612 LECTURE NOTES IN ECONOMICS AND MATHEMATICAL SYSTEMS

David Ardia

Financial Risk Management with Bayesian Estimation of GARCH Models

Theory and Applications

Springer

Lecture Notes in Economics
and Mathematical Systems 612

David Ardia

Financial Risk Management with Bayesian Estimation of GARCH Models

Theory and Applications

Springer

Dr. David Ardia
Department of Quantitative Economics
University of Fribourg
Bd. de Pérolles 90
1700 Fribourg
Switzerland
david.ardia@unifr.ch

ISBN 978-3-540-78656-6 e-ISBN 978-3-540-78657-3

DOI 10.1007/978-3-540-78657-3

Lecture Notes in Economics and Mathematical Systems ISSN 0075-8442

Library of Congress Control Number: 2008927201

This book is the Ph.D. dissertation with the original title "Bayesian Estimation of Single-Regime and Regime-Switching GARCH Models. Applications to Financial Risk Management" presented to the Faculty of Economics and Social Sciences at the University of Fribourg Switzerland by the author. Accepted by the Faculty Council on 19 February 2008. The Faculty of Economics and Social Sciences at the University of Fribourg Switzerland neither approves nor disapproves the opinions expressed in a doctoral dissertation. They are to be considered those of the author. (Decision of the Faculty Council of 23 January 1990).

Production: le-tex Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: WMX Design GmbH, Heidelberg

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

To my nonno, Riziero.

Preface

This book presents in detail methodologies for the Bayesian estimation of single-regime and regime-switching GARCH models. These models are widespread and essential tools in financial econometrics and have, until recently, mainly been estimated using the classical Maximum Likelihood technique. As this study aims to demonstrate, the Bayesian approach offers an attractive alternative which enables small sample results, robust estimation, model discrimination and probabilistic statements on nonlinear functions of the model parameters.

The author is indebted to numerous individuals for help in the preparation of this study. Primarily, I owe a great debt to Prof. Dr. Philippe J. Deschamps who inspired me to study Bayesian econometrics, suggested the subject, guided me under his supervision and encouraged my research. I would also like to thank Prof. Dr. Martin Wallmeier and my colleagues of the Department of Quantitative Economics, in particular Michael Beer, Roberto Cerratti and Gilles Kaltenrieder, for their useful comments and discussions.

I am very indebted to my friends Carlos Ordás Criado, Julien A. Straubhaar, Jérôme Ph. A. Taillard and Mathieu Vuilleumier, for their support in the fields of economics, mathematics and statistics. Thanks also to my friend Kevin Barnes who helped with my English in this work.

Finally, I am greatly indebted to my parents and grandparents for their support and encouragement while I was struggling with the writing of this thesis. Thanks also to Margaret for her support some years ago. Last but not least, thanks to you Sophie for your love which puts equilibrium in my life.

Fribourg, April 2008 *David Ardia*

Table of Contents

Summary

This book presents in detail methodologies for the Bayesian estimation of single-regime and regime-switching GARCH models. Our sampling schemes have the advantage of being fully automatic and thus avoid the time-consuming and difficult task of tuning a sampling algorithm. The study proposes empirical applications to real data sets and illustrates probabilistic statements on nonlinear functions of the model parameters made possible under the Bayesian framework.

The first two chapters introduce the work and give a short overview of the Bayesian paradigm for inference. The next three chapters describe the estimation of the GARCH model with Normal innovations and the linear regression models with conditionally Normal and Student-t-GJR errors. For these models, we compare the Bayesian and Maximum Likelihood approaches based on real financial data. In particular, we document that even for fairly large data sets, the parameter estimates and confidence intervals are different between the methods. Caution is therefore in order when applying asymptotic justifications for this class of models. The sixth chapter presents some financial applications of the Bayesian estimation of GARCH models. We show how agents facing different risk perspectives can select their optimal VaR point estimate and document that the differences between individuals can be substantial in terms of regulatory capital. Finally, the last chapter proposes the estimation of the Markov-switching GJR model. An empirical application documents the in- and out-of-sample superiority of the regime-switching specification compared to single-regime GJR models. We propose a methodology to depict the density of the one-day ahead VaR and document how specific forecasters' risk perspectives can lead to different conclusions on the forecasting performance of the MS-GJR model.

JEL Classification: C11, C13, C15, C16, C22, C51, C52, C53.

Keywords and phrases: Bayesian, MCMC, GARCH, GJR, Markov-switching, Value at Risk, Expected Shortfall, Bayes factor, DIC.

1

Introduction

> *(...) "skedasticity refers to the volatility or wiggle of a time series. Heteroskedastic means that the wiggle itself tends to wiggle. Conditional means the wiggle of the wiggle depends on its own past wiggle. Generalized means that the wiggle of the wiggle can depend on its own past wiggle in all kinds of wiggledy ways."*
>
> — *Kent Osband*

Volatility plays a central role in empirical finance and financial risk management and lies at the heart of any model for pricing derivative securities. Research on changing volatility (*i.e.*, conditional variance) using time series models has been active since the creation of the original ARCH (AutoRegressive Conditional Heteroscedasticity) model in 1982. From there, ARCH models grew rapidly into a rich family of empirical models for volatility forecasting during the last twenty years. They are now widespread and essential tools in financial econometrics.

In the ARCH(q) specification originally introduced by Engle [1982], the conditional variance at time t, denoted by h_t, is postulated to be a linear function of the squares of past q observations $\{y_{t-1}, y_{t-2}, \ldots, y_{t-q}\}$. More precisely:

$$h_t \doteq \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 \tag{1.1}$$

where the parameters $\alpha_0 > 0$ and $\alpha_i \geqslant 0$ $(i = 1, \ldots, q)$ in order to ensure a positive conditional variance. In many of the applications with the ARCH model, a long lag length and therefore a large number of parameters are called for. To circumvent this problem, Bollerslev [1986] proposed the *Generalized* ARCH, or GARCH(p, q), model which extends the specification of the conditional variance (1.1) as follows:

$$h_t \doteq \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$ $(i = 1, \ldots, q)$ and $\beta_j \geqslant 0$ $(j = 1, \ldots, p)$. In this case, the conditional variance depends on its past values which renders the model more parsimonious. Indeed, in most empirical applications it turns out that the simple specification $p = q = 1$ is able to reproduce the volatility dynamics of financial data. This has led the GARCH$(1, 1)$ model to become the "workhorse model" by both academics and practitioners.

Numerous extensions and refinements of the GARCH model have been proposed to mimic additional stylized facts observed in financial markets. These extensions recognize that there may be important nonlinearity, asymmetry, and long memory properties in the volatility process. Many of these models are surveyed in Bollerslev, Chou, and Kroner [1992], Bollerslev, Engle, and Nelson [1994], Engle [2004]. Among them, we may cite the popular *Exponential* GARCH model by Nelson [1991] as well as the GJR model by Glosten, Jaganathan, and Runkle [1993] which both account for the asymmetric relation between stock returns and changes in variance [see Black 1976]. An additional class of GARCH models, referred to as regime-switching GARCH, has gained particular attention in recent years. In these models, the scedastic function's parameters can change over time according to a latent (*i.e.*, unobservable) variable taking values in the discrete space $\{1, \ldots, K\}$. The interesting feature of these models lies in the fact that they provide an explanation of the high persistence in volatility, *i.e.*, nearly unit root process for the conditional variance, observed with single-regime GARCH models [see, *e.g.*, Lamoureux and Lastrapes 1990]. Furthermore, these models are apt to react quickly to changes in the volatility level which leads to significant improvements in volatility forecasts as shown by Dueker [1997], Klaassen [2002], Marcucci [2005]. Further details on regime-switching GARCH models can be found in Haas, Mittnik, and Paolella [2004], Hamilton and Susmel [1994].

The Maximum Likelihood (henceforth ML) estimation technique is the generally favored scheme of inference for GARCH models, although semi- and nonparametric techniques have also been applied by some authors [see, *e.g.*, Gallant and Tauchen 1989, Pagan and Schwert 1990]. The primary appeal of the ML technique stems from the well-known asymptotic optimality conditions of the resulting estimators under ideal conditions [see Bollerslev *et al.* 1994, Lee and Hansen 1994]. In addition, the ML procedure is straightforward to implement and is nowadays available in econometric packages. However, while conceptually simple, we may encounter practical difficulties when dealing with the ML estimation of GARCH models. First, the maximization of the likelihood function must be achieved via a constrained optimization technique. The model parameters must indeed be positive to ensure a positive conditional variance and it

is also common to require that the covariance stationarity condition holds (this condition is $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ for the GARCH(p, q) model [see Bollerslev 1986, Thm.1, p.310]). The optimization procedure subject to inequality constraints can be cumbersome and does not necessarily converge if the true parameter values are close to the boundary of the parameter space or if the process is nearly non-stationary. The maximization is even more difficult to achieve in the context of regime-switching GARCH models where the likelihood surface is multimodal. Depending on the numerical algorithm, ML estimates often prove to be sensitive with respect to starting values. Moreover, the covariance matrix at the optimum can be extremely tedious to obtain and ad-hoc approaches are often required to get reliable results (*e.g.*, Hamilton and Susmel [1994] fix some transition probabilities to zero in order to determine the variance estimates for some model parameters). Second, as noted by Geweke [1988, p.77], in classical 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, in the case of the GARCH(p, q) model, one might be interested in the unconditional variance, denoted by h_y, which is given by:

$$h_y \doteq \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}$$

provided that the covariance stationarity condition is satisfied. To assess the uncertainty of this quantity, classical inference involves tedious delta methods, simulation from the asymptotic Normal approximation of the parameter estimates or the bootstrap methodology. However, none of these techniques is completely satisfactory. The delta method is an approximation which can be crude if the function of interest is highly nonlinear. The simulation and the bootstrap approaches can deal with nonlinear functions of the model parameters and give a full description of their distribution. Nevertheless, the former technique relies on asymptotic justifications and the latter method is very demanding since at each step of the procedure, a GARCH model is fitted to the bootstrapped data. Finally, in the case of regime-switching GARCH models, testing the null of K versus K' states is not possible within the classical framework. The regularity conditions for justifying the χ^2 approximation of the likelihood ratio statistic do not hold as some parameters are undefined under the null hypothesis [see Frühwirth-Schnatter 2006, Sect.4.4].

Fortunately, difficulties disappear when Bayesian methods are used. First, any constraints on the model parameters can be incorporated in the modeling through appropriate prior specifications. Moreover, the recent development of computational methods based on Markov chain Monte Carlo (henceforth

MCMC) procedures can be used to explore the joint posterior distribution of the model parameters. These techniques avoid local maxima commonly encountered via ML estimation of regime-switching GARCH models. Second, exact distributions of nonlinear functions of the model parameters can be obtained at low cost by simulating from the joint posterior distribution. In particular, we will show in **Chap. 6** that, upon assuming that the underlying process is of GARCH type, the well known Value at Risk risk measure (henceforth VaR) can be expressed as a function of the model parameters. Therefore, the Bayesian approach gives an adequate framework to estimate the full density of the VaR. In conjunction with the decision theory framework, this allows to optimally choose a single point estimate within the density of the VaR, given our risk preferences. Hence, the Bayesian approach has a clear advantage in combining estimation and decision making. Lastly, in the Bayesian framework, the issue of determining the number of states can be addressed by means of model likelihood and Bayes factors. All these reasons strongly motivate the use of the Bayesian approach when estimating GARCH models.

The choice of the algorithm is the first issue when dealing with MCMC methods and it depends on the nature of the problem under study. In the case of GARCH models, due to the recursive nature of the conditional variance, the joint posterior and the full conditional densities are of unknown forms, whatever distributional assumptions are made on the model disturbances. Therefore, we cannot use the simple Gibbs sampler and need more elaborate estimation procedures. The initial approaches have been implemented using importance sampling [see Geweke 1988, 1989, Kleibergen and van Dijk 1993]. More recent studies include the Griddy-Gibbs sampler [see Ausín and Galeano 2007, Bauwens and Lubrano 1998] or the Metropolis-Hastings (henceforth M-H) algorithm with some specific choice of the proposal densities. The Normal random walk Metropolis is used in Müller and Pole [1998], Vrontos, Dellaportas, and Politis [2000], Adaptive Radial-Based Direction Sampling (henceforth ARDS) is proposed by Bauwens, Bos, van Dijk, and van Oest [2004] while Nakatsuma [1998, 2000] constructs proposal densities from an auxiliary process. In the context of regime-switching ARCH models, Kaufmann and Frühwirth-Schnatter [2002], Kaufmann and Scheicher [2006] use the method of Nakatsuma [1998, 2000] while Bauwens, Preminger, and Rombouts [2006], Bauwens and Rombouts [2007] rely on the Griddy-Gibbs sampler for regime-switching GARCH models.

In the importance sampling approach, a suitable importance density is required for efficiency which can be a bit of an art, especially if the posterior density is asymmetric or multimodal. In the random walk and independence M-

H strategies, preliminary runs and tuning are necessary. Therefore, the method cannot be completely automatic which is not a desirable property. The Griddy-Gibbs sampler of Ritter and Tanner [1992] is used by Bauwens and Lubrano [1998] in the context of GARCH models to get rid of these difficulties. This methodology consists in updating each parameter by inversion from the distribution computed by a deterministic integration rule. However, the procedure is time consuming and this can become a real burden for regime-switching models which involve many parameters. 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 the case of the ARDS algorithm of Bauwens *et al.* [2004], the method involves a reparametrization in order to enhance the efficiency of the estimation. This technique requires a large number of evaluations, which significantly slows down the estimation procedure compared to usual M-H approaches. Lastly, one could also use a Bayesian software such as BUGS [see Spiegelhalter, Thomas, Best, and Gilks 1995, Spiegelhalter, Thomas, Best, and Lunn 2007] for estimating GARCH models. However, this becomes extremely slow as the number of observations increases mainly due to the recursive nature of the conditional variance process. Moreover, the implementation of specific constraints on the model parameters is difficult and extensions to regime-switching specifications are limited.

In the rest of the book, we will use the approach suggested by Nakatsuma [1998, 2000] which relies on the M-H algorithm where some model parameters are updated by blocks. The proposal densities are constructed from an auxiliary ARMA process for the squared observations. This methodology has the advantage of being fully automatic and thus avoids the time-consuming and difficult task, especially for non-experts, of choosing and tuning a sampling algorithm. We obtained very high acceptance rates with this M-H algorithm, ranging from 89% to 95% for the single-regime GARCH$(1, 1)$ model, which indicates that the proposal densities are close to the full posteriors. In addition, the approach of Nakatsuma [1998, 2000] is easy to extend to regime-switching GARCH models. In this case, the parameters in each regime can be regrouped and updated by blocks which may enhance the sampler's efficiency.

Organization of the book

A short introduction to Bayesian inference and MCMC methods is given in **Chap. 2**. The rest of the book treats in detail the methodologies for the Bayesian estimation of single-regime and regime-switching GARCH models, proposes empirical applications to real data sets and illustrates some probabilistic state-

ments on nonlinear functions of the model parameters made possible under the Bayesian framework.

In **Chap. 3**, we propose the Bayesian estimation of the parsimonious but effective GARCH$(1,1)$ model with Normal innovations. We detail the MCMC scheme based on the methodology of Nakatsuma [1998, 2000]. An empirical application to a foreign exchange rate time series is presented where we compare the Bayesian and the ML estimates. In particular, we show that even for a fairly large data set, the point estimates and confidence intervals are different between the methods. Caution is therefore in order when applying the asymptotic Normal approximation for the model parameters in this case. We perform a sensitivity analysis to check the robustness of our results with respect to the choice of the priors and test the residuals for misspecification. Finally, we compare the theoretical and sample autocorrelograms of the process and test the covariance and strict stationarity conditions.

In **Chap. 4**, we consider the linear regression model with conditionally heteroscedastic errors and exogenous or lagged dependent variables. We extend the symmetric GARCH model to account for asymmetric responses to past shocks in the conditional variance process. To that aim, we consider the GJR$(1,1)$ model of Glosten *et al.* [1993]. We fit the model to the Standard and Poors 100 (henceforth S&P100) index log-returns and compare the Bayesian and the ML estimations. We perform a prior sensitivity analysis and test the residuals for misspecification. Finally, we test the covariance stationarity condition and illustrate the differences between the unconditional variance of the process obtained through the Bayesian approach and the delta method. In particular, we show that the Bayesian framework leads to a more precise estimate.

In **Chap. 5**, we extend the linear regression model with conditionally heteroscedastic errors by considering Student-t disturbances, which allows to model extreme shocks in a convenient manner. In the Bayesian approach, the heavy-tails effect is created by the introduction of latent variables in the variance process as proposed by Geweke [1993]. An empirical application based on the S&P100 index log-returns is proposed with a comparison between the estimated joint posterior and the asymptotic Normal approximation of the distribution of the estimates. We perform a prior sensitivity analysis and test the residuals for misspecification. Finally, we analyze the conditional and unconditional kurtosis of the underlying time series.

In **Chap. 6**, we present some financial applications of the Bayesian estimation of GARCH models. We introduce the concept of Value at Risk risk measure and propose a methodology to estimate the density of this quantity for different risk levels and time horizons. This gives us the possibility to determine the

VaR term structure and to characterize the uncertainty coming from the model parameters. Then, we review some basics in decision theory and use this framework as a rational justification for choosing a point estimate of the VaR. We show how agents facing different risk perspectives can select their optimal VaR point estimate and document, in an illustrative application, that the differences between individuals, in particular between fund managers and regulators, can be substantial in terms of regulatory capital. We show that the common testing methodology for assessing the performance of the VaR is unable to discriminate between the point estimates but the deviations are large enough to imply substantial differences in terms of regulatory capital. This therefore gives an additional flexibility to the user when allocating risk capital. Finally, we extend our methodology to the Expected Shortfall risk measure.

In **Chap. 7**, we extend the single-regime GJR model to the regime-switching GJR model (henceforth MS-GJR); more precisely, we consider an asymmetric version of the Markov-switching GARCH$(1,1)$ specification of Haas *et al.* [2004]. We introduce a novel MCMC scheme which can be viewed as an extension of the sampler proposed by Nakatsuma [1998, 2000]. Our approach allows to generate the parameters of the MS-GJR model by blocks which may enhance the sampler's efficiency. As an application, we fit a single-regime and a Markov-switching GJR model to the Swiss Market Index log-returns. We use the random permutation sampler of Frühwirth-Schnatter [2001b] to find suitable identification constraints for the MS-GJR model and show the presence of two distinct volatility regimes in the time series. The generalized residuals are used to test the models for misspecification. By using the Deviance information criterion of Spiegelhalter, Best, Carlin, and van der Linde [2002] and by estimating the model likelihoods using the bridge sampling technique of Meng and Wong [1996], we show the in-sample superiority of the MS-GJR model. To test the predictive performance of the models, we run a forecasting analysis based on the VaR. In particular, we compare the MS-GJR model to a single-regime GJR model estimated on rolling windows and show that both models perform equally well. However, contrary to the single-regime model, the Markov-switching model is able to anticipate structural breaks in the conditional variance process and needs to be estimated only once. Then, we propose a methodology to depict the density of the one-day ahead VaR by simulation and document how specific forecasters' risk perspectives can lead to different conclusions on forecasting performance of the model. A comparison with the traditional ML approach concludes the chapter.

Finally, we summarize the main results of the book and discuss future avenues of research in **Chap. 8**.

2

Bayesian Statistics and MCMC Methods

"The people who don't know they are Bayesian are called non-Bayesian."

— Irving J. Good

This chapter gives a short introduction to the Bayesian paradigm for inference and an overview of the Markov chain Monte Carlo (henceforth MCMC) algorithms used in the rest of the book. For a more thorough discussion on Bayesian statistics, the reader is referred to Koop [2003], for instance. Further details on MCMC methods can be found in Chib and Greenberg [1996], Smith and Roberts [1993], Tierney [1994]. The reader who is familiar with these topics can skip this part of the book and go to the first chapter dedicated to the Bayesian estimation of GARCH models, on page 17.

The plan of this chapter is as follows. The Bayesian paradigm is introduced in **Sect. 2.1**. MCMC techniques are presented in **Sect. 2.2** where we introduce the Gibbs sampler as well as the Metropolis-Hastings algorithm. We also briefly discuss some practical implementation issues.

2.1 Bayesian inference

As in the classical approach to inference, the Bayesian estimation assumes a $T \times 1$ vector $\mathbf{y} \doteq (y_1 \cdots y_T)'$ of observations described through a probability density $p(\mathbf{y} \mid \theta)$. The parameter $\theta \in \Theta$ serves as an index of the family of possible distributions for the observations. It represents the characteristics of interest one would wish to know in order to obtain a complete description of the generating process for $\mathbf{y}$. It can be a scalar, a vector, a matrix or even a set of these mathematical objects. For simplicity, we will consider θ as a d-dimensional vector, hence $\theta \in \Theta \subseteq \mathbb{R}^d$ in what follows.

The difference between the Bayesian and the classical approach lies in the mathematical nature of θ. In the classical framework, it is assumed that there exists a *true* and *fixed* value for parameter θ. Conversely, the Bayesian approach

considers θ as a random variable which is characterized by a *prior* density denoted by $p(\theta)$. The prior is specified with the help of parameters called *hyperparameters* which are initially assumed to be known and constant. Moreover, depending on the researcher's prior information, this density can be more or less informative. Then, by coupling the likelihood function of the model parameters, $\mathcal{L}(\theta \mid \mathbf{y}) \equiv p(\mathbf{y} \mid \theta)$, with the prior density, we can *invert* the probability density using Bayes' rule to get the *posterior* density $p(\theta \mid \mathbf{y})$ as follows:

$$p(\theta \mid \mathbf{y}) = \frac{\mathcal{L}(\theta \mid \mathbf{y})p(\theta)}{\int_{\Theta} \mathcal{L}(\theta \mid \mathbf{y})p(\theta)\mathrm{d}\theta} \ . \tag{2.1}$$

This posterior is a quantitative, probabilistic description of the knowledge about the parameter θ *after* observing the data.

It is often convenient to choose a prior density which is *conjugate* to the likelihood. That is, a density that leads to a posterior which belongs to the same distributional family as the prior. In effect, conjugate priors permit posterior densities to emerge without numerical integration. However, the easy calculations of this specification comes with a price due to the restrictions they impose on the form of the prior. In many cases, it is unlikely that the conjugate prior is an adequate representation of the prior state of knowledge. In such cases, the evaluation of (2.1) is analytically intractable, so asymptotic approximations or Monte Carlo methods are required. Deterministic techniques can provide good results for low dimensional models. However, when the dimension of the model becomes large, simulation is the only way to approximate the posterior density.

2.2 MCMC methods

The idea of MCMC sampling was first introduced by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller [1953] and was subsequently generalized by Hastings [1970]. For ease of exposition, we will restrict the presentation to the context of Bayesian inference. A general and detailed statistical theory of MCMC methods can be found in Tierney [1994].

The MCMC sampling strategy relies on the construction of a Markov chain with realizations $\theta^{[0]}, \theta^{[1]}, \ldots, \theta^{[j]}, \ldots$ in the parameter space Θ. Under appropriate regularity conditions [see Tierney 1994], asymptotic results guarantee that as j tends to infinity, then $\theta^{[j]}$ tends in distribution to a random variable whose density is $p(\theta \mid \mathbf{y})$. Hence, the realized values of the chain can be used to make inference about the joint posterior. All we require are algorithms for constructing appropriately behaved chains. The best known MCMC algorithms are the

Gibbs sampler and the Metropolis-Hastings (henceforth M-H) algorithm. These samplers are nowadays essential tools to perform realistic Bayesian inference.

2.2.1 The Gibbs sampler

The Gibbs sampler is possibly the MCMC sampling technique which is used most frequently. In the statistical physics literature, it is known as the *heat bath algorithm*. Geman and Geman [1984] christened it in the mainstream statistical literature as the Gibbs sampler. An elementary exposition can be found in Casella and George [1992]. See also Gelfand and Smith [1990], Tanner and Wong [1987] for practical examples.

The Gibbs sampler is an algorithm based on successive generations from the full conditional densities $p(\theta_i \mid \theta_{\neq i}, \mathbf{y})$, *i.e.*, the posterior density of the ith element of $\theta \doteq (\theta_1 \cdots \theta_d)'$, given all other elements, where elements of θ can be scalars or sub-vectors. In practice the sampler works as follows:

1. Initialize the iteration counter of the chain to $j = 1$ and set an initial value $\theta^{[0]} \doteq (\theta_1^{[0]} \cdots \theta_d^{[0]})'$;
2. Generate a new value $\theta^{[j]}$ from $\theta^{[j-1]}$ through successive generation values:

$$\begin{aligned}
\theta_1^{[j]} &\sim p(\theta_1 \mid \theta_{\neq 1}^{[j-1]}, \mathbf{y}) \\
\theta_2^{[j]} &\sim p(\theta_2 \mid \theta_1^{[j]}, \theta_3^{[j-1]}, \ldots, \theta_d^{[j-1]}, \mathbf{y}) \\
&\vdots \\
\theta_d^{[j]} &\sim p(\theta_d \mid \theta_{\neq d}^{[j]}, \mathbf{y});
\end{aligned}$$

3. Change counter j to $j + 1$ and go back to step 2 until convergence is reached.

As the number of iterations increases, the chain approaches its stationary distribution and convergence is then assumed to hold approximately [see Tierney 1994]. Sufficient conditions for the convergence of the Gibbs sampler are given in Roberts and Smith [1994, Sect.4]. As noted in Chib and Greenberg [1996, p.414], these conditions ensure that each full conditional density is well defined and that the support of the joint posterior is not separated into disjoint regions since this would prevent exploration of the full parameter space. Although these are only sufficient conditions for the convergence of the Gibbs sampler, they are extremely weak and are satisfied in most applications.

The Gibbs sampler is the most natural choice of MCMC sampling strategy when it is easy to write down full conditionals from which we can easily generate

draws. When the expression of $p(\theta_i \mid \theta_{\neq i}, \mathbf{y})$ is nonstandard, we might consider rejection methods [see, *e.g.*, Ripley 1987], the Griddy-Gibbs sampler when θ_i is univariate [see Ritter and Tanner 1992], adaptive rejection sampling [see Gilks and Wild 1992] or M-H sampling as shown in the next section.

2.2.2 The Metropolis-Hastings algorithm

Some complicated Bayesian problems cannot be solved by using the Gibbs sampler. This is the case when it is not easy to break down the joint density into full conditionals or when the full conditional densities are of unknown form. The M-H algorithm is a simulation scheme which allows to generate draws from any density of interest whose normalizing constant is unknown. The algorithm consists of the following steps.

1. Initialize the iteration counter to $j = 1$ and set an initial value $\theta^{[0]}$;
2. Move the chain to a new value $\theta^\star$ generated from a proposal (candidate) density $q(\bullet \mid \theta^{[j-1]})$;
3. Evaluate the acceptance probability of the move from $\theta^{[j-1]}$ to $\theta^\star$ given by:

$$\min\left\{\frac{p(\theta^\star \mid \mathbf{y})}{p(\theta^{[j-1]} \mid \mathbf{y})}\frac{q(\theta^{[j-1]} \mid \theta^\star)}{q(\theta^\star \mid \theta^{[j-1]})}, 1\right\} .$$

   If the move is accepted, set $\theta^{[j]} \doteq \theta^\star$, if not, set $\theta^{[j]} \doteq \theta^{[j-1]}$ so that the chain does not move;
4. Change counter from j to $j+1$ and go back to step 2 until convergence is reached.

As in the Gibbs sampler, the chain approaches its equilibrium distribution as the number of iterations increases [see Tierney 1994]. The power of the M-H algorithm stems from the fact that the convergence of the chain is obtained for any proposal q whose support includes the support of the joint posterior [see Roberts and Smith 1994, Sect.5]. It is however crucial that q approximates closely the posterior to guarantee an acceptance rate which is reasonable.

With no intention of being exhaustive, some comments are in order here. If we choose a symmetric proposal density, *i.e.*, $q(\theta^{[j]} \mid \theta^\star) = q(\theta^\star \mid \theta^{[j]})$, the acceptance probability of the M-H algorithm reduces to:

$$\min\left\{\frac{p(\theta^\star \mid \mathbf{y})}{p(\theta^{[j]} \mid \mathbf{y})}, 1\right\}$$

so that the proposal does not need to be evaluated. This simpler version of the M-H algorithm is known as the *Metropolis algorithm* because it is the original algorithm by Metropolis *et al.* [1953]. A special case consists of a proposal density which only depends on the distance between $\theta^\star$ and $\theta^{[j-1]}$, *i.e.*, $q(\theta^\star \mid \theta^{[j-1]}) = q(\theta^\star - \theta^{[j-1]})$. The resulting algorithm is referred to as the *random walk Metropolis algorithm*. For instance, q could be a multivariate Normal density centered at previous draw $\theta^{[j-1]}$ and whose covariance matrix is calibrated to take steps which are reasonably close to $\theta^{[j-1]}$ such that the probability of accepting the candidate is not too low, but with a step size large enough to ensure a sufficient exploration of the parameter space. The drawback of this method is that it is not fully automatic since the covariance matrix needs to be chosen carefully; thus preliminary runs are required.

Another special case of the M-H sampler is the *independence M-H algorithm*, in which proposal draws are generated independently of the current position of the chain, *i.e.*, $q(\theta^\star \mid \theta^{[j-1]}) = q(\theta^\star)$. This algorithm is often used with a Normal or a Student-t proposal density whose moments are estimated from previous runs of the MCMC sampler. This approach works well for well-behaved unimodal posterior densities but may be very inefficient if the posterior is asymmetric or multimodal.

Finally, we note that in the form of the M-H algorithm we have presented, the vector θ is updated in a single block at each iteration so that all elements are changed simultaneously. However, we could also consider *componentwise algorithms* where each component is generated by its own proposal density [see Chib and Greenberg 1995, Tierney 1994]. In fact, the Gibbs belongs to this class of samplers where each component is updated sequentially, and where proposal densities are the full conditionals. In this case, new draws are always accepted [see Chib and Greenberg 1995]. The M-H algorithm is often used in conjunction with the Gibbs sampler for those components of θ that have a conditional density that cannot be sampled from directly, typically because the density is known only up to a scale factor [see Tierney 1994].

2.2.3 Dealing with the MCMC output

Having examined the building-blocks for the standard MCMC samplers, we now discuss some issues associated with their practical implementation. In particular, we comment on the manner we can assess their convergence, the way we can account for autocorrelation in the chains and how we can obtain characteristics of the joint posterior from the MCMC output. Further details can be found in Kass, Carlin, Gelman, and Neal [1998], Smith and Roberts [1993].

Several statistics have been devised for assessing convergence of MCMC outputs. The basic idea behind most of them is to compare moments of the sampled parameters at different parts of the chain. Alternatively, we can compare several sequences drawn from different starting points and check that they are indistinguishable as the number of iterations increases. We refer the reader to Cowles and Carlin [1996], Gelman [1995] for a comparative review of these techniques. In the rest of the book, we will use a methodology based on the analysis of variance developed by Gelman and Rubin [1992]. More precisely, the approximate convergence is diagnosed when the variance between different sequences is no larger than the variance within each individual sequence. Apart from formal diagnostic tests, it is also often convenient to check convergence by plotting the parameters' draws over iterations (trace plots) as well as the cumulative or running mean of the drawings.

Regarding the Monte Carlo (simulation) error, it is crucial to understand that the draws generated by a MCMC method are not independent. The autocorrelation either comes from the fact that the new draw depends on the past value of the chain or that the old element is duplicated. When assessing the precision of an estimator, we must therefore rely on estimation techniques which account for this autocorrelation [see, *e.g.*, Geweke 1992, Newey and West 1987]. In the rest of the book, we will estimate the numerical standard errors, that is the variation of the estimates that can be expected if the simulations were to be repeated, by the method of Andrews [1991], using a Parzen kernel and AR(1) pre-whitening as presented in Andrews and Monahan [1992]. As noted by Deschamps [2006], this ensures easy, optimal, and automatic bandwidth selection.

After the run of a Markov chain and its convergence to the stationary distribution, a sample $\{\theta^{[j]}\}_{j=1}^{J}$ from the joint posterior density $p(\theta \mid \mathbf{y})$ is available. We can thus approximate the posterior expectation of any function $\xi(\theta)$ of the model parameters:

$$\mathbb{E}_{\theta|\mathbf{y}}\big[\xi(\theta)\big] = \int_{\Theta} \xi(\theta) p(\theta \mid \mathbf{y}) \mathrm{d}\theta \tag{2.2}$$

by averaging over the draws from the posterior distribution in the following manner:

$$\overline{\xi} \doteq \frac{1}{J} \sum_{j=1}^{J} \xi(\theta^{[j]}) \; .$$

Under mild conditions, the sample average $\overline{\xi}$ converges to the posterior expectation by the law of large numbers, even if the draws are generated by a MCMC sampler [see Tierney 1994]. Some particular cases of (2.2) allow to obtain characteristics of the joint posterior. For instance, when $\xi(\theta) \doteq \theta$ we obtain the

posterior mean vector $\overline{\theta}$; for $\xi(\theta) \doteq (\theta - \overline{\theta})(\theta - \overline{\theta})'$ we obtain the posterior covariance matrix; for $\xi(\theta) \doteq \mathbb{I}_{\{\theta \in C\}}$, where $\mathbb{I}_{\{\bullet\}}$ denotes the indicator function which is equal to one if the constraint holds and zero otherwise, we obtain the posterior probability of a set C. Finally, if we are interested in the marginal posterior density of a single component of θ, we can estimate it through a histogram or a kernel density estimate of sampled values [see Silverman 1986]. By contrast, deterministic numerical integration is often intractable.

3

Bayesian Estimation of the GARCH(1, 1) Model with Normal Innovations

"Large changes tend to be followed by large changes (of either sign) and small changes tend to be followed by small changes."

— Benoît Mandelbrot

(...) "it is remarkable how large a sample is required for the Normal distribution to be an accurate approximation."

— Robert McCulloch and Peter E. Rossi

In this chapter, we propose the Bayesian estimation of the parsimonious but effective GARCH(1, 1) model with Normal innovations. We sample the joint posterior distribution of the parameters using the approach suggested by Nakatsuma [1998, 2000]. As a first step, we fit the model to foreign exchange log-returns and compare the Bayesian and the Maximum Likelihood estimates. Next, we analyze the sensitivity of our results with respect to the choice of the priors and test the residuals for misspecification. Finally, we illustrate some appealing aspects of the Bayesian approach through probabilistic statements made on the parameters.

The plan of this chapter is as follows. We set up the model in **Sect. 3.1**. The MCMC scheme is detailed in **Sect. 3.2**. The empirical results are presented in **Sect. 3.3**. We conclude with some illustrative applications of the Bayesian approach in **Sect. 3.4**.

3.1 The model and the priors

A GARCH(1, 1) model with Normal innovations may be written as follows:

$$\begin{aligned} y_t &= \varepsilon_t h_t^{1/2} \quad \text{for } t = 1, \ldots, T \\ \varepsilon_t &\overset{iid}{\sim} \mathcal{N}(0,1) \\ h_t &\doteq \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1} \end{aligned} \tag{3.1}$$

where $\alpha_0 > 0$, $\alpha_1 \geqslant 0$ and $\beta \geqslant 0$ to ensure a positive conditional variance and $h_0 = y_0 \doteq 0$ for convenience; $\mathcal{N}(0,1)$ is the standard Normal density. In this setting, the conditional variance h_t is a linear function of the squared past observation and the past variance.

