Efficient Bayesian estimation and combination of GARCH-type models

Ardia, David and Hoogerheide, Lennart F.

aeris CAPITAL AG Switzerland, Econometric and Tinbergen Institutes, Erasmus University Rotterdam, the Netherlands

8 February 2010

Efficient Bayesian estimation and combination of GARCH-type models*

David Ardia[†]   Lennart F. Hoogerheide[‡]

This version: February 8, 2010

Abstract

This chapter proposes an up-to-date review of estimation strategies available for the Bayesian inference of GARCH-type models. The emphasis is put on a novel efficient procedure named AdMitIS. The methodology automatically constructs a mixture of Student-t distributions as an approximation to the posterior density of the model parameters. This density is then used in importance sampling for model estimation, model selection and model combination. The procedure is fully automatic which avoids difficult and time consuming tuning of MCMC strategies. The AdMitIS methodology is illustrated with an empirical application to S&P index log-returns where non-nested GARCH-type models are estimated and combined to predict the distribution of next-day ahead log-returns.

Keywords: GARCH, Bayesian inference, MCMC, marginal likelihood, Bayesian model averaging; adaptive mixture of Student-t distributions, importance sampling.

1   Introduction

Volatility forecasting plays an essential role in empirical finance, financial risk management and derivative pricing. Research on modeling volatility dynamics using time series models has been active since the creation of the original ARCH model by Engle (1982). From there, multiple extensions of the standard ARCH scedastic function have been proposed in order to reproduce additional stylized facts observed in financial markets. These so-called *GARCH-type models* recognize that there may be important nonlinearities, asymmetries, and long memory properties in the volatility process; see Bollerslev *et al.* (1992), Bollerslev *et al.* (1994) and Engle (2004) for a review. Well

---

[†]aeris CAPITAL AG Switzerland; `da@aeris-capital.com` (corresponding author). Disclaimer: The views expressed in this chapter are the sole responsibility of the author and do not necessarily reflect those of aeris CAPITAL AG or any of its affiliates.

[‡]Tinbergen Institute and Econometric Institute, Erasmus University Rotterdam, The Netherlands.

known extensions are the exponential GARCH model by Nelson (1991), the GJR model by Glosten *et al.* (1993) or the TGARCH model of Zakoian (1994) which account for the asymmetric relation between stock returns and changes in variance (see Black, 1976). In parallel to the development of alternative scedastic functions, numerous types of disturbances have been used. The Gaussian and Student-t distributions are common choices while more sophisticated parameterizations such as the skewed Student-t or the mixture of Gaussian distributions allow to model skewness and fat tails in the conditional distribution of returns (see Ausin and Galeano, 2007). More recently, interest has focused on regime-switching GARCH models. In this framework, the GARCH parameters are functions of an unobservable state variable and can change over time. These processes provide an explanation of the high persistence in volatility observed with single-regime GARCH models (see, e.g., Lamoureux and Lastrapes, 1990). Furthermore, as shown by Dueker (1997), Klaassen (2002) and Marcucci (2005) for instance, the regime-switching GARCH models allow for a quick change in the volatility level which can lead to significant improvements in risk forecasts.

Until recently GARCH-type models have mainly been estimated using the classical maximum likelihood technique. However, the Bayesian approach offers an attractive alternative. It enables small sample results, probabilistic statements on nonlinear functions of the model parameters, selection and combination of non-nested models. Due to these numerous advantages, the study of GARCH-type models from a Bayesian viewpoint can be considered very promising.

This chapter proposes an up-to-date review of simulation techniques used in the literature to perform the Bayesian estimation of GARCH-type models. The emphasis is put on a novel approach referred to as AdMitIS. The algorithm constructs a mixture of Student-t distributions to perform an efficient estimation of models via importance sampling. The methodology is fully automatic and has proved to require less computing time for precise estimation results than several well-known alternatives. We describe in some details the steps of the algorithm. In an empirical application to S&P500 index log-returns, we show how it can be used to estimate and combine distribution forecasts of non-nested GARCH-type models.

The chapter proceeds as follows. In Section 2, we survey the existing simulation techniques available for Bayesian GARCH-type models. In Section 3, we present the AdMitIS algorithm. In Section 4, we provide an empirical illustration. Section 5 concludes.

2 Bayesian estimation of GARCH-type models

The maximum likelihood (ML) estimation technique is the most commonly used scheme of inference for GARCH-type models. ML is easy to understand and to implement. Nowadays, several econometric softwares contain a GARCH toolbox and ML estimation of standard GARCH models takes less than one second on a modern computer. This is highly desirable for automated trading strategies, for instance, where several models are fitted on thousands of data many times per day. From a theoretical viewpoint, ML estimators benefit from being asymptotically optimal under certain conditions (see Bollerslev *et al.*, 1994; Lee and Hansen, 1994).

Despite these appealing features, we face practical difficulties when dealing with the ML estimation of GARCH-type models. First, the maximization of the likelihood function must be achieved via a constrained optimization technique since some (or all) model parameters must be positive to ensure a positive conditional variance. It is also common to require that the covariance stationarity condition holds. This leads to complicated non-linear inequality constraints which render the optimization procedure cumbersome. Moreover, the convergence of the optimization is hard to achieve if the true parameter values are close to the boundary of the parameter space and if the GARCH process is nearly non-stationary. Optimization results are often sensitive to the choice of starting values. Second, in standard applications of GARCH models, the interest usually does not center directly on the model parameters but on possibly complicated nonlinear functions of the parameters. For instance, a trader might be interested in the unconditional variance implied by a GARCH model, which is a (highly) non-linear function of the model parameters. In order to assess the uncertainty of such a quantity, classical inference involves tedious delta methods, simulation from the asymptotic Gaussian approximation of the parameter estimates or the time-consuming bootstrap methodology. Third, the conditions for the optimal asymptotic properties of ML estimators to hold are fairly difficult to prove, while often assumed to hold in practice. Moreover, since GARCH-type models are highly non-linear, the asymptotic argument would require a very large number of data to hold. This is obviously not always the case in practice. Finally, in the case of GARCH with mixture disturbances or regime-switching GARCH models, testing for the number of mixture components or the number of regimes is not possible within the classical framework due to the violation of regularity conditions (see Frühwirth-Schnatter, 2006, Section 4.4).

Fortunately, those difficulties disappear when Bayesian methods are used. First, any constraints on the model parameters can be incorporated in the modeling through appropriate prior specifications. Second, appropriate Markov chain Monte Carlo (MCMC) procedures can explore the joint posterior distribution of the model parameters. These techniques avoid local maxima (i.e., non convergence or convergence to the wrong values) encountered via ML estimation of sophisticated GARCH-type models. Third, exact distributions of nonlinear functions of the model parameters can be obtained at low cost by simulating from the joint posterior distribution. Fourth, the issue of determining the number of mixture components in mixtures disturbances or the number of regimes in regime-switching GARCH model can be addressed by means of marginal likelihoods and Bayes factors. Marginal likelihoods can also be used for model selection and model combination of non-nested GARCH-type models. The latter case is especially interesting in the context of financial risk management since it allows to integrate out model uncertainty, thus accounting for model risk in the forecasts.

The choice of the sampling algorithm is the first issue when dealing with MCMC methods and it depends on the nature of the problem under study. For a general introduction to MCMC methods, see Chapter 2 of this volume. In the case of GARCH-type models, due to the recursive nature of the conditional variance, the joint posterior and the full conditional densities are of unknown forms, whatever assumptions are made on the scedastic function or the model disturbances. That is, there exists no conjugate prior under which the parameters' (conditional) posterior distributions fall within a known class. Therefore, we cannot use the simple Gibbs sampler and need

more elaborated procedures. Hereafter, we shortly present an up-to-date review of the different techniques available to perform the Bayesian estimation of GARCH-type models and discuss their advantages and drawbacks. For a numerical comparison of some of these approaches in the context of GARCH-type models, we refer the reader to Asai (2006) and Ardia *et al.* (2009a).

Griddy-Gibbs The Griddy-Gibbs sampler of Ritter and Tanner (1992) is a variant of the Gibbs sampler where each parameter is updated by inversion from the full conditional distribution computed by a deterministic integration rule. The algorithm is intuitive and very simple to implement. However, the procedure is extremely time consuming due to the numerical integration steps required at each sweep of the sampler. While this can still be acceptable for simple GARCH specifications, this becomes a real burden for sophisticated and highly parametrized GARCH-type models. Moreover, for computational efficiency, we must limit the range where the probability mass is computed so that the prior density has to be somewhat informative. In our viewpoint, the Griddy-Gibbs is attractive for its simplicity and remains a useful tool for "one-shot" studies. It is however not relevant in real-world applications, where the models need to be estimated many times on a large number of times series. The Griddy-Gibbs approach is used in the context of GARCH-type models by Bauwens and Lubrano (1998), Bauwens *et al.* (2000), Bauwens and Lubrano (2002), Wago (2004), Tsay (2005, Chapter 10), Ausin and Galeano (2007), Bauwens and Rombouts (2007) and Bauwens *et al.* (in press).

