ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

MCMC APPLICATIONS IN TIME-VARYING VOLATILITY MODELS

By

Ioannis D. Vrontos

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfillment of the requirements for

the degree of PhD in Statistics

Athens, Greece

February 2001

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

MCMC ΕΦΑΡΜΟΓΕΣ ΣΕ ΜΟΝΤΕΛΑ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ

Ιωάννης Δ. Βρόντος

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Διδακτορικού Διπλώματος στη Στατιστική

Αθήνα
Φεβρουάριος 2001

To my family

Introduction

Econometric time series models of changing variance and covariance have attracted a lot of attention in modern economic theory. These models, which we call time-varying volatility models, can capture the stylized facts which are observed in the squares or absolute values of financial time series. The estimation of time-varying variances and covariances is crucial for asset pricing, risk management and portfolio analysis. In this thesis, we study and analyze existing univariate and multivariate time-varying volatility models and we propose a new class of multivariate models. A complete analysis of the models under consideration is presented by using Bayesian techniques. In particular, we use Markov Chain Monte Carlo (MCMC) method, a powerful computer-intensive statistical tool, which has been used extensively in a great variety of applications.

The thesis is organized in seven chapters. In the first chapter, we present Bayesian inference by using MCMC methods. We discuss how MCMC algorithms can be constructed to produce Markov chains with the desired stationary distribution, and present alternative updating schemes that have been proposed in the literature. We also discuss the Bayesian approach to model selection problem by using the Bayes Factor, and present Bayesian model averaging that accounts for model uncertainty. In chapter two, we propose a new diagnostic that can be used to detect when convergence of the MCMC algorithms is achieved. This convergence diagnostic is based on the subsampling methodology, and on the construction of $(1-a)\,100\%$ confidence regions for the mean and for the $90-th$ percentile of the first marginal distribution of the Markov chain.

In the third chapter, we present some well known univariate and multivariate time-

varying volatility models, and discuss their properties. In chapter four, a full Bayesian analysis of univariate GARCH and EGARCH models is proposed. We estimate the model parameters by using MCMC methods, provide posterior model probabilities for the analyzed models, and present a way to predict the future volatility via Bayesian model averaging. We provide implementation details and illustrations using the General index of the Athens stock exchange. An analysis of a multivariate class of ARCH and GARCH models is presented in chapter five. Bayesian and classical techniques are used for the estimation of the parameters of the models, and model comparisons are addressed via predictive distributions. We provide implementation details and illustrations using daily exchange rates of the Athens exchange market.

A new multivariate time series model with time-varying conditional variances and covariances is introduced and analyzed in chapter six. A complete analysis of the proposed model is presented consisting of parameter estimation, model selection and volatility prediction. Classical and Bayesian techniques are used for the estimation of the model parameters. It turns out that the construction of our proposed model allows easy maximum likelihood estimates and construction of well-mixing MCMC algorithms. Bayesian model selection is addressed using Markov chain Monte Carlo model composition. The problem of accounting for model uncertainty is considered using Bayesian model averaging. We provide implementation details and illustrations using daily rates of return on eight stocks of the US market. Chapter 7 concludes the thesis and discusses some points for future research.

Chapter 1

Bayesian Inference

1.1 Introduction

Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities of interest. Using a Bayesian framework, observables and model parameters are considered random quantities. However, the Bayesian paradigm requires the calculation of various, usually intractable, multivariate integrals. The purpose of this chapter is to present Markov Chain Monte Carlo methods, that have been suggested to deal with this problem. We present more complex ideas, and discuss some implementational issues associated with MCMC methods. We present the Bayes Factor, which is used for inference about the model selection problem, and discuss some model selection strategies based on MCMC methods. We present and discuss ideas that have been mainly used in our analysis of the time varying volatility models, and we do not provide a review on these topics.

1.2 Bayesian Inference for a given model

1.2.1 The problem

Bayesian inference for parametric statistical models is based on the posterior distribution. Given data $\mathbf{y} = (y_1, y_2, ..., y_T)$ from a distribution with likelihood function $l(\mathbf{y}|\boldsymbol{\theta})$ and a prior density $p(\boldsymbol{\theta})$ for the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)$, $\boldsymbol{\theta} \in \mathbf{R}^n$, Bayes' theorem can be applied to obtain the joint posterior density

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{l(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int l(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \partial\boldsymbol{\theta}}. \tag{1.1}$$

The posterior density can be thought as a description of what is known about $\boldsymbol{\theta}$ from the prior information $p(\boldsymbol{\theta})$ and the data $l(\mathbf{y}|\boldsymbol{\theta})$. The integral $\int l(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \partial\boldsymbol{\theta}$ does not depend on $\boldsymbol{\theta}$ and the posterior can be written as

$$p(\boldsymbol{\theta}|\mathbf{y}) = c(\mathbf{y})\, l(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}),$$

where $c(\mathbf{y}) = 1/\int l(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \partial\boldsymbol{\theta}$. Therefore, the posterior can be written up to a constant of proportionality as

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto l(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}).$$

The posterior density $p(\boldsymbol{\theta}|\mathbf{y})$ provides the basis for inference about $\boldsymbol{\theta}$. Marginal posterior distributions for a parameter of interest, say θ_i, can be obtained by integrating over the parameter vector $\boldsymbol{\theta}$ except the $i-th$ element. That is,

$$p(\theta_i|\mathbf{y}) = \int p(\boldsymbol{\theta}|\mathbf{y})\, \partial\boldsymbol{\theta}_{\setminus i},$$

where $\boldsymbol{\theta}_{\setminus i}$ is the $(n-1) \times 1$ parameter vector which contains all the elements except the $i-th$.

Posterior expectations such as means, variances, covariances, predictive distributions etc., can be derived via $p(\boldsymbol{\theta}|\mathbf{y})$. For example, the posterior expectation of a function of interest $f(\boldsymbol{\theta})$ can be written in the form

$$\begin{aligned} E[f(\boldsymbol{\theta})|\mathbf{y}] &= \int f(\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathbf{y})\, \partial\boldsymbol{\theta} \\ &= \frac{\int f(\boldsymbol{\theta})\, l(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \partial\boldsymbol{\theta}}{\int l(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \partial\boldsymbol{\theta}}. \end{aligned} \tag{1.2}$$

These quantities of interest (normalizing constant, marginal posterior distributions, posterior expectations) require the computation of high dimensional integrals which make the Bayesian inference difficult, from the computational point of view. In most applications, analytic evaluation of these quantities is impossible, especially in high dimensions. Different computational strategies have been suggested to deal with the calculation of integrals required: numerical integration techniques (see, for example, Naylor and Smith, 1982, Smith *et al.*, 1985, Dellaportas and Wright, 1991a, 1991b, 1992), analytical approximations (Tierney and Kadane, 1986, Tierney, Kass and Kadane, 1989a, 1989b, Kass, Tierney and Kadane, 1988), Monte Carlo methods (Stewart, 1983, 1987, Stewart and Davis, 1986, Geweke, 1988, 1989a), Sampling - Resampling, Markov Chain Monte Carlo methods. A description of these strategies is given, for example, in Bernardo and Smith (1994), Smith (1991), Evans and Swartz (1995), Dellaportas (1996). We describe sampling-based methods, such as Monte Carlo integration, Sampling-Resampling, Markov Chain Monte Carlo methods, because of their conceptual simplicity and ease of implementation in high dimensional problems. We present in detail Markov Chain Monte Carlo methods which are the main tools in our analysis.

1.2.2 Monte Carlo Integration

Monte Carlo methods provide a systematic approach that can be applied in order to evaluate the expectation of a function of interest $f(\boldsymbol{\theta})$, given in equation (1.2). The main idea is simple. Let $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_m\}$ be an iid sequence of n-dimensional random vectors

having common probability density $I(\boldsymbol{\theta})$. Then, the Monte Carlo approach estimates the expectation in (1.2) by

$$\widehat{E[f(\boldsymbol{\theta})|\mathbf{y}]} = \frac{\sum_{i=1}^{m} f(\boldsymbol{\theta}_i)\, w(\boldsymbol{\theta}_i)}{\sum_{i=1}^{m} w(\boldsymbol{\theta}_i)},$$

where, $w(\boldsymbol{\theta}_i) = l(\mathbf{y}|\boldsymbol{\theta}_i)\, p(\boldsymbol{\theta}_i) / I(\boldsymbol{\theta}_i)$. The density $I(\boldsymbol{\theta})$ is called *importance sampling density* and the process of generating $\boldsymbol{\theta}_i$, $i = 1, ..., m$ according to I is called *importance sampling*. The efficiency of the method depends on the choice of a suitable importance sampling density I. Importance sampling densities are usually derived from multivariate normal or Student t distributions. Monte Carlo methods have been extensively used in Bayesian inference in Econometric models. See, for example, Kloek and van Dijk (1978), van Dijk and Kloek (1983, 1985), van Dijk, Hop and Louter (1987), Geweke (1988, 1989a, 1991).

1.2.3 Sampling-Resampling approach

According to (1.1) the inference process is done by updating the prior density $p(\boldsymbol{\theta})$ to the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$ through the likelihood function $l(\mathbf{y}|\boldsymbol{\theta})$. Sampling - Resampling approach focuses on Bayes' theorem from a sampling perspective. That is, having a sample from the prior $p(\boldsymbol{\theta})$ we form a sample from the posterior $p(\boldsymbol{\theta}|\mathbf{y})$ through the likelihood function $l(\mathbf{y}|\boldsymbol{\theta})$. This can be done by using the Rejection method for generating random variates (see, for example, Ripley, 1987) or the weighted Bootstrap, which is a variant of the Bootstrap resampling procedure (Efron, 1982).

According to the rejection method, each $\boldsymbol{\theta}$ generated from the prior density $p(\boldsymbol{\theta})$, is accepted as a sample point from the posterior $p(\boldsymbol{\theta}|\mathbf{y})$ with probability

$$\frac{l(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{l\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}\right) p(\boldsymbol{\theta})} = \frac{l(\mathbf{y}|\boldsymbol{\theta})}{l\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}\right)},$$

otherwise $\boldsymbol{\theta}$ is rejected. $l\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}\right)$ is the value of the likelihood function at the maximum $\widehat{\boldsymbol{\theta}}$. That is, these $\boldsymbol{\theta}$ in the prior sample having high likelihood are more likely to be taken in the posterior sample.

According to weighted bootstrap, if $\boldsymbol{\theta}_i$, $i = 1, ..., m$ is a sample from the prior density $p(\boldsymbol{\theta})$, we can take a sample from the posterior $p(\boldsymbol{\theta}|\mathbf{y})$ as follows: calculate

$$w_i = l(\mathbf{y}|\boldsymbol{\theta}_i)\, p(\boldsymbol{\theta}_i)\, / p(\boldsymbol{\theta}_i) = l(\mathbf{y}|\boldsymbol{\theta}_i)$$

and then

$$q_i = \frac{w_i}{\sum_{i=1}^{m} w_i} = \frac{l(\mathbf{y}|\boldsymbol{\theta}_i)}{\sum_{i=1}^{m} l(\mathbf{y}|\boldsymbol{\theta}_i)}.$$

Draw $\boldsymbol{\theta}^*$ from the points $\{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_m\}$ with probability q_i, $i = 1, ..., m$. Then, $\boldsymbol{\theta}^*$ are approximately distributed according to $p(\boldsymbol{\theta}|\mathbf{y})$. See, for details, Smith and Gelfand (1992).

1.2.4 Markov Chain Monte Carlo methods

Markov Chain Monte Carlo methodology is the main computational tool in Bayesian Statistics, because it provides an idealized way to extract any posterior summary of interest such as functions of parameters. The idea of MCMC was first introduced by Metropolis *et al.* (1953) and was extended by Hastings (1970) for statistical problems. The basic idea of MCMC is similar of Monte Carlo but instead of drawing samplers from the posterior distribution via an importance sampling density, MCMC draws samples by running a cleverly constructed Markov chain for a long time.

The general formulation is as follows: suppose that, for a given parameter vector $\boldsymbol{\theta} \in \mathbf{R}^n$ and data $\mathbf{y}$, we want to generate a sample from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$, known up to a constant of proportionality. The idea is based on the construction of an irreducible and aperiodic Markov chain, which is easily simulated, with realizations $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, ..., \boldsymbol{\theta}^{(t)}, ...$ in the parameter space, equilibrium distribution $p(\boldsymbol{\theta}|\mathbf{y})$ and a transi-

tion probability $K\left(\boldsymbol{\theta}^{''},\boldsymbol{\theta}^{'}\right)=p\left(\boldsymbol{\theta}^{(t+1)}=\boldsymbol{\theta}^{''}|\boldsymbol{\theta}^{(t)}=\boldsymbol{\theta}^{'}\right)$, where $\boldsymbol{\theta}^{'}$ and $\boldsymbol{\theta}^{''}$ are the realized states at time t and $t+1$ respectively. Under appropriate regularity conditions, asymptotic results guarantee that as $t\rightarrow\infty$, $\boldsymbol{\theta}^{(t)}$ tends in distribution to a random variable with density $p\left(\boldsymbol{\theta}|\mathbf{y}\right)$, and the ergodic average of an integrable function of $\boldsymbol{\theta}$ is a consistent estimator of the (posterior) mean of the function. That is,

$$\boldsymbol{\theta}^{(t)}\rightarrow\boldsymbol{\theta}\sim p\left(\boldsymbol{\theta}|\mathbf{y}\right),\text{ in distribution}$$

and

$$\frac{1}{t}\sum_{i=1}^{t}f\left(\boldsymbol{\theta}^{(i)}\right)\overset{t\rightarrow\infty}{\rightarrow}E_{\boldsymbol{\theta}|\mathbf{y}}\left[f\left(\boldsymbol{\theta}\right)\right]\text{ almost surely.}$$

For more details about theory and application of MCMC methods, see, for example, Smith and Roberts (1993), Besag, Green, Higdon and Mengersen (1995), Gilks, Richardson and Spiegelhalter (1996).

In the MCMC methods the target distribution (the distribution from which we want to generate a sample) is known, and the transition probability or transition kernel K is unknown. There are many possible choices for the transition kernel K, each leading to different sampling schemes. There are two main ideas used for the construction of sampling schemes. The first idea proceeds by splitting the parameter space into a number of components and updating each in turn by using *conditional* distributions. It is the basis for the Gibbs Sampler. The second idea proceeds by using *proposal* distributions for the components of the parameter space and is the basis for the Metropolis-Hastings algorithm. A combination of these two sampling schemes is called Metropolis-within-Gibbs. A detailed description of these sampling schemes follows.

The Gibbs Sampler

The Gibbs Sampler was introduced by Geman and Geman (1984). Investigation of the applicability of the Gibbs sampler approach to Bayesian statistical problems was given by Gelfand and Smith (1990). The Gibbs sampler is a technique for generating samples from

the joint posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$, via iterated sampling from the full conditional distributions $p\left(\theta_i|\boldsymbol{\theta}_{\setminus i}, \mathbf{y}\right)$, $i = 1, 2, ..., n$ (i.e. the distribution of each individual element of $\boldsymbol{\theta}$ conditional on specified values of the data $\mathbf{y}$ and all the other elements of $\boldsymbol{\theta}$). It is an iterative sampling scheme that proceeds as follows: given arbitrary starting values $\boldsymbol{\theta}^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_n^{(0)}\right)$ for the parameters, draw

$$\begin{aligned}
&\theta_1^{(1)} \text{ from } p\left(\theta_1|\theta_2^{(0)}, \theta_3^{(0)}, ..., \theta_n^{(0)}, \mathbf{y}\right) \\
&\theta_2^{(1)} \text{ from } p\left(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, ..., \theta_n^{(0)}, \mathbf{y}\right) \\
&\theta_3^{(1)} \text{ from } p\left(\theta_3|\theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)} ..., \theta_n^{(0)}, \mathbf{y}\right) \\
&\vdots \\
&\theta_n^{(1)} \text{ from } p\left(\theta_n|\theta_1^{(1)}, \theta_2^{(1)}, ..., \theta_{n-1}^{(1)}, \mathbf{y}\right).
\end{aligned}$$

This completes one iteration from $\boldsymbol{\theta}^{(0)}$ to $\boldsymbol{\theta}^{(1)}$. After t iterations the chain will be at $\boldsymbol{\theta}^{(t)} = \left(\theta_1^{(t)}, \theta_2^{(t)}, ..., \theta_n^{(t)}\right)$. This sampling scheme produces a sequence $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\theta}^{(1)}$,..., $\boldsymbol{\theta}^{(t)}$,... which is a realization of a Markov chain with transition probability from $\boldsymbol{\theta}^{(t)}$ to $\boldsymbol{\theta}^{(t+1)}$ given by

$$K\left(\boldsymbol{\theta}^{(t+1)}, \boldsymbol{\theta}^{(t)}\right) = \prod_{i=1}^{n} p\left(\theta_i^{(t+1)}|\theta_j^{(t)}, j > i,\ \theta_j^{(t+1)}, j < i,\ \mathbf{y}\right).$$

If we replicate independently m times the above sampling scheme, we will take m replicates of the sampled vector $\boldsymbol{\theta}^{(t)}$. Then as $t \to \infty$, the replicates $\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}, ..., \boldsymbol{\theta}_m^{(t)}$ are approximately a random sample from the posterior $p(\boldsymbol{\theta}|\mathbf{y})$. For more details and applications about Gibbs sampling, see, for example, Gelfand and Smith (1990), Gelfand, Hills, Racine-Poon and Smith (1990), Casella and George (1992), Besag and Green (1993), Dellaportas and Smith (1993), Gilks, Thomas and Spiegelhalter (1994), Spiegelhalter, Thomas, Best and Gilks (1995a, 1995b) among several others.

The Metropolis-Hastings algorithm

The Metropolis-Hastings (M-H) algorithm was introduced by Metropolis *et al.* (1953) and extended by Hastings (1970). It is a clever method that can be used by itself or in conjuction with the Gibbs sampler to simulate intractable full conditional distributions. That is, when the full conditional distributions can be specified and are of known form the Gibbs sampler is used. In the cases where the full conditional distributions are not of known form (non standard distributions) or the full conditional distributions are complex enough, we use the Metropolis-Hastings algorithm.

Suppose that the parameter vector of interest is $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)$ and we want to generate a sample from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ by using the n full conditional distributions $p\left(\theta_i|\boldsymbol{\theta}_{\setminus i}, \mathbf{y}\right)$, $i = 1, 2, ..., n$, which are not of known form (non standard distributions). The update of the parameter vector $\boldsymbol{\theta}^{(t)} = \left(\theta_1^{(t)}, \theta_2^{(t)}, ..., \theta_n^{(t)}\right)$ at time t, to $\boldsymbol{\theta}^{(t+1)} = \left(\theta_1^{(t+1)}, \theta_2^{(t+1)}, ..., \theta_n^{(t+1)}\right)$ at time $t+1$, is done by updating all the components $\theta_i^{(t)}$, $i = 1, ..., n$ one by one. We refer to it as a *single component update* M-H. That is, an iteration of the single component update M-H is consisted of n updating steps. We describe one updating step of the algorithm and the other steps are done with an analogous way. For step i of iteration $t+1$, $\theta_i^{(t)}$ is updated to take $\theta_i^{(t+1)}$ using the Metropolis-Hastings algorithm as follows:

- a candidate $\theta_i^{'}$ is generated from a proposal distribution $q_i\left(\theta_i^{(t)}, \theta_i^{'}\right)$. Thus $q_i(.,.)$ generates a candidate only for the $i-th$ component of the parameter vector $\boldsymbol{\theta}$.

- calculate the probability of acceptance

$$a\left(\theta_i^{(t)}, \theta_i^{'}\right) = \min\left\{1, \frac{p\left(\theta_i^{'}|\boldsymbol{\theta}_{\setminus i}, \mathbf{y}\right) q_i\left(\theta_i^{'}, \theta_i^{(t)}\right)}{p\left(\theta_i^{(t)}|\boldsymbol{\theta}_{\setminus i}, \mathbf{y}\right) q_i\left(\theta_i^{(t)}, \theta_i^{'}\right)}\right\},$$

  where $p\left(\theta_i^{'}|\boldsymbol{\theta}_{\setminus i}, \mathbf{y}\right)$ is the full conditional distribution of the i component of $\boldsymbol{\theta}$ evaluated at $\theta_i^{'}$, $\boldsymbol{\theta}_{\setminus i} = \left(\theta_1^{(t+1)}, \theta_2^{(t+1)}, ..., \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, ..., \theta_n^{(t)}\right)$.

- update $\theta_i^{(t+1)} = \theta^{'}$ with probability $a\left(\theta_i^{(t)}, \theta_i^{'}\right)$, otherwise $\theta_i^{(t+1)} = \theta_i^{(t)}$.

Gibbs sampler is a special case of the single component update M-H algorithm since, when the proposal density $q_i\left(\theta_i^{(t)}, \theta_i^{'}\right)$ equals to the full conditional distribution of the $i-th$ component $p\left(\theta_i^{'}|\boldsymbol{\theta}_{\setminus i}, \mathbf{y}\right)$, the probability of acceptance $a\left(\theta_i^{(t)}, \theta_i^{'}\right)$ is one and therefore we always accept the proposed move for the $i-th$ component.

Another alternative is to use a *simultaneous component update* M-H algorithm, where all the components of the parameter vector $\boldsymbol{\theta}$ updated simultaneously, by using a multivariate proposal distribution. In this case, the update of the parameter vector $\boldsymbol{\theta}^{(t)} = \left(\theta_1^{(t)}, \theta_2^{(t)}, ..., \theta_n^{(t)}\right)$ at time t, to $\boldsymbol{\theta}^{(t+1)} = \left(\theta_1^{(t+1)}, \theta_2^{(t+1)}, ..., \theta_n^{(t+1)}\right)$ at time $t+1$, is done by using the following step:

- a candidate parameter vector $\boldsymbol{\theta}^{'}$ is generated from a multivariate proposal distribution $q\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right)$.
- calculate the probability of acceptance

$$a\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right) = \min\left\{1, \frac{p\left(\boldsymbol{\theta}^{'}|\mathbf{y}\right) q\left(\boldsymbol{\theta}^{'}, \boldsymbol{\theta}^{(t)}\right)}{p\left(\boldsymbol{\theta}^{(t)}|\mathbf{y}\right) q\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right)}\right\}, \tag{1.3}$$

  where $p\left(\boldsymbol{\theta}^{'}|\mathbf{y}\right)$ is the posterior distribution evaluated at $\boldsymbol{\theta}^{'}$.

- update $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{'}$ with probability $a\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right)$, otherwise $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$.

The proposal distribution $q\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right)$ can have any form and the stationary distribution of the chain will be $p\left(\boldsymbol{\theta}|\mathbf{y}\right)$. The transition kernel for this Metropolis-Hastings algorithm is given by

$$K\left(\boldsymbol{\theta}^{(t+1)}, \boldsymbol{\theta}^{(t)}\right) = q\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right) a\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right) + \left[1 - \int q\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right) a\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right) \partial\boldsymbol{\theta}^{'}\right] I\left(\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}\right),$$

where $I\left(.\right)$ denotes the indicator function, taking the value of 1 when the argument is true and 0 otherwise. Note that, the calculation of $a\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right)$ does not require knowledge

of the normalizing constant of $p(\boldsymbol{\theta}|\mathbf{y})$ because it appears both in the numerator and denominator.

An important implementation issue about Metropolis-Hastings algorithm is the choice of the proposal distribution $q(.,.)$. In the case where the candidate generating density is *symmetric*, i.e. $q\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right) = q\left(\boldsymbol{\theta}^{'}, \boldsymbol{\theta}^{(t)}\right)$, then the probability of acceptance (1.3) reduces to

$$a\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right) = \min\left\{1, \frac{p\left(\boldsymbol{\theta}^{'}|\mathbf{y}\right)}{p\left(\boldsymbol{\theta}^{(t)}|\mathbf{y}\right)}\right\}.$$

That is, if $p\left(\boldsymbol{\theta}^{'}|\mathbf{y}\right) \geqslant p\left(\boldsymbol{\theta}^{(t)}|\mathbf{y}\right)$, the chain moves to $\boldsymbol{\theta}^{'}$, otherwise, it moves with probability $p\left(\boldsymbol{\theta}^{'}|\mathbf{y}\right)/p\left(\boldsymbol{\theta}^{(t)}|\mathbf{y}\right)$.

If the candidate $\boldsymbol{\theta}^{'}$ is drawn according to $\boldsymbol{\theta}^{'} = \boldsymbol{\theta}^{(t)} + \mathbf{z}$, where $\mathbf{z} \sim f$, the algorithm is called *random walk* M-H. Possible choices for f are the multivariate normal or multivariate-t distributions. In this case, it is that $q\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right) = f\left(\boldsymbol{\theta}^{'} - \boldsymbol{\theta}^{(t)}\right)$.

Another kind of candidate generating densities has the form $q\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right) = f\left(\boldsymbol{\theta}^{'}\right)$. That is, the candidate values $\boldsymbol{\theta}^{'}$ are drawn independently of the current values $\boldsymbol{\theta}^{(t)}$. It is the *independence* M-H algorithm, and the probability of acceptance is given by

$$a\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{'}\right) = \min\left\{1, \frac{w\left(\boldsymbol{\theta}^{'}\right)}{w\left(\boldsymbol{\theta}^{(t)}\right)}\right\},$$

where $w\left(\boldsymbol{\theta}^{'}\right) = p\left(\boldsymbol{\theta}^{'}|\mathbf{y}\right)/f\left(\boldsymbol{\theta}^{'}\right)$ is the importance weight function that could be used in importance sampling if observations generated by f. More details about theory and applications of the Metropolis-Hastings algorithm are given by Smith and Roberts (1993), Tierney (1994), Chib and Greenberg (1995), Gilks, Richardson and Spiegelhalter (1996).

Metropolis-within-Gibbs

It is common, in practice, to use a combination of different strategies (combination of different transition kernels) to improve convergence of the MCMC algorithm. One simple

example is the combination of Metropolis-Hastings algorithm with the Gibbs sampler. Components of the parameter vector whose full conditional distributions are of known form are updated directly by using Gibbs sampling, while those with non standard conditional distributions are updated by using Metropolis-Hastings steps. This scheme is known as Metropolis-within-Gibbs.

1.2.5 Convergence of the MCMC simulation

The main problem that turns up is to gauge when convergence is achieved; that is, to assess at what point the chain gets in the target distribution, and to figure out how many points will have to be taken from this distribution in order to estimate, with the desired accuracy, the parameters of interest. The first part of the problem is that of determining the length of the required burn-in period or the point in which the Markov chain has 'forgotten' its starting point; parameter values within an initial transient phase are discarded, in order to reduce the bias caused by the effect of starting values. The second part is that of determining the number of sample points which adequately represent the posterior distribution and on which inference will be based.

To solve the problem of convergence, a number of different approaches have been proposed in the literature. From a theoretical point of view, there is an attempt to predetermine the number of iterations that will ensure convergence; see, for example, Schervish and Carlin (1992), Rosenthal (1993), Polson (1996). Due to the difficulty of this approach in practice, almost all of the applied works are based on the output produced by running Markov Chain Monte Carlo algorithms. Such techniques are known as convergence diagnostics and described in the review papers of Cowles and Carlin (1996), Robert and Mengersen (1998), and of Brooks and Roberts (1999). Some of the proposed methods try to assess convergence, while others attempt to 'measure' the performance of any particular sampler. The choice of diagnostic depends on the problem at hand. The theoretical background, the assumptions, the assessment of convergence of the joint or the marginal density of an MCMC output, the number of chains that are

needed, the range of the samplers to which the diagnostic is applied, the computational expense and the interpretability are some of the criteria for the choice of the diagnostic. Some well known convergence diagnostics are that of Gelman and Rubin (1992), Raftery and Lewis (1992), Geweke (1992) and of Heidelberger and Welch (1983) because they are easily implemented.

Cowles and Carlin (1996) proposed to use a combination of strategies aimed at evaluating and accelerating the convergence of the MCMC algorithms, including applying diagnostic procedures to a small number of parallel chains, monitoring autocorrelations and cross-correlations, and modifying parameterizations or sampling algorithms appropriately. Implementation issues such as the reparameterization of the target distribution, the introduction of auxiliary variables and the use of alternative updating schemes can improve convergence. A comprehensive review of MCMC methods and other related implementation issues is given by Brooks (1998) and a discussion of these issues is provided by Kass, Carlin, Gelman and Neal (1998). Although it is never possible to say with certainty that an MCMC output converges to its stationary distribution, convergence diagnostics provide a helpful tool in exploring the algorithm's performance.

1.2.6 Alternative Updating Schemes

Alternative updating schemes have been proposed to overcome problems associated with slow mixing Markov chains.

Metropolis-coupled Markov Chain Monte Carlo and Simulated Tempering

Metropolis-coupled MCMC (Geyer, 1991) and Simulated tempering (Marinari and Parisi, 1992, Geyer and Thompson, 1995) are MCMC methods which simulate a family of m distributions specified by the unnormalized densities $h_i(\boldsymbol{\theta})$, $i = 1, ..., m$, on the same parameter space. The index i is called 'temperature', ranging from the 'cold' distribution $h_1(\boldsymbol{\theta})$, which is usually the distribution of interest, to the 'hot' distribution $h_m(\boldsymbol{\theta})$, which is easily simulated. Geyer and Thompson (1995) suggested taking $h_i(\boldsymbol{\theta}) = [h(\boldsymbol{\theta})]^{1/i}$,

$i = 1, ..., m$, where h_∞ correspond to a uniform distribution over the entire parameter space. $h(\boldsymbol{\theta})$ can be the posterior distribution of interest $p(\boldsymbol{\theta}|\mathbf{y})$.

Simulated tempering works as follows. Suppose that for each i, $i = 1, ..., m$, there is a method (Metropolis-Hasting algorithm or Gibbs sampling) for updating $\boldsymbol{\theta}$ that has $h_i(\boldsymbol{\theta})$ as a stationary distribution. The state of simulated tempering is the pair $(\boldsymbol{\theta}, i)$, where $\boldsymbol{\theta}$ takes values in the common parameter space for all $h_i(\boldsymbol{\theta})$ and i is the temperature. One iteration of the simulated tempering algorithm is given by the following steps:

- starting from temperature i, update $\boldsymbol{\theta}$ using Metropolis-Hasting algorithm or Gibbs sampling.
- propose a move from temperature i to temperature j $(j = \pm 1)$ according to probabilities $q_{i,j}$, where $q_{1,2} = q_{m,m-1} = 1$ and $q_{i,i+1} = q_{i,i-1} = \frac{1}{2}$ if $1 < i < m$.
- accept the transition from i to j with probability

$$a(i,j) = \min\left\{1, \frac{h_j(\boldsymbol{\theta})\,\pi(j)\,q_{j,i}}{h_i(\boldsymbol{\theta})\,\pi(i)\,q_{i,j}}\right\}, \tag{1.4}$$

  where $\pi(i)$ and $\pi(j)$ are appropriate normalization constants for the densities h_i and h_j, respectively.

From the resulting Markov chain, we base our inference on all observations which were drawn from the 'cold' distribution, i.e. from the density h_1. The advantage of this method is that improves the mixing of the chain by using the 'hot' distributions. These distributions mix more rapidly than the 'cold' distribution. The main drawback of this method is the estimation of the normalization constants $\pi(i)$, $i = 1, ..., m$, appearing in equation (1.4).

Metropolis-coupled MCMC is based on a similar idea, but instead of running one chain we run m chains $\{\boldsymbol{\theta}_i\}$, $i = 1, ..., m$, where updates in chain i are based on h_i. One iteration of Metropolis-coupled MCMC algorithm proceeds as follows:

- update each chain i, $i = 1, ..., m$, using Metropolis-Hasting algorithm or Gibbs sampling.

- select two chains and attempt to swap the states of these chains. Suppose, that, at time t, we select chains i, j and propose the swap $\boldsymbol{\theta}_i^{'} = \boldsymbol{\theta}_j^{(t)}$ and $\boldsymbol{\theta}_j^{'} = \boldsymbol{\theta}_i^{(t)}$.

- accept the proposed swap with probability

$$\min\left\{1, \frac{h_i\left(\boldsymbol{\theta}_i^{'}\right) h_j\left(\boldsymbol{\theta}_j^{'}\right)}{h_i\left(\boldsymbol{\theta}_i^{(t)}\right) h_j\left(\boldsymbol{\theta}_j^{(t)}\right)}\right\}.$$

We base our inference only on the sample taken from the chain with stationary distribution h_1, that is the 'cold' distribution. The advantage of this algorithm is that the mixing of the 'cold' chain is improved by allowing the chains to swap states. The disadvantage is that we run m chains while we actually need only one for inference.

Delayed Rejection Algorithm

The idea of *Delayed rejection algorithm* was proposed by Tierney and Mira (1999). This strategy improves the Metropolis-Hastings algorithm in the Peskun sense (Peskun, 1973) that the resulting estimates have smaller asymptotic variance on a sweep by sweep basis.

When a Markov chain remains at the same state over successive iterations, the autocorrelation of the realized chain increases and thus the variance of the estimates. In a Metropolis-Hastings algorithm this happens when the proposed parameter value is rejected. Tierney and Mira (1999) improved the Metropolis-Hastings algorithm by reducing the probability of remaining at the current state using the following idea: when a candidate value is rejected, instead of staying at the current value, propose a new candidate from a different proposal distribution, and accept or reject that candidate using an adjusted acceptance probability in order to preserve the stationary distribution.

Suppose that we want to generate a sample from the posterior $p(\boldsymbol{\theta}|\mathbf{y})$ and that the chain at time t is at $\boldsymbol{\theta}^{(t)}$. According to the delayed rejection algorithm a candidate $\boldsymbol{\theta}_1^{'}$ is

generated from a proposal distribution $q_1\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}\right)$ and accepted with probability

$$a_1\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}\right) = \min\left\{1, \frac{p\left(\boldsymbol{\theta}_1^{'}|\mathbf{y}\right) q_1\left(\boldsymbol{\theta}_1^{'}, \boldsymbol{\theta}^{(t)}\right)}{p\left(\boldsymbol{\theta}^{(t)}|\mathbf{y}\right) q_1\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}\right)}\right\}.$$

If the candidate $\boldsymbol{\theta}_1^{'}$ is rejected, a new candidate $\boldsymbol{\theta}_2^{'}$ is generated from a new proposal $q_2\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}, \boldsymbol{\theta}_2^{'}\right)$ at the second stage. Tierney and Mira (1999) derived the probability of acceptance for this candidate by imposing detailed balance at the second stage in order to preserve the stationary distribution. The probability of acceptance is given by

$$a_2\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}, \boldsymbol{\theta}_2^{'}\right) = \min\left\{1, \frac{p\left(\boldsymbol{\theta}_2^{'}|\mathbf{y}\right) q_1\left(\boldsymbol{\theta}_2^{'}, \boldsymbol{\theta}_1^{'}\right) q_2\left(\boldsymbol{\theta}_2^{'}, \boldsymbol{\theta}_1^{'}, \boldsymbol{\theta}^{(t)}\right)\left[1 - a_1\left(\boldsymbol{\theta}_2^{'}, \boldsymbol{\theta}_1^{'}\right)\right]}{p\left(\boldsymbol{\theta}^{(t)}|\mathbf{y}\right) q_1\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}\right) q_2\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}, \boldsymbol{\theta}_2^{'}\right)\left[1 - a_1\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}\right)\right]}\right\}.$$

If the candidate $\boldsymbol{\theta}_2^{'}$ is rejected the algorithm can either stay at the current state $\boldsymbol{\theta}^{(t)}$, or move on the third stage, and so on. In order to maintain the stationary distribution, the probability of acceptance of new candidates is calculated by imposing detailed balance separately at each stage. The probability of acceptance for the candidate at stage i is presented by Mira (1999).

The *symmetric delayed rejection algorithm* is a special case of the above algorithm, where the proposal distribution is symmetric and depends only on the last rejected candidate. In this case the probability of acceptance at the first stage is

$$a_1\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}\right) = \min\left\{1, \frac{p\left(\boldsymbol{\theta}_1^{'}|\mathbf{y}\right)}{p\left(\boldsymbol{\theta}^{(t)}|\mathbf{y}\right)}\right\}$$

as in the regular Metropolis algorithm. If $\boldsymbol{\theta}_1^{'}$ is rejected, a new candidate $\boldsymbol{\theta}_2^{'}$ is generated

from $q_2\left(\boldsymbol{\theta}_1^{'}, \boldsymbol{\theta}_2^{'}\right)$ and accept it with probability

$$a_2\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}, \boldsymbol{\theta}_2^{'}\right) = \min\left\{1, \frac{\max\left\{0, \left[p\left(\boldsymbol{\theta}_2^{'}|\mathbf{y}\right) - p\left(\boldsymbol{\theta}_1^{'}|\mathbf{y}\right)\right]\right\}}{p\left(\boldsymbol{\theta}^{(t)}|\mathbf{y}\right) - p\left(\boldsymbol{\theta}_1^{'}|\mathbf{y}\right)}\right\}.$$

Generally, the $i-th$ stage of the symmetric delayed algorithm works as follows: if the $\boldsymbol{\theta}_{i-1}^{'}$ is rejected, generate a candidate $\boldsymbol{\theta}_i^{'}$ from $q_i\left(\boldsymbol{\theta}_{i-1}^{'}, \boldsymbol{\theta}_i^{'}\right)$, and accept it with probability

$$a_i\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}_1^{'}, ..., \boldsymbol{\theta}_i^{'}\right) = \min\left\{1, \frac{\max\left\{0, \left[p\left(\boldsymbol{\theta}_i^{'}|\mathbf{y}\right) - p\left(\boldsymbol{\theta}_*^{'}|\mathbf{y}\right)\right]\right\}}{p\left(\boldsymbol{\theta}^{(t)}|\mathbf{y}\right) - p\left(\boldsymbol{\theta}_*^{'}|\mathbf{y}\right)}\right\},$$

where $p\left(\boldsymbol{\theta}_*^{'}|\mathbf{y}\right) = \arg\max_{j<i} p\left(\boldsymbol{\theta}_j^{'}|\mathbf{y}\right)$. For more details about delayed rejection algorithm, see, Mira (1999).

1.3 Model Selection Strategies

1.3.1 Bayesian Model Comparison

Assume that we have a countable set M of competing models for a given set of data $\mathbf{y} = (y_1, y_2, ..., y_T)$. Let model $m \in M$ have a vector $\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m$ of unknown parameters, the dimension of which may vary from model to model. The posterior probability of model m is given by

$$p(m|\mathbf{y}) = \frac{p(m)\, p(\mathbf{y}|m)}{\sum_{m \in M} p(m)\, p(\mathbf{y}|m)}, \tag{1.5}$$

where $p(m)$ is the prior probability for model m, and $p(\mathbf{y}|m)$ is the marginal likelihood of the data for model m. According to (1.5) for models m_i and m_j we have that

$$\frac{p(m_i|\mathbf{y})}{p(m_j|\mathbf{y})} = \frac{p(m_i)\, p(\mathbf{y}|m_i)}{p(m_j)\, p(\mathbf{y}|m_j)}.$$

That is, the ratio of prior probabilities transformed to the ratio of posterior probabilities through consideration of the data.

Inference about the model selection problem may be done using the Bayes Factor (BF) of model m_i against model m_j given by

$$BF = \frac{p\left(\mathbf{y}|m_i\right)}{p\left(\mathbf{y}|m_j\right)} = \frac{p\left(m_i|\mathbf{y}\right)}{p\left(m_j|\mathbf{y}\right)} \frac{p\left(m_j\right)}{p\left(m_i\right)}. \tag{1.6}$$

The density $p\left(\mathbf{y}|m_i\right)$ is obtained by integrating over the parameter space, so that

$$p\left(\mathbf{y}|m_i\right) = \int\limits_{\boldsymbol{\Theta}_{m_i}} p\left(\mathbf{y}|\boldsymbol{\theta}_{m_i}, m_i\right) p\left(\boldsymbol{\theta}_{m_i}|m_i\right) d\boldsymbol{\theta}_{m_i}, \tag{1.7}$$

where $\boldsymbol{\theta}_{m_i}$ is the parameter vector for model m_i, $p\left(\mathbf{y}|\boldsymbol{\theta}_{m_i}, m_i\right)$ is the likelihood given the model m_i and the parameter vector $\boldsymbol{\theta}_{m_i}$, $p\left(\boldsymbol{\theta}_{m_i}|m_i\right)$ is the prior for the parameter vector $\boldsymbol{\theta}_{m_i}$ given the model m_i. Therefore, the Bayes Factor of model m_i against model m_j (1.6) can be written, using (1.7) as

$$BF = \frac{\int\limits_{\boldsymbol{\Theta}_{m_i}} p\left(\mathbf{y}|\boldsymbol{\theta}_{m_i}, m_i\right) p\left(\boldsymbol{\theta}_{m_i}|m_i\right) d\boldsymbol{\theta}_{m_i}}{\int\limits_{\boldsymbol{\Theta}_{m_j}} p\left(\mathbf{y}|\boldsymbol{\theta}_{m_j}, m_j\right) p\left(\boldsymbol{\theta}_{m_j}|m_j\right) d\boldsymbol{\theta}_{m_j}}. \tag{1.8}$$

Note that, all constants appearing in the likelihoods $p\left(\mathbf{y}|\boldsymbol{\theta}_{m_i}, m_i\right)$ and $p\left(\mathbf{y}|\boldsymbol{\theta}_{m_j}, m_j\right)$ must be retained when computing Bayes Factor in equation (1.8).

According to Kass and Raftery (1995), "the Bayes Factor is a summary of the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another". Possible interpretations of Bayes Factor are given by Tables 1.1 and 1.2 provided by Kass and Raftery (1995). The logarithm of the marginal likelihood of the data may also be viewed as a predictive score. In this case, the interpretation of the Bayes Factor does not depend on viewing one of the models as 'true'.

$log_{10}(B_{ij})$	B_{ij}	Evidence against m_j
0 to 0.5	1 to 3.2	Not worth more than a bare mention
0.5 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
>2	>100	Decisive

Table 1.1: Bayes Factor interpretation (log_{10} scale).

$2log_e(B_{ij})$	B_{ij}	Evidence against m_j
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very strong

Table 1.2: Bayes Factor interpretation (natural logarithm).

1.3.2 Calculation of the Bayes Factor

Bayes factor requires evaluation of the integrals in the numerator and denominator of (1.8) which are the marginal densities $p(\mathbf{y}|m_i)$ and $p(\mathbf{y}|m_j)$. These integrals are in general difficult to calculate; Kass and Raftery (1995) provide an extensive description and comparison of available numerical strategies. Here, we describe Laplace's method (see, for example, Tierney and Kadane, 1986, Tierney, Kass and Kadane, 1989a, Erkanli, 1994, Raftery, 1996) and variants of Laplace's method, which are used in the multivariate latent GARCH model we introduce in chapter 6.

The Laplace's method of approximation of (1.7) is given by

$$\widehat{p(\mathbf{y}|m_i)} = (2\pi)^{d_{m_i}/2} |\widetilde{\Sigma}|^{1/2} p\left(\mathbf{y}|\widetilde{\boldsymbol{\theta}}_{m_i}, m_i\right) p\left(\widetilde{\boldsymbol{\theta}}_{m_i}|m_i\right), \tag{1.9}$$

where d_{m_i} is the dimension of the parameter vector $\boldsymbol{\theta}_{m_i}$ of model m_i, $p\left(\mathbf{y}|\widetilde{\boldsymbol{\theta}}_{m_i}, m_i\right)$ and $p\left(\widetilde{\boldsymbol{\theta}}_{m_i}|m_i\right)$ are the likelihood and the prior, respectively, evaluated at the posterior mode $\widetilde{\boldsymbol{\theta}}_{m_i}$, $\widetilde{\Sigma} = \left[-D^2\widetilde{l}(\boldsymbol{\theta}_{m_i})\right]^{-1}$, $D^2\widetilde{l}(\boldsymbol{\theta}_{m_i})$ is the Hessian matrix of second derivatives of $\widetilde{l}(\boldsymbol{\theta}_{m_i})$, where $\widetilde{l}(\boldsymbol{\theta}_{m_i}) = \log\left[p\left(\mathbf{y}|\boldsymbol{\theta}_{m_i}, m_i\right) p(\boldsymbol{\theta}_{m_i}|m_i)\right]$. According to Kass and Raftery (1995), when Laplace's method is applied to both the numerator and denominator of

(1.8), the resulting approximation has relative error of order $O\left(T^{-1}\right)$.

An important variant of (1.9) is

$$\widehat{p\left(\mathbf{y}|m_i\right)}_{mle} = (2\pi)^{d_{m_i}/2} \left|\widehat{\Sigma}\right|^{1/2} p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_{m_i}, m_i\right) p\left(\widehat{\boldsymbol{\theta}}_{m_i}|m_i\right), \tag{1.10}$$

where $\widehat{\Sigma}$ is the inverse of the negative Hessian matrix of the log-likelihood evaluated at the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{m_i}$, $p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_{m_i}, m_i\right)$ and $p\left(\widehat{\boldsymbol{\theta}}_{m_i}|m_i\right)$ are the likelihood and the prior, respectively, evaluated at $\widehat{\boldsymbol{\theta}}_{m_i}$. This approximation is easily computed from any statistical package that reports the maximum likelihood estimator, the observed information matrix, and the value of the maximized likelihood. The relative error of this approximation is again of order $O\left(T^{-1}\right)$. Another alternative is to use the inverse of the expected information matrix in place of $\widehat{\Sigma}$ in equation (1.10). The resulting approximation has a larger asymptotic relative error of order $O\left(T^{-1/2}\right)$, but (see, Kass and Raftery, 1995) it remains sufficiently accurate to be of use in many problems.

Other methods can be used for the calculation of Bayes Factor such as simple Monte Carlo, Importance Sampling and Gaussian Quadrature, and methods for simulating from the posterior; see, for example, Kass and Raftery (1995), DiCiccio, Kass, Raftery and Wasserman (1997).

1.3.3 Bayesian Model Averaging

Assume that we have a countable set M of competing models m_1, m_2, ..., m_M for a given set of data $\mathbf{y}$. Each model is compared in turn with m_1, yielding Bayes factors B_{21}, B_{31}, ..., B_{M1}. Then, Bayes factors provide the posterior model probabilities. The posterior probability of model m_i is

$$p\left(m_i|\mathbf{y}\right) = \frac{a_i B_{i1}}{\sum_{j=1}^{M} a_j B_{j1}}, \tag{1.11}$$

where $a_i = p\left(m_i\right)/p\left(m_1\right)$ is the prior odds for model m_i against m_1, $i = 1, ..., M$, and $B_{11} = 1$.

Having been able to calculate the posterior probabilities of each model, it seems natural to account for model uncertainty in our predictive inferences. Rather than choosing a single "best" model and then making inferences as if the selected model was the true model, we can use the following model averaging approach, which provides composite predictions. Suppose that we are interested in a quantity Δ. For example, in time varying volatility models we analyze in the following sections, this quantity may be the volatility at a future time period. Then, its posterior distribution given data $\mathbf{y}$ is given by

$$p\left(\Delta|\mathbf{y}\right) = \sum_{i=1}^{M} p\left(\Delta|m_i,\mathbf{y}\right) p\left(m_i|\mathbf{y}\right), \tag{1.12}$$

which is an average of the posterior predictive distribution under each model weighted by their posterior model probabilities. Note that, the posterior predictive distribution of Δ given a particular model m_i is found by integrating out the model parameters $\boldsymbol{\theta}_{m_i}$:

$$p\left(\Delta|m_i,\mathbf{y}\right) = \int\limits_{\boldsymbol{\Theta}_{m_i}} p\left(\Delta|\boldsymbol{\theta}_{m_i},m_i,\mathbf{y}\right) p\left(\boldsymbol{\theta}_{m_i}|m_i,\mathbf{y}\right) d\boldsymbol{\theta}_{m_i}.$$

We can also use the maximum likelihood approximation:

$$p\left(\Delta|m_i,\mathbf{y}\right) \simeq p\left(\Delta|m_i,\mathbf{y},\widehat{\boldsymbol{\theta}}_{m_i}\right),$$

where $\widehat{\boldsymbol{\theta}}_{m_i}$ is the maximum likelihood estimator of the parameter vector $\boldsymbol{\theta}_{m_i}$ of model m_i. See, for example, Volinsky, Madigan, Raftery and Kronmal (1997). For discussion of the above approach as well as evidence that accounting for model uncertainty improves predictive performance see Kass and Raftery (1995), Draper (1995), Madigan, Gavrin and Raftery (1995), Raftery, Madigan and Hoeting (1997), Volinsky, Madigan, Raftery and Kronmal (1997) and Raftery, Madigan and Volinsky (1994).

The posterior mean and standard deviation of Δ are given by

$$E\left[\Delta|\mathbf{y}\right] = \sum_{i=1}^{M} E\left[\Delta|m_i, \mathbf{y}\right] p\left(m_i|\mathbf{y}\right)$$

and

$$V\left[\Delta|\mathbf{y}\right] = \sum_{i=1}^{M} \left(V\left[\Delta|m_i, \mathbf{y}\right] + \left(E\left[\Delta|m_i, \mathbf{y}\right]\right)^2\right) p\left(m_i|\mathbf{y}\right) - E\left[\Delta|\mathbf{y}\right]^2 .$$

Note that, equation (1.12) shows that choosing a single model and proceeding conditionally on it may be reasonable if one of the $p\left(m_i|\mathbf{y}\right)$ is close to unity or if the sum is dominated by models for which the values of $p\left(\Delta|m_i, \mathbf{y}\right)$ are similar. If this is not the case, the analysis will not take account for model uncertainty; see, for example, Kass and Raftery (1995).

1.3.4 MCMC model search methods

Implementation of Bayesian model averaging (BMA) is difficult for two reasons. First, the Bayes factor can be hard to compute, and second, the number of terms in equation (1.12) can be enormous. An ad hoc procedure for accounting for model uncertainty is the Occam's window algorithm of Madigan and Raftery (1994). We will not describe this algorithm; see, for example, Madigan and Raftery (1994), Raftery, Madigan and Hoeting (1997) among several others. In this section we describe some Markov chain Monte Carlo methods that provide posterior model probabilities and therefore can account for model uncertainty using Bayesian model averaging.

Markov chain Monte Carlo model Composition ($\mathbf{MC^3}$)

Markov chain Monte Carlo model Composition (MC^3) was introduced by Madigan and York (1995). MC^3 generates a stochastic process that moves through the class of models under consideration. They constructed a Markov chain $\left\{m\left(t\right),\ t = 1, 2, ...\right\}$ with state space the model space and equilibrium distribution $p\left(m_i|\mathbf{y}\right)$. If this Markov chain is

simulated for $t = 1, ..., N$, then under certain regularity conditions, for any function $f(m_i)$ defined on the model space, the average

$$\widehat{F} = \frac{1}{N} \sum_{i=1}^{N} f(m(i))$$

is a consistent estimate of the $E[f(m)]$; see, for example, Smith and Roberts (1993). To compute (1.12) in this way set $f(m) = p(\Delta|m, \mathbf{y})$.

To construct the Markov chain they defined a neighbourhood $nbd(m)$ for each model m. They also defined a transition matrix q by setting $q\left(m \to m^{'}\right) = 0$ for all $m^{'} \notin nbd(m)$ and $q\left(m \to m^{'}\right)$ constant for all $m^{'} \in nbd(m)$. If the current state of the chain is model m then $m^{'}$ is drawn from $q\left(m \to m^{'}\right)$ and accepted with probability

$$\min\left\{1, \frac{|nbd(m)|\, p\left(m^{'}|\mathbf{y}\right)}{|nbd(m^{'})|\, p(m|\mathbf{y})}\right\},$$

where $|nbd(m)|$ is the number of models that belong in the neighbourhood of model m. Otherwise, the chain stays in state m. Note that, if $|nbd(m)| = \left|nbd\left(m^{'}\right)\right|$ and all models are equally likely a priori then the probability of acceptance is given by

$$\min\left\{1, \frac{p\left(\mathbf{y}|m^{'}\right)}{p(\mathbf{y}|m)}\right\}.$$

This process is very flexible and runs of 10000 iterations or less are typically adequate. Model averaging by the Occam's window and MC^3 provide better predictions than using a single "best" model. The MC^3 method has better predictive performance than Occam's window, but is more computational expensive.

Reversible Jump MCMC

Green (1995) introduced a reversible jump MCMC strategy for generating from the joint posterior $p(m, \boldsymbol{\theta}_m|\mathbf{y})$, based on the standard Metropolis Hastings approach. During re-

versible jump MCMC sampling, the constructed Markov chain moves within and between models so that the limiting proportion of visits to a given model is the required $p(m|\mathbf{y})$ in (1.5). The reversible jump MCMC has been applied by Richardson and Green (1997) for an analysis of univariate normal mixtures, Nobile and Green (1997) for factorial experiments using mixture modelling, Troughton and Godsill (1997) for autoregressive time series, Dellaportas, Karlis and Xekalaki (1997) for an analysis of finite poisson mixtures, Dellaportas and Forster (1999) for analysis of contingency tables and Vrontos, Dellaportas and Politis (2000) for analysis of univariate GARCH and EGARCH models.

In general, suppose that the current state of the Markov chain at time t is $(m, \boldsymbol{\theta}_m)$, where $\boldsymbol{\theta}_m$ has dimension d_m, and a move is proposed at time $t+1$ to a new model $m^{'}$ with probability $j\left(m, m^{'}\right)$ and corresponding parameter vector $\boldsymbol{\theta}^{'}_{m^{'}}$ of dimension $d_{m^{'}}$. Then, a vector $\mathbf{u}$ is generated from a specified proposal density $q\left(\mathbf{u}|\boldsymbol{\theta}_m, m, m^{'}\right)$ and we set $\left(\boldsymbol{\theta}^{'}_{m^{'}}, \mathbf{u}^{'}\right) = g_{m,m^{'}}\left(\boldsymbol{\theta}_m, \mathbf{u}\right)$ for a specified invertible function $g_{m,m^{'}}$ such that $g_{m^{'},m} = g^{-1}_{m,m^{'}}$. Note that $d_m + d(\mathbf{u}) = d_{m^{'}} + d\left(\mathbf{u}^{'}\right)$. Green (1995) showed that if the new move is accepted as the next realization of the Markov chain with probability $a = \min\{1, r\}$ where

$$r = \frac{p\left(\mathbf{y}|m^{'}, \boldsymbol{\theta}^{'}_{m^{'}}\right) p\left(\boldsymbol{\theta}^{'}_{m^{'}}|m^{'}\right) p\left(m^{'}\right) j\left(m^{'}, m\right) q\left(\mathbf{u}^{'}|\boldsymbol{\theta}^{'}_{m^{'}}, m^{'}, m\right)}{p\left(\mathbf{y}|m, \boldsymbol{\theta}_m\right) p\left(\boldsymbol{\theta}_m|m\right) p(m)\, j\left(m, m^{'}\right) q\left(\mathbf{u}|\boldsymbol{\theta}_m, m, m^{'}\right)} |J|, \tag{1.13}$$

with $J = \partial\left(\boldsymbol{\theta}^{'}_{m^{'}}, \mathbf{u}^{'}\right) / \partial\left(\boldsymbol{\theta}_m, \mathbf{u}\right)$ denoting the Jacobian of the transformation, then the chain satisfies the condition of detailed balance and has the required limiting distribution $p(m, \boldsymbol{\theta}_m|\mathbf{y})$. The condition of detailed balance requires that the equilibrium probability of moving from a state $(m, \boldsymbol{\theta}_m)$ to $\left(m^{'}, \boldsymbol{\theta}^{'}_{m^{'}}\right)$ equals to that of moving from $\left(m^{'}, \boldsymbol{\theta}^{'}_{m^{'}}\right)$ to $(m, \boldsymbol{\theta}_m)$; see, for details, Green (1995). To implement the reversible jump MCMC we need to specify the probability $j\left(m, m^{'}\right)$ for every proposed move, the proposal distributions $q(\mathbf{u}|\boldsymbol{\theta}_m,\ m,\ m^{'})$, $q\left(\mathbf{u}^{'}|\boldsymbol{\theta}^{'}_{m^{'}}, m^{'}, m\right)$ and the function $g_{m,m^{'}}$. These choices do not affect the results but may be crucial for the convergence rate of the Markov chain.