In order to write the likelihood function, we define the vectors $\mathbf{y} \doteq (y_1 \cdots y_T)'$ and $\boldsymbol{\alpha} \doteq (\alpha_0 \ \alpha_1)'$ and we regroup the model parameters into $\psi \doteq (\boldsymbol{\alpha}, \beta)$ for notational purposes. In addition, we define the $T \times T$ diagonal matrix:

$$\Sigma \doteq \Sigma(\psi) = \operatorname{diag}\big(\{h_t(\psi)\}_{t=1}^{T}\big)$$

where:

$$h_t(\psi) \doteq \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1}(\psi) \ .$$

From there, the likelihood function of ψ can be expressed as follows:

$$\mathcal{L}(\psi \mid \mathbf{y}) \propto (\det \Sigma)^{-1/2} \exp\left[-\tfrac{1}{2}\mathbf{y}'\Sigma^{-1}\mathbf{y}\right] \ .$$

We propose the following proper priors on the parameters $\boldsymbol{\alpha}$ and β of the preceding model:

$$p(\boldsymbol{\alpha}) \propto \mathcal{N}_2(\boldsymbol{\alpha} \mid \boldsymbol{\mu}_\alpha, \Sigma_\alpha)\mathbb{I}_{\{\boldsymbol{\alpha} > \mathbf{0}\}}$$
$$p(\beta) \propto \mathcal{N}(\beta \mid \mu_\beta, \Sigma_\beta)\mathbb{I}_{\{\beta > 0\}}$$

where $\mu_\bullet$ and $\Sigma_\bullet$ are the hyperparameters, $\mathbb{I}_{\{\bullet\}}$ is the indicator function which equals unity if the constraint holds and zero otherwise, $\mathbf{0}$ is a 2×1 vector of zeros and $\mathcal{N}_d$ is the d-dimensional Normal density ($d > 1$). In addition, we assume prior independence between parameters $\boldsymbol{\alpha}$ and β which implies that $p(\psi) = p(\boldsymbol{\alpha})p(\beta)$. Then, we construct the joint posterior density via Bayes' rule:

$$p(\psi \mid \mathbf{y}) \propto \mathcal{L}(\psi \mid \mathbf{y})p(\psi) \ . \tag{3.2}$$

3.2 Simulating the joint posterior

The recursive nature of the variance equation in model (3.1) does not allow for conjugacy between the likelihood function and the prior density in (3.2). Therefore, we rely on the M-H algorithm to draw samples from the joint posterior distribution. The algorithm in this section is a special case of the algorithm described by Nakatsuma [1998, 2000]. We draw an initial value $\psi^{[0]} \doteq (\boldsymbol{\alpha}^{[0]}, \beta^{[0]})$ from the joint prior and we generate iteratively J passes for ψ. A single pass is

decomposed as follows:

$$\begin{aligned}\boldsymbol{\alpha}^{[j]} &\sim p(\boldsymbol{\alpha} \mid \beta^{[j-1]}, \mathbf{y}) \\ \beta^{[j]} &\sim p(\beta \mid \boldsymbol{\alpha}^{[j]}, \mathbf{y}) \ .\end{aligned}$$

Since no full conditional density is known analytically, we sample parameters $\boldsymbol{\alpha}$ and β from two proposal densities. These densities are obtained by noting that the GARCH$(1,1)$ model can be written as an ARMA$(1,1)$ model for $\{y_t^2\}$. Indeed, by defining $w_t \doteq y_t^2 - h_t$, we can transform the expression of the conditional variance as follows:

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1} \\ \Leftrightarrow y_t^2 &= \alpha_0 + (\alpha_1 + \beta) y_{t-1}^2 - \beta w_{t-1} + w_t \end{aligned} \tag{3.3}$$

where w_t can be written as:

$$w_t \doteq y_t^2 - h_t = \left(\frac{y_t^2}{h_t} - 1 \right) h_t = \left(\chi_1^2 - 1 \right) h_t$$

with χ_1^2 denoting a Chi-squared variable with one degree of freedom. Hence, by construction, $\{w_t\}$ is a Martingale Difference process with a conditional mean of zero and a conditional variance of $2h_t^2$ since a χ_1^2 variable has a unit mean and a variance equal to two.

Following Nakatsuma [1998, 2000], we construct an approximate likelihood for parameters $\boldsymbol{\alpha}$ and β from expression (3.3). The procedure consists in approximating first the variable w_t by a variable z_t which is Normally distributed with a mean of zero and a variance of $2h_t^2$. This leads to the following *auxiliary* model:

$$y_t^2 = \alpha_0 + (\alpha_1 + \beta) y_{t-1}^2 - \beta z_{t-1} + z_t \ .$$

Then, by noting that z_t is a function of ψ given by:

$$z_t(\psi) = y_t^2 - \alpha_0 - (\alpha_1 + \beta) y_{t-1}^2 + \beta z_{t-1}(\psi) \tag{3.4}$$

and by defining the $T \times 1$ vector $\mathbf{z} \doteq (z_1 \cdots z_T)'$ as well as the $T \times T$ diagonal matrix:

$$\Lambda \doteq \Lambda(\psi) = \operatorname{diag}\left(\{2h_t^2(\psi)\}_{t=1}^T \right)$$

we can approximate the likelihood function of ψ from the auxiliary model as follows:

$$\mathcal{L}(\psi \mid \mathbf{y}) \propto (\det \Lambda)^{-1/2} \exp\left[-\tfrac{1}{2} \mathbf{z}' \Lambda^{-1} \mathbf{z} \right] \ . \tag{3.5}$$

As will be shown hereafter, the construction of the proposal densities for parameters $\boldsymbol{\alpha}$ and β is based on this approximate likelihood function.

3.2.1 Generating vector $\boldsymbol{\alpha}$

Recursive transformations initially proposed by Chib and Greenberg [1994] allow to express the function $z_t(\psi)$ in (3.4) as a linear function of the 2×1 vector $\boldsymbol{\alpha}$. Let us define $v_t \doteq y_t^2$ for notational convenience. The recursive transformations are defined as follows:

$$\begin{aligned} l_t^* &\doteq 1 + \beta\, l_{t-1}^* \\ v_t^* &\doteq v_{t-1} + \beta\, v_{t-1}^* \end{aligned}$$

where $l_0^* = v_0^* \doteq 0$. As shown in **Prop. A.1** (see **App. A**), upon defining the 2×1 vector $\mathbf{c}_t \doteq (l_t^*\ v_t^*)'$, the function z_t can be expressed as $z_t = v_t - \mathbf{c}_t'\boldsymbol{\alpha}$. Then, by considering the $T \times 1$ vector $\mathbf{v} \doteq (v_1 \cdots v_T)'$ and the $T \times 2$ matrix C whose tth row is $\mathbf{c}_t'$, we get $\mathbf{z} = \mathbf{v} - C\boldsymbol{\alpha}$. Therefore, we can express the approximate likelihood function of parameter $\boldsymbol{\alpha}$ as follows:

$$\mathcal{L}(\boldsymbol{\alpha} \mid \beta, \mathbf{y}) \propto (\det \Lambda)^{-1/2} \exp\left[-\tfrac{1}{2}(\mathbf{v} - C\boldsymbol{\alpha})'\Lambda^{-1}(\mathbf{v} - C\boldsymbol{\alpha})\right] .$$

The proposal density to sample vector $\boldsymbol{\alpha}$ is obtained by combining this likelihood function and the prior density by the usual Bayes update:

$$q_\alpha(\boldsymbol{\alpha} \mid \widetilde{\boldsymbol{\alpha}}, \beta, \mathbf{y}) \propto \mathcal{N}_2(\boldsymbol{\alpha} \mid \widehat{\boldsymbol{\mu}}_\alpha, \widehat{\Sigma}_\alpha)\mathbb{I}_{\{\boldsymbol{\alpha}>\mathbf{0}\}}$$

with:

$$\begin{aligned} \widehat{\Sigma}_\alpha^{-1} &\doteq C'\widetilde{\Lambda}^{-1}C + \Sigma_\alpha^{-1} \\ \widehat{\boldsymbol{\mu}}_\alpha &\doteq \widehat{\Sigma}_\alpha(C'\widetilde{\Lambda}^{-1}\mathbf{v} + \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha) \end{aligned}$$

where the $T\times T$ diagonal matrix $\widetilde{\Lambda} \doteq \operatorname{diag}\big(\{2h_t^2(\widetilde{\boldsymbol{\alpha}}, \beta)\}_{t=1}^T\big)$ and $\widetilde{\boldsymbol{\alpha}}$ is the previous draw of $\boldsymbol{\alpha}$ in the M-H sampler. A candidate $\boldsymbol{\alpha}^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\alpha}^\star, \beta \mid \mathbf{y})}{p(\widetilde{\boldsymbol{\alpha}}, \beta \mid \mathbf{y})}\frac{q_\alpha(\widetilde{\boldsymbol{\alpha}} \mid \boldsymbol{\alpha}^\star, \beta, \mathbf{y})}{q_\alpha(\boldsymbol{\alpha}^\star \mid \widetilde{\boldsymbol{\alpha}}, \beta, \mathbf{y})}, 1\right\} .$$

3.2.2 Generating parameter $\boldsymbol{\beta}$

The function $z_t(\psi)$ in (3.4) could be expressed, in the previous section, as a linear function of parameter $\boldsymbol{\alpha}$ but cannot be expressed as a linear function of

β. To overcome this problem, we linearize $z_t(\beta)$ by a first order Taylor expansion at point $\widetilde{\beta}$:

$$z_t(\beta) \simeq z_t(\widetilde{\beta}) + \left.\frac{\mathrm{d}z_t}{\mathrm{d}\beta}\right|_{\beta=\widetilde{\beta}} \times (\beta - \widetilde{\beta})$$

where $\widetilde{\beta}$ is the previous draw of parameter β in the M-H sampler. Furthermore, let us define the following:

$$r_t \doteq z_t(\widetilde{\beta}) + \widetilde{\beta}\nabla_t \quad , \quad \nabla_t \doteq -\left.\frac{\mathrm{d}z_t}{\mathrm{d}\beta}\right|_{\beta=\widetilde{\beta}}$$

where the terms ∇_t can be computed by the following recursion:

$$\nabla_t \doteq y_{t-1}^2 - z_{t-1}(\widetilde{\beta}) + \widetilde{\beta}\nabla_{t-1}$$

with $\nabla_0 \doteq 0$. This recursion is simply obtained by differentiating (3.4) with respect to β. Then, we regroup these terms into the $T \times 1$ vectors $\mathbf{r} \doteq (r_1 \cdots r_T)'$ and $\boldsymbol{\nabla} \doteq (\nabla_1 \cdots \nabla_T)'$ and we approximate the term within the exponential in (3.5) by $\mathbf{z} \simeq \mathbf{r} - \beta\boldsymbol{\nabla}$. This yields the following approximate likelihood function for parameter β:

$$\mathcal{L}(\beta \mid \boldsymbol{\alpha}, \mathbf{y}) \propto (\det \Lambda)^{-1/2} \exp\left[-\tfrac{1}{2}(\mathbf{r} - \beta\boldsymbol{\nabla})'\Lambda^{-1}(\mathbf{r} - \beta\boldsymbol{\nabla})\right] .$$

The proposal density to sample β is obtained by combining this likelihood and the prior density by Bayes' update:

$$q_\beta(\beta \mid \boldsymbol{\alpha}, \widetilde{\beta}, \mathbf{y}) \propto \mathcal{N}(\beta \mid \widehat{\boldsymbol{\mu}}_\beta, \widehat{\Sigma}_\beta)\mathbb{I}_{\{\beta>0\}}$$

with:

$$\begin{aligned} \widehat{\Sigma}_\beta^{-1} &\doteq \boldsymbol{\nabla}'\widetilde{\Lambda}^{-1}\boldsymbol{\nabla} + \Sigma_\beta^{-1} \\ \widehat{\boldsymbol{\mu}}_\beta &\doteq \widehat{\Sigma}_\beta(\boldsymbol{\nabla}'\widetilde{\Lambda}^{-1}\mathbf{r} + \Sigma_\beta^{-1}\boldsymbol{\mu}_\beta) \end{aligned}$$

where the $T \times T$ diagonal matrix $\widetilde{\Lambda} \doteq \mathrm{diag}\big(\{2h_t^2(\boldsymbol{\alpha}, \widetilde{\beta})\}_{t=1}^T\big)$. A candidate $\beta^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\alpha}, \beta^\star \mid \mathbf{y})}{p(\boldsymbol{\alpha}, \widetilde{\beta} \mid \mathbf{y})} \frac{q_\beta(\widetilde{\beta} \mid \boldsymbol{\alpha}, \beta^\star, \mathbf{y})}{q_\beta(\beta^\star \mid \boldsymbol{\alpha}, \widetilde{\beta}, \mathbf{y})}, 1\right\} .$$

We end this section with some comments regarding the implementation of the MCMC scheme. The program is written in the `R` language [see R Development Core Team 2007] with some subroutines implemented in `C` in order to speed up

the simulation procedure. The validity of the algorithm as well as the correctness of the computer code are verified by a variant of the method proposed by Geweke [2004]. We sample ψ from a proper joint prior and generate some passes of the M-H algorithm. At each pass, we simulate the dependent variable $\mathbf{y}$ from the full conditional $p(\mathbf{y} \mid \psi)$ which is given by the conditional likelihood. This way, we draw a sample from the joint density $p(\mathbf{y}, \psi)$. If the algorithm is correct, the resulting replications of ψ should reproduce the prior. The Kolmogorov-Smirnov empirical distribution test does not reject this hypothesis at the 1% significance level.

3.3 Empirical analysis

We apply our Bayesian estimation method to daily observations of the Deutschmark vs British Pound (henceforth DEM/GBP) foreign exchange log-returns. The sample period is from January 3, 1985, to December 31, 1991, for a total of 1'974 observations. The nominal returns are expressed in percent as in Bollerslev and Ghysels [1996]. This data set has been proposed as an informal benchmark for GARCH time series software validation and is available from the Journal of Business and Economic Statistics at `ftp://www.amstat.org/`. From this time series, the first 750 observations, which is about three financial years, are used to illustrate the Bayesian approach. The data set is large enough to perform classical Maximum Likelihood (henceforth ML) estimation and apply asymptotic justifications. Hence, we have an interesting point of view from which to compare classical and Bayesian approaches. The remaining data set will be used in an empirical analysis proposed in **Chap. 6**.

The observation window excerpt from our data set is plotted in the upper part of **Fig. 3.1**. We test for autocorrelation in the time series by testing the joint nullity of autoregressive coefficients for $\{y_t\}$. We estimate the regression with autoregressive coefficients up to lag 20 and compute the covariance matrix using the White estimate. The p-values of the Wald test is 0.377 which does not support the presence of autocorrelation. However, from **Fig. 3.1**, we clearly observe clusters of high and low variability in the time series. This phenomenon is well known in financial data and is referred to as volatility clustering. This effect is emphasized in the lower part of the figure where the sample autocorrelogram of squared observations is displayed. In this case, the first autocorrelations are large and significant, indicating GARCH effects; the Wald test strongly rejects the null hypothesis of the absence of autocorrelation in the squares. As an additional data analysis, we test for unit root using the test by Phillips and Perron [1988]. The test strongly rejects the $I(1)$ hypothesis. From this preliminary

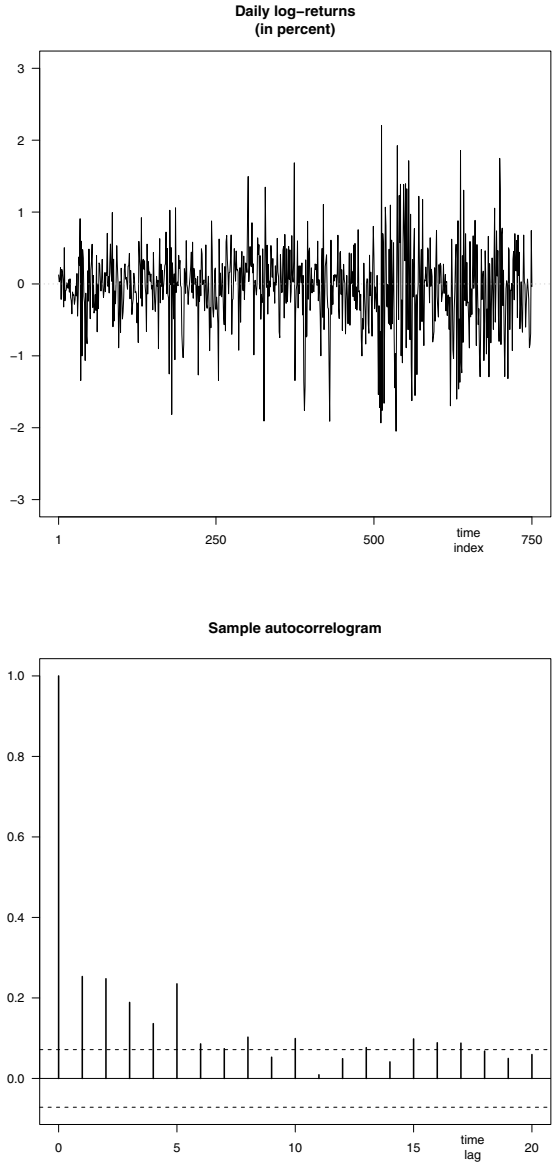

Fig. 3.1. DEM/GBP foreign exchange daily log-returns (upper graph) and sample autocorrelogram of the squared log-returns (lower graph).

analysis, we conclude that the time series is not integrated and does not exhibit autocorrelation. However, we strongly suspect the presence of GARCH effects in the data.

3.3.1 Model estimation

We fit the parsimonious GARCH(1, 1) model to the data for this observation window. As prior densities for the Bayesian estimation, we choose truncated Normal densities with zero mean vectors and diagonal covariance matrices. The variances are set to 10'000 so we do not introduce tight prior information in our estimation (see **Sect. 3.3.2** for a formal check). Finally, we recall that the joint prior is constructed by assuming prior independence between $\boldsymbol{\alpha}$ and β. We run two chains for 10'000 passes each. We emphasize the fact that only positivity constraints are implemented in the M-H algorithm, through the prior densities; no stationarity conditions are imposed in the simulation procedure. In addition, we estimate the model by the usual ML technique for comparison purposes.

In **Fig. 3.2**, the running means are plotted over iterations. For all parameters, we notice a convergence of the two chains toward a constant value after something like 5'000 iterations. As a formal check, we follow Gelman and Rubin [1992] where the authors elaborated the idea that the chain trajectories should be the same after convergence using analysis of variance techniques. Considering m parallel chains and a real function $\xi \doteq \xi(\psi)$ of the model parameters, there are m trajectories of length J given by $\{\xi_i^{[j]}\}_{j=1}^{J}$, $i = 1, \ldots, m$. The variances between chains and within chains, respectively, denoted by B and W, are then defined as follows:

$$B \doteq \frac{J}{m-1} \sum_{i=1}^{m} (\overline{\xi}_i - \overline{\xi})^2$$

$$W \doteq \frac{1}{m(J-1)} \sum_{i=1}^{m} \sum_{j=1}^{J} (\xi_i^{[j]} - \overline{\xi}_i)^2$$

where $\overline{\xi}_i$ is the average of observations of the ith chain and $\overline{\xi}$ is the average of these averages. After convergence, all these mJ values for ξ_i are drawn from the posterior distribution, and σ_ξ^2, the variance of ξ, can be consistently estimated by W, B as well as the following weighted average:

$$\widehat{\sigma}_\xi^2 \doteq \left(1 - \frac{1}{J}\right) W + \frac{1}{J} B \, .$$

If the chains have not yet converged, then initial values will still be influencing the trajectories and, due to their overdispersion, will force $\widehat{\sigma}_\xi^2$ to overestimate σ_ξ^2 until stationarity is reached. On the other hand, before convergence, W will tend to underestimate σ_ξ^2 because each chain will not have adequately traversed the complete state space. Following this reasoning, Gelman and Rubin [1992] construct an indicator of convergence; this is the estimator of *potential scale*

reduction factor given by:

$$\widehat{R} \doteq \sqrt{\frac{\widehat{\sigma}_{\xi}^2}{W}} \, .$$

As the simulation converges, the potential scale reduction declines to one, meaning that the m parallel chains are essentially overlapping. Gelman and Rubin [1992] suggests accepting convergence when the value of $\widehat{R}$ is below 1.2. Since this indicator is subject to estimation error, asymptotic confidence bands can be constructed and the 97.5th percentile is used as a conservative point estimate.

In our context, we test the convergence of the chains by using the following functions:

$$\xi(\psi) = \alpha_0 \quad , \quad \xi(\psi) = \alpha_1 \quad \text{and} \quad \xi(\psi) = \beta \, .$$

For these three functions, the diagnostic test by Gelman and Rubin [1992] does not lead to the rejection of the convergence if we consider the second half of the simulated values; the 97.5th percentile values for $\widehat{R}$ indeed belong to the interval $[1.04, 1.05]$. We can therefore be confident that the generated parameters are drawn from the joint posterior distribution.

Complementary analyses of the MCMC output are also worth mentioning at this point. In particular, we note that the one-lag autocorrelations in the chains range from 0.75 for parameter α_1 to 0.95 for β which is reasonable. Moreover, the sampling algorithm allows to reach very high acceptance rates ranging from 89% for vector $\boldsymbol{\alpha}$ to 95% for β, suggesting that the proposal densities are close to the full conditionals. On the basis of these results, we discard the first 5'000 draws from the overall MCMC output as a *burn-in* period and merge the two chains to get a final sample of length 10'000.

The posterior statistics as well as the ML results are reported in **Table 3.1**. First, we note that even though the number of observations is large, the ML estimates and the Bayesian posterior means are different; the ML point estimate is lower for components of vector $\boldsymbol{\alpha}$ and higher for parameter β. We also notice a difference between the 95% confidence intervals. Whereas the confidence band is symmetric in the ML case due to the asymptotic Normality assumption, this is not true for the posterior confidence intervals. The reason can be explained through **Fig. 3.3** where the marginal posterior densities of the parameters are displayed. We clearly notice the asymmetric shape of the histograms for parameters α_0 and α_1; the skewness values are 0.46 and 0.39, both significantly different from zero at the 1% significance level. Therefore the ML confidence band has a tendency to underestimate the right boundary of the 95% confidence interval for these parameters. In the case of parameter β, the skewness is -0.09, also significant; in this case, the Maximum Likelihood approach overestimates the

Table 3.1. Estimation results for the GARCH(1, 1) model with Normal innovations.★

ψ	ψ_{MLE}	$\overline{\psi}$	$\psi_{0.5}$	$\psi_{0.025}$	$\psi_{0.975}$	min	max	IF
α_0	0.039	0.048	0.047	0.022	0.080	0.011	0.119	9.79
	[0.014,0.064]	(0.448)						
α_1	0.198	0.226	0.223	0.128	0.337	0.083	0.499	5.85
	[0.102,0.294]	(1.284)						
β	0.686	0.636	0.636	0.476	0.795	0.338	0.849	40.79
	[0.538,0.833]	(5.021)						

★ ψ_{MLE}: Maximum Likelihood estimate; $\overline{\psi}$: posterior mean; ψ_ϕ: estimated posterior quantile at probability ϕ; min: minimum value; max: maximum value; IF: inefficiency factor (*i.e.*, ratio of the squared numerical standard error and the variance of the sample mean from a hypothetical *iid* sampler); [•]: Maximum Likelihood 95% confidence interval; (•): numerical standard error ($\times 10^3$). The posterior statistics are based on 10'000 draws from the joint posterior sample.

left boundary of the 95% confidence band. Moreover, as shown in the bottom right-hand side of the figure, the joint density of parameters α_0 and β is slightly different from the ellipsoid obtained with the asymptotic Normal approximation. Therefore, these results warn us against the abusive use of asymptotic justifications. In the present case, even 750 observations do not suffice to justify the asymptotic Normal approximation for the parameters estimates.

The last column of **Table 3.1** reports the inefficiency factors (IF) for the different parameters. Their values are computed as the ratio of the squared numerical standard error of the posterior sample and the variance estimate divided by the number of iterations (*i.e.*, the variance of the sample mean from a hypothetical *iid* sequence). The numerical standard errors are estimated by the method of Andrews [1991], using a Parzen kernel and AR(1) pre-whitening as presented in Andrews and Monahan [1992]. As noted by Deschamps [2006], this ensures easy, optimal, and automatic bandwidth selection. In our estimation, using 10'000 simulations out of the posterior distribution seems appropriate if we require that the Monte Carlo error in estimating the mean is smaller than 0.4% of the variation of the error due to the data. The larger inefficiency factor reported for parameter β is reflected in a larger autocorrelation in the simulated values.

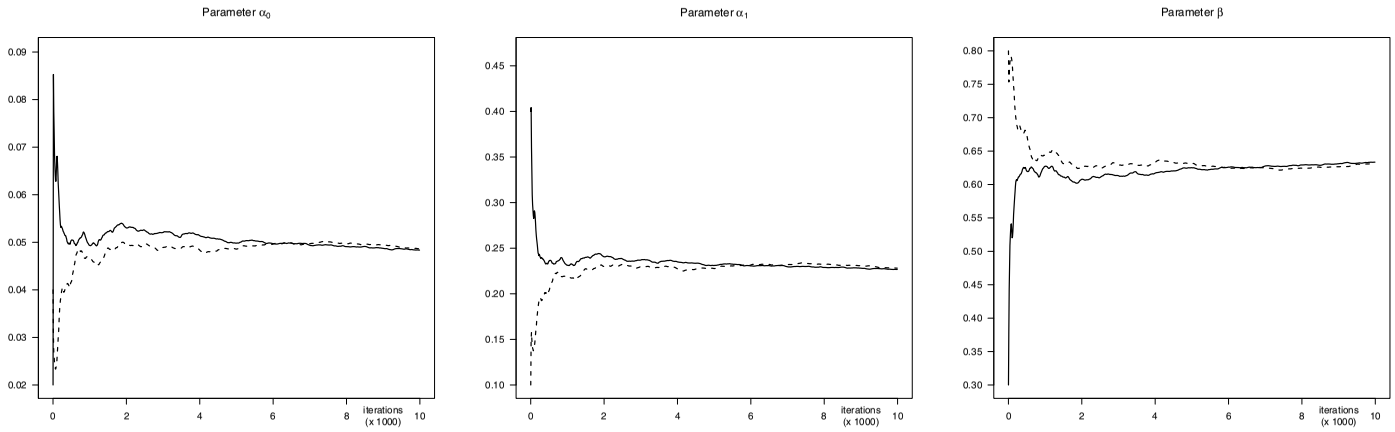

Fig. 3.2. Running means of the chains over iterations (up to 10'000). The acceptance rate ranges from 89% for vector $\boldsymbol{\alpha}$ to 95% for parameter β. The autocorrelations range from 0.75 for α_1 to 0.95 for β. The convergence diagnostic test by Gelman and Rubin [1992] indicates convergence of the chains from iteration 5'000; the 97.5th percentile of the potential reduction factor ranges from 1.04 to 1.05.

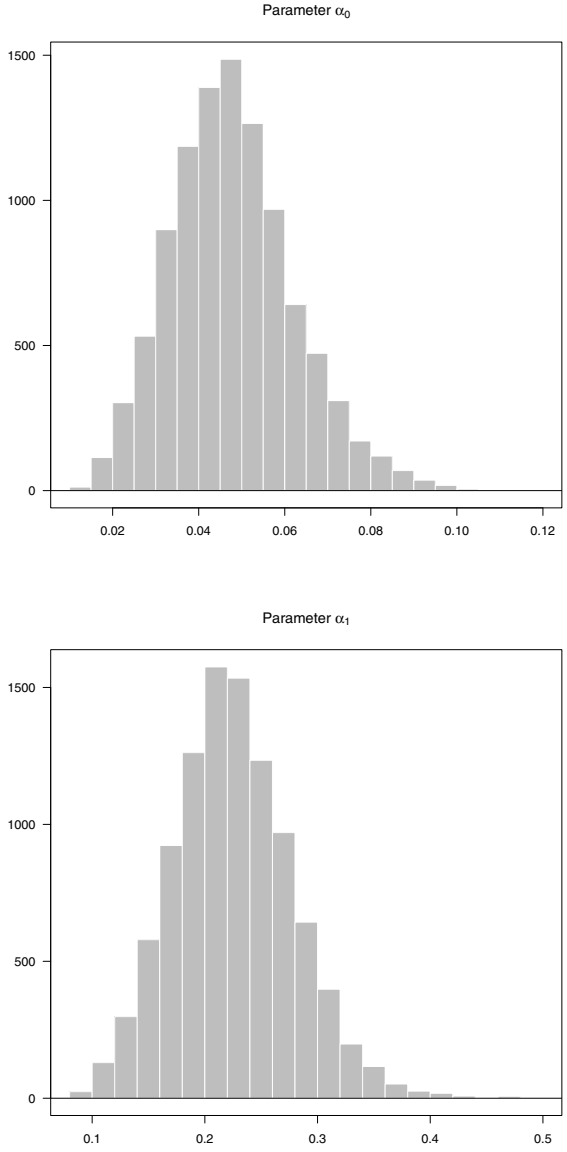

Fig. 3.3. Marginal posterior densities of the GARCH(1, 1) parameters; upper graph: parameter α_0; lower graph: parameter α_1. The histograms are based on 10'000 draws from the joint posterior sample.

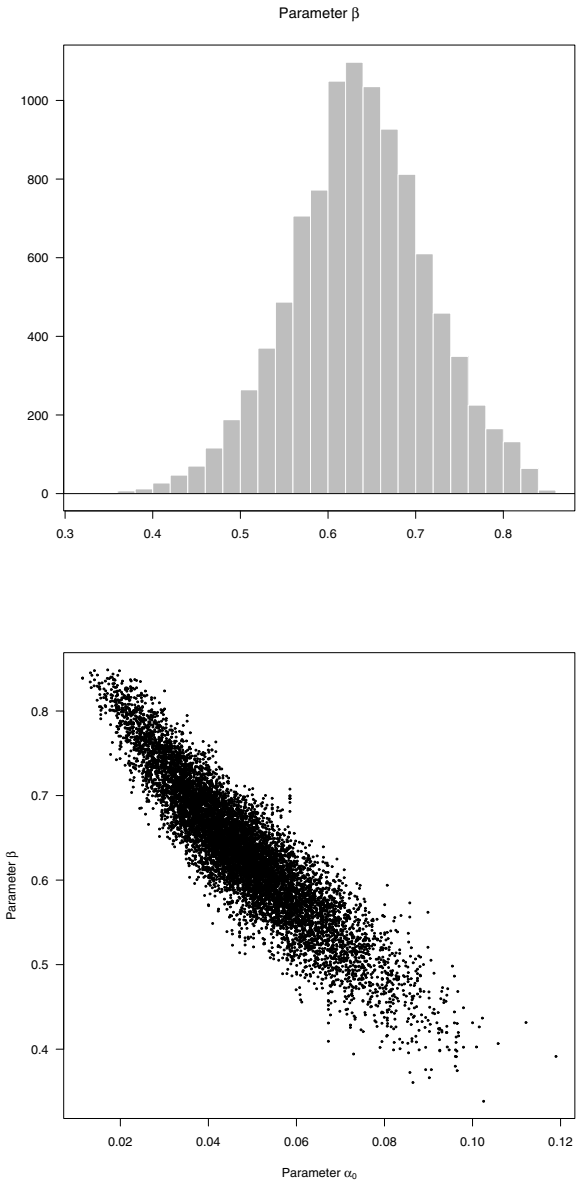

Fig. 3.3. (*cont.*) Marginal posterior densities of the GARCH$(1, 1)$ parameters; upper graph: parameter β; lower graph: scatter plot of (α_0, β). Both graphs are based on 10'000 draws from the joint posterior sample.

3.3.2 Sensitivity analysis

The Bayesian approach is often criticized on the grounds that the choice of the prior density may have a non negligible impact on the posterior density and, consequently, bias the posterior results. It is therefore important to determine the extent of this impact through a sensitivity analysis. To that aim, we follow Geweke [1999] who proposes a methodology to estimate the Bayes factors for the initial model against a model with an alternative prior. While the Bayes factor is a quantity which is often difficult to estimate, Geweke [1999, Sect.2] shows that it is possible to approximate the Bayes factor between two models differing only by their prior densities using the posterior simulation output from just one of the models. This approach provides an attractive way of performing sensitivity analysis since it does not require the estimation of the alternative model.

More precisely, let us denote by $p_{\mathrm{I}}(\psi)$ the initial prior density for $\psi \doteq (\boldsymbol{\alpha}, \beta)$ and by $p_{\mathrm{A}}(\psi)$ the alternative prior used to test the sensitivity of the posterior density. Based on the $T \times 1$ vector of observations $\mathbf{y} \doteq (y_1 \cdots y_T)'$, the Bayes factor in favor of the alternative model A over the initial model I can be expressed as follows:

$$\mathrm{BF}_{\mathrm{A}\succ\mathrm{I}} = \frac{p(\mathbf{y} \mid \mathrm{A})}{p(\mathbf{y} \mid \mathrm{I})}$$

where the marginal densities are found by integrating out the parameters:

$$p(\mathbf{y} \mid \bullet) = \int \mathcal{L}(\psi \mid \mathbf{y}) p_{\bullet}(\psi) \mathrm{d}\psi \; .$$

Developing the Bayes factor using the expression of the marginal densities yields:

$$\begin{aligned}
\mathrm{BF}_{\mathrm{A}\succ\mathrm{I}} &= \frac{\int \mathcal{L}(\psi \mid \mathbf{y}) p_{\mathrm{A}}(\psi) \mathrm{d}\psi}{\int \mathcal{L}(\psi \mid \mathbf{y}) p_{\mathrm{I}}(\psi) \mathrm{d}\psi} \\
&= \int \frac{\mathcal{L}(\psi \mid \mathbf{y}) p_{\mathrm{I}}(\psi) \frac{p_{\mathrm{A}}(\psi)}{p_{\mathrm{I}}(\psi)} \mathrm{d}\psi}{\int \mathcal{L}(\psi \mid \mathbf{y}) p_{\mathrm{I}}(\psi) \mathrm{d}\psi} \\
&= \int \frac{p_{\mathrm{A}}(\psi)}{p_{\mathrm{I}}(\psi)} \left[\frac{\mathcal{L}(\psi \mid \mathbf{y}) p_{\mathrm{I}}(\psi)}{\int \mathcal{L}(\psi \mid \mathbf{y}) p_{\mathrm{I}}(\psi) \mathrm{d}\psi} \right] \mathrm{d}\psi \\
&= \int \frac{p_{\mathrm{A}}(\psi)}{p_{\mathrm{I}}(\psi)} p(\psi \mid \mathbf{y}, \mathrm{I}) \mathrm{d}\psi \\
&= \mathbb{E}_{\psi|(\mathbf{y},\mathrm{I})} \left[\frac{p_{\mathrm{A}}(\psi)}{p_{\mathrm{I}}(\psi)} \right]
\end{aligned}$$

where the notation $\mathbb{E}_{\psi|(\mathbf{y},\mathrm{I})}$ emphasizes the fact that the posterior expectation is calculated with respect to the initial prior p_{I}. In this simple context, we thus notice that the Bayes factor is nothing else than the posterior expectation

under the initial prior of the ratio of prior densities. The posterior expectation can therefore be estimated using the joint posterior sample $\{\psi^{[j]}\}_{j=1}^{J}$ as follows:

$$\mathrm{BF}_{\mathrm{A}\succ\mathrm{I}} = \mathbb{E}_{\psi|(\mathbf{y},\mathrm{I})}\left[\frac{p_{\mathrm{A}}(\psi)}{p_{\mathrm{I}}(\psi)}\right] \approx \frac{1}{J}\sum_{j=1}^{J}\frac{p_{\mathrm{A}}(\psi^{[j]})}{p_{\mathrm{I}}(\psi^{[j]})}\ . \tag{3.6}$$

We test the sensitivity of our posterior results by considering three alternative priors which are truncated Normal densities as the initial prior. We choose however different hyperparameters, in particular larger variances in the covariance matrices. Formally, the alternative priors may be expressed as follows:

$$\begin{aligned} p(\boldsymbol{\alpha}) &\propto \mathcal{N}_2(\boldsymbol{\alpha} \mid \mu\,\boldsymbol{\iota}_2, \sigma^2 I_2)\mathbb{I}_{\{\boldsymbol{\alpha}>\mathbf{0}\}} \\ p(\beta) &\propto \mathcal{N}(\beta \mid \mu, \sigma^2)\mathbb{I}_{\{\beta>0\}} \end{aligned}$$

where $\boldsymbol{\iota}_2$ is a 2×1 vector of ones, I_2 is a 2×2 identity matrix, μ is the prior mean and σ^2 the prior variance; their values are given in the first two columns of **Table 3.2**. The Bayes factors are estimated using approximation (3.6) based on 10'000 draws from the joint posterior sample. The discrimination between models is then based on the Jeffrey's scale of evidence [see Kass and Raftery 1995, Sect.3.2] which can be summarized as follows:

- Strong evidence in favor of the initial prior compared to the alternative prior:
$$\mathrm{BF}_{\mathrm{A}\succ\mathrm{I}} < 0.1$$
- Moderate evidence in favor of the initial prior compared to the alternative prior:
$$0.1 \leqslant \mathrm{BF}_{\mathrm{A}\succ\mathrm{I}} < 0.3125$$
- Weak evidence in favor of the initial prior compared to the alternative prior:
$$0.3125 \leqslant \mathrm{BF}_{\mathrm{A}\succ\mathrm{I}} < 1\ .$$

Estimated BF are reported in the last column of **Table 3.2**. The numerical standard errors are not shown since their values are negligible. First, we note that a change in the prior mean has no impact on the BF. On the contrary, larger variances in the alternative covariance matrices diminishes the value of Bayes factors to 0.866; this indicates a weak evidence for the initial specification relative to the alternative priors. Therefore, for each alternative prior, the estimated BF confirms that our initial choice is vague enough and does not introduce significant information in our estimation.

Table 3.2. Results of the sensitivity analysis.★

Alternative priors		
μ	σ^2	BF
1.00	10'000	1.000
0.00	11'000	0.866
1.00	11'000	0.866

★ The alternative priors are truncated Normal densities; μ: prior mean; σ^2 prior variance; BF: Bayes factor.

3.3.3 Model diagnostics

We test the residuals for possible misspecification. The standardized residuals are defined by:

$$\widehat{\varepsilon}_t \doteq \frac{y_t}{\widehat{h}_t^{1/2}}$$

for $t = 1, \ldots, 750$, where $\widehat{h}_t$ is the conditional variance computed with $\psi_{0.5}$, the median of the posterior sample. If the statistical assumptions in (3.1) are satisfied, these residuals should be independent and Normally distributed asymptotically.