Importance sampling Importance sampling (IS), due to Hammersley and Handscomb (1965), was introduced in econometrics and statistics by Kloek and van Dijk (1978). The importance sampling approach relies on an *importance* density which approximates the posterior density of the model parameters. Draws are generated from this importance density and weighted, with higher weights given to draws for which the importance density is relatively small compared with the posterior. Then, quantities of interest of the posterior distribution (or predictive distribution) are estimated by weighted averages of (functions of) draws. This methodology leads to a fast estimation since it only requires generating and weighting draws. Moreover, it generates uncorrelated draws so that the whole posterior sample can be used for Bayesian inference (i.e., no burn-in or "thinning" is necessary), and so that the precision of the estimators is easily assessed. However, the key issue for applicability and efficiency is the choice of the importance density. Finding this can be a bit of an art, especially if the posterior density is asymmetric or multi-modal. For instance, in the case of a GARCH model with mixture disturbances, the posterior distribution is multimodal. In this case, a (standard) unimodal importance density may imply that some draws have huge weights leading to large inefficiencies in the estimation, or worse, that relevant parts of the parameters space are completely "missed". The importance distribution should be close to the posterior distribution and it is especially important that the tails of the importance should not be thinner than those of the posterior. Bayesian estimation of GARCH-type models using importance sampling is proposed by Geweke (1988), Geweke (1989b) and Kleibergen and van Dijk (1993).

Metropolis-Hastings The Metropolis-Hastings (MH) algorithm was introduced by Metropolis *et al.* (1953) and generalized by Hastings (1970). The approach constructs a Markov chain by generating draws from a *candidate* density; the candidate draw is then accepted (or rejected) based on an acceptance probability. If the candidate is accepted, the chain moves to the new value, otherwise the chain stays in the current state. After a *burn-in* period, which is required to make the influence of initial values negligible, draws from the Markov chain are considered as (correlated) draws from the joint posterior distribution of interest. Two variants of the MH approach are most common: (i) the independence chain MH and (ii) the random-walk MH. In the former case, candidate draws are generated from an unconditional candidate distribution whereas in the latter draws are generated from a distribution conditional on (and around) the current value of the chain. In both variants the candidate distribution must be tuned to achieve a reasonable acceptance rate and to explore sufficiently the domain of the posterior distribution. This tuning process requires preliminary runs and some knowledge of MCMC techniques from the user. Hence, the method is not automatic which is not a desirable property. In addition, for interpreted languages such as R (R Development Core Team, 2009), MATLAB$^©$ (The MathWorks Inc., 2009) or Mathematica$^©$ (Wolfram Research, 2009), the MH algorithm can be significantly slower than the importance sampling strategy. This is so because the probability of acceptance of the new draw depends on the current state of the Markov chain; hence, the algorithm cannot be vectorized. Finally, the MH algorithm creates a sequence of correlated draws from the posterior distribution. Therefore, robust techniques must be used to assess the precision of the estimators and more draws are required to achieve the same degree of (numerical) precision as the importance sampling approach. The MH algorithm is used in the context of GARCH-type models by Müller and Pole (1998), Vrontos *et al.* (2000), Tsay (2005, Chapter 10), Gerlach and Chen (2006), Miazhynskaia and Dorffner (2006), Aussenegg and Miazhynskaia (2006), Miazhynskaia *et al.* (2006), Chen *et al.* (2008), Chen and Gerlach (2008) and Chen *et al.* (2009). Another example is the chapter by Ausin and Lopes in this book.

An interesting strategy has been proposed to automate the MH algorithm and improve its efficiency. This approach is proposed by Nakatsuma (1998, 2000) and consists of a MH algorithm where the proposal distributions are constructed from auxiliary ARMA processes on the squared observations. In addition to be fully automatic and more efficient than naive MH approaches, the methodology can be extended to regime-switching GARCH models; see Ardia (2008b, Chapter 7) and Ardia (2009). Note however that the construction of the proposal distributions highly depends on the form of the scedastic function and is not applicable to all GARCH-type models. Moreover, the algorithm requires filtering steps which significantly increase the computational burden for highly parametrized models. The approach of Nakatsuma (1998, 2000) is used in Nakatsuma and Tsurumi (1999), Kaufmann and Frühwirth-Schnatter (2002), Kaufmann and Scheicher (2006), Rachev *et al.* (2008, Chapter 11), Henneke *et al.* (2009), Ardia (2008b) and Ardia (2009). The algorithm is implemented in the R package bayesGARCH (Ardia, 2008a; Ardia and Hoogerheide, 2009) for the GARCH(1,1) model with Student-t disturbances.

AdMitIS Recently, Ardia *et al.* (2009a) and Ardia *et al.* (2009c) relied on a special case of the adaptive approach proposed by Hoogerheide (2006); Hoogerheide *et al.*

(2007) for performing the Bayesian estimation of GARCH-type models. The methodology, named AdMitIS for importance sampling with adaptive mixture of Student-t distributions, consists of two steps. First, the algorithm fits adaptively a mixture of Student-t distributions to the kernel of the posterior density. Then, importance sampling is used to obtain quantities of interest for the target posterior distribution, using the fitted mixture as the importance density. The estimation procedure is fully automatic. Moreover, Ardia *et al.* (2009a) compared the methodology to standard cases of importance sampling and MH algorithm using a naive candidate and to the Griddy-Gibbs approach. Overall, they demonstrate the superiority of the AdMitIS approach both in terms of efficiency and reliability.

Due to its flexibility, the adaptive algorithm is able to provide a suitable importance distribution for non-elliptical, possibly multi-modal, posterior distributions. But it is not only useful for sophisticated GARCH-type models. For large samples and/or simple scedastic specifications, the posterior distribution of the model parameters is likely to be roughly elliptical. In these cases, the adaptive approach will stop with one or two mixture components, leading to a simple unimodal symmetric importance density. The adaptive fitting together with the importance sampling estimation is achieved fairly quickly. Therefore, models can be re-estimated many times on many data sets without practical issues. This is clearly an appealing aspect for practitioners. Moreover, Ardia *et al.* (2009c) show that the approach allows an efficient and reliable estimation of marginal likelihood, which lies at the heart of model selection and model combination. The approach can therefore be used to estimate, select and combine GARCH-type models in a simple, quick and efficient fashion.

3   Adaptive mixture of Student-t method

The adaptive mixture of Student-t distributions (AdMit) procedure has been developed by Hoogerheide (2006); Hoogerheide *et al.* (2007); see also Hoogerheide and van Dijk (2008). The AdMit methodology consists of the construction of a mixture of Student-t distributions which approximates a target distribution of interest. The fitting procedure relies only on a kernel of the target density, so that the normalizing constant is not required. In a second step this approximation is used as an importance function in importance sampling (AdMitIS) or as a candidate density in the independence chain MH algorithm (AdMitMH). Both AdMitIS and AdMitMH strategies have been implemented in the R package AdMit (Ardia *et al.*, 2008) which is freely available from the Comprehensive R Archive Network (CRAN) at `http://cran.r-project.org/package=AdMit`. The usage of the package with empirical examples is discussed in Ardia *et al.* (2009a) and Ardia *et al.* (2009b).

Hoogerheide *et al.* (2007) mention several reasons why mixtures of Student-t distributions are natural importance or candidate distributions. Indeed, a Student-t mixture

- provides an accurate approximation to a wide variety of target distributions, with substantial skewness and high kurtosis; it can deal with multi-modality and with non-elliptical shapes.

6

- can be constructed in a quick, iterative procedure and is easy to sample from.

- has fatter tails than the Gaussian distribution (especially if one specifies Student-t components with few degrees of freedom). Therefore, the risk is small that the tails of the importance or candidate are thinner than those of the target distribution.

Moreover, from a purely theoretical framework, mixture of Student-t densities can approximate any density function to arbitrary accuracy under certain conditions (Zeevi and Meir, 1997).