Another strategy, which does not require a function g but requires proposal densities q of higher dimension is as follows. First, we suggest that all the parameters of the proposed

model are generated from a proposal distribution. Consequently, $\left(\boldsymbol{\theta}_{m'}^{'}, \mathbf{u}^{'}\right) = (\mathbf{u}, \boldsymbol{\theta}_{m})$ with $d_m = d\left(\mathbf{u}^{'}\right)$, $d_{m'} = d(\mathbf{u})$, $q(\mathbf{u}|\boldsymbol{\theta}_m,\, m,\, m^{'}) = q\left(\mathbf{u}|m^{'}\right)$, $q\left(\mathbf{u}^{'}|\boldsymbol{\theta}_{m'}^{'}, m^{'}, m\right) = q\left(\mathbf{u}^{'}|m\right)$ and the Jacobian term in (1.13) is one. In this case, the probability of acceptance of the new move as the next realization of the Markov chain is given by $a = \min\{1, r\}$, where

$$r = \frac{p\left(\mathbf{y}|m^{'}, \boldsymbol{\theta}_{m'}^{'}\right) p\left(\boldsymbol{\theta}_{m'}^{'}|m^{'}\right) p\left(m^{'}\right) j\left(m^{'}, m\right) q\left(\mathbf{u}^{'}|m\right)}{p\left(\mathbf{y}|m, \boldsymbol{\theta}_{m}\right) p\left(\boldsymbol{\theta}_{m}|m\right) p(m)\, j(m, m^{'})\, q\left(\mathbf{u}|m^{'}\right)}. \tag{1.14}$$

The proposal densities $q\left(\mathbf{u}|m^{'}\right)$ and $q\left(\mathbf{u}^{'}|m\right)$ can be chosen by investigation of a "pilot run": we start a Markov chain for each model from the best available starting values (for example the maximum likelihood estimates if they exist) and simulate the "within model" Markov chain a large number of times, in order to obtain approximate marginal posterior means and covariance matrices for each model parameter vector. These estimates are then used to construct proposal densities $q\left(\mathbf{u}|m^{'}\right)$ and $q\left(\mathbf{u}^{'}|m\right)$ taken as multivariate normal densities. To complete specification of our reversible jump MCMC algorithm, we need to specify the probabilities $j\left(m, m^{'}\right)$. We have used $j\left(m, m^{'}\right) = (|M| - 1)^{-1}$ for all $m, m^{'} \in M$, which is the simplest choice, where $|M|$ is the number of different models which are used in the reversible jump MCMC algorithm. The assessment of the convergence of Reversible Jump MCMC simulations is addressed by Brooks and Guidici (1998). Different Bayesian model selection strategies using MCMC are presented in Ntzoufras (1999).

1.3.5 Model Determination using Predictive distributions

Predictive distributions can be used to address the issues of model adequacy and model selection. This seems natural since one of the most important practical use of models is to construct predictions. According to Box (1980) the posterior distribution is used for the estimation of the model parameters while the predictive distribution enables the comparison of the models.

Suppose that $\mathbf{y}$ is a vector of T observations, $\boldsymbol{\theta}$ is a n-dimensional vector of the

parameters of interest and Y_{T+1} is an unknown observable. The distribution of Y_{T+1} is the posterior predictive distribution and is given by

$$p\left(Y_{T+1}|\mathbf{y}\right) = \int p\left(Y_{T+1}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}|\mathbf{y}\right) d\boldsymbol{\theta},$$

where $p\left(\boldsymbol{\theta}|\mathbf{y}\right)$ is the posterior distribution of $\boldsymbol{\theta}$. $p\left(Y_{T+1}|\mathbf{y}\right)$ is an average of conditional predictions over the posterior distribution of $\boldsymbol{\theta}$. The predictive distributions $p\left(Y_{T+1}|\mathbf{y}\right)$ are to be checked against the real value y_{T+1} in the sense that, if the model holds, y_{T+1} may be viewed as a random value from $p\left(Y_{T+1}|\mathbf{y}\right)$. We assess the comparative validity of models under consideration by using the estimate of the predictive densities $\widehat{p}\left(y_{T+1}|\mathbf{y}\right)$. The MCMC methods, described in previews sections, produce observations from the joint posterior $p\left(\boldsymbol{\theta}|\mathbf{y}\right)$. Hence, the outputted $\boldsymbol{\theta}_s$, $s = 1, ..., B$ can be used to carry out computations needed for model comparison. The estimate of the predictive density is

$$\widehat{p}\left(y_{T+1}|\mathbf{y}\right) = B^{-1} \sum_{s=1}^{B} p\left(y_{T+1}|\boldsymbol{\theta}_s, \mathbf{y}\right).$$

To estimate P one-step-ahead predictive densities $\widehat{p}\left(y_{T+1}|\mathbf{y}\right)$ we use the previous $T, T+1, ..., T+P-1$ data points. The quantity $\prod_{t=T}^{T+P-1} \widehat{p}\left(y_{t+1}|\mathbf{y}\right)$ is used for model comparison. The model with the largest value of $\prod_{t=T}^{T+P-1} \widehat{p}\left(y_{t+1}|\mathbf{y}\right)$ may be viewed as the most preferable. Equivalently, between two models m_i and m_j, we choose model m_i if

$$\log\left[\frac{\prod^{(m_i)} \widehat{p}\left(y_{t+1}|\mathbf{y}\right)}{\prod^{(m_j)} \widehat{p}\left(y_{t+1}|\mathbf{y}\right)}\right] > 0,$$

where $\prod^{(m_i)} \widehat{p}\left(y_{t+1}|\mathbf{y}\right)$ is the product of all $\widehat{p}\left(y_{t+1}|\mathbf{y}\right)$, $t = T, ..., T+P-1$ under model m_i. This criterion is described in more detail in Gelfand, Dey and Chang (1992), but their approach is a cross validation one. Vrontos, Dellaportas and Politis (1999) have used this criterion for the comparison of a class of multivariate ARCH and GARCH models, and Vrontos, Giakoumatos, Dellaportas and Politis (2000) have used it for the comparison of

multivariate GARCH and Stochastic volatility models. Similar diagnostic measures are presented by Pitt and Shephard (1999) in order to compare stochastic volatility models. Other criteria based on predictive densities are presented, for example, by Laud and Ibrahim (1995).

Chapter 2

Using the Subsampling methodology

2.1 Introduction

In this chapter we introduce a new convergence diagnostic method which assesses the convergence of both marginal and joint posterior densities. It can be applied to any MCMC sampler and uses the output from a single chain to gauge convergence. Generalization for multiple chains is straightforward, and the diagnostic can be very useful when there is suspicion of very influential starting points; see Gelman (1996). The diagnostic can be used to detect MCMC convergence in great generality and does not need to be combined with other methods in order to be effectively implemented. The subsampling method for statistical inference as developed by Politis and Romano (1994) and Politis, Romano and Wolf (1997) is used in this diagnostic. The diagnostic is based on obtaining (via subsampling) $(1-\alpha)100\%$ confidence regions for the posterior mean and for the $90th$ percentile of the first marginal distribution of the Markov chain at hand; the assessment of convergence can be done by using the coefficient of determination of a weighted linear regression. An alternative diagnostic is also proposed that uses the asymptotic normal distribution together with a subsampling estimate of the asymptotic variance-covariance matrix. A recent approach, which also looks with different perspective at the use of confidence intervals as a means of convergence diagnosis, is given by Brooks and Gelman

(1998). The authors generalize the method of Gelman and Rubin (1992) and they use the (squared) ratio of the lengths of the empirical estimated confidence intervals for the parameters of interest as an alternative interpretation of the $\hat{R}$ diagnostic (Gelman and Rubin, 1992), which is defined as the ratio of the between and within variance of the MCMC sequences. This alternative calculation of $\hat{R}$ is simpler than the original ratio of variances and is free from the assumption of normality.

2.2 The Subsampling methodology

The basic ideas of subsampling methodology for time series (Politis and Romano, 1994 and Politis, Romano and Wolf, 1997) are described. Let ($\mathbf{y}_1$,$\mathbf{y}_2$, ...,$\mathbf{y}_N$) be an observed stretch of a multivariate time series $\{\mathbf{y}_s, s = 1, 2, ...\}$; each $\mathbf{y}_s$ is assumed to be a $(p \times 1)$ vector. The time series is also assumed to be strong mixing, and asymptotically stationary.

The assumption of strong mixing is an assumption of "asymptotic independence": for any positive integers i and m, the two sets of random variables $\left(\mathbf{y}_i, \mathbf{y}_{i+1}, ..., \mathbf{y}_{i+m}\right)$ and $\left(\mathbf{y}_{i+m+k}, \mathbf{y}_{i+m+k+1}, ...,\mathbf{y}_{i+2m+k}\right)$ should be approximately independent if k is large enough; see, for example, Rosenblatt (1956), Doukhan (1994), Politis, Romano and Wolf (1997) for a precise definition. The notion of asymptotic stationarity means that the sequence $\left(\mathbf{y}_k, \mathbf{y}_{k+1}, ...\right)$ is approximately stationary in the strict sense (Brockwell and Davis, 1991) if k is large enough.

The basic idea of subsampling is to approximate the sampling distribution of a statistic based on the values of the same statistic recomputed over smaller subsets of the data that retain the dependence structure of the observations. If we are willing to consider subsets of size $b(< N)$, where b is a positive integer that in general may depend on N, then we are led to consider the $B = N - b + 1$ "blocks" of consecutive observations of the type $\left(\mathbf{y}_i, \mathbf{y}_{i+1}, ...,\mathbf{y}_{i+b-1}\right)$, for $i = 1, ..., B$.

Let $\mathbf{T}_N$ be a statistic of interest that is a function of the data sequence $(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N)$.

The statistic $\mathbf{T}_N$ is generally a vector (say q-dimensional), and is employed as an estimator of an unknown parameter $\boldsymbol{\theta}$. In general, $\boldsymbol{\theta}$ can be a parameter of the whole (infinite-dimensional) joint distribution of $\{\mathbf{y}_s, s = 1, 2, ...\}$; however, in the MCMC case considered in the next sections, $\boldsymbol{\theta}$ will almost always be a parameter of the invariant distribution of the Markov chain, i.e., the "asymptotic" first marginal of the sequence $\{\mathbf{y}_s, s = 1, 2, ...\}$.

We will assume that $\mathbf{T}_N$ is consistent for $\boldsymbol{\theta}$ as the sample size $N \to \infty$. More specifically, we will assume that the statistic $\mathbf{T}_N$, suitably centered and normalized, possesses a nondegenerate large-sample distribution. To be more precise, let $\{\tau_n, n = 1, 2, ...\}$ be an increasing sequence that diverges to ∞ as $n \to \infty$, let $||\cdot||$ denote a norm on space $\mathbf{R}^q$, and let

$$J_N(y) = Prob\left[\tau_N ||\mathbf{T}_N - \boldsymbol{\theta}|| \leq y\right].$$

The assumption required is that there exists some nondegenerate continuous distribution function $J(\cdot)$ such that

$$J_N(y) \to J(y) \tag{2.1}$$

for all y as $N \to \infty$.

Let $\mathbf{T}_{i,b}$ be our statistic of interest computed from block $\left(\mathbf{y}_i, \mathbf{y}_{i+1}, ..., \mathbf{y}_{i+b-1}\right)$, with all other data temporarily ignored, and construct an "empirical" distribution of the "subsample values" $\{\mathbf{T}_{i,b}, i = 1, ..., B\}$ by

$$L_N(y) = \frac{1}{B}\sum_{i=1}^{B} 1\left\{\tau_b ||\mathbf{T}_{i,b} - \mathbf{T}_N|| \leqslant y\right\}.$$

If the series $\{\mathbf{y}_s, s = 1, 2, ...\}$ is strong mixing and asymptotically stationary, and if our statistic possesses a nondegenerate large-sample distribution (i.e., equation (2.1) holds), then $L_N(\cdot)$ is a consistent estimator of the limit distribution $J(\cdot)$, provided b is chosen in a way that: $b \to \infty$ as $N \to \infty$, but $b/N \to 0$ and $\tau_b/\tau_N \to 0$; see, for example, Politis, Romano and Wolf (1997). Perhaps more important is that consistent estimation

of the quantiles of $J(\cdot)$ can be achieved by looking at the quantiles of $L_N(\cdot)$; in other words, for any $t \in (0,1)$,

$$L_N^{-1}(t) \to J^{-1}(t) \tag{2.2}$$

in probability as $N \to \infty$, where $L_N^{-1}(t) \equiv \inf\{y : L_N(y) \geq t\}$ and $J^{-1}(t) = \inf\{y : J(y) \geq t\}$ are the t quantiles of $L_N(\cdot)$ and $J(\cdot)$ respectively.

Using the quantiles of $L_N(\cdot)$ we can now construct confidence regions for $\boldsymbol{\theta}$ with a prescribed coverage level (to be attained in large samples). Relation (2.2) implies that the set $\left\{\boldsymbol{\theta} : \tau_N ||\mathbf{T}_N - \boldsymbol{\theta}|| \leq L_N^{-1}(1-\alpha)\right\}$ is a confidence region for $\boldsymbol{\theta}$ with asymptotic coverage probability equal to the nominal $1-\alpha$.

Note that our choice of norm $||\cdot||$ will dictate the shape of those confidence regions. If $||\cdot||$ is the Euclidean norm, then the confidence regions for $\boldsymbol{\theta}$ will be spheres centered at $\mathbf{T}_N$. Notably, the choice of sup-norm for $||\cdot||$ (i.e., the l_∞ norm which is nothing other than the maximum absolute coordinatewise deviation) results in confidence regions for $\boldsymbol{\theta}$ that have the shape of hypercubes with edges that are perpendicular to the axis; the sup-norm will be our choice in what follows as confidence regions that are hypercubes have the useful alternative interpretation as confidence intervals for the coordinates of $\boldsymbol{\theta}$ with *simultaneous* coverage equal to the coverage level of the whole confidence region.

Note that if the variances of different coordinates of the multivariate statistic $\mathbf{T}_N$ are of different orders of magnitude it may be inefficient to construct a hypercube for a confidence region for $\boldsymbol{\mu}$; rather, a "hyper-parallelepiped" should be constructed instead. To achieve this, the subsampling methodology must be applied to a "studentized" version of our statistic $\mathbf{T}_N$ as discussed in Politis and Romano (1994). In other words, a new "studentized" multivariate statistic $\tilde{\mathbf{T}}_N$ is defined with the property that the coordinates of $\tilde{\mathbf{T}}_N$ have all approximately equal variances; for example, we can define $\tilde{T}_N^{(k)} = T_N^{(k)} / \hat{S}_{T_N^{(k)}}$, where $\hat{S}^2_{T_N^{(k)}}$ is a consistent estimate of the variance of $T_N^{(k)}$, and thus ensure that all coordinates of $\tilde{\mathbf{T}}_N$ have approximately variance equal to one. Notably, $\hat{S}^2_{T_N^{(k)}}$ may even be a subsampling estimate of variance so that an "iterated" subsampling takes effect. However, for practical purposes, even a rough preliminary variance estimate can be used

in this type of studentization with good ensuing results.

2.3 The MCMC Subsampling diagnostic

In this section, the proposed method for assessing convergence of the MCMC output is presented. Let $(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N)$ be the multivariate output of an MCMC simulation; as before, each $\mathbf{y}_s$ is assumed to be a $(p \times 1)$ vector with $k-th$ coordinate denoted by $y_s^{(k)}$.

Notably, both assumptions required for subsampling to work (i.e., strong mixing and asymptotic stationarity) hold true in the MCMC case that interests us where the sequence $\{\mathbf{y}_s, s = 1, 2, ...\}$ is a Markov chain that possesses a unique invariant (i.e., stationary) distribution but the starting value $\mathbf{y}_1$ may follow a different distribution; see, for example, Meyn and Tweedie (1993).

Let $\overline{\mathbf{y}}_N = N^{-1} \sum_{i=1}^{N} \mathbf{y}_i$ denote the sample mean of the observed sequence $(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N)$; $\overline{\mathbf{y}}_N$ is of course a $(p \times 1)$ vector as well, and its $k-th$ coordinate will be denoted by $\overline{y}_N^{(k)}$. For $t \in (0, 1)$, let $\mathbf{q}_N^{(t)}$ be a $(p \times 1)$ vector with $k-th$ coordinate denoted by $q_N^{(k,t)}$, where $q_N^{(k,t)}$ is the empirical t-quantile of the $k-th$ coordinate data sequence $\left(y_1^{(k)}, y_2^{(k)}, ..., y_N^{(k)}\right)$; in other words, if $\left(y_{(1)}^{(k)} \leq y_{(2)}^{(k)} \leq \cdots \leq y_{(N)}^{(k)}\right)$ are the order statistics of the $k-th$ coordinate data sequence $\left(y_1^{(k)}, y_2^{(k)}, ..., y_N^{(k)}\right)$, then $q_N^{(k,t)} = y_{(\lfloor tN+1 \rfloor)}^{(k)}$, where $\lfloor . \rfloor$ is the integer part.

Using the subsampling methodology presented in section 2.2, and choosing for our statistic $\mathbf{T}_N$ either $\overline{\mathbf{y}}_N$ or $\mathbf{q}_N^{(t)}$ (with some choice of t that is of interest, e.g. $t = 0.90$), we can construct confidence regions for the mean and the t quantile of the "asymptotic" first marginal, i.e., the unique invariant distribution of the Markov chain $\{\mathbf{y}_s, s = 1, 2, ...\}$. Note that in either case ($\overline{\mathbf{y}}_N$ or $\mathbf{q}_N^{(t)}$) we have in general that $\tau_N = N^{1/2}$, i.e., $\sqrt{N}$-convergent statistics. Thus, we can choose the block size b proportional to N^{γ} (for some constant $\gamma \in (0, 1)$), and thus ensure that the conditions for (2.2) –on which the construction of confidence regions is based– are fulfilled.

As mentioned before, the sup-norm is recommended for use, i.e., $\|\mathbf{T}_N\| = \sup_{k=1,...,q} |T_N^{(k)}|$,

where $T_N^{(k)}$ is the $k-th$ coordinate of vector $\mathbf{T}_N$; therefore, the confidence regions are hypercubes in $\mathbf{R}^q$. Note also that in both cases ($\overline{\mathbf{y}}_N$ and $\mathbf{q}_N^{(t)}$) the "observation" dimension p coincides with the "parameter" dimension q.

Since our confidence regions are hypercubes in $\mathbf{R}^q$, we can define the "range" of such a confidence region as the qth root of its volume. The proposed diagnostic can now be viewed as a consequence of the following fact:

- The "range" of a $(1-\alpha)100\%$ confidence region for either the mean or the t quantile of the "asymptotic" first marginal is (asymptotically) proportional to $1/\sqrt{N}$.

Based on the above fact, our diagnostic can be formulated as follows:

(A) **Estimation of "burn-in" time**. As the simulation is running and N increases, construct $(1-\alpha)100\%$ confidence regions for the t quantile based on different (increasing) values of N; we used $\alpha = 0.05$ and $t = 0.90$ but other choices are possible as well. Now plot the "range" of the confidence region versus $1/\sqrt{N}$. We would estimate the "burn-in" time to be N^* if the plot of "range" versus $1/\sqrt{N}$ is approximately linear for $N > N^*$. Linearity can be checked by visual inspection of the plot, but we also recommend to use a plot of the coefficient of determination of a weighted linear regression between the dependent variable "range" and $1/\sqrt{N}$. Having estimated the "burn-in" time to be N^*, observations $(\mathbf{y}_1, \ldots, \mathbf{y}_{N^*})$ are discarded from the simulation as they could introduce undesired bias.

(B) **Figuring out when to stop the simulation**. Again as the simulation is running and N increases, construct $(1-\alpha)100\%$ confidence regions for the mean $\boldsymbol{\mu}$ of the "asymptotic" first marginal of the Markov chain $\{\mathbf{y}_s\}$. Since the main objective of the simulation is to estimate $\boldsymbol{\mu}$ by Monte Carlo, and since the error in this estimation can be quantified by the range of a $(1-\alpha)100\%$ confidence region for $\boldsymbol{\mu}$, we would then propose to stop the MCMC simulation when the range of this $(1-\alpha)100\%$ confidence region (with $\alpha = 0.05$, say) is appropriately small, smaller

than some prespecified absolute or relative measure of accuracy; for example, we could stop when the range becomes smaller than 0.001 (say), thus obeying an absolute measure, or when it is smaller than 0.001 $||\overline{\mathbf{y}}_N||$ which is then a relative measure (relative to $||\boldsymbol{\mu}||$ which is estimated by $||\overline{\mathbf{y}}_N||$).

The proposed diagnostic is simple and easy to use, and has the significant advantage of being valid in *asymptotically* stationary settings such as the MCMC case of interest; see, for example, Yue and Chan (1996). Notably, other diagnostics are shown to be valid only under the assumption of stationarity, thus neglecting the fact that the MCMC output is *not* exactly stationary.

The reason that the t quantile (with a large t, say $t = 0.90$) is considered in part A of the diagnostic as opposed to a similar graphical plot of the confidence range of other statistics (e.g. the sample mean) is based on the notion that stabilization of estimates of the invariant distribution of the Markov chain (especially in the tails) is a reliable indicator of the target distribution having been achieved.

Other statistics are also possible even in part B of the diagnostic. For example, in a simulated annealing setting where the posterior mode is the objective, our statistic $\mathbf{T}_N$ could have as $k-th$ coordinate the sample mode of the $k-th$ coordinate data sequence $\left(y_1^{(k)}, y_2^{(k)}, ..., y_N^{(k)}\right)$; for concreteness however, we will focus in the sequel on the sample mean and t quantile of an MCMC output.

Given the burn-in sample of the MCMC chain and the sample size needed to estimate the parameters of interest with the desired accuracy, we can estimate the variance covariance matrix of these parameters using the whole sample (excluding the burn-in) with a subsampling estimate. For example, we could use the blocked sample variance as introduced by Politis and Romano (1993) and its formula is presented in the next section. This treatment could be beneficial as indicated by MacEachern and Berliner (1994).

2.4 An alternative diagnostic

In cases where the asymptotic distribution of our statistic $\mathbf{T}_N$ is known to be normal, then an alternative method that combines subsampling with the information regarding asymptotic normality may be used. Note that, the sample mean and sample 0.90 quantile, which are chosen as statistics of interest, are both asymptotically normal under standard regularity conditions; see, for example, Brockwell and Davis (1991). To elaborate, let ($\mathbf{y}_1$,$\mathbf{y}_2$, ...,$\mathbf{y}_N$) be the multivariate output of an MCMC simulation; as before, each $\mathbf{y}_s$ is assumed to be a $(p \times 1)$ vector with $k-th$ coordinate denoted by $y_s^{(k)}$. The assumptions that are required for subsampling (i.e., strong mixing and asymptotic stationarity) still are assumed to hold together with the additional assumption that $\sqrt{N}\,(\mathbf{T}_N - \boldsymbol{\theta}) \stackrel{L}{\Longrightarrow} N\,(\mathbf{0}, \boldsymbol{\Sigma}_\infty)$, for some nonnegative definite matrix $\boldsymbol{\Sigma}_\infty$.

The problem that now turns up is the estimation of the unknown asymptotic variance-covariance matrix $\boldsymbol{\Sigma}_\infty$. Subsampling can be used for this purpose as well, under some additional regularity conditions. Carlstein (1986) proposed the subsampling estimators of variance based on nonoverlapping blocks from a stationary sequence, while Künsch (1989) addressed the case of overlapping blocks. Recently, Fukuchi (1997) demonstrated the asymptotic consistency of the subsampling estimator of variance in the case of asymptotically stationary time series considered in this paper.

Thus, as an estimator of $\boldsymbol{\Sigma}_\infty$ we may use the blocked sample variance matrix $\hat{\mathbf{V}}_{b/N}$, where b is the block size as in the previous section. The blocked sampled variance is given from the following formula (Politis and Romano, 1993):

$$\hat{\mathbf{V}}_{b/N} = \begin{bmatrix} \hat{\sigma}^{11}_{b/N} & \hat{\sigma}^{12}_{b/N} & \cdots & \hat{\sigma}^{1p}_{b/N} \\ \hat{\sigma}^{21}_{b/N} & \hat{\sigma}^{22}_{b/N} & \cdots & \hat{\sigma}^{2p}_{b/N} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}^{p1}_{b/N} & \hat{\sigma}^{p2}_{b/N} & \cdots & \hat{\sigma}^{pp}_{b/N} \end{bmatrix}$$

where

$$\hat{\sigma}_{b/N}^{ij} = \frac{b}{N-b+1} \sum_{m=1}^{N-b+1} \left[T_{m,b}^{(i)} - T_N^{(i)} \right] \left[T_{m,b}^{(j)} - T_N^{(j)} \right], \; i,j = 1,2,...,p$$

Since $\hat{\mathbf{V}}_{b/N}$ is an asymptotically consistent estimator of $\mathbf{\Sigma}_\infty$ it follows that we can approximate the probability law of

$$\left\| \sqrt{N} \left(\mathbf{T}_N - \boldsymbol{\theta} \right) \right\| = \sup_{k=1,...,q} |\sqrt{N} \left(T_N^{(k)} - \theta^{(k)} \right)|,$$

by the probability law of $\sup_{k=1,...,q} |Z^{(k)}|$, where the multivariate random variable $\mathbf{Z}$ has the $N\left(\mathbf{0}, \hat{\mathbf{V}}_{b/N}\right)$ distribution. Although the latter is difficult to evaluate analytically, it is nevertheless very easy to approximate by Monte Carlo. For this purpose we let $\mathbf{Z}_1, \mathbf{Z}_2$,...,$\mathbf{Z}_M$ be iid random variables having distribution $N\left(\mathbf{0}, \hat{\mathbf{V}}_{b/N}\right)$, where M is large. Now we have

$$Prob[\sup_{k=1,...,q} |\sqrt{N} \left(T_N^{(k)} - \theta^{(k)} \right)| \leq y] \simeq M^{-1} \sum_{m=1}^{M} 1\{\sup_{k=1,...,q} |Z_m^{(k)}| \leq y\} \simeq \tilde{L}_N(y), \quad (2.3)$$

where we define

$$\tilde{L}_N(y) \equiv \lim_{M\to\infty} M^{-1} \sum_{m=1}^{M} 1\{\sup_{k=1,...,q} |Z_m^{(k)}| \leq y\};$$

note that the approximations " $\simeq$ " in equation (2.3) occur with high probability for large M and N, i.e. they are justified as convergences in probability. The bootstrap 'rule-of-thumb' is to take M of the order of a few thousands.

The alternative subsampling diagnostic proceeds the same way (parts A and B) as our first subsampling diagnostic, the only difference being that now our confidence regions for the mean and t-quantile are based on the quantiles of the distribution $\tilde{L}_N(\cdot)$, and not on the quantiles of the subsampling distribution $L_N(\cdot)$. In cases where the assumption of asymptotic normality is valid then the alternative method is expected to work better than the subsampling diagnostic of section 2.3 as it uses this additional information (on

the asymptotic normality). On the other hand, if the assumption of normality does not hold then the subsampling diagnostic of section 2.3 must be used.

2.5 Implementation and numerical simulations

2.5.1 Description of the simulations

Using the MCMC algorithm, we recursively create the Markov sequence $(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N)$, where each $\mathbf{y}_i$ is a $(p \times 1)$ vector; p denotes the number of parameters and N is the number of iterates (the total sample size). As before the quantity of interest is a function of the data sequence $(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N)$ which we denote by the $(q \times 1)$ vector $\mathbf{T}_N$; in our case we will consider $\overline{\mathbf{y}}_N$ or $\mathbf{q}_N^{(t)}$ for our $\mathbf{T}_N$ statistic, and therefore $q = p$.

To fix ideas, let $N_j = jN/100$, for $j = 1, ..., 100$. The simulation algorithm is now precisely described; for $j = 1, ..., 100$ do the following:

1) As discussed in section 2.2, identify the $B_j = N_j - b_j + 1$ subsamples $(\mathbf{y}_i, \mathbf{y}_{i+1}, ..., \mathbf{y}_{i+b_j-1})$, for $i = 1, ..., B_j$; note that the subsample size b_j depends on N_j. We used the simple choice $b_j = \sqrt{N_j}$, although other choices are possible as well; see Hall, Horowitz and Jing (1995), Politis, Romano and Wolf (1997).

2) From the sequence $\left(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{N_j}\right)$, calculate the quantity of interest $\mathbf{T}_{N_j}$ (sample mean or t quantile with $t = 0.90$).

3) From each subsample $\left(\mathbf{y}_i, \mathbf{y}_{i+1}, ..., \mathbf{y}_{i+b_j-1}\right)$, recalculate the quantity of interest $\mathbf{T}_{i,b_j}$, where $i = 1, 2, ..., B_j$.

4) Let $T_{i,b_j}^{(k)}$ and $T_{N_j}^{(k)}$ denote the $k-th$ coordinates of the vectors $\mathbf{T}_{i,b_j}$ and $\mathbf{T}_{N_j}$ respectively. For $i = 1, 2, ..., B_j$, compute the "maximum deviation" (i.e. sup-norm) from block i as

$$d_{i,j} = \max_{k=1,...,q} \left| T_{i,b_j}^{(k)} - T_{N_j}^{(k)} \right| \text{ and } D_{i,j} = \sqrt{b_j} d_{i,j}$$

5) In order to find the estimated quantile $L_{N_j}^{-1}(1-\alpha)$ (with $\alpha = 0.05$), we sort $D_{i,j}$,

$i = 1, 2, ..., B_j$ in ascending order to come up with the order statistics $D_{(1,j)} \leq D_{(2,j)} \leq \cdots \leq D_{(B_j,j)}$, and then let

$$L_{N_j}^{-1}(1-\alpha) = D_{(\lfloor (1-\alpha)\cdot B_j + 1 \rfloor, j)}$$

where $\lfloor . \rfloor$ is the integer part.

6) As discussed before, the confidence region at "time" N_j is a hypercube in q dimensions that is centered at the value $\mathbf{T}_{N_j}$ and has "sides" that are perpendicular to the coordinate axes. The length of each "edge" of the hypercube is identical to the "range" R_j of the hypercube region which is given by

$$R_j = \frac{2L_{N_j}^{-1}(1-\alpha)}{\sqrt{N_j}}.$$

A different way of describing this situation is to say that the q confidence intervals of the type

$$\left(T_{N_j}^{(k)} \pm \frac{L_{N_j}^{-1}(1-\alpha)}{\sqrt{N_j}} \right),$$

for $k = 1, ..., q$, have *simultaneous* coverage $1-\alpha$ for the q respective coordinate parameters, provided of course the sample size N_j is large.

7) Finally plot R_j versus $1/\sqrt{N_j}$ and take appropriate action based on the plot (quantile case), or just the magnitude of the range for the largest N_j considered (sample mean case).

The implementation of the alternative diagnostic is exactly like the implementation of the subsampling diagnostic above, the only difference being that the quantile $\tilde{L}_{N_j}^{-1}(1-\alpha)$ is put instead of $L_{N_j}^{-1}(1-\alpha)$ in all occurrences of the latter.

2.5.2 Estimation of the 'burn-in' period using the coefficient of determination

Although linearity can be visually assessed by inspecting the graph, a more objective/automatic method might be desirable. One way to automatically check the linearity between the range of the confidence region and the $1/\sqrt{N_j}$, in order to estimate the burn-in, is the coefficient of determination (R^2) between these two variables. In detail, the assumed linear relation is given by the linear model

$$\mathbf{y}^*_{j:j+c-1} = \beta \mathbf{x}^*_{j:j+c-1} + \boldsymbol{\varepsilon}$$

where $\mathbf{y}^*_{j:j+c-1}$, $\mathbf{x}^*_{j:j+c-1}$ denote the vectors $\left(y^*_j, ..., y^*_{j+c-1}\right), \left(x^*_j, ..., x^*_{j+c-1}\right)$ respectively and $\boldsymbol{\varepsilon}$ represents a mean-zero error term. Each element of the $\mathbf{y}^*_{j:j+c-1}$ is the estimated range of the confidence region of the 0.90 quantile (or of the mean) using N_j iterations and each element of $\mathbf{x}^*_{j:j+c-1}$ is the corresponding $1/\sqrt{N_j}$. The quantity c is chosen by the practitioner and represents the window of our regression, that is, the number of elements $\left(y^*_j, x^*_j\right)$ in the vectors $\mathbf{y}^*_{j:j+c-1}$, $\mathbf{x}^*_{j:j+c-1}$ that we use to calculate the coefficient of determination. For each pair of vectors $\left(\mathbf{y}^*_{j:j+c-1}, \mathbf{x}^*_{j:j+c-1}\right)$ we calculate the R^2, using the weighted least squares method with weights $w_j = \sqrt{b_j}/N_j$. We use weighted linear regression because the dependent variable is not homoscedastic and its standard deviation is proportional to the ratio $\sqrt{b_j}/N_j$; see, for example, Politis and Romano (1993) where the variance of the subsampling estimate of variance is calculated. We choose $c = 20$ and therefore there are $(100 - c + 1) = 81$ different $\mathbf{y}^*_{j:j+c-1}$, $\mathbf{x}^*_{j:j+c-1}$ vectors; for each of them we calculate the R^2_k, for $k = 1, 2, ..., 100 - c + 1$. If $R^2_k > d$ for all $k > k^*$ and some prespecified threshold d we discard the first N_{k^*-1} iterations as burn-in. Typical values of d we use are 0.998 or 0.999.

2.5.3 Examples and simulations

To illustrate how the proposed diagnostic performs in practice, the following three examples are presented. The first example concerns a trivariate normal distribution with high correlation, the second deals with a bimodal mixture of trivariate normals and the third refers to the stochastic search variable selection (SSVS) MCMC output introduced by George and McCulloch (1993).

Example 1

This example is taken from the MCMC diagnostics review paper of Cowles and Carlin (1996). We assume that we deal with a three parameter joint posterior density which is a zero-mean trivariate normal with correlations 0.90, 0.90 and 0.98 and covariance matrix

$$\begin{pmatrix} 1.0 & 4.5 & 9.0 \\ 4.5 & 25.0 & 49.0 \\ 9.0 & 49.0 & 100.0 \end{pmatrix}.$$

We initialize the Gibbs Sampler with $\theta_2 = 10, \theta_3 = -10$. The algorithm then proceeds by sampling from the full conditional densities, which generally are given by:

$$\theta_i|\boldsymbol{\theta}_j \sim N\left(\mu_i + \boldsymbol{\Sigma}_{ij}\boldsymbol{\Sigma}_{jj}^{-1}\left(\boldsymbol{\theta}_j - \boldsymbol{\mu}_j\right),\ \boldsymbol{\Sigma}_{ii} - \boldsymbol{\Sigma}_{ij}\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{ji}\right),\ i = 1, 2, 3$$

where, θ_i is the $i-th$ element of the vector of parameters, $\boldsymbol{\theta}_j$ is the vector of all the other parameters except the $i-th$, μ_i is the $i-th$ element of the mean vector, $\boldsymbol{\mu}_j$ is the vector of all the other elements except the $i-th$, and $\boldsymbol{\Sigma}_{ij}$, $\boldsymbol{\Sigma}_{jj}$, $\boldsymbol{\Sigma}_{ii}$, $\boldsymbol{\Sigma}_{ji}$ are the corresponding partitions of the variance-covariance matrix.

To illustrate our diagnostic test, we generate $N = 30000$ values from the above iterative Gibbs sampling scheme. A studentization of the values of the parameters has been made using the simple variance estimator. We estimate the range of the confidence interval for the 0.90 quantile for $N_j = jN/100$, $j = 1, ..., 100$, samples. The weighted

	Iterations				
	6000	12000	18000	24000	30000
θ_1	0.2783	0.2251	0.1824	0.1624	0.1481
θ_2	1.4894	1.2206	0.9867	0.8746	0.8023
θ_3	2.9927	2.4593	1.9854	1.7529	1.6083

Table 2.1: Accuracy of the posterior mean of the model parameters.

linear regression of the range versus $1/\sqrt{N_j}$ is our burn-in indicator. Adopting $d = 0.999$ as a threshold for R^2, we discard as burn-in the first 4800 iterations because after that $R^2 > d$; see Figure 2-1(a), where the values of R^2 are presented across iterations. The results of the alternative method are also illustrated in Figure 2-1(a) and are similar with the results of our subsampling diagnostic.

The second part of the problem is to estimate how many points are needed to estimate the parameters of interest with the desired accuracy. This can be done by using the range of the confidence interval for the mean. The accuracy of the parameters is the difference of the upper and lower limit of the confidence interval of the mean. Table 2.1 gives the accuracy of the posterior mean of these parameters for some iterations.

A diagnostic which is similar in spirit with ours is the one suggested by Raftery and Lewis (1992). To compare the two methodologies, we run the Raftery and Lewis diagnostic requiring the precision achieved by our diagnostic for the 0.90 quantile after 30000 iterations. The input values were $q = 0.90$, the 0.90 quantile, $r = \pm 0.085$, the precision of the 0.90 quantile and $s = 0.95$, the probability of estimating the 0.90 quantile within ± 0.085 for the parameter θ_1. The precision of the 0.90 quantile for parameters θ_2 and θ_3 was ± 0.458 and ± 0.924 respectively. The required burn-in and sample size results were 104 and 1105 iterations. It is evident that our diagnostic is much more conservative than the one by Raftery and Lewis; this is in part due to our high choice (0.999) as a linearity threshold in our R^2 criterion. Cowles and Carlin (1996) argued that the Raftery and Lewis' method is less conservative because more iterations are required for estimating quantiles near the median than extreme quantiles. Nevertheless, in this example, the Raftery and Lewis diagnostic seems to underestimate the sample size

required to get the claimed accuracy. To elaborate, after 1105 iterations and excluding the burn-in of 104 iterations the estimated 0.90 quantiles turn out to be 1.65, 9.02 and 17.78 for the parameters θ_1, θ_2 and θ_3 respectively, whereas after 30000 iterations the 0.90 quantiles are estimated to be 1.31, 6.69, and 13.34 respectively, within the desired accuracy of the true 0.90 quantiles 1.28, 6.40, and 12.81. Note that in both the Raftery and Lewis diagnostic and our subsampling diagnostic, convergence detection is related to precision required. By demanding smaller accuracy in the posterior quantity of interest, one needs more iterations for convergence.

We ran the subsampling and the alternative method on a Sun Ultra-2, Sparcstation. The computational time for the subsampling diagnostic we propose is around 3180 seconds, while the alternative method needs 3306 seconds. For both methods the above times were used to calculate confidence intervals for both the mean and the 0.90 quantile of the joint density. The Raftery and Lewis diagnostic required only a fraction of this time (7 seconds) being one of the cheapest diagnostic available; see Brooks and Roberts (1999).

Example 2

This example is also taken from Cowles and Carlin (1996) and refers to a bimodal target density consisting of a mixture of two trivariate normals with equal probability. These two normals have a common covariance matrix

$$\begin{pmatrix} 1 & 1.3 & 1.5 \\ 1.3 & 2 & 2 \\ 1.5 & 2 & 4 \end{pmatrix},$$

which produces correlations 0.919, 0.75, 0.707, and mean vectors (0 0 0) and (-6 -8.49 -12). We generate an MCMC output using a random walk Metropolis Hastings algorithm consisting of 8000 values.

Two independent chains with starting values ($\theta_1 = 5$, $\theta_2 = 15$, $\theta_3 = 10$) and ($\theta_1 =$

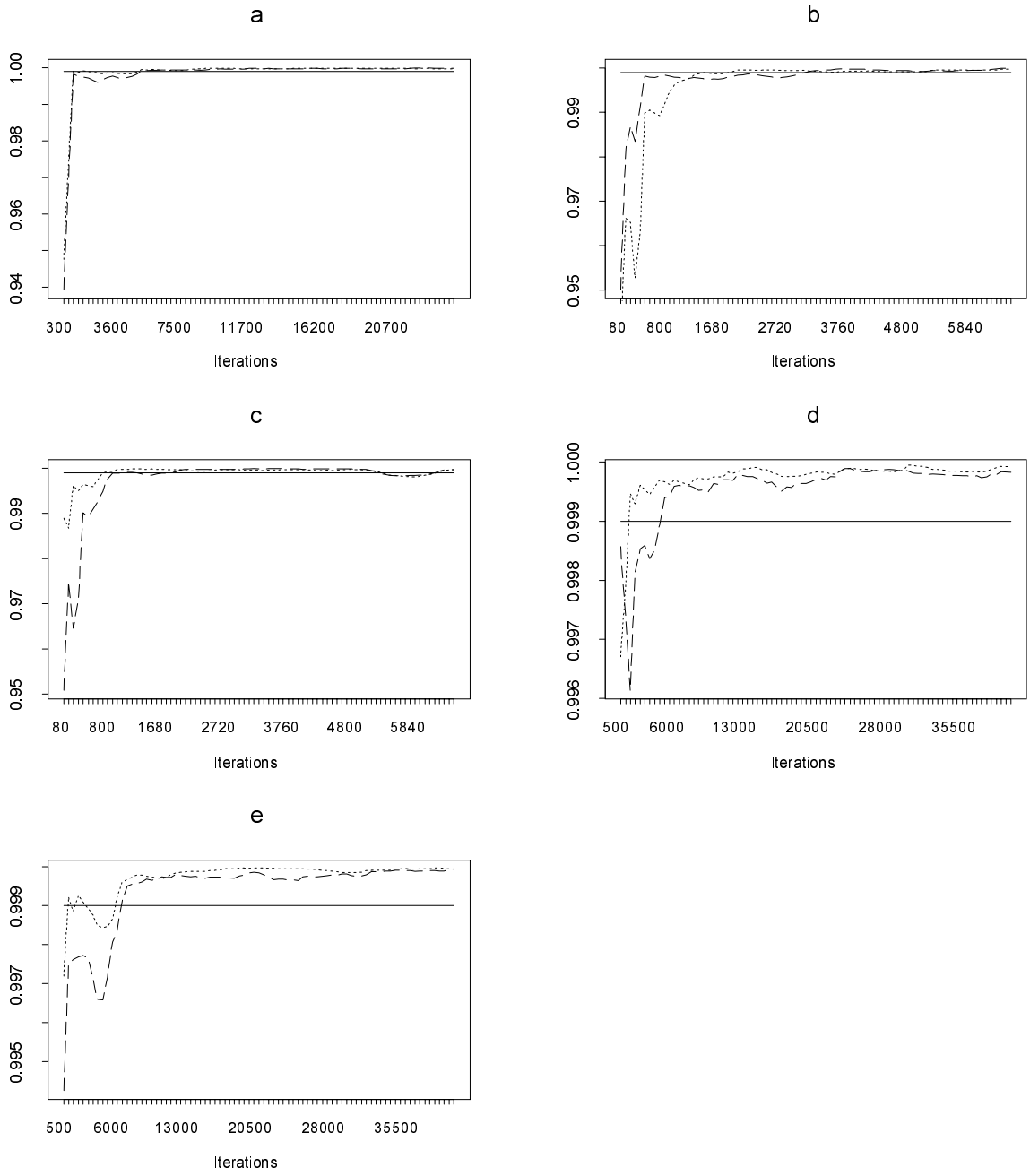

Figure 2-1: Coefficient of determination. Solid line: threshold d=0.999, dashed line: subsampling method, dotted line: alternative method. (a) Example 1, (b) Example 2, first chain, (c) Example 2, second chain, (d) Example 3, problem 1, (e) Example 3, problem 2.

-15, $\theta_2 = -20$, $\theta_3 = -25$) were used to illustrate our diagnostic. The Metropolis proposal density was chosen so that the two chains remain in the area of one mixture component and therefore the chain does not visit the whole parameter space. Figures 2-1(b) and 2-1(c) depict the resulting R^2 for both our suggested diagnostics for the two chains respectively. Although clearly the Markov chain has not converged to its stationary distribution, in Figure 2-1(b) we receive the wrong signal of "getting in" the target distribution, after 3040 iterations using the subsampling and after 1840 iterations using the alternative method. On the other hand, in Figure 2-1(c) we correctly detect that more iterations are needed because R^2 is not continuously higher than the threshold 0.999. The estimated accuracy of the mean of the parameters θ_1, θ_2, θ_3 is 0.3188, 0.4422, 0.6571, respectively for the first chain and 0.3387, 0.4802, 0.5613, for the second chain after 8000 iterations. Also, the estimated accuracy of the 0.90 quantile of the parameters θ_1, θ_2, θ_3 is 0.3873, 0.5161 and 0.7270 respectively for the first chain, and 0.3935, 0.5106 and 0.7073 for the second chain after 8000 iterations, which clearly suggests that only a crude estimate is available.

Example 3

There are cases in which the posterior summary of interest may be only the posterior mean with a corresponding confidence interval. These cases are particularly suited to our methodology. For example, take the usual model choice or variable selection approaches dealt with the MCMC algorithm; see for example Green (1995), George and McCulloch (1993). In these models the MCMC output contains a variable, say γ, which expresses the probability of a model or the probability that a variable is included in the model. This variable is a string of 0 or 1, and in stationarity, the mean and confidence intervals of γ are the desired posterior summaries of interest.

A major problem in the linear model theory is the choice of the appropriate set of regressors which explain satisfactory the variability of the dependent variable. Recently George and McCulloch (1993) developed the Stochastic Search Variable Selection which

Problem 1			Problem 2		
Model	Prob	Accuracy	Model	Prob	Accuracy
$\{\mathbf{X}_4, \mathbf{X}_5\}$	0.23	0.010311	$\{\mathbf{X}_4, \mathbf{X}_5\}$	0.10	0.010450
$\{\mathbf{X}_5\}$	0.16	0.008893	$\{\mathbf{X}_3, \mathbf{X}_4\}$	0.09	0.009580
$\{\mathbf{X}_2, \mathbf{X}_4, \mathbf{X}_5\}$	0.09	0.006071	$\{\mathbf{X}_4\}$	0.08	0.007388
$\{\mathbf{X}_2, \mathbf{X}_5\}$	0.06	0.005508	$\{\mathbf{X}_5\}$	0.07	0.007678
$\{\mathbf{X}_4\}$	0.06	0.005530	$\{\mathbf{X}_3\}$	0.06	0.006669
			$\{\mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5\}$	0.05	0.009486

Table 2.2: Posterior model probabilities and their accuracies.

enables the calculation of the posterior probability of inclusion of a regressor.

To illustrate our methodology, we use the example 4.1 of George and McCulloch (1993). There are five regressors $\mathbf{X}_1, ..., \mathbf{X}_5 \overset{iid}{\sim} \mathbf{N}(0,1)$ of size $n = 60$ which are used in two variable selection problems. In Problem 1, the dependent variable is generated according to the model

$$\mathbf{Y} = \mathbf{X}_4 + 1.2\mathbf{X}_5 + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon} \sim \mathbf{N}_{60}(0, \sigma^2\mathbf{I})$ with $\sigma = 2.5$. Problem 2 is identical to Problem 1, apart from the regressor $\mathbf{X}_3$ which is replaced by $\mathbf{X}_3^* = \mathbf{X}_4 + 0.15\mathbf{Z}$ where $\mathbf{Z} \sim \mathbf{N}(0,1)$, yielding $corr(\mathbf{X}_3, \mathbf{X}_5) = 0.99$.

For each of the potential models of the above problems we can construct, using the SSVS method, a variable which takes values 1 or 0, depending on whether the particular model is chosen or not in the current iteration. To obtain a sample from these posterior model probabilities we construct an MCMC chain that converges to the posterior distribution of interest. For more details of the above methodology and the example, see George and McCulloch (1993).

We focus our analysis on the models that have more than 0.05 posterior probability. Using an MCMC chain of 50000 iterations and choosing as threshold value $d = 0.999$, we estimate the burn-in period. This comes out to be 4500 and 6000 iterations for Problems 1 and 2 respectively (see Figures 2-1(d) and 2-1(e)). The resulting model probabilities and corresponding accuracies after 50000 iterations are presented in Table 2.2. The

Problem 1		Problem 2	
Model	Iterations	Model	Iterations
$\{\mathbf{X}_4, \mathbf{X}_5\}$	252286	$\{\mathbf{X}_4, \mathbf{X}_5\}$	1150652
$\{\mathbf{X}_5\}$	345571	$\{\mathbf{X}_3, \mathbf{X}_4\}$	1125107
$\{\mathbf{X}_2, \mathbf{X}_4, \mathbf{X}_5\}$	521015	$\{\mathbf{X}_4\}$	883304
$\{\mathbf{X}_2, \mathbf{X}_5\}$	817300	$\{\mathbf{X}_5\}$	1107036
$\{\mathbf{X}_4\}$	855504	$\{\mathbf{X}_3\}$	1359638
		$\{\mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5\}$	3941288

Table 2.3: Iterations for specific accuracy.

variables in the curly brackets are contained in the resulting models.

Assume that, in a hypothetical scenario, one needs to obtain posterior model probabilities with required accuracy of 2% of their estimated value. The results of our diagnostic can guide the MCMC to run for a number of iterations such that the above requirement is satisfied. In detail, discarding the burn-in period for each model, there is a strong linear relation between the range of the confidence region and the $1/\sqrt{N_j}$ as indicated in previous sections. Therefore, a weighted regression between the range, as dependent variable, and $1/\sqrt{N_j}$, could give an estimate for the required iterations that are needed to estimate the posterior model probabilities with the desired accuracy. Table 2.3 contains the estimated iterations that an MCMC needs in order to estimate the model probabilities for Problems 1 and 2 with accuracy of 2% of their estimated value. Note that if our desired accuracy was produced in less than 50000 iterations, this could be a significant drawback of our diagnostic: we produced unnecessary iterations to estimate that only a portion of them is needed!

2.6 Discussion

We have presented an MCMC convergence strategy which is based on subsampling from the MCMC output chain. The methodology we suggest is based on results that hold in asymptotically stationary settings such as the MCMC output.

We have demonstrated the use of our methodology for both single and multiple

MCMC chains. Naturally, combining information from many output chains can produce a safer diagnostic. For example, confidence intervals for parameters should approximately "coincide" in each MCMC chain. We believe that such extensions are straightforward so we have not pursued this issue further.

Chapter 3

Time-varying volatility models

3.1 Introduction

Univariate and multivariate ARCH-type time-varying volatility models are presented in this chapter. Research into time series models of changing variance and covariance has exploded in the last twenty years. The main reason for this activity is that uncertainty is central to much of modern finance theory. In option pricing the uncertainty associated with the future price of the underling asset is the most important determinant in the pricing formula. Many issues in asset pricing, portfolio allocation decisions, and risk management can only be meaningfully analyzed in a multivariate framework, where modelling of variances and covariances is crucial.

Empirical research on the statistical properties of asset returns dates back to the work of Mandelbrot (1963) and Fama (1965). Asset returns tend to be leptokurtic. This leads on modelling asset returns as iid draws from thick tailed distributions. Asset returns tend to exhibit nonnormal unconditional sampling distribution, in the form of skewness but more pronounced in the form of excess kurtosis. The assumption of conditional normality captures some degree of excess kurtosis, but typically less than adequate to fully account for the fat tailed properties of the data. Therefore, conditional distributions with fatter tails than normal distribution should be adopted. Some of these distributions

are the Student-t (see, for example, Bollerslev, 1987; Baillie and Bollerslev, 1989; Sharma, Mougoue and Kamath, 1996) and the Generalized error distribution (see, for example, Nelson, 1991) among others.

An empirical characteristic of asset returns is volatility clustering. The autocorrelations of squares and absolute returns have more significant values than the autocorrelations of returns. This is evidence that large (small) price changes tend to be followed by large (small) changes, of either sign, and this phenomenon is more marked for higher frequency series. Generally, the distribution of the next absolute or squared return depends not only on the current return but also on several previous returns.

In asset return series, there is a tendency that current returns are negatively correlated with future volatility. This characteristic is called "leverage effect". According to "leverage effect" a reduction in the asset value would raise the risk and would increase the future volatility. That is, volatility tends to rise in response to "bad news" (returns are lower than expected) and tend to fall in response to "good news" (returns are higher than expected).

It has been observed that there is a lot of commonality in volatility changes across assets. For example, when stock volatilities change, they all tend to change in the same direction. The fact that the volatilities move together indicates that a few common factors may explain much of the temporal variation in the conditional variances and covariances of asset returns. Co-movements in volatilities are the basis for the factor models.

The importance of risk and uncertainty in modern economic theory and the empirical evidence of the above characteristics of asset returns have necessitated the development of new econometric time series techniques that allow for modelling time-varying variances and covariances. Two classes of models, called ARCH-GARCH and Stochastic volatility models, have been developed. In ARCH-type models the information set is a function of lagged values of the process of interest. Reviews of the literature on this topic are given by Bollerslev, Chou and Kroner (1992), Bera and Higgins (1993), Bollerslev, Engle and Nelson (1994), Diebold and Lopez (1995), Gourieroux (1997), Hafner (1998), and a collec-

tion of ARCH papers is in Engle (1995). In Stochastic volatility models the information set is a function of some unobserved or latent components. Previous work and review on Stochastic volatility models has been reported by Harvey, Ruiz and Shephard (1994), Jacquier, Polson and Rossi (1994), Shephard (1996), Pitt and Shephard (1999), Chib, Nardari and Shephard (1999), Giakoumatos, Dellaportas and Politis (1999), Aguilar and West (2000) among others. In this study the analysis is concentrated in ARCH-type models. We present and discuss some of the most important ideas, focusing on the simplest forms of the techniques and models used in the literature. Univariate and multivariate models are presented in the next sections.