In the upper part of **Fig. 3.4**, we display the residuals over time. No autocorrelation or heteroscedasticity are visually apparent. We test for autocorrelation using the Ljung-Box test up to lag 20 [see Ljung and Box 1978]. The test does not reject the null hypothesis of absence of autocorrelation at the 5% significance level (p-value = 0.652). This is also true for the squared residuals (p-value = 0.961). Therefore, the GARCH(1, 1) process has been able to *filter* the heteroscedastic nature of the data. We form a quantile-quantile plot of the residuals against the Normal distribution in the lower graph of the figure. The distribution is almost Normal at its center whereas the tails are slightly fatter, especially the left one. The Kolmogorov-Smirnov Normality test rejects the null hypothesis at the 5% significance level (p-value = 0.008). The tails of the innovations' distribution are not fat enough to fully capture the distributional nature of the data. This point will be addressed in **Chap. 5** with the introduction of Student-t disturbances in the modeling.

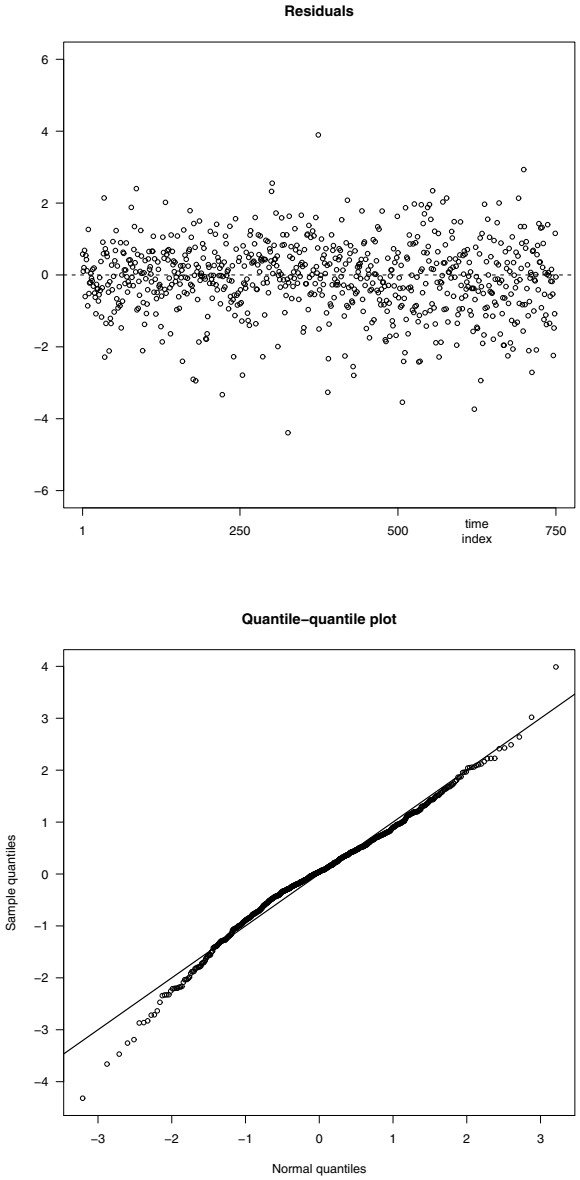

Fig. 3.4. Residuals time series (upper graph) and Normal quantile-quantile plot (lower graph).

3.4 Illustrative applications

In this section, we illustrate some probabilistic statements made possible under the Bayesian framework. The joint posterior sample is used to simulate nonlinear functions of the model parameters.

3.4.1 Persistence

As pointed out in **Sect. 3.2**, a GARCH(1, 1) process for $\{y_t\}$ is equivalent to an ARMA(1, 1) process for $\{y_t^2\}$ with an autoregressive coefficient $(\alpha_1 + \beta)$ and a moving average coefficient $-\beta$. Consequently, the autocorrelation function (henceforth ACF) of the squared observations comes from the standard formulae for the ARMA(1, 1) model. It is recursively given by:

$$\rho_i \doteq (\alpha_1 + \beta) \times \rho_{i-1}$$

for $i > 1$, where the first order autocorrelation is:

$$\rho_1 \doteq \frac{\alpha_1(1 - \beta^2 - \alpha_1\beta)}{1 - \beta^2 - 2\alpha_1\beta} \ .$$

The term $(\alpha_1 + \beta)$ is the degree of persistence in the autocorrelation of the squares which controls the intensity of the clustering in the variance process. With a value close to one, past shocks and past variances will have a longer impact on the future conditional variance. An autoregressive coefficient $(\alpha_1+\beta) = 1$ corresponds to a unit root process for squared observations.

To make inference on the persistence and ACF of the squared process, we simply use the posterior sample and generate $(\alpha_1^{[j]} + \beta^{[j]})$ as well as $\rho_i^{[j]}$ for $j = 1, \ldots, 10'000$ and $i = 1, \ldots, 20$. The posterior density of the persistence $(\alpha_1 + \beta)$ is plotted in the upper part of **Fig. 3.5**. The histogram is left-skewed with a median value of 0.865 and a maximum value of 0.992. In this case, the integration for the variance process is not supported by the data. In the lower part of the figure, we display the posterior ACF with its 95% and 99% confidence bands together with the sample autocorrelations of the squared observations. Although a single observation, at lag 11, lies outside the confidence bands, the autocorrelation structure of the estimated GARCH(1, 1) model is in line with the data.

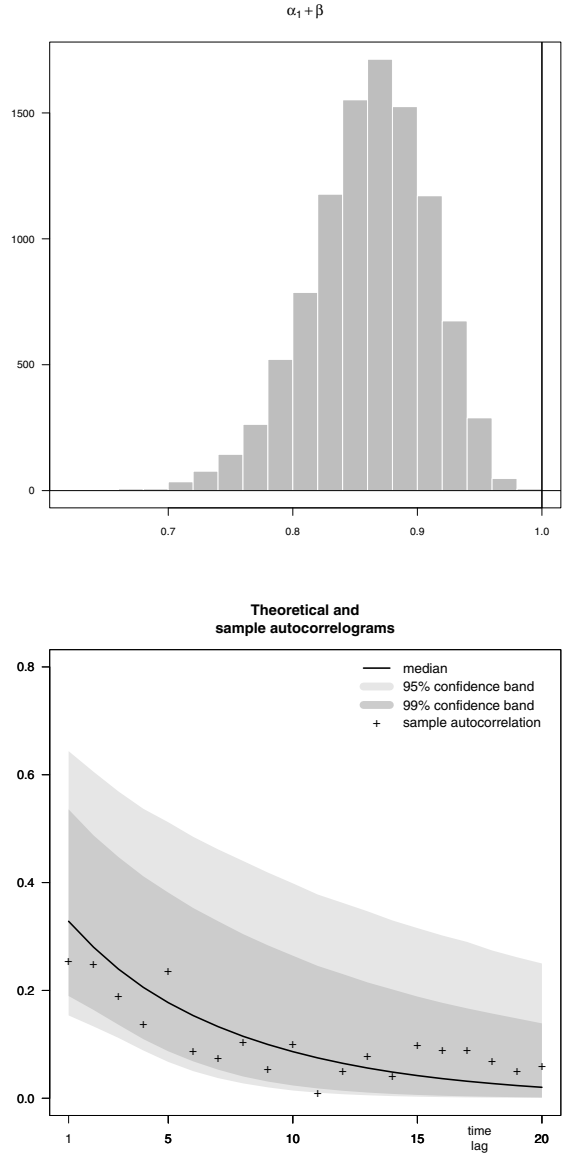

Fig. 3.5. Posterior density of the persistence (upper graph) and posterior autocorrelogram (lower graph) of the squared observations. Both graphs are based on 10'000 draws from the joint posterior sample.

3.4.2 Stationarity

In the case of the GARCH(1, 1) model with Normal innovations, Bollerslev [1986, Thm.1, p.310] and Nelson [1990, Thm.2, p.320] gave the conditions for covariance stationarity (CSC) and strict stationarity (SSC), respectively. These conditions are given by:

$$\begin{aligned} \text{CSC} &\doteq \alpha_1 + \beta - 1 < 0 \\ \text{SSC} &\doteq \mathbb{E}\big[\ln(\alpha_1 \varepsilon_t^2 + \beta)\big] < 0 \end{aligned}$$

where the error term ε_t is Normally distributed. As pointed out in **Sect. 3.3**, no stationarity condition has been imposed in the M-H algorithm. The joint posterior sample can therefore be used to estimate the posterior density of these functions by generating:

$$\begin{aligned} \text{CSC}^{[j]} &\doteq \alpha_1^{[j]} + \beta^{[j]} - 1 \\ \text{SSC}^{[j]} &\doteq \frac{1}{K} \sum_{k=1}^{K} \ln\left(\alpha_1^{[j]} (\eta^{[k]})^2 + \beta^{[j]}\right) \end{aligned}$$

for $j = 1, \ldots, 10'000$, where $\eta^{[k]}$ is a draw from a standard Normal distribution and K is set large enough (we choose $K = 1'000$ in our application). In **Fig. 3.6**, we present the Gaussian kernel density estimates of the posterior densities for CSC and SSC. As we can notice, none of these values exceed zero in our simulation study. Thus, the estimated model is covariance stationary and strictly stationary.

We conclude this section by noting that other probabilistic statements on interesting functions of the model parameters can be obtained using the joint posterior sample. For instance, the posterior median is 0.341 for the unconditional variance and 4.54 for the unconditional kurtosis. They approximately correspond to the sample estimations of 0.323 and 4.63.

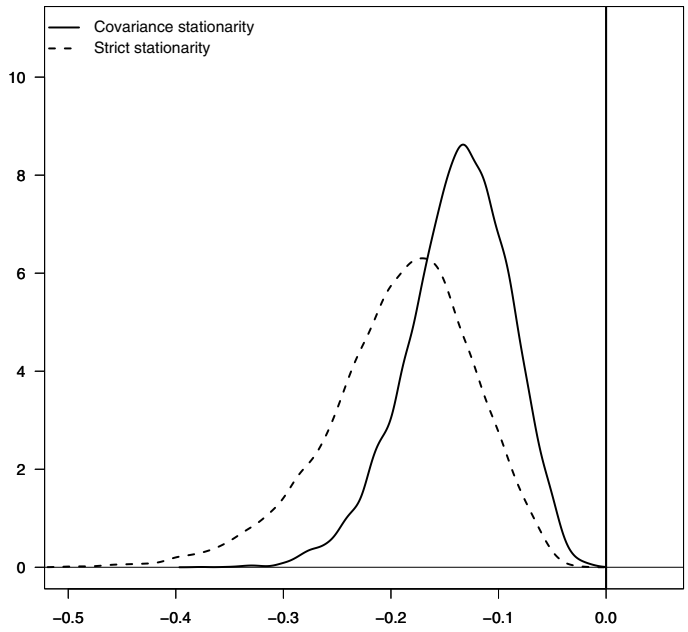

Fig. 3.6. Posterior densities of the covariance stationarity and strict stationarity conditions. Gaussian kernel density estimates with bandwidth selected by the "Silverman's rule of thumb" criterion [see Silverman 1986, p.48]. Both kernel density estimates are based on 10'000 draws from the joint posterior sample.

4

Bayesian Estimation of the Linear Regression Model with Normal-GJR(1, 1) Errors

"Overall, these results show a greater impact on volatility of negative, rather than positive return shocks."

— Robert F. Engle and Victor K. Ng

In this chapter, we propose the Bayesian estimation of the linear regression model with conditionally heteroscedastic errors. In the context of time series regressions, the regression part can include exogenous or lagged dependent variables. Moreover, we extend the traditional GARCH specification of the errors to account for asymmetric movements between the conditional variance and the underlying process. The volatility tends to rise more in response to bad news than to good news and this phenomenon is especially true on equity markets. This effect was first observed by Black [1976] and is referred to as the *leverage effect* in the financial literature. One explanation of this empirical fact is that negative returns increase financial leverage which extends the company's risk and therefore the variance. To cope with this stylized fact, we use the GJR model of Glosten *et al.* [1993]. In this setting, the conditional variance can react asymmetrically depending on the sign of the past shocks due to the introduction of dummy variables. The appealing aspect of this model is that it encompasses the symmetric GARCH. In addition, the MCMC scheme presented in **Sect. 3.2** can easily be extended for this asymmetric model in order to find proposal densities for the parameters. As a first illustration, we fit the model to the S&P100 index log-returns and compare the Bayesian and the Maximum Likelihood estimates. Next, we perform a prior sensitivity analysis and test the residuals for misspecification. Finally, we estimate the density of the unconditional variance of the process.

The plan of this chapter is as follows. We set up the model in **Sect. 4.1**. The MCMC scheme is detailed in **Sect. 4.2**. The empirical results are presented in **Sect. 4.3**. We conclude with some illustrations of the Bayesian approach in **Sect. 4.4**.

4.1 The model and the priors

A linear regression model with Normal-GJR(1,1) errors may be written as follows:

$$\begin{aligned} y_t &= \mathbf{x}_t'\boldsymbol{\gamma} + u_t \quad \text{for } t = 1, \ldots, T \\ u_t &= \varepsilon_t h_t^{1/2} \\ \varepsilon_t &\overset{iid}{\sim} \mathcal{N}(0,1) \\ h_t &\doteq \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}}) u_{t-1}^2 + \beta h_{t-1} \end{aligned} \tag{4.1}$$

where $\alpha_0 > 0$, $\alpha_1 \geqslant 0$, $\alpha_2 \geqslant 0$ and $\beta \geqslant 0$ to ensure a positive conditional variance and $h_0 = y_0 \doteq 0$ for convenience; y_t is a scalar dependent variable; $\mathbf{x}_t$ is a $m \times 1$ vector of exogenous or lagged dependent variables; $\boldsymbol{\gamma}$ is a $m \times 1$ vector of regression coefficients; $\mathcal{N}(0,1)$ is the standard Normal density. In this setting, the conditional variance h_t is a linear function of the squared past shock and the past variance but contrary to the GARCH model, the conditional variance can react asymmetrically to past shocks depending on their signs. The leverage effect is present if $\alpha_2 > \alpha_1$ so that the conditional variance is higher after a negative shock than a positive shock.

In order to write the likelihood function, we define the vectors $\mathbf{y} \doteq (y_1 \cdots y_T)'$ and $\boldsymbol{\alpha} \doteq (\alpha_0 \ \alpha_1 \ \alpha_2)'$ as well as the $T \times m$ matrix X whose tth row is given by $\mathbf{x}_t'$. We regroup the model parameters into $\psi \doteq (\boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta)$ for notational purposes and define the $T \times T$ diagonal matrix:

$$\Sigma \doteq \Sigma(\psi) = \operatorname{diag}\big(\{h_t(\psi)\}_{t=1}^T\big)$$

where:

$$\begin{aligned} h_t(\psi) &\doteq \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1}(\boldsymbol{\gamma}) \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1}(\boldsymbol{\gamma}) < 0\}}) u_{t-1}^2(\boldsymbol{\gamma}) + \beta h_{t-1}(\psi) \\ u_t(\boldsymbol{\gamma}) &\doteq y_t - \mathbf{x}_t'\boldsymbol{\gamma} \ . \end{aligned}$$

Then, we regroup the error terms $u_t(\boldsymbol{\gamma})$ into the $T \times 1$ vector $\mathbf{u} \doteq (u_1 \cdots u_T)'$ and express the likelihood function of ψ as follows:

$$\mathcal{L}(\psi \mid \mathbf{y}, \mathrm{X}) \propto (\det \Sigma)^{-1/2} \exp\left[-\tfrac{1}{2}\mathbf{u}'\Sigma^{-1}\mathbf{u}\right] \ . \tag{4.2}$$

We propose the following proper priors on the parameters $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and β of the preceding model:

$$
\begin{aligned}
p(\boldsymbol{\gamma}) &= \mathcal{N}_m(\boldsymbol{\gamma} \mid \boldsymbol{\mu}_\gamma, \Sigma_\gamma) \\
p(\alpha) &\propto \mathcal{N}_3(\boldsymbol{\alpha} \mid \boldsymbol{\mu}_\alpha, \Sigma_\alpha)\mathbb{I}_{\{\boldsymbol{\alpha}>\mathbf{0}\}} \\
p(\beta) &\propto \mathcal{N}(\beta \mid \mu_\beta, \Sigma_\beta)\mathbb{I}_{\{\beta>0\}} \ .
\end{aligned}
$$

where $\mu_\bullet$ and $\Sigma_\bullet$ are the hyperparameters, $\mathbf{0}$ is a 3×1 vector of zeros, $\mathbb{I}_{\{\bullet\}}$ is the indicator function and $\mathcal{N}_d$ is the d-dimensional Normal density ($d > 1$). In addition, we assume prior independence between parameters $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and β which yields the following joint prior:

$$
p(\psi) = p(\boldsymbol{\gamma})p(\boldsymbol{\alpha})p(\beta) \ .
$$

Then, we construct the joint posterior density via Bayes' rule:

$$
p(\psi \mid \mathbf{y}, \mathrm{X}) \propto \mathcal{L}(\psi \mid \mathbf{y}, \mathrm{X})p(\psi) \ .
$$

4.2 Simulating the joint posterior

As in the GARCH model of **Chap. 3**, the recursive nature of the variance equation does not allow for conjugacy between the likelihood function and the joint prior density. Hence, we rely again on the M-H algorithm to draw samples from the joint posterior distribution. We draw an initial value $\psi^{[0]} \doteq (\boldsymbol{\gamma}^{[0]}, \boldsymbol{\alpha}^{[0]}, \beta^{[0]})$ from the joint prior and we generate iteratively J passes for ψ. A single pass is decomposed as follows:

$$
\begin{aligned}
\boldsymbol{\gamma}^{[j]} &\sim p(\boldsymbol{\gamma} \mid \boldsymbol{\alpha}^{[j-1]}, \beta^{[j-1]}, \mathbf{y}, \mathrm{X}) \\
\boldsymbol{\alpha}^{[j]} &\sim p(\boldsymbol{\alpha} \mid \boldsymbol{\gamma}^{[j]}, \beta^{[j-1]}, \mathbf{y}, \mathrm{X}) \\
\beta^{[j]} &\sim p(\beta \mid \boldsymbol{\gamma}^{[j]}, \boldsymbol{\alpha}^{[j]}, \mathbf{y}, \mathrm{X}) \ .
\end{aligned}
$$

Since no full conditional density is known analytically, we sample the parameters $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and β from three proposal densities.

4.2.1 Generating vector $\boldsymbol{\gamma}$

The proposal density to sample the $m \times 1$ vector $\boldsymbol{\gamma}$ is obtained by combining the likelihood function (4.2) and the prior density by the usual Bayes update:

$$
q_\gamma(\boldsymbol{\gamma} \mid \widetilde{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \beta, \mathbf{y}, \mathrm{X}) = \mathcal{N}_m(\boldsymbol{\gamma} \mid \widehat{\boldsymbol{\mu}}_\gamma, \widehat{\Sigma}_\gamma)
$$

with:

$$\widehat{\Sigma}_\gamma^{-1} \doteq \mathrm{X}'\widetilde{\Sigma}^{-1}\mathrm{X} + \Sigma_\gamma^{-1}$$
$$\widehat{\boldsymbol{\mu}}_\gamma \doteq \widehat{\Sigma}_\gamma(\mathrm{X}'\widetilde{\Sigma}^{-1}\mathbf{y} + \Sigma_\gamma^{-1}\boldsymbol{\mu}_\gamma)$$

where the $T \times T$ diagonal matrix $\widetilde{\Sigma} \doteq \mathrm{diag}\big(\{h_t(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \beta)\}_{t=1}^T\big)$ and $\widetilde{\boldsymbol{\gamma}}$ is the previous draw of $\boldsymbol{\gamma}$ in the M-H sampler. A candidate $\boldsymbol{\gamma}^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\gamma}^\star, \boldsymbol{\alpha}, \beta \mid \mathbf{y}, \mathrm{X})}{p(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \beta \mid \mathbf{y}, \mathrm{X})} \frac{q_\gamma(\widetilde{\boldsymbol{\gamma}} \mid \boldsymbol{\gamma}^\star, \boldsymbol{\alpha}, \beta, \mathbf{y}, \mathrm{X})}{q_\gamma(\boldsymbol{\gamma}^\star \mid \widetilde{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \beta, \mathbf{y}, \mathrm{X})}, 1\right\} .$$

4.2.2 Generating the GJR parameters

The proposal densities to generate the parameters $\boldsymbol{\alpha}$ and β are obtained in the same manner as in **Sect. 3.2**. However, since we have a regression term which appears in the model, we estimate the GJR parameters from the errors $u_t \doteq y_t - \mathbf{x}_t'\boldsymbol{\gamma}$ instead of y_t. An approximate likelihood function for $(\boldsymbol{\alpha}, \beta)$ is then constructed from the process $\{u_t^2\}$. Note that in the case of a GJR model for $\{u_t\}$, we do not end up with an ARMA process for $\{u_t^2\}$ as in the GARCH model since we have two dummy variables which appear in the expression of the conditional variance. Indeed, by defining $w_t \doteq u_t^2 - h_t$, we can transform the expression of the conditional variance as follows:

$$\begin{aligned} h_t &= \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}}) u_{t-1}^2 + \beta h_{t-1} \\ \Leftrightarrow u_t^2 &= \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}} + \beta) u_{t-1}^2 - \beta w_{t-1} + w_t \end{aligned}$$

where w_t can be written as $w_t = (\chi_1^2 - 1)h_t$, χ_1^2 denoting a Chi-squared variable with one degree of freedom. As in the GARCH case, the sequence $\{w_t\}$ is a Martingale Difference process where the variable w_t has a conditional mean of zero and a conditional variance of $2h_t^2$.

Following the methodology of **Sect. 3.2**, we approximate the variable w_t by a variable z_t which is Normally distributed with a mean of zero and a variance of $2h_t^2$. This leads to the following *auxiliary* model:

$$u_t^2 = \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}} + \beta) u_{t-1}^2 - \beta z_{t-1} + z_t .$$

Then, by noting that z_t is a function of $(\boldsymbol{\alpha}, \beta)$ given by:

$$\begin{aligned} z_t(\boldsymbol{\alpha}, \beta) = u_t^2 &- \alpha_0 - (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}} + \beta) u_{t-1}^2 \\ &+ \beta z_{t-1}(\boldsymbol{\alpha}, \beta) \end{aligned} \tag{4.3}$$

and by defining the the $T \times 1$ vector $\mathbf{z} \doteq (z_1 \cdots z_T)'$ as well as the $T \times T$ diagonal matrix:

$$\Lambda \doteq \Lambda(\boldsymbol{\alpha}, \beta) = \operatorname{diag}\big(\{2h_t^2(\boldsymbol{\alpha}, \beta)\}_{t=1}^T\big)$$

we can express the approximate likelihood function of $(\boldsymbol{\alpha}, \beta)$ as follows:

$$\mathcal{L}(\boldsymbol{\alpha}, \beta \mid \boldsymbol{\gamma}, \mathbf{y}, \mathrm{X}) \propto (\det \Lambda)^{-1/2} \exp\left[-\tfrac{1}{2}\mathbf{z}'\Lambda^{-1}\mathbf{z}\right] \ .$$

As will be shown hereafter, the construction of the proposal densities for parameters $\boldsymbol{\alpha}$ and β is based on this approximate likelihood function.

Generating vector $\boldsymbol{\alpha}$

We extend the recursive transformations presented in **Sect. 3.2.1** to express the function $z_t(\boldsymbol{\alpha}, \beta)$ in (4.3) as a linear function of the 3×1 vector $\boldsymbol{\alpha}$. Since the conditional variance is an asymmetric function of past errors, we have to distinguish between positive and negative shocks in the recursive transformations. Let us define $v_t \doteq u_t^2$ for notational convenience. The appropriate recursive transformations are then defined as follows:

$$\begin{aligned} l_t^* &\doteq 1 + \beta\, l_{t-1}^* \\ v_t^* &\doteq v_{t-1}\mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \beta\, v_{t-1}^* \\ v_t^{**} &\doteq v_{t-1}\mathbb{I}_{\{u_{t-1} < 0\}} + \beta\, v_{t-1}^{**} \end{aligned}$$

where $l_0^* = v_0^* = v_0^{**} \doteq 0$. As shown in **Prop. A.2** (see **App. A**), upon defining the 3×1 vector $\mathbf{c}_t \doteq (l_t^*\ v_t^*\ v_t^{**})'$, the function z_t can be expressed as $z_t = v_t - \mathbf{c}_t'\boldsymbol{\alpha}$. Then, by considering the $T \times 1$ vector $\mathbf{v} \doteq (v_1 \cdots v_T)'$ as well as the $T \times 3$ matrix C whose tth row is $\mathbf{c}_t'$, it turns out that $\mathbf{z} = \mathbf{v} - C\boldsymbol{\alpha}$. Therefore, we can express the approximate likelihood function of parameter $\boldsymbol{\alpha}$ as follows:

$$\mathcal{L}(\boldsymbol{\alpha} \mid \boldsymbol{\gamma}, \beta, \mathbf{y}, \mathrm{X}) \propto (\det \Lambda)^{-1/2} \exp\left[-\tfrac{1}{2}(\mathbf{v} - C\boldsymbol{\alpha})'\Lambda^{-1}(\mathbf{v} - C\boldsymbol{\alpha})\right] \ .$$

The proposal density to sample vector $\boldsymbol{\alpha}$ is obtained by combining this likelihood function and the prior density by Bayes' update:

$$q_\alpha(\boldsymbol{\alpha} \mid \boldsymbol{\gamma}, \widetilde{\boldsymbol{\alpha}}, \beta, \mathbf{y}, \mathrm{X}) \propto \mathcal{N}_3(\boldsymbol{\alpha} \mid \widehat{\boldsymbol{\mu}}_\alpha, \widehat{\Sigma}_\alpha)\mathbb{I}_{\{\boldsymbol{\alpha} > \mathbf{0}\}}$$

with:

$$\begin{aligned} \widehat{\Sigma}_\alpha^{-1} &\doteq C'\widetilde{\Lambda}^{-1}C + \Sigma_\alpha^{-1} \\ \widehat{\boldsymbol{\mu}}_\alpha &\doteq \widehat{\Sigma}_\alpha(C'\widetilde{\Lambda}^{-1}\mathbf{v} + \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha) \end{aligned}$$

where the $T \times T$ diagonal matrix $\widetilde{\Lambda} \doteq \text{diag}\big(\{2h_t^2(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\alpha}}, \beta)\}_{t=1}^T\big)$ and $\widetilde{\boldsymbol{\alpha}}$ is the previous draw of $\boldsymbol{\alpha}$ in the M-H sampler. A candidate $\boldsymbol{\alpha}^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\gamma}, \boldsymbol{\alpha}^\star, \beta \mid \mathbf{y}, \mathrm{X})}{p(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\alpha}}, \beta \mid \mathbf{y}, \mathrm{X})} \frac{q_\alpha(\widetilde{\boldsymbol{\alpha}} \mid \boldsymbol{\gamma}, \boldsymbol{\alpha}^\star, \beta, \mathbf{y}, \mathrm{X})}{q_\alpha(\boldsymbol{\alpha}^\star \mid \boldsymbol{\gamma}, \widetilde{\boldsymbol{\alpha}}, \beta, \mathbf{y}, \mathrm{X})}, 1\right\} .$$

Generating parameter $\boldsymbol{\beta}$

The methodology is the same as the one presented in **Sect. 3.2.2**. The function $z_t(\beta)$ given by:

$$z_t(\beta) = u_t^2 - \alpha_0 - (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}} + \beta) u_{t-1}^2 + \beta z_{t-1}(\beta)$$

is approximated by a first order Taylor expansion at point $\widetilde{\beta}$, the previous draw of parameter β in the M-H sampler. The proposal density $q_\beta(\beta \mid \bullet)$ is then obtained by combining the approximate likelihood of β and the prior density via Bayes' update. A candidate $\beta^\star$ is generated from this density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta^\star \mid \mathbf{y}, \mathrm{X})}{p(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \widetilde{\beta} \mid \mathbf{y}, \mathrm{X})} \frac{q_\beta(\widetilde{\beta} \mid \boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta^\star, \mathbf{y}, \mathrm{X})}{q_\beta(\beta^\star \mid \boldsymbol{\gamma}, \boldsymbol{\alpha}, \widetilde{\beta}, \mathbf{y}, \mathrm{X})}, 1\right\} .$$

Finally, we conclude this section by noting that the validity of the algorithm and the correctness of the computer code are verified by the methodology detailed at the end of **Sect. 3.2.2**.

4.3 Empirical analysis

We apply our Bayesian estimation method to daily observations of the Standard & Poors 100 (henceforth S&P100) index log-returns. The sample period is from January 2, 1990, to December 17, 1992, for a total of 750 observations. The log-returns are expressed in percent. The S&P100 index is one of the major benchmarks of U.S. equity performance. It consists of one hundred stocks in a representative sample of leading companies chosen for market size, liquidity and industry group. We choose this data set since it is an equity index and is therefore susceptible to exhibit leverage effects. Moreover, a volatility index of the S&P100 index, the VIX, is computed by the Chicago Board of Exchange. This index aims to give a fear gauge to investors and can be viewed as a proxy for the conditional variance. This volatility index gained particular attention in recent years as it provides an interesting asset for hedging downside market movements.

The two data sets are freely available from `http://www.finance.yahoo.com`. Note that in 2003, the formula for the VIX's calculation has been modified and the underlying index has changed from the S&P100 index to the S&P500 index. This is of no consequence here since we consider the old VIX definition which is based on the S&P100 index.

The S&P100 index log-returns are displayed in the upper part of **Fig. 4.1**. In the lower part, the VIX index is plotted for the same time period. The correlation between log-returns and gross rates of squared VIX, which can be viewed as a proxy for the variance, is -0.52, indeed suggesting the presence of the leverage effect.

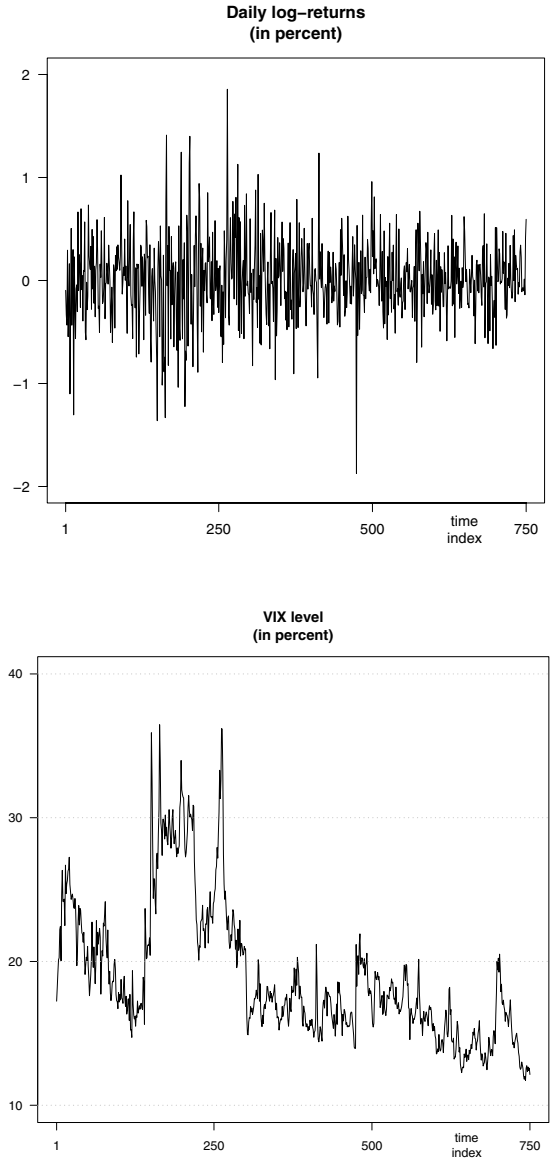

Fig. 4.1. S&P100 index log-returns (upper graph) and VIX level (lower graph).

4.3.1 Model estimation

As noted by Campbell, Lo, and MacKinlay [1996, p.104] for instance, financial time series such as equity indices sometimes present positive first order autocorrelation. This effect is stronger for high-frequency data such as intra-daily

or even daily time series due to market micro-structures. Based on that evidence, we estimate the model (4.1) with a constant term and an autoregressive parameter:

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + u_t \tag{4.4}$$

where the errors $\{u_t\}$ are modeled by the Normal-GJR$(1,1)$ process introduced in (4.1). As a prior density for the regression parameters, we choose a bi-dimensional Normal density. In the case of the GJR parameters, the priors are truncated Normal densities. Both priors have zero mean vectors and diagonal covariance matrices whose variances are set to 10'000 so we do not introduce tight prior information in our estimation (see **Sect. 4.3.2** for a formal check). Finally, we recall that the joint prior is constructed by assuming prior independence between $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and β.

We run two chains for 10'000 passes each where only positivity constraints for the GJR parameters are implemented in the M-H algorithm, through the prior densities. We test the convergence of the chains using the diagnostic test by Gelman and Rubin [1992]. The convergence diagnostic based on the two chains shows no evidence against convergence of the sampler for the last 5'000 iterations (the values of the 97.5th percentile of the potential scale reduction factor ranges from 1.01 to 1.09). The one-lag autocorrelations in the chains range from 0.30 for parameter γ_1 to 0.96 for β. The sampling algorithm allows to reach acceptance rates of 66% for vector $\boldsymbol{\alpha}$, 77% for $\boldsymbol{\gamma}$ and 95% for β. From the overall MCMC output, we discard the first 5'000 draws and merge the two chains to get a final sample of length 10'000. In addition, we estimate the model by the usual ML technique for comparison purposes.

The posterior statistics as well as the ML results are reported in **Table 4.1**. First, we note a difference between the Bayesian and ML point estimates for the GJR parameters; the ML estimates are lower for the components of vector $\boldsymbol{\alpha}$ and higher for parameter β. In the case of vector $\boldsymbol{\gamma}$, the difference between the Bayesian and the classical approaches is more pronounced for the component γ_1. For both components of $\boldsymbol{\gamma}$, the posterior 95% confidence bands are centered around zero which suggests a zero expectation for y_t and no need to model an autoregressive component for this data set. Second, we note strong asymmetries in the marginal posterior densities for the parameters $\boldsymbol{\alpha}$ and β, as shown in **Fig. 4.2**. The marginal posteriors for components of vector $\boldsymbol{\alpha}$ are right-skewed while left-skewed for parameter β. In the ML case, the asymptotic Normal approximation of the parameter estimates leads to negative values for the left boundary of α_0 and α_1. Caution is therefore in order when applying the asymptotic Normal approximation in this context. Finally, the values of the

inefficiency factor (IF) reported in the last column of **Table 4.1** indicate that using 10'000 simulations is appropriate if we require that the Monte Carlo error in estimating the mean is smaller than 0.33% of the variation of the error due to the data. The larger inefficiency factor reported for parameter β is reflected in a larger autocorrelation in the simulated values.

Table 4.1. Estimation results for the linear regression model with Normal-GJR(1, 1) errors.★

ψ	ψ_{MLE}	$\overline{\psi}$	$\psi_{0.5}$	$\psi_{0.025}$	$\psi_{0.975}$	min	max	IF
γ_0	0.001	0.002	0.002	-0.025	0.028	-0.048	0.047	1.92
	[-0.025,0.028]	(0.190)						
γ_1	0.003	0.004	0.004	-0.075	0.084	-0.160	0.163	1.84
	[-0.076,0.082]	(0.546)						
α_0	0.018	0.023	0.022	0.012	0.041	0.006	0.071	16.52
	[-0.008,0.028]	(0.299)						
α_1	0.022	0.038	0.034	0.002	0.096	0.000	0.193	3.45
	[-0.020,0.064]	(0.463)						
α_2	0.155	0.180	0.175	0.101	0.286	0.058	0.393	3.96
	[0.077,0.234]	(0.939)						
β	0.799	0.750	0.754	0.620	0.853	0.423	0.918	33.62
	[0.713,0.885]	(3.45)						

★ ψ_{MLE}: Maximum Likelihood estimate; $\overline{\psi}$: posterior mean; ψ_ϕ: estimated posterior quantile at probability ϕ; min: minimum value; max: maximum value; IF: inefficiency factor (*i.e.*, ratio of the squared numerical standard error and the variance of the sample mean from a hypothetical *iid* sampler); [•]: Maximum Likelihood 95% confidence interval; (•): numerical standard error ($\times 10^3$). The posterior statistics are based on 10'000 draws from the joint posterior sample.

Given the expression of the scedastic function in (4.1), the leverage effect can be measured by $\triangle\alpha \doteq (\alpha_2 - \alpha_1)$. The posterior density of $\triangle\alpha$ is displayed in **Fig. 4.3**. The mean value is 0.142 and the median is 0.139, indicating a stronger impact of negative shocks to the conditional variance, as expected. Using the posterior sample, we can also estimate the probability of the presence of the leverage effect, *i.e.*, $\mathbb{P}(\triangle\alpha > 0 \mid \mathbf{y}, \mathrm{X})$. The estimation gives a probability value of 0.998 with a 95% confidence band of [0.9976,0.9996]. Therefore, the data strongly support the presence of the leverage effect for the S&P100 index.