Because of its superiority compared to several alternative methods, both in terms of efficiency and reliability of posterior estimation (see Hoogerheide, 2006; Hoogerheide *et al.*, 2007; Ardia *et al.*, 2009a) and of marginal likelihood estimation (see Ardia *et al.*, 2009c), we will make use of the AdMitIS approach in the sequel.

For a given GARCH-type model, we denote by $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$ the vector of parameters, $p(\boldsymbol{\theta} \mid \mathbf{y})$ the posterior density of $\boldsymbol{\theta}$ and $\mathbf{y} \doteq (y_1, \ldots, y_T)'$ the vector of log-returns. The joint posterior density of $\boldsymbol{\theta}$ is then obtained by Bayes' theorem as

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int_\Theta p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}} \,, \tag{1}$$

where $p(\mathbf{y} \mid \boldsymbol{\theta})$ is the joint density of $\mathbf{y}$ given $\boldsymbol{\theta}$, i.e., the likelihood, and $p(\boldsymbol{\theta})$ is the exact prior density of $\boldsymbol{\theta}$, i.e., not merely a prior kernel. In expression (1), we define $k(\boldsymbol{\theta}) \doteq p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$ as the kernel function of the joint posterior and

$$p(\mathbf{y}) \doteq \int_\Theta k(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \tag{2}$$

as the marginal likelihood. It is clear that the marginal likelihood is equal to the normalizing constant of the joint posterior density (1). As the key ingredient in Bayes factors, the marginal likelihood lies at the heart of model selection and model combination in Bayesian statistics (see, e.g., Kass and Raftery (1995) and Section 5.1 of Chapter 1 in this book).

The AdMit methodology constructs a mixture of Student-t distributions in order to approximate the posterior density $p(\boldsymbol{\theta} \mid \mathbf{y})$. The density of a mixture of Student-t distributions can be written as

$$q(\boldsymbol{\theta}) = \sum_{h=1}^{H} \eta_h \, t_d(\boldsymbol{\theta} \mid \boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h, \nu) \,,$$

where $\{\eta_h\}$ are the mixing weights of the Student-t components ($0 \leqslant \eta_h \leqslant 1$ and $\sum_{h=1}^{H} \eta_h = 1$) and $t_d(\boldsymbol{\theta} \mid \boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h, \nu)$ is a d-dimensional Student-t density with mode vector $\boldsymbol{\mu}_h$, scale matrix $\boldsymbol{\Sigma}_h$, and ν degrees of freedom. The adaptive mixture approach determines H, $\{\eta_h\}$, $\{\boldsymbol{\mu}_h\}$ and $\{\boldsymbol{\Sigma}_h\}$ based on the kernel $k(\boldsymbol{\theta})$. It consists of the following steps:

Step 0 Compute the mode $\boldsymbol{\mu}_1$ and scale $\boldsymbol{\Sigma}_1$ of the first Student-t distribution in the mixture as $\boldsymbol{\mu}_1 = \arg\max_{\boldsymbol{\theta} \in \Theta} \log k(\boldsymbol{\theta})$, the mode of the log kernel function, and $\boldsymbol{\Sigma}_1$ as minus the Hessian of $\log k(\boldsymbol{\theta})$ evaluated at its mode $\boldsymbol{\mu}_1$. Then draw a set of N points $\{\boldsymbol{\theta}^{[i]}\}$ from this first stage candidate density $q(\boldsymbol{\theta}) = t_d(\boldsymbol{\theta} \mid \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu)$, with small ν to allow for fat tails. N is typically a very large number, e.g., $N = 100\,000$.

After that add iteratively components to the mixture by performing the following steps:

Step 1 Compute the importance sampling weights

$$w(\boldsymbol{\theta}^{[i]}) \doteq \frac{k(\boldsymbol{\theta}^{[i]})}{q(\boldsymbol{\theta}^{[i]})} \tag{3}$$

for $i = 1, \ldots, N$. In order to determine the number of components H of the mixture, we make use of a simple diagnostic criterion: the *coefficient of variation*, i.e., the standard deviation divided by the mean, of the importance sampling weights $\{w(\boldsymbol{\theta}^{[i]})\}$. If the relative change in the coefficient of variation of the importance sampling weights caused by adding one new Student-t component to the candidate mixture is small, e.g., less than 10%, then the algorithm stops and the current mixture $q(\boldsymbol{\theta})$ is the approximation. Otherwise, the algorithm goes to step 2.

Step 2 Add another Student-t distribution with density $t_d(\boldsymbol{\theta} \,|\, \boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h, \nu)$ to the current mixture. The new component is based on the ratio of the previous mixture of Student-t densities and the target density kernel $k(\boldsymbol{\theta})$. It is located where this ratio is *relatively* high, which does not depend on the normalizing constant of the target density. Since for most GARCH-type specifications the region of integration Θ is bounded, it may occur that $w(\boldsymbol{\theta})$ attains its maximum at a boundary of Θ. In this case, minus the inverse Hessian of $\log w(\boldsymbol{\theta})$ evaluated at its mode (which would otherwise provide $\boldsymbol{\mu}_h$ and $\boldsymbol{\Sigma}_h$) may be a very poor choice; in fact this Hessian may not even be positive definite. Therefore, in order to avoid any numerical problem, $\boldsymbol{\mu}_h$ and $\boldsymbol{\Sigma}_h$ are obtained as the estimated mean and covariance based on a subset of draws corresponding to a certain percentage of largest weights. More precisely, $\boldsymbol{\mu}_h$ and $\boldsymbol{\Sigma}_h$ are obtained as

$$\boldsymbol{\mu}_h = \sum_{j \in J_c} \frac{w(\boldsymbol{\theta}^{[j]})}{\sum_{j \in J_c} w(\boldsymbol{\theta}^{[j]})} \boldsymbol{\theta}^{[j]},$$

$$\boldsymbol{\Sigma}_h = \sum_{j \in J_c} \frac{w(\boldsymbol{\theta}^{[j]})}{\sum_{j \in J_c} w(\boldsymbol{\theta}^{[j]})} (\boldsymbol{\theta}^{[j]} - \boldsymbol{\mu}_h)(\boldsymbol{\theta}^{[j]} - \boldsymbol{\mu}_h)',$$

where J_c denotes the set of indices corresponding to the c percents of the largest weights using the sample $\{w(\boldsymbol{\theta}^{[j]})\}$ from $q(\boldsymbol{\theta})$ we already have. Since our aim is to detect regions with relatively too little candidate probability mass as compared with the target distribution (e.g., a distant mode), the percentage c is typically a low value, e.g., 5%.

Step 3 Choose the probabilities $\{\eta_h\}$ in the mixture $q(\boldsymbol{\theta}) = \sum_{h=1}^{H} \eta_h \, t_d(\boldsymbol{\theta} \,|\, \boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h, \nu)$ by minimizing the coefficient of variation of the importance sampling weights.

Step 4 Draw a sample of N points $\{\boldsymbol{\theta}^{[i]}\}$ from the new mixture of Student-t distributions $q(\boldsymbol{\theta}) = \sum_{h=1}^{H} \eta_h \, t_d(\boldsymbol{\theta} \,|\, \boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h, \nu)$ and go to Step 1.

The coefficient of variation of the importance sampling weights is a natural and intuitive measure of quality of the candidate as an approximation to the target. If the candidate and the target distributions coincide, all importance sampling weights are equal, so that the coefficient of variation is zero. For a poor candidate that not even roughly approximates the target, some importance sampling weights are huge while most are (almost) zero, so that the coefficient of variation is high. The better the candidate approximates the target, the more evenly the weight is divided among the candidate draws, and the smaller the coefficient of variation of the importance sampling weights. We refer the reader to Ardia *et al.* (2009a) for theoretical reasons justifying the coefficient of variation.

The mode and Hessian of the log kernel function (step 0), the coefficient of variation of the importance sampling weights (steps 1 and 3), and the scaled weights $w(\boldsymbol{\theta}^{[j]})/\sum_{j\in J_c} w(\boldsymbol{\theta}^{[j]})$ (step 2) do not depend on the normalizing constant of the target density, which explains why the whole AdMit procedure only requires a target density kernel (which is also a property of the MH algorithm).