3.2 Univariate ARCH-type models

Traditional econometric models assume a constant one step ahead forecast variance. A new class of models called autoregressive conditional heteroscedastic (ARCH) have been introduced by Engle (1982) to relax this assumption. Let $\{\varepsilon_t\}$ be a real valued discrete-time stochastic process, which is the innovation in the mean for some observable stochastic process $\{y_t\}$, where

$$y_t = g\left(x_{t-1}, \boldsymbol{\mu}\right) + \varepsilon_t,$$

and $g\left(x_{t-1}; \boldsymbol{\mu}\right)$ is a function of exogenous variable x and of parameter vector $\boldsymbol{\mu}$. For simplicity, we assume that the mean equation is of the form $y_t = \varepsilon_t$, and therefore the innovation process $\{\varepsilon_t\}$ is itself observable.

3.2.1 ARCH model

The ARCH(p) model (Engle, 1982) is given by the following two-stage formulation:

$$\varepsilon_t = z_t \sigma_t \tag{3.1}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2, \ t = 1, ..., T, \tag{3.2}$$

where z_t are iid with $E(z_t) = 0$ and $Var(z_t) = 1$; σ_t^2 is the variance of the $\{\varepsilon_t\}$ process at time t conditional on information available at time $t-1$, p is integer with $p > 0$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, ..., p$. These restrictions ensure a positive variance. Stationarity conditions impose that $\sum_{i=1}^{p} \alpha_i < 1$, and the stationary variance is given by

$$\sigma_\varepsilon^2 = E\left(\varepsilon_t^2\right) = \alpha_0 / \left[1 - \sum_{i=1}^{p} \alpha_i\right],$$

see, for details, Engle (1982). Solving with respect to α_0, we obtain $\alpha_0 = \sigma_\varepsilon^2 \left[1 - \sum_{i=1}^{p} \alpha_i\right]$. Substituting this expression into (3.2), the conditional variance is given by:

$$\sigma_t^2 = \sigma_\varepsilon^2 \left[1 - \sum_{i=1}^{p} \alpha_i\right] + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2.$$

Therefore, the conditional variance is a weighted average of the "global" variance σ_ε^2 and the "local" variances ε_{t-i}^2, $i = 1, ..., p$.

Under conditional normality, an equivalent representation of the ARCH model is $\varepsilon_t | \Phi_{t-1} \sim N(0, \sigma_t^2)$, where Φ_{t-1} is the information set up to time $t-1$, and σ_t^2 is given by (3.2). Clearly the model can be written as a non-Gaussian AR model in terms of ε_t^2:

$$\begin{aligned} \varepsilon_t^2 &= \sigma_t^2 + \left(\varepsilon_t^2 - \sigma_t^2\right) \\ &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + v_t, \end{aligned}$$

where $v_t = \varepsilon_t^2 - \sigma_t^2 = \sigma_t^2 (z_t^2 - 1)$, and $E(v_t | \Phi_{t-1}) = 0$.

Engle (1982) presents conditions for the existence of the moments of the ARCH(1) model under conditional normality. All odd moments are zero under a symmetric conditional distribution. The condition for the variance to be finite is $\alpha_1 < 1$, while to have a

finite fourth moment it is also required that $3\alpha_1^2 < 1$. These moments are, respectively,

$$E\left[\varepsilon_t^2\right] = \frac{\alpha_0}{1-\alpha_1},\ E\left[\varepsilon_t^4\right] = \frac{3\alpha_0^2}{(1-\alpha_1)^2}\frac{1-\alpha_1^2}{1-3\alpha_1^2}.$$

This implies that the kurtosis k of the unconditional distribution of ε_t,

$$k = \frac{E\left[\varepsilon_t^4\right]}{\left(E\left[\varepsilon_t^2\right]\right)^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2},$$

is always larger than 3, which is the value of kurtosis of the normal distribution. Therefore, under conditional normality for the ε_t process, the unconditional distribution of ε_t has fatter tails than the normal distribution and the ARCH model capture the "fat tails" property of the data. With financial data, ARCH(p) model captures the volatility clustering phenomenon, i.e. large (small) changes are followed by large (small) changes in the price of returns, but of unpredictable sign. This happens because ARCH models allow conditional variance to change over time as a function of past innovations ε_{t-i}, $i = 1, ..., p$.

Assuming that z_t are normally distributed, the parameter vector to be estimated in (3.1) and (3.2) is, for $p > 0$, $\boldsymbol{\theta} = (\alpha_0,\ \alpha_1, ...,\ \alpha_p)$. The likelihood for a sample of T observations $\mathbf{y} = (y_1, ..., y_T)$ can be written as

$$l_N\left(\mathbf{y}|\boldsymbol{\theta}\right) = (2\pi)^{-\frac{T}{2}} \prod_{t=1}^{T} \left\{ \left(\sigma_t^2\right)^{-1/2} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \right\},$$

where ε_t is given from the corresponding mean equation, and σ_t^2 is expressed via (3.2).

Although the likelihood function is highly nonlinear in the parameters, maximum likelihood estimates are derived using a scoring algorithm or any optimization routine. Under certain regularity conditions (see, for example, White, 1982; Gourieroux, Monfort and Trognon, 1984) the maximum likelihood estimates are strongly consistent and asymptotically normal. If the conditional distribution is correctly specified, then the asymptotic covariance matrix of the ML estimates is given by the inverse of the infor-

mation matrix. If the true conditional distribution is not normal, then Quasi MLE is still consistent (Weiss, 1986) but the asymptotic covariance matrix must be corrected. The small sample properties of ARCH estimators and tests have been analyzed by Engle, Hendry and Trumble (1985). ARCH-type models can also be estimated with Generalized Method of Moments; see, for example, Mark (1988), Bodurtha and Mark (1991). Bayesian inference procedures for ARCH models have been developed by Geweke (1989a, b), who used Monte Carlo methods to determine the exact posterior distribution, Polasek and Muller (1995) and Vrontos (1997), who used Markov Chain Monte Carlo methods.

A problem with ARCH models is that large lag orders are required in some applications, so that many parameters must be estimated under inequality restrictions. Engle (1982, 1983) used a linearly declining set of weights according to the formula

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sum_{i=1}^{p} w_i \varepsilon_{t-i}^2,$$

with

$$w_i = \frac{2\,(p+1-i)}{p\,(p+1)},$$

so that only two parameters must be estimated.

3.2.2 GARCH model

The extension of the ARCH model to the Generalized Autoregressive Conditional Heteroscedastic (GARCH) model resembles to the extension of the standard time series AR process to the general ARMA process. The GARCH(p,q) model (Bollerslev, 1986; Taylor, 1986) is given by the following formulation:

$$\varepsilon_t = z_t \sigma_t, \;\; t = 0, ..., T \tag{3.3}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \;\; t = 1, ..., T, \tag{3.4}$$

where z_t are iid with $E(z_t) = 0$ and $Var(z_t) = 1$; σ_t^2 is the (conditional) variance of the $\{\varepsilon_t\}$ process at time t, p, q are integers with $p > 0$, $q \geq 0$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, ..., p$, and $\beta_j \geq 0$, $j = 1, ..., q$. In (3.4) it is assumed that $\varepsilon_t = \sigma_t = 0$ for $t < 0$. These restrictions ensure a positive variance. However, Nelson and Cao (1992) showed that weaker sufficient conditions can be found. For example, in a GARCH(2,1) process, $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_1\beta_1 + \alpha_2 \geq 0$ are sufficient to guarantee that $\sigma_t^2 > 0$. They presented general results for GARCH(p,1) and GARCH (p,2), but a derivation for GARCH(p,q) with $q \geq 3$ is difficult. For $q = 0$, the process reduces to the ARCH(p) process, and for $p = q = 0$, ε_t is white noise.

In ARCH models the conditional variance is specified as a linear function of past innovations (ε_{t-i}). GARCH models allow the conditional variance to change over time as a function of past innovations (ε_{t-i}) and of past variances (σ_{t-j}^2). This family of models achieves more flexible lag structure than ARCH models and capture the volatility clustering phenomenon. In empirical applications of ARCH models a relatively long lag in the conditional variance equation is called for, and a large number of parameters must be estimated. GARCH models, due to the fact that include lagged conditional variance terms in the conditional variance equation, achieve a more parsimonious representation of higher order ARCH models.

Under conditional normality, an equivalent representation of the GARCH model is $\varepsilon_t|\Phi_{t-1} \sim N(0, \sigma_t^2)$, where Φ_{t-1} is the information set up to time $t-1$, and σ_t^2 is given by (3.4). The GARCH(p,q) model can be written as a non-Gaussian linear ARMA(max(p,q),q) model in the squares ε_t^2:

$$\begin{aligned}
\varepsilon_t^2 &= \sigma_t^2 + \left(\varepsilon_t^2 - \sigma_t^2\right) \\
&= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + v_t \\
&= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \left(\varepsilon_{t-j}^2 - v_{t-j}\right) + v_t
\end{aligned}$$

$$= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j}^2 - \sum_{j=1}^{q} \beta_j v_{t-j} + v_t,$$

where $v_t = \varepsilon_t^2 - \sigma_t^2 = \sigma_t^2 (z_t^2 - 1)$, and $E(v_t|\Phi_{t-1}) = 0$.

The unconditional moments of the GARCH process characterize the properties of the model. The derivation of the unconditional moments of the ARCH and GARCH processes is possible through extensive use of the Law of Iterated Expectations. The unconditional mean of a GARCH(p,q) error process ε_t with conditional variance (3.4) is given by $E(\varepsilon_t) = E[E(\varepsilon_t|\Phi_{t-1})]$. The GARCH model specifies that $E(\varepsilon_t|\Phi_{t-1}) = 0$ for all realizations of Φ_{t-1}, and therefore, the GARCH process has mean zero: $E(\varepsilon_t) = 0$. The unconditional variance of the GARCH(p,q) model is given by

$$\sigma_\varepsilon^2 = E(\varepsilon_t^2) = \alpha_0 / \left[1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j\right],$$

and the necessary and sufficient condition for the existence of the variance is $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$; see, for details, Bollerslev (1986). Although the variance of ε_t conditional on the information set Φ_{t-1} changes over time, unconditionally the GARCH process is homoscedastic. In many applications with high frequency financial data the estimate for $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j$ turns out to be very close to unity. This provides an empirical motivation for the integrated GARCH(p,q) or IGARCH(p,q) model (Engle and Bollerslev, 1986).

Bollerslev (1986) present the conditions for the existence of the moments of the GARCH(1,1) model under conditional normality. If $\alpha_1 + \beta_1 < 1$ and $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$, then the second and fourth order moment exist and are given by

$$E[\varepsilon_t^2] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}, \; E[\varepsilon_t^4] = \frac{3\alpha_0^2 (1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2)}.$$

Therefore, the kurtosis of the unconditional distribution of ε_t is given by

$$k = \frac{E\left[\varepsilon_t^4\right]}{\left(E\left[\varepsilon_t^2\right]\right)^2} = 3 + \frac{6\alpha_1^2}{1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2},$$

which is larger than 3, indicating that ε_t is leptokurtic. The condition on the parameters of the general GARCH(p,q) model for the existence of the fourth moment of ε_t is given by Karanasos (1999).

The autocovariances of the GARCH(p,q) process, for $k \geq 1$, are given by

$$\begin{aligned} E\left(\varepsilon_t \varepsilon_{t-k}\right) &= E\left[E\left(\varepsilon_t \varepsilon_{t-k} | \Phi_{t-1}\right)\right] \\ &= E\left[\varepsilon_{t-k} E\left(\varepsilon_t | \Phi_{t-1}\right)\right] = 0. \end{aligned}$$

Since the GARCH process is serially uncorrelated, with constant mean zero, the process is weakly stationary if the variance exist, that is if $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$.

An attractive characteristic of ARCH and GARCH models is that even though the conditional distribution of the errors is normal, the unconditional distribution is nonnormal with thicker tails than the normal distribution. In spite of this property, empirical work with ARCH and GARCH models indicated that the implied unconditional distribution of estimated ARCH and GARCH models were not sufficiently leptokurtic to represent the distribution of returns. Bollerslev (1987) used the conditional Student-t distribution, which allows for heavier tails than the normal distribution and, as the degrees of freedom goes to infinity, includes the normal distribution as a limiting case.

Maximum likelihood estimates for the parameters of GARCH model are obtained by the Berndt, Hall, Hall and Hausman (1974) algorithm; see, for example, Bollerslev (1986), Bollerslev (1987), Baillie and Bollerslev (1989) among several others. Fiorentini, Calzolari and Panattoni (1996) used analytic first and second derivatives of the log-likelihood for the estimation of the parameters of the GARCH model. They also made a comparison of various gradient algorithms that are used for the maximization of the GARCH Gaussian likelihood. The properties of the quasi-maximum likelihood estimator for dy-

namic models with time varying covariances and GARCH models have been investigated by Bollerslev and Wooldridge (1992) and Lee and Hansen (1994). The wide acceptance and the extended applicability of GARCH models is based on the fact that these models can accommodate many features of financial time series; see, for example, Akgiray (1989), Bera and Higgins (1997) among several others. A Bayesian analysis of GARCH models was proposed by Bauwens and Lubrano (1998), Muller and Pole (1999), Bos, Mahieu and van Dijk (1999), Vrontos, Dellaportas and Politis (2000) who used MCMC methods to extract the posterior distribution of the model parameters. Bauwens, Bos and van Dijk (1999) used Adaptive Polar Sampling within a Bayesian analysis of a GARCH-mixture model for the evaluation of the value-at-Risk of the return of the Dow Jones stock index.

3.2.3 EGARCH model

ARCH and GARCH models capture the volatility clustering phenomenon. This feature of GARCH models accounts for both their theoretical appeal and their empirical success. On the other hand, their functional forms impose important limitations. GARCH models assume that only the magnitude and not the sign (positivity or negativity) of the error process determines future σ_t^2. That is, the conditional variance σ_t^2 is symmetric in the lagged ε_t's. A symmetric conditional variance function may be inappropriate for modeling the volatility of returns of stocks because it cannot capture the "leverage effect", which is the negative correlation between returns and future volatility. In other words, volatility tend to rise in response to "bad news" and to fall in response to "good news". Other possible limitations of GARCH models concern the interpretation of the persistence of shocks to conditional variance and the nonnegativity constraints on the parameters of the model, which are imposed to ensure that σ_t^2 remains positive. These constraints can create difficulties in the estimation procedure.

Nelson (1991) introduced the Exponential GARCH (EGARCH) model, that relax the above limitations, and used the Generalized Error Distribution (Box and Tiao, 1973), which account for the fat tails property of financial data. The density of a Generalized

Error Distribution (GED) random variate with mean 0 and variance 1, is given by

$$p(z_t) = \frac{v \exp\left[-\frac{1}{2}\left|\frac{z_t}{\lambda}\right|^v\right]}{\lambda 2^{\left(1+\frac{1}{v}\right)}\Gamma\left(\frac{1}{v}\right)}, \quad -\infty < z_t < +\infty, \; v > 0$$

$$\lambda = \left[\frac{2^{(-2/v)}\Gamma\left(\frac{1}{v}\right)}{\Gamma\left(\frac{3}{v}\right)}\right]^{1/2},$$

where $\Gamma(.)$ is the gamma function, and v is a tail-thickness parameter. If $v = 2$ we obtain the normal distribution, and for $v < 2$ and $v > 2$ the distribution of z_t has thicker and thinner tails than the normal distribution respectively. The EGARCH(p,q) model is given by the following formulation

$$\varepsilon_t = z_t \sigma_t, \; t = 0, ..., T,$$

$$\ln\left(\sigma_t^2\right) = \alpha_0 + \sum_{j=1}^{q} \beta_j \ln\left(\sigma_{t-j}^2\right) + \sum_{k=1}^{p} \left[\theta_k z_{t-k} + \gamma_k \left(|z_{t-k}| - E|z_{t-k}|\right)\right], \tag{3.5}$$

where z_t are iid and follow GED with mean 0 and variance 1, σ_t^2 is the conditional variance of the $\{\varepsilon_t\}$ process at time t, $\sigma_t = z_t = 0$ for $t < 0$. Under the assumption of generalized error distribution for z_t, $E|z_{t-k}|$ is given by

$$E|z_{t-k}| = \frac{\Gamma\left(\frac{2}{v}\right)}{\left[\Gamma\left(\frac{1}{v}\right)\Gamma\left(\frac{3}{v}\right)\right]^{1/2}}.$$

If the coefficients β_j, $j = 1, ..., q$ are equal to zero then the model is called Exponential ARCH (EARCH), since the conditional variance equation contains only past standardized observations.

Modeling the logarithm of the conditional variance according to (3.5) solves the limitations of GARCH models. The EGARCH models capture the "leverage effect". To accommodate the asymmetric relation between stock returns and future volatility, a function of both the magnitude and the sign of z_t is used. This function is the linear

combination of z_{t-k} and of $|z_{t-k}|$, and is given by

$$g\left(z_{t-k}\right) = \theta_k z_{t-k} + \gamma_k \left(|z_{t-k}| - E\left|z_{t-k}\right|\right).$$

Assume, for example, that $\theta_k = 0$ and $\gamma_k > 0$. Then, the innovation in $\ln(\sigma_t^2)$ is positive (negative) when the absolute value of z_{t-k} is larger (smaller) than its expected value, in other words, when $|z_{t-k}| - E\left|z_{t-k}\right|$ is positive (negative). This term represents the magnitude effect. Assume now that $\theta_k < 0$ and $\gamma_k = 0$. The innovation in conditional variance is now positive (negative) when returns z_{t-k} are negative (positive). The term $\theta_k z_{t-k}$ represents the sign effect. The specification (3.5) allows the conditional variance to be positive, and therefore, there are no restrictions on the model parameters. Details for the properties and the stationarity of the Exponential ARCH model are presented in Nelson (1991).

The EGARCH model allows lagged shocks to have an asymmetric effect on the conditional volatility. The performance of this model has been studied by Pagan and Schwert (1990) using U.S. monthly stock returns. Engle and Ng (1993) measured how new information is incorporated into volatility estimates using daily Japanese stock return data, and a comparative study of different volatility models was presented by Taylor (1994) using daily exchange rates. The EGARCH model has been used by Poon and Taylor (1992), Day and Lewis (1992), Kuwahara and Marsh (1992), and a Bayesian analysis of EGARCH models is in Vrontos, Dellaportas and Politis (2000).

3.2.4 Extensions of the ARCH-GARCH model

In this section we present some alternative functional forms for the conditional variance that have been proposed in the literature. Engle (1982) considered that the conditional variance function was linear in the squared errors and that the conditional distribution was normal. However, he mentioned that "it is likely that other formulations of the variance model may be more appropriate for particular applications". He suggested that

two alternative formulations are the exponential and the absolute value models:

$$\sigma_t^2 = \exp\left[\alpha_0 + \alpha_1 \varepsilon_{t-1}^2\right] \text{ and } \sigma_t^2 = \alpha_0 + \alpha_1 \left|\varepsilon_{t-1}\right|.$$

In the absolute model the parameters must be positive, while in the exponential model no inequality restrictions are required for the α_i's to ensure that the conditional variance is strictly positive. Taylor (1986) suggested that the conditional standard deviation σ_t can be modeled by

$$\sigma_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \left|\varepsilon_{t-i}\right|.$$

Geweke (1986) and Pantula (1986) suggested the log ARCH model

$$\ln\left(\sigma_t^2\right) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \ln\left(\varepsilon_{t-i}^2\right), \tag{3.6}$$

which ensures that the conditional variance is positive for all values of α_i, $i = 0, 1, ..., p$, but there is problem when the value of ε_{t-i}, $i = 0, 1, ..., p$ is zero.

Higgins and Bera (1992) proposed a general class of models, called nonlinear ARCH (NARCH), which is given by

$$\sigma_t^2 = \left[\phi_0 \left(\sigma^2\right)^{\delta} + \phi_1 \left(\varepsilon_{t-1}^2\right)^{\delta} + ... + \phi_p \left(\varepsilon_{t-p}^2\right)^{\delta}\right]^{1/\delta}, \tag{3.7}$$

where $\sigma^2 > 0$, $\phi_i \geq 0$, $i = 0, 1, ..., p$, $\delta > 0$, and the ϕ_i are such that $\sum_{i=0}^{p} \phi_i = 1$. The conditional variance (3.7) has $p + 3$ parameters, but the restriction $\sum_{i=0}^{p} \phi_i = 1$ reduces the number of parameters by one. This model includes the linear ARCH (3.2) model as a special case when $\delta = 1$, and the log ARCH model (3.6) as a limiting case. See, for details, Higgins and Bera (1992). The above models can be extended to GARCH models.

Glosten, Jagannathan and Runkle (1993) used the following formulation

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-1}^2 I\left(\varepsilon_{t-1} > 0\right), \tag{3.8}$$

where I is the indicator function and thus the impact of ε_{t-1}^2 on the conditional varinace σ_t^2 is different when ε_{t-1} is positive than when ε_{t-1} is negative. This model can capture the asymmetric or "leverage effect", in which good news and bad news have different predictability for future volatility. The asymmetric GARCH model has been analyzed by Bauwens and Lubrano (1999) using Bayesian methodology. Zakoian (1994) adopted a different approach. He specified the conditional standard deviation based on a result of Davidian and Carroll (1987) that in the case of nonnormal distributions, absolute residuals yield more efficient variance estimates than squared residuals. He called his model Threshold GARCH(p,q) model which is given by

$$\sigma_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i^+ \varepsilon_{t-i}^+ - \alpha_i^- \varepsilon_{t-i}^- + \sum_{j=1}^{q} \beta_j \sigma_{t-j},$$

where $\varepsilon_t^+ = \max(\varepsilon_t, 0)$ and $\varepsilon_t^- = \min(\varepsilon_t, 0)$, $\alpha_0 > 0$, $\alpha_i^+ \geq 0$, $\alpha_i^- \geq 0$, $i = 1, ..., p$, and $\beta_j \geq 0$, $j = 1, ..., q$. This formulation is known as threshold GARCH because the coefficient of ε_{t-i} changes when ε_{t-i} crosses the threshold of zero. For $\alpha_i^+ = \alpha_i^- = \alpha_i$ for all $i = 1, ..., p$, the conditional standard deviation is given by

$$\sigma_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \left|\varepsilon_{t-i}\right| + \sum_{j=1}^{q} \beta_j \sigma_{t-j}.$$

Instead of modeling the conditional variance using a piecewise linear function, Gourieroux and Monfort (1992) used a piecewise constant function over a partition of the set of values of the data. They introduced the qualitative threshold ARCH, or QTARCH, model which is given by

$$\varepsilon_t = \sum_{j=1}^{J} \alpha_j I\left(\varepsilon_{t-1} \in A_j\right) + \sum_{j=1}^{J} \beta_j I\left(\varepsilon_{t-1} \in A_j\right) z_t,$$

where ε_t is the series of interest, A_j, $j = 1, ..., J$ is a partition over the support of the data, $I(.)$ is the indicator function, i.e. is 1 if $\varepsilon_{t-1} \in A_j$, α_j and β_j, $j = 1, ..., J$ are

parameters, and z_t is iid with mean 0 and variance 1. The α_j parameters govern the mean and the β_j parameters govern the variance of the $\{\varepsilon_t\}$ process.

Engle and Ng (1993) provided a summary of asymmetric models and proposed some new formulations. They defined the news impact curve which measures how new information is incorporated into volatility estimates. The proposed parametric models are

$$\sigma_t^2 = \alpha_0 + \alpha_1 \left(\varepsilon_{t-1} + \gamma\right)^2 + \beta_1 \sigma_{t-1}^2$$

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 \left(\varepsilon_{t-1} + \gamma \sigma_{t-1}\right)^2$$

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 \left(\varepsilon_{t-1}/\sigma_{t-1} + \gamma\right)^2 .$$

They also proposed a partially nonparametric model (PNP), which is similar in spirit with the model of Gourieroux and Monfort (1992), but they specified the variance function using a piecewise linear function. They divided the range of $\{\varepsilon_t\}$ into m intervals, where the boundaries of the interval are $\{\tau_{m^-}, ..., \tau_{-1}, 0, \tau_1, ..., \tau_{m^+}\}$, m^- is the number of intervals in the range where ε_{t-1} is negative, m^+ is the number of intervals in the range where ε_{t-1} is positive, and $m = m^- + m^+$. The variance function is given by

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \sum_{i=0}^{m^+} \theta_i P_{it-1} \left(\varepsilon_{t-1} - \tau_i\right) + \sum_{i=0}^{m^-} \delta_i N_{it-1} \left(\varepsilon_{t-1} - \tau_{-i}\right), \qquad (3.9)$$

where P_{it-1} is 1 if $\varepsilon_{t-1} > \tau_i$ and 0 otherwise, and N_{it-1} is 1 if $\varepsilon_{t-1} < \tau_{-i}$ and 0 otherwise. α_0, β_1, θ_i, $i = 0, 1, ..., m^+$, δ_i, $i = 0, 1, ..., m^-$ are constant parameters. From (3.9) σ_t^2 is linear with a different slope over each interval. For example, between 0 and τ_1 the slope is θ_0, between τ_1 and τ_2 the slope is $\theta_0 + \theta_1$, generally between τ_i and τ_{i+1} the slope is $\theta_1 + ... + \theta_i$. Engle and Ng (1993) chose the τ_i's to be multiples of the unconditional standard deviation of the dependent variable. The statistical properties of a family of GARCH models is presented in He and Terasvirta (1999).

Engle, Lilien and Robins (1987) introduced the ARCH-M model which extends the ARCH model and allows the conditional mean return to be a function of volatility. In a

simple form the model can be written as

$$\begin{aligned} y_t &= a + bf\left(\sigma_t^2\right) + \boldsymbol{\gamma}^{'}\mathbf{x}_t + \varepsilon_t \\ \varepsilon_t|\Phi_{t-1} &\sim N\left(0, \sigma_t^2\right) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2, \end{aligned}$$

where y_t is the observation process, $f\left(\sigma_t^2\right)$ is a function of the conditional variance, $\mathbf{x}_t$ is a vector of exogenous or lagged dependent variables, ε_t is the innovation process, Φ_{t-1} is the information set up to time $t-1$, σ_t^2 is the conditional variance of the process which follows a linear ARCH(p) model, and a, b, $\boldsymbol{\gamma}$, α_0, α_i, $i = 1, ..., p$ are parameters to be estimated. Any other formulation can be used to model the conditional variance. This model is useful in financial applications since it relates the returns of a stock and volatility. Empirical studies on this relationship have been done, for example, by French, Schwert and Stambaugh (1987), Nelson (1991), Poon and Taylor (1992), Glosten, Jagannathan and Runkle (1993) among several others.

3.3 Multivariate ARCH-type models

The extension of univariate ARCH-type models to a multivariate framework and the estimation of time-varying covariances between asset returns is crucial for asset pricing, portfolio analysis, and risk management. The development of multivariate models, with a view to modelling the covariance pattern in volatility, is lagging significantly behind the development of univariate time-varying volatility models. Two major problems related with multivariate ARCH and GARCH models are the large number of parameters to be estimated, and the difficulty of the estimation due to the positive definiteness restrictions of the covariance matrix. We consider having observed data of the form

$$\mathbf{y}_t,\ t = 1, ..., T,$$

where each $\mathbf{y}_t = (y_{1,t}, ..., y_{N,t})$ is a $N \times 1$ vector. Suppose that the mean equation and the conditional distribution of the innovation process are given by the following equations:

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t,$$

$$\boldsymbol{\varepsilon}_t | \Phi_{t-1} \sim N_N \left(\mathbf{0}, \boldsymbol{\Sigma}_t \right),$$

where $\boldsymbol{\mu}$ is a $N \times 1$ vector of constants, $\boldsymbol{\varepsilon}_t$ is a $N \times 1$ innovation vector, Φ_{t-1} is the information set up to time $t-1$, $\boldsymbol{\Sigma}_t$ is $N \times N$ covariance matrix with elements $\sigma_{i,t}^2$ and $\sigma_{ij,t}$, $i = 1, ..., N$, $j = i+1, ..., N$, where $\sigma_{i,t}^2$ is the variance of the $i-th$ variable at time t, and $\sigma_{ij,t}$ is the covariance between $i-th$ and $j-th$ variable at time t. Different time-varying covariance models impose different restrictions on how past shocks affect the forecasted covariance matrix. In this section we present some well known multivariate ARCH-GARCH models that have been proposed in the literature.

3.3.1 Multivariate ARCH model

Kraft and Engle (1982) introduced the multivariate ARCH(p) model which allows the elements of the conditional covariance matrix to change over time. The model can be written as:

$$vech \left(\boldsymbol{\Sigma}_t \right) = \mathbf{C} + \sum_{i=1}^{p} \mathbf{A}_i vech \left(\boldsymbol{\varepsilon}_{t-i} \boldsymbol{\varepsilon}_{t-i}^{'} \right), \tag{3.10}$$

where $\boldsymbol{\Sigma}_t$ is $N \times N$ covariance matrix, $vech\left(.\right)$ denotes the column stacking operator of the lower portion of a symmetric matrix, $\mathbf{C}$ is a $\frac{N(N+1)}{2} \times 1$ vector, $\mathbf{A}_i$, $i = 1, ..., p$ are $\frac{N(N+1)}{2} \times \frac{N(N+1)}{2}$ matrices. This model involves a total of $\frac{N(N+1)}{2} + p\left(\frac{N(N+1)}{2}\right)^2$ parameters in the covariance equation. Representation (3.10) is called 'vech representation' of a multivariate ARCH model. For $N = 2$ and $p = 1$, equation (3.10) can be written as:

$$vech \left(\boldsymbol{\Sigma}_t \right) = \mathbf{C} + \mathbf{A}_1 vech \left(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}^{'} \right)$$

or

$$\begin{bmatrix} \sigma_{1,t}^2 \\ \sigma_{21,t} \\ \sigma_{2,t}^2 \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix}. \tag{3.11}$$

The covariance matrix $\mathbf{\Sigma}_t$ must be positive definite for all realizations of $\boldsymbol{\varepsilon}_{t-1}$. The positive definiteness of $\mathbf{\Sigma}_t$ places restrictions on the elements of vector $\mathbf{C}$ and on the rows and columns of $\mathbf{A}_1$. Kraft and Engle (1982) showed that the necessary conditions for the positive definiteness of $\mathbf{\Sigma}_t$, in the bivariate ARCH model (3.11) are

$$\begin{aligned} c_{11} &> 0,\ c_{22} > 0,\ c_{11}c_{22} - c_{21}^2 > 0, \\ a_{11} &\geq 0,\ a_{13} \geq 0,\ a_{11}a_{13} - \frac{1}{4}a_{12}^2 \geq 0, \\ a_{31} &\geq 0,\ a_{33} \geq 0,\ a_{31}a_{33} - \frac{1}{4}a_{32}^2 \geq 0, \\ a_{11}a_{33} - a_{22}^2 &\geq 0,\ a_{11}a_{31} - a_{21}^2 \geq 0,\ a_{13}a_{33} - a_{23}^2 \geq 0. \end{aligned}$$

For multivariate systems of larger dimension, analogous constraints on the rows and columns of matrix $\mathbf{A}_1$ must be imposed. The derivation of these restrictions is given in Theorem 1 of Kraft and Engle (1982). Note, that in (3.11) and (3.10) each $\sigma_{i,t}^2$ and $\sigma_{ij,t}$ depends on lagged squared residuals and cross-products of all the variables in the system. A simple assumption is to specify that the variances depend only on its own past squared residuals and that the covariances depend only on past cross products. In this case the matrices $\mathbf{A}_i,\ i = 1, ..., p$ in (3.10) are diagonal. Then the number of parameters is $(p+1)\frac{N(N+1)}{2}$. For $N = 2$ and $p = 1$, the covariance equation can be written as

$$\begin{bmatrix} \sigma_{1,t}^2 \\ \sigma_{21,t} \\ \sigma_{2,t}^2 \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix}. \tag{3.12}$$

Positive definiteness of $\mathbf{\Sigma}_t$ in (3.12) requires that

$$\begin{aligned} c_{11} &> 0,\ c_{22} > 0,\ c_{11}c_{22} - c_{21}^2 > 0, \\ a_{11} &\geq 0,\ a_{33} \geq 0,\ a_{11}a_{33} - a_{22}^2 \geq 0. \end{aligned}$$

This 'diagonal representation' was used by Engle, Granger and Kraft (1984) for their analysis of the problem of combining competing inflation forecasts using time varying weights. Estimates of the model parameters was obtained by using the method of scoring.

3.3.2 Multivariate GARCH model

Bollerslev, Engle and Wooldridge (1988) introduced the multivariate GARCH(p,q) model. A version of this model can be written using the 'vech representation' as:

$$vech\left(\mathbf{\Sigma}_t\right) = \mathbf{C} + \sum_{i=1}^{p} \mathbf{A}_i vech\left(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-i}^{'}\right) + \sum_{j=1}^{q} \mathbf{B}_j vech\left(\mathbf{\Sigma}_{t-j}\right), \tag{3.13}$$

where $\mathbf{\Sigma}_t$ is $N \times N$ covariance matrix, $vech\left(.\right)$ denotes the column stacking operator of the lower portion of a symmetric matrix, $\mathbf{C}$ is a $\frac{N(N+1)}{2} \times 1$ vector, $\mathbf{A}_i$, $i = 1, ..., p$ and $\mathbf{B}_j$, $j = 1, ..., q$ are $\frac{N(N+1)}{2} \times \frac{N(N+1)}{2}$ matrices. This model has $\frac{N(N+1)}{2} + (p+q)\left(\frac{N(N+1)}{2}\right)^2$ parameters to be estimated. For $N = 2$ and $p = q = 1$, equation (3.13) can be written as:

$$vech\left(\mathbf{\Sigma}_t\right) = \mathbf{C} + \mathbf{A}_1 vech\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^{'}\right) + \mathbf{B}_1 vech\left(\mathbf{\Sigma}_{t-1}\right) \tag{3.14}$$

or

$$\begin{bmatrix} \sigma_{1,t}^2 \\ \sigma_{21,t} \\ \sigma_{2,t}^2 \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix} \begin{bmatrix} \sigma_{1,t-1}^2 \\ \sigma_{21,t-1} \\ \sigma_{2,t-1}^2 \end{bmatrix}.$$

Model (3.13) allows each element of the conditional covariance matrix to depend on lagged values of the squares and cross-products of all the variables in the system, as well

as lagged values of the elements of the covariance matrix. This model is very general and has a large number of parameters. A natural simplification is to assume that the variances depend only on its own past squared residuals and past variances and that the covariances depend only on past cross products and past covariances. Therefore, the matrices $\mathbf{A}_i,\ i = 1, ..., p$ and $\mathbf{B}_j,\ j = 1, ..., q$ in (3.13) are diagonal. Then the number of parameters is $(p + q + 1)\frac{N(N+1)}{2}$. For $N = 2$ and $p = q = 1$, the covariance equation can be written as

$$\begin{bmatrix} \sigma_{1,t}^2 \\ \sigma_{21,t} \\ \sigma_{2,t}^2 \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{22} & 0 \\ 0 & 0 & \beta_{33} \end{bmatrix} \begin{bmatrix} \sigma_{1,t-1}^2 \\ \sigma_{21,t-1} \\ \sigma_{2,t-1}^2 \end{bmatrix}.$$

Engle and Kroner (1995) presented the necessary and sufficient conditions for the covariance stationarity of the multivariate GARCH model (3.13). They proved that $\{\boldsymbol{\varepsilon}_t\}$ is covariance stationary if and only if the eigenvalues of $\sum_{i=1}^{p} \mathbf{A}_i + \sum_{j=1}^{q} \mathbf{B}_j$ are less than one in modulus. Focusing on the bivariate GARCH(1,1) model (3.14), $\{\boldsymbol{\varepsilon}_t\}$ is covariance stationary if and only if the eigenvalues of $\mathbf{A}_1 + \mathbf{B}_1$ are less than one in modulus, and the unconditional covariance matrix in a 'vech representation' is given by

$$E\left[vech\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{'}\right)\right] = \left[I - \mathbf{A}_1 - \mathbf{B}_1\right]^{-1}\mathbf{C}.$$

The 'diagonal representation' was used by Bollerslev, Engle and Wooldridge (1988), Baillie and Myers (1991), Bera, Garcia and Roh (1991), Kroner and Lastrapes (1993). Maximum likelihood estimates for the parameters of multivariate GARCH model (3.13) are obtained by the Berndt, Hall, Hall and Hausman (1974) algorithm along with numerical first order derivatives. Bollerslev and Engle (1993) extended the idea of Integrated GARCH model to a multivariate framework. In the IGARCH models shocks to the conditional variance are persistent, in the sense that they are important for forecasts of all horizons. According to Bollerslev and Engle (1993), although many financial time

series may exhibit persistence in their conditional variances, a linear combination of such variables may have no persistence in variance. In this case the variables are defined co-persistent in variance, and the co-persistent linear combination may be interpreted as a long-run relationship.

The covariance matrix $\mathbf{\Sigma}_t$ must be positive definite for all values of $\boldsymbol{\varepsilon}_t$. In the 'vech representation', and even in the 'diagonal representation', this restriction is not easy to check and also difficult to impose at the estimation procedure.

3.3.3 The BEKK model

Engle and Kroner (1995) proposed a new parameterization of multivariate GARCH model, known as 'BEKK representation'. It is characterized by the following equation:

$$\mathbf{\Sigma}_t = \mathbf{C}_0^{*'}\mathbf{C}_0^* + \sum_{k=1}^{K}\sum_{i=1}^{p}\mathbf{A}_{ik}^{*'}\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-i}^{'}\mathbf{A}_{ik}^* + \sum_{k=1}^{K}\sum_{i=1}^{q}\mathbf{B}_{ik}^{*'}\mathbf{\Sigma}_{t-i}\mathbf{B}_{ik}^*, \tag{3.15}$$

where $\mathbf{\Sigma}_t$ is $N \times N$ covariance matrix, $\mathbf{C}_0^*$, $\mathbf{A}_{ik}^*$, $\mathbf{B}_{ik}^*$ are $N \times N$ parameter matrices with $\mathbf{C}_0^*$ triangular, and the generality of the process is determined by K. To illustrate the BEKK model, consider the simple GARCH(1,1) model, with $K = 1$:

$$\mathbf{\Sigma}_t = \mathbf{C}_0^{*'}\mathbf{C}_0^* + \mathbf{A}_{11}^{*'}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^{'}\mathbf{A}_{11}^* + \mathbf{B}_{11}^{*'}\mathbf{\Sigma}_{t-1}\mathbf{B}_{11}^*. \tag{3.16}$$

In the bivariate case the model becomes

$$\begin{aligned}\begin{bmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{2,t}^2 \end{bmatrix} &= \mathbf{C}_0^{*'}\mathbf{C}_0^* + \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}\varepsilon_{1,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix} \\ &\quad + \begin{bmatrix} \beta_{11}^* & \beta_{12}^* \\ \beta_{21}^* & \beta_{22}^* \end{bmatrix}' \begin{bmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{21,t-1} & \sigma_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} \beta_{11}^* & \beta_{12}^* \\ \beta_{21}^* & \beta_{22}^* \end{bmatrix}. \end{aligned} \tag{3.17}$$

The BEKK model provides a solution to the positive definiteness problem. Due to the fact that the second and third term of the right hand side of equation (3.16) or

(3.17) are expressed in quadratic forms, the positive definiteness of the covariance matrix is guaranteed, provided that $\mathbf{C}_0^{*'}\mathbf{C}_0^*$ is positive definite. The number of parameters to be estimated in model (3.16) is $\frac{N(N+1)}{2} + 2N^2$. Engle and Kroner (1995) presented the relation between the BEKK model and the multivariate vector/diagonal GARCH model. They also developed necessary and sufficient conditions for the covariance stationarity of the BEKK model (3.15). $\{\boldsymbol{\varepsilon}_t\}$ is covariance stationary if and only if the eigenvalues of $\sum_{k=1}^{K}\sum_{i=1}^{p}\left(\mathbf{A}_{ik}^* \otimes \mathbf{A}_{ik}^*\right)' + \sum_{k=1}^{K}\sum_{i=1}^{q}\left(\mathbf{B}_{ik}^* \otimes \mathbf{B}_{ik}^*\right)'$ are less than one in modulus. In the case of the BEKK model (3.16) using GARCH(1,1) and $K = 1$, $\{\boldsymbol{\varepsilon}_t\}$ is covariance stationary if and only if the eigenvalues of $\left(\mathbf{A}_{11}^* \otimes \mathbf{A}_{11}^*\right)' + \left(\mathbf{B}_{11}^* \otimes \mathbf{B}_{11}^*\right)'$ are less than one in modulus, and the unconditional covariance matrix in a 'vec representation' is given by

$$E\left[vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\right] = \left[I - \left(\mathbf{A}_{11}^* \otimes \mathbf{A}_{11}^*\right)' - \left(\mathbf{B}_{11}^* \otimes \mathbf{B}_{11}^*\right)'\right]^{-1} vec\left(\mathbf{C}_0^{*'}\mathbf{C}_0^*\right),$$

where $vec\,(.)$ is the vector operator that stacks the columns of a matrix. The estimation of the model parameters is addressed by using the Berndt, Hall, Hall and Hausman (1974) algorithm.

3.3.4 Multivariate GARCH model with constant conditional correlations

Bollerslev (1990) proposed a simple multivariate conditional heteroscedastic time series model. This model has time-varying conditional variances and covariances, but constant conditional correlations. He expressed the conditional covariance matrix $\boldsymbol{\Sigma}_t$ as

$$\boldsymbol{\Sigma}_t = \mathbf{D}_t\mathbf{R}\mathbf{D}_t, \tag{3.18}$$

where $\mathbf{R}$ is the $N \times N$ time invariant correlation matrix with elements ρ_{ij}, $i = 1, ..., N$, $j = i+1, ..., N$, $i \neq j$, and $\mathbf{D}_t$ is the $N \times N$ diagonal matrix with elements $\sigma_{i,t}$, $i = 1, ..., N$. The individual variances are assumed to be standard univariate GARCH(p,q) models

given by

$$\sigma_{i,t}^2 = a_{0i} + \sum_{j=1}^{p} a_{ij}\varepsilon_{i,t-j}^2 + \sum_{j=1}^{q} \beta_{ij}\sigma_{i,t-j}^2,\ i = 1, ..., N. \tag{3.19}$$

The covariance matrix $\mathbf{\Sigma}_t$ will be positive definite for all t if and only if the constant conditional correlation matrix is positive definite and the conditional variances are well defined. These restrictions are very easy to impose, compared to other parameterizations for the varying covariance matrix. The number of parameters in the covariance equation of this multivariate GARCH model is the $\frac{N(N-1)}{2} + N\,(1 + p + q)$, constructed by the ρ_{ij}, a_{0i}, a_{ij}, and β_{ij}. To illustrate the multivariate GARCH model (3.18) and (3.19) with constant conditional correlations consider the simple case with $N = 2$, and $p = q = 1$. The elements of the conditional covariance matrix $\mathbf{\Sigma}_t$ are given by:

$$\sigma_{i,t}^2 = a_{0i} + a_{i1}\varepsilon_{i,t-1}^2 + \beta_{i1}\sigma_{i,t-1}^2,\ i = 1, 2$$

$$\sigma_{12,t} = \rho_{12}\left(\sigma_{1,t}^2\sigma_{2,t}^2\right)^{1/2},\ -1 \leq \rho_{12} \leq 1.$$

For positive definiteness of $\mathbf{\Sigma}_t$ in this bivariate case, we need $a_{0i} > 0$, $a_{i1} \geq 0$, $\beta_{i1} \geq 0$, $i = 1, 2$ and $-1 < \rho_{12} < 1$. For finite variance and stationarity, it is also necessary to impose that $a_{i1} + \beta_{i1} < 1$ for $i = 1, 2$. Many applications of multivariate models use the above representation; see, for example, Giovannini and Jorion (1989), Baillie and Bollerslev (1990), Schwert and Seguin (1990), Kroner and Claessens (1991), Kroner and Sultan (1991), Ng (1991), Bekaert and Hodrick (1993), Turtle, Buse and Korkie (1994) among others.

Jeantheau (1998) allowed the conditional variances to depend on lagged values of the squares of the error process of all variables in the system, as well as lagged values of its own variance and on lagged values of the other variances. He expressed the conditional covariance matrix $\mathbf{\Sigma}_t$ as in (3.18), where the elements $\sigma_{i,t}$, $i = 1, ..., N$ of the $N \times N$

diagonal matrix $\mathbf{D}_t$ satisfy, for all i, the following relation:

$$\begin{bmatrix} \sigma_{1,t}^2 \\ \vdots \\ \sigma_{N,t}^2 \end{bmatrix} = \mathbf{A}_0 + \sum_{i=1}^{p} \mathbf{A}_i \begin{bmatrix} \varepsilon_{1,t-i}^2 \\ \vdots \\ \varepsilon_{N,t-i}^2 \end{bmatrix} + \sum_{j=1}^{q} \mathbf{B}_j \begin{bmatrix} \sigma_{1,t-j}^2 \\ \vdots \\ \sigma_{N,t-j}^2 \end{bmatrix}, \tag{3.20}$$

where $\mathbf{A}_0$ is a $N \times 1$ vector of constants, $\mathbf{A}_i$, $i = 1, ..., p$ and $\mathbf{B}_j$, $j = 1, ..., q$ are $N \times N$ parameter matrices, whose elements are positive. The number of parameters in the covariance equation of this multivariate GARCH model is the $\frac{N(N-1)}{2} + N(1 + Np + Nq)$, constructed by the ρ_{ij}, $\mathbf{A}_0$, $\mathbf{A}_i$, and $\mathbf{B}_j$. To illustrate the multivariate GARCH model (3.20) with constant conditional correlations consider the simple case with $N = 2$, and $p = q = 1$. The elements of the conditional covariance matrix $\mathbf{\Sigma}_t$ are given by:

$$\sigma_{1,t}^2 = a_{10} + a_{11}\varepsilon_{1,t-1}^2 + a_{12}\varepsilon_{2,t-1}^2 + \beta_{11}\sigma_{1,t-1}^2 + \beta_{12}\sigma_{2,t-1}^2$$

$$\sigma_{2,t}^2 = a_{20} + a_{21}\varepsilon_{1,t-1}^2 + a_{22}\varepsilon_{2,t-1}^2 + \beta_{21}\sigma_{1,t-1}^2 + \beta_{22}\sigma_{2,t-1}^2$$

$$\sigma_{12,t} = \rho_{12}\left(\sigma_{1,t}^2\sigma_{2,t}^2\right)^{1/2}, \quad -1 \leq \rho_{12} \leq 1.$$

For positive definiteness of $\mathbf{\Sigma}_t$ in this bivariate case, we need $a_{0i} > 0$, $a_{ij} \geq 0$, $\beta_{ij} \geq 0$, $i, j = 1, 2$ and $-1 < \rho_{12} < 1$.

The assumption of constant conditional correlation is found to be reasonable for some empirical studies. However, this assumption may seem restrictive since the time variation of the conditional correlations of asset returns is a well established stylized fact. For example, correlations between stock markets tend to increase during periods of high volatility, while some economic variables such as dividend yields and interest rates may contribute to the information set about correlations. Christodoulakis and Satchell (1998) proposed a new bivariate model with time varying conditional variances and correlations, the evolution of which is generated by a discrete time stochastic process. Their model can be considered as an expansion of the constant conditional correlation model of Bollerslev

(1990). According to their model the conditional covariance matrix is given by (3.18), where the conditional variances are assumed to follow univariate ARCH processes (of any type), and the element of correlation matrix $\rho_{12,t}$ evolves over time. In order to ensure that the conditional correlation $\left|\rho_{12,t}\right|$ is always less than one, they adopted the Fisher-z transformation of correlation coefficient

$$z_{12,t}\left(\rho_{12,t}\right) = \frac{1}{2}\ln\left(\frac{1+\rho_{12,t}}{1-\rho_{12,t}}\right)$$

which implies that

$$\rho_{12,t}\left(z_{12,t}\right) = \frac{\exp\left(2z_{12,t}\right)-1}{\exp\left(2z_{12,t}\right)+1}.$$

They change $z_{12,t}$ over time according to a linear function of the available information set. That is

$$z_{12,t} = \phi_0 + \sum_{i=1}^{p}\phi_i\overline{\varepsilon}_{t-i},$$

where $\overline{\varepsilon}_t = \left(\varepsilon_{1,t}\varepsilon_{2,t}\right)/\left(\sigma_{1,t}\sigma_{2,t}\right)$. This model was called Correlated ARCH (CorrARCH) of order p. An expansion to Correlated GARCH model is obvious. They estimated the model parameters using maximum likelihood techniques, under alternatives distributional assumptions of conditional normality and of bivariate-t distribution.

3.3.5 Latent Factor ARCH Model

Diebold and Nerlove (1989) proposed a multivariate Latent Factor ARCH model for the analysis of exchange rates. The motivation for this model is that the ARCH effects capture the volatility clustering phenomenon and the factor structure captures the commonality in volatility movements across exchange rates. The model can be written as

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\beta}F_t + \mathbf{v}_t$$

$$F_t|\Phi_{t-1} \sim N\left(0,\sigma_t^2\right)$$

$$\sigma_t^2 = a_0 + \theta \sum_{i=1}^{12} (13 - i) F_{t-i}^2,$$

where $\boldsymbol{\varepsilon}_t$ is a $N \times 1$ innovation vector, $\boldsymbol{\beta}$ is a $N \times 1$ parameter vector, F is a common factor which influences all the time series of interest, $\mathbf{v}_t$ is a $N \times 1$ vector of 'unique factors' which reflect the specific shocks, Φ_{t-1} is the information set up to time $t-1$, which contains the values of the common factor F at previous times $t-1$, $t-2$, ..., σ_t^2 is the variance of the common factor at time t, which is given by an ARCH(12) model. Other ARCH-type specifications can be used for modeling the variance of common factor. In addition $E(F_t|\Phi_{t-1}) = E(v_{jt}|\Phi_{t-1}) = 0$, for all j and t, $E(F_t F_{t'}|\Phi_{t-1}) = 0$, for $t \neq t'$, $E\left(F_t v_{jt'}|\Phi_{t-1}\right) = 0$, for all j, t, t', $E\left(v_{it} v_{jt'}|\Phi_{t-1}\right) = \gamma_j$ if $i = j$, $t \neq t'$, and zero, otherwise. Therefore, the conditional covariance matrix is given by

$$\boldsymbol{\Sigma}_t = \sigma_t^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Gamma},$$

where $\boldsymbol{\Gamma} = cov(\mathbf{v}_t)$, is a $N \times N$ diagonal matrix with elements γ_j, $j = 1, ..., N$. For example, the $j-th$ conditional variance at time t is given by

$$\sigma_{j,t}^2 = \beta_j^2 \sigma_t^2 + \gamma_j = \beta_j^2 a_0 + \gamma_j + \beta_j^2 \theta \sum_{i=1}^{12} (13 - i) F_{t-i}^2$$

and the $jk-th$ conditional covariance is given by

$$\sigma_{jk,t}^2 = \beta_j \beta_k \sigma_t^2 = \beta_j \beta_k a_0 + \beta_j \beta_k \theta \sum_{i=1}^{12} (13 - i) F_{t-i}^2.$$

The number of parameters to be estimated in this latent factor model is $2N + 2$, which is a very small number in comparison with that of alternative multivariate models. In this model all the conditional variances and covariances are influenced by the common factor F. The impact of the common factor on the $j-th$ time series is given by the value of β_j. The impact of the 'unique factors' on the $j-th$ conditional variance is represented by the value of γ_j. The estimation of the model parameters is done by using either a

two-stage procedure or a simultaneous parameter estimation using Kalman filter. See for details Diebold and Nerlove (1989).

3.3.6 Factor GARCH model

King, Sentana and Wadhwani (1994) proposed a multivariate Factor model in which the time-varying volatility of returns is induced by changing the volatility of the underlying factors. This allows a parsimonious representation of the conditional covariance matrix of returns as a function of the variances of a small number of factors. They used 'observable' and 'unobservable' factors, whose conditional variances vary over time according to a univariate GARCH model. The model can be formulated as

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{y}_t$ is a $N \times 1$ vector containing the returns, $\boldsymbol{\mu}_t = E\left(\mathbf{y}_t|\Phi_{t-1}\right)$ is a $N \times 1$ parameter vector of conditional means or risk premia, $\boldsymbol{\varepsilon}_t$ is a $N \times 1$ innovation vector or unanticipated returns, Φ_{t-1} is the information set up to time $t-1$. Using a dynamic model for asset risk premia in terms of changing volatility of factors $\boldsymbol{\mu}_t = \mathbf{B}\boldsymbol{\Lambda}_t\boldsymbol{\tau}$, and a conditional factor model for the innovation returns $\boldsymbol{\varepsilon}_t = \mathbf{B}\mathbf{F}_t + \mathbf{v}_t$, the model can be written as:

$$\mathbf{y}_t = \mathbf{B}\boldsymbol{\Lambda}_t\boldsymbol{\tau} + \mathbf{B}\mathbf{F}_t + \mathbf{v}_t, \tag{3.21}$$

where $\mathbf{B}$ is a $N \times k$ matrix of factor loadings which measure the sensitivity of the assets to the common factors, $\boldsymbol{\Lambda}_t$ is a $k \times k$ diagonal positive definite matrix of factor variances which change over time, $\boldsymbol{\tau}$ is a $k \times 1$ parameter vector which can be interpreted as the price of risk for each factor, $\mathbf{F}_t$ is a $k \times 1$ vector of common factors which affect all assets, $\mathbf{v}_t$ is a $N \times 1$ vector of idiosyncratic terms which reflect unsystematic risk. To guarantee $E\left(\varepsilon_{it}|\Phi_{t-1}\right) = 0$ they assume that $E\left(F_{it}|\Phi_{t-1}\right) = 0$, and $E\left(v_{it}|\Phi_{t-1}\right) = 0$, for all i and t. They also assume that $E\left(F_{it}F_{jt}|\Phi_{t-1}\right) = 0$, for $i \neq j$, that is the factors are orthogonal, and that they have time varying conditional variances $\lambda_{it} > 0$, $i = 1, ..., k$. Let also $\boldsymbol{\Omega}_t$ is

a $N \times N$ positive definite covariance matrix of the idiosyncratic terms, $\mathbf{v}_t$. Given that the common factor and the idiosyncratic terms are conditionally orthogonal, the conditional covariance matrix $\mathbf{\Sigma}_t$ of $\mathbf{y}_t$ is:

$$\mathbf{\Sigma}_t = \mathbf{B}\mathbf{\Lambda}_t\mathbf{B}^{'} + \mathbf{\Omega}_t. \tag{3.22}$$

King, Sentana and Wadhwani (1994) divided the k common factors into k_1 'observable' factors $\mathbf{F}_{1t}$, which attempt to capture the correlation of the unanticipated innovations of some economic variables with stock returns, and k_2 'unobservable' factors $\mathbf{F}_{2t}$, which are correlated only with the return process. Using the orthogonality of the factors we can write equations (3.21) and (3.22) as:

$$\mathbf{y}_t = \mathbf{B}_1\mathbf{\Lambda}_{1t}\boldsymbol{\tau}_1+\mathbf{B}_2\mathbf{\Lambda}_{2t}\boldsymbol{\tau}_2+\mathbf{B}_1\mathbf{F}_{1t}+\mathbf{B}_2\mathbf{F}_{2t} + \mathbf{v}_t \tag{3.23}$$

and

$$\mathbf{\Sigma}_t = \mathbf{B}_1\mathbf{\Lambda}_{1t}\mathbf{B}_1^{'} + \mathbf{B}_2\mathbf{\Lambda}_{2t}\mathbf{B}_2^{'} + \mathbf{\Omega}_t. \tag{3.24}$$

Equation (3.24) express the covariance of asset returns into three components related to 'observable' factors, 'unobservable' factors, and idiosyncratic terms. The above model formulation given by (3.23) and (3.24) contains the models of Engle, Ng and Rothscild (1990) and Ng, Engle and Rothschild (1992) who used only 'unobservable' factors ($k_1 = 0$) and constant idiosyncratic variances ($\mathbf{\Omega}_t = \mathbf{\Omega}$), and of Diebold and Nerlove (1989) who used only one 'unobservable' factor and a zero risk premium ($k_1 = 0$, $k_2 = 1$, $\boldsymbol{\tau}_1 = 0$, $\boldsymbol{\tau}_2 = 0$).