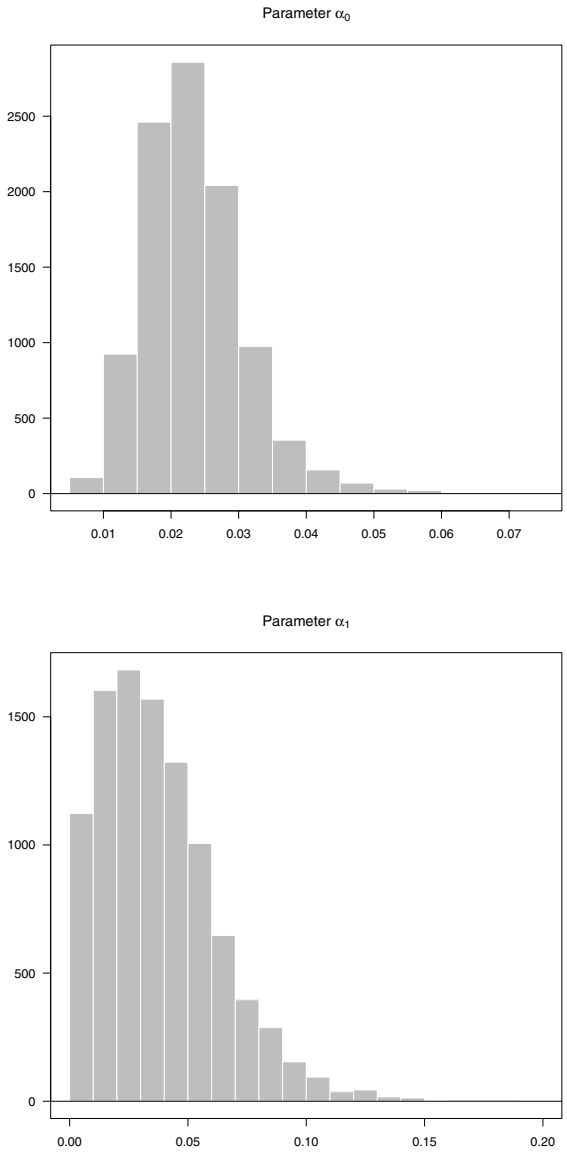

Fig. 4.2. Marginal posterior densities of the GJR$(1, 1)$ parameters; upper graph: parameter α_0; lower graph: parameter α_1. The histograms are based on 10'000 draws from the joint posterior sample.

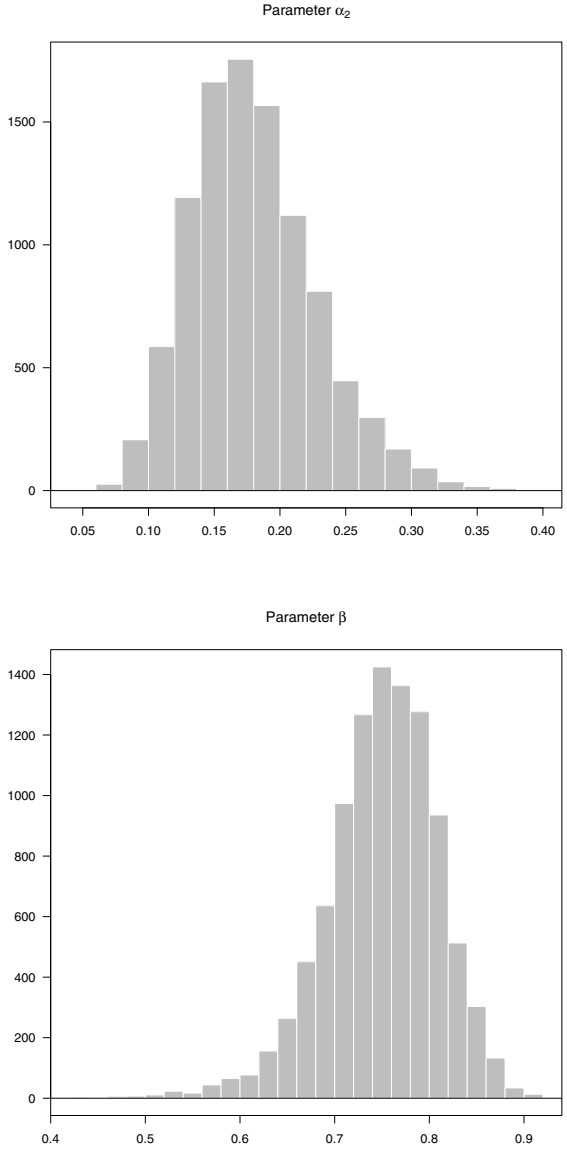

Fig. 4.2. (*cont.*) Marginal posterior densities of the GJR(1, 1) parameters; upper graph: parameter α_2; lower graph: parameter β. The histograms are based on 10'000 draws from the joint posterior sample.

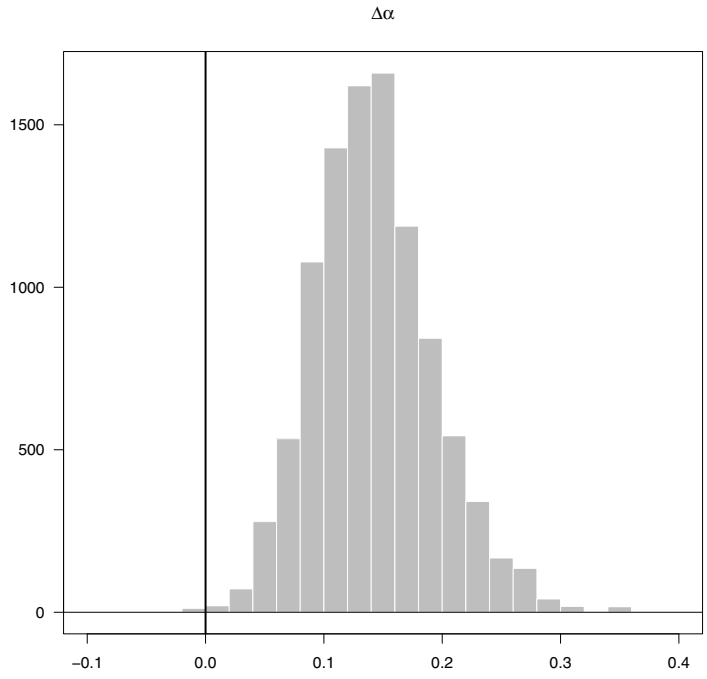

Fig. 4.3. Posterior density of the leverage effect parameter $\triangle\alpha \doteq (\alpha_2 - \alpha_1)$. The vertical line is set at $\triangle\alpha = 0$. The histogram is based on 10'000 draws from the joint posterior sample.

4.3.2 Sensitivity analysis

As in **Sect. 3.3.2**, we test the sensitivity of our posterior results with respect to the choice of the prior density. We consider three alternative priors by either modifying the mean and/or increasing the variance relative to our initial prior. Formally, the alternative prior densities can be expressed as follows:

$$
\begin{aligned}
p(\boldsymbol{\gamma}) &\propto \mathcal{N}_2(\boldsymbol{\gamma} \mid \mu\, \boldsymbol{\iota}_2, \sigma^2 I_2) \\
p(\boldsymbol{\alpha}) &\propto \mathcal{N}_3(\boldsymbol{\alpha} \mid \mu\, \boldsymbol{\iota}_3, \sigma^2 I_3) \mathbb{I}_{\{\boldsymbol{\alpha}>0\}} \\
p(\beta) &\propto \mathcal{N}(\beta \mid \mu, \sigma^2) \mathbb{I}_{\{\beta>0\}}
\end{aligned}
$$

where $\boldsymbol{\iota}_d$ is a $d \times 1$ vector of ones, I_d is a $d \times d$ identity matrix, μ is the prior mean and σ^2 the prior variance.

The sensitivity results are reported in **Table 4.2**; the first two columns give the hyperparameters' values of the alternative priors while the last column report the estimated Bayes factors. In all cases, the Bayes factors belong to the interval $[0.3125, 1]$ which implies a weak evidence in favor of our initial specification relative to the alternative priors. This indicates that our initial prior is vague enough and does not introduce significant information in our estimation.

Table 4.2. Results of the sensitivity analysis.★

Alternative priors		
μ	σ^2	BF
1.00	10'000	0.999
0.00	11'000	0.751
1.00	11'000	0.751

★ The alternative priors are (truncated) Normal densities; μ prior mean; σ^2 prior variance; BF: Bayes factor.

4.3.3 Model diagnostics

We test the residuals for possible misspecification. The standardized residuals are defined by:

$$
\widehat{\varepsilon}_t \doteq \frac{y_t - \mathbf{x}_t' \widehat{\boldsymbol{\gamma}}}{\widehat{h}_t^{1/2}}
$$

for $t = 1, \ldots, 750$, where $\widehat{\boldsymbol{\gamma}}$ is the posterior median of vector $\boldsymbol{\gamma}$ and $\widehat{h}_t$ is the conditional variance computed with the median of the posterior sample. If the

statistical assumptions in (4.1) are satisfied, these residuals should be independent and Normally distributed asymptotically.

We test the residuals for autocorrelation using the Ljung-Box test up to lag 20 [see Ljung and Box 1978]. The test does not reject the null hypothesis of the absence of autocorrelation at the 5% significance level (p-value = 0.365). This is also true for the squared residuals (p-value = 0.780). The Kolmogorov-Smirnov Normality test does not reject the null hypothesis at the 5% significance level with a p-value of 0.0514. On the contrary, the Jarque-Bera Normality test strongly rejects the null. Hence, while the model is able to filter the heteroscedasticity, it is not flexible enough to account for the high kurtosis of the residuals. This point will be addressed in **Chap. 5** with the introduction of Student-t errors in the modeling.

4.4 Illustrative applications

We end this chapter with the estimation of the unconditional variance of the underlying process. Under model specification (4.4), the process is covariance stationary if the following conditions are satisfied:

$$\begin{aligned} \mathrm{CSC}_1 &\doteq \gamma_1^2 - 1 < 0 \\ \mathrm{CSC}_2 &\doteq (\overline{\alpha} + \beta) - 1 < 0 \end{aligned}$$

where we define $\overline{\alpha} \doteq \frac{\alpha_1+\alpha_2}{2}$ for notational purposes. If both conditions are satisfied, the unconditional variance h_y exists and is given by:

$$h_y \doteq \frac{\alpha_0}{\mathrm{CSC}_1 \times \mathrm{CSC}_2} \,.$$

The joint posterior sample can be used to estimate the posterior density of these functions by generating:

$$\begin{aligned} \mathrm{CSC}_1^{[j]} &\doteq \left(\gamma_1^{[j]}\right)^2 - 1 \\ \mathrm{CSC}_2^{[j]} &\doteq (\overline{\alpha}^{[j]} + \beta^{[j]}) - 1 \end{aligned}$$

and then:

$$h_y^{[j]} \doteq \frac{\alpha_0^{[j]}}{\mathrm{CSC}_1^{[j]} \times \mathrm{CSC}_2^{[j]}}$$

for $j = 1, \ldots, 10'000$. In our simulation study, none of the values CSC_1 and CSC_2 exceed zero, thus indicating that the process is covariance stationary and that

the unconditional variance exists. The posterior density of the unconditional variance is displayed in **Fig. 4.4** together with the ML asymptotic Normal approximation. The posterior mean and the posterior median are respectively 0.169 and 0.1653. The value of the unconditional variance computed from the ML point estimates is slightly lower with a value of 0.1608. The 95% confidence band given by the Bayesian approach is [0.1373,0.2197]. In the classical approach, the confidence band computed via the delta method is [0.0725,0.2491]. In this case, the asymptotic Normal approximation highly overestimates the size of the confidence band, especially the left part of the interval. As shown in **Fig. 4.4**, the asymptotic approximation is flat and symmetric whereas the posterior density is more peaked and exhibits a positive skewness (the skewness is 2.01 and significantly different from zero).

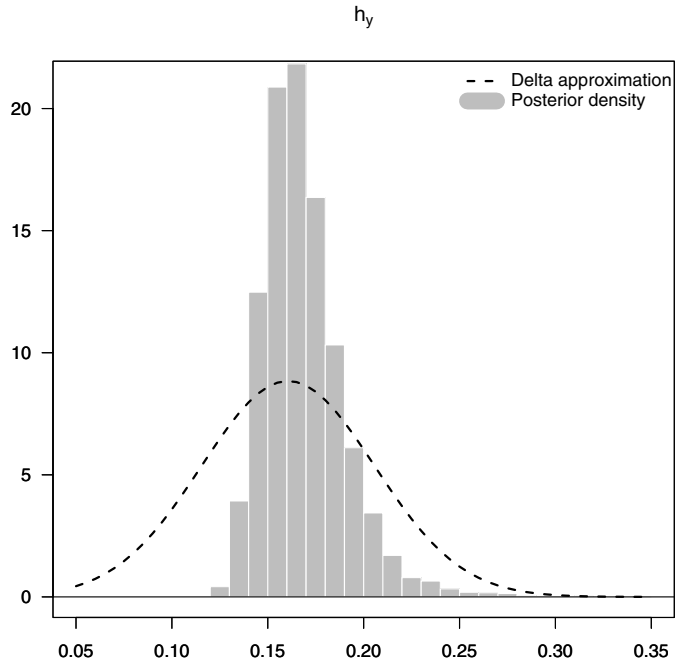

Fig. 4.4. Posterior density of the unconditional variance and asymptotic Normal approximation. The histogram is based on 10'000 draws from the joint posterior sample.

5

Bayesian Estimation of the Linear Regression Model with Student-t-GJR(1, 1) Errors

> *"This development (i.e., the Student-t distribution) permits a distinction between conditional heteroskedasticity and a conditional leptokurtic distribution, either of which could account for the observed unconditional kurtosis in the data."*
>
> — *Tim Bollerslev*

In this chapter, we extend the linear regression model with conditionally heteroscedastic errors. The conditional variance is again described by the GJR process introduced in **Chap. 4**. However, in the new specification, the errors are no longer Normally distributed but follow a Student-t distribution. Therefore, the model incorporates the possibility of heavy-tailed disturbances. Indeed, while the Normal distribution is used routinely, it has been widely recognized that financial markets exhibit significant non-Normalities, in particular asset returns exhibit heavy tails. A distribution with fat tails makes extreme outcomes such as crashes relatively more likely than does a Normal distribution which assigns virtually zero probability to events that are greater than three standard deviations. Since one of the objectives of financial risk management models is to measure severe losses, *i.e.*, events appearing in the tails of the distribution, this is a serious shortcoming and the alternative of the Student-t distribution is a parsimonious way to incorporate fat tails in the modeling. In the Bayesian approach, the heavy-tails effect is created by the introduction of latent variables in the variance process as proposed by Geweke [1993]; this approach allows the Bayesian estimation of the degrees of freedom parameter in a convenient manner. As a first illustration, we fit the model to the S&P100 index log-returns and compare the Bayesian and the Maximum Likelihood estimations. Next, we perform a prior analysis and test the residuals for misspecification. Finally, we estimate the conditional and unconditional kurtosis of the underlying time series.

The plan of this chapter is as follows. We set up the model in **Sect. 5.1**. The MCMC scheme is detailed in **Sect. 5.2**. The empirical results are presented in **Sect. 5.3**. We conclude with some illustrations of the Bayesian approach in **Sect. 5.4**.

5.1 The model and the priors

A linear regression model with Student-t-GJR(1, 1) errors may be written as follows:

$$\begin{aligned} y_t &= \mathbf{x}_t'\boldsymbol{\gamma} + u_t \quad \text{for } t = 1, \ldots, T \\ u_t &= \varepsilon_t(\varrho h_t)^{1/2} \\ \varepsilon_t &\overset{iid}{\sim} \mathcal{S}(0, 1, \nu) \\ \varrho &\doteq \frac{\nu - 2}{\nu} \\ h_t &\doteq \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}}) u_{t-1}^2 + \beta h_{t-1} \end{aligned} \tag{5.1}$$

where $\alpha_0 > 0$, $\alpha_1 \geqslant 0$, $\alpha_2 \geqslant 0$, $\beta \geqslant 0$, $\nu > 2$ and $h_0 = y_0 \doteq 0$ for convenience; y_t is a scalar dependent variable; $\mathbf{x}_t$ is a $m \times 1$ vector of exogenous or lagged dependent variables; $\boldsymbol{\gamma}$ is a $m \times 1$ vector of regression coefficients; $\mathcal{S}(0, 1, \nu)$ is the standard Student-t density with ν degrees of freedom, *i.e.*, its variance is $\frac{\nu}{\nu-2}$. From model specification (5.1) we note that ϱ is a scaling factor which normalizes the variance of the Student-t density so that h_t is the variance of y_t given by the GJR scedastic function. The restriction on the degrees of freedom parameter ensures the conditional variance to be finite and the restrictions on the GJR parameters guarantee its positivity.

In order to write the likelihood function, we define the vectors $\mathbf{y} \doteq (y_1 \cdots y_T)'$ and $\boldsymbol{\alpha} \doteq (\alpha_0 \ \alpha_1 \ \alpha_2)'$ as well as the $T \times m$ matrix X of observations whose tth row is $\mathbf{x}_t'$. For notational purposes, we regroup the model parameters into $\psi \doteq (\boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta, \nu)$. In addition, we define the $T \times T$ diagonal matrix:

$$\Sigma \doteq \Sigma(\psi) = \operatorname{diag}\big(\{\varrho h_t(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta)\}_{t=1}^T\big)$$

where:

$$\begin{aligned} h_t(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta) &\doteq \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1}(\boldsymbol{\gamma}) \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1}(\boldsymbol{\gamma}) < 0\}}) u_{t-1}^2(\boldsymbol{\gamma}) \\ &\quad + \beta h_{t-1}(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta) \\ u_t(\boldsymbol{\gamma}) &\doteq y_t - \mathbf{x}_t'\boldsymbol{\gamma} \ . \end{aligned} \tag{5.2}$$

Then, we regroup the error terms $u_t(\boldsymbol{\gamma})$ into the $T \times 1$ vector $\mathbf{u} \doteq (u_1 \cdots u_T)'$ and express the likelihood function of ψ as follows:

$$\mathcal{L}_{\mathcal{S}}(\psi \mid \mathbf{y}, \mathrm{X}) \propto \left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\nu^{1/2}}\right]^T (\det \Sigma)^{-1/2} \prod_{t=1}^{T} \left(1 + \frac{u_t^2}{\nu \varrho h_t}\right)^{-\frac{(\nu+1)}{2}} . \tag{5.3}$$

The notation $\mathcal{L}_{\mathcal{S}}$ emphasizes the fact that the likelihood function is constructed from the Student-t density.

While the likelihood function (5.3) can be used in classical inference, it is not convenient in the Bayesian framework. It is indeed difficult to find proposal densities for the parameters, especially if we aim to sample the degrees of freedom ν as well. To overcome this problem, we express the innovations process $\{u_t\}$ in another specification as proposed by Geweke [1993]. In this new setting, the variable u_t is expressed as follows:

$$\begin{aligned} u_t &= \varepsilon_t (\varpi_t \varrho h_t)^{1/2} \quad \text{for } t = 1, \ldots, T \\ \varepsilon_t &\overset{iid}{\sim} \mathcal{N}(0,1) \\ \varpi_t &\overset{iid}{\sim} \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) . \end{aligned} \tag{5.4}$$

Hence, the error term u_t, conditional on ϖ_t, follows a Normal distribution with a mean of zero and a variance of $\varpi_t \varrho h_t$; the scedastic function ϱh_t is multiplied by a latent variable ϖ_t which is Inverted Gamma distributed. The degrees of freedom parameter ν characterizes the density of ϖ_t as follows:

$$p(\varpi_t \mid \nu) = \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} \varpi_t^{-\frac{\nu}{2}-1} \exp\left[-\frac{\nu}{2\varpi_t}\right] . \tag{5.5}$$

Let us now regroup the latent variables into the $T \times 1$ vector $\boldsymbol{\varpi} \doteq (\varpi_1 \cdots \varpi_T)'$ and define the augmented set of parameters $\Theta \doteq (\psi, \boldsymbol{\varpi})$. Upon defining the $T \times T$ diagonal matrix:

$$\Sigma \doteq \Sigma(\Theta) = \mathrm{diag}\big(\{\varpi_t \varrho h_t(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta)\}_{t=1}^{T}\big)$$

where h_t is given in expression (5.2), we can express the likelihood function of Θ as follows:

$$\mathcal{L}(\Theta \mid \mathbf{y}, \mathrm{X}) \propto (\det \Sigma)^{-1/2} \exp\left[-\tfrac{1}{2}\mathbf{u}'\Sigma^{-1}\mathbf{u}\right] . \tag{5.6}$$

This specification is equivalent to (5.3). However, in that case, the Bayesian estimation can be handled for all parameters in a convenient manner.

We propose the following proper priors on the parameters $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$, β of the preceding model:

$$\begin{aligned} p(\boldsymbol{\gamma}) &= \mathcal{N}_m(\boldsymbol{\gamma} \mid \boldsymbol{\mu}_\gamma, \Sigma_\gamma) \\ p(\boldsymbol{\alpha}) &\propto \mathcal{N}_3(\boldsymbol{\alpha} \mid \boldsymbol{\mu}_\alpha, \Sigma_\alpha)\mathbb{I}_{\{\boldsymbol{\alpha}>\mathbf{0}\}} \\ p(\beta) &\propto \mathcal{N}(\beta \mid \mu_\beta, \Sigma_\beta)\mathbb{I}_{\{\beta>0\}} \end{aligned}$$

where we recall that $\mu_\bullet$ and $\Sigma_\bullet$ are the hyperparameters, $\mathbb{I}_{\{\bullet\}}$ is the indicator function, $\mathbf{0}$ is a 3×1 vector of zeros and $\mathcal{N}_d$ is the d-dimensional Normal density ($d > 1$).

The prior density of vector $\boldsymbol{\varpi}$ conditional on ν is found by noting that the components ϖ_t are independent and identically distributed from (5.5), which yields:

$$p(\boldsymbol{\varpi} \mid \nu) = \left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \left(\prod_{t=1}^{T} \varpi_t\right)^{-\frac{\nu}{2}-1} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}\frac{\nu}{\varpi_t}\right] .$$

We follow Deschamps [2006] in the choice of the prior density on the degrees of freedom parameter. The density is a translated Exponential with parameters $\lambda > 0$ and $\delta \geqslant 2$:

$$p(\nu) = \lambda \exp\big[-\lambda(\nu-\delta)\big]\mathbb{I}_{\{\nu>\delta\}} .$$

For large values of λ, the mass of the prior is concentrated in the neighborhood of δ and a constraint on the degrees of freedom can be imposed in this manner. The Normality of the errors is obtained when δ becomes large. As pointed out by Deschamps [2006], this prior density is useful for two reasons. First, it is potentially important, for numerical reasons, to bound the degrees of freedom parameter away from two to avoid explosion of the conditional variance. Second, we can approximate the Normality of the errors while maintaining a reasonably tight prior which can improve the convergence of the MCMC sampler.

Finally, we assume prior independence between $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$, β and $(\boldsymbol{\varpi}, \nu)$ which yields the following joint prior:

$$p(\Theta) = p(\boldsymbol{\gamma})p(\boldsymbol{\alpha})p(\beta)p(\boldsymbol{\varpi} \mid \nu)p(\nu)$$

and, by combining the likelihood function (5.6) and the joint prior, we construct the posterior density via Bayes' rule:

$$p(\Theta \mid \mathbf{y}, \mathrm{X}) \propto \mathcal{L}(\Theta \mid \mathbf{y}, \mathrm{X})p(\Theta) .$$

5.2 Simulating the joint posterior

Once again, we rely on the M-H algorithm to draw samples from the joint posterior distribution. We draw an initial value:

$$\Theta^{[0]} \doteq (\boldsymbol{\gamma}^{[0]}, \boldsymbol{\alpha}^{[0]}, \beta^{[0]}, \boldsymbol{\varpi}^{[0]}, \nu^{[0]})$$

from the joint prior and we generate iteratively J passes for Θ. A single pass is decomposed as follows:

$$\begin{aligned}
\boldsymbol{\gamma}^{[j]} &\sim p(\boldsymbol{\gamma} \mid \boldsymbol{\alpha}^{[j-1]}, \beta^{[j-1]}, \boldsymbol{\varpi}^{[j-1]}, \nu^{[j-1]}, \mathbf{y}, \mathrm{X}) \\
\boldsymbol{\alpha}^{[j]} &\sim p(\boldsymbol{\alpha} \mid \boldsymbol{\gamma}^{[j]}, \beta^{[j-1]}, \boldsymbol{\varpi}^{[j-1]}, \nu^{[j-1]}, \mathbf{y}, \mathrm{X}) \\
\beta^{[j]} &\sim p(\beta \mid \boldsymbol{\gamma}^{[j]}, \boldsymbol{\alpha}^{[j]}, \boldsymbol{\varpi}^{[j-1]}, \nu^{[j-1]}, \mathbf{y}, \mathrm{X}) \\
\boldsymbol{\varpi}^{[j]} &\sim p(\boldsymbol{\varpi} \mid \boldsymbol{\gamma}^{[j]}, \boldsymbol{\alpha}^{[j]}, \beta^{[j]}, \nu^{[j-1]}, \mathbf{y}, \mathrm{X}) \\
\nu^{[j]} &\sim p(\nu \mid \boldsymbol{\varpi}^{[j]}) \ .
\end{aligned}$$

Only vector $\boldsymbol{\varpi}$ can be simulated from a known expression. Draws of parameters $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and β are made using a method similar to the one presented in **Sect. 4.2**. Sampling parameter ν is more technical and relies on an optimized rejection technique.

5.2.1 Generating vector $\boldsymbol{\gamma}$

The proposal density to sample the $m \times 1$ vector $\boldsymbol{\gamma}$ is obtained by combining the likelihood function (5.3) and the prior density by Bayes' update:

$$q_\gamma(\boldsymbol{\gamma} \mid \widetilde{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \beta, \boldsymbol{\varpi}, \nu, \mathbf{y}, \mathrm{X}) = \mathcal{N}_m(\boldsymbol{\gamma} \mid \widehat{\boldsymbol{\mu}}_\gamma, \widehat{\Sigma}_\gamma)$$

with:

$$\begin{aligned}
\widehat{\Sigma}_\gamma^{-1} &\doteq \mathrm{X}'\widetilde{\Sigma}^{-1}\mathrm{X} + \Sigma_\gamma^{-1} \\
\widehat{\boldsymbol{\mu}}_\gamma &\doteq \widehat{\Sigma}_\gamma(\mathrm{X}'\widetilde{\Sigma}^{-1}\mathbf{y} + \Sigma_\gamma^{-1}\boldsymbol{\mu}_\gamma)
\end{aligned}$$

where the $T \times T$ diagonal matrix $\widetilde{\Sigma} \doteq \mathrm{diag}\big(\{\varpi_t \varrho h_t(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \beta)\}_{t=1}^T\big)$, $\widetilde{\boldsymbol{\gamma}}$ is the previous draw of $\boldsymbol{\gamma}$ in the M-H sampler and $\varrho \doteq \frac{\nu-2}{\nu}$. A candidate $\boldsymbol{\gamma}^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\gamma}^\star, \boldsymbol{\alpha}, \beta, \boldsymbol{\varpi}, \nu \mid \mathbf{y}, \mathrm{X})}{p(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \beta, \boldsymbol{\varpi}, \nu \mid \mathbf{y}, \mathrm{X})} \frac{q_\gamma(\widetilde{\boldsymbol{\gamma}} \mid \boldsymbol{\gamma}^\star, \boldsymbol{\alpha}, \beta, \boldsymbol{\varpi}, \nu, \mathbf{y}, \mathrm{X})}{q_\gamma(\boldsymbol{\gamma}^\star \mid \widetilde{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \beta, \boldsymbol{\varpi}, \nu, \mathbf{y}, \mathrm{X})}, 1\right\} \ .$$

5.2.2 Generating the GJR parameters

The methodology is similar to the one exposed in **Sect. 4.2.2**. Let us define:

$$w_t \doteq \frac{u_t^2}{\tau_t} - h_t$$

where $\tau_t \doteq \varpi_t \varrho$ for convenience. From there, we can transform the expression of the conditional variance as follows:

$$\begin{aligned} h_t &= \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}}) u_{t-1}^2 + \beta h_{t-1} \\ \Leftrightarrow \frac{u_t^2}{\tau_t} &= \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}}) u_{t-1}^2 \\ &\quad + \beta \frac{u_{t-1}^2}{\tau_{t-1}} - \beta w_{t-1} + w_t \\ \Leftrightarrow v_t &= \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}}) \tau_{t-1} v_{t-1} \\ &\quad + \beta v_{t-1} - \beta w_{t-1} + w_t \end{aligned}$$

where we define $v_t \doteq \frac{u_t^2}{\tau_t}$ for notational purposes. Moreover, we note that the variable w_t can also be expressed as follows:

$$w_t \doteq \frac{u_t^2}{\tau_t} - h_t = \left(\frac{u_t^2}{\tau_t h_t} - 1 \right) h_t = \left(\chi_1^2 - 1 \right) h_t$$

where χ_1^2 denotes a Chi-squared variable with one degree of freedom. The last equality results from the expression for u_t in (5.4). Therefore, the variable w_t has a conditional mean of zero and a conditional variance of $2h_t^2$. In addition, we note that the sequence $\{w_t\}$ is again a Martingale Difference process.

Approximating the variable w_t by a variable z_t which is Normally distributed with a mean of zero and a variance of $2h_t^2$ yields the following *auxiliary model* for v_t:

$$\begin{aligned} v_t &= \alpha_0 + (\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}}) \tau_{t-1} v_{t-1} + \beta v_{t-1} \\ &\quad - \beta z_{t-1} + z_t \ . \end{aligned}$$

Then, by noting that z_t is a function of $(\boldsymbol{\alpha}, \beta)$ given by:

$$\begin{aligned} z_t(\boldsymbol{\alpha}, \beta) &= v_t - \alpha_0 - \left[(\alpha_1 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{u_{t-1} < 0\}}) \tau_{t-1} + \beta \right] v_{t-1} \\ &\quad + \beta z_{t-1}(\boldsymbol{\alpha}, \beta) \end{aligned} \tag{5.7}$$

and by defining the $T \times 1$ vector $\mathbf{z} \doteq (z_1 \cdots z_T)'$ as well as the $T \times T$ diagonal matrix:

$$\Lambda \doteq \Lambda(\boldsymbol{\alpha}, \beta) = \operatorname{diag}\big(\{2h_t^2(\boldsymbol{\alpha}, \beta)\}_{t=1}^T\big)$$

we can approximate the likelihood function of $(\boldsymbol{\alpha}, \beta)$ as follows:

$$\mathcal{L}(\boldsymbol{\alpha}, \beta \mid \boldsymbol{\gamma}, \boldsymbol{\varpi}, \mathbf{y}, \mathrm{X}) \propto (\det \Lambda)^{-1/2} \exp\left[-\tfrac{1}{2}\mathbf{z}'\Lambda^{-1}\mathbf{z}\right] \ . \tag{5.8}$$

As will be shown hereafter, the construction of the proposal densities for parameters $\boldsymbol{\alpha}$ and β is based on this approximate likelihood function.

Generating vector $\boldsymbol{\alpha}$

Our aim is to express the function $z_t(\boldsymbol{\alpha}, \beta)$ in (5.7) as a linear function of the 3×1 vector $\boldsymbol{\alpha}$. To that aim, let us define the following recursive transformations:

$$\begin{aligned} l_t^* &\doteq 1 + \beta\, l_{t-1}^* \\ v_t^* &\doteq u_{t-1}^2 \mathbb{I}_{\{u_{t-1} \geqslant 0\}} + \beta\, v_{t-1}^* \\ v_t^{**} &\doteq u_{t-1}^2 \mathbb{I}_{\{u_{t-1} < 0\}} + \beta\, v_{t-1}^{**} \end{aligned}$$

where $l_0^* = v_0^* = v_0^{**} \doteq 0$. As shown in **Prop. A.3** (see **App. A**), upon defining the 3×1 vector $\mathbf{c}_t \doteq (l_t^* \ v_t^* \ v_t^{**})'$, it turns out that the function z_t can be expressed as $z_t = v_t - \mathbf{c}_t'\boldsymbol{\alpha}$. Hence, by defining the $T \times 1$ vectors $\mathbf{z} \doteq (z_1 \cdots z_T)'$ and $\mathbf{v} \doteq (v_1 \cdots v_T)'$ as well as the $T \times 3$ matrix C whose tth row is $\mathbf{c}_t'$, we get $\mathbf{z} = \mathbf{v} - C\boldsymbol{\alpha}$. Therefore, we can express the approximate likelihood function of parameter $\boldsymbol{\alpha}$ as follows:

$$\mathcal{L}(\boldsymbol{\alpha} \mid \boldsymbol{\gamma}, \beta, \boldsymbol{\varpi}, \nu, \mathbf{y}, \mathrm{X}) \propto (\det \Lambda)^{-1/2} \exp\left[-\tfrac{1}{2}(\mathbf{v} - C\boldsymbol{\alpha})'\Lambda^{-1}(\mathbf{v} - C\boldsymbol{\alpha})\right] \ .$$

The proposal density to sample vector $\boldsymbol{\alpha}$ is obtained by combining this likelihood function and the prior density by Bayes' update:

$$q_\alpha(\boldsymbol{\alpha} \mid \widetilde{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \beta, \boldsymbol{\varpi}, \nu, \mathbf{y}, \mathrm{X}) \propto \mathcal{N}_3(\boldsymbol{\alpha} \mid \widehat{\boldsymbol{\mu}}_\alpha, \widehat{\Sigma}_\alpha)\mathbb{I}_{\{\boldsymbol{\alpha} > \mathbf{0}\}}$$

with:

$$\begin{aligned} \widehat{\Sigma}_\alpha^{-1} &\doteq C'\widetilde{\Lambda}^{-1}C + \Sigma_\alpha^{-1} \\ \widehat{\boldsymbol{\mu}}_\alpha &\doteq \widehat{\Sigma}_\alpha(C'\widetilde{\Lambda}^{-1}\mathbf{v} + \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha) \end{aligned}$$

where the $T \times T$ diagonal matrix $\widetilde{\Lambda} \doteq \operatorname{diag}\big(\{2h_t^2(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\alpha}}, \beta)\}_{t=1}^T\big)$ and $\widetilde{\boldsymbol{\alpha}}$ is the previous draw of $\boldsymbol{\alpha}$ in the M-H sampler. A candidate $\boldsymbol{\alpha}^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\alpha}^\star,\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\varpi},\nu\mid\mathbf{y},\mathrm{X})}{p(\widetilde{\boldsymbol{\alpha}},\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\varpi},\nu\mid\mathbf{y},\mathrm{X})}\frac{q_\alpha(\widetilde{\boldsymbol{\alpha}}\mid\boldsymbol{\gamma},\boldsymbol{\alpha}^\star,\beta,\boldsymbol{\varpi},\nu,\mathbf{y},\mathrm{X})}{q_\alpha(\boldsymbol{\alpha}^\star\mid\boldsymbol{\gamma},\widetilde{\boldsymbol{\alpha}},\beta,\boldsymbol{\varpi},\nu,\mathbf{y},\mathrm{X})},1\right\}\ .$$

Generating parameter $\boldsymbol{\beta}$

Contrary to the parameter $\boldsymbol{\alpha}$, we cannot express the function $z_t(\boldsymbol{\alpha},\beta)$ in (5.7) as a linear function of β. To bypass this problem, we approximate the function $z_t(\beta)$ by a first order Taylor expansion at point $\widetilde{\beta}$:

$$z_t(\beta)\simeq z_t(\widetilde{\beta})+\left.\frac{\mathrm{d}z_t}{\mathrm{d}\beta}\right|_{\beta=\widetilde{\beta}}\times(\beta-\widetilde{\beta})$$

where $\widetilde{\beta}$ is the previous draw of β in the M-H sampler. From there, we define the following:

$$r_t\doteq z_t(\widetilde{\beta})+\widetilde{\beta}\nabla_t\quad,\quad\nabla_t\doteq-\left.\frac{\mathrm{d}z_t}{\mathrm{d}\beta}\right|_{\beta=\widetilde{\beta}}$$

where the terms ∇_t can be computed by the following recursion:

$$\nabla_t\doteq v_{t-1}^2-z_{t-1}(\widetilde{\beta})+\widetilde{\beta}\nabla_{t-1}$$

with $\nabla_0\doteq 0$. This recursion is simply obtained by differentiating (5.7) with respect to β. Then, we regroup these terms into the $T\times 1$ vectors $\mathbf{r}\doteq(r_1\cdots r_T)'$ and $\boldsymbol{\nabla}\doteq(\nabla_1\cdots\nabla_T)'$ and we approximate the term within the exponential in (5.8) by $\mathbf{z}\simeq\mathbf{r}-\beta\boldsymbol{\nabla}$. This yields the following approximate likelihood function for parameter β:

$$\mathcal{L}(\beta\mid\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\varpi},\nu,\mathbf{y},\mathrm{X})\propto(\det\Lambda)^{-1/2}\exp\left[-\tfrac{1}{2}(\mathbf{r}-\beta\boldsymbol{\nabla})'\Lambda^{-1}(\mathbf{r}-\beta\boldsymbol{\nabla})\right]\ .$$

This likelihood function is combined with the prior density by Bayes' update to construct the proposal $q_\beta(\beta\mid\bullet)$. A candidate $\beta^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\gamma},\boldsymbol{\alpha},\beta^\star,\boldsymbol{\varpi},\nu\mid\mathbf{y},\mathrm{X})}{p(\boldsymbol{\gamma},\boldsymbol{\alpha},\widetilde{\beta},\boldsymbol{\varpi},\nu\mid\mathbf{y},\mathrm{X})}\frac{q_\beta(\widetilde{\beta}\mid\boldsymbol{\gamma},\boldsymbol{\alpha},\beta^\star,\boldsymbol{\varpi},\nu,\mathbf{y},\mathrm{X})}{q_\beta(\beta^\star\mid\boldsymbol{\gamma},\boldsymbol{\alpha},\widetilde{\beta},\boldsymbol{\varpi},\nu,\mathbf{y},\mathrm{X})},1\right\}\ .$$

5.2.3 Generating vector $\boldsymbol{\varpi}$

The components of $\boldsymbol{\varpi}$ are independent a posteriori and the full conditional posterior of ϖ_t is obtained as follows:

$$p(\varpi_t \mid \boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta, \nu, \mathbf{y}, \mathrm{X}) \propto \mathcal{L}(\Theta \mid \mathbf{y}, \mathrm{X}) p(\varpi_t \mid \nu)$$
$$\propto \varpi_t^{-\frac{(\nu+3)}{2}} \exp\left[-\frac{b_t}{\varpi_t}\right] \tag{5.9}$$

with:

$$b_t \doteq \frac{1}{2}\left[\frac{(y_t - \mathbf{x}_t'\boldsymbol{\gamma})^2}{\varrho h_t} + \nu\right]$$

where we recall that $h_t \doteq h_t(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta)$ and $\varrho \doteq \frac{\nu-2}{\nu}$. Expression (5.9) is the kernel of an Inverted Gamma density with parameters $\frac{\nu+1}{2}$ and b_t.