Once the adaptive mixture of Student-t distributions has been fitted to the target density $p(\boldsymbol{\theta}\,|\,\mathbf{y})$ through the kernel function $k(\boldsymbol{\theta})$, the approximation $q(\boldsymbol{\theta})$ is used in importance sampling to obtain quantities of interest of the posterior $p(\boldsymbol{\theta}\,|\,\mathbf{y})$ or perform model selection and model combination. The importance sampling technique is based on the relationship

$$\mathbb{E}_p\big[g(\boldsymbol{\theta})\big] = \frac{\int_\Theta g(\boldsymbol{\theta})p(\boldsymbol{\theta}\,|\,\mathbf{y})\mathrm{d}\boldsymbol{\theta}}{\int_\Theta p(\boldsymbol{\theta}\,|\,\mathbf{y})\mathrm{d}\boldsymbol{\theta}} = \frac{\int_\Theta g(\boldsymbol{\theta})w(\boldsymbol{\theta})q(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}{\int_\Theta w(\boldsymbol{\theta})q(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} = \frac{\mathbb{E}_q\big[g(\boldsymbol{\theta})w(\boldsymbol{\theta})\big]}{\mathbb{E}_q\big[w(\boldsymbol{\theta})\big]}, \quad (4)$$

where $g(\boldsymbol{\theta})$ is a given function (integrable with respect to $p(\boldsymbol{\theta}\,|\,\mathbf{y})$), $\mathbb{E}_p$ denotes the expectation with respect to the posterior density $p(\boldsymbol{\theta}\,|\,\mathbf{y})$ and $\mathbb{E}_q$ denotes the expectation with respect to the Student-t mixture $q(\boldsymbol{\theta})$. The importance sampling estimator of $\mathbb{E}_p\big[g(\boldsymbol{\theta})\big]$ is given by

$$\hat{\mathbb{E}}_p\big[g(\boldsymbol{\theta})\big] = \frac{\sum_{l=1}^L g(\boldsymbol{\theta}^{[l]})w(\boldsymbol{\theta}^{[l]})}{\sum_{l=1}^L w(\boldsymbol{\theta}^{[l]})}, \quad (5)$$

where $\{\boldsymbol{\theta}^{[l]}\}$ is a sample of L *i.i.d.* draws from the importance density $q(\boldsymbol{\theta})$. Under certain conditions, $\hat{\mathbb{E}}_p$ is a consistent estimator of $\mathbb{E}_p$ (see Geweke, 1989a). The choice of the function $g(\boldsymbol{\theta})$ allows to obtain different quantities of interest; for instance, the posterior mean is obtained with $g(\boldsymbol{\theta}) \doteq \boldsymbol{\theta}$ and the posterior probability of a set $S \subseteq \mathbb{R}^d$ is obtained with $g(\boldsymbol{\theta}) \doteq 1_{\{\boldsymbol{\theta}\in S\}}$, where $1_{\{\cdot\}}$ denotes the indicator function.

As for any time series model, prediction is essential. The Bayesian framework allows for obtaining *predictive* densities that by construction incorporate parameter uncertainty when forecasting future observations. For instance, the one-step ahead predictive density is obtained by setting $g(\boldsymbol{\theta})$ to the density of the one-step ahead observation y_{T+1} in expression (4). Formally,

$$p(y_{T+1}\,|\,\mathbf{y}) = \int_\Theta p(y_{T+1}\,|\,\boldsymbol{\theta},\mathbf{y})p(\boldsymbol{\theta}\,|\,\mathbf{y})\mathrm{d}\boldsymbol{\theta}\,. \quad (6)$$

This quantity is easily estimated using (5) since $p(y_{T+1}\,|\,\boldsymbol{\theta},\mathbf{y})$ is known in closed-form for any GARCH-type model. For multi-step ahead forecasts, the predictive density is

of unknown form and we must rely on simulation in conjunction with the method of composition. We refer the reader to Ardia (2008b, Chapter 6) for an illustration.

The Bayesian framework is also appealing for selecting and combining models, which is achieved through the computation of the posterior model probability. Assume that we have M (possibly non-nested) GARCH-type models $\mathcal{M}_i$ with marginal likelihood $p_i(\mathbf{y})$ and prior model probability $p(\mathcal{M}_i)$. Usually, the prior model probability is equal for all models or favors parsimonious specifications. The posterior model probability of model $\mathcal{M}_i$ is then given by

$$p(\mathcal{M}_i \,|\, \mathbf{y}) = \frac{p_i(\mathbf{y})p(\mathcal{M}_i)}{\sum_{i=1}^{M} p_i(\mathbf{y})p(\mathcal{M}_i)} \,,$$

where the importance sampling approach immediately provides the estimator for the marginal likelihood $p_i(\mathbf{y})$

$$\hat{p}_i(\mathbf{y}) = \frac{1}{L} \sum_{l=1}^{L} w_i(\boldsymbol{\theta}_i^{[l]}) \,, \tag{7}$$

as shown by Kloek and van Dijk (1978) for instance. In expression (7), $w_i(\boldsymbol{\theta}_i)$ denotes the weight function (3) based on the kernel $k_i(\boldsymbol{\theta}_i)$ and the Student-t mixture density $q_i(\boldsymbol{\theta}_i)$ corresponding to model $\mathcal{M}_i$, and $\{\boldsymbol{\theta}_i^{[l]}\}$ are L i.i.d. draws generated from $q_i(\boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i \in \Theta_i \subseteq \mathbb{R}^{d_i}$. Therefore, estimator (7) is simply obtained as the sample counter-part of (2) for $\mathcal{M}_i$, which can be expressed as

$$p_i(\mathbf{y}) = \int_{\Theta_i} k_i(\boldsymbol{\theta}_i)\mathrm{d}\boldsymbol{\theta}_i = \int_{\Theta_i} \frac{k_i(\boldsymbol{\theta}_i)}{q_i(\boldsymbol{\theta}_i)} q_i(\boldsymbol{\theta}_i)\mathrm{d}\boldsymbol{\theta}_i = \int_{\Theta_i} w_i(\boldsymbol{\theta}_i)q_i(\boldsymbol{\theta}_i)\mathrm{d}\boldsymbol{\theta}_i \,.$$

For model discrimination, the model with the largest posterior probability will then be selected. The posterior probability can can also be used to produce a combination of the predictive distributions, an approach referred to as Bayesian model averaging (BMA) in the literature (see Kass and Raftery, 1995). For instance, the one-step ahead BMA predictive density is given by

$$p_{\mathrm{BMA}}(y_{T+1} \,|\, \mathbf{y}) = \sum_{i=1}^{M} p_i(y_{T+1} \,|\, \mathbf{y})p(\mathcal{M}_i \,|\, \mathbf{y}) \,, \tag{8}$$

where $p_i(y_{T+1} \,|\, \mathbf{y})$ denotes the predictive density of model $\mathcal{M}_i$, which is easily obtained using (6) with the density $p_i(y_{T+1} \,|\, \boldsymbol{\theta}_i, \mathbf{y})$ and the posterior $p_i(\boldsymbol{\theta}_i \,|\, \mathbf{y})$ corresponding to $\mathcal{M}_i$. Expression (8) is nothing else than the weighted average of the M single-model one-step ahead predictive densities, where the weights are the posterior model probabilities. This distribution of the one-step ahead observation accounts for uncertainty on both parameter values and model choice.

4   Empirical illustration

This section proposes an illustration of the AdMitIS strategy with the Bayesian estimation of two non-nested GARCH-type models. The posterior model probabilities

are estimated and used to combine the predictive densities of the one-day ahead log-returns. This case study aims at describing in a real-life example the mechanics of the AdMitIS strategy and demonstrating its effectiveness.

We apply our Bayesian estimation methods to daily observations of the S&P500 index log-returns. The sample period is from April 28, 1995, to October 27, 1997, for a total of 633 observations. The reason for this particular data window of two and a half years, which is long enough to estimate GARCH-type models, is that it ends with an extremely negative return, which makes the differences between forecasts from the two models more clearly visible. Further, for this data set the typically imposed restrictions on the models (for ensuring stationarity and positivity of the volatility process) seem to be supported by the data. The analysis of the correctness of the model restrictions falls outside the scope of this chapter. The time series has been demeaned and the nominal returns are expressed in percent. Robust autocorrelation tests do not exhibit any autocorrelation in the returns, whereas significant autocorrelation is detected for the squared log-returns, thus suggesting GARCH effects in the data.

The two models are based on two non-nested scedastic functions. For the variance dynamics, we use the parsimonious but effective GJR(1,1) (see Glosten *et al.*, 1993) and EGARCH(1,1) (see Nelson, 1991) specifications. These models are well-known in univariate GARCH modeling for their ability to reproduce the asymmetric behavior of the conditional variance observed in equity markets. For the model disturbances, we consider Student-t innovations, which allow to reproduce fat tails in the conditional distribution.