A set of economic variables is used by King, Sentana and Wadhwani (1994) in order to generate the 'observable' factors. They estimated a vector autoregressive process for the economic variables, and extracted common observable factors from the innovations of these processes. Their model is completed by using the following equations

$$\mathbf{x}_t \;=\; \sum_{j=1}^{p} \mathbf{A}_j\mathbf{x}_{t-j}+\mathbf{e}_t$$

$$\mathbf{e}_t \;=\; \mathbf{C}_1\mathbf{F}_{1t} + \mathbf{w}_t,$$

where $\mathbf{x}_t$ is a $M \times 1$ vector of economic variables, $\mathbf{A}_j$ is a $M \times M$ matrix of coefficients, $\mathbf{e}_t$ is a $M \times 1$ innovation vector, $\mathbf{C}_1$ is a $M \times k_1$ matrix of factor loadings for the economic variables, and $\mathbf{w}_t$ is a $M \times 1$ innovation vector of idiosyncratic error terms. It is also that $E\left(\mathbf{w}_t|\Phi_{t-1}\right) = 0$, $E\left(\mathbf{F}_{1t}\mathbf{w}_t'|\Phi_{t-1}\right) = 0$, $E\left(\mathbf{w}_t\mathbf{v}_t'|\Phi_{t-1}\right) = 0$, and $E\left(\mathbf{w}_t\mathbf{w}_t'|\Phi_{t-1}\right) = \mathbf{\Gamma}_t$, a positive semidefinite diagonal matrix.

To complete the model, the conditional variances of the common factors and of idiosyncratic terms are given by univariate GARCH(1,1) models. For example, the diagonal elements of $\mathbf{\Lambda}_t$, $\mathbf{\Omega}_t$ are given by:

$$\lambda_{i,t} = \alpha_{i0} + \alpha_{i1}E\left(F_{i,t-1}^2|J_{t-1}\right) + b_{i1}\lambda_{i,t-1},\ \ i = 1, ..., k$$

$$\omega_{i,t} = \phi_{i0} + \phi_{i1}E\left(v_{i,t-1}^2|J_{t-1}\right) + \varphi_{i1}\omega_{i,t-1},\ \ i = 1, ..., N$$

and an analogous formula is given for the elements of $\mathbf{\Gamma}_t$. In the above formulations J_{t-1} is the available information set, and for the variances $\lambda_{i,t}$ of the common factors an unconditional variance equal to unity is imposing by $\alpha_{i0} = 1 - \alpha_{i1} - b_{i1}$. Details about the specification of these variance equations is given in King, Sentana and Wadhwani (1994). The number of parameters in this model is $Nk_1 + Nk_2 + k_1 + k_2 + Mk_1 + 2\left(k_1 + k_2\right) + 3N + 3M$ corresponding to $\mathbf{B}_1$, $\mathbf{B}_2$, $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$, $\mathbf{C}_1$, and the GARCH parameters of the diagonal elements of $\mathbf{\Lambda}_{1t}$, $\mathbf{\Lambda}_{2t}$, $\mathbf{\Omega}_t$ and $\mathbf{\Gamma}_t$. This number is very small comparing to that of other multivariate GARCH models especially when the dimensionality of the problem is large.

Different algorithms have been proposed in the literature for the estimation of factor models. Watson and Engle (1983) used the method of scoring and the EM algorithm for the estimation of dynamic factor, mimic and varying coefficient regression models. They suggested, for practical methods, a mixed EM and scoring algorithm, and the use of scoring algorithm for inference. Lin (1992) proposed four estimators for factor GARCH models and examined their finite sample properties. Demos and Sentana (1998) presented

an EM algorithm for conditionally heteroscedastic factor models and proposed a quasi-newton algorithm at last iterations. The study of Engle and Susmel (1993) investigates if some international stock markets share the same volatility process. Generally speaking, different factor models have been proposed in the literature, and have been analyzed by many researchers either using ARCH - GARCH framework (see, for example, Diebold and Nerlove, 1989, Laux and Ng, 1993, Engle, Ng and Rothschild, 1990, King, Sentana and Wadhwani, 1994 among several others) or using Stochastic Volatility framework (see, for example, Chib, Nardari and Shephard, 1999, Giakoumatos, Dellaportas and Politis, 1999, Aguilar and West, 2000) for the specification of the variances of the factors.

3.3.7 Asymmetric models of time-varying covariance matrix

It appears that volatility in asset return series is asymmetric: current returns are negatively correlated with future volatility. This is known as "leverage effect" or "asymmetric volatility effect". A negative return shock (unexpected price drop) will lead to a higher future volatility than a positive return shock (unexpected price increase) of the same magnitude. A number of sophisticated univariate models have been developed to accommodate asymmetric volatility; see, for example, Nelson (1991), Engle and Ng (1993), Glosten, Jagannathan and Runkle (1993), Hentschel (1995) among several others. In a multivariate framework, Kroner and Ng (1998) proposed a general asymmetric dynamic covariance matrix model which allows covariance asymmetry in stock returns on portfolios of small and large firms. This model encompasses various asymmetric extensions of some well known multivariate models. Their proposed multivariate model can be written as:

$$\mathbf{\Sigma}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t + \mathbf{\Phi} \circ \mathbf{\Theta}_t, \tag{3.25}$$

where $\mathbf{D}_t$ is a $N \times N$ diagonal matrix with elements the square root of the variance functions, that is $d_{ii,t} = \sqrt{\theta_{ii,t}}$ for all $i = 1, ..., N$, $d_{ij,t} = 0$, for all $i \neq j$, $\mathbf{R}$ is the $N \times N$ time invariant correlation matrix with elements ρ_{ij}, $i, j = 1, ..., N$, $\rho_{ii} = 1$ for all i, $\mathbf{\Phi}$ is

a $N \times N$ symmetric matrix with elements ϕ_{ij}, $\phi_{ii} = 0$ for all $i = 1, ..., N$, and $\phi_{ij} \neq 0$ for $i \neq j$, $\mathbf{\Theta}_t$ is a $N \times N$ matrix with elements $\theta_{ij,t}$ which are given by

$$\theta_{ij,t} = w_{ij} + \mathbf{b}_i^{'}\mathbf{\Sigma}_{t-1}\mathbf{b}_j + \mathbf{a}_i^{'}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^{'}\mathbf{a}_j + \mathbf{g}_i^{'}\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{'}\mathbf{g}_j, \tag{3.26}$$

where $\boldsymbol{\varepsilon}_{t-1}$ is a $N \times 1$ innovation vector, $\boldsymbol{\eta}_{t-1}$ is a $N \times 1$ vector with elements $\eta_{i,t-1} = \max[0, -\varepsilon_{i,t-1}]$, $\mathbf{a}_i$, $\mathbf{b}_i$, and $\mathbf{g}_i$, $i = 1, ..., N$ are $N \times 1$ vectors of parameters, and ρ_{ij}, ϕ_{ij}, and w_{ij}, $i, j = 1, ..., N$ are parameters.

The asymmetric covariance matrix model (3.25) has two components: the first component, $\mathbf{D}_t\mathbf{R}\mathbf{D}_t$, is like the multivariate GARCH model with constant conditional correlation, but the variance function is specified as in the BEKK model. The second term, $\mathbf{\Phi} \circ \mathbf{\Theta}_t$, has zero diagonal elements, and non zero off-diagonal elements which are given by the BEKK model, scaled by the ϕ_{ij} parameters. Note that in (3.26) the term $\mathbf{g}_i^{'}\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{'}\mathbf{g}_j$ allows for asymmetric effects in the conditional variances and covariances. This asymmetric covariance matrix model (3.25) and (3.26) has $\frac{N(N-1)}{2} + \frac{N(N-1)}{2} + \frac{N(N+1)}{2} + 3N^2$ parameters to be estimated corresponding to $\mathbf{R}$, $\mathbf{\Phi}$, w_{ij}, $\mathbf{a}_i$, $\mathbf{b}_i$, and $\mathbf{g}_i$, $i = 1, ..., N$. This model is very general and under restrictions on the parameters other well known multivariate models can be derived; see for details Kroner and Ng (1998).

To illustrate the model consider the following bivariate case:

$$\begin{aligned}
\sigma_{i,t}^2 &= \theta_{ii,t}, \text{ for } i = 1, 2 \\
\sigma_{12,t} &= \rho_{12}\sqrt{\theta_{11,t}}\sqrt{\theta_{22,t}} + \phi_{12}\theta_{12,t} \\
\theta_{ij,t} &= w_{ij} + \mathbf{b}_i^{'}\mathbf{\Sigma}_{t-1}\mathbf{b}_j + \mathbf{a}_i^{'}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^{'}\mathbf{a}_j + \mathbf{g}_i^{'}\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{'}\mathbf{g}_j,
\end{aligned}$$

where $\mathbf{b}_i^{'} = [b_{i1}, b_{i2}]$, $\mathbf{a}_i^{'} = [a_{i1}, a_{i2}]$, $\mathbf{g}_i^{'} = [g_{i1}, g_{i2}]$, $i = 1, 2$. Note that the parameter ϕ_{12} allows asymmetry in the covariance $\sigma_{12,t}$ that is not driven by the asymmetry in the variances. Kroner and Ng (1998) illustrated the above bivariate asymmetric model in stock return on portfolios of small and large firms, while an extensive study on the relation of volatility and risk in equity markets using an asymmetric BEKK model is

presented by Bekaert and Wu (2000).

3.3.8 Alternative multivariate ARCH-type models

Except for these well known multivariate models, alternative formulations have been proposed in the literature. For example, a bivariate version of the EGARCH model has been introduced by Braun, Nelson and Sunier (1995) in order to model the "leverage effect". Sentana (1995) introduced the univariate and multivariate Quadratic ARCH models. These multivariate models can capture the dynamic asymmetries of financial time series data. Tsay (1987) proposed the conditional heteroscedastic autoregressive moving-average model which includes both the random coefficient autoregressive (RCA) models (see for example, Nicholls and Quinn, 1982) and the autoregressive conditional heteroscedastic (ARCH) models as special cases. A multivariate conditional heteroscedastic autoregressive moving average model is presented and analyzed by Wong and Li (1997). Alexander (2000) introduced the Orthogonal GARCH model, which is a generalization of the factor GARCH model proposed by Engle, Ng and Rothschild (1990) to a multifactor model with orthogonal factors. A number of principal components is used to explain a large part of the variation in the system, and a GARCH(1,1) model is used to estimate the variances of the principal components. Engle (2000) introduced the dynamic conditional correlation (DCC) model to capture the stylized fact of time-varying conditional correlation of financial time series. This class of multivariate models can be estimated using a two step estimation strategy. The first is a series of univariate GARCH estimates and the second step is the correlation estimates.

3.3.9 Stationarity conditions and asymptotic properties

The problem of establishing stationarity conditions in univariate ARCH-type models has been considered by many researchers. Engle (1982) and Bollerslev (1986) presented the weakly stationarity conditions for ARCH and GARCH models, respectively, using the assumption that the process starts indefinitely far in the past with finite variance.

Nelson (1990) demonstrated that GARCH(1,1) may be strongly stationary, under weak conditions on the existence of moments, without being weakly stationary. Bougerol and Picard (1992) extended Nelson's result to the GARCH(p,q) process and reported necessary and sufficient condition for strong stationarity. That the GARCH process may be strong stationary without being weakly stationary comes from the fact that weak stationarity requires the mean, variance and autocovariances to be finite and time invariant. Strong stationarity requires the distribution function of any finite set of ε_t to be invariant under time translations, but finite moments are not required to exist. That is the unconditional variance may be infinite and yet the GARCH process may still be strongly stationary. The empirical studies suggest that some financial time series (for example exchange rates and interest rates) seem to be characterized by such a process. From a theoretical point of view, a related problem is that of statistical inference for these models. Weiss (1986) proved the consistency and asymptotic normality of the maximum likelihood estimator for the univariate ARCH model under strong conditions on the existence of the moments of the error term. Lumsdaine (1991), Lee and Hansen (1994), and Elie and Jeantheau (1995) present the asymptotic theory for GARCH models under weaker assumptions.

Stationarity conditions for multivariate ARCH and GARCH models are more complex, and results are only available for a few special cases. For example, Bollerslev and Engle (1993) and Engle and Kroner (1995) presented necessary and sufficient conditions for covariance stationary for the multivariate GARCH model, following an analogous way as in the univariate models. However, much more work remains establishing the conditions for strong stationarity and ergodicity for the multivariate GARCH(p,q) model. Jeantheau (1998) studied the problem of statistical inference for multivariate heteroscedastic models. He dealt with the asymptotic properties of quasi maximum likelihood of the multivariate GARCH model with constant conditional correlation and gave weak conditions under which strong consistency can be obtained.

Chapter 4

Bayesian inference for GARCH and EGARCH models

4.1 Introduction

In this chapter, we consider the general problem of Bayesian inference, prediction and model criticism of univariate ARCH-type models. We demonstrate that adoption of the Bayesian framework can be advantageous on grounds of generality, accuracy and flexibility. Moreover, the MCMC sampling-based approach provides an idealized way to extract any posterior summary of interest such as functions of parameters, and in addition to construct predictive densities which take into account model uncertainty.

Two of the most useful ARCH parameterizations are the generalized ARCH (GARCH) model and the exponential GARCH (EGARCH) model. Although the list of other proposed ARCH parameterizations is long (see, for example, the description of univariate time-varying volatility models in chapter 3), in our analysis in this chapter we focus on GARCH and EGARCH models, without restricting the methodological potential of our suggested procedures. GARCH models are chosen because of their extended applicability to many financial data and EGARCH models because of both their good performance in a series of comparative studies (see Pagan and Schwert, 1990, Engle and Ng, 1993,

and Shephard, 1996), as well as their complexity which can be viewed as a challenge for our methodology. In ARCH-type models the estimation of the parameters can be done by using maximum likelihood, quasi-maximum likelihood or generalized method of moments; see, for details, the review paper of Bollerslev, Chou and Kroner (1992).

The key steps in our proposed framework are as follows. First, within a certain GARCH/EGARCH model, we construct a Markov chain which has as a stationary distribution the posterior distribution of the model parameters. Simulation of this Markov chain provides, after some burn-in period and adequately many iterations, samples from the posterior distribution of interest; see, for details, Smith and Roberts (1993) or Besag, Green, Higdon and Mengersen (1995). Second, for a given set of competing models, we propose modeling each GARCH or EGARCH model jointly and base our inference about the models on their posterior probabilities or Bayes factors. Thus, we avoid the usual approach which considers the models separately and chooses the best model via significance tests. We believe that the joint estimation of parameters and model probabilities not only provides a probabilistically sound way to overcome the awkward model selection problem in GARCH/EGARCH models, but also introduces a new way to predict the future volatility via "model averaging". To obtain a sample of the joint posterior density of models and model parameters, we extend the MCMC strategy so that the sampler jumps between parameter subspaces of different dimensionality corresponding to different models. This idea is based on reversible jump MCMC introduced by Green (1995).

A Bayesian strategy similar to the one we propose for the parameter estimation of our models is described in Muller and Pole (1999). It is an MCMC strategy which handles successfully a GARCH model with covariates by using Metropolis-Hasting steps with a sophisticated choice of proposal distributions. A pioneer Bayesian implementation strategy for such models is due to Geweke (1989b) who proposed a Monte Carlo strategy to derive the desired posterior summaries of interest. Finally, in a similar context, Jacquier, Polson and Rossi (1994) proposed a Metropolis-Hastings algorithm to derive posterior

distributions of the parameters of stochastic volatility models.

4.2 The analyzed models

Let $\{\varepsilon_t\}$ be a real valued discrete-time stochastic process, which is the error process. Assume that the mean equation is of the form $y_t = \varepsilon_t$, where y_t is the observation process. The GARCH(p,q) model (Bollerslev, 1986) is given by the following two-stage formulation

$$\varepsilon_t = z_t \sigma_t, \ t = 0, ..., T \tag{4.1}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \ t = 1, ..., T, \tag{4.2}$$

where z_t are iid with $E(z_t) = 0$ and $Var(z_t) = 1$; σ_t^2 is the (conditional) variance of the $\{\varepsilon_t\}$ process at time t, p, q are integers with $p > 0$, $q \geq 0$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, ..., p$, and $\beta_j \geq 0$, $j = 1, ..., q$. In (4.2) it is assumed that $\varepsilon_t = \sigma_t = 0$ for $t < 0$.

Assuming that z_t are normally distributed, the parameter vector to be estimated in (4.1) and (4.2) is, for $q > 0$, $\boldsymbol{\theta} = (\alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, \sigma_0^2)$. The likelihood for a sample of $T + 1$ observations $\mathbf{y} = (y_0, ..., y_T)$ can be written as

$$l_N(\mathbf{y}|\boldsymbol{\theta}) = (2\pi)^{-\frac{T+1}{2}} \prod_{t=0}^{T} \left\{ \left(\sigma_t^2\right)^{-1/2} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \right\},$$

where ε_t is given from the corresponding mean equation, and σ_t^2 is expressed via (4.2). Under the assumption of a Student-t distribution (Bollerslev, 1987, Baillie and Bollerslev, 1989) for the error process $\{\varepsilon_t\}$, the likelihood for a sample of $T + 1$ observations $\mathbf{y}$ can be written as

$$l_T(\mathbf{y}|\boldsymbol{\theta}) = \prod_{t=0}^{T} \left\{ \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \left[(n-2)\sigma_t^2\right]^{1/2}} \left(1 + \frac{\varepsilon_t^2}{(n-2)\sigma_t^2}\right)^{-\frac{n+1}{2}} \right\},$$

where $n > 2$ denotes the degrees of freedom of the Student-t distribution, ε_t is given from

the corresponding mean equation, σ_t^2 is expressed via (4.2), and the parameter vector to be estimated is $\boldsymbol{\theta} = (\alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, \sigma_0^2, n)$.

A modification of GARCH models is the EGARCH models introduced by Nelson (1991). They are defined by assuming that z_t in (4.1) follows a generalized error distribution with mean 0, variance 1 and parameter v, so that if $v = 2$ we obtain the normal distribution, and for $v < 2$ and $v > 2$ the distribution of z_t has thicker and thinner tails than the normal distribution respectively. Moreover, the EGARCH formulation in the second stage equation (4.2) is given by

$$\ln\left(\sigma_t^2\right) = \alpha_0 + \sum_{j=1}^{q} \beta_j \ln\left(\sigma_{t-j}^2\right) + \sum_{k=1}^{p} \left[\theta_k z_{t-k} + \gamma_k \left(|z_{t-k}| - E\left|z_{t-k}\right|\right)\right] \tag{4.3}$$

where $\sigma_t = z_t = 0$ for $t < 0$ and $E\left|z_{t-k}\right|$ is given, under the assumption of generalized error distribution for z_t, by

$$E\left|z_{t-k}\right| = \frac{\Gamma\left(\frac{2}{v}\right)}{\left[\Gamma\left(\frac{1}{v}\right)\Gamma\left(\frac{3}{v}\right)\right]^{1/2}}.$$

The EGARCH(p,q) formulation above has parameters $\boldsymbol{\theta} = (\alpha_0, \beta_1, ..., \beta_q, \theta_1, ..., \theta_p, \gamma_1, ..., \gamma_p, v, \sigma_0^2)$ and the likelihood function is given by

$$l\left(\mathbf{y}|\boldsymbol{\theta}\right) = \left[\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma\left(\frac{1}{v}\right)}\right]^{T+1} \prod_{t=0}^{T} \left\{\left(\sigma_t^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left|\frac{\varepsilon_t}{\sigma_t \lambda}\right|^v\right)\right\}. \tag{4.4}$$

Under the usual assumption that the past variance σ_0^2 is specified in advance, then the EGARCH(p,q) model has a parameter vector $\boldsymbol{\theta}^* = \boldsymbol{\theta} \backslash \{\sigma_0^2\}$ and the likelihood for a sample of $T + 1$ observations can be written as

$$l\left(\mathbf{y}|\boldsymbol{\theta}^*\right) = \left[\frac{v}{\lambda 2^{1+\frac{1}{v}}\Gamma\left(\frac{1}{v}\right)}\right]^{T-p+1} \prod_{t=p}^{T} \left\{\left(\sigma_t^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left|\frac{\varepsilon_t}{\sigma_t \lambda}\right|^v\right)\right\}.$$

4.3 Bayesian formulation and implementation

4.3.1 Inferences for a given model

Bayesian inferences about the parameter vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)$ conditional on data $\mathbf{y}$ are made via the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$. Using Bayes theorem, this density takes the form $p(\boldsymbol{\theta}|\mathbf{y}) = c\, l(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$ for some normalizing constant c, likelihood function $l(\mathbf{y}|\boldsymbol{\theta})$ and prior density $p(\boldsymbol{\theta})$.

For many realistic problems, evaluation of $p(\boldsymbol{\theta}|\mathbf{y})$ is analytically intractable so numerical or asymptotic methods are necessary to obtain posterior summaries of interest; see Evans and Swartz (1995) for a recent review of possible avenues. In this study, we adopt the MCMC sampling strategies as our tool for this purpose. A simple strategy is to use n independent Metropolis steps (Tierney, 1994, Chib and Greenberg, 1995) for all θ_i, $i = 1, 2, ..., n$, and a usual approach is to adopt a random walk chain with an increment normal density $N(0, \sigma^2)$, where the variance σ^2 is appropriately chosen so that the convergence of the MCMC sampler is as fast as possible. Metropolis steps are probably not the best choice for constructing the required Markov chain, but if the full conditional posterior densities are not of known form, or if they do not have exploitable properties such as log-concavity of known maximum, Metropolis steps give the easiest black-box sampling strategy yielding the required realizations of $p(\boldsymbol{\theta}|\mathbf{y})$.

Despite the fact that we propose this simple sampling scheme, where we update the elements of the parameter vector one at a time (single component update), we emphasize in the application section that other alternatives may drastically improve the efficiency of the algorithm. For example, our empirical data analysis indicated that simultaneous sampling of a carefully chosen subvector of $\boldsymbol{\theta}$, after a possible transformation to a subvector taking values on $(-\infty, \infty)$, may be extremely appealing. To do that, we update the highly correlated elements of the parameter vector simultaneously (simultaneous component update), and take a sample from this subvector of $\boldsymbol{\theta}$ using multivariate Metropolis steps as follows. We first estimate the sample covariance matrix $\boldsymbol{\Sigma}$ related to this sub-

vector from an initial exploratory run of the Markov chain. Then, the update from time t to time $t+1$ is achieved by using a multivariate normal proposal density $N(\boldsymbol{\mu}^t, c\boldsymbol{\Sigma})$, with $\boldsymbol{\mu}^t$ denoting the vector at time t and c a constant to tune the acceptance rate.

Muller and Pole (1999) deal with the correlation problem in estimating the parameters of a GARCH(1,1) model in another way. They choose their proposal distribution to mimic the true full conditional posterior distribution by using related results from regression problems. An extra cost is an additional rejection step which is needed to maintain the joint posterior density as the stationary distribution of the simulated Markov chain. The authors report problems in the above methodology when the proposal distribution is significantly thinner than the true full conditional, and they propose a rather sophisticated periodic change of the shape of the proposal density.

4.3.2 Inferences under model uncertainty

Assume that we have a countable set M of competing models for a given set of data $\mathbf{y}$. Let model $m \in M$ have a vector $\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m$ of unknown parameters, the dimension of which may vary from model to model. The posterior probability of model m is given by

$$p(m|\mathbf{y}) = \frac{p(m) \int\limits_{\boldsymbol{\Theta}_m} p(\mathbf{y}|m, \boldsymbol{\theta}_m)\, p(\boldsymbol{\theta}_m|m)\, d\boldsymbol{\theta}_m}{\sum\limits_{m \in M} p(m) \int\limits_{\boldsymbol{\Theta}_m} p(\mathbf{y}|m, \boldsymbol{\theta}_m)\, p(\boldsymbol{\theta}_m|m)\, d\boldsymbol{\theta}_m} \tag{4.5}$$

where $p(\mathbf{y}|m, \boldsymbol{\theta}_m)$ is the likelihood given the model m, and the parameter vector $\boldsymbol{\theta}_m$, $p(m)$ is the prior probability for model m, and $p(\boldsymbol{\theta}_m|m)$ is the prior of the parameter vector $\boldsymbol{\theta}_m$ given the model m. Inference about the model selection problem may be done using the Bayes Factor (BF) of model m_i against model m_j given by

$$BF = \frac{\int\limits_{\boldsymbol{\Theta}_{m_i}} p(\mathbf{y}|m_i, \boldsymbol{\theta}_{m_i})\, p(\boldsymbol{\theta}_{m_i}|m_i)\, d\boldsymbol{\theta}_{m_i}}{\int\limits_{\boldsymbol{\Theta}_{m_j}} p\left(\mathbf{y}|m_j, \boldsymbol{\theta}_{m_j}\right) p\left(\boldsymbol{\theta}_{m_j}|m_j\right) d\boldsymbol{\theta}_{m_j}}. \tag{4.6}$$

Kass and Raftery (1995) gave a series of arguments that make Bayes factors appealing when compared with other model selection strategies such as AIC or BIC. However, Bayes factors require evaluation of the integrals in the numerator and denominator of (4.6) which are the marginal densities $p(\mathbf{y}|m_i)$ and $p(\mathbf{y}|m_j)$. Green (1995) introduced a reversible jump MCMC strategy for generating from the joint posterior $p(m, \boldsymbol{\theta}_m|\mathbf{y})$, based on the standard Metropolis Hastings approach. During reversible jump MCMC sampling, the constructed Markov chain moves within and between models so that the limiting proportion of visits to a given model is the required $p(m|\mathbf{y})$ in (4.5).

We presented a detailed description of the Reversible Jump MCMC methodology in the first chapter. Using the same notation and in order to illustrate the algorithm we present, in this section, an example using GARCH and EGARCH models. Suppose we deal with only two models; a GARCH(1,1) model, denoted by m_1, with parameter vector $\boldsymbol{\theta}_1 = (\alpha_0,\ \beta_1,\ \sigma_0^2,\ \alpha_1)$ and an EGARCH(1,1) model, denoted by m_2, with parameter vector $\boldsymbol{\theta}_2 = (\widetilde{\alpha}_0,\ \widetilde{\beta}_1,\ \widetilde{\sigma}_0^2,\ \widetilde{\theta}_1,\ \widetilde{\gamma}_1,\ \widetilde{v})$. We shall illustrate in turn two reversible jump strategies. For convenience assume that $j(m_1, m_2) = j(m_2, m_1) = 1$, i.e. we always propose to move from one model to the other. For the first strategy, we need a 3-dimensional invertible function g which transforms the "common" elements of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$:

- To move from GARCH(1,1) to EGARCH(1,1)

1. set $\mathbf{u} = \left(\widetilde{\theta}_1,\ \widetilde{\gamma}_1,\ \widetilde{v}\right)$ and $\mathbf{u}' = \alpha_1$ $\left[\text{so } d(\mathbf{u}) + d_{m_1} = d\left(\mathbf{u}'\right) + d_{m_2} = 7\right]$
2. choose proposal densities $q(\mathbf{u}|\boldsymbol{\theta}_1, m_1, m_2)$ and $q\left(\mathbf{u}'|\boldsymbol{\theta}_2, m_2, m_1\right)$
3. generate $\mathbf{u}$ from $q(\mathbf{u}|\boldsymbol{\theta}_1, m_1, m_2)$
4. set $\boldsymbol{\theta}_2' = \left(g(\alpha_0, \beta_1, \sigma_0^2), \mathbf{u}\right)$ for some invertible function g
5. set $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2'$ and move to model 2 with probability $a = \min\{1, r\}$ where r is given by (1.13)

- To move from EGARCH(1,1) to GARCH(1,1)

1. set $\mathbf{u} = (\alpha_1)$ and $\mathbf{u}^{'} = \left(\widetilde{\theta}_1,\ \widetilde{\gamma}_1,\ \widetilde{v} \right)$ $\left[\text{so } d\left(\mathbf{u}\right) + d_{m_2} = d\left(\mathbf{u}^{'}\right) + d_{m_1} = 7\right]$
2. choose proposal densities $q(\mathbf{u}|\boldsymbol{\theta}_2, m_2, m_1)$ and $q\left(\mathbf{u}^{'}|\boldsymbol{\theta}_1, m_1, m_2\right)$
3. generate $\mathbf{u}$ from $q(\mathbf{u}|\boldsymbol{\theta}_2, m_2, m_1)$
4. set $\boldsymbol{\theta}_1^{'} = \left(g^{-1}\left(\widetilde{\alpha}_0, \widetilde{\beta}_1, \widetilde{\sigma}_0^2\right), \mathbf{u}\right)$
5. set $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{'}$ and move to model 1 with probability $a = \min\{1, r\}$ where r is given by (1.13).

The second strategy, which we propose in such problems, does not require a function g but requires proposal densities q of higher dimension:

- To move from GARCH(1,1) to EGARCH(1,1)

1. set $\mathbf{u} = \left(\widetilde{\alpha}_0,\ \widetilde{\beta}_1,\ \widetilde{\sigma}_0^2,\ \widetilde{\theta}_1,\ \widetilde{\gamma}_1,\ \widetilde{v}\right)$ and $\mathbf{u}^{'} = (\alpha_0,\ \beta_1,\ \sigma_0^2,\ \alpha_1)$
   $\left[\text{so } d\left(\mathbf{u}\right) + d_{m_1} = d_{m_2} + d\left(\mathbf{u}^{'}\right) = 10\right]$
2. choose proposal densities $q(\mathbf{u}|m_2)$ and $q\left(\mathbf{u}^{'}|m_1\right)$
3. generate $\mathbf{u}$ from $q(\mathbf{u}|m_2)$
4. set $\boldsymbol{\theta}_2^{'} = \mathbf{u}$
5. set $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^{'}$ and move to model 2 with probability $a = \min\{1, r\}$ where r is given by (1.14)

- To move from EGARCH(1,1) to GARCH(1,1)

1. set $\mathbf{u} = (\alpha_0,\ \beta_1,\ \sigma_0^2,\ \alpha_1)$ and $\mathbf{u}^{'} = \left(\widetilde{\alpha}_0,\ \widetilde{\beta}_1,\ \widetilde{\sigma}_0^2,\ \widetilde{\theta}_1,\ \widetilde{\gamma}_1,\ \widetilde{v}\right)$
   $\left[\text{so } d\left(\mathbf{u}\right) + d_{m_2} = d_{m_1} + d\left(\mathbf{u}^{'}\right) = 10\right]$
2. choose proposal densities $q(\mathbf{u}|m_1)$ and $q\left(\mathbf{u}^{'}|m_2\right)$

3. generate $\mathbf{u}$ from $q(\mathbf{u}|m_1)$

4. set $\boldsymbol{\theta}_1^{'} = \mathbf{u}$

5. set $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{'}$ and move to model 1 with probability $a = \min\{1, r\}$ where r is given by (1.14).

4.3.3 Bayesian model averaging for volatility prediction

In time-varying volatility models such as GARCH and EGARCH, prediction of the future volatility is of particular interest. Having been able to calculate the posterior probabilities of each model, it seems natural to account for this uncertainty in our predictive inferences. Rather than choosing a single "best" model and then make inferences as if the selected model was the true model, we can use the following model averaging approach which provides composite predictions. Suppose that we are interested in σ_{T+1}^2, the predictive volatility at time $T+1$. Then, its posterior distribution given data $\mathbf{y}$ is given by

$$p\left(\sigma_{T+1}^2|\mathbf{y}\right) = \sum_{m \in M} p\left(\sigma_{T+1}^2|m, \mathbf{y}\right) p\left(m|\mathbf{y}\right) \tag{4.7}$$

which is an average of the posterior predictive distribution under each model weighted by their posterior model probabilities.

Computation of (4.7) is straightforward after the reversible jump MCMC sampling algorithm has been implemented. First, given a model m, a posterior sample of $p(\sigma_{T+1}^2|m, \mathbf{y})$ is just obtained by calculating, for each sampled point in $\boldsymbol{\theta}$, the variances $\sigma_0^2, \sigma_1^2, ..., \sigma_{T+1}^2$ in (4.2) or (4.3). Then (4.7) suggests that in order to obtain a sample of $p(\sigma_{T+1}^2|\mathbf{y})$, each sampled point under model m should be taken with probability $p\left(m|\mathbf{y}\right)$. Thus, the derived sample of $p(\sigma_{T+1}^2|\mathbf{y})$ is obtained by weighting all samples of $p(\sigma_{T+1}^2|m, \mathbf{y})$ by the corresponding $p\left(m|\mathbf{y}\right)$. We emphasize that we advocate the above procedure when the model selection goal is prediction rather than the understanding of the data generation mechanism. Moreover, all probability calculations are conditioned on a set of possible

models, M, and therefore, $p(m|\mathbf{y})$ should be interpreted with care. In fact, we just average, using appropriate weights, over different available mechanisms which generate predictive distributions.

4.4 A simulation study

In this section we perform a simulation study to illustrate the relative merits of Bayesian versus classical approaches. We focus on one of the most important practical issues related to heteroscedastic models, the construction of volatility predictions. We first simulate data $y_1, y_2, ..., y_{1025}$ from a GARCH(1,1) model of the form

$$\sigma_t^2 = 0.0005 + 0.25\varepsilon_{t-1}^2 + 0.7\sigma_{t-1}^2,\ \sigma_0^2 = 0.0003,\ \varepsilon_t \sim N\left(0, \sigma_t^2\right)$$

and from an EGARCH(1,1) model of the form

$$\ln\left(\sigma_t^2\right) = -0.3 + 0.7\ln\left(\sigma_{t-1}^2\right) + 0.1\frac{\varepsilon_{t-1}}{\sigma_{t-1}} + 0.5\left(\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - E\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right|\right)$$

$$\sigma_0^2 = 0.003,\ \varepsilon_t \sim GED_2\left(0, \sigma_t^2\right),$$

where GED_2 denotes the Generalized Error distribution with mean 0, variance σ_t^2 and tail-thickness parameter $v = 2$. Our simulation hypothetical scenario consists of comparisons on a daily basis, of the last 25 "true" volatilities with the classical and Bayesian volatility predictions. Thus the predictions $\widehat{\sigma}_{t+1}^2$ for $t = 1000, ..., 1024$ are based on the previous t points.

A first simple criterion to compare the two prediction approaches is to use some point estimates $\widehat{\sigma}_{t+1}^2$ and construct the mean absolute percentage error

$$MAPE = \frac{1}{25}\sum_{t=1001}^{1025}\frac{\left|\sigma_t^2 - \widehat{\sigma}_t^2\right|}{\sigma_t^2},$$

where σ_t^2 is the "true" volatility obtained from the simulated data. Table 4.1 provides the

	GARCH	*EGARCH*
Classical	0.0332	0.0468
Posterior mean	0.0318	0.0593
Posterior median	0.0308	0.0581

Table 4.1: Mean absolute percentage error for GARCH and EGARCH models.

values of $MAPE$ for both GARCH and EGARCH models and for both posterior mean and medians taken as Bayesian point estimators. The classical estimators are based on the function *garch()* of Splus. It is evident that the differences between the 2 predictive approaches are in general minimal.

A better insight into the performance of the estimators can be achieved by a graphical inspection of the true and the predicted volatilities. Moreover, this exercise reveals the comparative merits of the classical and Bayesian approaches. On the one hand, construction of 95% posterior credible intervals is straightforward by just calculating the 0.025 and 0.975 quantiles of the predictive density sample. On the other hand, classical estimates of the dispersion of $\widehat{\sigma}_{t+1}^2$ seem to be unavailable in the literature. In any case, it is reassuring that the 95% Bayesian credible intervals for the predictive volatility contain the "true" volatilities for both GARCH and EGARCH models and for all time periods; see Figures 4-1 and 4-2 respectively.

The above promising aspect of Bayesian inference provides a further advantage for real financial applications, for example, the value of a call option C on a nondividend-paying stock is given (see Dubofsky, 1992) by

$$C = S\Phi\left(d_1\right) - ke^{-rt}\Phi\left(d_2\right)$$

where

$$d_1 = \frac{\ln\left(S/k\right) + \left(r + \sigma^2/2\right)/T}{\sigma\sqrt{T}},\ d_2 = d_1 - \sigma\sqrt{T},$$

$\Phi\left(.\right)$ is the cumulative standard normal distribution function, S is the price of the underlying asset, k is the strike price of the call option, r is the risk-free rate, T is the time

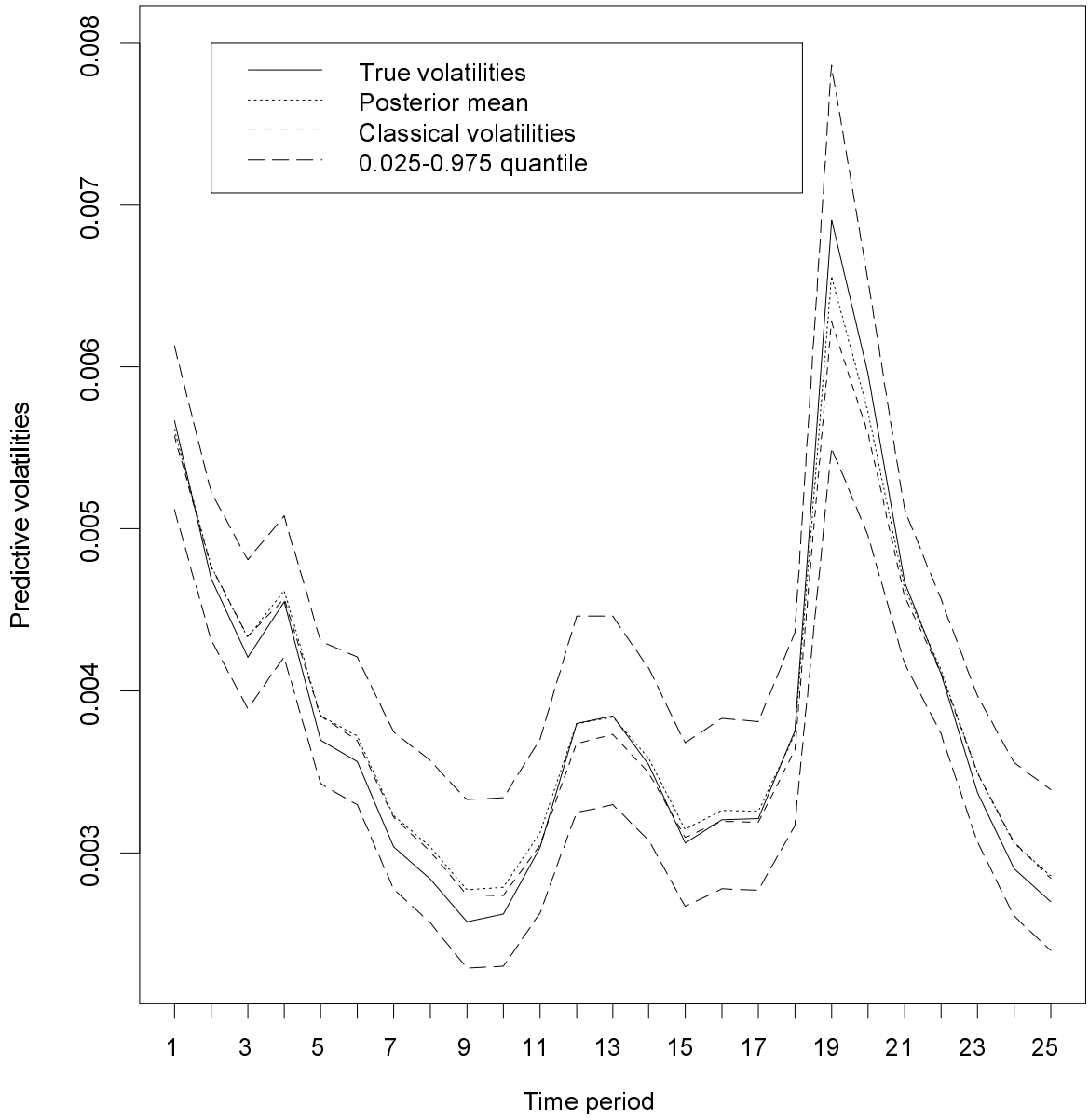

Figure 4-1: GARCH(1,1) model

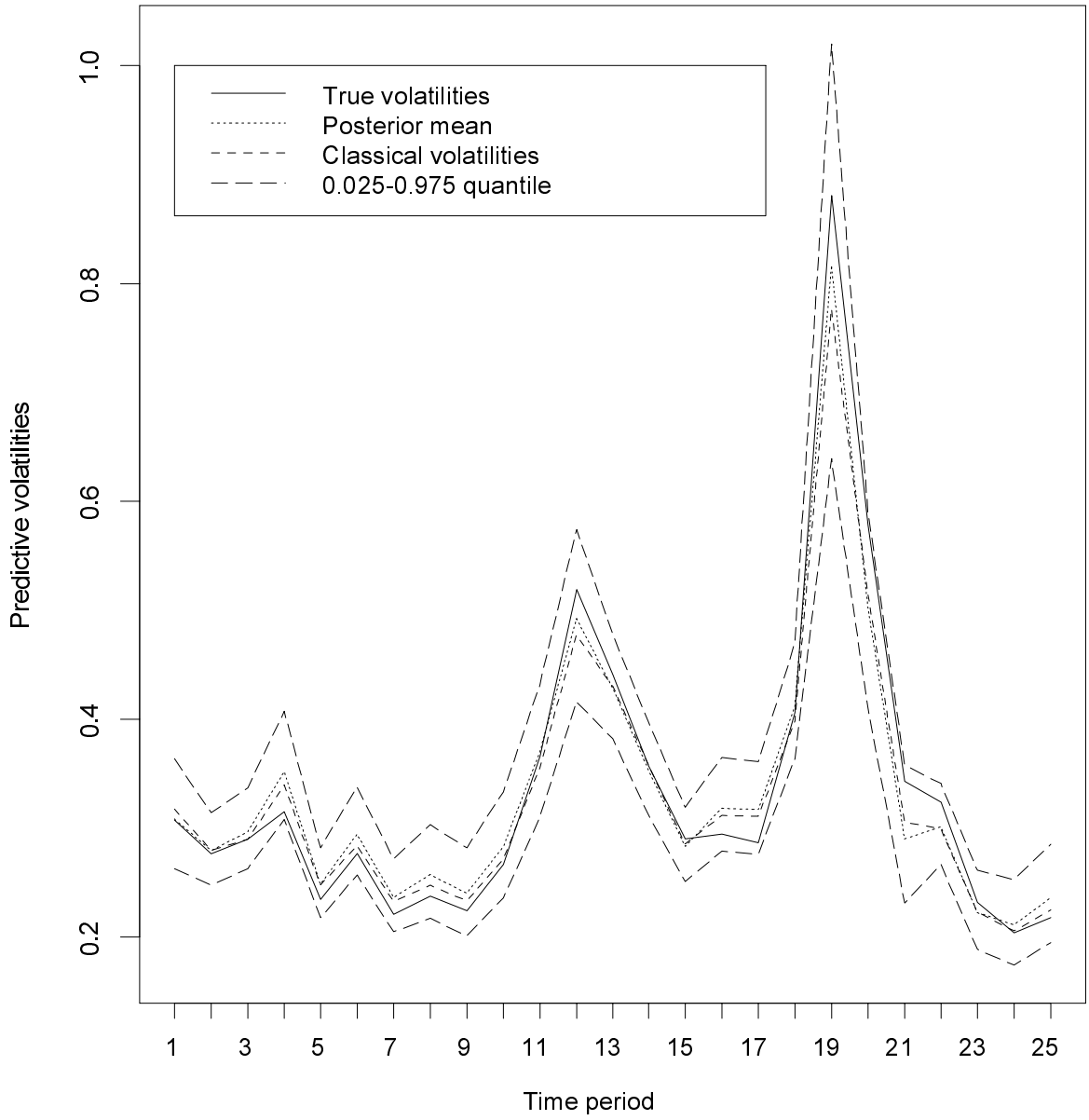

Figure 4-2: EGARCH(1,1) model

to expiration, σ is the standard deviation of the underlying asset's return at time to expiration. The volatility σ^2 of the stock is an important determinant of an option's value and is the only determinant of the value of the call option that is not directly observable. The prediction obtained from an ARCH-type model may be used as an estimate for σ^2 (Noh, Engle and Kane, 1994). It is immediately evident that we can utilize the predictive density of σ^2, using GARCH models, and construct the posterior density of C. Moreover, if the shape of $p(\sigma^2|\mathbf{y})$ is not normal-like then this feature will be realistically represented in the distribution of C.

4.5 An application: The Athens stock exchange index

We illustrate our proposed methodology using $T = 490$ weekly rates of the General Index of the Athens stock exchange over the period 1986-1996. If G_t is the value of the General Index at time t, then we model the weekly rate $y_t = \ln\left(\frac{G_t}{G_{t-1}}\right)$, $t = 1, ..., T$.

First, we apply GARCH(1,1) and EGARCH(p, q), $p, q = 1, 2$ models to the Athens stock exchange data. For our illustration we chose non-informative priors for all model parameters. For the parameters of the GARCH model, we used $p(\alpha_0) = \alpha_0^{-1}$, $p(\sigma_0^2) = \sigma_0^{-2}$ and $U(0, 1)$ priors for α_1 and β_1. Under the Student-t distribution the degree of freedom n $(n > 2)$ is a parameter to be estimated, and we used as a prior the non-informative $p(n) = (n-2)^{-1}$. Stationarity conditions impose that $\alpha_1 + \beta_1 < 1$ for the GARCH(1,1) model; this was taken into account by just rejecting, in the MCMC algorithm, all pairs of (α_1, β_1) which did not obey the above restriction; see, for example, Gelfand, Smith and Lee (1992). For the parameters of the EGARCH models, we used $U(-1, 1)$ priors for β_j, $j = 1, 2$ and normal or lognormal noninformative priors for the other parameters of the model taken as, $N(0, 10)$ for α_0, θ_i, γ_i, $i = 1, 2$, $LN(1.04 \cdot 10^{22}, 2.93 \cdot 10^{87})$ for v and $LN(4.72 \cdot 10^{18}, 6.008 \cdot 10^{80})$ for σ_0^2, where $LN(\mu, \sigma^2)$ denotes the lognormal distribution with mean μ and variance σ^2. The priors for both models turned out to be practically

$Model$	α_0	α_1	β_1	σ_0^2	n				
$Garch_N(1,1)$	1.22	0.15	−0.47	0.93	−				
$Garch_T(1,1)$	−1.56	−0.24	0.67	0.58	−0.003				
$Model$	α_0	β_1	θ_1	γ_1	v	β_2	θ_2	γ_2	σ_0^2
$Egarch\,(1,1)$	0.23	0.19	0.15	−0.94	−1.180	−	−	−	1.42
$Egarch\,(1,2)$	0.06	−0.33	−0.13	1.02	−1.670	0.35	−	−	−0.11
$Egarch\,(2,1)$	1.07	0.99	−0.25	−1.04	−0.607	−	−0.36	0.06	−0.47
$Egarch\,(2,2)$	0.69	−0.03	0.67	0.10	1.180	0.11	−1.00	−0.41	−1.35
$Model$	α_0	β_1	θ_1	γ_1	v	β_2	θ_2	γ_2	
$Egarch^*\,(1,1)$	1.64	1.60	−1.14	−1.33	0.214	−	−	−	
$Egarch^*\,(1,2)$	−0.03	−0.98	1.07	−1.59	−0.806	0.90	−	−	
$Egarch^*\,(2,1)$	1.46	1.42	−1.60	0.20	−1.750	−	0.17	−1.33	
$Egarch^*\,(2,2)$	0.09	−0.65	0.26	−0.65	−0.711	0.75	−0.12	1.47	

Table 4.2: Geweke's convergence z-scores for the parameters of the GARCH and EGARCH models.

noninformative as their effective range is from 2.5 to 70 times larger than the effective range of the resulting posterior densities. Stationarity conditions also were taken into account as in the GARCH model. We use two EGARCH specifications: The first, denoted by EGARCH(p,q) is the model formulation given by (4.4). The second, denoted by EGARCH*(p,q), is an EGARCH model with σ_0^2 estimated from the unconditional variance of the sample data.

The output sample of every MCMC run was constructed as follows. First a large sample was taken and an initial (burn-in) part of it was discarded after a visual inspection of the time series plots of each parameter. Then the autocorrelation function of each parameter was investigated and a decision was made about the lag intervals with which the sample should be collected, in order to achieve a nearly non-correlated sample. And finally, the resulting samples were checked for convergence by using the tests proposed by Geweke (1992) and Heidelberger and Welch (1983). Table 4.2 presents the z-scores from the former diagnostic. These z-scores indicate that the convergence of the Markov chain has been achieved. Estimated posterior means and standard deviations for the parameters of each model we considered are illustrated in Table 4.3. We present in Figures 4-3 and 4-4 the convergence diagrams and the histograms of the posterior sample

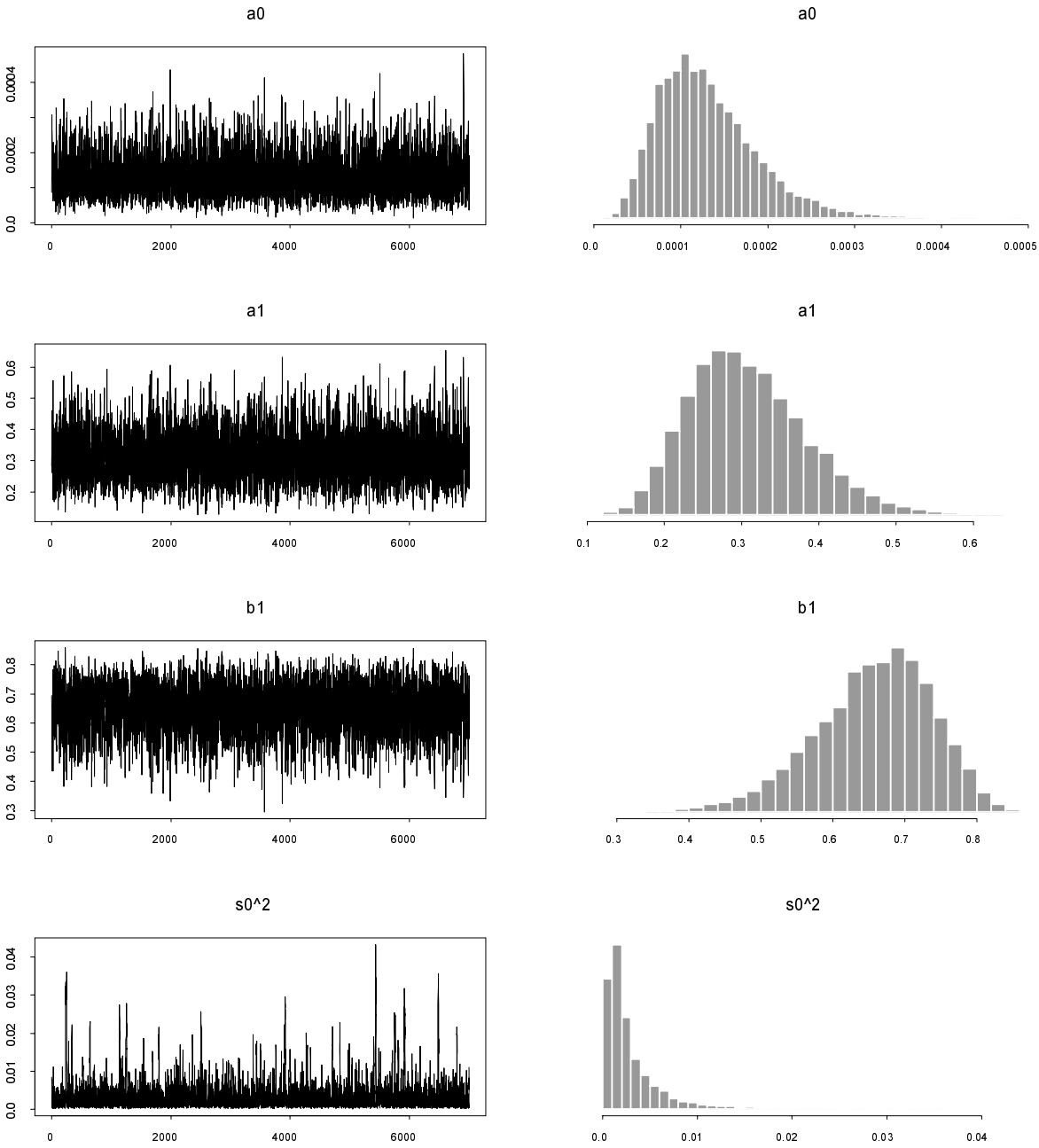

Figure 4-3: Convergence diagrams and histograms of the posterior sample of the parameters of the GARCH(1,1) model.