5.2.4 Generating parameter $\boldsymbol{\nu}$

Draws from $p(\nu \mid \boldsymbol{\varpi})$ are made by optimized rejection sampling from a translated Exponential source density. The target density is:

$$p(\nu \mid \boldsymbol{\varpi}) \propto p(\boldsymbol{\varpi} \mid \nu) p(\nu)$$
$$\propto \left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \exp\left[-\varphi\nu\right] \mathbb{I}_{\{\nu>\delta\}}$$

with:

$$\varphi \doteq \frac{1}{2}\sum_{t=1}^{T}\left(\ln \varpi_t + \varpi_t^{-1}\right) + \lambda\,.$$

Following Deschamps [2006], we sample a candidate $\nu^\star$ from a translated Exponential source density:

$$g(\nu; \bar{\mu}, \delta) \doteq \bar{\mu} \exp\big[-\bar{\mu}(\nu - \delta)\big] \mathbb{I}_{\{\nu>\delta\}}$$

where $\bar{\mu}$ maximizes the acceptance probability. The choice of $\bar{\mu}$ is found by solving:

$$\frac{T}{2}\left[\ln\left(\frac{1+\mu\delta}{2\mu}\right) + 1 - \Psi\left(\frac{1+\mu\delta}{2\mu}\right)\right] + \mu - \varphi = 0$$

for μ, where $\Psi(z) \doteq \frac{\mathrm{d}\ln\Gamma(z)}{\mathrm{d}z}$ is the Digamma function. The candidate $\nu^\star$ is accepted with probability:

$$p^\star \doteq \frac{k(\nu^\star)}{s(\bar{\mu}, \delta) g(\nu^\star; \bar{\mu}, \delta)} \tag{5.10}$$

where $k(\nu)$ is the kernel of the target density:

$$k(\nu) \doteq \left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \exp\left[-\varphi\nu\right]$$

and $s(\mu, \delta)$ is given by:

$$\begin{aligned} s(\mu,\delta) &\doteq k\left(\frac{1+\mu\delta}{\mu}\right)\left[g\left(\frac{1+\mu\delta}{\mu};\mu,\delta\right)\right]^{-1} \\ &= \left(\frac{1+\mu\delta}{2\mu}\right)^{\frac{T(1+\mu\delta)}{2\mu}}\left[\Gamma\left(\frac{1+\mu\delta}{2\mu}\right)\right]^{-T}\mu^{-1}\exp\left[1-\frac{\varphi(1+\mu\delta)}{\mu}\right] . \end{aligned}$$

Substituting for $k(\nu^\star)$, $s(\bar{\mu}, \delta)$ and $g(\nu^\star; \bar{\mu}, \delta)$ in expression (5.10) yields:

$$\begin{aligned} p^\star = &\left[\frac{\Gamma\left(\frac{1+\bar{\mu}\delta}{2\bar{\mu}}\right)}{\Gamma\left(\frac{\nu^\star}{2}\right)}\right]^T\left(\frac{\nu^\star}{2}\right)^{\frac{T\nu^\star}{2}}\left(\frac{1+\bar{\mu}\delta}{2\bar{\mu}}\right)^{\frac{-T(1+\bar{\mu}\delta)}{2\bar{\mu}}} \\ &\times\exp\left[(\nu^\star-\delta)(\bar{\mu}-\varphi)+\frac{\varphi}{\bar{\mu}}-1\right] . \end{aligned}$$

To end this section, we note that a slight modification of Geweke [1993] allows to generate draws from a Student-t distribution with conditional variance h_t without requiring the introduction of a scaling parameter $\varrho \doteq \frac{\nu-2}{\nu}$. This is done by replacing the specification for the latent variable ϖ_t in (5.4) by:

$$\varpi_t \overset{iid}{\sim} \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu-2}{2}\right) .$$

The use of this new specification requires some modifications of the efficient rejection scheme. We refer the reader to **App. B** for further details.

Finally, we note that the validity of the algorithm and the correctness of the computer code are verified by the methodology detailed at the end of **Sect. 3.2.2**.

5.3 Empirical analysis

To illustrate our Bayesian estimation method, we fit the Student-t-GJR(1, 1) model to the data set used in the empirical analysis of **Chap. 4**. Based on previous results, we do not include the regression part in the current estimation.

5.3.1 Model estimation

As prior densities for the GJR parameters, we choose truncated Normal densities with zero mean vectors and diagonal covariance matrices whose variances are set to 10'000. For the prior on the degrees of freedom parameter, we set the hyperparameters to $\lambda = 0.01$ and $\delta = 2$; the prior mean is therefore 102 and the

prior variance 10'000. Note that the value of the hyperparameter δ is determined so that the conditional variance exists. Moreover, we recall that the joint prior is constructed by assuming prior independence between $\boldsymbol{\alpha}$, β and (ϖ, ν).

We run two chains for 10'000 passes each and control the convergence of the sampler using the diagnostic test by Gelman and Rubin [1992]. The convergence diagnostic shows no evidence against convergence for the last 5'000 iterations (the value of the 97.5th percentile of the potential scale reduction factor ranges from 1.001 to 1.1). The one-lag autocorrelations in the chains range from 0.59 for parameter α_1 to 0.97 for parameter ν. The acceptance rate is 73% for vector $\boldsymbol{\alpha}$ and 95% for parameter β. The optimized rejection technique allows to draw a new value of ν at each pass in the M-H algorithm. From the overall MCMC output, we discard the first 5'000 draws and merge the two chains to get a final sample of length 10'000.

The posterior statistics as well as the ML results are reported in **Table 5.1**. First, we note that results for the GJR parameters are close to the results of **Table 4.1** (see p.48). The posterior means of the parameters are slightly higher in the Student-t case (except for parameter α_0) as well as the numerical standard errors. Second, the marginal posterior densities (not shown) are still clearly skewed and the 95% confidence band of the parameters obtained through the asymptotic Normal approximation leads to a negative left boundary for component α_1. The ML point estimate for the degrees of freedom parameter is 9.9 while the posterior mean is 7.15 and the posterior median is 7.11. This low value indicates a departure from Normality for the errors. In addition, the 95% confidence band given by the ML approach is much wider than the one estimated via the Bayesian approach. The left boundary is 0.14 which rejects the existence of the conditional variance. In the case of the Bayesian estimation, the minimum value for the degrees of freedom is 3.84, which supports the existence of the conditional variance. The values of the inefficiency factor (IF) range from 3.65 for parameter α_1 to 111.98 for parameter ν, indicating that in the worst case, the numerical errors represent about 1.12% of the variation of the errors due to the data.

In **Fig. 5.1**, we present a comparison between the classical and the Bayesian approaches. The upper graphs show a scatter plot of the draws from the asymptotic Normal approximation of the model parameters; the Normal density is centered at the ML estimates ψ_{MLE} and its covariance matrix is estimated as the inverse of the Hessian matrix evaluated at ψ_{MLE}. The lower graphs present a scatter plot of draws from the joint posterior sample. In both cases, the number of draws is 10'000. The first part of the figure depicts the draws for (α_0, β). By comparing the ML and Bayesian outputs, we can notice a clear difference in the

Table 5.1. Estimation results for the Student-t-GJR(1, 1) model.*

ψ	ψ_{MLE}	$\overline{\psi}$	$\psi_{0.5}$	$\psi_{0.025}$	$\psi_{0.975}$	min	max	IF
α_0	0.012	0.018	0.017	0.008	0.036	0.003	0.060	18.15
	[0.002,0.021]	(0.314)						
α_1	0.026	0.041	0.037	0.008	0.107	0.000	0.194	3.65
	[-0.015,0.067]	(0.516)						
α_2	0.153	0.203	0.194	0.105	0.350	0.055	0.533	6.93
	[0.063,0.242]	(1.641)						
β	0.834	0.776	0.785	0.634	0.870	0.408	0.908	54.98
	[0.740,0.929]	(4.636)						
ν	9.90	7.549	7.11	4.54	13.60	3.84	18.85	111.98
	[0.14,19.65]	(236.50)						

* ψ_{MLE}: Maximum Likelihood estimate; $\overline{\psi}$: posterior mean; ψ_ϕ: estimated posterior quantile at probability ϕ; min: minimum value; max: maximum value; IF: inefficiency factor (*i.e.*, ratio of the squared numerical standard error and the variance of the sample mean from a hypothetical *iid* sampler); [•]: Maximum Likelihood 95% confidence interval; (•): numerical standard error ($\times 10^3$). The posterior statistics are based on 10'000 draws from the joint posterior sample.

tails of the joint density. Indeed, the Bayesian posterior exhibits larger values for parameter α_0 together with lower values for parameter β. In addition, when drawing a vertical line at $\alpha_0 = 0$, we note that some draws are negative with the asymptotic Normal approximation. In the posterior sample, the draws are positive as this is required by the prior density. In the second part of **Fig. 5.1**, the two graphs show the draws for (α_2, β). For these parameters, the posterior sample exhibits a clear departure from the ellipsoid shape obtained with the Normal approximation.

In **Fig. 5.2**, we display the prior and the posterior densities of the degrees of freedom parameter. While the prior density is almost flat (we recall that the hyperparameters are set to $\lambda = 0.01$ and $\delta = 2$ so that the prior mean is 102 and the prior variance 10'000), the shape of the posterior density is peaked and concentrated around its mean value. In addition, the density is significantly right-skewed.

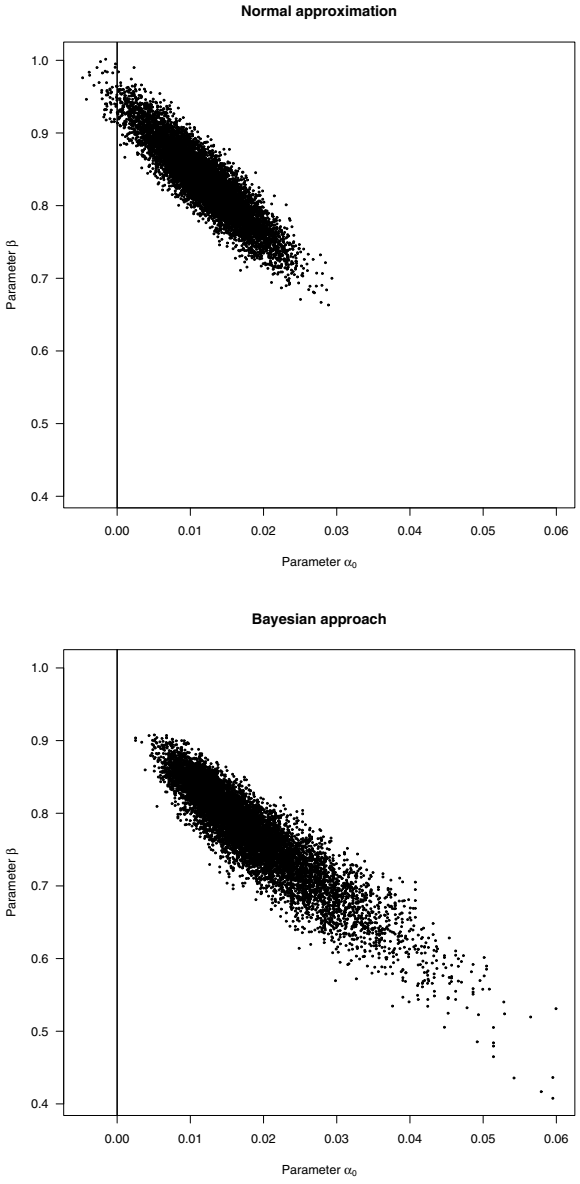

Fig. 5.1. Comparison between the ML (upper graph) and the Bayesian (lower graph) approaches. For both graphs, the number of draws is 10'000.

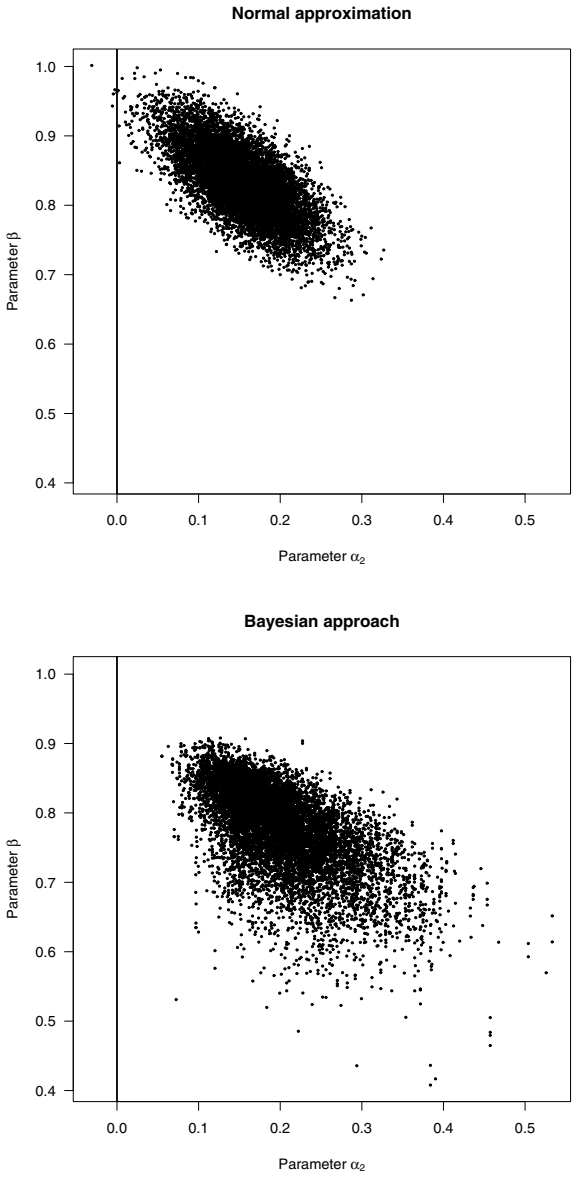

Fig. 5.1. (*cont.*) Comparison between the ML (upper graph) and the Bayesian (lower graph) approaches. For both graphs, the number of draws is 10'000.

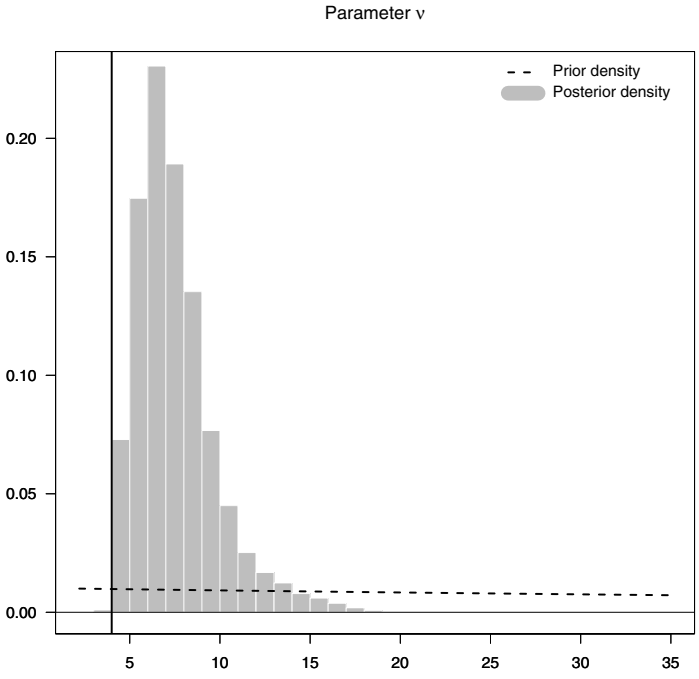

Fig. 5.2. Prior and posterior densities of the degrees of freedom parameter. The vertical line is centered at $\nu = 4$, the value required for the conditional kurtosis of the errors to exist. The histogram is based on 10'000 draws from the posterior sample.

5.3.2 Sensitivity analysis

As in previous chapters, we test the robustness of our results with respect to the choice of the prior density. To that aim, we consider the same alternative priors of **Sect. 4.3.2** for parameters $\boldsymbol{\alpha}$ and β:

$$p(\boldsymbol{\alpha}) \propto \mathcal{N}_3(\boldsymbol{\alpha} \mid \mu\, \boldsymbol{\iota}_3, \sigma^2 I_3)\mathbb{I}_{\{\boldsymbol{\alpha}>\mathbf{0}\}}$$
$$p(\beta) \propto \mathcal{N}(\beta \mid \mu, \sigma^2)\mathbb{I}_{\{\beta>0\}}$$

where we recall that $\boldsymbol{\iota}_3$ is a 3×1 vector of ones, I_3 is a 3×3 identity matrix, μ is the prior mean and σ^2 the prior variance. For the alternative prior on the degrees of freedom parameter, we consider a translated Exponential with $\lambda = 0.008$ and $\delta = 2$, which implies a prior mean of 127 and a prior variance of 15'625.

The results of **Table 5.2** indicate that the prior on the degrees of freedom has the largest impact on Bayes factors. Moreover, in all cases we conclude to a weak evidence in favor of the initial specification relative to the alternative priors since the Bayes factors belong to the interval $[0.3125, 1]$. This indicates that our initial prior is vague enough and does not introduce significant information in our estimation.

Table 5.2. Results of the sensitivity analysis.★

Alternative priors			
μ	σ^2	λ	BF
1.00	10'000	0.01	1.000
0.00	11'000	0.01	0.826
0.00	10'000	0.008	0.809
1.00	11'000	0.008	0.668

★ The alternative priors on the parameters $\boldsymbol{\alpha}$ and β are truncated Normal densities; μ prior mean; σ^2 prior variance; The alternative prior on the parameter ν is a translated Exponential with hyperparameters λ and $\delta = 2$; BF: Bayes factor.

5.3.3 Model diagnostics

We test the standardized residuals for possible model misspecification. The Ljung-Box test does not reject the absence of autocorrelation in the residuals at the 5% significance level (p-value = 0.4727). This is also true for the squared residuals (p-value = 0.8724). The Kolmogorov-Smirnov Normality test slightly

rejects the Normality assumption at the 5% significance level with a p-value of 0.0437. However, when comparing the standardized residuals to a Student-t distribution whose degrees of freedom parameter is set to the posterior median $\widehat{\nu} = 7.11$, the Kolmogorov-Smirnov empirical distribution test does not reject the null hypothesis at the 5% significance level (p-value = 0.4296). Hence, the model accounts for the conditional heteroscedasticity and for the high kurtosis in the residuals.

5.4 Illustrative applications

We end this chapter by illustrating some probabilistic statements made on the conditional and unconditional kurtosis of the underlying process. Under specification (5.1), the conditional kurtosis κ_ε is defined as follows:

$$\kappa_\varepsilon \doteq \frac{3(\nu-2)}{\nu-4}$$

provided that $\nu > 4$. Using the joint posterior sample, we estimate the posterior probability of the existence for the conditional kurtosis, $\mathbb{P}(\nu > 4 \mid \mathbf{y}, \mathrm{X})$, to 0.999. Therefore, the existence is clearly supported by the data. The posterior mean of the kurtosis is 6.82 and the 95% confidence interval is [3.72,13.8], indicating heavier tails than for the Normal distribution.

Finally, we extend the analysis to the unconditional kurtosis of the process. Let us define $\overline{\alpha} \doteq \frac{\alpha_1+\alpha_2}{2}$ for notational convenience. As demonstrated by He and Teräsvirta [1999], the expression of the unconditional kurtosis κ_y is given by:

$$\kappa_y \doteq \frac{\kappa_\varepsilon(1+\overline{\alpha}+\beta)(1-\overline{\alpha}-\beta)}{1-\kappa_\varepsilon\frac{(\alpha_1^2+\alpha_2^2)}{2}-2\beta(\overline{\alpha}+\beta)}$$

provided that κ_ε is finite and:

$$\frac{\kappa_\varepsilon(\alpha_1^2+\alpha_2^2)}{2}+2\beta(\overline{\alpha}+\beta)<1\ .$$

The posterior probability of the latter condition is 0.007, meaning that there is a 0.7% chance that the unconditional kurtosis exists.

6

Value at Risk and Decision Theory

"Density forecasting is fast becoming an important tool for decision makers in situations where loss functions are asymmetric and forecast errors follow non-Gaussian distributions."

— Allan Timmermann

6.1 Introduction

Since the Group of Thirty report in 1996, the Value at Risk (henceforth VaR) has become the corner-stone in any risk management framework and is essential in allocating capital as a cushion for market risk exposures. This measure gives, for a given time horizon and a given confidence level ϕ, the portfolio's loss that is expected to be exceeded with probability $\phi^c \doteq (1 - \phi)$. The VaR is in many aspects an attractive measure of risk, being relatively easy to implement and easy to explain to non-expert audiences. While primarily designed for market risk exposures, the VaR methodology now underpins the credit and operational risk recommendations. From the *internal models approach* endorsed by the Basel Committee on Banking and Supervision of Banks for Internal Settlement [see Basel Committee on Banking Supervision 1995] and later adopted by US bank regulators, banks are allowed to use their own models to estimate the VaR and keep aside regulatory capital.

From a statistical viewpoint, the VaR is nothing else than a given percentile of the profit and loss (henceforth P&L) distribution over a fixed horizon. To be acceptable by regulators, the confidence level must be 99% and the holding period must be two weeks (*i.e.*, ten trading days). This is motivated by the fear of a liquidity crisis where a financial institution might not be able to liquidate its holdings for ten days straights. However, market participants consider the 99% confidence level and the two weeks horizon to be too conservative. As an additional tool for internal risk controlling, both the holding period and the confidence level can be selected to fit the needs of analysts; in practice, it is common to limit the confidence level to 95% and the holding period to one day.

Evidently VaR can only be constructed by statistical methods. But in most applications, the true P&L distribution is not known and VaR can only be estimated from sample data. The underlying assumption of all the VaR estimation methods is that the risk associated with a particular portfolio for a fixed time horizon is encapsulated within the P&L distribution. If this distribution is known, the VaR can be obtained directly by reading the appropriate percentile value from this distribution. If the P&L is unknown, it must be estimated. As noted by McNeil and Frey [2000, p.272]:

> (...) "the existing approaches for estimating the P&L distribution can be divided into three groups: the *non-parametric historical simulation* method; *fully parametric* methods based on an econometric model for volatility dynamics and the assumption of the conditional distribution, *e.g.*, GARCH models; and finally methods based on *extreme value theory*."

We focus on the second approach in the current application.

Within the fully parametric literature, many papers either attempt to forecast the VaR at different time horizons or use the VaR to assess the forecasting performance of a particular model. In both cases, the methodology is the same. First, a statistical model which describes the P&L dynamics is determined. The model parameters are estimated by the Maximum Likelihood technique for a given *estimation window*. Then, based on these estimations, a VaR point forecast is determined for a given horizon. The procedure is repeated again over a *testing window* by rolling the estimation window; in this manner, we obtain a time series of VaR forecasts. Then, the model is *backtested*, *i.e.*, the predictions are compared with the realized P&Ls. Often, a statistical test is used to assess the performance of the model, *i.e.*, to determine whether the model captures the true VaR [see, *e.g.*, Christoffersen 1998, Kupiec 1995].

While this methodology is accepted by academics and is widely implemented in practice, we note that few empirical studies account for the uncertainty in the VaR predictions. Nevertheless, this issue is important in a risk management framework where some measure of the forecasts' accuracy is also needed; assessing the uncertainty of the VaR will allow the portfolio managers to make more informed decisions when dictating a portfolio re-balance, for instance. We may distinguish two sources of uncertainty which can influence the VaR accuracy:

- The parameter uncertainty within the context of a given model;
- The model uncertainty; via a probability function defined on a class of M possibly non-nested models $\mathcal{M}_i$ $(i = 1, \ldots, M)$.

The former source of uncertainty, also referred to as estimation risk, is straightforwardly handled in Bayesian inference since the complete characterization of the parameter uncertainty is contained in the joint posterior. The latter, known as model risk, is a natural concept in the Bayesian framework. However, in practice, its estimation involves many difficulties. In effect, the methodology requires the estimation of the model likelihood $p(\mathbf{y} \mid \mathcal{M}_i)$ which can be difficult to estimate. Several estimation methods have been proposed but their cost is far from negligible. In addition, the method may not work properly and can be sensitive to the choice of the prior density. For these reasons, we concentrate our attention on the estimation risk where the parameter uncertainty is used to determine the VaR density instead of a single VaR point estimate.

Some approaches have been proposed to quantify the VaR uncertainty when the P&L dynamics is described by GARCH models. Basically, these techniques rely on the bootstrap methodology as in Christoffersen and Gonçalves [2004] or on some asymptotic justifications as in Bams, Lehnert, and Wolff [2005]. The former approach is computationally very demanding since at each step in the procedure, a GARCH model is fitted to the bootstrapped data. While technically more convenient, the latter approach relies on an asymptotic approximation of the distribution of the parameter estimates. The Bayesian approach gives a natural answer to these problems, as noted by Miazhynskaia and Aussenegg [2006]. As will be shown hereafter, the s-day ahead VaR ($s \geqslant 1$) can be expressed as a function of the model parameters; hence, for each parameter in the joint posterior sample, we can obtain a VaR point forecast. By repeating the estimation for each draw in the posterior sample, we obtain an estimation for the VaR density itself. When the forecast horizon is one day, the Bayesian approach gives the exact VaR density. For forecasting horizons larger than one day, an approximation based on the first four moments of the future P&L density can be obtained.

In the Bayesian framework, we can either integrate out the parameter uncertainty or choose a Bayes point estimate within the VaR density. The former case is achieved by simulating from the predictive density and estimating the VaR from empirical percentiles. The latter case yields an interesting problem of decision theory: the choice of a Bayes point estimate which is optimal given a particular loss function. In decision theory, the common practice is to use a symmetric squared error loss function. While this loss function is appropriate in many statistical applications, it may however not be flexible enough for financial purposes, where over- and underestimation may have different consequences. Hence a flexible asymmetric loss function is required.

The contributions of this chapter to the existing literature are as follows. First, we provide a manner to approximate the multi-day ahead VaR density

when the underlying process is described by a GARCH model. Since this class of models is a workhorse in financial risk management, we therefore give the possibility to determine the VaR term structure and to characterize the uncertainty coming from the parameters. Second, we give a rational justification to the choice of a point estimate within the VaR density based on the decision theory framework. We document how agents facing different risk perspectives can select their optimal VaR point estimate and show that the differences across agents (*e.g.*, fund and risk managers) can be substantial in terms of regulatory capital. Lastly, we extend our methodology to the Expected Shortfall alternative risk measure, and show that our simulation procedure can also be applied in a straightforward manner.

The plan of this chapter is as follows. In **Sect. 6.2**, we formally define the concept of VaR and derive the s-day ahead VaR expression under GARCH dynamics. In **Sect. 6.3**, we review some fundamentals of Bayesian decision theory and introduce the asymmetric Linex loss function. In **Sect. 6.4**, we propose an empirical application with the estimation of the VaR term structure. Finally, we extend the methodology to the Expected Shortfall risk measure in **Sect. 6.5**.

6.2 The concept of Value at Risk

In this section, we formally define the concept of VaR and determine the density of the one-day ahead VaR under the GARCH$(1,1)$ dynamics with both Normal and Student-t disturbances. The density of the VaR for time horizons larger than one day is obtained by explicitly estimating the first four moments of the conditional P&L density and approximating the percentile of interest by either using the Cornish-Fisher expansion [see Cornish and Fisher 1937] or a Student-t approximation. We consider the GARCH$(1,1)$ model for ease of exposition but the methodology can be extended, upon modifications, to higher order GARCH models as well as asymmetric specifications.

Definition 6.1 (Value at Risk). *Let Y be a univariate random variable (not necessarily continuous) with a distribution function F_Y. For a given risk level ϕ, which belongs for risk management purposes to the interval* $[0.90, 0.995]$*, the* VaR *of Y is defined by:*

$$\text{VaR}^{\phi} \doteq \inf\left\{y \in \mathbb{R} \mid F_Y(y) \geqslant \phi^c\right\}$$

where $\phi^c \doteq (1-\phi)$ for notational purposes.

Hence, the VaR is nothing else than a percentile of the distribution of Y. When the variable Y follows a standard Normal distribution, the VaR with confidence level ϕ is the ϕ^cth percentile denoted by z_{ϕ^c}; *e.g.*, $z_{0.95} = -1.64$. When the variable Y follows a standard Student-t distribution with ν degrees of freedom, the ϕ^cth percentile is denoted by $t_{\phi^c}(\nu)$; *e.g.*, $t_{0.95}(5) = -2.01$. We emphasize the notation in the Student-t case where the VaR depends on the parameter ν.

6.2.1 The one-day ahead VaR under the GARCH(1, 1) dynamics

Under a GARCH(1, 1) model with Normal disturbances, the one-day ahead VaR at risk level ϕ, estimated at time t, is given by:

$$\text{VaR}_t^{\phi}(\psi) = h_{t+1}^{1/2}(\boldsymbol{\alpha}, \beta) \times z_{\phi^c}$$

where $\psi \doteq (\boldsymbol{\alpha}, \beta)$ and h_{t+1} is the conditional variance which is computed by recursion given $\mathcal{F}_t$, the information set at time t. Hence, under Normal disturbances, the one-day ahead VaR is nothing else than a given percentile of the standard Normal distribution scaled by the conditional standard deviation.

In the case of Student-t disturbances, the one-day ahead VaR at risk level ϕ, estimated at time t, is given by:

$$\text{VaR}_t^{\phi}(\psi) = \left[\varrho(\nu) \times h_{t+1}(\boldsymbol{\alpha}, \beta)\right]^{1/2} \times t_{\phi^c}(\nu)$$

where in this case $\psi \doteq (\boldsymbol{\alpha}, \beta, \nu)$. In addition to the scale factor $\varrho(\nu) \doteq \frac{\nu-2}{\nu}$, the ϕ^cth percentile of the standard Student-t distribution depends on the model parameter ν. For both Normal and Student-t cases, the joint posterior sample can be used to simulate the density of the one-day ahead VaR at any confidence level ϕ.

6.2.2 The s-day ahead VaR under the GARCH(1, 1) dynamics

If the horizon is larger than one day, predictions for the *cumulative returns* are needed, which in turn requires multi-step predictions. The cumulative returns over an s-day horizon (starting at time t) henceforth denoted as $y_{t,s}$, are straightforwardly calculated from the single period log-returns y_{t+i} $(i = 1, \ldots, s)$ as:

$$y_{t,s} \doteq y_{t+1} + y_{t+2} + \ldots + y_{t+s} \ .$$

This follows from the definition of the one-day log-return, calculated as the logarithmic difference of asset prices.

As in the one-day ahead case, our aim is to express the VaR of the variable $y_{t,s}$ as a function of the model parameters ψ. However, it is well known that

under GARCH dynamics, no expression in closed form exists for the density of y_{t+i} when $i > 1$; hence no closed form expression is available for the density of $y_{t,s}$ either. To overcome this problem, we might use Monte Carlo simulations to generate the density of interest. That is, for a given set of parameters ψ and information set $\mathcal{F}_t$, we could simulate B paths for the P&L over s days. The VaR would then be approximated by the empirical percentile of the distribution. In order to obtain a density for the VaR itself, this evaluation would have to be handled for each ψ in the joint posterior sample. However, since the quantity of interest describes the tail of the distribution, a large amount of simulations B would be required to get an accurate VaR estimate, which might lead to an extremely costly simulation scheme.

Therefore, in order to simplify and accelerate the estimation procedure, we propose an approximation of the VaR based on the first four conditional moments of the variable $y_{t,s}$ which can be calculated analytically when ψ is known. To that aim, let us define the pth conditional moment of $y_{t,s}$ as follows:

$$\kappa_p(\psi) \doteq \mathbb{E}_{t,\psi}\left(y_{t,s}^p\right)$$

where $\mathbb{E}_{t,\psi}(\bullet) \doteq \mathbb{E}(\bullet \mid \psi, \mathcal{F}_t)$ is the conditional expectation given ψ and $\mathcal{F}_t$. The notation $\kappa_p(\psi)$ emphasizes the fact that the pth conditional moment is a function of ψ and the time index is suppressed to simplify the notation. The explicit calculation of the first four moments is possible using the multinomial formula which gives the pth power of the cumulative return as follows:

$$\begin{aligned} y_{t,s}^p &= \left(\sum_{i=1}^{s} y_{t+i}\right)^p \\ &= \sum_{\substack{i_1,\ldots,i_s \\ i_1+\ldots+i_s=p}} \frac{p\,!}{i_1!\cdots i_s!} \times y_{t+1}^{i_1}\cdots y_{t+s}^{i_s}\,. \end{aligned}$$

Calculations given in **Props. C.1** and **C.3** (see **App. C**) show that under the GARCH$(1,1)$ specification, the first and third conditional moments are zero which implies that the conditional density of $y_{t,s}$ is symmetric around zero. Calculations for the second conditional moment of $y_{t,s}$ in **Prop. C.2** (see **App. C**) yield:

$$\kappa_2(\psi) = \sum_{i=1}^{s} \mathbb{E}_{t,\psi}(h_{t+i})$$

where:

$$\mathbb{E}_{t,\psi}(h_{t+i}) = \alpha_0 + \rho_1 \, \mathbb{E}_{t,\psi}(h_{t+i-1}) \tag{6.1}$$

and $\rho_1 \doteq (\alpha_1 + \beta)$. Expression (6.1) can be evaluated recursively from $\mathbb{E}_{t,\psi}(h_{t+1}) = h_{t+1}(\psi)$ since this value is known given ψ and $\mathcal{F}_t$. For the fourth conditional moment, the calculations in **Prop. C.4** (see **App. C**) yield the following expression:

$$\kappa_4(\psi) = \kappa_\varepsilon \sum_{i=1}^{s} \mathbb{E}_{t,\psi}(h_{t+i}^2) + 6 \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} \mathbb{E}_{t,\psi}(y_{t+i}^2 \, y_{t+j}^2) \tag{6.2}$$

where:

$$\mathbb{E}_{t,\psi}(h_{t+i}^2) = \alpha_0^2 + \tau_1 \, \mathbb{E}_{t,\psi}(h_{t+i-1}) + \tau_2 \, \mathbb{E}_{t,\psi}(h_{t+i-1}^2) \tag{6.3}$$

and:

$$\mathbb{E}_{t,\psi}(y_{t+i}^2 \, y_{t+j}^2) = \alpha_0 \left(\frac{1 - \rho_1^{j-i}}{1 - \rho_1} \right) \mathbb{E}_{t,\psi}(h_{t+i}) + \rho_1^{j-i-1} \, \rho_2 \, \mathbb{E}_{t,\psi}(h_{t+i}^2) \; . \tag{6.4}$$

In expression (6.2), the parameter κ_ε denotes the fourth moment of the disturbances in the GARCH$(1,1)$ process; in the case of Normal disturbances, $\kappa_\varepsilon = 3$, while for (scaled) Student-t disturbances, $\kappa_\varepsilon = \frac{3(\nu-2)}{\nu-4}$. The parameters τ_1, τ_2 and ρ_2 are functions of the set of parameters ψ, respectively given by:

$$\begin{aligned} \tau_1 &\doteq 2\alpha_0(\alpha_1 + \beta) \\ \tau_2 &\doteq \kappa_\varepsilon \alpha_1^2 + \beta(2\alpha_1 + \beta) \end{aligned}$$

and:

$$\rho_2 \doteq \kappa_\varepsilon \alpha_1 + \beta \; .$$

The conditional expectations of expressions (6.3) and (6.4) can be evaluated recursively from $\mathbb{E}_{t,\psi}(h_{t+1}) = h_{t+1}(\psi)$ and $\mathbb{E}_{t,\psi}(h_{t+1}^2) = h_{t+1}^2(\psi)$ since these values are known given ψ and $\mathcal{F}_t$.