Formally, the log-returns $\{y_t\}$ can be expressed as

$$y_t = \sigma_t \, \varrho \, \varepsilon_t \quad (t = 1, \ldots, T) \,,$$

where the scedastic function σ_t^2 can be of either type:

GJR $\qquad \sigma_t^2 \doteq \omega + \alpha \, y_{t-1}^2 + \gamma \, y_{t-1}^2 \, 1_{\{y_{t-1}<0\}} + \beta \, \sigma_{t-1}^2 \quad (\omega > 0, \alpha, \gamma, \beta \geq 0)$

EGARCH $\quad \log(\sigma_t^2) \doteq \omega + \alpha \left[|\frac{y_{t-1}}{\sigma_{t-1}}| - \mathbb{E}(|\frac{y_{t-1}}{\sigma_{t-1}}|) \right] + \gamma \, \frac{y_{t-1}}{\sigma_{t-1}} + \beta \, \log(\sigma_{t-1}^2) \,,$

and where the disturbances ε_t are *i.i.d.* Student-t variates,

$$p(\varepsilon_t) \doteq \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) (\pi \nu)^{1/2}} \left(1 + \frac{\varepsilon_t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad (\nu > 2) \,.$$

The scalar ϱ is the normalizing factor $\sqrt{(\nu-2)/\nu}$ to ensure that the conditional variance is σ_t^2. For the Student-t variates y_{t-1}/σ_{t-1} that are normalized to have variance one, we have

$$\mathbb{E}(|\frac{y_{t-1}}{\sigma_{t-1}}|) = \frac{\sqrt{\nu-2} \, \Gamma(\frac{\nu-1}{2})}{\sqrt{\pi} \, \Gamma(\frac{\nu}{2})} \,.$$

Note the positivity constraints on the model parameters in order to ensure a positive conditional variance and the constraint on the degrees of freedom parameter to ensure the existence of a finite conditional variance. Moreover, we require the process to be covariance stationary, i.e., $\alpha + \gamma/2 + \beta < 1$ for the GJR(1,1) model and $|\beta| < 1$ for the EGARCH(1,1) model. Note that the model specifications are used for illustrative

purposes only; checking for possible model misspecification is beyond the scope of the present paper.

For both models we specify weakly informative, proper prior distributions on the parameters that can roughly be interpreted as the proper versions of the improper priors used by Vrontos *et al.* (2000). We use proper priors for computing posterior model probabilities and performing Bayesian model averaging. For the GJR(1,1) model, we specify a Gaussian distribution for $\log(\omega)$ with mean $\log(0.01)$ and standard deviation $\log(10)$, which amounts to a 95% prior interval for ω between 0.0001 and 1. For α, γ and β we use a uniform prior on the subspace with $\alpha > 0$, $\gamma > 0$, $\beta > 0$, $\alpha + \gamma/2 + \beta < 1$. We specify an exponential prior distribution with mean 20 for $(\nu - 2)$. For the EGARCH(1,1) model we choose a uniform prior on [-1,1] for β. For ω, α and γ we use Gaussian priors with zero mean and standard deviation 0.1. Again we specify an exponential prior distribution with mean 20 for $(\nu - 2)$.

The priors are combined with the likelihood function, leading to the posterior kernel functions. These kernels are used in the AdMit algorithm. For each model, AdMit was applied with $N = 100\,000$ draws. For both models, the algorithm led to a mixture of three Student-t distributions as the importance sampling distribution. Figure 1 illustrates the steps of the AdMit algorithm for the marginal posterior of ν in the EGARCH model. The first candidate is a Student-t distribution around the posterior mode. Second, a Student-t component is added that is located in the right tail, yielding a right-skewed candidate. This consists of a substantial improvement of the approximation to the posterior, shown in Figure 3, so that the AdMit algorithm continues. Third, a Student-t component is added that is placed in the short left tail. This is only a minor improvement of the candidate, so that the AdMit algorithm stops at three components. Notice that the AdMit algorithm provides an approximation of the joint posterior in the five-dimensional parameter space; Figure 1 displays the one-dimensional marginal candidate distribution of ν only for illustration.

The densities constructed by AdMit for each models are used to perform the Bayesian estimation of the model parameters via importance sampling (AdMitIS) using $L = 100\,000$ draws. Results are reported in Table 1 and marginal posterior densities are shown in Figures 2 and 3. We notice the strong evidence for the leverage effect in the time series ($\gamma > 0$ in GJR, $\gamma < 0$ in EGARCH), as well as conditional leptokurticity in the data with a rather small ν. The standard deviations of the parameters are quite large. We also report the numerical standard errors (NSE), i.e., the square root of the variance of the estimates that can be expected if the simulations were to be repeated with different random numbers. The RNE is the relative numerical efficiency of the estimate, i.e., the ratio between an estimate of the variance of a hypothetical estimator based on direct sampling and the importance sampling estimator's estimated variance with the same number of draws. RNE is an indicator of efficiency of the chosen importance function; if the target and importance densities coincide, RNE equals one, whereas a very poor importance density will have a RNE close to zero. Both NSE and RNE are estimated by the method given in Geweke (1989a). The numerical standard error and relative numerical efficiency indicate a reasonable degree of efficiency for the methods, which may be difficult to achieve for non-elliptical, skewed, multidimensional posterior distributions. Table 2 reports estimation results for importance sampling using a naive importance density, i.e., a Student-t distribution around the posterior mode. Note the lower RNE for all coefficients in both models, as compared
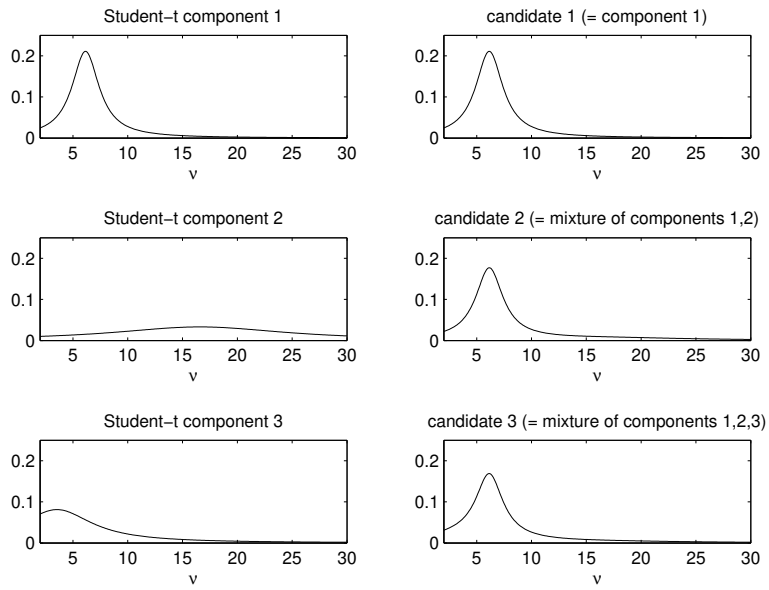
Figure 1: The AdMit algorithm automatically and iteratively approximates the skewed shapes of the marginal posterior of ν in the EGARCH model.

with AdMitIS. For the GJR parameter ω and EGARCH parameter ν, for which the marginal posteriors are substantially skewed, the RNEs are 3.4 and 2.6 times higher in the AdMitIS approach. Further, note the lower estimate of the posterior standard deviation of the EGARCH parameter ν. Even for $L = 100\,000$ draws, the naive approach may hardly cover some relevant regions of the parameter space. This failure of the naive approach does not show up in the NSE and RNE, as these only concern the parameter subspace that is visited by the draws. It should be mentioned here that the models that are used in practice often have more than five parameters, and that posterior shapes can be much further from normality than those in our application (e.g., not only skewness but also curved ridges or even multimodality). In such cases, the naive simulation method's problem of neglecting relevant parts of the parameter space can be even much larger. Then estimation, prediction and risk measurement can suffer from substantial biases.

As explained in the previous section, the AdMit algorithm also delivers a suitable importance density for marginal likelihood estimation. The natural logarithm of the AdMitIS estimate of the marginal likelihood is given by -725.6930 and -724.5382 for the GJR and EGARCH model, respectively. This implies a Bayes factor of 3.1734 in favor of the EGARCH model, given the observed data. Under equal prior model probabilities, i.e., a prior odds ratio of one, the posterior probabilities for the GJR and EGARCH models are estimated as 0.2396 and 0.7604, respectively.