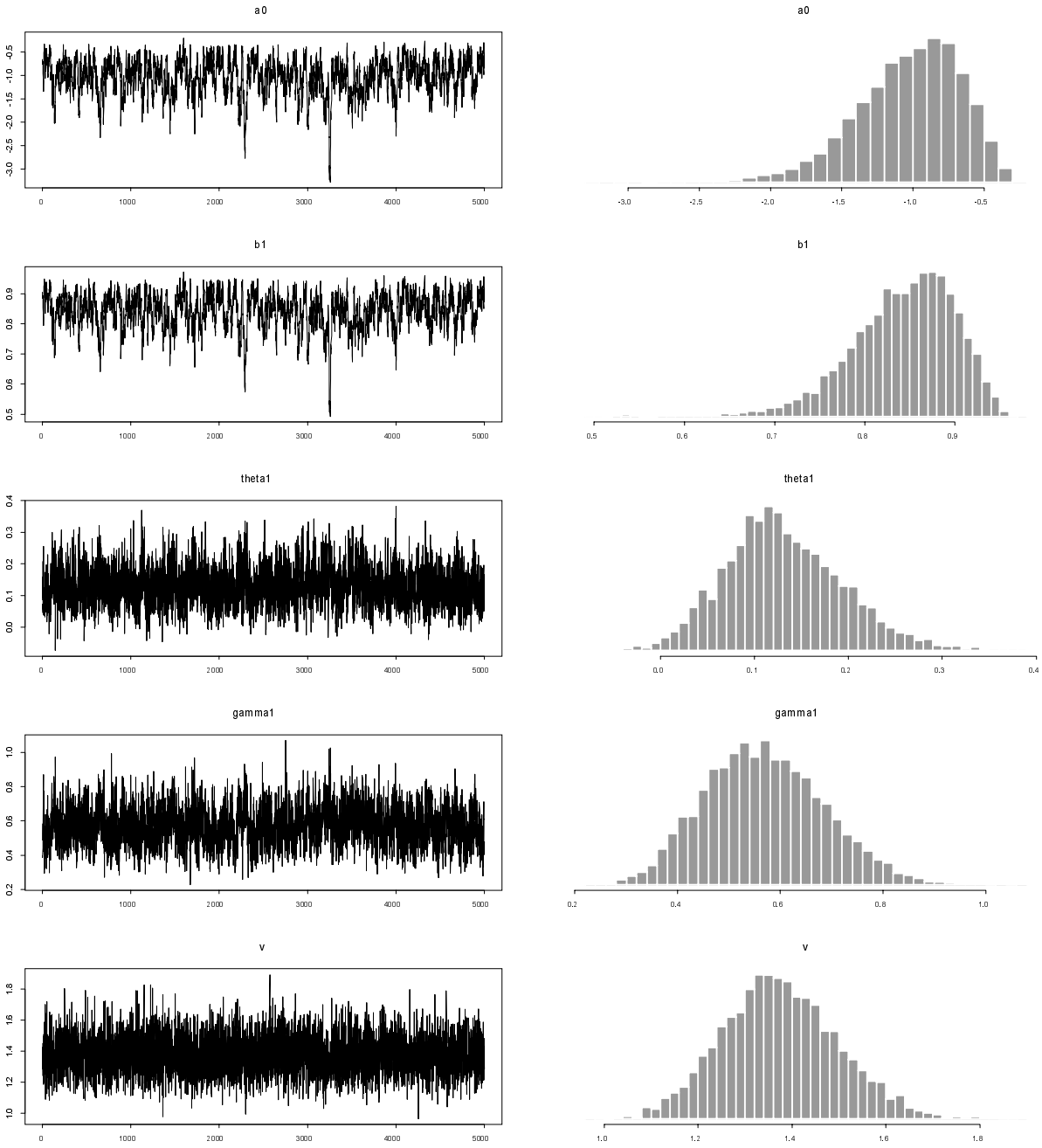

Figure 4-4: Convergence diagrams and histograms of the posterior sample of the parameters of the EGARCH*(1,1) model.

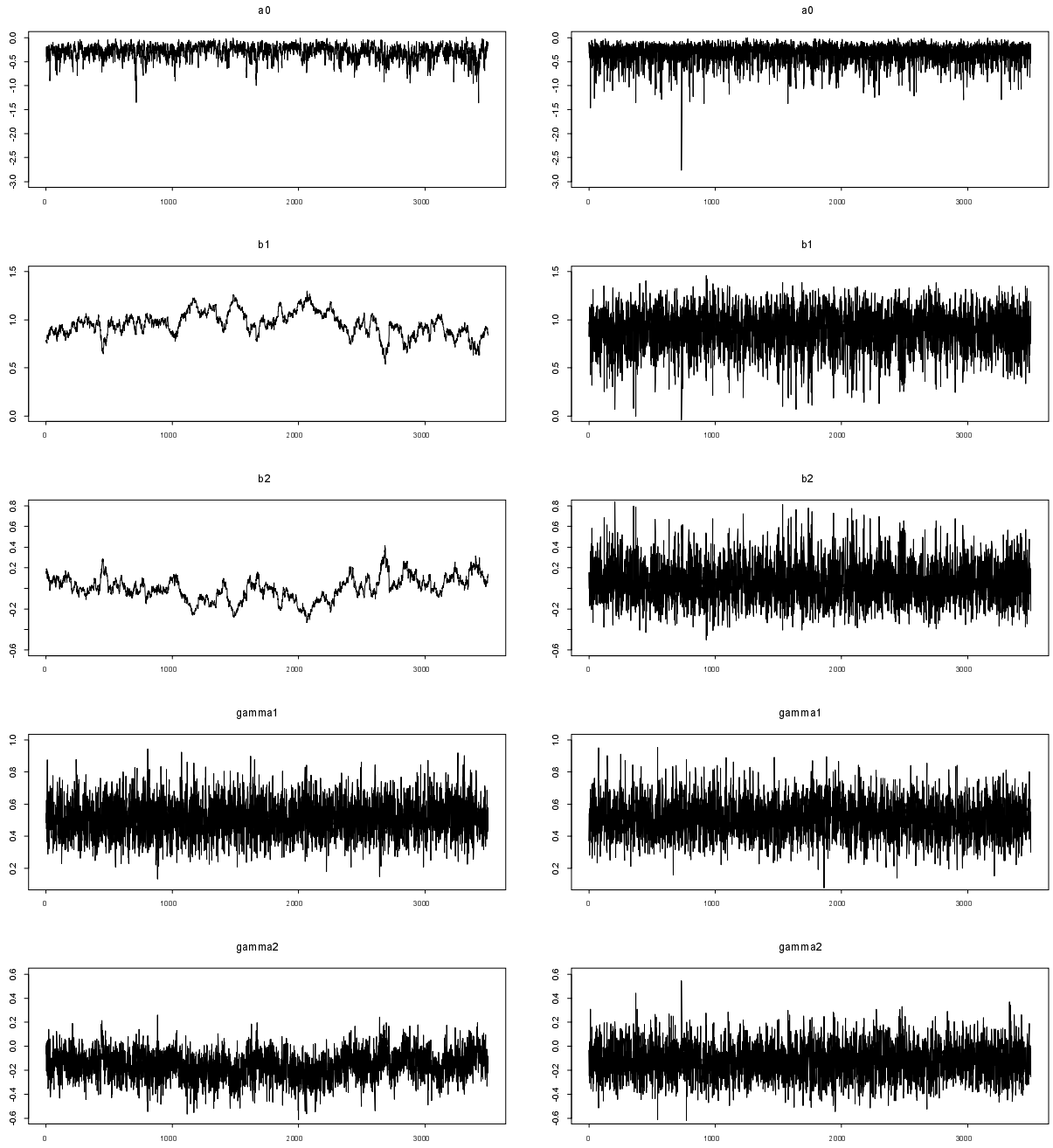

Figure 4-5: MCMC output of the highly correlated parameters of the EGARCH(2,2) model: first column: single component update, second column: simultaneous component update.

Model	$\alpha_0 10^{-3}$	α_1	β_1	$\sigma_0^2 10^{-2}$	n				
$Garch_N(1,1)$	0.13	0.31	0.66	0.32	–				
	0.06	0.08	0.08	0.51	–				
$Garch_T(1,1)$	0.10	0.26	0.71	0.45	6.59				
	0.06	0.07	0.08	0.63	2.14				
Model	α_0	β_1	θ_1	γ_1	v	β_2	θ_2	γ_2	σ_0^2
$Egarch\,(1,1)$	-0.46	0.93	0.08	0.41	1.40	–	–	–	0.01
	0.24	0.04	0.04	0.10	0.12	–	–	–	0.04
$Egarch\,(1,2)$	-0.54	0.55	0.12	0.52	1.43	0.37	–	–	0.02
	0.23	0.19	0.05	0.10	0.13	0.18	–	–	0.07
$Egarch\,(2,1)$	-0.29	0.96	0.27	0.51	1.49	–	-0.21	-0.16	0.01
	0.14	0.02	0.08	0.11	0.14	–	0.08	0.10	0.01
$Egarch\,(2,2)$	-0.33	0.89	0.26	0.51	1.48	0.06	-0.20	-0.12	0.01
	0.19	0.21	0.08	0.11	0.14	0.19	0.14	0.14	0.03
Model	α_0	β_1	θ_1	γ_1	v	β_2	θ_2	γ_2	
$Egarch^*\,(1,1)$	-1.03	0.84	0.13	0.57	1.37	–	–	–	
	0.37	0.05	0.06	0.11	0.12	–	–	–	
$Egarch^*\,(1,2)$	-0.92	0.46	0.15	0.61	1.42	0.40	–	–	
	0.28	0.15	0.06	0.09	0.12	0.16	–	–	
$Egarch^*\,(2,1)$	-0.73	0.89	0.28	0.58	1.44	–	-0.20	-0.08	
	0.30	0.04	0.08	0.11	0.13	–	0.08	0.11	
$Egarch^*\,(2,2)$	-0.99	0.43	0.21	0.57	1.43	0.42	-0.08	0.14	
	0.40	0.26	0.10	0.11	0.13	0.23	0.12	0.16	

Table 4.3: Estimated posterior means and standard deviations for the parameters of the GARCH and EGARCH models.

of the parameters of GARCH(1,1) and EGARCH*(1,1) model respectively. The shapes of the posterior distribution of all the GARCH(1,1) parameters (Figure 4-3) and of the α_0 and β_1 parameters in EGARCH*(1,1) model (Figure 4-4) indicate deviation from normality.

The above strategy enabled us to investigate more deeply the MCMC behavior of GARCH/EGARCH models. Indeed, it soon became evident that, in some models, pairs of parameters exhibit strong posterior correlation. As it is well known (Hills and Smith, 1992) this phenomenon reduces the performance of MCMC algorithm. To deal with this, we used (for the highly correlated parameters) the Metropolis multivariate step described in section 4.3.1. As an example, we report that for the EGARCH(2,2) model we used the

above strategy for the parameters α_0, β_1, β_2, γ_1, γ_2, with $c = 0.90$, and we achieved a reduction in the required number of iterations of at least 30%. In Figure 4-5, we illustrate the MCMC output of the parameters α_0, β_1, β_2, γ_1, γ_2 of the EGARCH(2,2) model using a single component update (Figure 4-5, first column) and a simultaneous component update (Figure 4-5, second column). The results are based on 3500 iterations taken with lag 200 and illustrate the improvement achieved with the multivariate Metropolis algorithm.

Our model selection exercise consists of specifying the order of an EGARCH model as well as testing whether the (usual) prespecification of σ_0^2 affects the inferences made. For the 8 competing models we applied the reversible jump MCMC algorithm and the posterior probabilities in each model are illustrated in Table 4.4, together with the Bayes factors of all models against the (least probable) model EGARCH*(2,1). Based on these results , the "best" model for the Athens Stock market is the EGARCH(2,1) model with posterior probability 0.4772. Proposal densities $q\left(\mathbf{u}|m^{'}\right)$, $q\left(\mathbf{u}^{'}|m\right)$ for each parameter were constructed by using the MCMC output of separate model runs described earlier. These proposals are taken as multivariate normal densities with mean vector consisting of the sample mean values, and covariance matrix equal to the corresponding sample covariance matrix of the parameters in each model. The behavior of the MCMC chain was good with rapid convergence of the probabilities $p\left(m|\mathbf{y}\right)$. We ran the reversible jump MCMC algorithm for 1000000 iterations and in Figure 4-6, we illustrate the probabilities of the 8 different EGARCH models across the sweeps calculated ergodically every 10000 iterations. Due to the large number of models the time for 1000000 iterations was around 8000 minutes in a Pentium II 300 MHz with 128 MB RAM. However, note that this is an extremely conservative run; Figure 4-6 illustrates that a tenth or a fifth of the runs would provide essentially the same results. In the reversible jump MCMC algorithm we used the non-informative priors we suggested for the estimation of the parameters of the EGARCH models.

In order to get some assessment of the robustness of the empirical results we tried five

Models	*Posterior probability*	*Bayes Factor*
$Egarch\,(2,1)$	0.4772	210.5118
$Egarch\,(1,2)$	0.1786	78.7814
$Egarch\,(1,1)$	0.1579	69.6343
$Egarch\,(2,2)$	0.1066	47.0078
$Egarch^{*}\,(1,2)$	0.0620	27.3466
$Egarch^{*}\,(1,1)$	0.0128	5.6520
$Egarch^{*}\,(2,2)$	0.0026	1.1508
$Egarch^{*}\,(2,1)$	0.0023	1.0000

Table 4.4: Posterior probabilities and Bayes Factors of 8 competing EGARCH models.

different Cauchy(μ, σ^2) prior specifications for the model selection of the 4 EGARCH*(p,q) models. The Cauchy(μ, σ^2) prior is preferred by Jeffreys (1961) and the normal analogue of the Cauchy(μ, σ^2) density is a $N\left(\mu, \pi\sigma^2/2\right)$ density according to Berger and Delampady (1987). Therefore, the analogue of the priors we used in the previous analysis is $U\left(-1,1\right)$ for β_1, β_2, Cauchy$(0, 6.4)$ for α_0, θ_i, γ_i, $i = 1, 2$ and Cauchy$(0.7, 6.4)$ for $v^{'} = \ln\left(v\right)$. We used Cauchy$(\mu, 30)$, Cauchy$(\mu, 50)$, Cauchy$(\mu, 100)$, Cauchy$(\mu, 500)$ and Cauchy$(\mu, 5000)$ priors. We ran the reversible jump MCMC algorithm for 1000000 iterations and in Table 4.5 we illustrate the posterior model probabilities using these alternative prior specifications. The time needed was 2310 minutes in a Pentium II 300 MHz with 128 MB RAM. Based on the results of Table 4.5, we gather that the results are robust to alternative prior specifications and in all cases the most probable model is EGARCH*(1,2). The reversible jump MCMC algorithm supports the simplest model EGARCH*(1,1) as the priors become more flat. This characteristic is well known in the Bayesian model selection literature; see Lindley (1957), Bartlett (1957), Kass and Raftery (1995), Berger and Delampady (1987) among others.

Some models are not visited very often in the reversible jump MCMC algorithm; for example, just the 4 top models are visited with probability more than 0.92. Had we required to obtain parameter estimates such as those in Table 4.3 together with posterior model probabilities, we should have tuned the reversible jump MCMC accordingly. For example, it is straightforward to tune the Markov chain so that only a subset of models

$Models$	$Cauchy(\mu, 6.4)$	$Cauchy(\mu, 30)$	$Cauchy(\mu, 50)$
$Egarch^*(1,1)$	0.16080	0.16304	0.16393
$Egarch^*(1,2)$	0.77801	0.71327	0.68659
$Egarch^*(2,1)$	0.02845	0.05381	0.05634
$Egarch^*(2,2)$	0.03274	0.06988	0.09315
$Models$	$Cauchy(\mu, 100)$	$Cauchy(\mu, 500)$	$Cauchy(\mu, 5000)$
$Egarch^*(1,1)$	0.17238	0.18670	0.22080
$Egarch^*(1,2)$	0.67388	0.61706	0.60558
$Egarch^*(2,1)$	0.05751	0.04471	0.04057
$Egarch^*(2,2)$	0.09624	0.15154	0.13306

Table 4.5: Posterior probabilities of the 4 EGARCH* models using alternative prior specifications.

is visited by just changing the probabilities $j\left(m, m^{'}\right)$.

Having obtained estimates of the posterior probabilities of each model, we can apply a Bayesian model averaging procedure to derive estimates of the composite posterior predictive volatility σ^2_{T+1}. In our illustration we calculated the predictive density σ^2_{T+1} based on the 8 EGARCH models. To achieve this, we constructed all predictive densities $p\left(\sigma^2_{T+1}|m, \mathbf{y}\right)$ under each model, for each sampled point in $\boldsymbol{\theta}$, and then we weighted all samples of $p\left(\sigma^2_{T+1}|m, \mathbf{y}\right)$ by the corresponding posterior model probabilities taken from the reversible jump MCMC algorithm. This curve, constructed by using the Splus kernel density estimation command *density()*, is depicted in Figure 4-7 together with the corresponding estimates obtained by the three most probable models.

Finally, note that the need to impose stationarity conditions in a Bayesian context is not well understood and not broadly accepted. In our data application, we relaxed these conditions for our most probable model EGARCH(2,1) and we constructed the posterior predicted volatility σ^2_{T+1}. The posterior sample of the parameters gives probability (0.0033) to non-stationarity (i.e. 0.0033 of the sampled points of β_1 are equal or greater than 1). The density of predicted volatility σ^2_{T+1} is also illustrated in Figure 4-7 and is almost identical with the density of EGARCH(2,1) model when we use stationarity conditions.

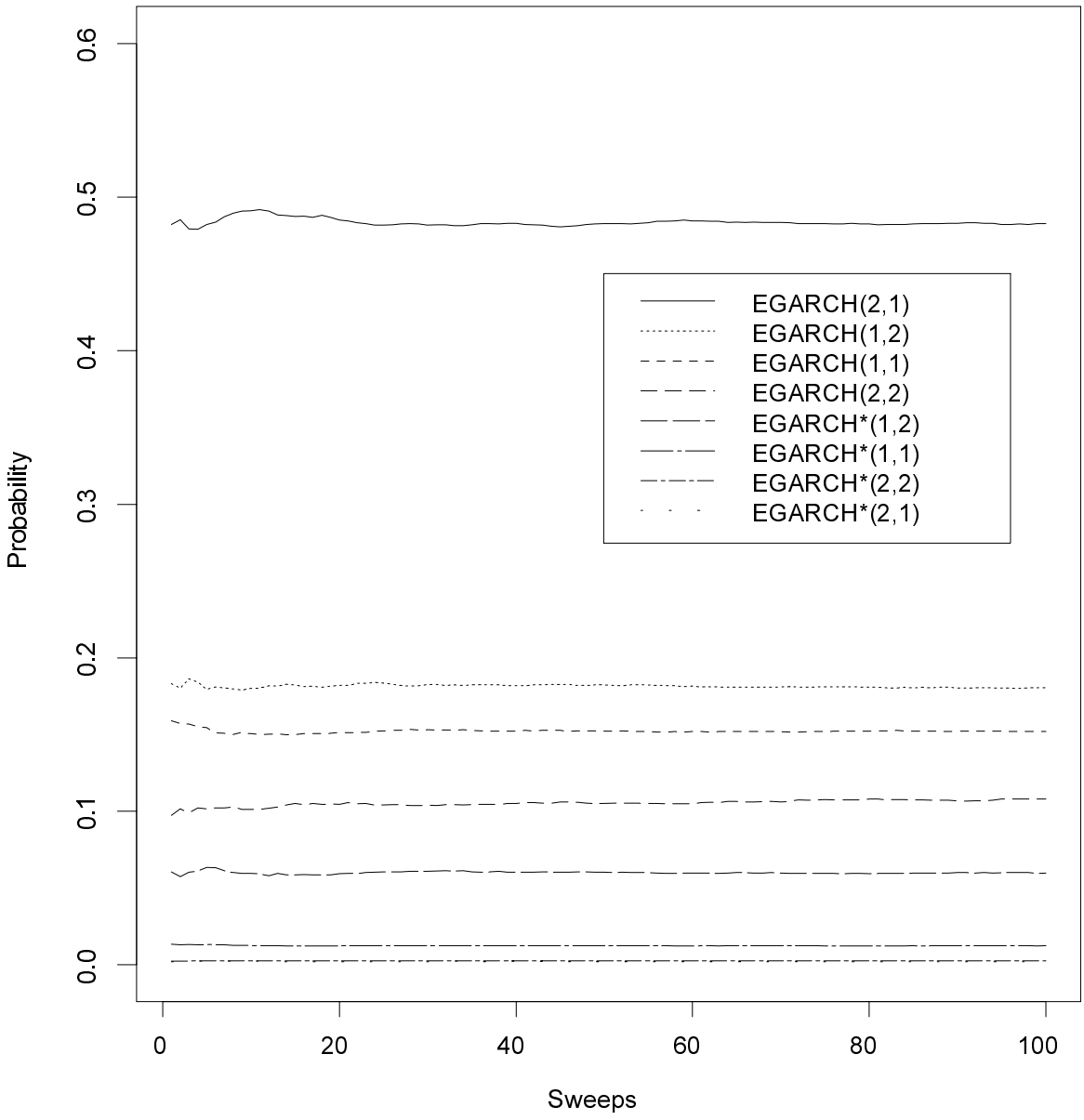

Figure 4-6: Convergence Behaviour of the EGARCH models

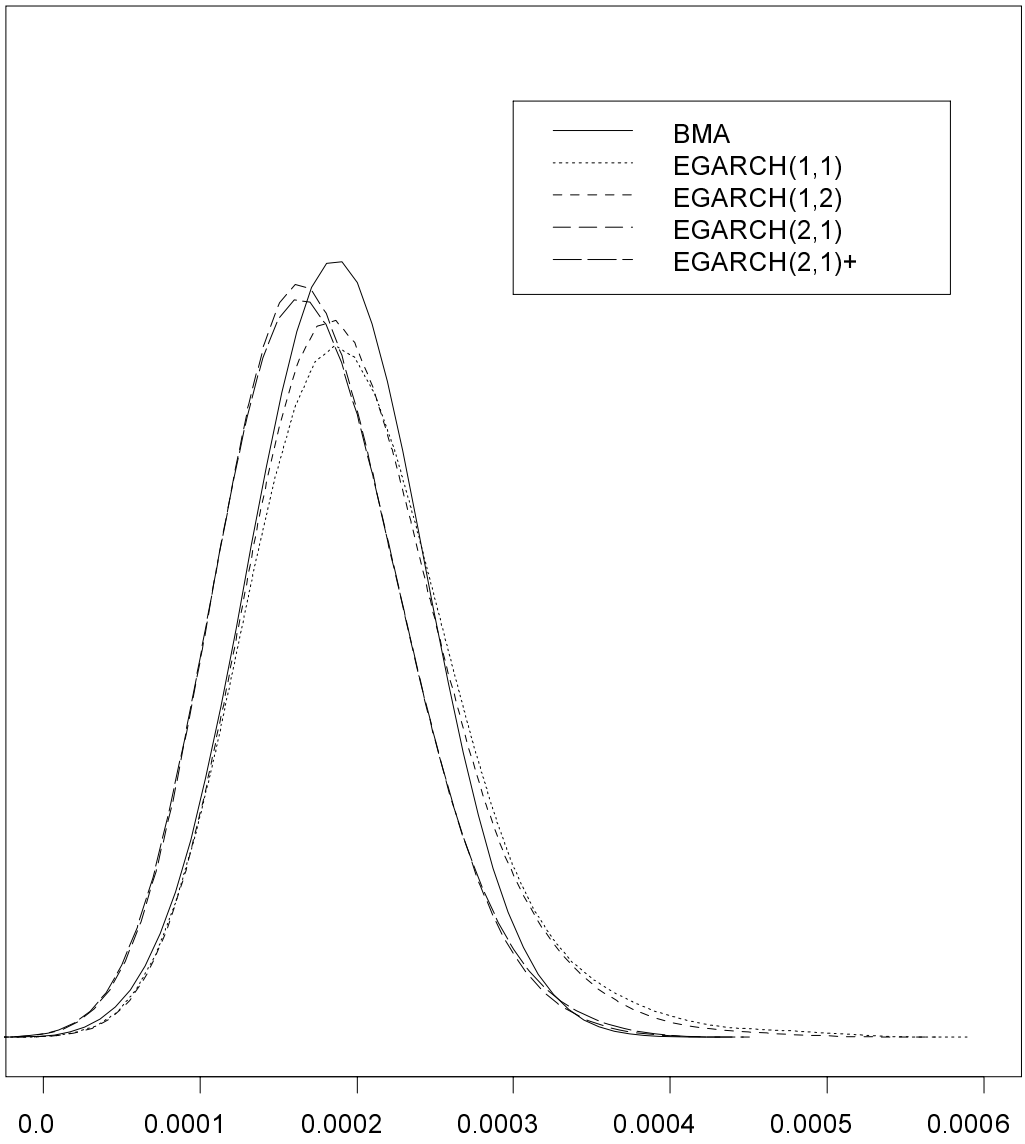

Figure 4-7: Posterior density of Predictive volatility using Bayesian model averaging (BMA), EGARCH(1,1) model with stationarity conditions, EGARCH(1,2) model with stationarity conditions, EGARCH(2,1) model with stationarity conditions, EGARCH(2,1)$^{+}$ model without stationarity conditions.

4.6 Discussion

In this chapter two important issues are considered. The first is related to the Bayesian inference of time-varying volatility models which, we believe, offers some advantages compared to the classical approaches. For example, local maxima do not present a problem and posterior densities of functions of the parameters are easily available. The second is related to the model selection problem, which is usually done via BIC or AIC. These approaches provide no direct information on the reliability of the estimates, and they do not allow prior input for model choice; moreover, their definitions and/or calibrations rely on asymptotic considerations. Our proposed strategy is to use the reversible jump MCMC algorithm to calculate the posterior probability of every proposed model; subsequently, we allow for richer inferences through model averaging. Finally, we note that although we only analyzed the methodology by using GARCH and EGARCH models, our analysis can be extended to variants of these models in a straightforward way.

Chapter 5

Analysis of a class of multivariate ARCH and GARCH models

5.1 Introduction

Univariate and multivariate ARCH-type models were described briefly in chapter three. The multivariate time-varying volatility models have attracted a lot of attention in the statistics/econometrics community. An analysis of a class of multivariate ARCH and GARCH models is proposed consisting of parameter estimation and model comparison. We study the multivariate ARCH model (Kraft and Engle, 1982), the GARCH model with constant conditional correlation (Bollerslev, 1990) which has attracted a lot of practical interest, a generalization of that model (Jeantheau, 1998), and some models that are special cases of model of Jeantheau (1998). Bayesian and classical techniques are used for the estimation of the parameters of the models analyzed, and model comparisons are addressed via predictive distributions. We use a bivariate example to apply the models and provide implementation details and illustrations using daily exchange rates of the Athens exchange market.

5.2 The analyzed models

We consider having observed data of the form

$$\mathbf{y}_t,\ t = 1, ..., T,$$

where $\mathbf{y}_t = (y_{1t},\ y_{2t})'$.

5.2.1 Bivariate ARCH model

The multivariate ARCH(p) model was introduced by Kraft and Engle (1982). In our study we analyze the first order bivariate diagonal ARCH model, which can be written as

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t \\
\boldsymbol{\varepsilon}_t|\Phi_{t-1} &\sim N_2(\mathbf{0}, \boldsymbol{\Sigma}_t) \\
\boldsymbol{\Sigma}_t &= \begin{bmatrix} \sigma^2_{1,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma^2_{2,t} \end{bmatrix} \\
\sigma^2_{1,t} &= c_{11} + a_{11}\varepsilon^2_{1,t-1} \\
\sigma_{21,t} &= c_{21} + a_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} \\
\sigma^2_{2,t} &= c_{22} + a_{33}\varepsilon^2_{2,t-1},
\end{aligned} \tag{5.1}$$

where $\boldsymbol{\mu}$ is a 2×1 vector of constants, $\boldsymbol{\varepsilon}_t$ is a 2×1 innovation vector, Φ_{t-1} is the information set up to time $t-1$, $\boldsymbol{\Sigma}_t$ is 2×2 covariance matrix with elements $\sigma^2_{i,t}$, $i = 1, 2$, where $\sigma^2_{i,t}$ is the variance of the $i - th$ variable at time t, and $\sigma_{12,t}$ is the covariance of the first with the second variable at time t. Positive definiteness of $\boldsymbol{\Sigma}_t$ requires that

$$\begin{aligned}
c_{11} &> 0,\ c_{22} > 0,\ c_{11}c_{22} - c^2_{21} > 0, \\
a_{11} &\geq 0,\ a_{33} \geq 0,\ a_{11}a_{33} - a^2_{22} \geq 0.
\end{aligned}$$

$\{\boldsymbol{\varepsilon}_t\}$ is covariance stationary if and only if the eigenvalues of $\mathbf{A}_1$ in (3.12) are less than one in modulus. That is, if and only if the parameters a_{11}, a_{22}, and a_{33} are less than one in modulus. The parameter vector to be estimated in (5.1) is $\boldsymbol{\theta}$ =(μ_1, μ_2, c_{11}, c_{21}, c_{22}, a_{11}, a_{22}, a_{33}). Under the assumption of conditional normality for the error process, the likelihood for model (5.1) for a sample of T observations $\mathbf{y}$ can be written as

$$l\left(\mathbf{y}|\boldsymbol{\theta}\right) = \left(2\pi\right)^{-T} \prod_{t=1}^{T} \left\{ \left|\boldsymbol{\Sigma}_t\right|^{-\frac{1}{2}} \right\} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \left(\mathbf{y}_t - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}_t^{-1} \left(\mathbf{y}_t - \boldsymbol{\mu}\right) \right\}. \tag{5.2}$$

5.2.2 Bivariate GARCH model with constant conditional correlation

The bivariate Generalized ARCH model with constant conditional correlation (Bollerslev, 1990) can be written as

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t \\
\boldsymbol{\varepsilon}_t|\Phi_{t-1} &\sim N_2\left(\mathbf{0}, \boldsymbol{\Sigma}_t\right) \\
\boldsymbol{\Sigma}_t &= \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{2,t}^2 \end{bmatrix} \\
\sigma_{1,t}^2 &= a_{01} + a_{11}\varepsilon_{1,t-1}^2 + \beta_{11}\sigma_{1,t-1}^2 \\
\sigma_{2,t}^2 &= a_{02} + a_{21}\varepsilon_{2,t-1}^2 + \beta_{21}\sigma_{2,t-1}^2 \\
\sigma_{12,t} &= \rho_{12}\left(\sigma_{1,t}^2\sigma_{2,t}^2\right)^{1/2},
\end{aligned} \tag{5.3}$$

where $\boldsymbol{\mu}$ is a 2×1 vector of constants, $\boldsymbol{\varepsilon}_t$ is a 2×1 innovation vector, Φ_{t-1} is the information set up to time $t-1$, $\boldsymbol{\Sigma}_t$ is 2×2 covariance matrix where $\sigma_{i,t}^2$ is the variance of the $i-th$ variable at time t, $\sigma_{12,t}$ is the covariance of the first with the second variable at time t, ρ_{12} is the correlation of the first to the second variable which is constant over time. For positive definiteness of $\boldsymbol{\Sigma}_t$, we need $a_{0i} > 0$, $a_{i1} \geq 0$, $\beta_{i1} \geq 0$, $i = 1, 2$ and $-1 < \rho_{12} < 1$. For finite variance and stationarity, it is also necessary to impose that

$a_{i1}+\beta_{i1} < 1$ for $i = 1, 2$. The parameter vector to be estimated for model (5.3) is $\boldsymbol{\theta} = (\mu_1,$ $\mu_2,\ a_{01},\ a_{02},\ a_{11},\ a_{21},\ \beta_{11},\ \beta_{21},\ \rho_{12})$.

5.2.3 Other bivariate ARCH and GARCH models

The bivariate version of the multivariate GARCH model of Jeantheau (1998) is also analyzed. The model is a generalization of the previous GARCH model and can be written as

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t \\
\boldsymbol{\varepsilon}_t|\Phi_{t-1} &\sim N_2\left(\mathbf{0}, \boldsymbol{\Sigma}_t\right) \\
\boldsymbol{\Sigma}_t &= \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{2,t}^2 \end{bmatrix} \\
\sigma_{1,t}^2 &= a_{10} + a_{11}\varepsilon_{1,t-1}^2 + a_{12}\varepsilon_{2,t-1}^2 + \beta_{11}\sigma_{1,t-1}^2 + \beta_{12}\sigma_{2,t-1}^2 \\
\sigma_{2,t}^2 &= a_{20} + a_{21}\varepsilon_{1,t-1}^2 + a_{22}\varepsilon_{2,t-1}^2 + \beta_{21}\sigma_{1,t-1}^2 + \beta_{22}\sigma_{2,t-1}^2 \\
\sigma_{12,t} &= \rho_{12}\left(\sigma_{1,t}^2\sigma_{2,t}^2\right)^{1/2},
\end{aligned} \tag{5.4}$$

where $\boldsymbol{\mu}$ is a 2×1 vector of constants, $\boldsymbol{\varepsilon}_t$ is a 2×1 innovation vector, Φ_{t-1} is the information set up to time $t-1$, $\boldsymbol{\Sigma}_t$ is 2×2 covariance matrix where $\sigma_{i,t}^2$ is the variance of the $i-th$ variable at time t, $\sigma_{12,t}$ is the covariance of the first with the second variable at time t, ρ_{12} is the correlation of the first to the second variable which is constant over time. The covariance matrix is positive definite if $a_{i0} > 0$, $a_{ij}, \beta_{ij} \geq 0$, for $i, j = 1, 2$ and $-1 < \rho_{12} < 1$. The parameter vector is $\boldsymbol{\theta} = (\mu_1,\ \mu_2,\ a_{10},\ a_{11},\ a_{12},\ \beta_{11},\ \beta_{12},\ a_{20},\ a_{21},\ a_{22},$ $\beta_{21},\ \beta_{22},\ \rho_{12})$.

Some special cases of the above model are also analyzed: an ARCH model with diagonal covariance matrix, an ARCH model with constant conditional correlation, and a GARCH model with diagonal covariance matrix. That is, we analyze the following

bivariate ARCH model

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t \\
\boldsymbol{\varepsilon}_t|\Phi_{t-1} &\sim N_2\left(\mathbf{0}, \boldsymbol{\Sigma}_t\right) \\
\boldsymbol{\Sigma}_t &= \begin{bmatrix} \sigma_{1,t}^2 & 0 \\ 0 & \sigma_{2,t}^2 \end{bmatrix} \\
\sigma_{1,t}^2 &= a_{10} + a_{11}\varepsilon_{1,t-1}^2 + a_{12}\varepsilon_{2,t-1}^2 \\
\sigma_{2,t}^2 &= a_{20} + a_{21}\varepsilon_{1,t-1}^2 + a_{22}\varepsilon_{2,t-1}^2,
\end{aligned} \tag{5.5}$$

where $\boldsymbol{\mu}$ is a 2×1 vector of constants, $\boldsymbol{\varepsilon}_t$ is a 2×1 innovation vector, Φ_{t-1} is the information set up to time $t-1$, $\boldsymbol{\Sigma}_t$ is 2×2 covariance matrix where $\sigma_{i,t}^2$ is the variance of the $i-th$ variable at time t, $a_{10} > 0$, $a_{20} > 0$, $a_{11} \geq 0$, $a_{12} \geq 0$, $a_{21} \geq 0$, $a_{22} \geq 0$. Under these restrictions the variances are well defined and the covariance matrix is positive definite. The parameter vector to be estimated in (5.5) is $\boldsymbol{\theta} = (\mu_1, \mu_2, a_{10}, a_{11}, a_{12}, a_{20}, a_{21}, a_{22})$. The likelihood under conditional normality for the error process $\boldsymbol{\varepsilon}_t$ for a sample of T observations $\mathbf{y}$ for model (5.5) is again given by formula (5.2).

To capture the time varying conditional covariance we also analyze the following constant conditional correlation ARCH model:

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t \\
\boldsymbol{\varepsilon}_t|\Phi_{t-1} &\sim N_2\left(\mathbf{0}, \boldsymbol{\Sigma}_t\right) \\
\boldsymbol{\Sigma}_t &= \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{2,t}^2 \end{bmatrix} \\
\sigma_{1,t}^2 &= a_{10} + a_{11}\varepsilon_{1,t-1}^2 + a_{12}\varepsilon_{2,t-1}^2 \\
\sigma_{2,t}^2 &= a_{20} + a_{21}\varepsilon_{1,t-1}^2 + a_{22}\varepsilon_{2,t-1}^2 \\
\sigma_{12,t} &= \rho_{12}\left(\sigma_{1t}^2\sigma_{2t}^2\right)^{1/2},
\end{aligned} \tag{5.6}$$

where $\boldsymbol{\mu}$ is a 2×1 vector of constants, $\boldsymbol{\varepsilon}_t$ is a 2×1 innovation vector, Φ_{t-1} is the

information set up to time $t-1$, $\mathbf{\Sigma}_t$ is 2×2 covariance matrix where $\sigma^2_{i,t}$ is the variance of the $i-th$ variable at time t, $\sigma_{12,t}$ is the covariance of the first with the second variable at time t, ρ_{12} is the correlation of the first to the second variable which is constant over time. The conditions for positive definiteness of the matrix $\mathbf{\Sigma}_t$ are $a_{10} > 0$, $a_{20} > 0$, $a_{11} \geq 0$, $a_{12} \geq 0$, $a_{21} \geq 0$, $a_{22} \geq 0$ and $-1 \leq \rho_{12} \leq 1$. The parameter vector of this model (5.6) is $\boldsymbol{\theta} = (\mu_1,\, \mu_2,\, a_{10},\, a_{11},\, a_{12},\, a_{20},\, a_{21},\, a_{22},\, \rho_{12})$.

We also consider a bivariate GARCH model of the form

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t \\
\boldsymbol{\varepsilon}_t|\Phi_{t-1} &\sim N_2\left(\mathbf{0}, \mathbf{\Sigma}_t\right) \\
\mathbf{\Sigma}_t &= \begin{bmatrix} \sigma^2_{1,t} & 0 \\ 0 & \sigma^2_{2,t} \end{bmatrix} \\
\sigma^2_{1,t} &= a_{10} + a_{11}\varepsilon^2_{1,t-1} + a_{12}\varepsilon^2_{2,t-1} + \beta_{11}\sigma^2_{1,t-1} + \beta_{12}\sigma^2_{2,t-1} \\
\sigma^2_{2,t} &= a_{20} + a_{21}\varepsilon^2_{1,t-1} + a_{22}\varepsilon^2_{2,t-1} + \beta_{21}\sigma^2_{1,t-1} + \beta_{22}\sigma^2_{2,t-1},
\end{aligned} \tag{5.7}$$

where $\boldsymbol{\mu}$ is a 2×1 vector of constants, $\boldsymbol{\varepsilon}_t$ is a 2×1 innovation vector, Φ_{t-1} is the information set up to time $t-1$, $\mathbf{\Sigma}_t$ is 2×2 covariance matrix where $\sigma^2_{i,t}$ is the variance of the $i-th$ variable at time t. The covariance matrix is positive definite if $a_{i0} > 0$, $a_{ij}, \beta_{ij} \geq 0$, for $i, j = 1, 2$. The parameter vector is $\boldsymbol{\theta} = (\mu_1,\, \mu_2,\, a_{10},\, a_{11},\, a_{12},\, \beta_{11},\, \beta_{12},\, a_{20},\, a_{21},\, a_{22},\, \beta_{21},\, \beta_{22})$. The likelihood under conditional normality for the error process $\boldsymbol{\varepsilon}_t$ for a sample of T observations $\mathbf{y}$ for model (5.7) is given by formula (5.2).

5.3 Inference and model determination

5.3.1 Bayesian inference

We adopt Bayesian methodology by constructing MCMC algorithms. These algorithms allow us to simulate from the joint posterior distribution of the parameters of the above models. We use Metropolis-Hastings algorithm; see, for example, Chib and Greenberg

(1995). We update simultaneously the elements of the parameter vector $\boldsymbol{\theta}$ using a multivariate normal proposal based on pilot run estimates. That is, we estimate from an initial explanatory run of the Markov chain (based on univariate random walk metropolis) the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$ related to these parameters. Then, the update from time t to time $t+1$ is achieved by using a multivariate normal proposal density $N\left(\boldsymbol{\mu}^t, c\widehat{\boldsymbol{\Sigma}}\right)$, with $\boldsymbol{\mu}^t$ denoting the vector at time t and c a constant to tune the acceptance rate.

5.3.2 Classical inference

We also consider the estimation of the parameters of the multivariate ARCH and GARCH models by using classical approaches. By using different computational methods we check the results of the models we analyzed and compare the estimates under the different approaches. Maximum likelihood estimates are taken by using numerical optimization algorithms such as Newton-Raphson, Fisher scoring, and the method proposed by Mak (1993) and developed further by Mak, Wong and Li (1997) for nonlinear time series with conditional heteroscedastic variances.

5.3.3 Model determination

Our implementation is also concerned with model comparison by using predictive distributions of the time-varying volatilities. This is sensible both because of the predictive nature of the applicability of the models and because volatilities are of primary interest. We choose to work with a criterion suggested by Gelfand, Dey and Chang (1992). Similar diagnostic measures are presented by Pitt and Shephard (1999) in order to compare stochastic volatility models. The MCMC techniques produce a sample from the joint posterior $p(\boldsymbol{\theta}|\mathbf{y})$. Hence, the outputted $\boldsymbol{\theta}_s$, $s = 1, ..., B$ can be used to carry out computations needed for model comparison. The estimate of the predictive density is $\widehat{p}(y_{T+1}|\mathbf{y}) = B^{-1}\sum_{s=1}^{B} p(y_{T+1}|\boldsymbol{\theta}_s, \mathbf{y})$. To estimate P one-step-ahead predictive densities $\widehat{p}(y_{T+1}|\mathbf{y})$ we use the previous $T, T+1, ..., T+P-1$ data points. Between two models

Model	"Burn-in" period	Total No. of iterations	No. of iterations to keep
ARCH (5.1) (sc)	1000	5000	4000
ARCH (5.1) (wsc)	2150	5000	2850
ARCH (5.5) (sc)	1800	5000	3200
ARCH (5.5) (wsc)	750	5000	4250
ARCH (5.6) (sc)	650	5000	4350
ARCH (5.6) (wsc)	900	5000	4100
GARCH (5.7) (sc)	2250	5000	2750
GARCH (5.7) (wsc)	450	5000	4550
GARCH (5.3) (sc)	1050	5000	3950
GARCH (5.3) (wsc)	1050	5000	3950
GARCH (5.4) (sc)	300	5000	4700
GARCH (5.4) (wsc)	800	5000	4200

Table 5.1: Estimated burn-in period using the subsampling diagnostic. sc: imposing stationarity conditions in the MCMC algorithms, wsc: without using stationarity conditions in the MCMC algorithms.

m_i and m_j , we choose model m_i if

$$\log\left[\frac{\prod^{(m_i)}\widehat{p}\,(y_{t+1}|\mathbf{y})}{\prod^{(m_j)}\widehat{p}\,(y_{t+1}|\mathbf{y})}\right] > 0,$$

where $\prod^{(m_i)}\widehat{p}\,(y_{t+1}|\mathbf{y})$ is the product of all $\widehat{p}\,(y_{t+1}|\mathbf{y})$, $t = T, ..., T+P-1$ under model m_i . A more detailed description of the criterion is presented in section 1.3.5.

5.4 An application to the Athens Exchange Market

We illustrate the multivariate models under consideration using daily exchange rates of the USA dollar and German marc with respect to the Greek drachma over the 15/12/1993 - 24/2/1997 period. If x_t is the value of one exchange return at time t, then we model the daily exchange rate $y_t = \ln(x_t/x_{t-1})$, $t = 1, ..., T$, $T = 800$. In Figure 5-1, we present the analyzed exchange rates.

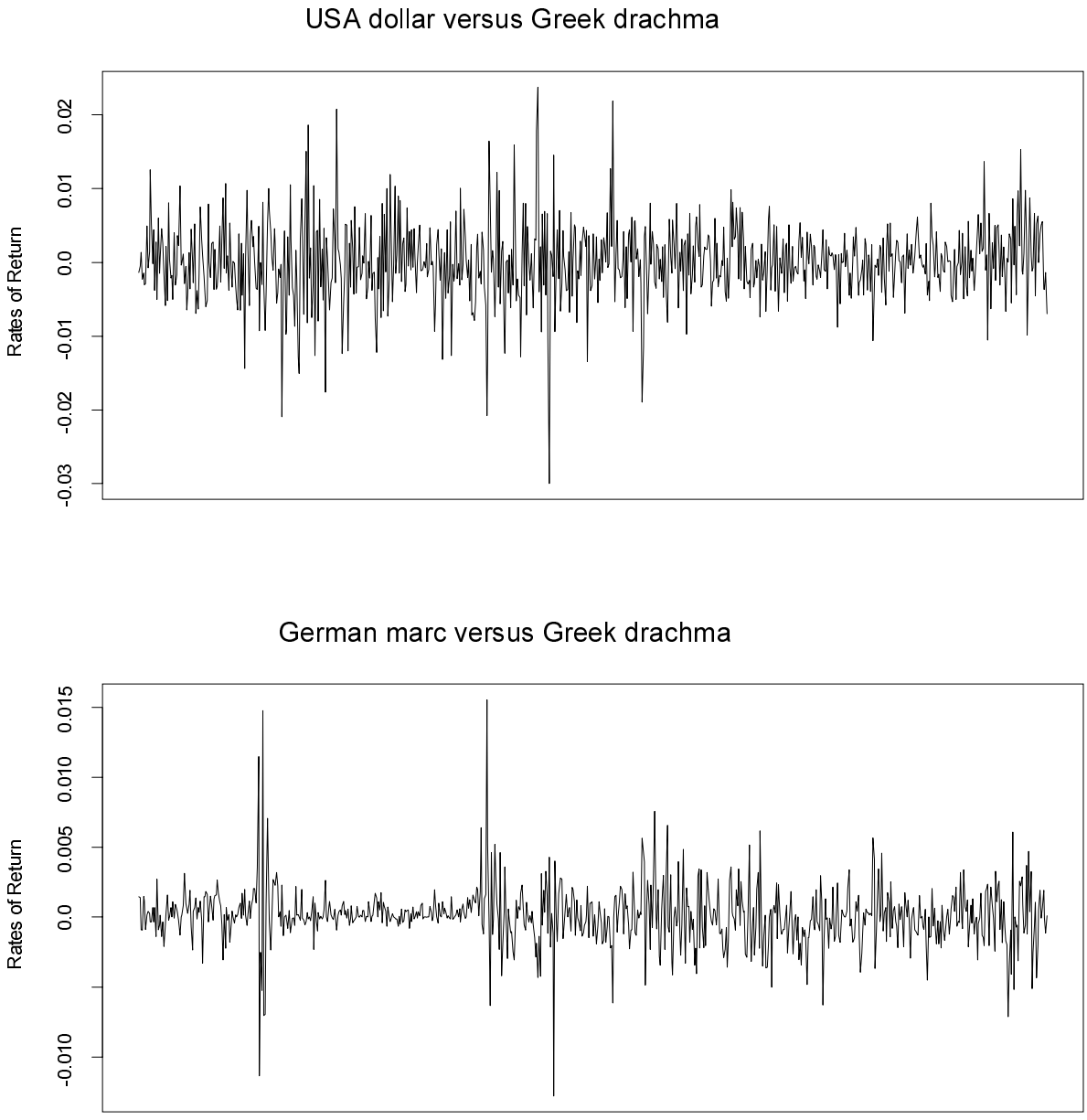

Figure 5-1: The analyzed exchange rates

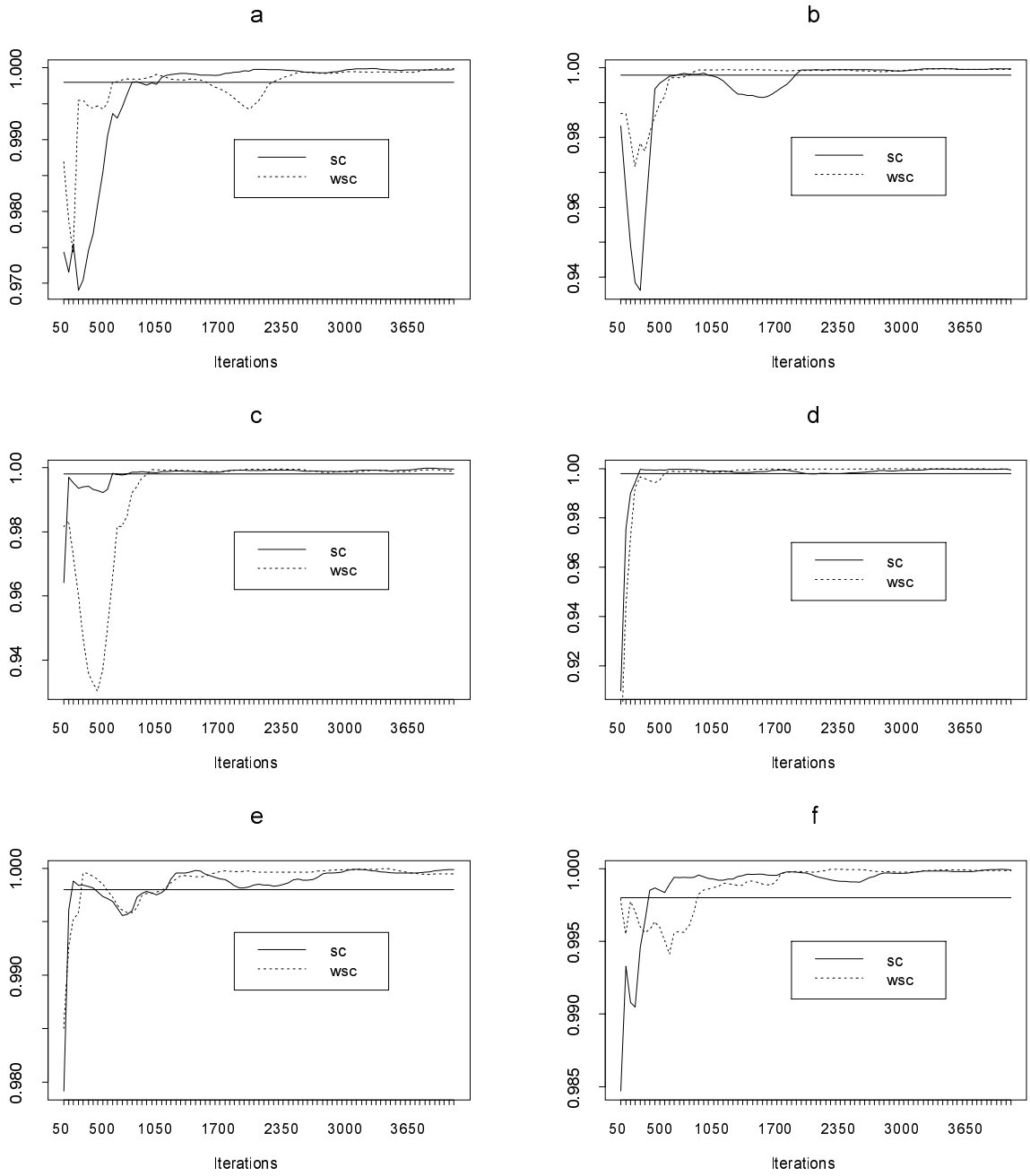

Figure 5-2: Coefficient of determination in the subsampling convergence diagnostic: a: ARCH model (5.1), b: ARCH model (5.5), c: ARCH model (5.6), d: GARCH model (5.7), e: GARCH model (5.3), f: GARCH model (5.4).

Parameter	Bayesian approach		Classical approach
	sc	wsc	
μ_1	$1.5 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$	$1.6 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$	$1.3 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$
μ_2	$1.2 \cdot 10^{-4}$ $(6 \cdot 10^{-5})$	$1.2 \cdot 10^{-4}$ $(6 \cdot 10^{-5})$	$1.2 \cdot 10^{-4}$ $(5 \cdot 10^{-5})$
c_{11}	$2.4 \cdot 10^{-5}$ $(1 \cdot 10^{-6})$	$2.4 \cdot 10^{-5}$ $(1 \cdot 10^{-6})$	$2.4 \cdot 10^{-5}$ $(1 \cdot 10^{-6})$
c_{21}	$-3.2 \cdot 10^{-6}$ $(4 \cdot 10^{-7})$	$-3.2 \cdot 10^{-6}$ $(4 \cdot 10^{-7})$	$-3.1 \cdot 10^{-6}$ $(3 \cdot 10^{-7})$
c_{22}	$2.5 \cdot 10^{-6}$ $(2 \cdot 10^{-7})$	$2.5 \cdot 10^{-6}$ $(2 \cdot 10^{-7})$	$2.4 \cdot 10^{-6}$ $(9 \cdot 10^{-8})$
a_{11}	$1.4 \cdot 10^{-1}$ $(4 \cdot 10^{-2})$	$1.4 \cdot 10^{-1}$ $(4 \cdot 10^{-2})$	$1.3 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$
a_{22}	$1.9 \cdot 10^{-1}$ $(6 \cdot 10^{-2})$	$1.9 \cdot 10^{-1}$ $(5 \cdot 10^{-2})$	$2.0 \cdot 10^{-1}$ $(5 \cdot 10^{-2})$
a_{33}	$6.2 \cdot 10^{-1}$ $(1 \cdot 10^{-1})$	$6.2 \cdot 10^{-1}$ $(1 \cdot 10^{-1})$	$5.8 \cdot 10^{-1}$ $(5 \cdot 10^{-2})$

Table 5.2: Estimates and standard deviation (in brackets) for the parameters of bivariate ARCH model (5.1). sc: imposing stationarity conditions, wsc: without using stationarity conditions.

5.4.1 Bayesian approach

For our illustration we chose non-informative priors for all model parameters. For the parameters of the bivariate ARCH model (5.1) we used non-informative priors $p(a_{11}) = a_{11}^{-1}$, $p(a_{33}) = a_{33}^{-1}$, $p(c_{11}) = c_{11}^{-1}$, $p(c_{22}) = c_{22}^{-1}$, $p(\mu_1,\ \mu_2,\ c_{21},\ a_{22}) = 1$. In the other bivariate ARCH and GARCH models, we transformed the positive parameters to "near normality" using the logarithmic transformation. These transformations improved the behavior of our MCMC algorithm. Then, we applied a constant prior to the transformed parameters. For the parameter ρ_{12} of these models we used uniform $U(-1, 1)$ prior. The need to impose stationarity conditions in a Bayesian context is not broadly accepted. For this reason, in our study, we estimated the model parameters using stationarity conditions by rejecting, in the MCMC algorithm, the parameters which did not obey these restrictions (see, for example, Gelfand, Smith and Lee, 1992), but we also estimate the model parameters without imposing stationarity conditions.