As previously stated, the conditional moments κ_i $(i = 1, \ldots, 4)$ are used to estimate the percentile of the conditional density of the cumulative return over an s-day horizon. We propose two approaches to determine this percentile. The first method is the well-known Cornish-Fisher expansion by [see Cornish and Fisher 1937] which consists in a transformation of the percentile of the standard Normal density to account for non-zero skewness (*i.e.*, asymmetry) and excess

kurtosis (*i.e.*, fat tails). In our context, κ_1 and κ_3 are zero so that the Cornish-Fisher formula simplifies. We thus obtain the following approximation for the s-day ahead VaR ($s > 1$) at risk level ϕ, estimated at time t:

$$\text{VaR}_{t,s}^{\phi}(\psi) \approx \kappa_2^{1/2}(\psi) \times \left[z_{\phi^c} + \frac{1}{24}(z_{\phi^c}^3 - 3z_{\phi^c}) \left(\frac{\kappa_4(\psi)}{\kappa_2^2(\psi)} - 3 \right) \right] \tag{6.5}$$

where we recall that $\phi^c \doteq (1 - \phi)$ and z_{ϕ^c} is the ϕ^cth percentile of the standard Normal distribution. From expression (6.5), we can notice the impact of the excess kurtosis $\left(\frac{\kappa_4}{\kappa_2^2} - 3\right)$ on the VaR. Since conditional moments are functions ψ, so is the VaR.

The Cornish-Fisher expansion is widely used in practice due to its simplicity, and its accuracy is sufficient in many situations, especially when the distribution of interest is close to the Normal. In this case, the Cornish-Fisher expansion provides a small correction for the non-zero skewness and excess kurtosis. However, as pointed out by Jaschke [2002], the Cornish-Fisher expansion may suffer from important deficiencies in pathological situations, for instance, when the kurtosis of the distribution we aim to approximate is high. We note in particular that:

- The approximation may yield a distribution which is not necessarily monotone. Hence, we may be faced with situations where the risk capital allocated for a 1% chance event would be lower than the capital allocated for a 5% chance event!

- The approximation has the *wrong tail behavior*, *i.e.*, the Cornish-Fisher approximation for the VaR at risk level ϕ becomes less and less reliable for $\phi \to \{0, 1\}$.

These drawbacks can have serious consequences for risk management systems and we propose therefore a second method to approximate the percentiles of interest. Since the density we aim to approximate is symmetric around zero, we simply fit a Student-t density to the second and fourth conditional moments κ_2 and κ_4. First, we determine the conditional kurtosis of $y_{t,s}$, denoted by $\widehat{\kappa}$, as follows:

$$\widehat{\kappa}(\psi) = \frac{\kappa_4(\psi)}{\kappa_2^2(\psi)} \, .$$

From there, we estimate the degrees of freedom parameter $\widehat{\nu}$ of the Student-t density. The relation between $\widehat{\kappa}$ and $\widehat{\nu}$ is given by:

$$\widehat{\nu}(\psi) = \frac{6 - 4\,\widehat{\kappa}(\psi)}{3 - \widehat{\kappa}(\psi)} \, .$$

Finally, the VaR is estimated by the appropriate percentile of the standard Student-t density scaled by the conditional standard deviation $\kappa_2^{1/2}$. This yields the following approximation for the s-day ahead VaR ($s > 1$) at risk level ϕ, estimated at time t:

$$\mathrm{VaR}_{t,s}^{\phi}(\psi) \approx \left(\frac{\widehat{\nu}(\psi)-2}{\widehat{\nu}(\psi)}\right)^{1/2} \times t_{\phi^c}\big(\widehat{\nu}(\psi)\big) \times \kappa_2^{1/2}(\psi)\ . \tag{6.6}$$

This approximation is a function of the set of parameters ψ. Hence, as with the Cornish-Fisher approximation, the density of the VaR can be estimated at low cost by simulating from the joint posterior sample.

In the Bayesian context, we can integrate out the parameter uncertainty to end up with a single VaR point estimate. This problem is solved by the estimation of the *predictive density* which is defined as the density of future s-day ahead observations, $\mathbf{y}_{t:s} \doteq (y_{t+1} \cdots y_{t+s})'$ for $s \geqslant 1$, conditioned on past observations $\mathbf{y}_{0:t} \doteq (y_1 \cdots y_t)'$ (also denoted by $\mathcal{F}_t$), but marginalized over ψ. More formally, the predictive density is defined as follows:

$$p(\mathbf{y}_{t:s} \mid \mathbf{y}_{0:t}) = \int p(\mathbf{y}_{t:s} \mid \psi, \mathbf{y}_{0:t}) p(\psi \mid \mathbf{y}_{0:t}) \mathrm{d}\psi \tag{6.7}$$

where $p(\mathbf{y}_{t:s} \mid \psi, \mathbf{y}_{0:t})$ is the conditional density of $\mathbf{y}_{t:s}$ given $(\psi, \mathbf{y}_{0:t})$ and the marginalization is with respect to the posterior density $p(\psi \mid \mathbf{y}_{0:t})$. In general, the predictive density is not available in closed form. However, one can use the posterior sample in conjunction with the method of composition to produce a sample of draws from the predictive density. We simulate a draw $\mathbf{y}_{t:s}^{[j]}$ from the density $p(\mathbf{y}_{t:s}, \psi \mid \mathbf{y}_{0:t})$ as follows:

$$\begin{aligned} \psi^{[j]} &\sim p(\psi \mid \mathbf{y}_{0:t}) \\ \mathbf{y}_{t:s}^{[j]} &\sim p(\mathbf{y}_{t:s} \mid \psi^{[j]}, \mathbf{y}_{0:t}) \end{aligned} \tag{6.8}$$

where the second step in the simulation process is possible by using the method of composition:

$$p(\mathbf{y}_{t:s} \mid \psi, \mathbf{y}_{0:t}) = \prod_{i=1}^{s} p(y_{t+i} \mid \psi, \mathbf{y}_{0:(t+i-1)})\ . \tag{6.9}$$

The collection of simulated values $\{\mathbf{y}_{t:s}^{[j]}\}_{j=1}^{J}$ is generated from the predictive density in (6.7) and the predictive VaR is a percentile of this density. For the one-day ahead VaR, we only need to consider the first component of vector $\mathbf{y}_{t:s}^{[j]}$ whereas for the s-day ahead VaR, we must sum the components of $\mathbf{y}_{t:s}^{[j]}$ to simulate the predictive density for $y_{t,s}$. From (6.7), we notice that the predictive

VaR is a quantile of a mixture density. Therefore, it can be viewed as an extension of the case where there is no parameter uncertainty (*i.e.*, ψ is constant); in this case, the predictive VaR would simply be estimated by a percentile of $p(\mathbf{y}_{t:s} \mid \psi, \mathbf{y}_{0:t})$.

To end this section, we illustrate the quality of the Cornish-Fisher and Student-t approximations through a simulation study. To that aim, we estimate the GARCH$(1,1)$ model with Student-t disturbances for the 750 Deutschmark vs British Pound foreign exchange log-returns used in the empirical analysis of **Chap. 3**. We arbitrarily select a set of parameters $\psi \doteq (\boldsymbol{\alpha}, \beta, \nu)$ in the joint posterior sample:

$$\boldsymbol{\alpha} = \begin{pmatrix} 0.036 \\ 0.297 \end{pmatrix} \quad , \quad \beta = 0.626 \quad \text{and} \quad \nu = 5.4$$

and estimate the VaR using formulae (6.5) and (6.6) for $s = 10$ and ϕ ranging from 0.001 to 0.999 with a step size of 0.001. This procedure allows to draw two approximations for the distribution of $y_{750,10}$. These distributions are compared with the distribution obtained by simulating 10'000 paths of the process over ten days, using (6.9). Under the model specification and the selected ψ, we obtain $\kappa_2 = 3.8$ and $\kappa_4 = 492$, implying a conditional kurtosis $\widehat{\kappa} = 34$ and a degrees of freedom parameter $\widehat{\nu} = 4.2$. The simulated distribution clearly exhibits heavier tails than the Normal distribution.

On the left-hand side of **Fig. 6.1**, we display the two approximations together with the distribution obtained by simulation. The Cornish-Fisher approximation is shown in dotted line, the Student-t approximation in dashed line and the empirical distribution in solid line. From this figure, it is almost impossible to distinguish the approximation based on the Student-t distribution from the simulated distribution. In contrast to this, the Cornish-Fisher expansion produces a S-shaped, non-monotone distribution. The four shaded regions delimit the extreme quantiles, at risk level $\phi \in \{0.01, 0.05, 0.95, 0.99\}$. We notice that the approximations for $\phi \in \{0.05, 0.95\}$ are quite similar for the Cornish-Fisher and the Student-t approaches. However, the difference is substantial in the case where $\phi \in \{0.01, 0.99\}$. In the middle graph of **Fig. 6.1**, we show a zoom of the previous graph over the domain $[2, 6] \times [0.94, 1]$. We can see that the Student-t approximation fits the distribution of interest well. Hence, in this particular example, the graphical comparison indicates that the Cornish-Fisher expansion fails in approximating the distribution of interest. On the other hand, the approximation based on the Student-t distribution seems to provide an adequate approximation of the whole distribution. To complete the simulation study, we display, on the right-hand side of **Fig. 6.1**, the difference between

the simulated distribution and the Student-t approximation as a function of ϕ^c. The dotted lines delimit the 95% confidence band for the simulation, estimated by replicating 500 times the empirical distribution. We note that the difference lies within the $[-0.1, 0.1]$ interval for risk levels ranging from 0.05 to 0.95. For other percentiles, the error increases together with the width of the confidence band. However, the confidence interval still contains the value of zero, indicating a good approximation in the tails too.

Finally, we note that other approximation methods of the whole density for $y_{t,s}$ can be obtained [see, *e.g.*, Highfield and Zellner 1988]. This is of interest when the density we aim to approximate is skewed, since in this case, the Student-t approximation would fail. Such asymmetric densities arise, *e.g.*, with asymmetric GARCH models [see Engle 2004, p.415].

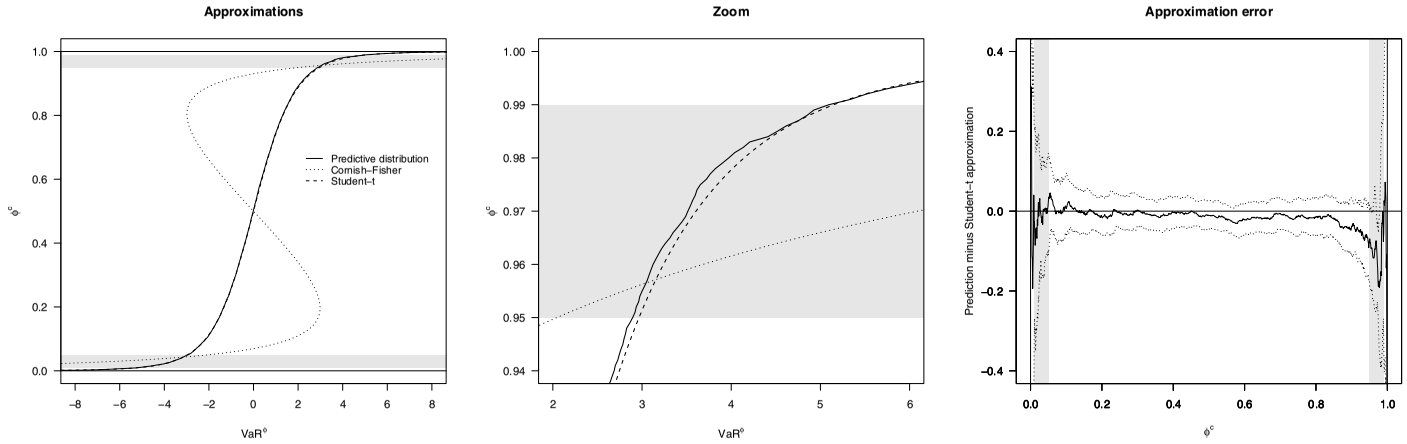

Fig. 6.1. Cornish-Fisher and Student-t approximations. On the left-hand side, we display the distribution given by the Cornish-Fisher (in dotted line) and the Student-t (in dashed line) approximations together with the simulated distribution based on 10'000 paths (in solid line). In the middle graph, we zoom the plot over the $[2, 6] \times [0.94, 1]$ domain. On the right-hand side, we plot the difference between the simulated distribution and the Student-t approximation. The dotted lines delimit the confidence band obtained by bootstrapping the simulated distribution 500 times. The shaded regions indicate the 1st, 5th, 95th and 99th percentiles.

6.3 Decision theory

Using the Bayesian approach leads to an interesting problem of decision theory: the choice of a Bayes point estimate within the whole VaR posterior density. In this section, we present a short review of decision theory and introduce the asymmetric Linex loss function. The use of asymmetric loss functions better characterizes the views of market participants where the impact of underestimation and overestimation can be significantly different. The Linex loss function has proved to be advantageous in many fields, especially for financial applications.

To keep the notation as general as possible, the decisions are formulated in terms of ω which can either be viewed as a one-dimensional parameter or a point forecast.

6.3.1 Bayes point estimate

As is the case in economics, statistical decisions are made based on expected ranking. In economics this ranking is achieved with the help of a utility function while we use a loss function in statistics. The Bayesian statistical decision consists in the choice of a point estimate over the posterior density of the parameters.

Let us assume that a decision maker needs to choose a point estimate $\widehat{\omega}$ and the true state of nature is ω. In the Bayesian framework, the parameter ω is random and its uncertainty is fully characterized by its posterior density $p(\omega \mid \mathbf{y})$. Furthermore, we define the loss function $\mathscr{L}(\widehat{\omega}, \omega)$ which is the loss incurred when ω is the true state of nature and $\widehat{\omega}$ is a point estimate. Then, the *Bayes estimate*, also referred to as the *optimal point estimate*, denoted by $\widehat{\omega}_{\mathscr{L}}$, is the parameter which minimizes the posterior risk $R_{\mathscr{L}}(\widehat{\omega} \mid \mathbf{y})$. Formally, the Bayes estimate is defined as follows:

$$\widehat{\omega}_{\mathscr{L}} \doteq \arg\min_{\widehat{\omega}} R_{\mathscr{L}}(\widehat{\omega} \mid \mathbf{y}) \tag{6.10}$$

where:

$$R_{\mathscr{L}}(\widehat{\omega} \mid \mathbf{y}) \doteq \int \mathscr{L}(\widehat{\omega}, \omega) p(\omega \mid \mathbf{y}) \mathrm{d}\omega \quad .$$

The problem of point estimation of a location parameter or forecast is most often treated as a symmetric problem in which positive and negative estimation errors of the same magnitude are considered to be equally serious; thus, the loss function $\mathscr{L}$ is symmetric. The most used loss functions are the squared error loss

(henceforth SEL), $\mathscr{L}(\widehat{\omega}, \omega) = (\widehat{\omega} - \omega)^2$, and the absolute error loss (henceforth AEL), $\mathscr{L}(\widehat{\omega}, \omega) = |\widehat{\omega} - \omega|$. Indeed most of the existing VaR literature ignores the asymmetric loss relevant for different economic agents. However, the impact of overestimating or underestimating VaR can be quite different. As quoted by Knight, Satchell, and Wang [2003, p.335]:

> "From the perspective of the fund manager in a bank, the loss of overestimating is usually much greater than that of underestimating, as the reserve capital exceeds the capital required by regulation and earns little or no return at all. On the other hand, from the regulator's perspective, systematic failure would be increasing in the degree to which each bank's losses actually exceed their capital reserves. So underestimating will result in more loss for the regulator."

This suggests the need of an appropriate asymmetric loss function when choosing a point estimate within the VaR density.

We point out that, in light of the capital structure theory, the relevance of an asymmetric loss function for banks is questionable since capital reserves do not have to be held in cash. In this case, if regulators demand a higher capital reserve, the bank will just have to rearrange its capital structure, which does not necessarily increase capital cost. This suggests that bank managers should in fact not be interested in minimizing regulatory capital. While this argument is valid for the bank as a whole, it does not hold at the trading desk level since it is common that traders and fund managers need a buffer in cash for facing market risk exposures.

6.3.2 The Linex loss function

The Linex loss function is employed in the analysis of several central statistical estimation and prediction problems. Varian [1974] motivates the use of the Linex loss function on the basis of an example in which there is a natural imbalance in the economic results of estimation errors of the same magnitude. Varian argues that the Linex loss is a rational manner to formulate the consequences of estimation errors in real estate assessment. Christoffersen and Diebold [1996, 1997] use the Linex loss function in a study of optimal point prediction where different asymmetric loss functions are tested. More recently, Hwang, Knight, and Satchell [1999, 2001] derive the Linex one-day ahead volatility forecast for various volatility models. The empirical results of these authors suggest the Linex loss function to be particularly well-suited in financial applications. In addition, we note that other fields than quantitative finance make use of the Linex loss function. An example is given in the field of hydrology with the

estimation of peak water level in the construction of dams or levies. In that case, overestimation represents a conservative error which increases construction costs, while underestimation corresponds to the much more serious error in which overflows might lead to huge damages in the adjacent communities.

In its reduced form, the Linex loss function is given by:

$$\mathscr{L}(\widehat{\omega}, \omega) = \exp[a\Delta] - a\Delta - 1 \tag{6.11}$$

where $a \in \mathbb{R}^*$ and $\Delta \doteq (\widehat{\omega} - \omega)$ denotes the scalar estimation error in using $\widehat{\omega}$ when estimating ω. From expression (6.11), we note that:

- $\mathscr{L}$ is a convex function of Δ;
- $\mathscr{L}$ is decreasing for $\Delta \in]-\infty, 0[$ and increasing for $\Delta \in]0, \infty[$;
- for $a > 0$, $\mathscr{L}$ grows exponentially in positive Δ but behaves approximately linearly for negative values of Δ. In this case, the Linex loss function imposes a substantial penalty for overestimation, *i.e.*, when $\widehat{\omega} > \omega$;
- for $|a| \simeq 0$, $\mathscr{L}$ is almost symmetric and not far from a squared error loss function. Indeed, on expanding:

  $$\exp[a\Delta] \simeq 1 + a\Delta + \frac{(a\Delta)^2}{2}$$

  and replacing it in expression (6.11), the loss function becomes proportional to the SEL function. Thus for small values of a, the SEL function is approximately nested within the Linex function.

In the upper graph of **Fig. 6.2**, we display the Linex loss function for parameter $a = 0.5$ in dotted line, $a = 1$ in dashed line and $a = 2$ in solid line. We can notice the impact on the shape of the loss function of larger values of parameter a. Indeed, as a increases, the asymmetry accentuates. In the lower part of the figure, we show the Linex loss function for parameter $a = 0.1$ together with the appropriately scaled SEL function. The loss functions are almost the same on the interval.

As verified by Zellner [1986], the derivation of the Bayes estimator of ω is straightforward under the Linex loss function (6.11). The optimization problem (6.10) yields:

$$\widehat{\omega}_{\mathscr{L}} = -\frac{1}{a} \ln \left(\int \exp[-a\omega] p(\omega \mid \mathbf{y}) \mathrm{d}\omega \right)$$

provided that the integral is finite. Hence, the key to Linex estimation is to find the moment generating function of $p(\omega \mid \mathbf{y})$, which can be estimated using the posterior sample.

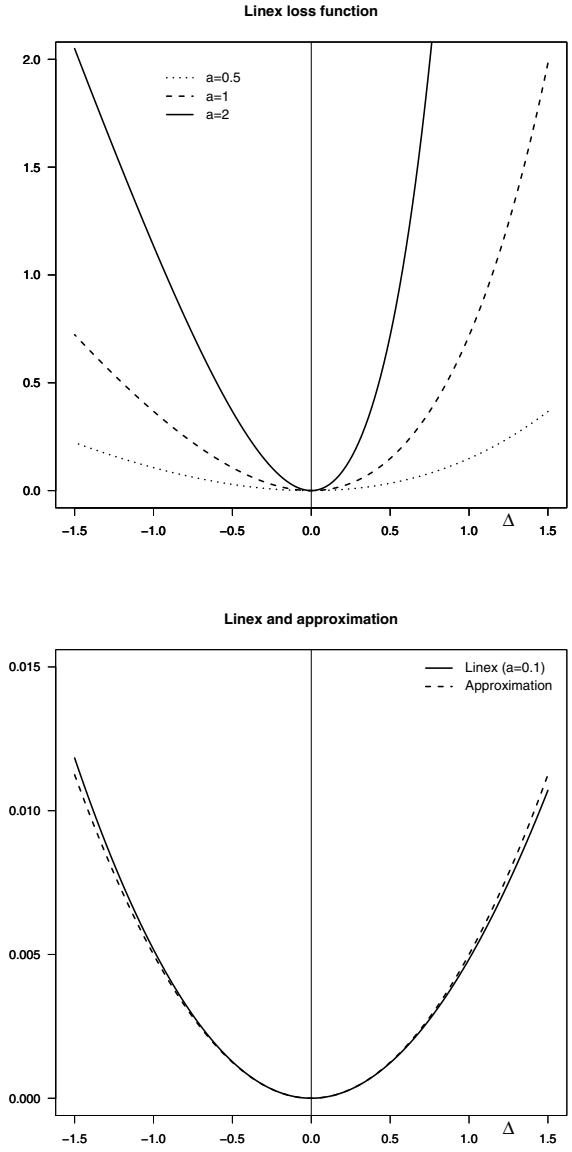

Fig. 6.2. Linex loss function. In the upper graph we plot the Linex function for different values of parameter a. In the lower part, we show the Linex loss function for parameter $a = 0.1$ in solid line together with the SEL function in dashed line. We recall that $\Delta \doteq (\widehat{\omega} - \omega)$ where $\widehat{\omega}$ is the point estimate and ω is the true parameter value.

6.3.3 The Monomial loss function

We end this section by noting that other interesting asymmetric loss functions are readily available in the statistical literature. A general class of asymmetric loss functions, referred to as *Monomial-splined* functions by Thompson and Basu [1995], is defined as follows:

$$\mathscr{L}(\Delta) \doteq \begin{cases} a_1 \times |\Delta|^p & \text{if} \quad \Delta \geqslant 0 \\ a_2 \times |\Delta|^p & \text{if} \quad \Delta < 0 \end{cases} \tag{6.12}$$

where $\Delta \doteq (\widehat{\omega} - \omega)$, $a_i \in \mathbb{R}^+$ $(i = 1, 2)$ and $p \in \mathbb{N}^*$. This class provides asymmetric loss functions for $a_1 \neq a_2$; when $a_1 > a_2$, an overestimation incurs more loss than an underestimation and inversely when $a_1 < a_2$. From expression (6.12), we note the two following special cases:

- $p = 1$: linear-linear loss function; when $a_1 = a_2$ we obtain the AEL function;
- $p = 2$: quadratic-quadratic loss function; when $a_1 = a_2$ we obtain the SEL function.

It is often easier to work with a reparametrization of expression (6.12). Let us define $q \doteq \frac{a_1}{a_1 + a_2}$ and make use of the homogeneity property so that we obtain the following loss function:

$$\mathscr{L}(\Delta) = \left(q + (1 - 2q)\mathbb{I}_{\{\Delta < 0\}}\right) |\Delta|^p \, . \tag{6.13}$$

This function is particularly interesting when $p = 1$. In this case, it turns out that the optimal point estimate $\hat{\omega}_{\mathscr{L}}$ is nothing else than the qth percentile of the posterior density $p(\omega \mid \mathbf{y})$. Hence, expression (6.13) gives a statistical justification to the choice of a posterior percentile as Bayes point estimate. *E.g.*, choosing the 95th percentile of the VaR density would be optimal for an agent whose loss function is given by (6.13) with $q = 0.95$. Finally, we note that numerical methods are needed to find Bayes point estimates for the loss function (6.13) when $p > 1$.

In **Fig. 6.3**, we display the loss function given in expression (6.13) for parameter $q = 0.5$ in dotted line, $q = 0.75$ in dashed line and $q = 0.95$ in solid line. As parameter q increases, the asymmetry becomes more pronounced and the impact of an overestimation, *i.e.*, $\Delta > 0$, is larger compared to the impact of an underestimation. The function is symmetric for $q = 0.5$.

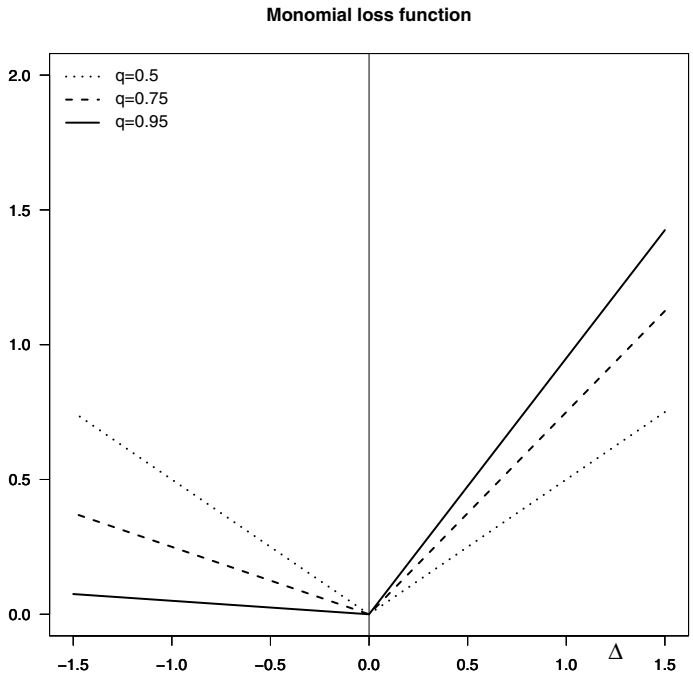

Fig. 6.3. Monomial loss function for different values of parameter q. We recall that $\Delta \doteq (\widehat{\omega} - \omega)$ where $\widehat{\omega}$ is the point estimate and ω is the true parameter value.

6.4 Empirical application: the VaR term structure

In this section, we estimate the term structure of the VaR when the P&L dynamics is described by a GARCH$(1,1)$ model with Normal and Student-t disturbances. Our analysis is inspired by the paper of Guidolin and Timmermann [2006] which considers the impact of different econometric specifications to the shape of the VaR term structure. While the authors report significant differences between the models, they do not account for parameter uncertainty in their analysis. This is indeed a weakness of their approach, as recognized by the authors [see Guidolin and Timmermann 2006, p.307]:

> "We ignored parameter estimation uncertainty in our analysis, but this could have important effects on the results."

The Bayesian approach provides a natural framework for investigating this point. As shown in **Sects. 6.2.1** and **6.2.2**, the VaR can be expressed as a function of the GARCH$(1,1)$ parameters under both Normal and Student-t

specifications. Consequently, the parameter uncertainty estimated by the joint posterior sample can be used to estimate the density of the VaR in a convenient manner.

An additional justification for the use of the Bayesian approach in our context is given by Miazhynskaia and Aussenegg [2006] who compare the Bayesian and traditional techniques for estimating GARCH models. In particular, they conclude that the Bayesian approach is an adequate framework with less uncertainty in VaR estimates compared to other VaR methods such as resampling technique and asymptotic Normal approximation. They also mention the interesting issue of determining a single VaR point estimate [see Miazhynskaia and Aussenegg 2006, p.262]:

> "Open questions for future research are how the total VaR distribution can be used in market risk management and how to account for VaR uncertainty in choosing traditional VaR point estimates used to calculate capital requirements for financial institution."

This is precisely what we aim to achieve in a rational manner through the decision theory framework.

6.4.1 Data set and estimation design

Our empirical analysis uses the Deutschmark vs British Pound exchange rate daily log-returns over a sample period ranging from January 3, 1985, to December 31, 1991, for a total of 1'974 observations. This data set was used in the empirical analysis of **Chap. 3**.

We consider daily log-returns so that the VaR term structure focuses on short-term horizons. This is of primary interest for traders and risk managers who adjust the bank's portfolios on a daily basis. The methodology can also be applied to longer time span log-returns as this is done in Guidolin and Timmermann [2006]. In this manner, a term structure for mid- and long-term horizons is obtained. Note however that modeling monthly or quarterly financial data would probably require more complicated models than the GARCH$(1,1)$ specification, to account for structural breaks in the time series, for instance. This could drastically complicate the approximation methodology developed in **Sect. 6.2.2**, in particular to find the first four moments conditioned on the model parameters and information set.

The estimation of the GARCH$(1,1)$ models is achieved by using the rolling window methodology. This procedure is heavily used in finance and financial risk management. The rationale behind it is to act as if we were moving over time, using past observations to estimate the model and test the performance

over a prediction window. This is based on the assumption that older data are not available or are irrelevant due to structural breaks, which are so complicated that they cannot be modeled. Conceptually, this method aims to take account for more recent information in a simplified framework and it has proved to be effective in many financial applications.

We structure the estimation procedure as follows: 750 log-returns, which is about three trading years, are used to estimate the models. Then, the next 50 log-returns, which is slightly less than one quarter, are used as a forecasting window. In the next step, the estimation and forecasting windows are moved together by 50 days ahead, so that the forecasting windows do not overlap. In this manner, the model parameters are updated every quarter and the estimation methodology fulfills the recommendations of the Basel Committee in the use of internal models [see Basel Committee on Banking Supervision 1996b]. When applied to our data set, the estimation design leads to the generation of 24 estimation windows. The non-overlapping forecasting windows represent a total of $24 \times 50 = 1'200$ observations. An illustration of the methodology is shown in **Fig. 6.4** where we plot the first three observation windows excerpt from our data set; the vertical lines separate the estimation and the forecasting windows.

Note that the standard practice when using the rolling window methodology in the context of GARCH models consists in moving the window by a single day ahead. While this procedure can be achieved quite rapidly when estimating the model by the Maximum Likelihood technique, this can become a computational burden with the Bayesian approach, since at each step, we need to run the MCMC scheme again. This problem is however only relevant in an *ex-post* framework; a portfolio or risk manager could run the Bayesian estimation of the model every day, without encountering these computational difficulties.

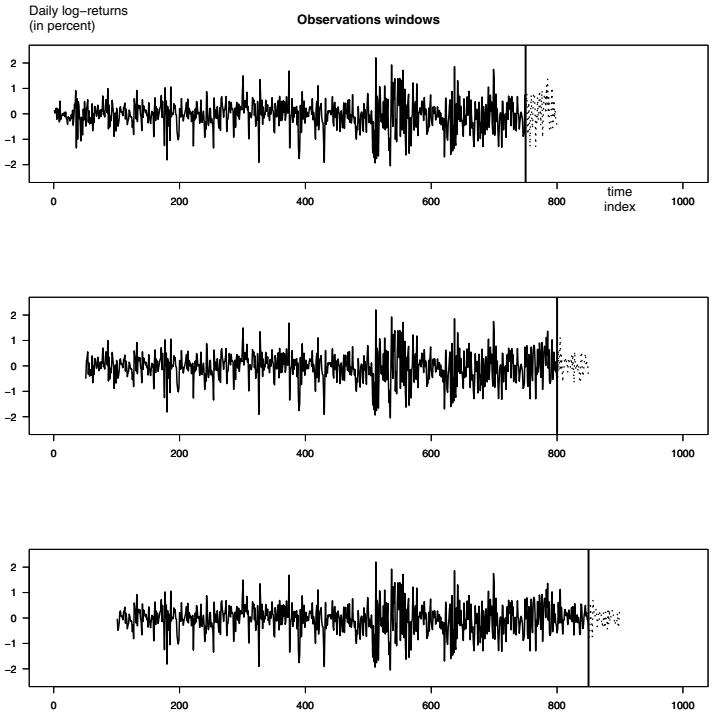

Fig. 6.4. Estimation and forecasting windows (first 3 windows out of 24). The 750 log-returns used for the estimation are shown in solid line and the 50 out-of-sample log-returns are shown in dotted line. The vertical line separate the estimation and the forecasting windows. At each step in the procedure, both windows are moved together by 50 days ahead.

6.4.2 Bayesian estimation

As prior densities for the scedastic function's parameters $\boldsymbol{\alpha}$ and β, we choose truncated Normal densities with zero mean vectors and diagonal covariance matrices whose variances are set to 10'000. In the case of Student-t disturbances, we use the translated Exponential as a prior density for the degrees of freedom parameter; the hyperparameters are set to $\lambda = 0.01$ and $\delta = 4$; the prior mean is therefore 104 and the prior variance 10'000. The parameter δ is set so that the conditional variance and conditional fourth moment exist, which allows the use of the approximation for the predictive density based on the first four moments (see **Sect. 6.2.2** for details). For each estimation window, two chains are run for 10'000 passes each and the convergence diagnostic test by Gelman and Rubin [1992] is applied to guarantee a good convergence of the algorithm. From the

overall MCMC output, we discard the first 5'000 draws and merge the two chains to get a final sample of length 10'000.

6.4.3 The term structure of the VaR density

As a preliminary analysis, we consider the first observation window excerpt from our data set and estimate the (conditional) term structure of the VaR at risk level $\phi = 0.95$. We consider risk horizons ranging from one day to fifteen days for the VaR density estimated under both GARCH$(1, 1)$ Normal and Student-t models using approximation (6.6). The two term structures are depicted in **Fig. 6.5**; the lines give the median point estimates while the shaded regions depict the 95% confidence intervals of the densities. From this graph, we note that the VaR is a monotone decreasing function of the time horizon for both models. The Student-t specification leads to lower median point estimates for time horizons larger than five days while the differences for smaller horizons are less pronounced. We also notice that the VaR uncertainty increases for both models with respect to the time horizon. Furthermore, the GARCH$(1, 1)$ model with Student-t disturbances leads to higher uncertainty in VaR at each horizon compared to the Normal specification. Finally, we note that the VaR density is almost symmetric for all horizons in the Normal case while the density is left-skewed for the Student-t model.

This first static analysis indicates that the density of the VaR is influenced by the time horizon as well as the specification of the model disturbances; leptokurtic disturbances lead to a larger uncertainty in the VaR as well as a left-skewed density.

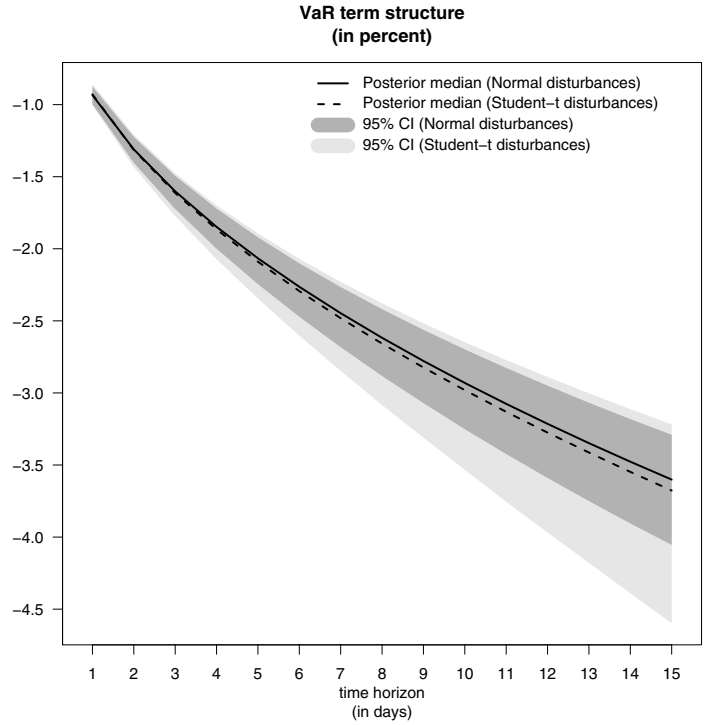

Fig. 6.5. Term structures of the VaR density at risk level $\phi = 0.95$ for the GARCH$(1,1)$ model with Normal and Student-t disturbances. Both densities are based on 10'000 draws from the joint posterior sample of the models' parameters.

6.4.4 VaR point estimates

We now investigate the differences in VaR point estimates under different loss functions of the forecasters. The comparison of point estimates over the out-of-sample window has two purposes. First, it will provide a statistical counterpart to the graphical findings observed in the preceding section. Second, since the VaR point estimates are used to calculate capital requirements for financial institutions, this analysis will give a first idea on how large the differences in risk capital are between agents facing different risk perspectives. In what follows, we concentrate the analysis on horizon $s \in \{1, 5, 10\}$ and risk level $\phi \in \{0.95, 0.99\}$. Note that the particular case $(s = 1, \phi = 0.95)$ corresponds to the criterion employed by the popular RiskMetrics benchmark [see RiskMetrics Group 1996]. The case $(s = 10, \phi = 0.99)$ is recommended by the Basel Committee on Banking Supervision [1996a] and aims to take in consideration liquidity constraints encountered by the bank.

To compare the VaR point estimates resulting from the use of different loss functions, we use the following methodology: for each point in the out-of-sample data set, we estimate the density of the VaR for the three different time horizons and the two risk levels using approximation (6.6). Then, for each density, we determine a point estimate for the VaR which solves the optimization problem (6.10) for a given loss function $\mathscr{L}$; this point estimate is denoted by $\widehat{\mathrm{VaR}}^{\phi}_{\mathscr{L},t,s}$. In what follows, we use the asymmetric Linex loss, the absolute error loss (AEL) as well as the squared error loss (SEL) functions for $\mathscr{L}$, the latter being considered as the benchmark in our analysis. As an additional point estimate, we use the predictive VaR defined in (6.7). In this case, for each draw in the joint posterior sample, we generate a draw from the predictive density using (6.8) and the predictive VaR is obtained by calculating the appropriate percentile of the simulated density.

