the GARCH models for the stock index returns from the three largest financial markets. As expected we proved the overwhelming superiority of the fat-tailed conditional distribution for all data considered.

Acknowledgements

The authors would like to thank Regina Tüchler for valuable discussions. This work was funded by the Austrian Science Fund (FWF) under grant SFB#010: “Adaptive Information Systems and Modeling in Economics and Management Science”. The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry for Education, Science and Culture and by the Austrian Federal Ministry for Transport, Innovation and Technology.

References

Bauwens, L. and Lubrano, M. (1998). Bayesian inference on GARCH models usin Gibbs sampler, *Econometrics Journal* **1**: 23–46.

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

Bollerslev, T., Engle, R. and Nelson, D. (1994). ARCH models, *in* R. Engle and D. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier Science B.V., Amsterdam.

Carlin, B. and Louis, T. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman&Hall, London.

Chib, S. (1995). Marginal likelihood from the Gibbs output, *Journal of American Statistical Association* **90**(432): 1313–1321.

Chib, S. and Jeliazhov, I. (2001). Marginal likelihood from the Metropolis-Hastings output, *Journal od American Stattistical Association* **96**(453): 270–281.

Frühwirth-Schnatter, S. (1995). Bayesian model discrimination and bayes factor for linear gaussian state space models, *Journal of the Royal Statistical Society, series B* **57**: 237–246.

Frühwirth-Schnatter, S. (2002). Model likelihoods and bayes factors for switching and mixture models, *Technical report*, SFB Adaptive Information Systems and Modelling in Economics and Management Science. can be downloaded from http://www.wu-wien.ac.at/am/reports.htm.

Gelfand, A. and Dey, D. (1994). Bayesian model choice: Asymptotic and exact calculations, *Journal of Royal Statistical Society, Ser. B* **56**: 501–514.

Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal densities, *Journal of American Stattistical Association* **85**: 398–409.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences (with discussion), *Statistical Science* **7**: 457–511.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* J. Bernardo, J. Berger, A. Dawid and A. Smith (eds), *Bayesian Statistics*, Vol. 4, Oxford University Press, pp. 169–193.

Geweke, J. (1993). Bayesian treatment of the independent student-t linear model, *Journal of Applied Econometrics* **8**: 19–40.

Geweke, J. (1995). Bayesian comparison of econometric models. Working Papers 532, Federal Reserve Bank of Minneapolis.

Green, P. (1995). Reversible jump MCMC computation and Bayesian model determination, *Biometrika* **82**: 711–732.

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

Jeffreys, H. (1961). *Theory of Probability, 3rd edition*, Oxford Univiersity Press, Oxford.

Kass, R. and Raftery, A. (1995). Bayes factor, *Journal od American Statistical Association* **90**: 773–792.

Kaufmann, S. and Frühwirth-Schnatter, S. (2002). Bayesian analysis of switching ARCH models, *Journal of Time Series Analysis* **23**(4): 425–458.

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

Kleibergen, F. and van Dijk, H. (1993). Non-stationarity in GARCH models: a bayesian analysis, *Journal of Applied Econometrics* **8**: 41–61.

Lavine, M. and Scherrish, M. (1999). Bayes factors : What they are and what they are not, *The American Statistician* **53**: 119–122.

Meng, X. and Wong, W. (1996). Simulating ratios of normalizing constants via a simple identity, *Statistical Sinica* **6**: 831–860.

Müller, P. and Pole, A. (1998). Monte carlo posterior integration in GARCH models, *Sankhya - The Indian Journal of Statistics* **60**: 127–144.

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

Newton, M. and Raftery, A. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap, *Journal of Royal Statistical Society, Ser. B* **56**: 1–48.

Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics* **6**: 461–464.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility, *in* D. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen (eds), *Time Series Models in Econometrics, Finance and Other Fields*, London: Chapman & Hall, pp. 1–67.

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

Vrontos, I., Dellaportas, P. and Politis, D. (2000). Full Bayesian inference for GARCH and EGARCH models, *Journal of Business & Economic Statistics* **18**(2): 187–198.

Wasserman, L. (1997). Bayesian model selection and model averaging, *Technical report*, Statistics Department, Carnegie Mellon University.