For our data set we ran the algorithms for 500000 iterations, and we kept one value every 100 iterations (to save computer space). The resulting samples of 5000 values were checked for convergence by using the subsampling diagnostic proposed by Giakoumatos, Vrontos, Dellaportas and Politis (1999). The method is based on the use of subsampling for the construction of confidence regions for the t-quantile ($t = 0.90$) of the unique

Parameter	Bayesian approach		Classical approach
	sc	wsc	
μ_1	$1.6 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$	$1.6 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$	$1.6 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$
μ_2	$1.4 \cdot 10^{-4}$ $(6 \cdot 10^{-5})$	$1.4 \cdot 10^{-4}$ $(6 \cdot 10^{-5})$	$1.4 \cdot 10^{-4}$ $(6 \cdot 10^{-5})$
a_{10}	$2.3 \cdot 10^{-5}$ $(2 \cdot 10^{-6})$	$2.3 \cdot 10^{-5}$ $(2 \cdot 10^{-6})$	$2.4 \cdot 10^{-5}$ $(2 \cdot 10^{-6})$
a_{11}	$1.6 \cdot 10^{-1}$ $(5 \cdot 10^{-2})$	$1.5 \cdot 10^{-1}$ $(5 \cdot 10^{-2})$	$1.5 \cdot 10^{-1}$ $(5 \cdot 10^{-2})$
a_{12}	$1.7 \cdot 10^{-1}$ $(1 \cdot 10^{-1})$	$1.7 \cdot 10^{-1}$ $(1 \cdot 10^{-1})$	$3.9 \cdot 10^{-2}$ $(1 \cdot 10^{-1})$
a_{20}	$2.3 \cdot 10^{-6}$ $(2 \cdot 10^{-7})$	$2.3 \cdot 10^{-6}$ $(2 \cdot 10^{-7})$	$2.4 \cdot 10^{-6}$ $(2 \cdot 10^{-7})$
a_{21}	$5.1 \cdot 10^{-3}$ $(4 \cdot 10^{-3})$	$5.0 \cdot 10^{-3}$ $(4 \cdot 10^{-3})$	$2.8 \cdot 10^{-3}$ $(4 \cdot 10^{-3})$
a_{22}	$6.4 \cdot 10^{-1}$ $(1 \cdot 10^{-1})$	$6.4 \cdot 10^{-1}$ $(1 \cdot 10^{-1})$	$6.3 \cdot 10^{-1}$ $(9 \cdot 10^{-2})$

Table 5.3: Estimates and standard deviation (in brackets) for the parameters of bivariate ARCH model (5.5). sc: imposing stationarity conditions, wsc: without using stationarity conditions.

invariant distribution of the Markov chain. We construct the $(1-a)\,100\%$ confidence regions for the 0.90 quantile ($a = 0.05$) based on different (increasing) values $N_j = jN/100$, $j = 1, 2, ..., 100$. We estimate the "burn-in" to be N^* if the "range" of the confidence regions versus $1/\sqrt{N_j}$ is approximately linear for $N > N^*$. Linearity is checked by using the coefficient of determination of the weighted linear regression between the dependent variable "range" and $1/\sqrt{N_j}$, $j = 1, 2, ..., 100$. The reason that the t-quantile (with a large t, say $t = 0.90$) is considered, is based on the notion that stabilization of estimates of the invariant distribution of the Markov chain (especially in the tails) is a reliable indicator of the target distribution having been achieved. We stop the MCMC simulation when the range of this $(1-a)\,100\%$ confidence region for the mean is appropriately small. Using the MCMC chains of $N = 5000$ iterations and choosing as threshold value $d = 0.998$ for the coefficient of determination in the subsampling convergence diagnostic we estimate the burn-in period. A detailed description of the subsampling diagnostic is presented in chapter two. The "burn-in" period of the MCMC algorithms of the analyzed models is presented in Table 5.1, and in Figure 5-2 the coefficient of determination of the subsampling diagnostic is illustrated for all the analyzed models. Estimated posterior means and standard deviations, after we have discarded the "burn-in" period, for the parameters of the models under consideration are presented in Tables 5.2 to 5.7,

Parameter	Bayesian approach		Classical approach
	sc	wsc	
μ_1	$1.4 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$	$1.4 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$	$1.4 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$
μ_2	$1.2 \cdot 10^{-4}$ $(6 \cdot 10^{-5})$	$1.3 \cdot 10^{-4}$ $(6 \cdot 10^{-5})$	$1.2 \cdot 10^{-4}$ $(6 \cdot 10^{-5})$
a_{10}	$2.3 \cdot 10^{-5}$ $(1 \cdot 10^{-6})$	$2.3 \cdot 10^{-5}$ $(2 \cdot 10^{-6})$	$2.3 \cdot 10^{-5}$ $(1 \cdot 10^{-6})$
a_{11}	$1.4 \cdot 10^{-1}$ $(5 \cdot 10^{-2})$	$1.4 \cdot 10^{-1}$ $(5 \cdot 10^{-2})$	$1.3 \cdot 10^{-1}$ $(4 \cdot 10^{-2})$
a_{12}	$2.9 \cdot 10^{-1}$ $(2 \cdot 10^{-1})$	$2.9 \cdot 10^{-1}$ $(2 \cdot 10^{-1})$	$1.9 \cdot 10^{-1}$ $(2 \cdot 10^{-1})$
a_{20}	$2.2 \cdot 10^{-6}$ $(2 \cdot 10^{-7})$	$2.2 \cdot 10^{-6}$ $(2 \cdot 10^{-7})$	$2.2 \cdot 10^{-6}$ $(2 \cdot 10^{-7})$
a_{21}	$4.3 \cdot 10^{-3}$ $(3 \cdot 10^{-3})$	$4.1 \cdot 10^{-3}$ $(3 \cdot 10^{-3})$	$2.1 \cdot 10^{-3}$ $(3 \cdot 10^{-3})$
a_{22}	$7.2 \cdot 10^{-1}$ $(1 \cdot 10^{-1})$	$7.3 \cdot 10^{-1}$ $(1 \cdot 10^{-1})$	$7.2 \cdot 10^{-1}$ $(1 \cdot 10^{-1})$
ρ_{12}	$-3.6 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$-3.6 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$-3.6 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$

Table 5.4: Estimates and standard deviation (in brackets) for the parameters of bivariate ARCH model with constant conditional correlation (5.6). sc: imposing stationarity conditions, wsc: without using stationarity conditions.

while the convergence diagrams and the histograms of the posterior sample of the model parameters are illustrated in Figures 5-3 to 5-8.

5.4.2 Classical approach

In this section we describe the way we estimate the parameters of the multivariate ARCH and GARCH models by using maximum likelihood methods. For the bivariate ARCH model (5.1), we estimate the model parameters by using the function *mgarch()* of the Splus program. The parameter values of this model and their standard deviations are presented in Table 5.2. As we can see, from Table 5.2, there is no difference in the estimation of the parameters of bivariate ARCH model (5.1) using Bayesian and classical techniques.

The parameters of the bivariate ARCH model (5.5) are estimated by using Newton-Raphson method, Fisher scoring algorithm, the method proposed by Mak (1993), and by using the subroutine *dfpmin()* of numerical recipes (Press, Teukolsky, Vetterling and Flannery, 1992). Our general impression is that the subroutine *dfpmin()* is more robust to the choice of initial values, and the routines of Fisher scoring and of Mak (1993) are in general faster than Newton-Raphson if the initial values are good enough. The

Parameter	Bayesian approach		Classical approach
	sc	wsc	
μ_1	$7.0 \cdot 10^{-5}$ $(2 \cdot 10^{-4})$	$5.7 \cdot 10^{-5}$ $(2 \cdot 10^{-4})$	$8.1 \cdot 10^{-5}$ $(2 \cdot 10^{-4})$
μ_2	$1.5 \cdot 10^{-4}$ $(5 \cdot 10^{-5})$	$1.7 \cdot 10^{-4}$ $(5 \cdot 10^{-5})$	$1.7 \cdot 10^{-4}$ $(4 \cdot 10^{-5})$
a_{10}	$1.2 \cdot 10^{-6}$ $(4 \cdot 10^{-7})$	$1.2 \cdot 10^{-6}$ $(4 \cdot 10^{-7})$	$8.5 \cdot 10^{-7}$ $(3 \cdot 10^{-7})$
a_{11}	$7.4 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$	$7.5 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$	$6.2 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$
a_{12}	$2.4 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$	$2.4 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$	$7.2 \cdot 10^{-3}$ $(6 \cdot 10^{-2})$
a_{20}	$1.0 \cdot 10^{-7}$ $(4 \cdot 10^{-8})$	$4.1 \cdot 10^{-8}$ $(2 \cdot 10^{-8})$	$3.8 \cdot 10^{-8}$ $(4 \cdot 10^{-8})$
a_{21}	$2.3 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$	$1.9 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$	$3.8 \cdot 10^{-6}$ $(8 \cdot 10^{-4})$
a_{22}	$2.2 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$2.7 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$2.6 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$
β_{11}	$8.8 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$8.8 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$9.1 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$
β_{12}	$2.8 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$	$2.3 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$	$3.1 \cdot 10^{-3}$ $(5 \cdot 10^{-2})$
β_{21}	$5.3 \cdot 10^{-4}$ $(5 \cdot 10^{-4})$	$4.2 \cdot 10^{-4}$ $(4 \cdot 10^{-4})$	$8.1 \cdot 10^{-6}$ $(2 \cdot 10^{-3})$
β_{22}	$7.8 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$7.9 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$	$8.0 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$

Table 5.5: Estimates and standard deviation (in brackets) for the parameters of bivariate GARCH model (5.7). sc: imposing stationarity conditions, wsc: without using stationarity conditions.

parameters converge to the same values using these methods, and are presented in Table 5.3. Comparing the Bayesian and the classical estimates we notice a difference on the parameters a_{12} and a_{21} of the bivariate ARCH model (5.5). This is clearly explained by noting that the shape of the posterior distribution for these parameters is far from normality; see, Figure 5-4.

The parameters of bivariate ARCH model (5.6) with constant conditional correlation are estimated by using the subroutine *dfpmin()* and are presented in Table 5.4. Comparing the Bayesian and the classical estimates we notice a difference on a_{12} and a_{21} parameters. This can be explained by noting the shape of the posterior distribution of these parameters (Figure 5-5).

In Tables 5.5 and 5.7 we present the estimated parameter values of bivariate GARCH model (5.7) and bivariate GARCH model (5.4), respectively, obtained by using the subroutine *dfpmin()* of numerical recipes. The maximum likelihood estimates and the corresponding standard deviations for some parameters of these models are different from the estimates taken from Bayesian techniques, and this can be again explained by looking

Parameter	Bayesian approach		Classical approach
	sc	wsc	
μ_1	$3.4 \cdot 10^{-5}$ $(2 \cdot 10^{-4})$	$1.7 \cdot 10^{-5}$ $(2 \cdot 10^{-4})$	$2.2 \cdot 10^{-5}$ $(2 \cdot 10^{-4})$
μ_2	$1.5 \cdot 10^{-4}$ $(4 \cdot 10^{-5})$	$1.8 \cdot 10^{-4}$ $(4 \cdot 10^{-5})$	$1.8 \cdot 10^{-4}$ $(4 \cdot 10^{-5})$
a_{01}	$1.2 \cdot 10^{-6}$ $(4 \cdot 10^{-7})$	$1.2 \cdot 10^{-6}$ $(4 \cdot 10^{-7})$	$9.8 \cdot 10^{-7}$ $(3 \cdot 10^{-7})$
a_{11}	$7.8 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$	$8.1 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$	$7.1 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$
β_{11}	$8.8 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$	$8.8 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$	$9.0 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$
a_{02}	$9.5 \cdot 10^{-8}$ $(3 \cdot 10^{-8})$	$3.6 \cdot 10^{-8}$ $(2 \cdot 10^{-8})$	$2.9 \cdot 10^{-8}$ $(1 \cdot 10^{-8})$
a_{21}	$2.2 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$	$2.9 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$2.8 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$
β_{21}	$7.8 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$	$7.8 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$	$7.9 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$
ρ_{12}	$-3.8 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$-4.1 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$-4.1 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$

Table 5.6: Estimates and standard deviation (in brackets) for the parameters of bivariate GARCH model with constant conditional correlation (5.3). sc: imposing stationarity conditions, wsc: without using stationarity conditions.

the histograms of the posterior sample of the model parameters (Figures 5-6, 5-8). Due to the fact that the posterior distribution for some parameters is far from normality, the estimates of the parameters and the corresponding standard deviations taken from classical approach should be interpreted carefully, but we present them for completeness.

For the bivariate GARCH model (5.3) with constant conditional correlations, we estimate the model parameters by using the subroutine *dfpmin()* of numerical recipes. From Table 5.6 we see that the classical and Bayesian parameter estimates are almost similar.

As a final implementation remark, we note here that classical estimates are very hard to be evaluated under stationarity conditions. Therefore, all our maximum likelihood estimates were produced by ignoring parameter constraints. As a result of this, bivariate GARCH model (5.3) with constant conditional correlations and bivariate GARCH models (5.7) and (5.4) produced estimates which do not satisfy the stationarity conditions and are naturally compared with the corresponding Bayesian estimates.

Parameter	Bayesian approach		Classical approach
	sc	wsc	
μ_1	$3.0 \cdot 10^{-5}$ $(2 \cdot 10^{-4})$	$7.0 \cdot 10^{-6}$ $(2 \cdot 10^{-4})$	$1.9 \cdot 10^{-5}$ $(2 \cdot 10^{-4})$
μ_2	$1.5 \cdot 10^{-4}$ $(5 \cdot 10^{-5})$	$1.9 \cdot 10^{-4}$ $(4 \cdot 10^{-5})$	$1.8 \cdot 10^{-4}$ $(4 \cdot 10^{-5})$
a_{10}	$1.2 \cdot 10^{-6}$ $(4 \cdot 10^{-7})$	$1.2 \cdot 10^{-6}$ $(4 \cdot 10^{-7})$	$9.1 \cdot 10^{-7}$ $(3 \cdot 10^{-7})$
a_{11}	$8.0 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$	$8.2 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$	$6.7 \cdot 10^{-2}$ $(2 \cdot 10^{-2})$
a_{12}	$4.4 \cdot 10^{-2}$ $(3 \cdot 10^{-2})$	$5.0 \cdot 10^{-2}$ $(4 \cdot 10^{-2})$	$5.2 \cdot 10^{-2}$ $(5 \cdot 10^{-2})$
a_{20}	$8.7 \cdot 10^{-8}$ $(3 \cdot 10^{-8})$	$2.8 \cdot 10^{-8}$ $(2 \cdot 10^{-8})$	$2.7 \cdot 10^{-8}$ $(3 \cdot 10^{-8})$
a_{21}	$1.9 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$	$1.7 \cdot 10^{-4}$ $(2 \cdot 10^{-4})$	$6.1 \cdot 10^{-9}$ $(1 \cdot 10^{-3})$
a_{22}	$2.2 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$	$2.9 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$2.8 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$
β_{11}	$8.7 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$8.7 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$8.9 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$
β_{12}	$3.8 \cdot 10^{-2}$ $(3 \cdot 10^{-2})$	$3.1 \cdot 10^{-2}$ $(3 \cdot 10^{-2})$	$1.2 \cdot 10^{-4}$ $(4 \cdot 10^{-2})$
β_{21}	$4.1 \cdot 10^{-4}$ $(4 \cdot 10^{-4})$	$3.4 \cdot 10^{-4}$ $(3 \cdot 10^{-4})$	$2.1 \cdot 10^{-8}$ $(6 \cdot 10^{-4})$
β_{22}	$7.7 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$	$7.8 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$	$7.9 \cdot 10^{-1}$ $(2 \cdot 10^{-2})$
ρ_{12}	$-3.8 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$-4.2 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$	$-4.1 \cdot 10^{-1}$ $(3 \cdot 10^{-2})$

Table 5.7: Estimates and standard deviation (in brackets) for the parameters of bivariate GARCH model with constant conditional correlation (5.4). sc: imposing stationarity conditions, wsc: without using stationarity conditions.

5.4.3 Model comparison via predictive distributions

We have analyzed six multivariate ARCH and GARCH models using 800 daily exchange rates. In order to compare these models we used Bayesian analysis by obtaining predictive distributions $p\left(y_{T+1}|\mathbf{y}\right)$ under each model we considered. In this implementation we used 835 real data points, and calculate $P = 35$ one-step-ahead predictive distributions $p\left(y_{T+1}|\mathbf{y}\right)$ based on the previous $T = 801, ..., 835$ data points. For each one of these 35 time periods we estimate the predictive densities $p\left(y_{T+1}|\mathbf{y}\right)$ at the real data point y_{T+1}, denoted by $\widehat{p}\left(y_{T+1}|\mathbf{y}\right)$, under each model, and running the MCMC method, described in section 5.3.1. Therefore, the algorithm can be constructed as follows. For each time period (based on the previous $T = 801, ..., 835$ data points) and for each model

- run the MCMC algorithm and obtain a sample of $\boldsymbol{\theta}_s$, $s = 1, ..., B$ from the joint posterior $p\left(\boldsymbol{\theta}|\mathbf{y}\right)$
- obtain a sample of size B of the covariance matrix $\boldsymbol{\Sigma}_{T+1}$, by using the outputted $\boldsymbol{\theta}_s$, $s = 1, ..., B$

- calculate, for each of the B outputted values of $\mathbf{\Sigma}_{T+1}$, the densities $\widehat{p}\,(y_{T+1}|\boldsymbol{\theta}_s, \mathbf{y})$
- estimate the predictive density $\widehat{p}\,(y_{T+1}|\mathbf{y}) = \frac{1}{B}\sum_{s=1}^{B} \widehat{p}\,(y_{T+1}|\boldsymbol{\theta}_s, \mathbf{y})$.

We use the quantity $\prod \widehat{p}\,(y_{T+1}|\mathbf{y})$ for assessing comparative validity of the models analyzed. In particular, we compare the multivariate ARCH and GARCH models under consideration using

$$D = \log\left[\frac{\prod^{(M_i)} \widehat{p}\,(y_{T+1}|\mathbf{y})}{\prod^{(M_j)} \widehat{p}\,(y_{T+1}|\mathbf{y})}\right]$$

or

$$D = \sum\nolimits^{(M_i)} \log\left[\widehat{p}\,(y_{T+1}|\mathbf{y})\right] - \sum\nolimits^{(M_j)} \log\left[\widehat{p}\,(y_{T+1}|\mathbf{y})\right],$$

and we choose the model M_i (M_j) if $D > 0$ ($D < 0$). The quantity

$$\sum_{801}^{835} \log\left[\widehat{p}\,(y_{T+1}|\mathbf{y})\right]$$

is estimated for all the models under consideration and the results are 130.42, 131.14, 130.32, 131.57, 130.85, and 130.81 for the bivariate ARCH models (5.1), (5.5), (5.6), and bivariate GARCH models (5.7), (5.3), and (5.4), respectively. Figure 5-9 (a, b) presents the estimates of the $\log\left[\widehat{p}\,(y_{T+1}|\mathbf{y})\right]$ for the period under consideration for the analyzed ARCH and GARCH models. According to these results the multivariate GARCH model (5.7) seems to be preferable for the one step ahead predictions for the analyzed dataset.

5.5 Discussion

An analysis of a class of multivariate ARCH and GARCH models was performed. The parameters of these models have been estimated by using Bayesian and classical approaches. In the parameters where the posterior distribution is far from normality the parameter values turned out to be different. In particular, although the maximum likelihood optimization algorithms can be used for maximizing the conditional likelihood, we believe that their implementation is not without problems in practice. Predictive distributions

are used to address model comparison. For the data set we analyzed, the multivariate GARCH model (5.7) appears preferable for one-step-ahead prediction as compared to the other analyzed multivariate models.

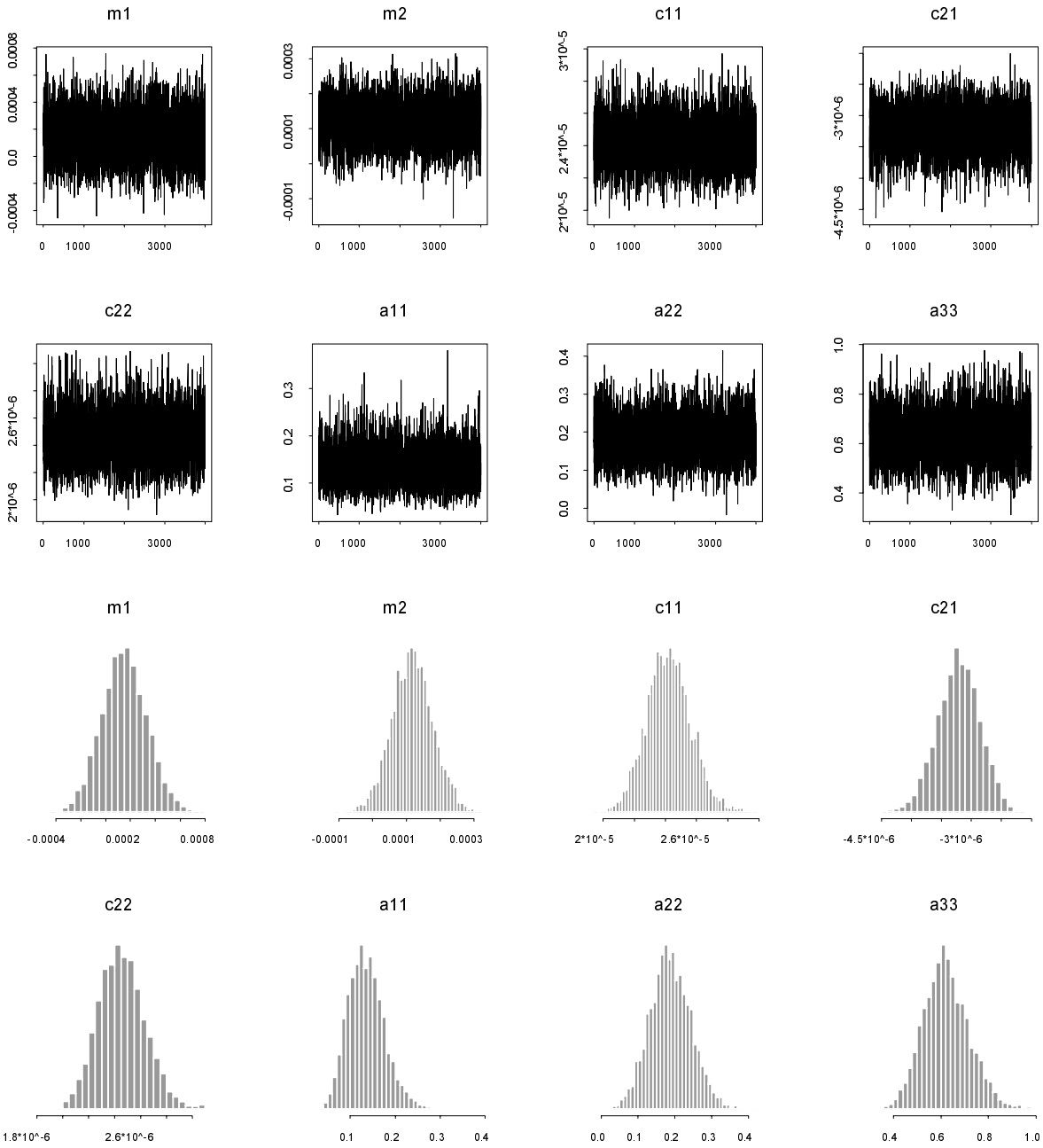

Figure 5-3: Convergence diagrams and histograms of the posterior sample of the parameters of ARCH model (5.1).

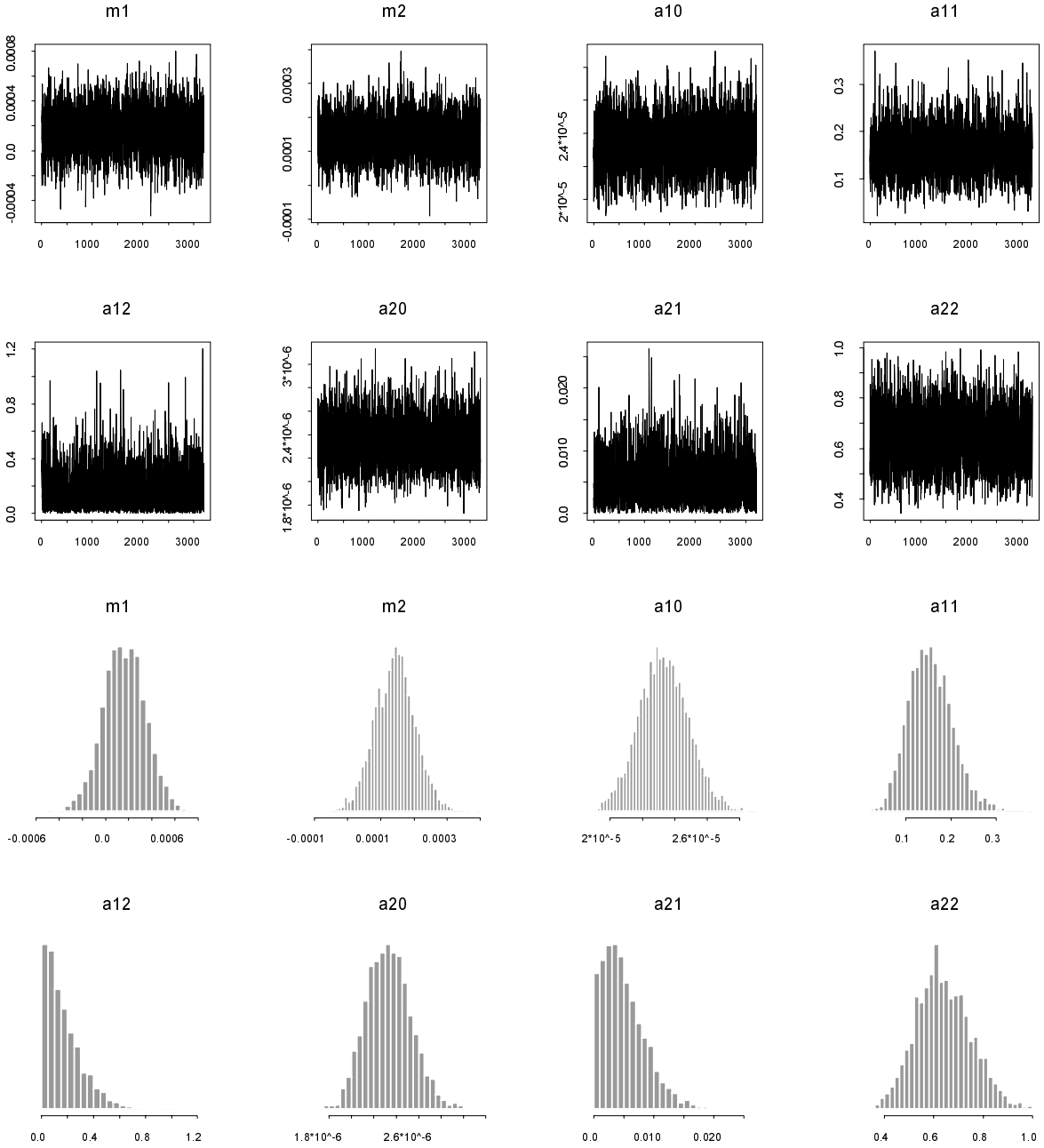

Figure 5-4: Convergence diagrams and histograms of the posterior sample of the parameters of ARCH model (5.5).

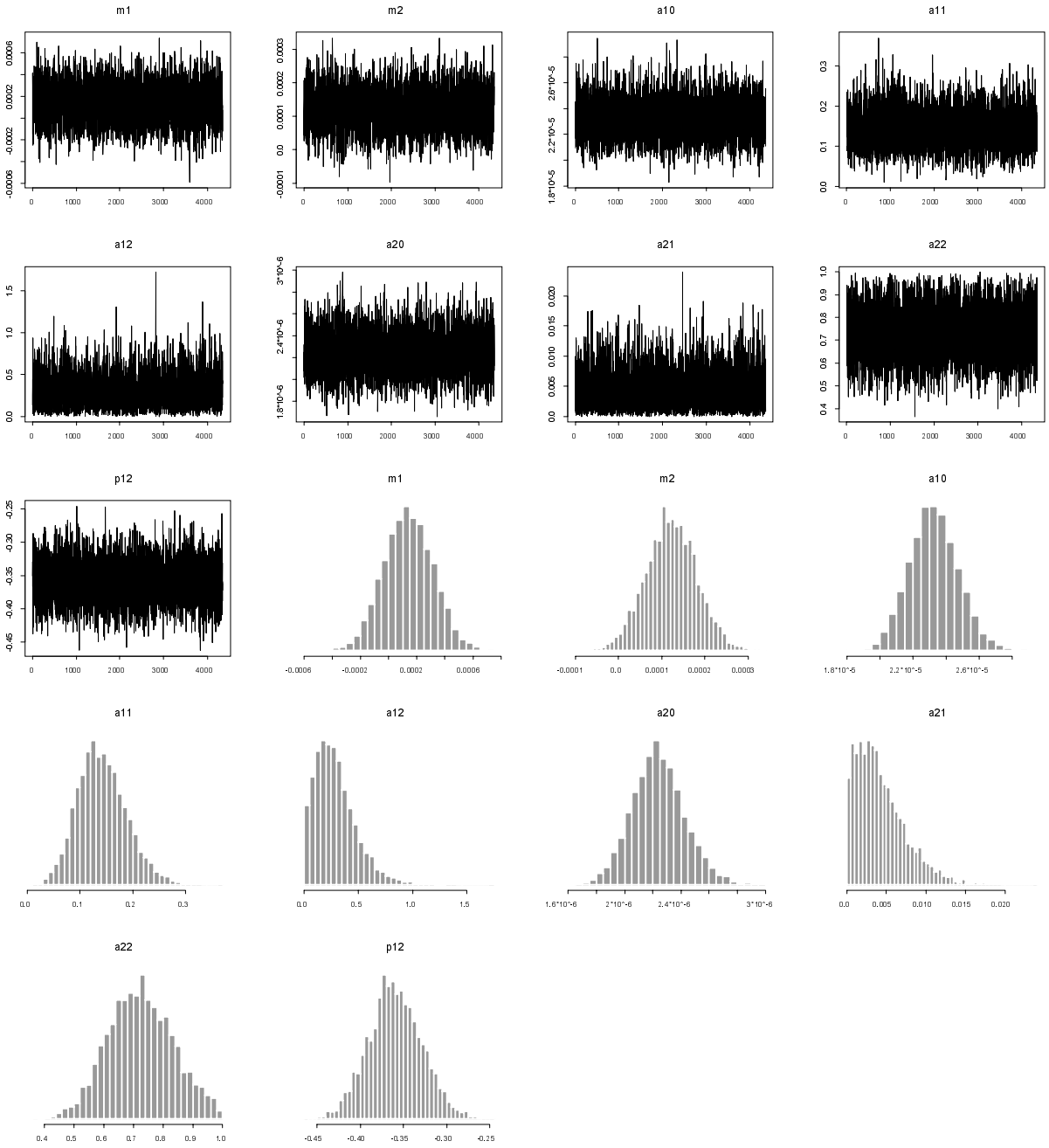

Figure 5-5: Convergence diagrams and histograms of the posterior sample of the parameters of ARCH model (5.6).

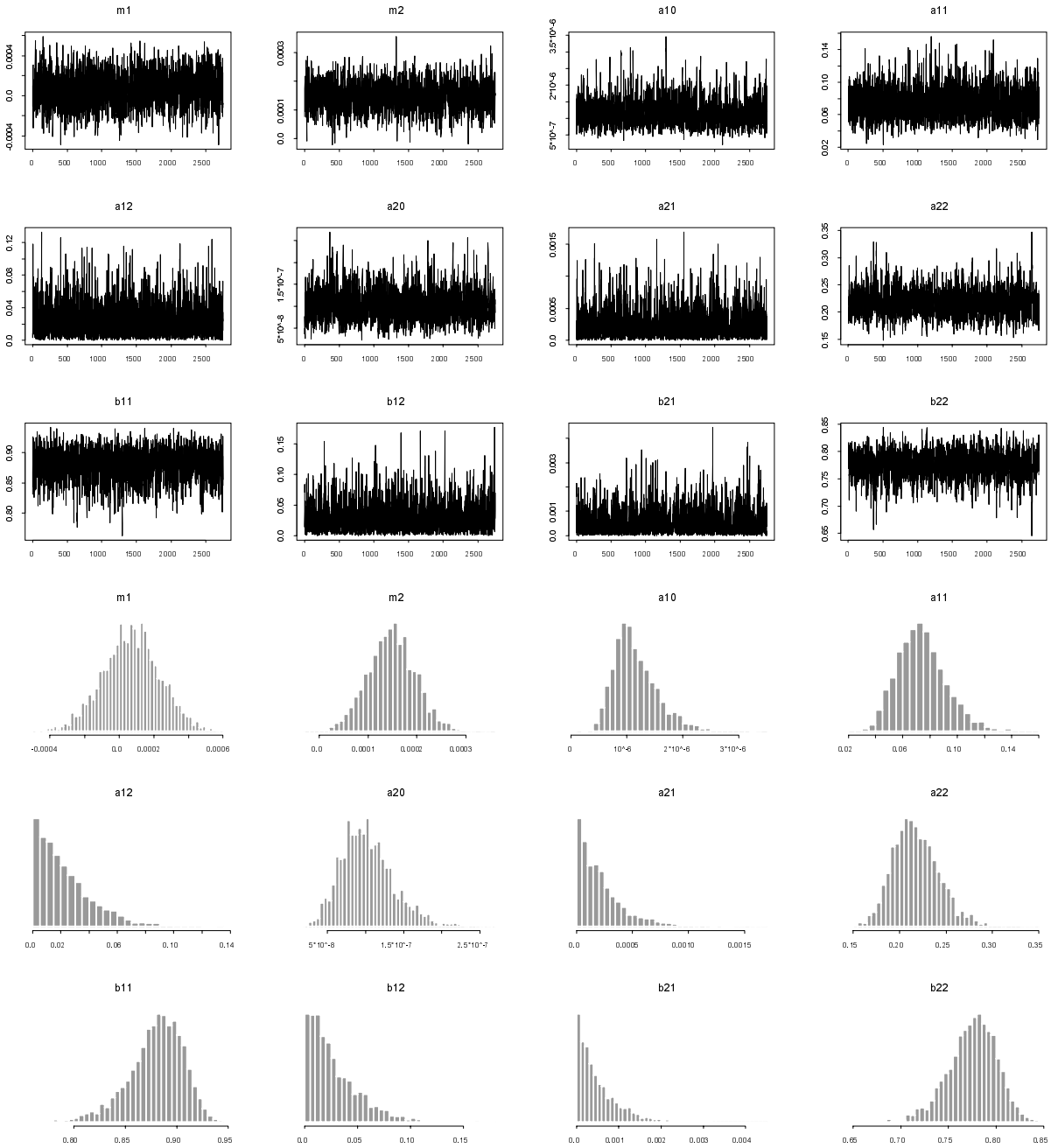

Figure 5-6: Convergence diagrams and histograms of the posterior sample of the parameters of GARCH model (5.7).

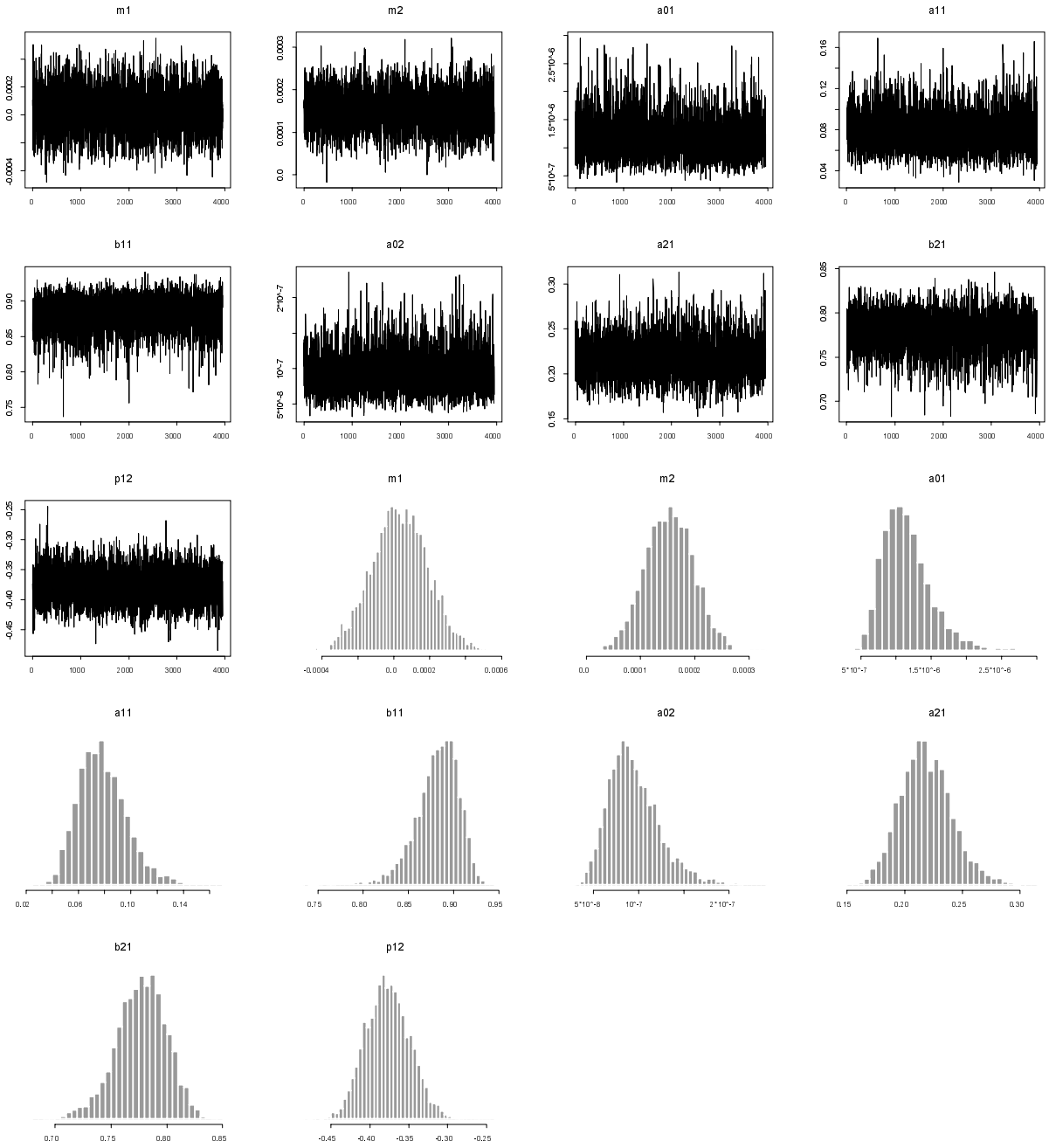

Figure 5-7: Convergence diagrams and histograms of the posterior sample of the parameters of GARCH model (5.3).

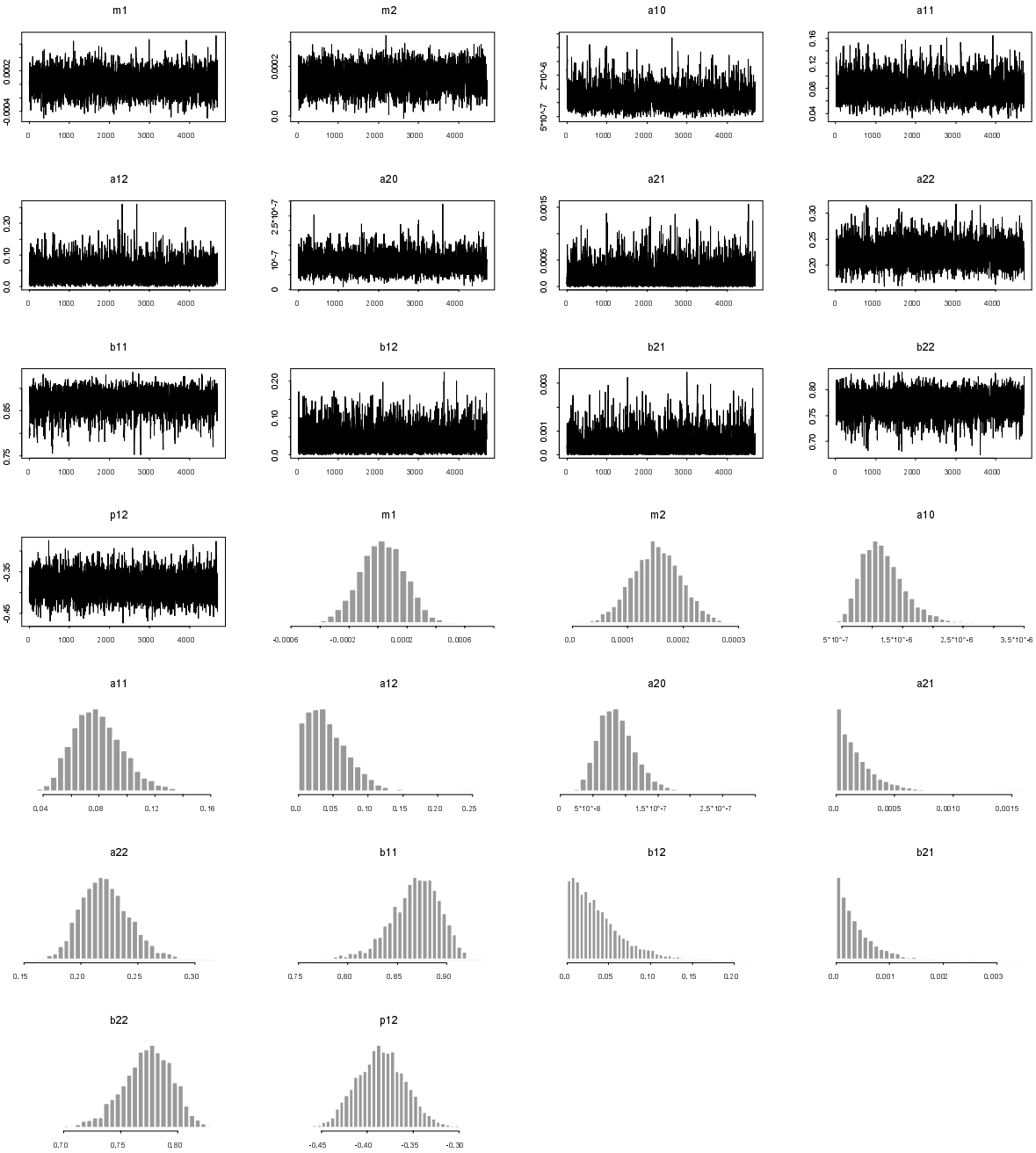

Figure 5-8: Convergence diagrams and histograms of the posterior sample of the parameters of GARCH model (5.4).

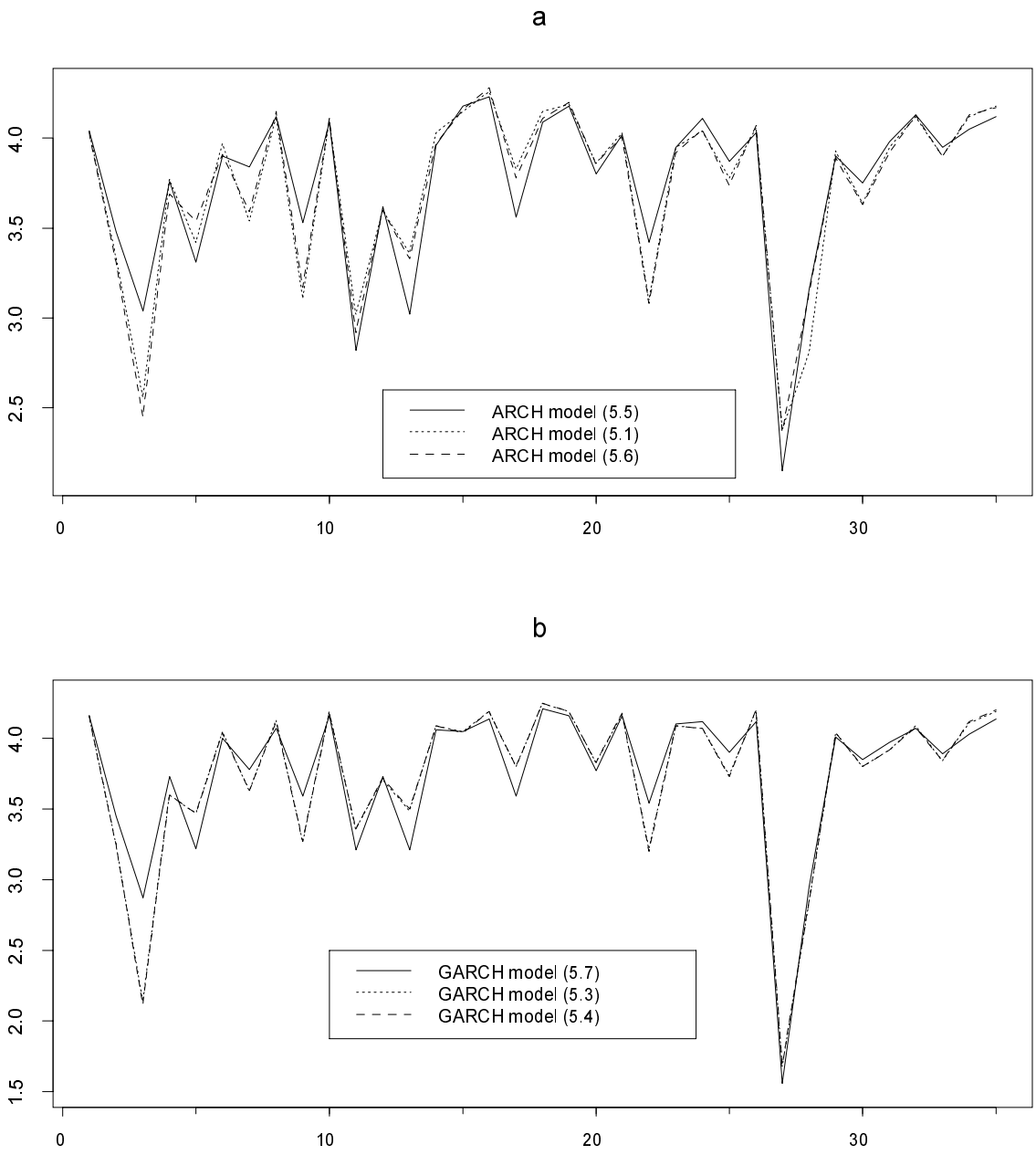

Figure 5-9: Ordinates of the estimated one-step-ahead log predictive densities at the 35 observed data points, a: ARCH models, b: GARCH models.

Chapter 6

A multivariate latent GARCH model

6.1 Introduction

Two major problems related with multivariate ARCH and GARCH models are the large number of parameters to be estimated, and the difficulty of the estimation due to the positive definiteness restrictions of the covariance matrix. Factor ARCH - GARCH models have been introduced to solve these problems providing a parsimonious parameterization and a positive definite covariance matrix. The motivation for the factor models is the commonality in the conditional variance movements. Although, in some cases, financial and economic theories suggest such a characteristic, the restrictions imposed by the factor ARCH - GARCH models on the dynamic behavior of the covariances and the correlations are strong enough, and may depend on the number of factors. For example, in a two factor model, one would expect more dynamics in the covariances and in the correlations than in a one factor model; see, for details, Kroner and Ng (1998).

In this chapter, a multivariate time series model with time-varying conditional variances and covariances is presented and analyzed. In the proposed multivariate latent GARCH model the covariance matrix is always positive definite, the number of para-

meters is relatively small, and the model can be applied very easily to high dimensional time series data. A complete analysis of the proposed model is presented consisting of parameter estimation, model selection and volatility prediction. Classical and Bayesian techniques are used for the estimation of the model parameters. Maximum likelihood estimation is implemented by using Fisher scoring algorithm. According to the Bayesian approach, a Markov chain which has as a stationary distribution the posterior distribution of the model parameters is constructed using a blocking sampling scheme in the Metropolis-Hastings algorithm. Bayesian model selection is addressed using Markov chain Monte Carlo model composition (MC^3) method of Madigan and York (1995) together with the delayed rejection algorithm of Tierney and Mira (1999). The problem of accounting for model uncertainty is considered using Bayesian model averaging. We provide implementation details and illustrations using daily rates of return on eight stocks of the US market.

6.2 The proposed multivariate latent GARCH model

6.2.1 Description and properties of the model

We consider having observed data of the form

$$\mathbf{y}_t,\ t = 1, ..., T,$$

where each $\mathbf{y}_t = (y_{1,t}, ..., y_{N,t})$ is a $N \times 1$ vector. The multivariate latent GARCH model is given by the following equations:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\varepsilon}_t &= W\mathbf{X}_t \\ \mathbf{X}_t|\Phi_{t-1} &\sim N_N\left(\mathbf{0}, \Sigma_t\right) \end{aligned} \tag{6.1}$$

where $\boldsymbol{\mu}$ is a $N \times 1$ vector of constants, $\boldsymbol{\varepsilon}_t$ is a $N \times 1$ innovation vector, W is $N \times N$ parameter matrix, Φ_{t-1} is the information set up to time $t-1$, $\mathbf{X}_t$ is a $N \times 1$ vector of "latent" variables with elements $x_{i,t}$, $i = 1, ..., N$ and Σ_t is $N \times N$ diagonal variance covariance matrix, which is given by $\Sigma_t = diag\left(\sigma_{1,t}^2, ..., \sigma_{N,t}^2\right)$ with

$$\sigma_{i,t}^2 = \alpha_i + bx_{i,t-1}^2 + g\sigma_{i,t-1}^2,\ i = 1, ..., N,\ t = 1, ..., T$$

and $\sigma_{i,t}^2$, $i = 1, ..., N$ is the variance of the $i-th$ "latent" variable at time t, $\alpha_i > 0$, $i = 1, ..., N$, $b \geq 0$, $g \geq 0$. In other words the "latent" variables $x_{i,t}$, $i = 1, ..., N$ are GARCH(1,1) processes. According to the above model, the vector $\boldsymbol{\varepsilon}_t$ is a linear combination of the "latent" variables $x_{i,t}$, $i = 1, ..., N$.

Assuming that the vector $\mathbf{X}_t$ follows a conditional multivariate Normal distribution, $\mathbf{X}_t|\Phi_{t-1} \sim N(\mathbf{0}, \Sigma_t)$, then the vector $\boldsymbol{\varepsilon}_t|\Phi_{t-1} \sim N(\mathbf{0}, H_t)$, where

$$\begin{aligned} H_t &= W\Sigma_t W' = W\Sigma_t^{1/2}\Sigma_t^{1/2}W' \\ &= \left(W\Sigma_t^{1/2}\right)\left(W\Sigma_t^{1/2}\right)' = LL'. \end{aligned} \tag{6.2}$$

Equation (6.2) presents the conditional covariance matrix H_t of the vector $\boldsymbol{\varepsilon}_t$. $\Sigma_t^{1/2}$ is the diagonal matrix with elements $\sigma_{1,t}$, $\sigma_{2,t}$, ..., $\sigma_{N,t}$. It is well known (Dhrymes, 1984, pp. 68-69) that the decomposition of a positive definite matrix into the product of a triangular matrix and its transpose always exists, and this decomposition is unique if the diagonal elements are restricted to be positive. So we can take W triangular with elements $w_{ij} = 0$ for $j > i$ and $w_{ii} > 0$ for $i = 1, ..., N$. In order to reduce the number of parameters in our model, a natural restriction is to assume that $w_{ii} = 1$, for $i = 1, ..., N$. Note that, similar constraints have adopted by Geweke and Zhou (1996), Chib, Nardari and Shephard (1999), and Aguilar and West (2000) among others for the factor model. Under the assumption that the matrix W is triangular with diagonal elements equal to

unity, the conditional covariance matrix H_t can be written as

$$H_t = W\Sigma_t W' = \begin{bmatrix} h_{11,t} & h_{12,t} & h_{13,t} & \cdots & h_{1N,t} \\ h_{21,t} & h_{22,t} & h_{23,t} & \cdots & h_{2N,t} \\ h_{31,t} & h_{32,t} & h_{33,t} & \cdots & h_{3N,t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1,t} & h_{N2,t} & h_{N3,t} & \cdots & h_{NN,t} \end{bmatrix} \tag{6.3}$$

$$= \begin{bmatrix} \sigma^2_{1,t} & w_{21}\sigma^2_{1,t} & w_{31}\sigma^2_{1,t} & \cdots & w_{N1}\sigma^2_{1,t} \\ w_{21}\sigma^2_{1,t} & \sum_{i=1}^{2} w^2_{2i}\sigma^2_{i,t} & \sum_{i=1}^{2} w_{2i}w_{3i}\sigma^2_{i,t} & \cdots & \sum_{i=1}^{2} w_{2i}w_{Ni}\sigma^2_{i,t} \\ w_{31}\sigma^2_{1,t} & \sum_{i=1}^{2} w_{3i}w_{2i}\sigma^2_{i,t} & \sum_{i=1}^{3} w^2_{3i}\sigma^2_{i,t} & \cdots & \sum_{i=1}^{3} w_{3i}w_{Ni}\sigma^2_{i,t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{N1}\sigma^2_{1,t} & \sum_{i=1}^{2} w_{Ni}w_{2i}\sigma^2_{i,t} & \sum_{i=1}^{3} w_{Ni}w_{3i}\sigma^2_{i,t} & \cdots & \sum_{i=1}^{N} w^2_{Ni}\sigma^2_{i,t} \end{bmatrix}.$$

From the construction of the model, the variance covariance matrix H_t is always positive definite if the variances $\sigma^2_{i,t}$, $i = 1, ..., N$ of the "latent" variables are well defined. Note also that the "latent" variables $x_{i,t}$, $i = 1, ..., N$ are not parameters to be estimated but are given by $\mathbf{X}_t = W^{-1}\boldsymbol{\varepsilon}_t$. Therefore, for $N = 1$ the model reduces to the GARCH(1,1) model.

Under the assumption of multivariate Normal distribution for the vector $\mathbf{X}_t$, the likelihood for model (6.1) for a sample of T observations $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T)$ can be written as

$$l(\mathbf{y}|\boldsymbol{\theta}) = (2\pi)^{-\frac{TN}{2}} \prod_{t=1}^{T} |H_t|^{-1/2} \exp\left[-\frac{1}{2}\sum_{t=1}^{T} (\mathbf{y}_t - \boldsymbol{\mu})' H_t^{-1} (\mathbf{y}_t - \boldsymbol{\mu})\right]$$

and the number of parameters to be estimated for a N dimensional problem is $2N + 2 + \frac{N(N-1)}{2}$. That is, the parameter vector is $\boldsymbol{\theta} = (\mu_1, \mu_2, ..., \mu_N, \alpha_1, \alpha_2, ..., \alpha_N, b, g, w_{21}, w_{31}, w_{32}, ..., w_{N1}, ..., w_{N,N-1})'$.