Some comments regarding the different perspectives of VaR point estimation are in order here. In the first approach, we estimate the density of the VaR by simulating from the joint posterior sample and choose a point estimate which is optimal for a given loss function (Linex, SEL and AEL). The parameter uncertainty is integrated out in the second step of the procedure, when the posterior risk is minimized (see **Sect. 6.3.1**). This methodology is natural in combining estimation and decision making, and gives therefore an additional flexibility to the user. In the second approach, the parameter uncertainty is integrated out by averaging the conditional densities of the cumulated returns over the joint posterior density of the model parameters. In this case, the VaR point estimate (*i.e.*, the predictive VaR) is not related to the risk preferences of an agent and is the same for both regulators and fund managers. Hence, agents differ not in the estimation of the VaR, but in the way they would make use of the point estimate afterwards. This approach is natural in Bayesian statistics but it is not the most sophisticated. It can be viewed as an extension of the case where there is no parameter uncertainty (*i.e.*, ψ is constant); in this case, the predictive VaR would simply be estimated by a percentile of the conditional density of future observations. The justification of the predictive VaR on the grounds of the decision theory still needs to be established.

To find an optimal VaR point estimate under the Linex loss function, we need first to choose a value for the parameter a in expression (6.11). The most direct way is to elicit the parameter through detailed discussion with a fund or risk manager. Due to the unavailability of this type of data, we will thus rely on the estimation of Knight *et al.* [2003] where the authors found $a \approx 3$ based on Standard and Poors 500 index data. In our framework, since the VaR estimates are negative percentages, this positive parameter a implies a larger

penalty when the estimated VaR is underestimated (in absolute value) compared to the true VaR, *i.e.*, $\widehat{\text{VaR}} > \text{VaR}$. Hence, the Linex optimal point estimate will be conservative in the sense that it will be located in the left tail of the density to avoid underestimation. Such loss function can thus be attributed to a regulator or a risk manager whose aim is to avoid systematic failure in VaR estimation. For comparison purposes, we also consider the Linex function with parameter $a = -3$. In that case, overestimating (in absolute value) the VaR, *i.e.*, $\widehat{\text{VaR}} < \text{VaR}$, leads to a larger penalty so that the Linex point estimate will be located in the right tail of the VaR density. This is the loss perspective of a trader or fund manager whose aim is to save regulatory capital since it earns little or no return at all (as pointed out previously, traders hold a buffer in cash for facing market risk exposures). The AEL and SEL functions correspond to the perspective of an agent for which under- and overestimation are equally serious; the SEL leads however to a larger penalty for large deviations from the true VaR, compared to the AEL function.

Once the time series of VaR point estimates for a given loss function has been obtained, we compare its values to the SEL benchmark. More precisely, for a given risk level ϕ and time horizon s we compute a time series of differences between the VaR point estimates obtained with the loss function $\mathscr{L}$ and the SEL benchmark. Then, we estimate the average of deviations over the $N = 1$'200 out-of-sample observations as follows:

$$\frac{1}{N}\sum_{t=1}^{N}\left(\widehat{\text{VaR}}^{\phi}_{\mathscr{L},t,s} - \widehat{\text{VaR}}^{\phi}_{\text{SEL},t,s}\right) .$$

Results of the average deviations are reported in **Table 6.1** where the table entries are given in hundredth percent, *i.e.*, multiplied by 100, for convenience. The upper panel gives the results for the GARCH$(1,1)$ model with Normal disturbances while the lower panel presents the results for the Student-t case. From this table, we first note that the average deviations vary considerably between table entries; the minimum value is 0.0001% in the case of the predictive VaR for Normal disturbances with $(s = 1, \phi = 0.95)$ while the maximum is 0.496% in the case of the Linex VaR $(a = 3)$ for Student-t disturbances with $(s = 10, \phi = 0.99)$. Moreover, we note that the average size of deviations increase with respect to the forecasting horizon. The deviations are also larger for risk level $\phi = 0.99$ and Student-t disturbances (except for the predictive VaR at risk level $\phi = 0.99$).

Deviations from the SEL benchmark are expected for the two Linex functions. Indeed, the asymmetric nature of the function leads to point estimates which are necessarily lower or larger than the mean point estimate. In the case

of the AEL function, departure from the SEL indicates asymmetric shapes for the VaR densities over the out-of-sample window. More precisely, the AEL point estimates are less conservative than the SEL on average, indicating left-skewed densities for the VaR. This asymmetry can also be captured by comparing the average deviations for the Linex loss functions. While a symmetric density for the VaR would imply similar values (of opposite sign) for the two Linex functions, this is clearly not the case here, especially for the Student-t density at time horizons $s = 5$ and $s = 10$. Finally, we can notice that the predictive VaR point estimates are close to the SEL point estimates for almost all risk levels and time horizons; for these cases, choosing a quantile of the predictive distribution or choosing the posterior mean of the VaR leads to the same VaR point estimates, on average.

In summary, the VaR is left-skewed and the asymmetry as well as the uncertainty increase with respect to the time horizon and the risk level. The average deviations are also larger when the GARCH$(1, 1)$ model disturbances are Student-t distributed. At first sight, the deviations seem negligible (we recall that the maximum deviation is half a percent). As will be shown later in this chapter, the common testing methodology for assessing the performance of the VaR is unable to discriminate between the point estimates but the deviations are large enough to imply substantial differences in terms of regulatory capital. This therefore gives an additional flexibility to the user when allocating risk capital.

Table 6.1. Average deviations of the VaR point estimates from the SEL benchmark.★

GARCH(1, 1) with Normal disturbances						
	$\phi = 0.95$			$\phi = 0.99$		
Loss $\mathscr{L}$	$s = 1$	$s = 5$	$s = 10$	$s = 1$	$s = 5$	$s = 10$
Linex ($a = 3$)	−0.233	−1.385	−3.605	−0.467	−4.041	−12.790
Linex ($a = -3$)	0.230	1.332	3.226	0.459	3.602	9.470
AEL[a]	0.066	0.256	0.604	0.093	0.637	1.628
Predictive[b]	0.011	−0.148	−0.145	−0.344	−1.603	−2.309
GARCH(1, 1) with Student-t disturbances						
	$\phi = 0.95$			$\phi = 0.99$		
Loss $\mathscr{L}$	$s = 1$	$s = 5$	$s = 10$	$s = 1$	$s = 5$	$s = 10$
Linex ($a = 3$)	−0.318	−2.496	−11.598	−1.087	−9.911	−49.603
Linex ($a = -3$)	0.312	2.185	6.695	1.043	7.739	22.548
AEL[a]	0.091	0.570	1.863	0.241	1.204	3.498
Predictive[b]	0.013	−0.104	1.353	−0.027	0.645	0.847

★ The tables entries are given in hundredth percent, *i.e.*, multiplied by 100. $\mathscr{L}$: loss function; s: time horizon (in days); ϕ: risk level.
[a] Absolute error loss.
[b] In the case of the predictive VaR, the point estimate is the ϕ^cth percentile of the predictive density for the s-day ahead cumulative return; $\phi^c \doteq 1 - \phi$.

6.4.5 Regulatory capital

In this section, we assess the financial consequences resulting from the use of a particular loss function when determining a VaR point estimate. To that aim, we will base our analysis on the notion of the regulatory capital as defined by the Basel II approach for market risk [see Basel Committee on Banking Supervision 1996b]. This capital is a cushion for market risk exposures and its value is based on the ten-day ahead VaR at risk level $\phi = 0.99$. Formally, the regulatory capital allocated at time t by an agent facing a loss function $\mathscr{L}$ can be expressed as follows:

$$\widehat{\mathrm{RC}}_{\mathscr{L},t} \doteq \min\left\{\widehat{\mathrm{VaR}}^{0.99}_{\mathscr{L},t,10}, \frac{\zeta}{60}\sum_{i=0}^{59}\widehat{\mathrm{VaR}}^{0.99}_{\mathscr{L},t-i,10}\right\} \tag{6.14}$$

where $\widehat{\mathrm{VaR}}^{0.99}_{\mathscr{L},t,10}$ denotes the ten-day ahead VaR point estimate at time t, for risk level $\phi = 0.99$ and loss function $\mathscr{L}$. The value $\zeta \in [3, 4]$ is a stress factor determined by the quality of the model; it is fixed by the regulators and is based on the forecasting performance of the model. We will set ζ to 3.5 for simplicity in what follows. From formula (6.14), we note that the regulatory capital is

smoothed over time in order to avoid frequent adjustments of the balance sheet (which is costly for the bank) but can also react quickly enough to market news such as crashes.

As this was done in **Sect. 6.4.4** for the VaR point estimates, we calculate the time series of differences between the regulatory capital obtained under loss $\mathscr{L}$ and the SEL benchmark and then compute the average of the deviations over the out-of-sample window. Results are reported in **Table 6.2** where the table entries are given in percent. First, we can notice that the deviations are much larger than for the VaR point estimates; the average deviations range from 0.023% in the case of the predictive VaR to 1.741% in the case of the Linex ($a = 3$), both for the GARCH$(1, 1)$ model with Student-t disturbances. In general, deviations from the SEL are larger when the disturbances are Student-t distributed. The percentage of capital obtained with the AEL function is, on average, lower than with the SEL, indicating a left-skewed density for the regulatory capital. The largest deviation is obtained for the Linex function with parameter $a = 3$; in this case, a risk manager or regulator will keep aside a capital which is 1.741% larger than the SEL agent, for which under- or overestimation are equally serious. In contrast to this, a fund manager will be able to invest 0.79% more capital on financial markets. For this special case, there is a differential of about 2.5% in risk capital allocation between a risk manager and a fund manager; this is substantial if we imagine the amounts invested on financial markets by financial institutions.

Table 6.2. Average deviations of the regulatory capital point estimates from the SEL benchmark.★

	GARCH$(1, 1)$ disturbances	
Loss $\mathscr{L}$	Normal	Student-t
Linex ($a = 3$)	−0.446	−1.741
Linex ($a = -3$)	0.331	0.790
AEL[a]	0.056	0.122
Predictive[b]	−0.084	0.023

★ The tables entries are given in percent. $\mathscr{L}$: loss function.
[a] Absolute error loss.
[b] In the case of the predictive VaR, the point estimate is the ϕ^cth percentile of the predictive density for the s-day ahead cumulative return; $\phi^c \doteq 1 - \phi$.

6.4.6 Forecasting performance analysis

To test the ability of our models to capture the true VaR, we compare the realization of the cumulated returns $\{y_{t,s}\}_t$ with our VaR estimates for time horizon $s \in \{1, 5, 10\}$ and risk level $\phi \in \{0.95, 0.99\}$. To that aim, we adopt the (back-testing) methodology proposed by Christoffersen [1998] which has become the standard practice in financial risk management. When the forecasting horizon is one day, this approach is based on the study of the random sequence $\{V_t^\phi\}$ where:

$$V_t^\phi \doteq \begin{cases} 1 & \text{if} \quad y_{t+1} < \text{VaR}_t^\phi \\ 0 & \text{else .} \end{cases}$$

A sequence of VaR forecasts at risk level ϕ has correct conditional coverage if $\{V_t^\phi\}$ is an independent and identically distributed sequence of Bernoulli random variables with parameter $\phi^c \doteq (1-\phi)$. In practice, this hypothesis can be verified by testing jointly the independence on the series and the unconditional coverage of the VaR forecasts, *i.e.*, $\mathbb{E}(V_t^\phi) = \phi^c$.

In order to test the performance of the s-day ahead VaR ($s > 1$), we use a similar methodology, based now on the study of the random sequence $\{V_{t,s}^\phi\}_t$ where:

$$V_{t,s}^\phi \doteq \begin{cases} 1 & \text{if} \quad y_{t,s} < \text{VaR}_{t,s}^\phi \\ 0 & \text{else .} \end{cases}$$

In this case however, since the cumulative returns $y_{t,s}$ and $y_{\tau,s}$ overlap for $|t - \tau| \leqslant s$, the variables $V_{t,s}^\phi$ and $V_{\tau,s}^\phi$ are not independent and the usual test by Christoffersen [1998] cannot be applied directly. However, we can exploit the structure of dependence between $y_{t,s}$ to $y_{\tau,s}$ to get rid of this difficulty. Indeed, the construction of cumulative returns leads to the creation of spurious moving average effects of order $(s-1)$ in the time series $\{y_{t,s}\}_t$. We can therefore follow Diebold and Mariano [1995] and correct the test by Christoffersen [1998] for serial correlation via Bonferroni bounds. To that aim, we partition the series $\{y_{t,s}\}_t$ into groups for which we expect independence and correct unconditional coverage. Under the assumption that the series $\{V_{t,s}^\phi\}_t$ is $(s-1)$ dependent, each of the following s sub-series:

$$\{V^{\phi}_{1,s}, V^{\phi}_{1+s,s}, V^{\phi}_{1+2s,s}, \ldots\}$$
$$\{V^{\phi}_{2,s}, V^{\phi}_{2+s,s}, V^{\phi}_{2+2s,s}, \ldots\}$$
$$\vdots$$
$$\{V^{\phi}_{s,s}, V^{\phi}_{2s,s}, V^{\phi}_{3s,s}, \ldots\}$$

will be *iid* Bernoulli distributed if the model for the underlying process is correct. Thus, a formal test with size bounded by α can be obtained by performing s tests, each of size α/s, on each of the s sub-series, and rejecting the null hypothesis if the null is rejected for any of the s sub-series.

Forecasting results are reported in **Table 6.3** where we give the p-values of the unconditional coverage (UC), independence (IND) as well as conditional coverage (CC) tests; for time horizons $s = 5$ and $s = 10$, we report the lowest p-value computed from the s series of VaR forecasts. Our results indicate that

Table 6.3. Forecasting results of the VaR point estimates.★

GARCH(1, 1) model with Normal disturbances						
	$\phi = 0.95$			$\phi = 0.99$		
s	UC	IND	CC	UC	IND	CC
1	0.026	0.761	0.081	0.018	0.052	0.009
5	0.392	0.062	0.121	0.306	NA	NA
10	0.412	0.544	0.594	0.162	NA	NA
GARCH(1, 1) model with Student-t disturbances						
	$\phi = 0.95$			$\phi = 0.99$		
s	UC	IND	CC	UC	IND	CC
1	0.222	0.903	0.470	0.572	NA	NA
5	0.392	0.050	0.101	0.344	NA	NA
10	0.983	0.280	0.558	0.162	NA	NA

★ Forecasting test by Christoffersen [1998] based on the SEL point estimates. ϕ: risk level; s time horizon (in days); UC: p-value for the correct uncoverage test; IND: p-value for the independence test; CC: p-value for the correct conditional coverage test; NA: not applicable.

the GARCH(1, 1) model with Normal disturbances fails, at the 5% significance level, in forecasting the one-day ahead VaR for both risk levels. Indeed, the unconditional coverage gives a p-value of 0.026 for $\phi = 0.95$ and 0.018 for $\phi = 0.99$. The joint test of correct unconditional coverage and independence is however only rejected for risk level $\phi = 0.99$. In contrast to this, the GARCH(1, 1) model with Student-t disturbances performs well. For longer time horizons, the models behave similarly well and neither the unconditional coverage nor the indepen-

dence tests are rejected at the 5% significance level. We point out, however, that the test by Christoffersen [1998] is powerful when the number of observations is large. In our context of 1'200 observations, the test of the ten-day ahead VaR is based on ten sequences of (only) 120 observations. At risk level $\phi = 0.99$, a single violation is thus expected. The forecasting results should therefore be taken with caution in this case.

We emphasize the fact that the test has been applied to the time series of SEL point estimates. For comparison purposes, we have also analyzed the forecasting performance of the alternative VaR point estimates, obtained with the Linex and AEL functions. In all cases, the testing methodology gave similar p-values for the different risk levels and time horizons. This is not surprising. Indeed, the differences between the VaR point estimates are small (we recall that the largest deviation is -0.496%) and the test by Christoffersen [1998] focuses on the number of times the VaR is exceeded instead of testing the size of discrepancy between predictions and realizations. In addition, the case where the differences between point estimates are important was observed for a forecasting horizon of ten days at risk level $\phi = 0.99$, precisely the case where the power of the test is weak. Therefore, alternative (more powerful) tests should be developed, as recently pursued by Zumbach [2006]. In **Sect. 7.6** of the next chapter, we will document that the differences between the one-day ahead VaR point estimates are large when the P&L dynamics is described by a Markov-switching GJR model. In this context, the loss function of the forecaster leads to different conclusions on the forecasting performance of the model, even when relying on the common testing methodology of Christoffersen [1998].

6.5 The Expected Shortfall risk measure

While being now a standard tool in financial risk management, the VaR has been criticized in the research literature for several reasons, in particular:

- the VaR does not tell anything about the potential size of loss that exceeds its level and, as a result, it is flawed;

- the VaR is not a *coherent measure* of risk in the sense of Artzner, Delbaen, Eber, and Heath [1999]. In particular, it lacks the property of *sub-additivity*.

To circumvent these problems, the concept of Expected Shortfall (henceforth ES) also known as Conditional VaR or CVaR has been introduced by Artzner *et al.* [1999].

Definition 6.2 (Expected Shortfall). *Let Y be a univariate random variable with distribution F_Y, assumed continuous for simplicity. Then the Expected Shortfall at risk level ϕ is defined as the expected value of Y below the* VaR^{ϕ} *level. Formally:*

$$\mathrm{ES}^{\phi} \doteq \mathbb{E}(Y \mid Y \leqslant \mathrm{VaR}^{\phi}) = \frac{\int_{-\infty}^{\mathrm{VaR}^{\phi}} y \, \mathrm{d}F_Y(y)}{\phi^c} \tag{6.15}$$

where we recall that $\phi^c \doteq (1-\phi)$ for convenience and VaR^{ϕ} *is given in* ***Def. 6.1*** *(see p.76).*

Basically, the ES risk measure is the expectation of the P&L below the VaR level.

In the case of the GARCH$(1,1)$ model with Normal and Student-t distributions, the integral on the right-hand side of expression (6.15) can be calculated explicitly given the set of parameters ψ. Indeed, when the model disturbances are Normally distributed, the one-day ahead ES at risk level ϕ, estimated at time t, is given by:

$$\mathrm{ES}_t^{\phi}(\psi) = h_{t+1}^{1/2}(\boldsymbol{\alpha}, \beta) \times \frac{-\exp\left[-\frac{z_{\phi^c}^2}{2}\right]}{\sqrt{2\pi}\phi^c} \tag{6.16}$$

where we recall that $\psi \doteq (\boldsymbol{\alpha}, \beta)$, h_{t+1} is the conditional variance which is computed by recursion given the information set $\mathcal{F}_t$ and z_{ϕ^c} is the ϕ^c percentile of the standard Normal distribution. In the case of a Student-t distribution with ν degrees of freedom, the one-day ahead ES at risk level ϕ, estimated at time t, can be expressed as follows:

$$\mathrm{ES}_t^{\phi}(\psi) = \left[\varrho(\nu) \times h_{t+1}(\boldsymbol{\alpha}, \beta)\right]^{1/2} \times \Psi_{\phi^c}(\nu) \tag{6.17}$$

with:

$$\Psi_{\phi^c}(\nu) \doteq \frac{\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}} \left(\frac{\nu}{\nu-1}\right) \left(1 + \frac{t_{\phi^c}^2(\nu)}{\nu}\right)^{\frac{1-\nu}{2}}}{\phi^c}$$

where in this case $\psi \doteq (\boldsymbol{\alpha}, \beta, \nu)$, $\varrho(\nu) \doteq \frac{\nu-2}{\nu}$ and t_{ϕ^c} is the ϕ^cth percentile of the Student-t distribution. Once a joint posterior sample of the model parameters is obtained, expressions (6.16) and (6.17) can be used to simulate the density of the one-day ahead ES risk measure at any risk level ϕ.

In order to find the expression for the s-day ahead ES^{ϕ} ($s > 1$), we note that the Expected Shortfall can also be viewed as the average of VaR below the risk level ϕ^c.

Proposition 6.3. *Assuming that $\mathbb{E}(|Y|) < \infty$ and F_Y is continuous, we can express the Expected Shortfall as follows:*

$$\mathbb{E}(Y \mid Y \leqslant \mathrm{VaR}^{\phi}) = \frac{\int_0^{\phi^c} \mathrm{VaR}^u \,\mathrm{d}u}{\phi^c} \,.$$

Proof. The integral in (6.15) is transformed by the change of variable $y \mapsto u \doteq F_Y(y)$, so that:

$$\int_{-\infty}^{\mathrm{VaR}^{\phi}} y \,\mathrm{d}F_Y(y) = \int_0^{\phi^c} \mathrm{VaR}^u \,\mathrm{d}u$$

since $F_Y(-\infty) = 0$, $F_Y(\mathrm{VaR}^{\phi}) = \phi^c$, $\mathrm{d}u = \mathrm{d}F_Y(y)$ and $y = \mathrm{VaR}^u$ by **Def. 6.1** (see p.76). □

Using **Prop. 6.3**, we can estimate the s-day ahead ES at any risk level ϕ by integrating the s-day ahead VaR over the $(0, \phi^c]$ interval. Formally:

$$\mathrm{ES}_{t,s}^{\phi}(\psi) = \frac{\int_0^{\phi^c} \mathrm{VaR}_{t,s}^u(\psi)\mathrm{d}u}{\phi^c} \tag{6.18}$$

where $\mathrm{VaR}_{t,s}^u(\psi)$ is calculated using approximation (6.6) on page 81. The integral in expression (6.18) can be estimated by conventional quadrature methods. As in the one-day ahead VaR, the joint posterior sample can be used to simulate the density for the s-day ahead ES using formula (6.18).

In **Fig. 6.6**, we display the (conditional) term structure of the ES density at risk level $\phi = 0.95$ for the first observation window excerpt from our data set. As in the VaR illustration of **Sect. 6.4.3**, the lines give the median point estimates and the shaded regions depict the 95% confidence intervals of the ES density. From this graph, we note that the GARCH$(1, 1)$ model with Student-t disturbances leads to lower median point estimates for every time horizon. In addition, the ES uncertainty increases for both models with respect to the time horizon. The Student-t model leads to higher uncertainty in ES at each horizon compared to the Normal specification. Finally, the left asymmetry of the density is visually apparent for horizons larger than five days for both models.

A comparison with the VaR term structure displayed in **Fig. 6.5** (see p.96) indicates that the ES density has heavier tails and a skewness which is more

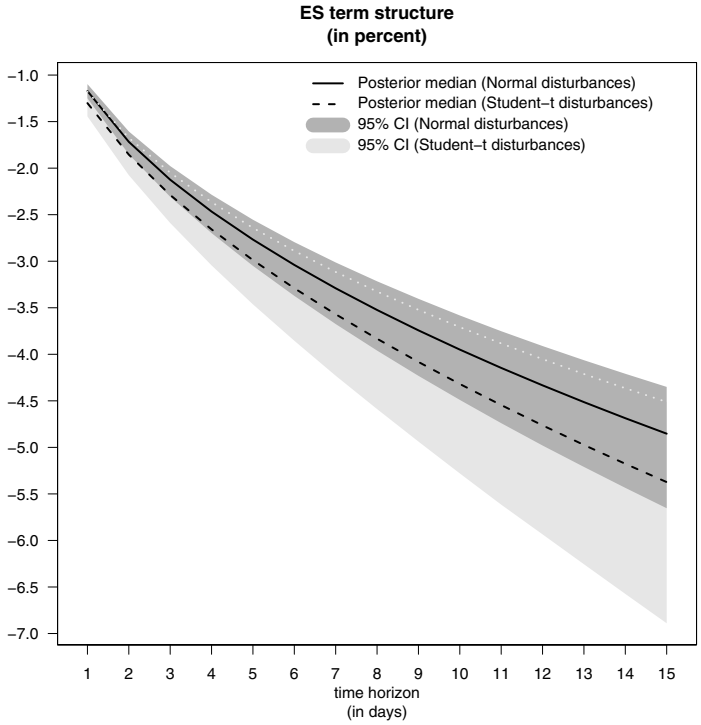

Fig. 6.6. Term structures of the ES density at risk level $\phi = 0.95$ for the GARCH$(1, 1)$ model with Normal and Student-t disturbances. The density is based on 10'000 draws from the joint posterior sample of the models' parameters.

pronounced. Therefore, given preferences in risk perspectives lead to larger differences in ES point estimates.

7

Bayesian Estimation of the Markov-Switching GJR(1, 1) Model with Student-t Innovations

> *(...) "the application of GARCH to long time series of stock-return data will yield a high measure of persistence because of the presence of deterministic shifts in the unconditional variance and the subsequent failure of the econometrician to model these shifts."*
>
> — *Christopher G. Lamoureux and William D. Lastrapes*

In this chapter, we address the problem of estimating GARCH models subject to structural changes in the parameters; namely, the Markov-switching GARCH models (henceforth MS-GARCH). In this framework, a hidden Markov sequence $\{s_t\}$ with state space $\{1, \ldots, K\}$ allows discrete changes in the model parameters. Such processes have received a lot of attention in recent years as they provide an explanation of the high persistence in volatility (*i.e.*, nearly unit root process for the conditional variance) observed with single-regime GARCH models [see, *e.g.*, Lamoureux and Lastrapes 1990]. Furthermore, the MS-GARCH models allow for a quick change in the volatility level which leads to significant improvements in volatility forecasts, as shown by Dueker [1997], Klaassen [2002], Marcucci [2005].

Following the seminal work of Hamilton and Susmel [1994], different parametrizations have been proposed to account for discrete changes in the GARCH parameters [see, *e.g.*, Dueker 1997, Gray 1996, Klaassen 2002]. However, these parametrizations for the conditional variance process lead to computational difficulties. Indeed, the evaluation of the likelihood function for a sample of length T requires the integration over all K^T possible paths, rendering the estimation infeasible. As a remedy, approximation schemes have been proposed to shorten the dependence on the state variable's history. While this difficulty is not present in ARCH type models, lower order GARCH specification of the conditional variance offers a more parsimonious representation than higher order ARCH models.

In order to avoid any difficulties related to the past infinite history of the state variable, we adopt a recent parametrization due to Haas *et al.* [2004]. In their model, the authors hypothesize K separate GARCH$(1, 1)$ processes for the conditional variance of the MS-GARCH process $\{y_t\}$. The conditional variances at time t can be written in vector form as follows:

$$\begin{pmatrix} h_t^1 \\ h_t^2 \\ \vdots \\ h_t^K \end{pmatrix} \doteq \begin{pmatrix} \alpha_0^1 \\ \alpha_0^2 \\ \vdots \\ \alpha_0^K \end{pmatrix} + \begin{pmatrix} \alpha_1^1 \\ \alpha_1^2 \\ \vdots \\ \alpha_1^K \end{pmatrix} y_{t-1}^2 + \begin{pmatrix} \beta^1 \\ \beta^2 \\ \vdots \\ \beta^K \end{pmatrix} \odot \begin{pmatrix} h_{t-1}^1 \\ h_{t-1}^2 \\ \vdots \\ h_{t-1}^K \end{pmatrix} \tag{7.1}$$

where $\odot$ denotes the Hadamard product, *i.e.*, element-by-element multiplication. The MS-GARCH process $\{y_t\}$ is then simply obtained by setting:

$$y_t = \varepsilon_t (h_t^{s_t})^{1/2}$$

where ε_t is an error term with zero mean and unit variance. The parameters α_0^k, α_1^k and β^k are therefore the GARCH$(1, 1)$ parameters related to the kth state of the nature. Under this specification, the conditional variance is solely a function of the past data and current state s_t, which avoids the problem of infinite history. In the context of the Bayesian estimation, this allows to simulate the state process in a multi-move manner which enhances the sampler's efficiency.

In addition to its appealing computational aspects, the MS-GARCH model of Haas *et al.* [2004] has conceptual advantages. In effect, one reason for specifying Markov-switching models that allow for different GARCH behavior in each regime is to capture the difference in the variance dynamics in low- and high-volatility periods. As pointed out by Haas *et al.* [2004, p.498]:

> (...) "a relatively large value of α_1^k and relatively low values of β^k in high-volatility regimes may indicate a tendency to over-react to news, compared to *regular* periods, while there is less memory in these subprocesses. Such an interpretation requires a parametrization of Markov-switching GARCH models that implies a clear association between the GARCH parameters within regime k, that is α_0^k, α_1^k and β^k and the corresponding $\{h_t^k\}$ process."

The specification of the conditional variance in equation (7.1) allows for a clear-cut interpretation of the variance dynamics in each regime. Moreover, Haas *et al.* [2004] show that results on the single-regime GARCH$(1, 1)$ model can be extended to their specification; in particular, they derive explicit formulae for the covariance stationarity condition, the unconditional variance as well as the dependence structure of the squared process $\{y_t^2\}$.

To account for additional stylized facts observed in financial time series, especially for stock indices (see **Chap. 4**), we will consider an asymmetric extension of (7.1) in which the GARCH$(1,1)$ processes are replaced by GJR$(1,1)$ processes. More precisely, in this Markov-switching GJR model (henceforth MS-GJR), the conditional variances at time t can be written in vector form as follows:

$$\begin{pmatrix} h_t^1 \\ h_t^2 \\ \vdots \\ h_t^K \end{pmatrix} \doteq \begin{pmatrix} \alpha_0^1 \\ \alpha_0^2 \\ \vdots \\ \alpha_0^K \end{pmatrix} + \left[\begin{pmatrix} \alpha_1^1 \\ \alpha_1^2 \\ \vdots \\ \alpha_1^K \end{pmatrix} \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \begin{pmatrix} \alpha_2^1 \\ \alpha_2^2 \\ \vdots \\ \alpha_2^K \end{pmatrix} \mathbb{I}_{\{y_{t-1} < 0\}} \right] y_{t-1}^2 + \begin{pmatrix} \beta^1 \\ \beta^2 \\ \vdots \\ \beta^K \end{pmatrix} \odot \begin{pmatrix} h_{t-1}^1 \\ h_{t-1}^2 \\ \vdots \\ h_{t-1}^K \end{pmatrix} \tag{7.2}$$

where $\mathbb{I}_{\{\bullet\}}$ denotes the indicator function. In this setting, the conditional variance in every regime can react asymmetrically depending on the sign of the past shocks due to the introduction of dummy variables. The leverage effect is present for a given state k as soon as $\alpha_2^k > \alpha_1^k$. An interesting feature of the parametrization (7.2) lies in the fact that we can estimate whether the response to past negative shock on the conditional variance is different across regimes.

The plan of this chapter is as follows. We set up the model in **Sect. 7.1**. The MCMC scheme is detailed in **Sect. 7.2**. The MS-GJR model as well as a single-regime GJR model are applied to the Swiss Market Index log-returns in **Sect. 7.3**. In **Sect. 7.4**, we test the models for misspecification by using the generalized residuals and assess the goodness-of-fit through the calculation of the Deviance information criterion and the model likelihoods. In **Sect. 7.5**, we test the predictive performance of the models by running a forecasting analysis based on the VaR. In **Sect. 7.6**, we propose a methodology to depict the one-day ahead VaR density and document how specific forecasters' risk perspectives can lead to different conclusions in terms of the forecasting performance of the model. We conclude with some comments regarding the ML estimation of the MS-GJR model in **Sect. 7.7**.

7.1 The model and the priors

A Markov-switching GJR$(1,1)$ model with Student-t innovations may be written as follows:

$$\begin{aligned} y_t &= \varepsilon_t(\varrho h_t)^{1/2} \quad \text{for } t = 1, \ldots, T \\ \varepsilon_t &\overset{iid}{\sim} \mathcal{S}(0, 1, \nu) \\ \varrho &\doteq \frac{\nu - 2}{\nu} \\ h_t &\doteq \mathbf{e}_t' \mathbf{h}_t \end{aligned} \tag{7.3}$$

where $\mathbf{e}_t$ is a $K \times 1$ vector defined by $\mathbf{e}_t \doteq \left(\mathbb{I}_{\{s_t=1\}} \cdots \mathbb{I}_{\{s_t=K\}}\right)'$, $\mathbb{I}_{\{\bullet\}}$ is the indicator function; the sequence $\{s_t\}$ is assumed to be a stationary, irreducible Markov process with discrete state space $\{1, \ldots, K\}$ and transition matrix $P \doteq [P_{ij}]$ where $P_{ij} \doteq \mathbb{P}(s_{t+1} = j \mid s_t = i)$; $\mathcal{S}(0, 1, \nu)$ denotes the standard Student-t density with ν degrees of freedom and ϱ is a scaling factor which ensures that the conditional variance is given by h_t. Moreover, we define the $K \times 1$ vector of GJR(1, 1) conditional variances in a compact form as follows:

$$\mathbf{h}_t \doteq \boldsymbol{\alpha}_0 + (\boldsymbol{\alpha}_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \boldsymbol{\alpha}_2 \mathbb{I}_{\{y_{t-1} < 0\}}) y_{t-1}^2 + \boldsymbol{\beta} \odot \mathbf{h}_{t-1}$$

where $\mathbf{h}_t \doteq (h_t^1 \cdots h_t^K)'$, $\boldsymbol{\alpha}_j \doteq (\alpha_j^1 \cdots \alpha_j^K)'$ for $j = 0, 1, 2$ and $\boldsymbol{\beta} \doteq (\beta^1 \cdots \beta^K)'$. In addition, we require that $\boldsymbol{\alpha}_0 > \mathbf{0}$, $\boldsymbol{\alpha}_1 \geqslant \mathbf{0}$, $\boldsymbol{\alpha}_2 \geqslant \mathbf{0}$ and $\boldsymbol{\beta} \geqslant \mathbf{0}$, where $\mathbf{0}$ is a $K \times 1$ vector of zeros, in order to ensure the positivity of the conditional variance in every regime and set $\mathbf{h}_0 \doteq \mathbf{0}$ and $y_0 \doteq 0$ for convenience.

The use of a Student-t instead of a Normal distribution is quite popular in standard single-regime GARCH literature. For regime-switching models, a Student-t distribution might be seen as superfluous since the switching regime can account for large unconditional kurtosis in the data [see, *e.g.*, Haas *et al.* 2004]. However, as empirically observed by Klaassen [2002], allowing for Student-t innovations within regimes can enhance the stability of the states and allows to focus on the conditional variance's behavior instead of capturing some outliers. Moreover, the Student-t distribution includes the Normal distribution as the limiting case where the degrees of freedom parameter goes to infinity. We have therefore an additional flexibility in the modeling and can impose Normality by constraining the lower boundary for the degrees of freedom parameter through the prior distribution.

As pointed out in **Sect. 5.1**, the Student-t specification (7.3) needs to be re-expressed to perform a convenient Bayesian estimation. This is achieved as follows:

$$\begin{aligned} y_t &= \varepsilon_t(\varpi_t \varrho h_t)^{1/2} \quad \text{for } t = 1, \ldots, T \\ \varepsilon_t &\overset{iid}{\sim} \mathcal{N}(0, 1) \\ \varpi_t &\overset{iid}{\sim} \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \end{aligned}$$

where $\mathcal{N}(0,1)$ is the standard Normal and $\mathcal{IG}$ denotes the Inverted Gamma density. The degrees of freedom parameter ν characterizes the density of ϖ_t as follows:

$$p(\varpi_t \mid \nu) = \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} \varpi_t^{-\frac{\nu}{2}-1} \exp\left[-\frac{\nu}{2\varpi_t}\right] . \tag{7.4}$$

For a parsimonious expression of the likelihood function, we define the $T \times 1$ vectors $\mathbf{y} \doteq (y_1 \cdots y_T)'$, $\boldsymbol{\varpi} \doteq (\varpi_1 \cdots \varpi_T)'$ as well as $\mathbf{s} \doteq (s_1 \cdots s_T)'$ and regroup the ARCH parameters into the $3K \times 1$ vector $\boldsymbol{\alpha} \doteq (\boldsymbol{\alpha}_0' \ \boldsymbol{\alpha}_1' \ \boldsymbol{\alpha}_2')'$. The model parameters are then regrouped into the augmented set of parameters $\Theta \doteq (\psi, \boldsymbol{\varpi}, \mathbf{s})$ where $\psi \doteq (\boldsymbol{\alpha}, \boldsymbol{\beta}, \nu, P)$. Finally, we define the $T \times T$ diagonal matrix:

$$\Sigma \doteq \Sigma(\Theta) = \operatorname{diag}\big(\{\varpi_t \varrho\, \mathbf{e}_t' \mathbf{h}_t\}_{t=1}^T\big)$$

where we recall that ϱ, $\mathbf{e}_t$ and $\mathbf{h}_t$ are both functions of the model parameters, respectively given by:

$$\begin{aligned} \varrho(\nu) &\doteq \frac{\nu - 2}{\nu} \\ \mathbf{e}_t(s_t) &\doteq \big(\mathbb{I}_{\{s_t=1\}} \cdots \mathbb{I}_{\{s_t=K\}}\big)' \end{aligned}$$

and:

$$\mathbf{h}_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) \doteq \boldsymbol{\alpha}_0 + (\boldsymbol{\alpha}_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \boldsymbol{\alpha}_2 \mathbb{I}_{\{y_{t-1} < 0\}}) y_{t-1}^2 + \boldsymbol{\beta} \odot \mathbf{h}_{t-1}(\boldsymbol{\alpha}, \boldsymbol{\beta}) .$$

We can now express the likelihood function of Θ as follows:

$$\mathcal{L}(\Theta \mid \mathbf{y}) \propto (\det \Sigma)^{-1/2} \exp\left[-\tfrac{1}{2} \mathbf{y}' \Sigma^{-1} \mathbf{y}\right] . \tag{7.5}$$

In the Bayesian approach, the vector of hidden states is considered as a parameter as implied by expression (7.5).