	$\mathcal{M}_1$: GJR Student-t					$\mathcal{M}_2$: EGARCH Student-t			
θ	$\hat{\mathbb{E}}_p(\theta)$	NSE	RNE	$\hat{\mathbb{V}}_p^{1/2}(\theta)$	θ	$\hat{\mathbb{E}}_p(\theta)$	NSE	RNE	$\hat{\mathbb{V}}_p^{1/2}(\theta)$
ω	0.0205	0.0002	0.0927	0.0147	ω	-0.0105	0.0001	0.1198	0.0117
α	0.0349	0.0002	0.1232	0.0240	α	0.1384	0.0002	0.2525	0.0367
γ	0.1124	0.0005	0.1404	0.0566	γ	-0.0737	0.0002	0.2230	0.0318
β	0.8898	0.0004	0.1207	0.0424	β	0.9733	0.0002	0.1201	0.0201
ν	6.4843	0.0201	0.0693	1.6745	ν	6.6905	0.0140	0.1844	1.9667

Table 1: Posterior results for the two non-nested models using AdMitIS. $\hat{\mathbb{E}}_p(\theta)$ posterior mean estimate. NSE: numerical standard error of the posterior mean estimate. RNE: relative numerical efficiency of the posterior mean estimate; $\hat{\mathbb{V}}_p^{1/2}(\theta)$: posterior standard deviation estimate. The number of importance sampling draws is $L = 100\,000$.

|	$\mathcal{M}_1$: GJR Student-t					$\mathcal{M}_2$: EGARCH Student-t			
θ	$\hat{\mathbb{E}}_p(\theta)$	NSE	RNE	$\hat{\mathbb{V}}_p^{1/2}(\theta)$	θ	$\hat{\mathbb{E}}_p(\theta)$	NSE	RNE	$\hat{\mathbb{V}}_p^{1/2}(\theta)$
ω	0.0202	0.0003	0.0271	0.0148	ω	-0.0106	0.0001	0.0608	0.0117
α	0.0350	0.0003	0.0890	0.0241	α	0.1389	0.0003	0.1722	0.0372
γ	0.1112	0.0008	0.0516	0.0565	γ	-0.0734	0.0002	0.1710	0.0318
β	0.8903	0.0006	0.0427	0.0424	β	0.9732	0.0003	0.0467	0.0202
ν	6.5077	0.0291	0.0337	1.6884	ν	6.6508	0.0180	0.1028	1.8265

Table 2: Posterior results for the two non-nested models using naive importance sampling (i.e., using a Student-t distribution around the posterior mode). See Table 1 for explanations.

In Figures 4 and 5, we display the predictive densities for the volatility σ_{T+1} and log-return y_{T+1} for the date October 28, 1997, for the GJR and EGARCH models as well as for Bayesian model averaging (BMA) which combines the predictive densities

14
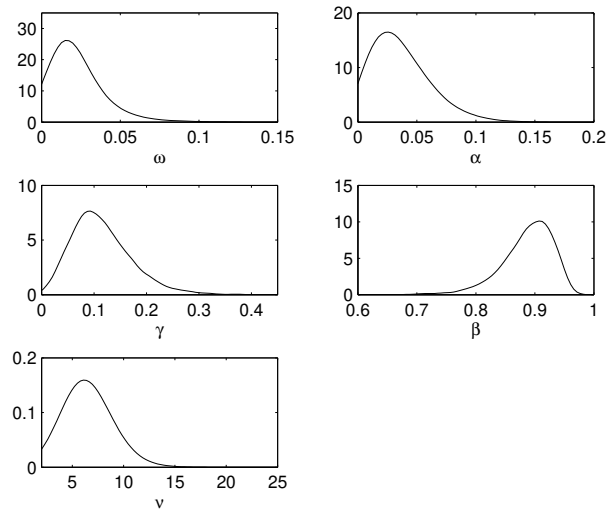
Figure 2: Marginal posterior densities for parameters in GJR model.
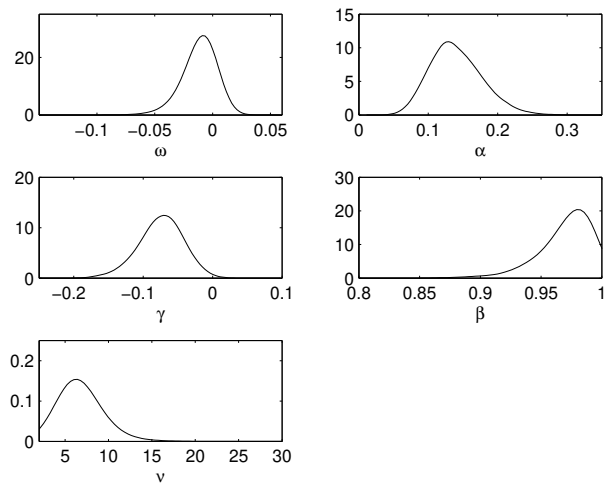
Figure 3: Marginal posterior densities for parameters in EGARCH model.

from the models via the posterior model probabilities. Approximate values of risk measures such as the value-at-risk (VaR) can be read from the figure.

Precise estimation of the VaR at the $100\alpha\%$ risk level can be obtained via importance sampling as follows. First, L draws $\{\boldsymbol{\theta}^{[l]}\}$ are simulated from the importance density, and the corresponding importance weights $\{w(\boldsymbol{\theta}^{[l]})\}$ are computed. Second, for the $\tilde{L}$ $(\tilde{L} \leq L)$ draws with non-zero weights a log-return $y_{T+1}^{[l]}$ is simulated from its distribution given the model and parameter values. Third, the simulated log-returns $y_{T+1}^{[l]}$ are sorted ascending as $y_{T+1}^{(i)}$ $(i = 1, \ldots, \tilde{L})$ and the VaR is estimated as $y_{T+1}^{(k)}$ with $S_k = (1 - \alpha)$ where $S_k \doteq \sum_{j=1}^{k} \tilde{w}(\boldsymbol{\theta}^{(j)})$ is the cumulative sum of scaled weights $\tilde{w}(\boldsymbol{\theta}^{(j)}) \doteq \frac{w(\boldsymbol{\theta}^{(j)})}{\sum_{i=1}^{\tilde{L}} w(\boldsymbol{\theta}^{(i)})}$ (i.e., scaled to add to one) corresponding to the ascending log-returns. In general there will be no k such that $S_k = (1 - \alpha)$, so that one interpolates between the values of $y_{T+1}^{(k)}$ and $y_{T+1}^{(k+1)}$ where S_{k+1} is the smallest value with $S_{k+1} > (1 - \alpha)$. For importance sampling estimation of VaR in case of more general profit and loss functions, see Hoogerheide and van Dijk (in press). Hoogerheide and van Dijk (in press) also introduce an additional approximation step, targeted particularly on generating high loss scenarios, that can make the simulation-based estimation of VaR even more efficient.

Table 3 reports the values for the one-day-ahead VaR for October 28, 1997, for both models and BMA. Figures 4 and 5 and Table 3 show that the GJR model predicts higher risk than EGARCH, with BMA results naturally taking a position in between. Accounting for both the uncertainty on estimated parameters and the uncertainty on model choice, we obtain BMA estimates of VaR (at the 95% risk level) at -3.7524 percent.

α	$\mathcal{M}_1$: GJR Student-t	$\mathcal{M}_2$: EGARCH Student-t	BMA
0.95	-4.7124	-3.4304	-3.7524
0.99	-7.6982	-5.6465	-6.3128

Table 3: Estimates of the $100\alpha\%$ one-day-ahead value-at-risk for the GJR model, EGARCH model, and Bayesian model averaging (BMA) of the two models.

Finally, note that instead of integrating out the parameter (and model) uncertainty in the VaR forecast, we could also determine the impact of the parameter (and model) uncertainty on the VaR estimate. We do not investigate this point here but refer the reader to the chapter by Ausin and Lopes in this volume; see also Ardia (2008b, Chapter 6) for an illustration of the impact of the parameter uncertainty to the VaR and expected shorfall term structures.