Bollerslev (1986) presents the necessary and sufficient condition for the existence of

the $2m-th$ moment of the GARCH(1,1) model and the conditions for the wide-sense stationarity of the GARCH model (see theorems 1 and 2). We can take advantage from these results for the multivariate latent GARCH model we propose. First note that $H_t = W\Sigma_t W^{'}$ and $vec\left(H_t\right) = vec\left(W\Sigma_t W^{'}\right) = \left(W \otimes W\right) vec\left(\Sigma_t\right)$, where $vec\left(.\right)$ denotes the vector operator that stacks the columns of the matrix. That is, the variances and the covariances in our model are linear combinations of the variances $\sigma_{i,t}^2$, $i = 1, ..., N$ of the "latent" variables. But according to our model the "latent" variables $x_{i,t}$, $i = 1, ..., N$ are independent GARCH(1,1) processes. Therefore, the unconditional variances and covariances of the multivariate latent GARCH model are linear combinations of the unconditional variances of the GARCH(1,1) processes $x_{i,t}$, $i = 1, ..., N$. Define $\mathbf{w}_t = vech\left(H_t\right)$, where $vech\left(.\right)$ denotes the column stacking operator of the lower portion of a symmetric matrix. That is, $\mathbf{w}_t = vech\left(H_t\right) = (h_{11,t}, h_{21,t}, h_{22,t}, h_{31,t}, h_{32,t}, h_{33,t}, ..., h_{N1,t}, h_{N2,t}, h_{N3,t}, ..., h_{NN,t})^{'}$ is a $N\left(N+1\right)/2 \times 1$ vector. After straightforward calculations and based on the results of Bollerslev (1986) the unconditional variances and covariances are given by $E\left(\mathbf{w}_t\right) = \frac{1}{1-b-g}(\alpha_1, w_{21}\alpha_1, \sum_{i=1}^{2} w_{2i}^2\alpha_i, w_{31}\alpha_1, \sum_{i=1}^{2} w_{3i}w_{2i}\alpha_i, \sum_{i=1}^{3} w_{3i}^2\alpha_i, ..., w_{N1}\alpha_1, \sum_{i=1}^{2} w_{Ni}w_{2i}\alpha_i, \sum_{i=1}^{3} w_{Ni}w_{3i}\alpha_i, ..., \sum_{i=1}^{N} w_{Ni}^2\alpha_i)^{'}$.

6.2.2 Some comments on the model

The proposed model can be considered as a latent factor model. Different factor models have been proposed in the literature, and have been analyzed by many researchers either using ARCH - GARCH framework (see, for example, Diebold and Nerlove, 1989, Engle, Ng and Rothschild, 1990, King, Sentana and Wadhwani, 1994, Alexander, 2000 among several others) or using Stochastic Volatility framework (see, for example, Chib, Nardari and Shephard, 1999, Aguilar and West, 2000) for the specification of the variances of the factors.

The structure of our model and especially the matrix W resembles with that of Aguilar and West (2000) and of Chib, Nardari and Shephard (1999). However, unlike the previous

authors, we use a full factor representation and zero idiosyncratic variance terms. Such a representation produces a more dynamic behavior for the variances and especially for the covariances and the correlations than a k factor model ($k < N$). The "latent" variables are not entered in the estimation procedure but are given by $\mathbf{X}_t = W^{-1}\boldsymbol{\varepsilon}_t$. For the specification of the variances of the "latent" variables a GARCH(1,1) model is used.

The structure of the conditional covariance matrix H_t, in equation (6.3), implies that the order of the univariate time series in the $\mathbf{y}_t$ vector affects the conditional variances and covariances. The ordering has an impact in model fitting and assessment; see, for example, Aguilar and West (2000). We solve this important practical issue using the Markov chain Monte Carlo model composition (MC^3) method that generates a process that moves through model space and the Delayed rejection algorithm (DRA). Moreover, these methods provide an idealized way to extract posterior model probabilities and in addition to construct predictive densities that take into account model uncertainty.

6.3 Inference for a given model

In this section we consider classical and Bayesian techniques for the estimation of the parameters of the multivariate latent GARCH model.

6.3.1 Classical approach

Maximum likelihood estimates, for heteroscedastic models, are usually taken by using numerical optimization algorithms such as scoring algorithm, the method proposed by Mak (1993) and developed further by Mak, Wong and Li (1997) and by Berndt, Hall, Hall and Hausman (1974) algorithm. We compute the maximum likelihood estimates by using the Fisher scoring algorithm. The $k-th$ iteration of the algorithm takes the form

$$\widehat{\boldsymbol{\theta}}^{k} = \widehat{\boldsymbol{\theta}}^{k-1} + \left\{ -E\left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \right\}^{-1} \frac{\partial L_T}{\partial \boldsymbol{\theta}} \tag{6.4}$$

where $\widehat{\boldsymbol{\theta}}^{k-1}$ is the estimate of the parameter vector obtained after $k-1$ iterations, L_T is the log-likelihood function, $-E\left[\frac{\partial^2 L_T}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]$ is the expected Information matrix I computed at $\widehat{\boldsymbol{\theta}}^{k-1}$, and $\frac{\partial L_T}{\partial\boldsymbol{\theta}}$ is the gradient computed at $\widehat{\boldsymbol{\theta}}^{k-1}$. One of the great advantages of our model is, we believe, the fact that the algorithm (6.4) can be computed analytically, that is the gradient and the estimated information matrix are available in closed forms. This is not surprising since by construction, our model consists of a linear combination of univariate GARCH models in which such a property exists. The motivation for using the fisher scoring algorithm in our multivariate latent GARCH model comes from the experimental results of researchers to GARCH models and to factor GARCH models. For example, Fiorentini, Calzolari, and Panattoni (1996) computed the analytic first and second derivatives of the log-likelihood for the GARCH(p,q) model, and constructed a mixed-gradient algorithm in order to accelerate the convergence of the parameters of the GARCH model. According to their results, the superiority of gradient algorithms, which use the estimated information matrix, is clear in early iterations. Watson and Engle (1983) used the method of scoring and the EM algorithm for the estimation of dynamic factor, mimic and varying coefficient regression models. They suggest, for practical methods, a mixed EM and scoring algorithm, and the use of scoring algorithm for inference. Similar are the results of the experiments of Demos and Sentana (1998), who present an EM algorithm for conditionally heteroscedastic factor models and propose a quasi-newton algorithm at last iterations.

For the multivariate latent GARCH model (6.1) the log likelihood function is given by

$$
\begin{aligned}
L_T\left(\mathbf{y}|\boldsymbol{\theta}\right) &= -\frac{TN}{2}\ln\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{T}\ln\left|H_t\right| - \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{y}_t-\boldsymbol{\mu}\right)' H_t^{-1}\left(\mathbf{y}_t-\boldsymbol{\mu}\right) \\
&= -\frac{TN}{2}\ln\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{T}\ln\left|W\Sigma_t W'\right| - \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{y}_t-\boldsymbol{\mu}\right)'\left(W\Sigma_t W'\right)^{-1}\left(\mathbf{y}_t-\boldsymbol{\mu}\right) \\
&= -\frac{TN}{2}\ln\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{T}\ln\left|\Sigma_t\right| - \frac{1}{2}\sum_{t=1}^{T}\mathbf{X}_t'\Sigma_t^{-1}\mathbf{X}_t, \text{ where } \mathbf{X}_t = W^{-1}\left(\mathbf{y}_t-\boldsymbol{\mu}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{TN}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln\left[\prod_{i=1}^{N}\sigma_{i,t}^2\right] - \frac{1}{2}\sum_{t=1}^{T}\mathbf{X}_l'\Sigma_t^{-1}\mathbf{X}_t \\
&= -\frac{TN}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\left[\ln\left(\sigma_{i,t}^2\right)\right]\right] - \frac{1}{2}\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\left[\frac{x_{i,t}^2}{\sigma_{i,t}^2}\right]\right],
\end{aligned}
$$

where $\boldsymbol{\theta} = (\mu_1,\ \mu_2,\ ...,\ \mu_N,\ \alpha_1,\ \alpha_2,\ ...,\ \alpha_N,\ b,\ g,\ w_{21},\ w_{31},\ w_{32},\ ...,\ w_{N1},\ ...,\ w_{N,N-1})'$, and $\alpha_i > 0,\ i = 1, ..., N,\ b \geq 0,\ g \geq 0$. In order to avoid these positivity restrictions, we transform the positive parameters using the logarithmic transformation, that is, $\alpha_i^* = \ln(\alpha_i)$, $b_i^* = \ln(b_i)$, and $g_i^* = \ln(g_i)$. We also divide the parameter vector into three blocks. The first block contains the parameters of the mean equation, that is $\boldsymbol{\theta}_1 = (\mu_1,\ \mu_2,\ ...,\ \mu_N)'$, the second block contains the transformed parameters of the variance equation, that is, $\boldsymbol{\theta}_2 = (\alpha_1^*,\ \alpha_2^*,\ ...,\ \alpha_N^*,\ b^*,\ g^*)'$, and the third block contains the parameters in matrix W, that is, $\boldsymbol{\theta}_3 = (w_{21},\ w_{31},\ w_{32},\ ...,\ w_{N1},\ ...,\ w_{N,N-1})'$. The information matrix is block diagonal (see Bollerslev, 1986) and the three diagonal blocks are estimated by $-E\left[\frac{\partial^2 L_T}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1'}\right]$, $-E\left[\frac{\partial^2 L_T}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2'}\right]$ and $-E\left[\frac{\partial^2 L_T}{\partial\boldsymbol{\theta}_3\partial\boldsymbol{\theta}_3'}\right]$. After the transformation of the positive parameters, the variances $\sigma_{i,t}^2$ of the of "latent" variables $x_{i,t}$, $i = 1, ..., N$, are given by

$$
\sigma_{i,t}^2 = e^{\alpha_i^*} + e^{b^*}x_{i,t-1}^2 + e^{g^*}\sigma_{i,t-1}^2, i = 1, ..., N, t = 1, ..., T.
$$

Some assumptions are also required for the initial values of the variances $\sigma_{i,t}^2$ and the squared "latent" variables $x_{i,t}^2$, as the variance equation of GARCH(1,1) model is dynamic. For $t \leq 0$, $\sigma_{i,t}^2$ are zero, while the $x_{i,t}^2$, for $t \leq 0$, are calculated by using a sufficient number of observations from the sample.

Differentiating with respect to the mean parameters $\boldsymbol{\theta}_1 = (\mu_1,\ \mu_2,\ ...,\ \mu_N)'$ yields

$$
\frac{\partial L_T}{\partial\boldsymbol{\theta}_1} = \sum_{t=1}^{T}\left\{\sum_{i=1}^{N}\left[\frac{1}{2\sigma_{i,t}^2}\frac{\partial\sigma_{i,t}^2}{\partial\boldsymbol{\theta}_1}\left(\frac{x_{i,t}^2}{\sigma_{i,t}^2} - 1\right) - \left(\frac{x_{i,t}}{\sigma_{i,t}^2}\frac{\partial x_{i,t}}{\partial\boldsymbol{\theta}_1}\right)\right]\right\}
$$

and the Information matrix for the first block is given by

$$\begin{aligned} I_1 &= -E\left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^{'}}\right] \\ &= \sum_{t=1}^{T}\left\{\sum_{i=1}^{N}\left[\frac{1}{2\left(\sigma_{i,t}^2\right)^2}\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_1}\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_1^{'}} + \frac{1}{\sigma_{i,t}^2}\frac{\partial x_{i,t}}{\partial \boldsymbol{\theta}_1}\frac{\partial x_{i,t}}{\partial \boldsymbol{\theta}_1^{'}}\right]\right\} \end{aligned}$$

where

$$\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_1} = 2e^{b^*}x_{i,t-1}\frac{\partial x_{i,t-1}}{\partial \boldsymbol{\theta}_1} + e^{g^*}\frac{\partial \sigma_{i,t-1}^2}{\partial \boldsymbol{\theta}_1},\ i = 1, ..., N$$

and the derivatives of $x_{i,t}$, $i = 1, ..., N$, with respect to the mean parameters $\boldsymbol{\theta}_1$ are given by the rows of $-W^{-1}$ matrix. That is, the $\frac{\partial x_{1,t}}{\partial \boldsymbol{\theta}_1}$ is given by the first row of $-W^{-1}$ matrix, the $\frac{\partial x_{2,t}}{\partial \boldsymbol{\theta}_1}$ is given by the second row of $-W^{-1}$ matrix, and so on.

Differentiating with respect to the variance parameters $\boldsymbol{\theta}_2 = (\alpha_1^*,\ \alpha_2^*,\ ...,\ \alpha_N^*,\ b^*,\ g^*)^{'}$ yields

$$\frac{\partial L_T}{\partial \boldsymbol{\theta}_2} = \sum_{t=1}^{T}\left\{\sum_{i=1}^{N}\left[\frac{1}{2\sigma_{i,t}^2}\left(\frac{x_{i,t}^2}{\sigma_{i,t}^2} - 1\right)\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_2}\right]\right\}$$

while the Information matrix for the second block is given by

$$I_2 = -E\left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^{'}}\right] = \sum_{t=1}^{T}\left\{\sum_{i=1}^{N}\left[\frac{1}{2\left(\sigma_{i,t}^2\right)^2}\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_2}\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_2^{'}}\right]\right\}$$

where

$$\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_2} = \mathbf{c}_{i,t} + e^{g^*}\frac{\partial \sigma_{i,t-1}^2}{\partial \boldsymbol{\theta}_2},\ i = 1, ..., N$$

and the vectors $\mathbf{c}_{i,t}$, $i = 1, ..., N$, can be calculated very easily. For example, $\mathbf{c}_{1,t} = (e^{\alpha_1^*},\ 0,\ ...,\ 0,\ e^{b^*}x_{1,t-1}^2,\ e^{g^*}\sigma_{1,t-1}^2)^{'}$, $\mathbf{c}_{2,t} = (0,\ e^{\alpha_2^*},\ 0,\ ...,\ 0,\ e^{b^*}x_{2,t-1}^2,\ e^{g^*}\sigma_{2,t-1}^2)^{'}$, ..., $\mathbf{c}_{N,t} = (0,\ ...,\ 0,\ e^{\alpha_N^*},\ e^{b^*}x_{N,t-1}^2,\ e^{g^*}\sigma_{N,t-1}^2)^{'}$.

Differentiating with respect to the parameters in matrix W, that is, with respect to $\boldsymbol{\theta}_3 = (w_{21},\ w_{31},\ w_{32},\ ...,\ w_{N1},\ ...,\ w_{NN-1})^{'}$ yields

$$\frac{\partial L_T}{\partial \boldsymbol{\theta}_3} = \sum_{t=1}^{T}\left\{\sum_{i=1}^{N}\left[\frac{1}{2\sigma_{i,t}^2}\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_3}\left(\frac{x_{i,t}^2}{\sigma_{i,t}^2} - 1\right) - \left(\frac{x_{i,t}}{\sigma_{i,t}^2}\frac{\partial x_{i,t}}{\partial \boldsymbol{\theta}_3}\right)\right]\right\}$$

and the Information matrix for the third block is given by

$$\begin{aligned} I_3 &= -E\left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta}_3 \partial \boldsymbol{\theta}_3^{'}}\right] \\ &= \sum_{t=1}^{T}\left\{\sum_{i=1}^{N}\left[\frac{1}{2\left(\sigma_{i,t}^2\right)^2}\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_3}\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_3^{'}} + \frac{1}{\sigma_{i,t}^2}\frac{\partial x_{i,t}}{\partial \boldsymbol{\theta}_3}\frac{\partial x_{i,t}}{\partial \boldsymbol{\theta}_3^{'}}\right]\right\} \end{aligned}$$

where

$$\frac{\partial \sigma_{i,t}^2}{\partial \boldsymbol{\theta}_3} = 2e^{b^*} x_{i,t-1}\frac{\partial x_{i,t-1}}{\partial \boldsymbol{\theta}_3} + e^{g^*}\frac{\partial \sigma_{i,t-1}^2}{\partial \boldsymbol{\theta}_3},\ i = 1, ..., N.$$

The "latent" variables $x_{i,t}$, $i = 1, ..., N$, are given by $\mathbf{X}_t = W^{-1}\boldsymbol{\varepsilon}_t$. Because $\mathbf{X}_t = W^{-1}\boldsymbol{\varepsilon}_t$,

$$\frac{\partial \mathbf{X}_t}{\partial w_{ij}} = \left[-W^{-1}\frac{\partial W}{\partial w_{ij}}W^{-1}\right]\boldsymbol{\varepsilon}_t$$

and therefore, the derivatives of $x_{i,t}$, $i = 1, ..., N$, with respect to w_{ij} are given by the $i-th$ element of the vector $\left[-W^{-1}\frac{\partial W}{\partial w_{ij}}W^{-1}\right]\boldsymbol{\varepsilon}_t$. For example, the derivative of $x_{1,t}$ with respect to parameter w_{N1} is given by the first element of the vector $\left[-W^{-1}\frac{\partial W}{\partial w_{N1}}W^{-1}\right]\boldsymbol{\varepsilon}_t$, and so on.

Having calculated the three blocks $-E\left[\frac{\partial^2 L_T}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^{'}}\right]$, $i = 1, 2, 3$ of the block diagonal information matrix I, and the gradients $\frac{\partial L_T}{\partial \boldsymbol{\theta}_i}$, $i = 1, 2, 3$, one can find the maximum likelihood estimates by applying the fisher scoring algorithm of equation (6.4). From our experience to different datasets, the estimates taken from the fisher scoring algorithm (6.4) are robust to different initial values for the model parameters, and the inverse of the information matrix provides estimates of the covariance matrix of the parameters. The above strategy enables us to investigate more deeply the behavior of covariance matrix estimators. For example, we can very easily evaluate the covariance estimate proposed for dynamic and conditional heteroscedastic models by Bollerslev and Wooldridge (1992). This covariance estimate (BW) can be given by

$$BW = \left(I^{-1}\right)(OP)\left(I^{-1}\right) \tag{6.5}$$

where I^{-1} is the inverse of the information matrix, and OP is the matrix of outer products of the first derivatives of the log-likelihood, which can be constructed by

$$OP = \sum_{t=1}^{T} \left[\frac{\partial L_t}{\partial \boldsymbol{\theta}}\right] \left[\frac{\partial L_t}{\partial \boldsymbol{\theta}'}\right]$$

where $\frac{\partial L_t}{\partial \boldsymbol{\theta}}$ is the derivative of the log-likelihood with respect to the parameter vector $\boldsymbol{\theta}$ at time t. Under misspesification of the conditional density of the process (non Gausian distribution) the estimate of the variance covariance matrix given by Bollerslev and Wooldridge (1992) is more robust and superior with respect to other covariance estimators; see, for example, Fiorentini, Calzolari, and Panattoni (1996).

6.3.2 Bayesian approach

This section presents the Bayesian approach for estimating the parameters of the proposed multivariate latent GARCH model by using Markov chain Monte Carlo (MCMC) methods. A detailed description of MCMC methods is presented in section 1.2.4. We adopt the Metropolis Hastings algorithm to obtain a sample from the posterior distribution of interest. In our proposed model, the convergence of the MCMC algorithm is accelerated by reparametrizing the positive parameters to "near normality" and by using a blocking sampling scheme as in Vrontos, Dellaportas and Politis (2000). We use (as in the maximum likelihood approach) three blocks consisting of the mean parameters, the variance parameters and the parameters in matrix W. In this blocking sampling scheme, we use the results from the classical approach. That is, we start from the maximum likelihood estimates and update the parameters in each block from time t to time $t+1$ by using three multivariate Normal proposal densities $N\left(\boldsymbol{\theta}_i^t,\, c\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}_i}\right)$, $i = 1, 2, 3$ with $\boldsymbol{\theta}_i^t$ denoting the vector of parameters in block i with values at time t, c is a constant to tune the acceptance rate, and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}_i}$ is the variance covariance estimate of the parameters in block i taken from fisher scoring algorithm or taken from the method proposed for dynamic and conditional heteroscedastic models by Bollerslev and Wooldridge (1992). We

found that for the first two blocks the algorithm works very well with either proposals, whereas for the third block the latter proposal density performed better. This occurred because Fisher scoring produced many zero values in the off-diagonal elements of the covariance matrix of $\boldsymbol{\theta}_3$.

6.4 Inference under model uncertainty

6.4.1 Bayesian model comparison and model averaging

As described in section 6.2.2, the order of the univariate time series in the $\mathbf{y}_t$ vector has an impact in model fitting. We consider the problem of finding the "best" ordering of the individual time series under the proposed model. That is, the ordering becomes a modelling decision to be made on the basis of model fit. Given the N observed univariate time series in the $\mathbf{y}_t$ vector, the number of all possible models (all possible different orderings) is $N!$. Let $M = \{m_1, ..., m_K\}$ to be the set of all models, so $K = N!$. A typical approach is to carry out a model selection exercise leading to a single "best" model ("best" ordering) and then make inferences as if the selected model was the true model. However, this ignores the uncertainty involved in model selection.

A Bayesian solution to this problem involves the calculation of the posterior probabilities of all the competing models. Inference about the model selection problem may be done using the Bayes Factor (BF) of model m_i against model m_j given by

$$BF = \frac{p\left(\mathbf{y}|m_i\right)}{p\left(\mathbf{y}|m_j\right)}, \tag{6.6}$$

where $p\left(\mathbf{y}|m_i\right)$ is the marginal likelihood of model m_i. We calculate the Bayes Factor by applying to both the numerator and denominator of (6.6) a variant of Laplace approximation:

$$p\left(\widehat{\mathbf{y}|m_i}\right) = (2\pi)^{d_{m_i}/2} \left|\widehat{\Sigma}\right|^{1/2} p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_{m_i}, m_i\right) p\left(\widehat{\boldsymbol{\theta}}_{m_i}|m_i\right), \tag{6.7}$$

where $\widehat{\Sigma}$ is the inverse of the negative Hessian matrix of the log-likelihood evaluated at the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{m_i}$, $p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_{m_i}, m_i\right)$ and $p\left(\widehat{\boldsymbol{\theta}}_{m_i}|m_i\right)$ are the likelihood and the prior, respectively, evaluated at $\widehat{\boldsymbol{\theta}}_{m_i}$. In our analysis, we use the inverse of the expected information matrix in place of $\widehat{\Sigma}$ in equation (6.7).

We account for model uncertainty in our predictive inference for a quantity of interest Δ by using its posterior distribution given data $\mathbf{y}$

$$p\left(\Delta|\mathbf{y}\right) = \sum_{i=1}^{K} p\left(\Delta|m_i, \mathbf{y}\right) p\left(m_i|\mathbf{y}\right), \tag{6.8}$$

where

$$p\left(\Delta|m_i, \mathbf{y}\right) = \int\limits_{\boldsymbol{\Theta}_{m_i}} p\left(\Delta|\boldsymbol{\theta}_{m_i}, m_i, \mathbf{y}\right) p\left(\boldsymbol{\theta}_{m_i}|m_i, \mathbf{y}\right) d\boldsymbol{\theta}_{m_i}$$

or using a maximum likelihood approximation

$$p\left(\Delta|m_i, \mathbf{y}\right) \simeq p\left(\Delta|m_i, \mathbf{y}, \widehat{\boldsymbol{\theta}}_{m_i}\right), \tag{6.9}$$

where $\widehat{\boldsymbol{\theta}}_{m_i}$ is the maximum likelihood estimator of the parameter vector $\boldsymbol{\theta}_{m_i}$ of model m_i. For details about these issues see section 1.3.

6.4.2 MCMC model search methods

In the application section we use two Markov chain Monte Carlo methods that provide posterior model probabilities and therefore can account for model uncertainty using Bayesian model averaging.

Markov chain Monte Carlo model Composition ($\mathbf{MC^3}$)

Markov chain Monte Carlo model Composition (MC^3) was presented in section 1.3.4. MC^3 generates a stochastic process that moves through model space. To construct the Markov chain we define a neighbourhood $nbd\left(m\right)$ for each model m. We also define a

transition matrix q by setting $q\left(m \to m^{'}\right) = 0$ for all $m^{'} \notin nbd\left(m\right)$ and $q\left(m \to m^{'}\right)$ constant for all $m^{'} \in nbd\left(m\right)$. If the current state of the chain is model m then $m^{'}$ is drawn from $q\left(m \to m^{'}\right)$ and accepted with probability

$$\min\left\{1, \frac{\left|nbd\left(m\right)\right| p\left(m^{'}|\mathbf{y}\right)}{\left|nbd\left(m^{'}\right)\right| p\left(m|\mathbf{y}\right)}\right\}, \tag{6.10}$$

where $\left|nbd\left(m\right)\right|$ is the number of models that belong in the neighbourhood of model m. Otherwise, the chain stays in state m. Note that, if $\left|nbd\left(m\right)\right| = \left|nbd\left(m^{'}\right)\right|$ and all models are equally likely a priori then the probability of acceptance is given by

$$\min\left\{1, \frac{p\left(\mathbf{y}|m^{'}\right)}{p\left(\mathbf{y}|m\right)}\right\}. \tag{6.11}$$

In the application of the MC^3 algorithm in section 6.5 we take $\left|nbd\left(m\right)\right| = \left|nbd\left(m^{'}\right)\right|$ and assume that all models are equally likely a priori, and therefore the probability of acceptance is given by (6.11).

Different models come from the particular ordering of the univariate time series in the $\mathbf{y}_t$ vector. After extensive searching for proposal densities or, equivalently, neighbour definitions, that provide an MCMC algorithm with good mixing in a series of problems, we found that multimodalities in model space is a very frequent phenomenon, so we suggest the following neighbourhood definitions. Assume that we are looking for the neighbours of model $m = \{m_1, ..., m_i, ..., m_j, ..., m_k\}$. Let us define $nbd_1^3\left(m\right)$ all models of the form $\{m_1, ..., m_j, ..., m_i, ..., m_k\}$ where m_i and m_j are at the most 3 positions apart. To ensure reversibility, using a cyclic fashion we can also swap m_1 with m_k or m_{k-1} or m_{k-2}, and so on for $i \leq 3$ and $j \geq k-3$. Moreover, define $nbd_2^3\left(m\right)$ as all models of the form

$$\{m_1, ..., m_{i-2}, m_i, m_{i-1}, m_{i+1,}..., m_k\} \cup$$
$$\cup \{m_1, ..., m_{i-1}, m_{i+1,}m_i, m_{i+2}, ..., m_k\} \cup$$

$$\cup \{m_1, ..., m_{i-3}, m_i, m_{i-2}, m_{i-1}, m_{i+1,}..., m_k\} \cup$$
$$\cup \{m_1, ..., m_{i-1}, m_{i+1}, m_{i+2}, m_i, m_{i+3}, ..., m_k\} \cup$$
$$\cup \{m_1, ..., m_{i-4}, m_i, m_{i-3}, m_{i-2}, m_{i-1}, m_{i+1,}..., m_k\} \cup$$
$$\cup \{m_1, ..., m_{i-1}, m_{i+1}, m_{i+2}, m_{i+3}, m_i, m_{i+4}, ..., m_k\}.$$

Caring with an obvious cyclic fashion for the cases $i \leq 3$ and $i \geq k-3$. The superscripts in $nbd_1^3(m)$ and $nbd_2^3(m)$ denote that the neighbourhoods are based of distances of length 3. In our example with 8 time series we chose a neighbourhood as $nbd^4 = nbd_1^4 \cup nbd_2^4$, proposing each model within nbd^4, and taking care to have $|nbd^4(m)| = \left|nbd\left(m'\right)\right|$ in (6.10).

The neighbourhood $nbd(m)$ of model m consists of the set of models with either a change in the position of two univariate time series or a move of time series to a different position. We change a randomly chosen time series one or two or three or four positions to the left or to the right, and we move a randomly chosen time series two or three or four positions to the left or to the right. As an example, suppose that there are 8 stocks in the $\mathbf{y}_t$ vector, and that a model m is given by the following ordering of the univariate time series 12345678. That is, the first stock is at position 1, the second stock is at position 2, and so on. A change in the position of two univariate time series could be, for example, 123**7**56**4**8, where we alter the positions of time series 4 and 7. A move of time series to a different position is, for example, **234**1567, where we move the time series of position 1 three positions to the right and the time series of positions 2, 3 and 4 one position to the left.

Delayed Rejection Algorithm

The idea of Delayed rejection algorithm was proposed by Tierney and Mira (1999). This strategy improves the Metropolis-Hastings algorithm in the sense (Peskun, 1973) that the resulting estimates have smaller asymptotic variance on a sweep by sweep basis. In our case, it is also useful as it increases the probability of moving between local modes of

the posterior density. We want to construct a Markov chain that moves through model space. In order to avoid cases where the chain remains at the same model over successive iterations, due to multimodalities of the model space, we can use the delayed rejection algorithm. Suppose that the current state of the chain is model m. Then, at the first stage, model $m^{'}$ is drawn from $q\left(m \rightarrow m^{'}\right)$ and accepted with probability

$$\min\left\{1, \frac{\pi\left(\mathbf{y}|m^{'}\right)}{\pi\left(\mathbf{y}|m\right)}\right\}.$$

If the candidate model $m^{'}$ is rejected, a new candidate model $m^{''}$ is proposed from $q\left(m^{'} \rightarrow m^{''}\right)$ at the second stage. That is, the new candidate model $m^{''}$ depends only on the last rejected candidate $m^{'}$. This is the symmetric delayed rejection algorithm. Note that the neighbourhood of models m and $m^{'}$, and the transitions from model m to $m^{'}$, and from $m^{'}$ to $m^{''}$ are defined as in the previous section. All models are assumed equally likely a priori. Tierney and Mira (1999) and Mira (2000) derived the probability of acceptance for this candidate by imposing detailed balance at each stage in order to preserve the stationary distribution. The probability of acceptance at this stage is given by

$$\min\left\{1, \frac{\max\left\{0, \left[\pi\left(\mathbf{y}|m^{''}\right) - \pi\left(\mathbf{y}|m^{'}\right)\right]\right\}}{\pi\left(\mathbf{y}|m\right) - \pi\left(\mathbf{y}|m^{'}\right)}\right\}.$$

This is a two stage symmetric delayed rejection algorithm. For the general formulation see Mira (2000).

6.5 Application to eight stocks from the US market

We illustrate the proposed multivariate latent GARCH model using 2350 daily data of eight stocks from the US stock market over the 1/1/1990 -1/1/1999 period. If S_t is the value of the stock at time t, then we model the rates of return $y_t = \ln\left(\frac{S_t}{S_{t-1}}\right)$, $t = 1, ..., T = 2349$. In Table 6.1, we present the summary statistics for the rates of return of the analyzed stocks, together with the Ljung-Box statistic computed for the

		Summary Statistics					
		Rates of return y_t				$\lvert y_t \rvert$	y_t^2
Order	*Stocks*	*Mean*	*Stdev*	*Kurtosis*	$LB(50)$	$LB(50)$	$LB(50)$
1	ATT	0.000374	0.01523	5.462	60.650	636.265	301.920
2	AXP	0.000514	0.01976	3.101	70.564	1107.453	1035.969
3	C	0.001000	0.02165	4.118	63.102	599.633	484.489
4	GE	0.000785	0.01355	2.536	82.068	928.415	877.401
5	JPM	0.000371	0.01650	3.279	80.431	1935.176	1893.695
6	PG	0.000702	0.01455	2.390	74.112	731.037	708.877
7	RAL	0.000443	0.01477	6.017	99.135	527.327	289.006
8	WMT	0.000844	0.01785	2.034	69.490	541.026	498.996

Table 6.1: Summary statistics for the rates of return of the analysed stocks.

rates of return y_t, for the absolute rates and for the squares of the rates. The Ljung-Box statistic is computed using 50 lags and show high level of autocorrelation in the squares values of rates of return and mainly in the absolute values. In Figure 6-1, we present the analyzed rates of return for the eight stocks. We also present, in Figure 6-2 the autocorrelations of the squares and of the absolute values of the rates of return. The autocorrelations for the cross product of the analyzed series are illustrated in Figure 6-3. These Figures (6-2 and 6-3) indicate that a multivariate model for time varying variances and covariances should be used.

The key steps in our analysis are as follows. First, we run the MC^3 algorithm without delayed rejection proposals (MC^3 without DRA) and the two stage symmetric delayed rejection algorithm (MC^3 with DRA) to find the "best" ordering of the individual time series and the posterior model probabilities. Second, we estimate the parameters of the "best" model using classical and Bayesian approach. Finally, we consider the problem of accounting for model uncertainty in the proposed time varying volatility model. We make inferences about quantities of interest such as future variances and covariances by using only the "best" model, using a set of most probable models and using Bayesian model averaging over all possible models.

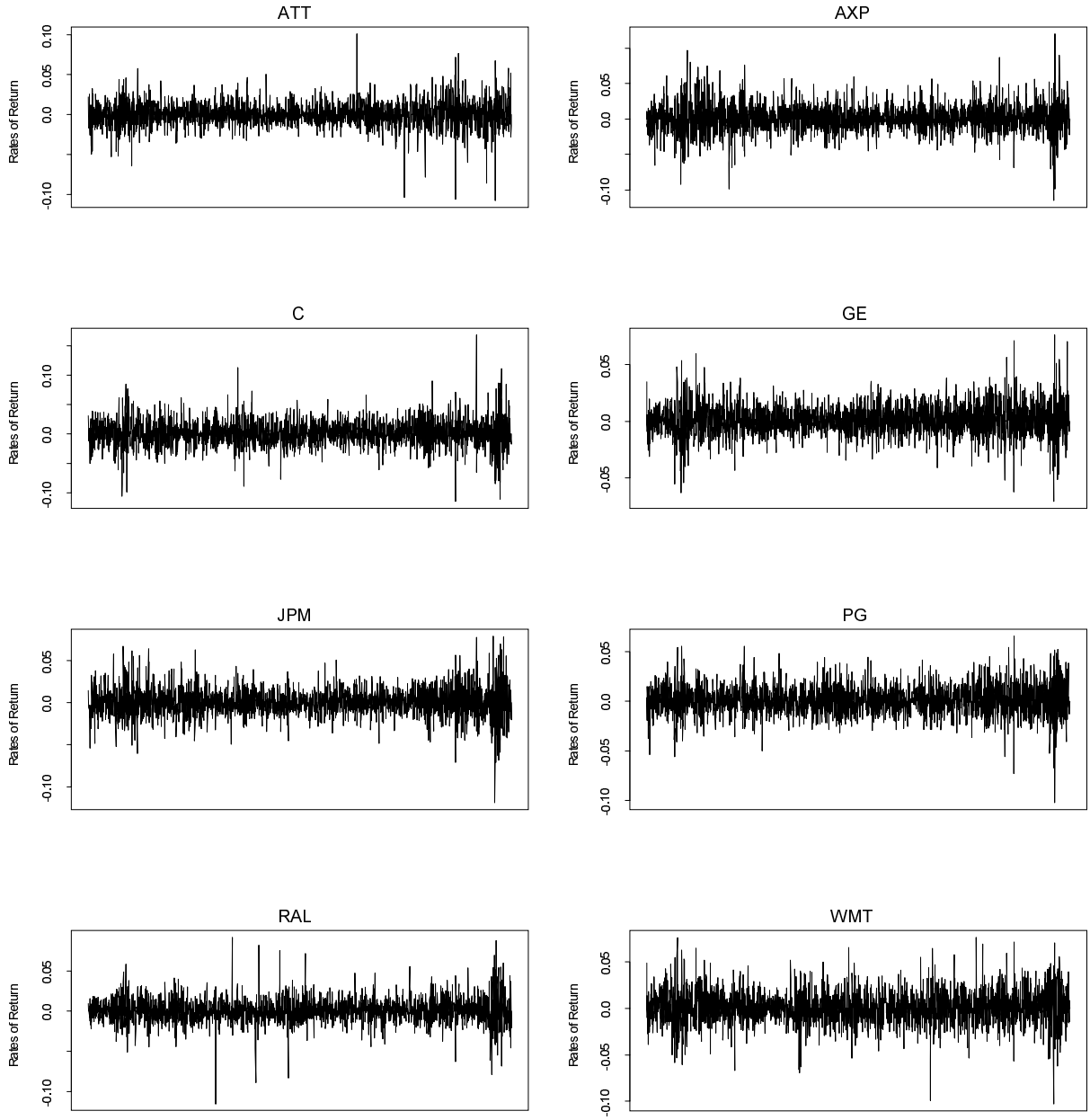

Figure 6-1: The analysed rates of return for the eight stocks of US market

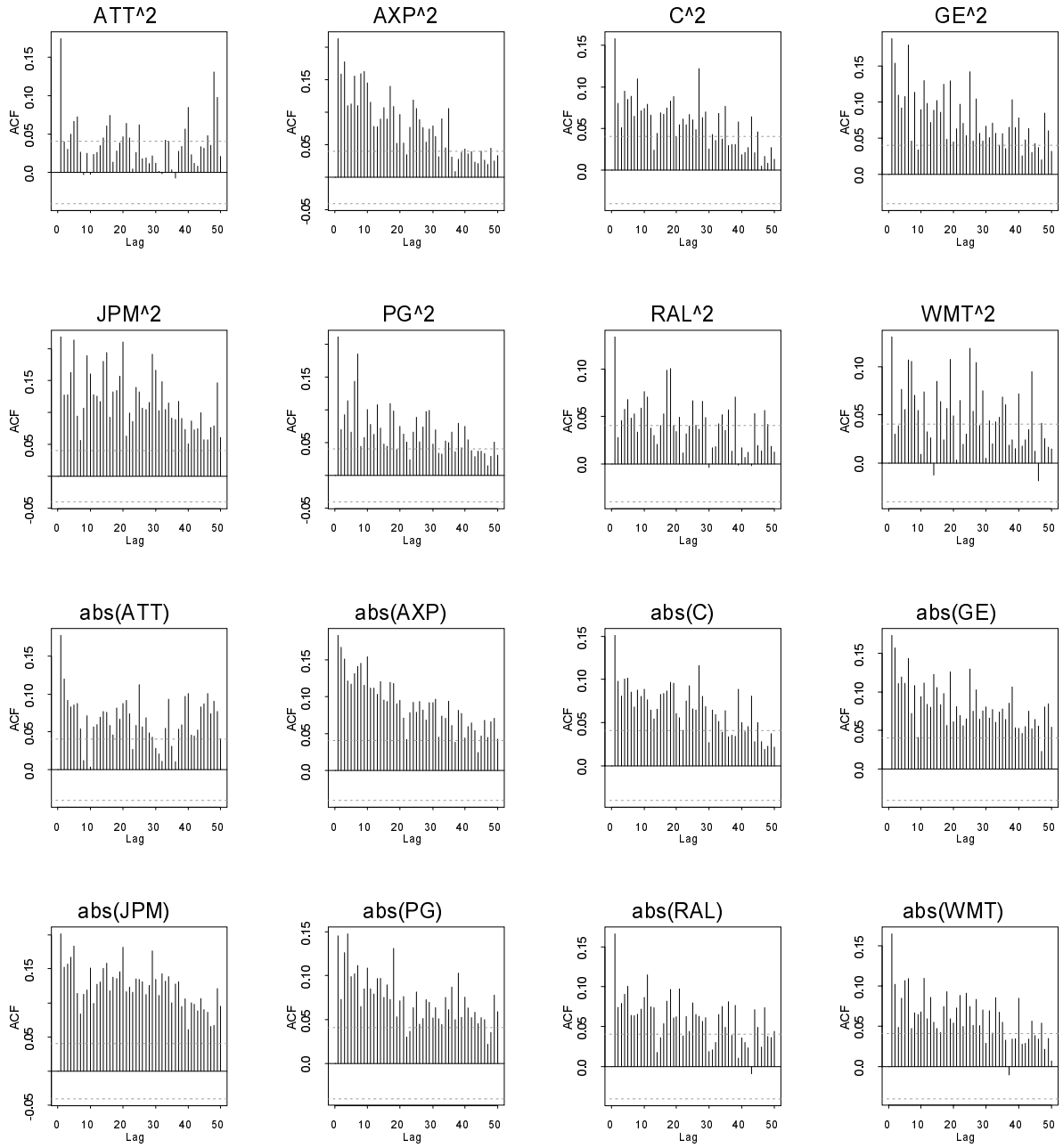

Figure 6-2: Autocorrelations of the squares and of the absolute values of the rates of return

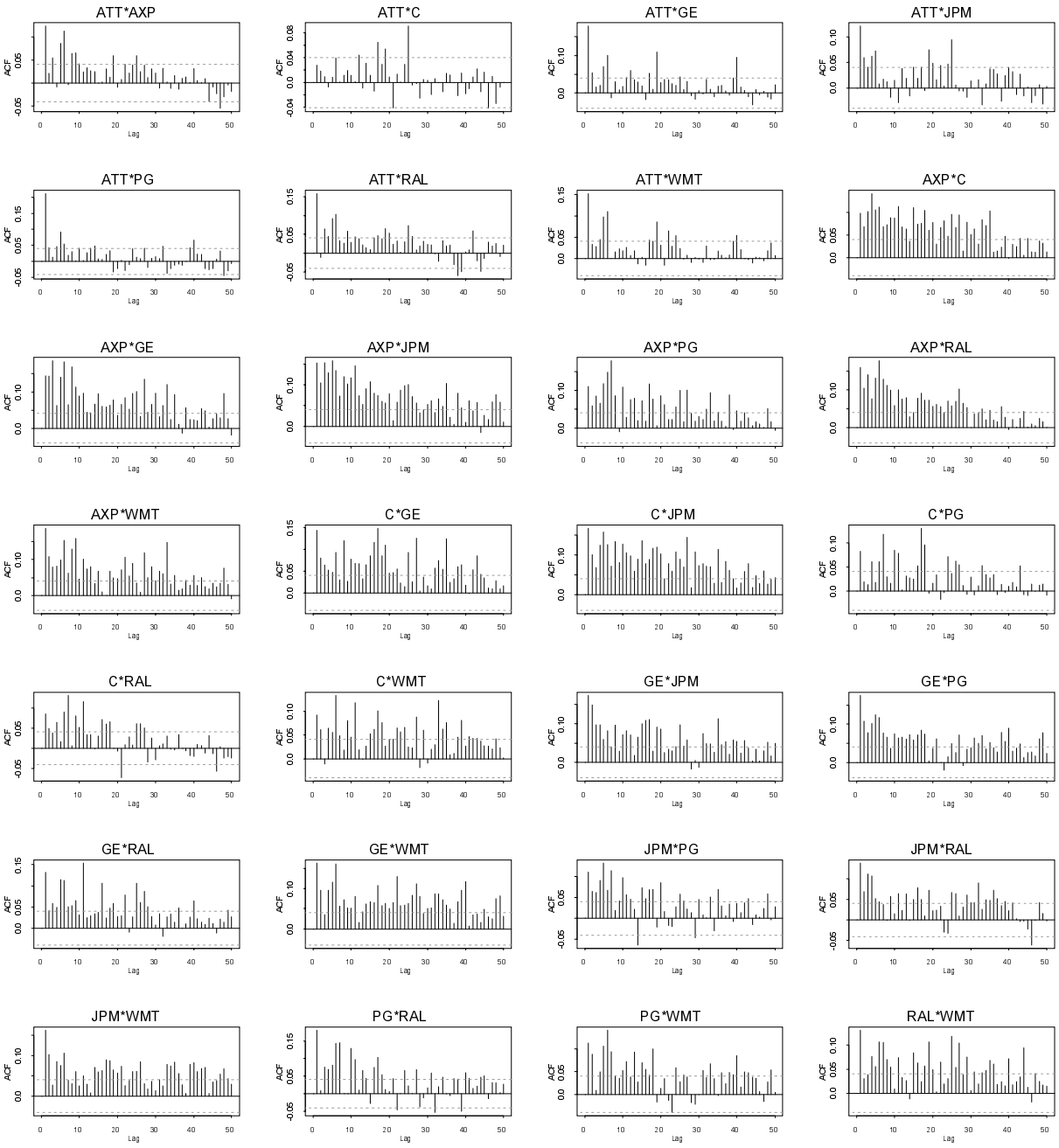

Figure 6-3: Autocorrelation of the cross product of the rates of return

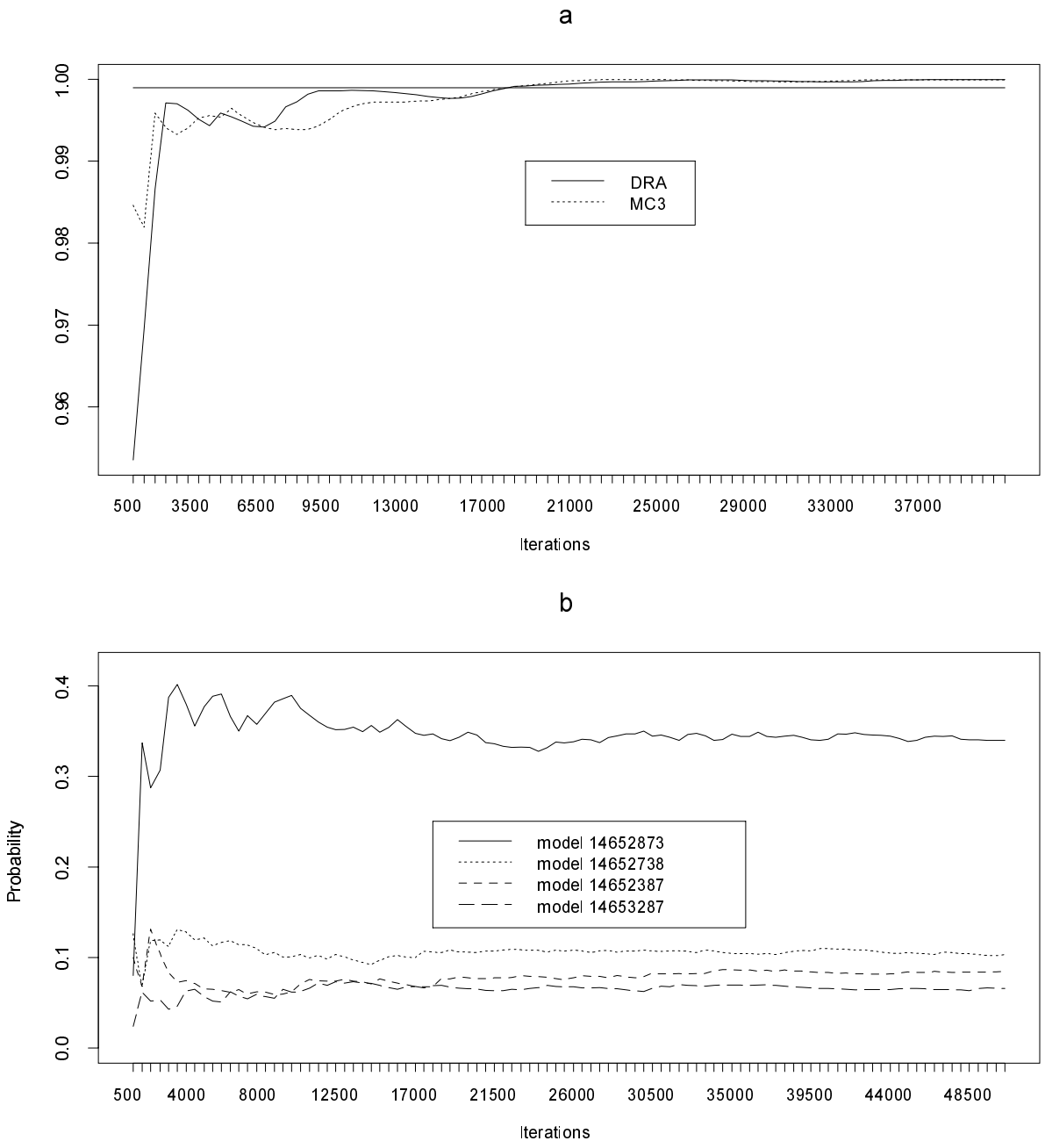

Figure 6-4: Convergence behavior of the MCMC model search algorithms: a: coefficient of determination in the subsampling diagnostic, b: probability across sweeps for the 4 most probable models using MC^3

6.5.1 MCMC model search and convergence assessment

We apply the MC^3 algorithm without and with DRA for 50000 iterations in order to find the posterior model probabilities. The subsampling methodology proposed by Giakoumatos, Vrontos, Dellaportas and Politis (1999) is used to check the convergence of the MCMC output taken from the above algorithms. The method is based on the use of subsampling for the construction of confidence regions for the mean (in our case) of the unique invariant distribution of the Markov chain. We construct the $(1-a)\,100\%$ confidence regions for the mean ($a = 0.05$) based on different (increasing) values $N_j = jN/100$, $j = 1, 2, ..., 100$, and $N = 50000$ iterations. We estimate the "burn-in" to be N^* if the "range" of the confidence regions versus $1/\sqrt{N_j}$ is approximately linear for $N > N^*$. Linearity can be checked by using the coefficient of determination of a weighted linear regression between the dependent variable "range" and $1/\sqrt{N_j}$, $j = 1, 2, ..., 100$. According to the subsampling convergence diagnostic, we stop the MCMC simulation when the range of this $(1-a)\,100\%$ confidence region for the mean is appropriately small, smaller than some prespecified absolute or relative measure of accuracy; see, for details, Giakoumatos, Vrontos, Dellaportas and Politis (1999). We focus our analysis on the 10 most probable models. Using the MCMC chains of $N = 50000$ iterations and choosing as threshold value $d = 0.999$ for the coefficient of determination in the subsampling convergence diagnostic we estimate the burn-in period. This comes out to be 17500 iterations for both the MC^3 without and with DRA (see Figure 6-4). Note however that the accuracy, that is the range of the 95% confidence region for the mean, is 0.0727 and 0.0640 for the MC^3 without and with DRA, respectively. These findings confirm that the MC^3 with DRA improves the Metropolis-Hastings algorithm in the sense that the resulting estimates have smaller asymptotic variance (smaller accuracy) on a sweep by sweep basis.

The resulting posterior model probabilities of the MC^3 without and with DRA are presented in Table 6.2. The posterior model probabilities are calculated by using 32500 iterations; that is, we have discarded the burn-in period of 17500 iterations. The accuracy

Model	MC^3 without DRA	MC^3 with DRA
14652873	0.34 (7)	0.35 (6)
14652738	0.10 (3)	0.10 (3)
14652387	0.09 (3)	0.08 (2)
14653287	0.06 (2)	0.08 (3)
14652837	0.06 (2)	0.06 (2)
14657328	0.05 (3)	0.07 (3)
14652783	0.05 (2)	0.05 (2)
14562873	0.04 (3)	0.03 (2)
14657238	0.03 (2)	0.03 (2)
14657283	0.02 (1)	0.03 (2)

Table 6.2: Posterior model probabilities of the 10 most probable models using the MC^3 algorithm without and with DRA. The order of time series is given in Table 6.1. Figures in brackets are accuracy$\times$100. Accuracy is the range of the 95% confidence region for the mean.

is the range of the 95% confidence region of the mean. They are calculated by running the subsampling diagnostic for each model using the MCMC output of 50000 iterations. The "best" model (ordering) is 14652873 with posterior probability 0.336 and 0.349 for the MC^3 without and with DRA, respectively. In total, 85 and 73 different models were visited during 50000 iterations of MC^3 without and with DRA, respectively.

For these estimates to be useful, we need to be confident that the reason we are only observing a small fraction of possible models in M, 85 and 73 out of 40320, is that other models have negligible posterior probability. To reassure our results we run the MC^3 algorithm starting from 50 different randomly chosen models. We also run the two stage symmetric delayed rejection algorithm (MC^3 with DRA) starting from the same 50 models. The results of these two approaches are presented in Table 6.3. We present the starting models and the number of iterations needed for the MC^3 without and with DRA to reach the most probable model 14652873.

Both methods seem to be very flexible since the algorithms arrive at the "best" ordering very fast. There was no evidence of the existence of any other regions of model space of high probability. While this is no guarantee that such regions do not exist, it does provide some reassurance. MC^3 with DRA seems to be significantly better

Initial model	Iterations MC^3	Iterations DRA	Initial model	Iterations MC^3	Iterations DRA	Initial model	Iterations MC^3	Iterations DRA
36187425	134	75	73846512	556	183	21637458	555	81
74165283	317	179	76352814	317	232	87261435	154	233
64852713	220	178	57264183	221	343	76531284	316	178
12367854	252	178	28671534	317	75	72861534	319	252
31725648	221	234	26538471	320	256	51378624	48	348
81567243	319	125	34856712	222	144	18576423	220	176
15246837	53	228	15638274	220	75	58374162	131	229
82637541	209	256	78612453	220	64	87351426	251	255
25381764	212	231	64135278	333	231	51732846	351	40
72431586	250	70	62574183	132	392	47235816	220	145
51746823	251	397	83142756	208	43	85623714	88	231
81356742	319	85	72145386	350	181	27814563	252	394
15364872	317	228	71852463	317	257	14827563	316	178
47186352	351	229	24563178	316	144	73461285	221	394
47238165	130	396	47813256	34	43	25741386	221	30
63745218	212	180	38426751	251	231	43518762	89	68
26178345	208	63	65748123	249	230			

Table 6.3: Number of iterations needed for the MC^3 without and with DRA to reach the most probable model 14652873.

than MC^3 without DRA. The mean values of the number of iterations needed for the MC^3 without DRA and for MC^3 with DRA to reach the most probable model 14652873 are approximately 246 and 194 iterations, respectively, and their corresponding standard deviations are approximately 105 and 104. Having in mind that multimodalities in model space is a very frequent phenomenon, MC^3 with DRA is useful as it increases the probability of moving between local modes of the posterior density.

6.5.2 Inference for a given model

Having been able to find the "best" ordering we present the estimates for the parameters of matrix W and the corresponding standard errors in Table 6.4. These standard errors (in brackets) are given by the square root of the diagonal elements of the inverse of the information matrix. The robust estimates of the covariance matrix (6.5) are also

calculated and the standard error are presented in Table 6.4 (in square brackets). We also estimate the parameters of the multivariate model for the "best" ordering by using Bayesian analysis and MCMC methods. We transform the positive parameters to "near normality" using the logarithmic transformation. These transformations improve the behavior of our MCMC algorithm. For our illustration we choose non-informative (constant) priors for all the parameters. The blocking sampling scheme was used in order to update the model parameters. We ran the algorithm for 260000 iterations, and we kept one value every 100 iterations (to save computer space). The resulting samples of 2600 values were checked for convergence by using the subsampling diagnostic proposed by Giakoumatos, Vrontos, Dellaportas and Politis (1999). The method is based on the use of subsampling for the construction of confidence regions for the t-quantile ($t = 0.90$) of the unique invariant distribution of the Markov chain. We construct the $(1-a)\,100\%$ confidence regions for the 0.90 quantile ($a = 0.05$) based on different (increasing) values $N_j = jN/100$, $j = 1, 2, ..., 100$, and $N = 2600$ iterations. We estimate the "burn-in" to be N^* if the "range" of the confidence regions versus $1/\sqrt{N_j}$ is approximately linear for $N > N^*$. Linearity is checked by using the coefficient of determination of the weighted linear regression between the dependent variable "range" and $1/\sqrt{N_j}$, $j = 1, 2, ..., 100$. The reason that the t-quantile (with a large t, say $t = 0.90$) is considered, is based on the notion that stabilization of estimates of the invariant distribution of the Markov chain (especially in the tails) is a reliable indicator of the target distribution having been achieved. We stop the MCMC simulation when the range of this $(1-a)\,100\%$ confidence region for the mean is appropriately small. Using the MCMC chains of $N = 2600$ iterations and choosing as threshold value $d = 0.999$ for the coefficient of determination in the subsampling convergence diagnostic we estimate the burn-in period. This comes out to be 780 iterations. The accuracy, that is the range of the 95% confidence region for the mean, is 0.013. The convergence of the parameters was also checked by using the tests proposed by Heidelberger and Welch (1983) and Raftery and Lewis (1992). The first diagnostic indicates that the convergence has been achieved after a burn-in period of 780

	Classical	*Bayesian*		*Classical*	*Bayesian*
w_{21}	0.30 (0.02) [0.02]	0.30 (0.02)	w_{65}	0.10 (0.02) [0.02]	0.10 (0.02)
w_{31}	0.25 (0.02) [0.03]	0.25 (0.02)	w_{71}	0.19 (0.02) [0.02]	0.19 (0.02)
w_{32}	0.42 (0.02) [0.02]	0.42 (0.02)	w_{72}	0.33 (0.02) [0.02]	0.33 (0.02)
w_{41}	0.27 (0.02) [0.03]	0.27 (0.02)	w_{73}	0.22 (0.02) [0.02]	0.22 (0.02)
w_{42}	0.40 (0.02) [0.03]	0.41 (0.02)	w_{74}	0.09 (0.02) [0.02]	0.09 (0.02)
w_{43}	0.14 (0.02) [0.02]	0.14 (0.02)	w_{75}	0.05 (0.02) [0.02]	0.05 (0.02)
w_{51}	0.27 (0.02) [0.03]	0.27 (0.02)	w_{76}	0.08 (0.02) [0.02]	0.08 (0.02)
w_{52}	0.47 (0.03) [0.03]	0.48 (0.03)	w_{81}	0.36 (0.03) [0.04]	0.36 (0.03)
w_{53}	0.20 (0.03) [0.03]	0.21 (0.03)	w_{82}	0.57 (0.03) [0.03]	0.57 (0.03)
w_{54}	0.35 (0.02) [0.02]	0.35 (0.02)	w_{83}	0.30 (0.03) [0.03]	0.30 (0.03)
w_{61}	0.30 (0.02) [0.03]	0.30 (0.02)	w_{84}	0.42 (0.02) [0.04]	0.42 (0.03)
w_{62}	0.50 (0.02) [0.03]	0.50 (0.03)	w_{85}	0.21 (0.02) [0.02]	0.21 (0.02)
w_{63}	0.21 (0.02) [0.03]	0.21 (0.02)	w_{86}	0.07 (0.02) [0.03]	0.07 (0.02)
w_{64}	0.15 (0.02) [0.02]	0.15 (0.02)	w_{87}	0.07 (0.03) [0.03]	0.07 (0.03)

Table 6.4: Estimates for the parameters in matrix W of multivariate latent GARCH model. Classical: estimates using Fisher scoring, figures in brackets are standard deviations taken using Fisher scoring, figures in square brackets are standard deviations proposed by Bollerslev and Wooldridge (1992); Bayesian: posterior means and posterior standard deviations (in brackets).

iterations, where the latter diagnostic gives values for the dependent factor around one. Estimated posterior means and standard deviations for the parameters of matrix W of the latent GARCH model (6.1) are illustrated in Table 6.4. Note that, all the parameters of matrix W are significant. We also present in Figures 6-5 and 6-6 the convergence diagrams of the posterior sample of the parameters of the multivariate latent GARCH model.

6.5.3 Model uncertainty and prediction

In multivariate financial models, prediction of the future covariance matrix is of particular interest. Having been able to calculate the posterior model probabilities, it seems natural to account for model uncertainty in our predictive inferences. Suppose that we are interested in H_{T+1}, the predictive covariance matrix at time $T+1$. Then, its posterior

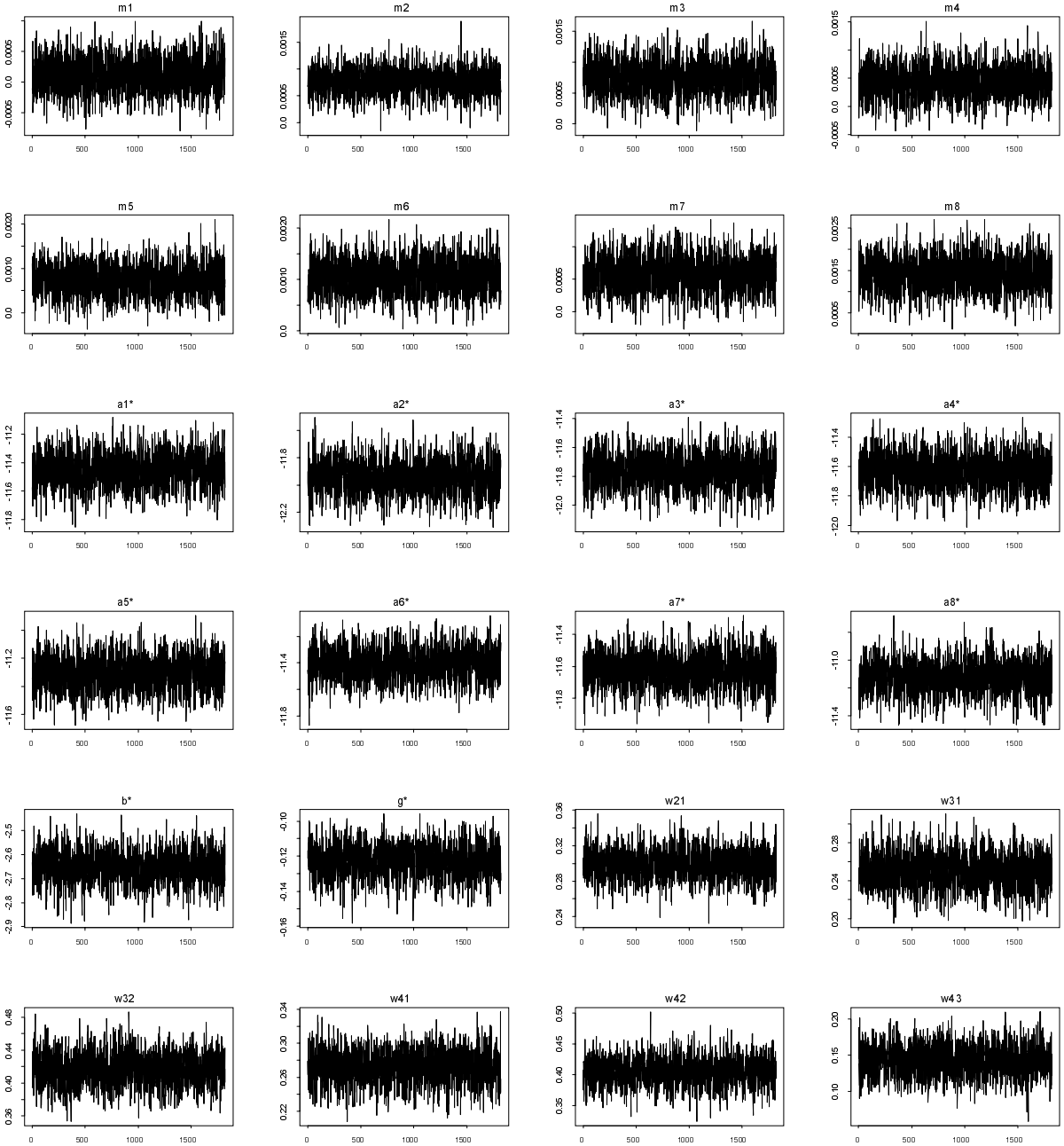

Figure 6-5: Convergence diagrams for the model parameters

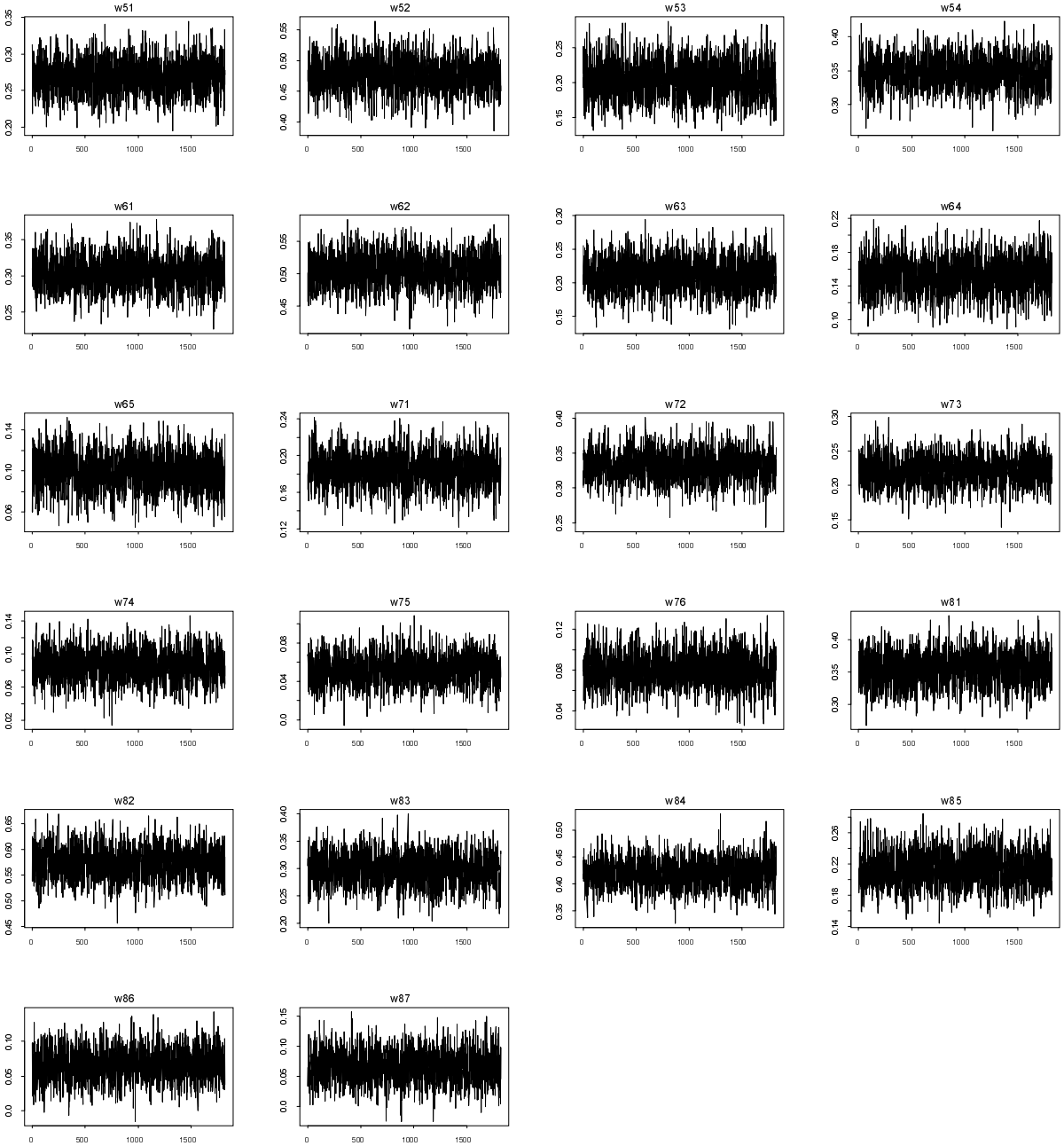

Figure 6-6: Convergence diagrams for the model parameters

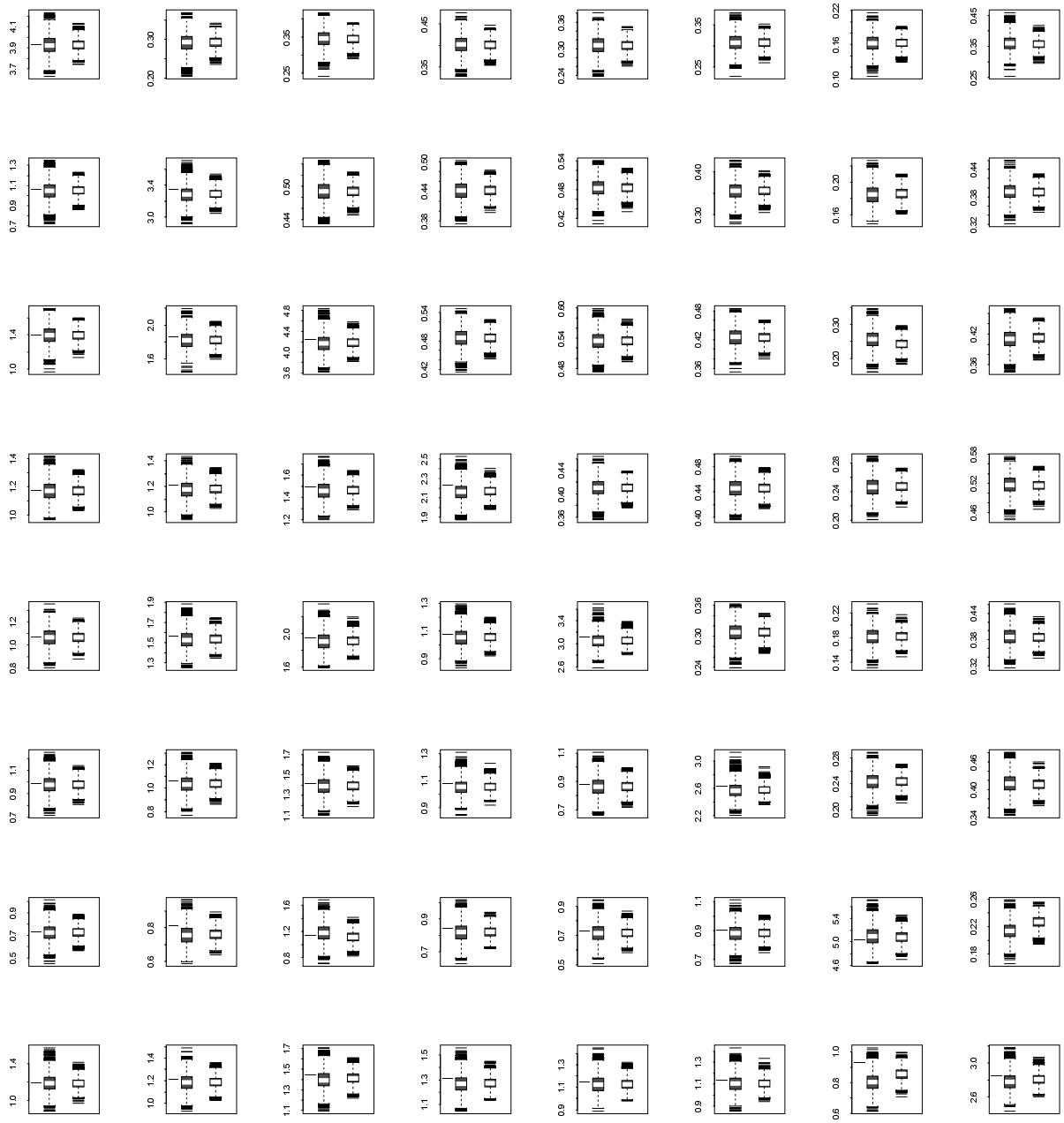

Figure 6-7: Posterior boxplots of the one step-ahead forecast for the variances (diagonal part of the graph), covariances (lower diagonal part of the graph) and correlations (upper diagonal part of the graph) based on the "best" model, on Bayesian model average of the 4 most probable models, and on Bayesian model average over all models that MC^3 algorithm has visited.

distribution given data $\mathbf{y}$ is given by

$$p\left(H_{T+1}|\mathbf{y}\right) = \sum_{i=1}^{K} p\left(H_{T+1}|m_i, \mathbf{y}\right) p\left(m_i|\mathbf{y}\right), \tag{6.12}$$

which is an average of the posterior predictive distribution under each model weighted by their posterior model probabilities. Computation of (6.12) is straightforward after the implementation of MC^3 or of delayed rejection algorithm. First, given a model m_i, a posterior sample of $p\left(H_{T+1}|m_i, \mathbf{y}\right)$ is just obtained by calculating, for each sampled point in $\boldsymbol{\theta}$, the covariance matrices $H_1, H_2, ..., H_{T+1}$. Then (6.12) suggests that in order to obtain a sample of $p\left(H_{T+1}|\mathbf{y}\right)$, each sampled point under model m_i should be taken with probability $p\left(m_i|\mathbf{y}\right)$. Thus, the derived sample of $p\left(H_{T+1}|\mathbf{y}\right)$ is obtained by weighting all samples of $p\left(H_{T+1}|m_i, \mathbf{y}\right)$ by the corresponding $p\left(m_i|\mathbf{y}\right)$.

We made inference about the predictive covariance matrix at time $T+1$ using the "best" ordering, Bayesian model averaging based on the 4 most probable models and Bayesian model average based on all possible models using equations (6.8) and (6.9). In our predictive exercise, we ran the algorithms for 50000 iterations, we discarded the first 10000 iterations as burn-in, and for the rest 40000 iterations we kept one value every 10 iterations, obtaining a sample of 4000 iterations. A posterior sample of $p\left(H_{T+1}|m, \mathbf{y}\right)$ is obtained using the "best" model by calculating, for each one of the 4000 sampled points in $\boldsymbol{\theta}$, the covariance matrix H_{T+1}. Then, we calculated the predictive density of H_{T+1} based on the four most probable models. To achieve this, we constructed all predictive densities $p\left(H_{T+1}|m_i, \mathbf{y}\right)$, $i = 1, ..., 4$, under each model, for each one of the 4000 sampled points in $\boldsymbol{\theta}$, and then we weighted all samples of $p\left(H_{T+1}|m_i, \mathbf{y}\right)$ by the corresponding normalized posterior model probabilities. Finally, we calculated the predictive covariance matrix using Bayesian model averaging over all models that have been visited during 50000 iterations of MC^3 algorithm. For each one of these models we estimated their parameters, evaluated the predictive covariance matrix H_{T+1} based on the maximum likelihood estimates, and then we weighted these values using the posterior model prob-

abilities $p(m_i|\mathbf{y})$. We present in Figure 6-7 the posterior boxplots of the one step-ahead forecasts for the volatility of the eight stocks (diagonal part of the graph), for the covariances of the eight stocks (lower diagonal part of the graph), and for the correlations of the stocks (upper diagonal part of the graph). The first posterior boxplot in each box of the graph is based on the "best" model, while the second posterior boxplot is based on Bayesian model average of the 4 most probable models. All the predictive values for the volatilities and the covariances have been multiplied by 10000. The predictive distribution of the elements of the covariance matrix based on Bayesian model averaging is narrower than the corresponding predictive distribution based on the "best" model, since BMA accounts for model uncertainty. The 'line' in Figure 6-7 illustrates the one step-ahead forecast for the volatility and the covariances of the analyzed stocks based on Bayesian model averaging over all models that MC^3 algorithm has visited.

6.6 Discussion

In this chapter, we propose a new multivariate GARCH model, where the covariance matrix is always positive definite and the number of parameters is relatively small with respect to other multivariate models. The model can be thought as a factor model with full factor representation. This allows a dynamic behavior of the covariances and the correlations.

The estimation of the parameters of the multivariate model is done by using classical and Bayesian techniques. Maximum likelihood estimation is implemented by using the method of Fisher scoring, while the MCMC algorithm is based on a blocking sampling scheme which accelerates the convergence of the model parameters. Due to the fact that the covariance matrix is guaranted to be positive definite, and that the estimation of the parameters is easily implemented, we believe that the model can be applied very easily to high dimensional problems.

We address the problem of model selection among different models (orderings) of

the analyzed time series. We apply a Markov chain Monte Carlo model composition (MC^3) method without and with a second delayed rejection stage. This improves the Metropolis-Hastings algorithm in the sense that the resulting estimates have smaller asymptotic variance (smaller accuracy) on a sweep by sweep basis. We believe that this algorithm is very flexible and useful as it increases the probability of moving between local modes of the posterior density.

We also consider the problem of accounting for model uncertainty in the proposed multivariate GARCH model. Conditioning on a single selected model ignores model uncertainty. We make inferences about quantities of interest such as prediction of future variances and covariances using Bayesian model averaging over a set or over all possible models.

Chapter 7

Future research

The thesis presents aspects of univariate and multivariate ARCH-type time-varying volatility models. We deal with the general problem of inference, prediction and model comparison of ARCH-type models by using Bayesian techniques and in particular Markov Chain Monte Carlo methods. We demonstrate that the MCMC methods provide an idealized way to extract any posterior summary of interest such as functions of parameters, to address the problem of model selection, and in addition to construct predictive densities that take into account model uncertainty by using Bayesian model averaging.

In the analyzed ARCH-type models estimates of the model parameters are obtained by using the Metropolis-Hastings algorithm. The efficiency of the algorithm is improved by using a simultaneous vector update or a blocking sampling scheme. Further investigation is needed in order to construct alternative sampling schemes which are faster mixing or easier to simulate. For example, the Auxiliary variable (AV) sampling techniques (Swendsen and Wang, 1987, Edwards and Sokal, 1988, Besag and Green, 1993, Higdon, 1998, Damien, Wakefield and Walker, 1999) could be adopted. The basic idea of AV sampling is that the parameter space of the posterior density can be increased by including extra latent variables which make the resulting posterior density more tractable by sampling methods. Properties of AV sampling has been examined by Mira and Tierney (1998), Roberts and Rosenthal (1997), while Giakoumatos, Dellaportas and Politis

(1998) has used these techniques in the analysis of the unobserved ARCh model.

The model selection problem in univariate and multivariate ARCH-type models is considered by using the Bayes Factor. MCMC methods such as the Reversible Jump and the MC^3 together with the delayed rejection algorithm were used in order to obtain posterior model probabilities in univariate (chapter four) and multivariate (chapter six) models. Model determination for multivariate ARCH-type models (chapter five) was also addressed by using predictive distributions. However, other methods (see for example, Chib, 1995, Chib and Jeliazkov, 1999) can be used for estimating the marginal likelihood. In particular, the approach of Chib and Jeliazkov (1999) which is based on the output of Metropolis-Hastings algorithm, seems a very promising alternative in the time-varying volatility models we analyzed.

In chapter six, a multivariate time series model with time-varying conditional variances and covariances was introduced and analyzed. For the specification of the variances of the "latent" variables a GARCH(1,1) model was used. The model can be extended to a general asymmetric multivariate model by allowing the variances of the "latent" variables to follow a "leverage" or "asymmetric" model, in which good news and bad news have different predictability for future variance. Such formulations are, for example, the EGARCH model of Nelson (1991) or the model of Glosten, Jagannathan and Runkle (1993).

Bibliography

[1] Aguilar, O. and West, M. (2000). Bayesian Dynamic Factor Models and Portfolio Allocation. *Journal of Business and Economic Statistics*, **18**, 338-357.

[2] Akgiray, V. (1989). Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts. *Journal of Business*, **62**, 55-80.

[3] Alexander, C.O. (2000). A Primer on the Orthogonal GARCH Model. Manuscript, ISMA Centre, The Business School for Financial Markets, University of Reading, UK.

[4] Baillie, R.T. and Bollerslev, T. (1989). The message in daily exchange rates: A conditional variance tale. *Journal of Business and Economic Statistics*, **7**, 297-305.

[5] Baillie, R.T. and Bollerslev, T. (1990). A multivariate generalized ARCH approach to modeling risk premia in forward foreign exchange rate markets. *Journal of International Money and Finance*, **9**, 309-324.

[6] Baillie, R.T. and Myers, R.J. (1991). Bivariate GARCH estimation of optimal commodity futures hedge. *Journal of Applied Econometrics*, **16**, 109-124.

[7] Bartlett, M.S. (1957). Comment on "a statistical paradox" by D. V. Lindley. *Biometrika*, **44**, 533-534.

[8] Bauwens, L. and Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler. CORE discussion paper, Universite catholique de Louvain, Belgium.

[9] Bauwens, L. and Lubrano, M. (1999). Bayesian option pricing using asymmetric GARCH. *Sankya*, to appear.

[10] Bauwens, L., Bos, C.S. and van Dijk, H.K (1999). Adaptive Polar Sampling with an application to a Bayes measure of Value-at-Risk. Econometric Institute Report EI-9940/A, Erasmus University Rotterdam, The Netherlands.

[11] Bekaert, G. and Hodrick, R.J. (1993). On Biases in the Measurement of Foreign Exchange Risk Premiums. *Journal of International Money and Finance*, **12**, 115-138.

[12] Bekaert, G. and Wu, G. (2000). Asymmetric Volatility and Risk in Equity Markets. *The Review of Financial Studies*, **13**, 1-42.

[13] Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian Theory*. Chichester: Wiley.

[14] Berndt, E.K., Hall, B.H., Hall, R.E. and Hausman, J.A. (1974). Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement,* **3/4**, 653-665.

[15] Bera, A.K. and Higgins, M.L. (1993). ARCH models: properties, estimation and testing. *Journal of Economic Surveys*, **7**, 305-366.

[16] Bera, A.K. and Higgins, M.L. (1997). ARCH and Bilinearity as Competing Models for Nonlinear Dependence. *Journal of Business and Economic Statistics*, **15**, 43-50.

[17] Bera, A.K., Garcia, P. and Roh, J-S. (1991). Estimation of time-varying hedge ratios for agricultural commodities: BGARCH and random coefficient approaches, Mimeo, Department of Economics, University of Illinois at Urbana-Champaign.

[18] Berger, J.O. and Delampady, M. (1987). Testing Precise Hypotheses (with discussion). *Statistical Science*, **2**, 317-352.

[19] Besag, J. and Green, P.J. (1993). Spatial Statistics and Bayesian Computation. *Journal of the Royal Statistical Society, B*, **55**, 25-37.

[20] Besag, J., Green, P.J., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10**, 3-66.

[21] Bodurtha, J.N. and Mark, N.C. (1991). Testing the CAPM with time varying risks and returns. *Journal of Finance,* **46**, 1485-1505.

[22] Bollerslev, T. (1986). Generalized Autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.

[23] Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, **69**, 542-547.

[24] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economic Statistics*, **72**, 498-505.

[25] Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modeling in finance - A review of the theory and empirical evidence. *Journal of Econometrics*, **52**, 5-59.

[26] Bollerslev, T. and Engle, R.F. (1993). Common persistence in conditional variances. *Econometrica*, **61**, 167-186.

[27] Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994). ARCH models. *The Handbook of Econometrics 4* (R.F. Engle, D.L. McFadden, eds.), North-Holland, 2959-3038.

[28] Bollerslev, T., Engle, R.F. and Wooldridge, J.M. (1988). A Capital Asset Pricing Model with Time-varying Covariances. *Journal of Political Economy*, **96**,116-131.

[29] Bollerslev, T. and Wooldridge, J.M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, **11**, 143-172.

[30] Bos, C.S., Mahieu, R.J. and van Dijk, H.K. (1999). Daily Exchange Rate Behaviour and Hedging of Currency Risk. Econometric Institute Report EI-9936/A, Erasmus University Rotterdam, The Netherlands.

[31] Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, **52**, 115-127.

[32] Box, G. (1980). Sampling and Bayes' inference in scientific modeling and robustness (with discussion). *Journal of the Royal Statistical Society*, A, **143**, 382-430.

[33] Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

[34] Braun, P.A., Nelson, D.B. and Sunier, A.M. (1995). Good News, Bad News, Volatility, and Betas. *The Journal of Finance*, **50**, 1575-1603.

[35] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods.* 2nd edition, New York: Springer-Verlag.

[36] Brooks, S.P. (1998). Markov Chain Monte Carlo method and its application. *The Statistician*, **47**, 69-100.

[37] Brooks, S.P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics,* **7**, 434-455.

[38] Brooks, S.P. and Guidici, P. (1998). Convergence Assessment for Reversible Jump MCMC simulations. *Bayesian Statistics 6* (J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford: Oxford University Press.

[39] Brooks, S.P. and Roberts, G.O. (1999). Assessing Convergence of Markov Chain Monte Carlo Algorithms. *Statistics and Computing,* **8**, 319-335.

[40] Carlstein, E. (1986). The use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence. *The Annals of Statistics,* **14**, 1171–1179.

[41] Casella, G. and George, E.I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, **46**, 167-174.

[42] Chib, S. (1995). Marginal Likelihood From the Gibbs Output. *Journal of the American Statistical Association*, **90**, 1313-1321.

[43] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, **49**, 327-335.

[44] Chib, S. and Jeliazkov, I. (1999). Marginal Likelihood from the Metropolis-Hastings Output. Manuscript, John M. Olin School of Business, Washington University, USA.

[45] Chib, S., Nardari, F. and Shephard, N. (1999). Analysis of High Dimensional Multivariate Stochastic Volatility Models. Manuscript, Department of Economics, Nuffield College, University of Oxford, UK.

[46] Christodoulakis, G.A. and Satchell, S.E. (1998). Correlated ARCH (CorrARCH): a new model for the time-varying Conditional Correlation between financial asset returns. *Technical report*, Department of Economics, Birkbeck College, University of London, UK.

[47] Cowles, M.K. and Carlin, B.P. (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association,* **91**, 883-904.

[48] Damien, P., Wakefield, J. and Walker, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of Royal Society, B*, **61**, 331 - 344.

[49] Davidian, M. and Carroll, R.J. (1987). Variance function estimation. *Journal of the American Statistical Association*, **82**, 1079-1091.

[50] Day, T.E. and Lewis, C.M. (1992). Stock market volatility and the information content of stock index options. *Journal of Econometrics*, **52**, 267-287.

[51] Dellaportas, P. (1996). Computational strategies for the implementation of the Bayesian paradigm. Hermis '96: 3rd Hellenic European Conference on Mathematics and Informatics.

[52] Dellaportas, P. and Smith, A.F.M. (1993). Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling. *Applied Statistics,* **42**, 443-459.

[53] Dellaportas, P., Karlis, D. and Xekalaki, E. (1997). Bayesian Analysis of Finite Poisson Mixtures. *Technical Report*, no. 29, Department of Statistics, Athens University of Economics and Business, Athens, Greece.

[54] Dellaportas, P. and Forster, J.J. (1999). Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika*, **86**, 615-633.

[55] Dellaportas, P. and Wright, D.E. (1991a). Positive embedded integration in Bayesian analysis. *Statistics and Computing*, **1**, 1-12.

[56] Dellaportas, P. and Wright, D.E. (1991b). Numerical prediction for the two parameter Weibull distribution. *The Statistician*, **40**, 365-372.

[57] Dellaportas, P. and Wright, D.E. (1992). A numerical integration strategy in Bayesian analysis. *Bayesian Statistics 4* (J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford: Oxford University Press, 601-606.

[58] Demos, A. and Sentana, E. (1998). An EM algorithm for conditionally heteroscedastic factor models. *Journal of Business and Economic Statistics*, **16**, 357-361.

[59] Dhrymes, P.J. (1984). *Mathematics for Econometrics*. New York: Springer-Verlag.

[60] DiCiccio, T.J., Kass, R.E., Raftery, A. and Wasserman, L. (1997). Computing Bayes Factors by Combining Simulation and Asymptotic Approximations. *Journal of the American Statistical Association*, **92**, 903-915.

[61] Diebold, F.X. and Lopez, J.A. (1995). ARCH models. *Macroeconomics: Developments, Tensions and Prospects* (K. Hoover, ed).

[62] Diebold, F.X. and Nerlove, M. (1989). The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model. *Journal of Applied Econometrics*, **4**, 1-21.

[63] Doukhan, P. (1994). *Mixing: properties and examples.* Lecture Notes in Statistics, 85, New York: Springer-Verlag.

[64] Draper, D. (1995). Assessment and Propagation of Model Uncertainty (with discussion). *Journal of the Royal Statistical Society, B*, **57**, 45-97.

[65] Dubofsky, D.A. (1992). *Options and Financial Futures: valuation and uses*. Singapore: McGraw-Hill.

[66] Edwards, R.G. and Sokal, A.D. (1988). Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo Algorithm. *Physical Review Letters*, **38**, 2009-2012.

[67] Efron, B. (1982). *The Jacknife, the Bootstrap and other Resampling Plans*. Philadelphia, PA: SIAM.

[68] Elie, L. and Jeantheau, T. (1995). Consistance dans les modeles heteroscedastiques. *Comptes Rendus de l' Academie des Sciences Serie I*, **320**, 1255-1258.

[69] Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, 987-1008.

[70] Engle, R.F. (1983). Estimates of the Variance of U.S. Inflation Based upon the ARCH Model. *Journal of Money, Credit, and Banking*, **15**, 286-301.

[71] Engle, R.F. (1995). *ARCH selected readings.* New York: Oxford University Press.

[72] Engle, R.F. (2000). Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models. Manuscript, Department of Economics, University of California, San Diego, USA.

[73] Engle, R.F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, **5**, 1-50, 81-87.

[74] Engle, R.F., Granger, C.W.J. and Kraft, D. (1984). Combining competing forecasts of inflation using a bivariate ARCH model. *Journal of Economic Dynamics and Control*, **8**, 151-165.

[75] Engle, R.F., Hendry, D.F. and Trumble, D. (1985). Small sample properties of ARCH estimatorsand tests. *Canadian Journal of Economics*, **18**, 66-93.

[76] Engle, R.F. and Kroner K.F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, **11**, 122-150.

[77] Engle, R.F., Lilien, D.M. and Robins, R.P. (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica*, **55**, 391-407.

[78] Engle, R.F. and Ng, V.K. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, **48**, 1749-1778.

[79] Engle, R.F. and Susmel, R. (1993). Common Volatility in international equity markets. *Journal of Business and Economic Statistics*, **11**, 167-176.

[80] Engle, R.F., Ng, V.K. and Rothschild, M. (1990). Asset pricing with a Factor-ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics*, **45**, 213-237.

[81] Erkanli, A. (1994). Laplace Approximations for Posterior Expectation when the model occurs at the boundary of the parameter space. *Journal of the American Statistical Association*, **89**, 250-258.

[82] Evans, M. and Swartz, T. (1995). Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science*, **10**, 254-272.

[83] Fama, E.F. (1965). The behavior of stock market prices, *Journal of Business*, **38**, 34-105.

[84] Fiorentini, G., Calzolari, G. and Panattoni, L. (1996). Analytic Derivatives and the computation of GARCH Estimates. *Journal of Applied Econometrics*, **11**, 399-417.

[85] French, K.R., Schwert, G.W. and Stambaugh, R.F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, **19**, 3-29.

[86] Fukuchi, J. (1997). Subsampling and Model Selection in Time Series Analysis. Technical report 97-2, Statistical Research Group, Hiroshima University, Japan.

[87] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, **85**, 398-409.

[88] Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, **85**, 972-985.

[89] Gelfand, A.E., Dey, D.K. and Chang, H. (1992). Model Determination using Predictive Distributions with Implementation via Sampling-Based Methods (with discussion). *Bayesian Statistics 4* (J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford: Oxford University Press, 147-167.

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

[91] Gelman, A. (1996). Inference and Monitoring Convergence. *Markov Chain Monte Carlo in Practice* (W.R. Gilks, S. Richardson and D.J. Spiegelhalter, eds), Chapman and Hall, 131-143.

[92] Gelman, A. and Rubin, D. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science,* **7**, 457-511.

[93] Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

[94] George, E.I. and McCulloch, R.E. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, **88**, 881-889.

[95] Geweke, J. (1986). Modelling the persistence of conditional variances: a comment. *Econometric Reviews*, **5**, 57-61.

[96] Geweke, J. (1988). Antithetic acceleration of Monte Carlo integration in Bayesian inference. *Journal of Econometrics*, **38**, 73-89.

[97] Geweke, J. (1989a). Bayesian Inference in Econometric models using Monte Carlo Integration. *Econometrica*, **57**, 1317-1339.

[98] Geweke, J. (1989b). Exact Predictive Densities for Linear Models with ARCH Disturbances. *Journal of Econometrics*, **40**, 63-86.

[99] Geweke, J. (1991). Generic, Algorithmic Approaches to Monte Carlo Integration in Bayesian Inference. *Contemporary Mathematics*, **115**, 117-135.

[100] Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *Bayesian Statistics 4* (J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford: Oxford University Press, 169-193.

[101] Geweke, J. and Zhou, G. (1996). Measuring the Pricing Error of the Arbitrage Pricing Theory. *The Review of Financial Studies*, **9**, 557-587.

[102] Geyer, C.J. (1991). Markov Chain Monte Carlo Maximum Likelihood. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, 156-163.

[103] Geyer, C.J. and Thompson, E.A. (1995). Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. *Journal of the American Statistical Association*, **90**, 909-920.

[104] Giakoumatos, S.G., Dellaportas, P. and Politis, D.N. (1999). Bayesian analysis of some multivariate time-varying volatility models. *Technical Report*, no. 71, Department of Statistics, Athens University of Economics and Business, Athens, Greece.

[105] Giakoumatos, S.G., Dellaportas, P. and Politis, D.N. (1998). Bayesian Analysis of the Unobserved ARCH Model Using Auxiliary Variable Sampling. *Technical Report*, no. 43, Department of Statistics, Athens University of Economics and Business, Athens, Greece.

[106] Giakoumatos, S.G., Vrontos, I.D., Dellaportas, P. and Politis, D.N. (1999). A Markov Chain Monte Carlo Convergence Diagnostic Using Subsampling. *Journal of Computational and Graphical Statistics*, **8**, 431-451.

[107] Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

[108] Gilks, W.R., Thomas, A. and Spiegelhalter, D.J. (1994). A Language and Program for Complex Bayesian Modelling. *The Statistician*, **43**, 169-177.

[109] Giovannini, A. and Jorion, P. (1989). The time variation of risk and return in the foreign exchange and stock markets. *The Journal of Finance*, **44**, 307-325.

[110] Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, **48**, 1779-1801.

[111] Gourieroux, C. (1997). *ARCH Models and Financial Applications.* New York: Springer-Verlag.

[112] Gourieroux, C. and Monfort, A. (1992). Qualitative threshold ARCH models. *Journal of Econometrics*, **52**, 159-199.

[113] Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Theory. *Econometrica*, **52**, 681-700.

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

[115] Hafner, C.M. (1998). *Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility.* Heidelberg: Physica-Verlag.

[116] Hall, P., Horowitz, J.L., and Jing, B. (1995). On Blocking Rules for the Bootstrap with Dependent Data. *Biometrika,* **82**, 561–574.

[117] Harvey, A.C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, **61**, 247-264.

[118] Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, **57**, 97-109.

[119] He, C. and Terasvirta, T. (1999). Properties of moments of a family of GARCH processes. *Journal of Econometrics*, **92**, 173-192.

[120] Heidelberger, P. and Welch, P.W. (1983). Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research*, **31**, 1109-1144.

[121] Hentschel, L. (1995). All in the Family: Nesting Symmetric and Asymmetric GARCH models. *Journal of Financial Economics*, **39**, 71-104.

[122] Higdon, D.M. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of the American Statistical Association*, **93**, 585-595.

[123] Higgins, M.L. and Bera, A.K. (1992). A Class of Nonlinear ARCH models. *International Economic Review*, **33**, 137-158.

[124] Hills, S.E. and Smith, A.F.M. (1992). Parameterization issues in Bayesian inference. *Bayesian Statistics 4* (J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford: Oxford University Press, 227-246.

[125] Jacquier, E., Polson, N.G. and Rossi, P.E. (1994). Bayesian Analysis of Stochastic Volatility Models (with discussion). *Journal of Business and Economic Statistics*, **12**, 371-417.

[126] Jeffreys, H. (1961). *Theory of Probability.* London: Oxford University Press.

[127] Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, **14**, 70-86.

[128] Karanasos, M. (1999). The second moment and the autocovariance function of the squared errors of the GARCH model. *Journal of Econometrics*, **90**, 63-76.

[129] Kass, R.E. and Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**, 773-795.

[130] Kass, R.E., Tierney, L. and Kadane, J.B. (1988). Asymptotics in Bayesian computation (with discussion). *Bayesian Statistics 3* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds.), Oxford: Oxford University Press, 261-278.

[131] Kass, R.E., Carlin, B.P., Gelman, A. and Neal, R.M. (1998). Markov Chain Monte Carlo in Practice: A Roundtable Discussion. *The American Statistician*, **52**, 93-100.

[132] King, M., Sentana, E. and Wadhwani, S. (1994). Volatility and links between national stock markets. *Econometrica*, **62**, 901-933.

[133] Kloek, T. and van Dijk, H.K. (1978). Bayesian estimates of system equation parameters: an application of integration by Monte Carlo. *Econometrica*, **46**, 1-19.

[134] Kraft, D.F. and Engle, R.F. (1982). Autoregressive conditional heteroscedasticity in multiple time series models. Discussion paper 82-23, Department of Economics, University of California, San Diego, USA.

[135] Kroner, K.F. and Claessens, S. (1991). Optimal dynamic hedging portfolios and the currency composition of external debt. *Journal of International Money and Finance*, **10**, 131-148.

[136] Kroner, K.F. and Lastrapes, W.D. (1993). The impact of exchange rate volatility on international trade: reduced form estimates using the GARCH-in-mean model. *Journal of International Money and Finance*, **12**, 298-318.

[137] Kroner, K.F. and Ng, V.K. (1998). Modeling Asymmetric Comovements of Asset Returns. *The Review of Financial Studies*, **11**, 817-844.

[138] Kroner, K.F. and Sultan, J. (1991). Exchange rate volatility and time varying hedge ratios. *Pacific-Basin Capital Markets Research 11* (S.G. Rhee and R.P. Chang, eds.), North Holland, Amsterdam.

[139] Künsch, H.R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics,* **17**, 1217–1241.

[140] Kuwahara, H. and Marsh, T.A. (1992). The Pricing of Japanese Equity Warrants. *Management Science*, **38**, 1610-1641.

[141] Laux, P.A. and Ng, L.K. (1993). The sources of GARCH: empirical evidence from an intraday returns model incorporating systematic and unique risks. *Journal of International Money and Finance*, **12**, 543-560.

[142] Laud, P.W. and Ibrahim, J.G. (1995). Predictive Model Selection. *Journal of the Royal Statistical Society*, *B*, **57**, 247-262.

[143] Lee, S.W. and Hansen, B.E. (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, **10**, 29-52.

[144] Lin, W.L. (1992). Alternative estimators for factor GARCH models-A monte carlo comparison. *Journal of Applied Econometrics,* **7**, 259-279.

[145] Lindley, D.V. (1957). A statistical paradox. *Biometrika*, **44**, 187-192.

[146] Lumsdaine, R.L. (1991). Asymptotic Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) models. Manuscript, Princeton University and National Bureau of Economic Research.

[147] MacEachern, S.N., and Berliner, L.M. (1994). Subsampling the Gibbs Sampler. *The American Statistician,* **48**, 188-190.

[148] Madigan, D. and Raftery, A.E. (1994). Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window. *Journal of the American Statistical Association*, **89**, 1535-1546.

[149] Madigan, D. and York, J. (1995). Bayesian Graphical Models for Discrete Data. *International Statistical Review*, **63**, 215-232.

[150] Madigan, D., Gavrin, J. and Raftery, A.E. (1995). Eliciting Prior Information to Enhance the Predictive Performance of Bayesian Graphical Models. *Communications in Statistics, Part A-Theory and Methods*, **24**, 2271-2292.

[151] Mak, T.K. (1993). Solving Non-linear estimation equations. *Journal of Royal Statistical Society*, B, **55**, 945-955.

[152] Mak, T.K., Wong, H. and Li, W.K. (1997). Estimation of nonlinear time series with conditional heteroscedastic variances by iteratively weighted least squares. *Journal of Computational Statistics and Data Analysis*, **24**, 169-178.

[153] Mandelbrot, B. (1963). The variation of certain speculative prices, *Journal of Business*, **36**, 394-419.

[154] Marinari, E. and Parisi, G. (1992). Simulated Tempering: a new Monte Carlo Scheme. *Europhysics Letters*, **19**, 451-458.

[155] Mark, N. (1988). Time-varying betas and risk premia in the pricing of forward foreign exchange contracts, *Journal of Financial Economics*, **22**, 335-354.

[156] Metropolis, N., Rosenbluth, A., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of State Calculations by Fast Computing Machine. *The Journal of Chemical Physics*, **21**, 1087-1092.

[157] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability.* London: Springer-Verlag.

[158] Mira, A. (1999). On Metropolis-Hastings algorithms with delayed rejection. *Technical Report*, Univerita dell'Insubria, Varese, Italy.

[159] Mira, A. and Tierney, L. (1998). On the use of auxiliary variables in Markov chain Monte Carlo sampling. *Scandinavian Journal of Statistics,* to appear.

[160] Muller, P. and Pole, A. (1999). Monte Carlo Posterior Integration in GARCH models. *Sankhya*, **62**.

[161] Naylor, J.C. and Smith, A.F.M. (1982). Applications of a Method for the Efficient Computation of Posterior Distributions. *Applied Statistics,* **31**, 214-225.

[162] Nelson, D.B. (1990). Stationarity and persistencein the GARCH(1,1) model. *Econometric Theory*, **6**, 318-334.

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

[164] Nelson, D.B. and Cao, C.Q. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics*, **10**, 229-235.

[165] Nicholls, D.F. and Quinn, B.G. (1982). *Random Coefficient Autoregressive Models: An Introduction*. New York: Springer Verlag.

[166] Ng, L. (1991). Tests of the CAPM with Time-Varying Covariances: A Multivariate GARCH Approach. *The Journal of Finance*, **46**, 1507-1521.

[167] Ng, V., Engle, R.F. and Rothschild, M. (1992). A multi-dynamic-factor model for stock returns. *Journal of Econometrics,* **52**, 245-266.

[168] Nobile, A. and Green, P. J. (1997). Bayesian Analysis of Factorial Experiments by Mixture Modelling. *Biometrika*, **87**.

[169] Noh, J., Engle, R.F. and Kane, A. (1994). Forecasting volatility and option prices of the S&500 index. *Journal of Derivatives,* 17-30.

[170] Ntzoufras, I. (1999). Aspects of Bayesian model and variable selection using MCMC. Phd thesis, Athens University of Economics and Business, Department of Statistics, Greece, ISBN: 960-7929-26-8.

[171] Pagan, A.R. and Schwert, G.W. (1990). Alternative Models for Conditional Stock Volatility. *Journal of Econometrics,* **45**, 267-290.

[172] Pantula, S.G. (1986). Modelling the persistence of conditional variances: a comment. *Econometric Reviews*, **5**, 71-74.

[173] Peskun, P.H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika*, **60**, 607-612.

[174] Pitt, M.K. and Shephard, N. (1999). Time varying covariances: a factor stochastic volatility approach (with discussion). *Bayesian Statistics 6,* (J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford: Oxford University Press, 547-570.

[175] Polasek, W. and Muller, P. (1995). Gibbs Sampling for ARCH models in Finance. *Proceedings of MODA 4* (C.P. Kitsos and W.G. Muller, eds.), Heidelberg, Physica Verlag.

[176] Politis, D.N. and Romano, J.P. (1993). On the Sample Variance of Linear Statistics Derived from Mixing Sequences. *Stochastic Processes and their Applications,* **45**, 155-167.

[177] Politis, D.N. and Romano, J.P. (1994). Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions. *The Annals of Statistics,* **22**, 2031–2050.

[178] Politis, D.N., Romano, J.P. and Wolf, M. (1997). Subsampling for Heteroskedastic Time Series. *Journal of Econometrics,* **81**, 281-318.

[179] Polson, N.G. (1996). Convergence of Markov Chain Monte Carlo Algorithms. *Bayesian Statistics 5,* (J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith, eds), Oxford, Oxford University Press, 297-322.

[180] Poon, S. and Taylor, S.J. (1992). Stock returns and volatility: an empirical study of the UK stock market. *Journal of Banking and Finance*, **16**, 37-59.

[181] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992). *Numerical Recipes in FORTRAN: The art of Scientific Computing*, 2nd edition, New York: Cambridge University Press.

[182] Raftery, A.E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika*, **83**, 251-266.

[183] Raftery, A.E. and Lewis, S. (1992). How Many Iterations in the Gibbs Sampler?. *Bayesian Statistics 4 (*J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford: Oxford University Press, 763-773.

[184] Raftery, A.E., Madigan, D. and Hoeting, J.A. (1997). Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association*, **92**, 179-191.

[185] Raftery, A.E., Madigan, D. and Volinsky, C.T. (1994). Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance. *Technical Report*, Department of Statistics, University of Washington, Seattle, USA.

[186] Richardson, S. and Green, P.J. (1997). On Bayesian Analysis of Mixtures with an Unknown Number of Components (with discussion). *Journal of the Royal Statistical Society, B*, **59**, 731-792.

[187] Ripley, B.D. (1987). *Stochastic Simulation*. Chichester: Wiley.

[188] Robert, C.P. and Mengersen, K.L. (1998). MCMC convergence Diagnostics: A "Reviewww". *Bayesian Statistics 6* (J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford: Oxford University Press.

[189] Roberts, G.O. and Rosenthal, J. S. (1997). Convergence of slice sampler Markov chains. *Journal of Royal Statistical Society, A*, **61**, 643-660.

[190] Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition. *Proceedings of National Academy of Science,* **42**, 43–47.

[191] Rosenthal, J.S. (1993). Rates of Convergence for Data Augmentation on the Finite Sample Space. *Annals of Applied Probability,* **3**, 819-839.

[192] Schervish, M.J. and Carlin, B.P. (1992). On the Convergence of Successive Substitution Sampling. *Journal of Computational and Graphical Statistics,* **1**, 111-127.

[193] Schwert, G.W. and Seguin, P.J. (1990). Heteroskedasticity in Stock Returns. *Journal of Finance,* **45**, 1129-1155.

[194] Sentana, E. (1995). Quadratic ARCH Models. *Review of Economic Studies*, **62**, 639-661.

[195] Sharma, J.L., Mougoue, M. and Kamath, R. (1996). Heteroscedasticity in stock market indicator return data: volume versus GARCH effects. *Applied Financial Economics*, **6**, 337-342.

[196] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. *Time Series Models in econometrics, finance and other fields*, (D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, eds.), London: Chapman & Hall, pp. 1-67.

[197] Smith, A.F.M. (1991). Bayesian computational methods. *Phil. Trans. Roy. Soc. London A*, **337**, 369-386.

[198] Smith, A.F.M. and Gelfand, A.E. (1992). Bayesian Statistics Without Tears: A Sampling - Resampling Perspective. *The American Statistician*, **46**, 84-88.

[199] Smith, A.F.M. and Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, B*, **55**, 3-24.

[200] Smith, A.F.M., Skene, A.M., Shaw, J.E.H., Naylor, J.C. and Dransfield, M. (1985). The implementation of the Bayesian paradigm. *Comm. Statist. Theory and Methods*, **14**, 1079-1109.

[201] Spiegelhalter, D.J., Thomas, A., Best, N. and Gilks, W.R. (1995a). BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50. *Technical Report*, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.

[202] Spiegelhalter, D.J., Thomas, A., Best, N. and Gilks, W.R. (1995b). BUGS: Examples, Version 0.50. *Technical Report*, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.

[203] Stewart, L. (1983). Bayesian analysis using Monte Carlo integration, a powerful methodology for handling some difficult problems. *The Statistician*, **32**, 195-200.

[204] Stewart, L. (1987). Hierarchical Bayesian analysis using Monte Carlo integration: computing posterior distributions when there are many models. *The Statistician*, **36**, 211-219.

[205] Stewart, L. and Davis, W.W. (1986). Bayesian posterior distributions over sets of possible models with inferences computed by Monte Carlo integration. *The Statistician*, **35**, 175-182.

[206] Swendsen, R.H. and Wang, J.S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, **58**, 86-88.

[207] Taylor, S.J. (1986). *Modelling Financial Time Series*, Chichester: Wiley.

[208] Taylor, S.J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, **4**, 183-204.

[209] Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions (with discussion). *The Annals of Statistics*, **22**, 1701-1762.

[210] Tierney, L. and Kadane, J.B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, **81**, 82-86.

[211] Tierney, L., Kass, R.E. and Kadane, J.B. (1989a). Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions. *Journal of the American Statistical Association*, **84**, 710-716.

[212] Tierney, L., Kass, R.E. and Kadane, J.B. (1989b). Approximate marginal densities of nonlinear functions. *Biometrika*, **76**, 425-433.

[213] Tierney, L. and Mira, A. (1999). Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine*, **18**, 2507-2515.

[214] Troughton, P.T. and Godsill, S.J. (1997). A Reversible Jump Sampler for Autoregressive Time Series, Employing Full Conditionals to Achieve Efficient Model Space Moves. *Technical Report*, Department of Engineering, University of Cambridge, Cambridge, UK.

[215] Tsay, R.S. (1987). Conditional Heteroscedastic Time Series Models. *Journal of the American Statistical Association*, **82**, 590-604.

[216] Turtle, H., Buse, A. and Korkie, B. (1994). Tests of Conditional Asset Pricing with Time-Varying Moments and Risk Prices. *Journal of Financial and Quantitative Analysis*, **29**, 15-29.

[217] Yue, H. and Chan, K.S. (1996). Asymptotic Efficiency of the Sample Mean in MCMC Schemes. *Journal of the Royal Statistical Society,* B, **58**, 525-539.

[218] van Dijk, H.K. and Kloek, T. (1983). Monte Carlo analysis of skew posterior distributions: An illustrative econometric example. *The Statistician*, **32**, 216-223.

[219] van Dijk, H.K. and Kloek, T. (1985). Experiments with some alternatives for simple importance sampling in Monte Carlo integration (with discussion). *Bayesian Statistics 2* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds.), Amsterdam: North-Holland, 511-530.

[220] van Dijk, H.K., Hop, J.P. and Louter, A.S. (1987). An algorithm for the computation of posterior moments and densities using simple importance sampling. *The Statistician*, **36**, 83-90.

[221] Volinsky, C.T., Madigan, D., Raftery, A.E. and Kronmal, R.A. (1997). Bayesian Model Averaging in Proportional Hazard Models: Assessing the Risk of a Stroke. *Applied Statistics*, **46**, 433-448.

[222] Vrontos, I.D. (1997). Bayesian Autoregressive Conditional Heteroscedasticity Models, Msc thesis, Department of Statistics, Athens University of Economic and Business, Athens, Greece, ISBN: 960-7929-01-2.

[223] Vrontos, I.D., Dellaportas, P. and Politis, D.N. (1999). Analysis of some multivariate ARCH and GARCH models. *Technical Report*, no. 86, Department of Statistics, Athens University of Economics and Business, Athens, Greece.

[224] Vrontos, I.D., Dellaportas, P. and Politis, D.N. (2000). Full Bayesian Inference for GARCH and EGARCH Models. *Journal of Business and Economic Statistics*, **18**, 187-198.

[225] Vrontos, I.D., Giakoumatos, S.G., Dellaportas, P. and Politis, D.N. (2000). An application of three bivariate time varying volatility models. *Applied Stochastic Models in Business and Industry*, to appear.

[226] Watson, M.W. and Engle, R.F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics,* **23**, 385-400.

[227] Weiss, A.A. (1986). Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory*, **2**, 107-131.

[228] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1-25.

[229] Wong, H. and Li, W.K. (1997). On a multivariate conditional heteroscedastic model. *Biometrika*, **84**, 111-123.

[230] Zakoian, J. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, **18**, 931-955.