5 Conclusion

The study of GARCH-type models from a Bayesian viewpoint is relatively recent and can be considered very promising due to the advantages of the Bayesian approach compared to classical techniques. In particular, the Bayesian framework enables small sample results, robust estimation, probabilistic statements on nonlinear functions of the model parameters and model discrimination. Moreover, the Bayesian paradigm
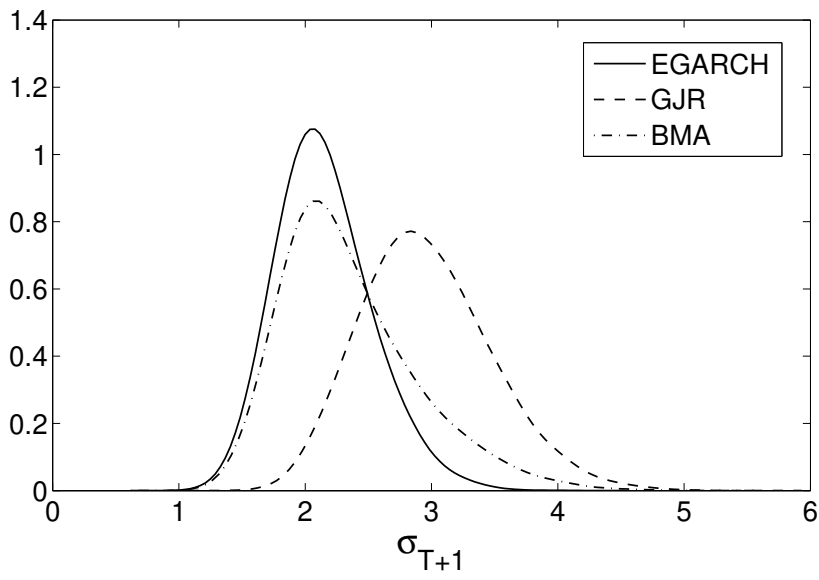
Figure 4: Predictive density of σ_{T+1} for EGARCH model, GJR model and Bayesian model averaging (BMA) of the two models.
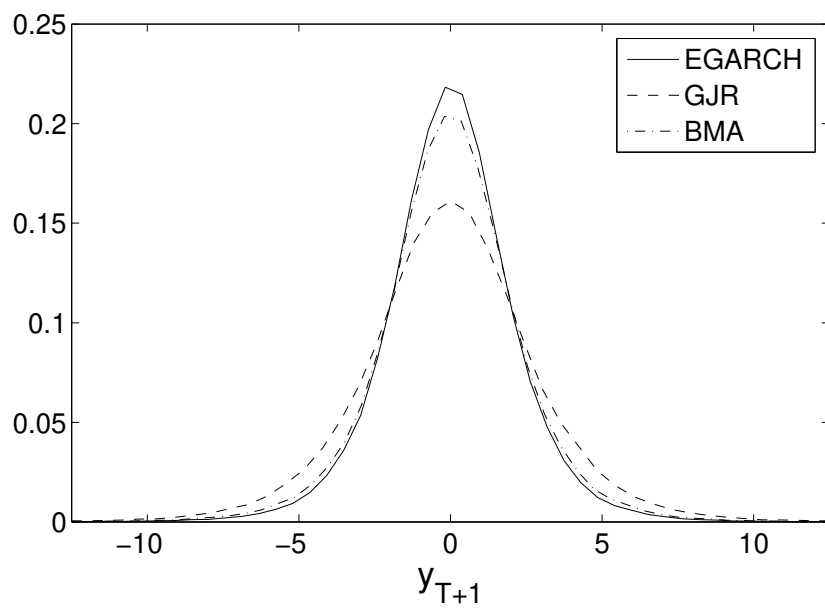
Figure 5: Predictive density of y_{T+1} for EGARCH model, GJR model and Bayesian model averaging (BMA) of the two models.

19

allows to combine model forecasts, thus accounting for model risk in the predictions, which is crucial from a risk management perspective.

This chapter reviewed existing methods for the Bayesian estimation of GARCH-type models. We focused our presentation on a novel approach, named AdMitIS, which performs importance sampling with an adaptive mixture of Student-t distribution as the importance distribution. The methodology allows a quick and efficient estimation of any kind of GARCH-type models. Moreover, with this approach, it is easy to combine forecasts of non-nested models. The estimation procedure is fully automatic and is achieved within a reasonable computational time compared to alternative MCMC techniques. This is of primary importance for automated trading systems for instance, where models are estimated frequently and for numerous datasets.

The AdMitIS algorithm was described in details and we provided an application to S&P500 index log-returns. We illustrated how two non-nested GARCH-type models can be estimated and combined in order to forecast the next-day ahead log-returns distribution.

References

Ardia D (2008a). *'bayesGARCH': Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations in R.* URL `http://CRAN.R-project.org/package=bayesGARCH`.

Ardia D (2008b). *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications*, volume 612 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, Germany. ISBN 978-3-540-78656-6. doi:10.1007/978-3-540-78657-3.

Ardia D (2009). "Bayesian Estimation of a Markov-Switching Threshold Asymmetric GARCH Model with Student-t Innovations." *Econometrics Journal*, **12**, 105–126. doi:10.1111/j.1368-423X.2008.00253.x.

Ardia D, Hoogerheide LF (2009). "Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations in R." MPRA working paper. URL `http://mpra.ub.uni-muenchen.de/17414/`.

Ardia D, Hoogerheide LF, van Dijk HK (2008). *'AdMit': Adaptive Mixture of Student-t Distributions for Efficient Simulation in R.* URL `http://CRAN.R-project.org/package=AdMit`.

Ardia D, Hoogerheide LF, van Dijk HK (2009a). "Adaptive Mixture of Student-t Distributions as a Flexible Candidate Distribution for Efficient Simulation: The R Package AdMit." *Journal of Statistical Software*, **29**(3), 1–32. URL `http://www.jstatsoft.org/v29/i03/`.

Ardia D, Hoogerheide LF, van Dijk HK (2009b). "AdMit: Adaptive Mixtures of Student-t Distributions." *The R Journal*, **1**(1), 25–30.

Ardia D, Hoogerheide LF, van Dijk HK (2009c). "To Bridge, to Warp or to Wrap? A Comparative Study of Monte Carlo Methods for Efficient Evaluation of Marginal Likelihoods." Tinbergen Institute report 09-017/4.

Asai M (2006). "Comparison of MCMC Methods for Estimating GARCH Models." *Journal of the Japan Statistical Society*, **36**(2), 199–212.

Ausin MC, Galeano P (2007). "Bayesian Estimation of the Gaussian Mixture GARCH Model." *Computational Statistics & Data Analysis*, **51**(5), 2636–2652. doi:10.1016/j.csda.2006.01.006.

Aussenegg W, Miazhynskaia T (2006). "Uncertainty in Value at Risk Estimates under Parametric and Non-Parametric Modeling." *Financial Markets & Portfolio Management*, **20**(3), 243–264. doi:10.1007/s11408-006-0020-8.

Bauwens L, Lubrano M (1998). "Bayesian Inference on GARCH Models Using the Gibbs Sampler." *Econometrics Journal*, **1**(1), C23–C46. doi:10.1111/1368-423X.11003.

Bauwens L, Lubrano M (2002). "Bayesian Option Pricing Using Asymmetric GARCH Models." *Journal of Empirical Finance*, **9**(3), 321–342. doi:10.1016/S0927-5398(01)00058-5.

Bauwens L, Lubrano M, Richard JF (2000). *Bayesian Inference in Dynamic Econometric Models*. Advanced Texts in Econometrics. Oxford University Press, Oxford, UK, first edition. ISBN 0198773129.

Bauwens L, Preminger A, Rombouts JVK (in press). "Theory and Inference for a Markov Switching GARCH Model." Forthcoming in The Econometrics Journal.

Bauwens L, Rombouts JVK (2007). "Bayesian Inference for the Mixed Conditional Heteroskedasticity Model." *Econometrics Journal*, **10**(2), 408–425. doi:10.1111/j.1368-423X.2007.00213.x.

Black F (1976). "The Pricing of Commodity Contracts." *Journal of Financial Economics*, **3**(1–2), 167–179. doi:10.1016/0304-405X(76)90024-6.

Bollerslev T, Chou RY, Kroner K (1992). "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence." *Journal of Econometrics*, **52**(1–2), 5–59. doi:10.1016/0304-4076(92)90064-X.

Bollerslev T, Engle RF, Nelson DB (1994). "ARCH Models." In "Handbook of Econometrics," chapter 49, pp. 2959–3038. North Holland.

Chen CW, Gerlach R (2008). "Bayesian Inference and Model Comparison for Asymmetric Smooth Transition Heteroscedastic Models." *Statistics & Computing*, **18**(4), 391–408. doi:10.1007/s11222-008-9063-1.

Chen CW, Gerlach R, So MKP (2008). "Bayesian Model Selection for Heteroscedastic Models." In S Chib, G Koop, B Griffiths, D Terrell (eds.), "Bayesian Econometrics," Advanced in Econometrics, pp. 567–594. Emerald Group Publishing Ltd. ISBN 1848553080.

Chen CWS, So MKP, Lin EMH (2009). "Volatility Forecasting with Double Markov Switching GARCH Models." *Journal of Forecasting*, **28**(8), 681–697. doi:10.1002/for.1119.

Dueker MJ (1997). "Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility." *Journal of Business & Economic Statistics*, **15**(1), 26–34.

Engle RF (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, **50**(4), 987–1008.

Engle RF (2004). "Risk and Volatility: Econometric Models and Financial Practice." *The American Economic Review*, **94**(3), 405–420. doi:10.1257/0002828041464597.

Frühwirth-Schnatter S (2006). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. Springer-Verlag, New York/Berlin/Heidelberg, first edition. ISBN 0387329099.

Gerlach R, Chen CWS (2006). "Comparison of Nonnested Asymmetric Heteroscedastic Models." *Computational Statistics & Data Analysis*, **51**(15), 2164–2178. doi:10.1016/j.csda.2006.07.025.

Geweke JF (1988). "Exact Inference in Models with Autoregressive Conditional Heteroscedasticity." In ER Berndt, HL White, WA Barnett (eds.), "Dynamic Econometric Modeling," Number 3 in International Symposium in Economic Theory and Econometrics, pp. 73–103. Cambridge University Press, New York, USA. ISBN 0521333954.

Geweke JF (1989a). "Bayesian Inference in Econometric Models Using Monte Carlo Integration." *Econometrica*, **57**(6), 1317–1339. Reprinted in: Bayesian Inference, G. C. Box and N. Polson (Eds.), Edward Elgar Publishing, 1994.

Geweke JF (1989b). "Exact Predictive Densities in Linear Models with ARCH Disturbances." *Journal of Econometrics*, **40**(1), 63–86. doi:10.1016/0304-4076(89)90030-4.

Glosten LR, Jaganathan R, Runkle DE (1993). "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." *Journal of Finance*, **48**(5), 1779–1801.

Hammersley JM, Handscomb DC (1965). *Monte Carlo Methods*. Chapman and Hall. ISBN 0412158701.

Hastings WK (1970). "Monte Carlo Sampling Methods Using Markov Chains and their Applications." *Biometrika*, **57**(1), 97–109. doi:10.1093/biomet/57.1.97.

Henneke JS, Rachev Svetlozar T, Fabozzi FJ, Nikolov M (2009). "MCMC-based Estimation of Markov Switching ARMA-GARCH Models." *Applied Economics*, **iFirst**, 1–13. doi:10.1080/00036840802552379.

Hoogerheide LF (2006). *Essays on Neural Network Sampling Methods and Instrumental Variables*. Ph.D. thesis, Tinbergen Institute, Erasmus University Rotterdam. Book nr. 379 of the Tinbergen Institute Research Series.

Hoogerheide LF, Kaashoek JF, van Dijk HK (2007). "On the Shape of Posterior Densities and Credible Sets in Instrumental Variable Regression Models with Reduced Rank: An Application of Flexible Sampling Methods using Neural Networks." *Journal of Econometrics*, **139**(1), 154–180. doi:10.1016/j.jeconom.2006.06.009.

Hoogerheide LF, van Dijk HK (2008). "Possibly Ill-Behaved Posteriors in Econometric Models: On the Connection between Model Structures, Non-elliptical Credible Sets and Neural Network Simulation Techniques." Tinbergen Institute discussion paper 2008-036/4. URL http://www.tinbergen.nl/discussionpapers/08036.pdf.

Hoogerheide LF, van Dijk HK (in press). "Bayesian Forecasting of Value at Risk and Expected Shortfall using Adaptive Importance Sampling." Forthcoming in International Journal of Forecasting.

Kass RE, Raftery AE (1995). "Bayes Factors." *Journal of the American Statistical Association*, **90**(430), 773–795.

Kaufmann S, Frühwirth-Schnatter S (2002). "Bayesian Analysis of Switching ARCH Models." *Journal of Time Series Analysis*, **23**(4), 425–458. doi:10.1111/1467-9892. 00271.

Kaufmann S, Scheicher M (2006). "A Switching ARCH Model for the German DAX Index." *Studies in Nonlinear Dynamics & Econometrics*, **10**(4), 1–35. Article nr. 3, URL http://www.bepress.com/snde/vol10/iss4/art3/.

Klaassen F (2002). "Improving GARCH Volatility Forecasts with Regime-Switching GARCH." *Empirical Economics*, **27**(2), 363–394. doi:10.1007/s001810100100.

Kleibergen F, van Dijk HK (1993). "Non-Stationarity in GARCH Models: A Bayesian Analysis." *Journal of Applied Econometrics*, **8**(S1), S41–S61. doi: 10.1002/jae.3950080505.

Kloek T, van Dijk HK (1978). "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo." *Econometrica*, **46**(1), 1–19.

Lamoureux CG, Lastrapes WD (1990). "Persistence in Variance, Structural Change, and the GARCH Model." *Journal of Business & Economic Statistics*, **8**(2), 225–243.

Lee SW, Hansen BE (1994). "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator." *Econometric Theory*, **10**(1), 29–52.

Marcucci J (2005). "Forecasting Stock Market Volatility with Regime-Switching GARCH Models." *Studies in Nonlinear Dynamics & Econometrics*, **9**(4), 1–53. Article nr. 6, URL http://www.bepress.com/snde/vol9/iss4/art6/.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953). "Equations of State Calculations by Fast Computing Machines." *Journal of Chemical Physics*, **21**(6), 1087–1092.

Miazhynskaia T, Dorffner G (2006). "A Comparison of Bayesian Model Selection Based on MCMC with Application to GARCH-type Models." *Statistical Papers*, **47**(4), 525–549. doi:10.1007/s00362-006-0305-z.

Miazhynskaia T, Frühwirth-Schnatter S, Dorffner G (2006). "Bayesian Testing for Non-linearity in Volatility Modeling." *Computational Statistics & Data Analysis*, **51**(3), 2029–2042. doi:10.1016/j.csda.2005.12.014.

Müller P, Pole A (1998). "Monte Carlo Posterior Integration in GARCH Models." *Sankhya: The Indian Journal of Statistics*, **60**, 127–144.

Nakatsuma T (1998). "A Markov-Chain Sampling Algorithm for GARCH Models." *Studies in Nonlinear Dynamics & Econometrics*, **3**(2), 107–117. Algorithm nr.1, URL http://www.bepress.com/snde/vol3/iss2/algorithm1/.

Nakatsuma T (2000). "Bayesian Analysis of ARMA-GARCH Models: A Markov Chain Sampling Approach." *Journal of Econometrics*, **95**(1), 57–69. doi:10.1016/S0304-4076(99)00029-9.

Nakatsuma T, Tsurumi H (1999). "Bayesian Estimation of ARMA-GARCH Model of Weekly Foreign Exchange Rates." *Asia-Pacific Financial Markets*, **6**, 71–84.

Nelson DB (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica*, **59**(2), 347–370.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. URL http://www.R-project.org/.

Rachev ST, Hsu JSJ, Bagasheva Biliana S, Fabozzi Frank J (2008). *Bayesian Methods in Finance*. John Whiley and Sons Ltd.

Ritter C, Tanner MA (1992). "Facilitating the Gibbs Sampler: the Gibbs Stopper and the Griddy-Gibbs Sampler." *Journal of the American Statistical Association*, **87**(419), 861–868.

The MathWorks Inc (2009). *MATLAB*.

Tsay RS (2005). *Analysis of Financial Time Series*. Wiley Series in Probability and Statistics. John Wiley and Sons Ltd., Hoboken, New Jersey, USA, second edition. ISBN 0471690740.

Vrontos ID, Dellaportas P, Politis DN (2000). "Full Bayesian Inference for GARCH and EGARCH Models." *Journal of Business & Economic Statistics*, **18**(2), 187–198.

Wago H (2004). "Bayesian Estimation of Smooth Transition GARCH Model Using Gibbs Sampling." *Mathematics & Computer in Simulation*, **64**(1), 63–78. doi:10.1016/S0378-4754(03)00121-6.

Wolfram Research (2009). *Mathematica*.

Zakoian JM (1994). "Threshold Heteroskedastic Models." *Journal of Economic Dynamic & Control*, **18**(5), 931–955. doi:10.1016/0165-1889(94)90039-6.

Zeevi AJ, Meir R (1997). "Density Estimation Through Convex Combinations of Densities: Approximation and Estimation Bounds." *Neural Networks*, **10**(1), 99–109. doi:10.1016/S0893-6080(96)00037-8.