The likelihood function (7.5) is invariant with respect to relabeling the states (*i.e.*, the labeling of the states can be interchanged without affecting the likelihood value), which leads to a lack of identification of the state-specific parameters. So, without a prior inequality restriction on some state-specific parameters, a multimodal posterior is obtained and is difficult to interpret and summarize. To overcome this problem, we make use of the permutation sampler of Frühwirth-Schnatter [2001b] to find suitable identification constraints. The permutation sampler requires priors that are labeling invariant. Furthermore, we cannot be completely non-informative about the state specific parameters since, from a theoretical viewpoint, this would result in improper posteriors [see

Diebolt and Robert 1994]. These points have therefore to be taken into account when choosing the prior densities.

Conditionally on the $K \times K$ transition probabilities matrix $P \doteq [P_{ij}]$ where:

$$P_{ij} \doteq \mathbb{P}(s_{t+1} = j \mid s_t = i)$$

the prior on vector $\mathbf{s}$ is Markov:

$$p(\mathbf{s} \mid P) = \pi(s_1) \prod_{i=1}^{K} \prod_{j=1}^{K} P_{ij}^{N_{ij}}$$

where $N_{ij} \doteq \#\{s_{t+1} = j \mid s_t = i\}$ is the number of one-step transitions from state i to j in the $T \times 1$ vector $\mathbf{s}$. The mass function for the initial state, $\pi(s_1)$, is obtained by calculating the ergodic probabilities of the Markov chain. The vector of ergodic probabilities can be obtained as the sum of the columns of matrix $(A'A)^{-1}$, where the matrix A is defined as follows:

$$A \doteq \begin{pmatrix} I_K - P' \\ \boldsymbol{\iota}_K' \end{pmatrix}$$

where I_K is a $K \times K$ identity matrix and $\boldsymbol{\iota}_K$ a $K \times 1$ vector of ones [see Hamilton 1994, Sect.22.2].

The prior density for the $K \times K$ transition matrix P is obtained by assuming that the K rows are independent and that the density of the ith row is Dirichlet with parameter $\boldsymbol{\eta}_i \doteq (\eta_{i1} \cdots \eta_{iK})$:

$$\begin{aligned} p(P) &= \prod_{i=1}^{K} \mathcal{D}(\boldsymbol{\eta}_i) \\ &\propto \prod_{i=1}^{K} \prod_{j=1}^{K} P_{ij}^{\eta_{ij}-1} \ . \end{aligned}$$

Due to the labeling invariance assumption, we require that $\eta_{ii} \doteq \eta_p$ for $i = 1, \ldots, K$ and $\eta_{ij} \doteq \eta_q$ for $i, j \in \{1, \ldots, K; i \neq j\}$. A prior density with $\eta_p > \eta_q$ could be used to model the belief that the probability of persistence is bigger than the probability of transition.

For the scedastic function's parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we use truncated Normal densities:

$$\begin{aligned} p(\boldsymbol{\alpha}) &\propto \mathcal{N}_{3K}(\boldsymbol{\alpha} \mid \boldsymbol{\mu}_\alpha, \Sigma_\alpha) \mathbb{I}_{\{\boldsymbol{\alpha} > \mathbf{0}\}} \\ p(\boldsymbol{\beta}) &\propto \mathcal{N}_{K}(\boldsymbol{\beta} \mid \boldsymbol{\mu}_\beta, \Sigma_\beta) \mathbb{I}_{\{\boldsymbol{\beta} > \mathbf{0}\}} \end{aligned}$$

where we recall that $\boldsymbol{\mu}_\bullet$ and $\Sigma_\bullet$ are the hyperparameters, $\mathbf{0}$ is a vector of zeros of appropriate size and $\mathcal{N}_d$ is the d-dimensional Normal density ($d > 1$). The assumption of labeling invariance is fulfilled if we assume further that the hyperparameters are the same for all states. In particular, we set:

$$[\boldsymbol{\mu}_\alpha]_i \doteq \mu_{\alpha_0} \ , \quad [\Sigma_\alpha]_{ii} \doteq \sigma^2_{\alpha_0} \quad , \quad \left[\boldsymbol{\mu}_\beta\right]_i \doteq \mu_\beta \quad , \quad [\Sigma_\beta]_{ii} \doteq \sigma^2_\beta$$

for $i = 1, \ldots, K$, and:

$$[\boldsymbol{\mu}_\alpha]_i \doteq \mu_{\alpha_1} \quad , \quad [\Sigma_\alpha]_{ii} \doteq \sigma^2_{\alpha_1}$$

for $i = K+1, \ldots, 2K$, and:

$$[\boldsymbol{\mu}_\alpha]_i \doteq \mu_{\alpha_2} \quad , \quad [\Sigma_\alpha]_{ii} \doteq \sigma^2_{\alpha_2}$$

for $i = 2K+1, \ldots, 3K$, where μ_{α_j}, $\sigma^2_{\alpha_j}$ $(j = 0, 1, 2)$, and μ_β, σ^2_β are fixed hyperparameters. We note that matrices Σ_α and Σ_β are diagonal in this case.

The prior density of the $T \times 1$ vector $\boldsymbol{\varpi}$ conditional on ν is found by noting that ϖ_t are independent and identically distributed from (7.4), which yields:

$$p(\boldsymbol{\varpi} \mid \nu) = \left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \left(\prod_{t=1}^{T} \varpi_t\right)^{-\frac{\nu}{2}-1} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}\frac{\nu}{\varpi_t}\right] .$$

Following Deschamps [2006], we choose a translated Exponential with parameters $\lambda > 0$ and $\delta \geqslant 2$ for the degrees of freedom parameter:

$$p(\nu) = \lambda \exp[-\lambda(\nu - \delta)]\mathbb{I}_{\{\delta < \nu < \infty\}} .$$

Finally, we form the joint prior by assuming prior independence between $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $(\boldsymbol{\varpi}, \nu)$ and $(\mathbf{s}, P)$ as follows:

$$p(\Theta) = p(\boldsymbol{\alpha})p(\boldsymbol{\beta})p(\boldsymbol{\varpi} \mid \nu)p(\nu)p(\mathbf{s} \mid P)p(P)$$

and by combining the likelihood function (7.5) with the joint prior above, we obtain the posterior density via Bayes' rule:

$$p(\Theta \mid \mathbf{y}) \propto \mathcal{L}(\Theta \mid \mathbf{y})p(\Theta) . \tag{7.6}$$

7.2 Simulating the joint posterior

We draw an initial value:

$$\Theta^{[0]} \doteq (\boldsymbol{\alpha}^{[0]}, \boldsymbol{\beta}^{[0]}, \boldsymbol{\varpi}^{[0]}, \nu^{[0]}, \mathbf{s}^{[0]}, P^{[0]})$$

from the joint prior and we generate iteratively J passes for Θ. A single pass is decomposed as follows:

$$\begin{aligned}
\mathbf{s}^{[j]} &\sim p(\mathbf{s} \mid \boldsymbol{\alpha}^{[j-1]}, \boldsymbol{\beta}^{[j-1]}, \boldsymbol{\varpi}^{[j-1]}, \nu^{[j-1]}, P^{[j-1]}, \mathbf{y}) \\
P^{[j]} &\sim p(P \mid \mathbf{s}^{[j]}) \\
\boldsymbol{\alpha}^{[j]} &\sim p(\boldsymbol{\alpha} \mid \boldsymbol{\beta}^{[j-1]}, \boldsymbol{\varpi}^{[j-1]}, \nu^{[j-1]}, \mathbf{s}^{[j]}, \mathbf{y}) \\
\boldsymbol{\beta}^{[j]} &\sim p(\boldsymbol{\beta} \mid \boldsymbol{\alpha}^{[j]}, \boldsymbol{\varpi}^{[j-1]}, \nu^{[j-1]}, \mathbf{s}^{[j]}, \mathbf{y}) \\
\boldsymbol{\varpi}^{[j]} &\sim p(\boldsymbol{\varpi} \mid \boldsymbol{\alpha}^{[j]}, \boldsymbol{\beta}^{[j]}, \nu^{[j-1]}, \mathbf{s}^{[j]}, \mathbf{y}) \\
\nu^{[j]} &\sim p(\nu \mid \boldsymbol{\varpi}^{[j]}) \, .
\end{aligned} \tag{7.7}$$

In (7.7), only $\boldsymbol{\varpi}$ and P can be simulated from known expressions. Draws of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are achieved by a multivariate extension of the methodology proposed by Nakatsuma [1998, 2000]. The generation of state vector $\mathbf{s}$ is made by using the Forward Filtering Backward Sampling (henceforth FFBS) algorithm described in Chib [1996]. Finally, sampling ν is achieved by an efficient rejection technique.

As pointed out previously, the likelihood function and the joint prior are labeling invariant. Consequently, the joint posterior density in (7.6) will also be invariant and hence exhibit, at least theoretically, $K!$ different modes. Therefore, it is important to carefully select constraints to identify the model. In effect, a constraint that ignores the geometry of the posterior density will not lead to a unique labeling and can introduce a bias toward the constraint, as shown in Frühwirth-Schnatter [2001b]. If a suitable identifying restriction is not available or is not known a priori, an elegant solution to determine these constraints is to use the *random permutation sampler* proposed by Frühwirth-Schnatter [2001b]. In this version of the permutation sampler, each pass of the MCMC scheme is followed by a random permutation of the regime definitions. Formally, a random permutation $\{\Pi_1, \dots \Pi_K\}$ of $\{1, \dots, K\}$ is selected with probability $\frac{1}{K!}$. Then, for $i, j \in \{1, \dots, K\}$, the element (i, j) of P is replaced by the element with indices (Π_i, Π_j). The hidden states process $\{s_t\}$ is substituted by $\{\Pi_{s_t}\}$. Finally, for $k = 1, \dots, K$, parameter α_0^k is replaced by $\alpha_0^{\Pi_k}$, parameter α_1^k by $\alpha_1^{\Pi_k}$, parameter α_2^k by $\alpha_2^{\Pi_k}$ and parameter β^k by β^{Π_k}. Hence, relabeling only affects the scedastic function's parameters, the state process and the transition probabilities while the vector $\boldsymbol{\varpi}$ and the degrees of freedom parameter ν remain unchanged.

The random permutation sampler by Frühwirth-Schnatter [2001b] is used to improve the mixing of the MCMC sampler and to explore the full unconstrained parameter space. Then post-processing the MCMC output of the random per-

mutation sampler in an exploratory way, by plotting scatter plots for instance, can suggest an appropriate identification constraint, such as:

$$\beta^1 \; < \; \ldots \; < \; \beta^K \tag{7.8}$$

meaning, in this particular case, that the MS-GJR model can be identified through inequalities on the parameter β between regimes. At this stage, the model is estimated again under the constraint (7.8) by enforcing the corresponding permutation of the regimes. This version of the permutation sampler is referred to as the *constrained permutation sampler*. At each sweep of the sampler, we test whether the constraint is fulfilled. If not, we order the pairs $\{1, \beta^1\}, \ldots, \{K, \beta^K\}$ with respect to the second component. The first component $\{\Pi_1, \ldots, \Pi_K\}$ of the ordered pairs defines the correct permutation of reordering the state parameters and this permutation is applied to the state-specific components, as this was done for the random permutation sampler. If the model is identifiable up to permutations of the states and satisfies certain regularity conditions, the constrained posterior density will exhibit a single mode. Note that the selection of the constraint (7.8) is arbitrary, because there exist $K!$ different ways of formulating constraints which render the model identified, namely:

$$\beta^{\Pi_1} \; < \; \ldots \; < \; \beta^{\Pi_K}$$

for all permutations $\{\Pi_1, \ldots, \Pi_K\}$ of $\{1, \ldots, K\}$. At this stage, if label switching still occurs, this might indicate that the inequality restriction (7.8) is not well suited or that the number K of chosen regimes is too large [see Frühwirth-Schnatter 2006, Sect.4.2]. We will now present the derivation for the full conditionals appearing in the MCMC scheme (7.7).

7.2.1 Generating vector s

The generation of posterior samples for the $T \times 1$ vector $\mathbf{s}$ is carried out in a multi-move manner by using the FFBS algorithm. We refer the reader to Chib [1996] and Frühwirth-Schnatter [2006, Chap.11] for a detailed presentation of this procedure. We mention however that the FFBS approach can be used since the conditional density of y_t only depends on the current regime which is a consequence of the definition for the conditional variance $h_t \doteq \mathbf{e}_t' \mathbf{h}_t$. Other specifications for the conditional variance in Gray [1996] or Klaassen [2002] for instance, do not allow for such an approach, as noted in Kaufmann and Frühwirth-Schnatter [2002, Sect.6.3]. The application of the FFBS algorithm has the potential advantage that the states are updated as a single block,

which avoids superfluous correlation in the vector's components, and therefore enhances the sampler's efficiency [see Frühwirth-Schnatter 2006, Sect.11.5.6].

7.2.2 Generating matrix $\boldsymbol{P}$

The full conditional density of the transition matrix can be derived without regard to the sampling model since P becomes independent of Θ and $\mathbf{y}$ given the vector of states. Indeed, the posterior density is obtained as follows:

$$\begin{aligned} p(P \mid \mathbf{s}) &\propto p(\mathbf{s} \mid P)p(P) \\ &\propto \left(\pi(s_1) \prod_{i=1}^{K} \prod_{j=1}^{K} P_{ij}^{N_{ij}} \right) \times \left(\prod_{i=1}^{K} \prod_{j=1}^{K} P_{ij}^{\eta_{ij}-1} \right) \\ &\propto \prod_{i=1}^{K} \prod_{j=1}^{K} P_{ij}^{N_{ij}+\eta_{ij}-1} \\ &\propto \prod_{i=1}^{K} \mathcal{D}(\widehat{\boldsymbol{\eta}}_i) \end{aligned} \tag{7.9}$$

where $\widehat{\boldsymbol{\eta}}_i \doteq (N_{i1}+\eta_{i1} \cdots N_{iK}+\eta_{iK})$ and $N_{ij} \doteq \#\{s_{t+1} = j \mid s_t = i\}$ is the total number of one-step transitions from state i to state j in the vector $\mathbf{s}$. The rows of matrix P are independent a posteriori and the ith row follows a Dirichlet density with parameter $\widehat{\boldsymbol{\eta}}_i$.

7.2.3 Generating the GJR parameters

The methodology used to draw vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be viewed as a multivariate extension of the approach proposed in **Chap. 4** for the single-regime GJR model. Let us consider the following $K \times 1$ vector:

$$\mathbf{w}_t \doteq \frac{y_t^2 \boldsymbol{\iota}_K}{\tau_t} - \mathbf{h}_t$$

where we define $\tau_t \doteq \varpi_t \rho$ for convenience and recall that $\boldsymbol{\iota}_K$ is a $K \times 1$ vector of ones. In order to simplify the notations further, we define $v_t \doteq \frac{y_t^2}{\tau_t}$ which yields $\mathbf{w}_t = v_t \boldsymbol{\iota}_K - \mathbf{h}_t$. From there, we can transform the expression for the vector of conditional variances as follows:

$$
\begin{aligned}
\mathbf{h}_t &= \boldsymbol{\alpha}_0 + (\boldsymbol{\alpha}_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \boldsymbol{\alpha}_2 \mathbb{I}_{\{y_{t-1}<0\}}) y_{t-1}^2 + \boldsymbol{\beta} \odot \mathbf{h}_{t-1} \\
\Leftrightarrow (v_t \boldsymbol{\iota}_K - \mathbf{w}_t) &= \boldsymbol{\alpha}_0 + (\boldsymbol{\alpha}_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \boldsymbol{\alpha}_2 \mathbb{I}_{\{y_{t-1}<0\}}) y_{t-1}^2 \\
&\quad + \boldsymbol{\beta} \odot (v_{t-1} \boldsymbol{\iota}_K - \mathbf{w}_{t-1}) \\
\Leftrightarrow v_t \boldsymbol{\iota}_K &= \boldsymbol{\alpha}_0 + \left[\tau_{t-1}(\boldsymbol{\alpha}_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \boldsymbol{\alpha}_2 \mathbb{I}_{\{y_{t-1}<0\}}) + \boldsymbol{\beta}\right] \odot v_{t-1} \boldsymbol{\iota}_K \\
&\quad - \boldsymbol{\beta} \odot \mathbf{w}_{t-1} + \mathbf{w}_t \\
\Leftrightarrow \mathbf{w}_t &= v_t \boldsymbol{\iota}_K - \boldsymbol{\alpha}_0 - \left[\tau_{t-1}(\boldsymbol{\alpha}_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \boldsymbol{\alpha}_2 \mathbb{I}_{\{y_{t-1}<0\}}) + \boldsymbol{\beta}\right] \odot v_{t-1} \boldsymbol{\iota}_K \\
&\quad + \boldsymbol{\beta} \odot \mathbf{w}_{t-1} \ .
\end{aligned}
$$

Moreover, let us define $w_t \doteq \mathbf{e}_t' \mathbf{w}_t$ and note that w_t can be written as follows:

$$
\begin{aligned}
w_t &\doteq \mathbf{e}_t' \mathbf{w}_t \\
&= v_t - h_t = \left(\frac{y_t^2}{\varpi_t \varrho h_t} - 1 \right) h_t \\
&= \left(\chi_1^2 - 1\right) h_t
\end{aligned}
$$

where χ_1^2 denotes a Chi-squared variable with one degree of freedom. This comes from the fact that the conditional distribution of y_t is Normal with zero mean and variance $\varpi_t \varrho h_t$. Therefore, the conditional mean of w_t is zero and the conditional variance is $2h_t^2$. As in the single-regime GJR model, this variable can be approximated by z_t, a Normal variable with a mean of zero and a variance of $2h_t^2$. The variable z_t can be further expressed as $z_t \doteq \mathbf{e}_t' \mathbf{z}_t$ where $\mathbf{z}_t$ is a function of vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ given by:

$$
\begin{aligned}
\mathbf{z}_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= v_t \boldsymbol{\iota}_K - \boldsymbol{\alpha}_0 - \left[\tau_{t-1}(\boldsymbol{\alpha}_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \boldsymbol{\alpha}_2 \mathbb{I}_{\{y_{t-1}<0\}}) + \boldsymbol{\beta}\right] \odot v_{t-1} \boldsymbol{\iota}_K \\
&\quad + \boldsymbol{\beta} \odot \mathbf{z}_{t-1}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \ .
\end{aligned}
\tag{7.10}
$$

Then, we construct the $T \times 1$ vector $\mathbf{z} \doteq (z_1 \cdots z_T)'$ where $z_t \doteq \mathbf{e}_t' \mathbf{z}_t$ as well as the $T \times T$ diagonal matrix:

$$
\Lambda \doteq \Lambda(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \operatorname{diag}\left(\{2\mathbf{e}_t' \mathbf{h}_t^2(\boldsymbol{\alpha}, \boldsymbol{\beta})\}_{t=1}^T\right)
$$

and express the approximate likelihood function of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ as follows:

$$
\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \varpi, \nu, \mathbf{s}, \mathbf{y}) \propto (\det \Lambda)^{-1/2} \exp\left[-\tfrac{1}{2} \mathbf{z}' \Lambda^{-1} \mathbf{z}\right] \ . \tag{7.11}
$$

As will be shown hereafter, the construction of the proposal densities for parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is based on this likelihood function.

Generating vector $\boldsymbol{\alpha}$

First, we note that the function $\mathbf{z}_t(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in (7.10) can be expressed as a linear function of vector $\boldsymbol{\alpha}$. To show this, we simply extend the argument of the single-regime GJR model by using appropriate recursive transformations. More precisely, the ith component of the $K \times 1$ vector $\mathbf{z}_t$ can be written as follows:

$$[\mathbf{z}_t]_i = v_t - \begin{pmatrix} l_t^*(\beta^i) & v_t^*(\beta^i) & v_t^{**}(\beta^i) \end{pmatrix} \begin{pmatrix} \alpha_0^i \\ \alpha_1^i \\ \alpha_2^i \end{pmatrix}$$

with the recursive transformations l_t^*, v_t^* and v_t^{**} given by:

$$\begin{aligned} l_t^*(\beta^i) &\doteq 1 + \beta^i l_{t-1}^*(\beta^i) \\ v_t^*(\beta^i) &\doteq y_{t-1}^2 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \beta^i v_{t-1}^*(\beta^i) \\ v_t^{**}(\beta^i) &\doteq y_{t-1}^2 \mathbb{I}_{\{y_{t-1} < 0\}} + \beta^i v_{t-1}^{**}(\beta^i) \end{aligned} \tag{7.12}$$

where $l_0^* = v_0^* = v_0^{**} \doteq 0$. We notice that $l_t^*(\bullet)$, $v_t^*(\bullet)$ and $v_t^{**}(\bullet)$ in (7.12) are similar to the recursive transformations used for the single-regime GJR model. Let us now regroup the recursive values into a $K \times 3K$ matrix C_t as follows:

$$C_t \doteq \begin{pmatrix} l_t^*(\beta^1) & 0 & \cdots & 0 & v_t^*(\beta^1) & 0 & \cdots & 0 & v_t^{**}(\beta^1) & 0 & \cdots & 0 \\ 0 & l_t^*(\beta^2) & 0 & \vdots & 0 & v_t^*(\beta^2) & 0 & \vdots & 0 & v_t^{**}(\beta^2) & 0 & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & l_t^*(\beta^K) & 0 & \cdots & 0 & v_t^*(\beta^K) & 0 & \cdots & 0 & v_t^{**}(\beta^K) \end{pmatrix}.$$

It is straightforward to show that $\mathbf{z}_t = v_t \boldsymbol{\iota}_K - C_t \boldsymbol{\alpha}$, and since $z_t \doteq \mathbf{e}_t' \mathbf{z}_t$ we get $z_t = v_t - \mathbf{e}_t' C_t \boldsymbol{\alpha}$. Then, by defining the $T \times 1$ vectors $\mathbf{z} \doteq (z_1 \cdots z_T)'$ and $\mathbf{v} \doteq (v_1 \cdots v_T)'$ as well as the $T \times 3K$ matrix C whose tth row is $\mathbf{e}_t' C_t$, we end up with $\mathbf{z} = \mathbf{v} - C\boldsymbol{\alpha}$ which is the desired linear expression for $\mathbf{z}$. The proposal density to sample vector $\boldsymbol{\alpha}$ is obtained by combining the approximate likelihood (7.11) and the prior density by Bayes' update:

$$q_\alpha(\boldsymbol{\alpha} \mid \widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\varpi}, \nu, \mathbf{s}, \mathbf{y}) \propto \mathcal{N}_{3K}(\boldsymbol{\alpha} \mid \widehat{\boldsymbol{\mu}}_\alpha, \widehat{\Sigma}_\alpha) \mathbb{I}_{\{\boldsymbol{\alpha} > \mathbf{0}\}} \tag{7.13}$$

with:

$$\begin{aligned} \widehat{\Sigma}_\alpha^{-1} &\doteq C' \widetilde{\Lambda}^{-1} C + \Sigma_\alpha^{-1} \\ \widehat{\boldsymbol{\mu}}_\alpha &\doteq \widehat{\Sigma}_\alpha (C' \widetilde{\Lambda}^{-1} \mathbf{v} + \Sigma_\alpha^{-1} \boldsymbol{\mu}_\alpha) \end{aligned}$$

where the $T \times T$ diagonal matrix $\widetilde{\Lambda} \doteq \text{diag}\big(\{2\mathbf{e}_t'\mathbf{h}_t^2(\widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta})\}_{t=1}^T\big)$ and $\widetilde{\boldsymbol{\alpha}}$ is the previous draw of $\boldsymbol{\alpha}$ in the M-H sampler. A candidate $\boldsymbol{\alpha}^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\alpha}^\star, \boldsymbol{\beta}, \boldsymbol{\varpi}, \nu, \mathbf{s}, P \mid \mathbf{y})}{p(\widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\varpi}, \nu, \mathbf{s}, P \mid \mathbf{y})} \frac{q_\alpha(\widetilde{\boldsymbol{\alpha}} \mid \boldsymbol{\alpha}^\star, \boldsymbol{\beta}, \boldsymbol{\varpi}, \nu, \mathbf{s}, \mathbf{y})}{q_\alpha(\boldsymbol{\alpha}^\star \mid \widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\varpi}, \nu, \mathbf{s}, \mathbf{y})}, 1\right\} .$$

Generating vector $\boldsymbol{\beta}$

The function $\mathbf{z}_t(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in (7.10) could be expressed, in the previous section, as a linear function of $\boldsymbol{\alpha}$ but cannot be expressed as a linear function of vector $\boldsymbol{\beta}$. To overcome this problem, we linearize the $K \times 1$ vector $\mathbf{z}_t(\boldsymbol{\beta})$ by a first order Taylor expansion at point $\widetilde{\boldsymbol{\beta}}$:

$$\mathbf{z}_t(\boldsymbol{\beta}) \simeq \mathbf{z}_t(\widetilde{\boldsymbol{\beta}}) + \left.\frac{\mathrm{d}\mathbf{z}_t}{\mathrm{d}\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}} \times (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})$$

where $\widetilde{\boldsymbol{\beta}}$ is the previous draw of $\boldsymbol{\beta}$ in the M-H sampler. Furthermore, let us define the following:

$$\mathbf{r}_t \doteq \mathbf{z}_t(\widetilde{\boldsymbol{\beta}}) + G_t\widetilde{\boldsymbol{\beta}} \quad , \quad G_t \doteq -\left.\frac{\mathrm{d}\mathbf{z}_t}{\mathrm{d}\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}} \tag{7.14}$$

where the $K \times K$ matrix G_t can be computed by the following recursion:

$$G_t \doteq v_{t-1} I_K - Z_{t-1} + G_{t-1}\widetilde{\boldsymbol{\beta}}$$

where Z_{t-1} is a $K \times K$ diagonal matrix with $\mathbf{z}_{t-1}(\widetilde{\boldsymbol{\beta}})$ in its diagonal, I_K is a $K \times K$ identity matrix and G_0 is a $K \times K$ matrix of zeros. This recursion is simply obtained by differentiating (7.10) with respect to $\boldsymbol{\beta}$. From the definitions in (7.14) we get $\mathbf{z}_t \simeq \mathbf{r}_t - G_t\boldsymbol{\beta}$ and the approximation for z_t is obtained as $z_t \simeq r_t - \mathbf{e}_t'G_t\boldsymbol{\beta}$ where $r_t \doteq \mathbf{e}_t'\mathbf{r}_t$. Let us now define the $T \times 1$ vector $\mathbf{r} \doteq (r_1 \cdots r_T)'$ as well as the $T \times K$ matrix G whose tth row is $\mathbf{e}_t'G_t$. It turns out that $\mathbf{z} \simeq \mathbf{r} - G\boldsymbol{\beta}$, thus we can approximate the exponential of the approximate likelihood (7.11) with:

$$\exp\left[-\tfrac{1}{2}(\mathbf{r} - G\boldsymbol{\beta})'\Lambda^{-1}(\mathbf{r} - G\boldsymbol{\beta})\right] .$$

The proposal density to sample vector $\boldsymbol{\beta}$ is obtained by combining this approximation with the prior density by Bayes' update:

$$q_\beta(\boldsymbol{\beta} \mid \boldsymbol{\alpha}, \widetilde{\boldsymbol{\beta}}, \boldsymbol{\varpi}, \nu, \mathbf{s}, \mathbf{y}) \propto \mathcal{N}_K(\boldsymbol{\beta} \mid \widehat{\boldsymbol{\mu}}_\beta, \widehat{\Sigma}_\beta)\mathbb{I}_{\{\boldsymbol{\beta}>\mathbf{0}\}} \tag{7.15}$$

with:

$$\widehat{\Sigma}_{\beta}^{-1} \doteq G'\widetilde{\Lambda}^{-1}G + \Sigma_{\beta}^{-1}$$
$$\widehat{\mu}_{\beta} \doteq \widehat{\Sigma}_{\beta}(G'\widetilde{\Lambda}^{-1}\mathbf{r} + \Sigma_{\beta}^{-1}\boldsymbol{\mu}_{\beta})$$

where the $T \times T$ diagonal matrix $\widetilde{\Lambda} \doteq \text{diag}\big(\{2\mathbf{e}_t'\mathbf{h}_t^2(\boldsymbol{\alpha}, \widetilde{\boldsymbol{\beta}})\}_{t=1}^T\big)$. A candidate $\boldsymbol{\beta}^\star$ is sampled from this proposal density and accepted with probability:

$$\min\left\{\frac{p(\boldsymbol{\alpha}, \boldsymbol{\beta}^\star, \boldsymbol{\varpi}, \nu, \mathbf{s} \mid \mathbf{y})}{p(\boldsymbol{\alpha}, \widetilde{\boldsymbol{\beta}}, \boldsymbol{\varpi}, \nu, \mathbf{s} \mid \mathbf{y})} \frac{q_\beta(\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}^\star, \boldsymbol{\varpi}, \nu, \mathbf{s}, \mathbf{y})}{q_\beta(\boldsymbol{\beta}^\star \mid \boldsymbol{\alpha}, \widetilde{\boldsymbol{\beta}}, \boldsymbol{\varpi}, \nu, \mathbf{s}, \mathbf{y})}, 1\right\} .$$

7.2.4 Generating vector $\boldsymbol{\varpi}$

The components of $\boldsymbol{\varpi}$ are independent a posteriori and the full conditional posterior of ϖ_t is obtained as follows:

$$\begin{aligned} p(\varpi_t \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \nu, \mathbf{s}, \mathbf{y}) &\propto \mathcal{L}(\Theta \mid \mathbf{y}) p(\varpi_t \mid \nu) \\ &\propto \varpi_t^{-\frac{(\nu+3)}{2}} \exp\left[-\frac{b_t}{\varpi_t}\right] \end{aligned} \tag{7.16}$$

with:

$$b_t \doteq \frac{1}{2}\left[\frac{y_t^2}{\varrho h_t} + \nu\right]$$

where we recall that $h_t \doteq \mathbf{e}_t'\mathbf{h}_t(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\varrho \doteq \frac{\nu-2}{\nu}$. Expression (7.16) is the kernel of an Inverted Gamma density with parameters $\frac{\nu+1}{2}$ and b_t.

7.2.5 Generating parameter $\boldsymbol{\nu}$

Draws from $p(\nu \mid \boldsymbol{\varpi})$ are made by optimized rejection sampling from a translated Exponential source density. This is achieved by following the lines of **Sect. 5.2.4**.

Finally, we note that the computer code and the correctness of the algorithm are tested as in previous chapters; the testing methodology is applicable to the constrained as well as unconstrained versions of the permutation sampler.

7.3 An application to the Swiss Market Index

We apply our Bayesian estimation method to demeaned daily log-returns $\{y_t\}$ of the Swiss Market Index (henceforth SMI). The sample period is from November 12, 1990, to December 16, 2005 for a total of 3'800 observations and the log-returns are expressed in percent. The data set is freely available from the website

`http://www.finance.yahoo.com`. Note that September 11, 2001, has not been recorded by the data provider since the stock markets closed after the terrorist attacks for a few days. From this time series, the first 2'500 observations (up to November 2001), which represent slightly less than two third of the data set, are used to estimate the model while the remaining 1'300 log-returns are used in a forecasting performance analysis.

The time series under investigation is plotted in the upper part of **Fig. 7.1** where the vertical line delimits the in- and out-of-sample observation windows. We test for autocorrelation in the times series by testing the joint nullity of autoregressive coefficients for $\{y_t\}$. We estimate the regression with autoregressive coefficients up to lag 15 and compute the covariance matrix using the White estimate. The p-value of the Wald test is 0.5299 which does not support the presence of autocorrelation. When testing for the autocorrelation in the series of squared observations $\{y_t^2\}$, we strongly reject the absence of autocorrelation. This is in line with the autocorrelogram of $\{y_t^2\}$ plotted in the lower part of **Fig. 7.1**. The autocorrelations are large and significantly different from zero up to lag 70. As an additional data analysis, we test for unit root using the test by Phillips and Perron [1988]. The test strongly rejects the $I(1)$ hypothesis.

We estimate the single-regime GJR$(1,1)$ model as well as the two-state Markov-switching GJR$(1,1)$ model henceforth referred to as GJR and MS-GJR for convenience. Both models are estimated using the MCMC scheme presented in **Sect. 7.2**. The estimation of the GJR model is obtained as a simplified version of the algorithm when $K = 1$ by setting the $T \times 1$ vector $\mathbf{s}$ to a vector of ones and omitting the generation of the transition matrix. For the hyperparameters on priors $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$, we set $\mu_{\alpha_i} (i = 0, 1, 2)$ and μ_β to zero mean vectors and choose diagonal covariance matrices for $\Sigma_{\alpha_i} (i = 0, 1, 2)$ and Σ_β. The variances are set to $\sigma^2_{\alpha_i} = \sigma^2_\beta = 10'000$ $(i = 0, 1, 2)$ so we do not introduce tight prior information in our estimation. In the case of the prior on the degrees of freedom parameter, we choose $\lambda = 0.01$ and $\delta = 2$; this therefore ensures the existence for the conditional variance. Finally, the hyperparameters for the prior on the transition probabilities are set to $\eta_{ii} = 2$ and $\eta_{ij} = \eta_{ji} = 1$ for $i, j \in \{1, 2\}$ so that we have a prior belief that the probabilities of persistence are bigger than the probabilities of transition.

For both models, we run two chains for 50'000 iterations each and assess the convergence of the sampler by using the diagnostic test by Gelman and Rubin [1992]. The convergence appears rather quickly, but we nevertheless consider the first half of the iterations as a burn-in phase for precaution. For the GJR model, the acceptance rates range from 88% for vector $\boldsymbol{\alpha}$ to 97% for β indicating that the proposal densities are close to the exact conditional posteriors. The one-

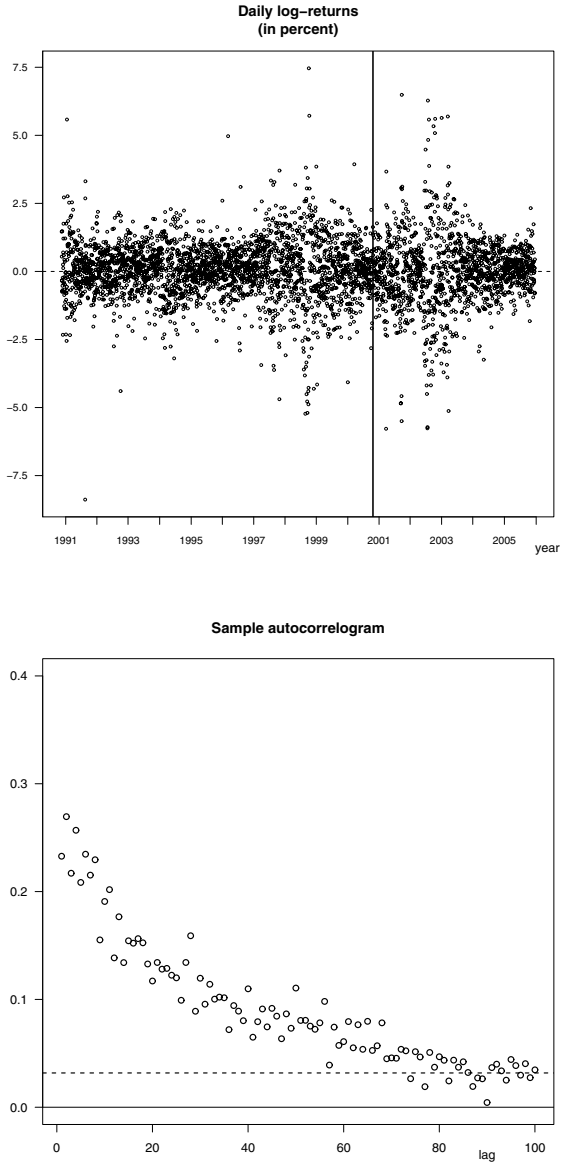

Fig. 7.1. SMI daily log-returns (upper graph) and sample autocorrelogram of the squared log-returns up to lag 100 (lower graph). The vertical line in the upper graph delimits the in-sample and out-of-sample observation windows.

lag autocorrelations in the chain range from 0.52 for α_1 to 0.96 for β which is reasonable. For the MS-GJR model, the random permutation sampler is run first to determine suitable identification constraints. In **Fig. 7.2**, we show the contour

plots of the posterior density for (β^k, α_0^k), (β^k, α_1^k) and (β^k, α_2^k), respectively. Note that the state value k is arbitrary since all marginal densities contain the same information [see Frühwirth-Schnatter 2001b]. As we can notice, the bimodality of the posterior density is clear for the parameter β^k on the three graphs, suggesting a constraint of the type $\beta^1 < \beta^2$ for identification. Therefore, the model is estimated again by imposing this constraint at each sweep in the sampler and the definition of the states is permuted if the constraint is violated. In that case, label switching only appeared 16 times after the burn-in phase thus confirming the suitability of the identification constraint. The acceptance rates obtained with the constrained version of the permutation sampler range from 22% for the vector $\boldsymbol{\alpha}$ to 93% for $\boldsymbol{\beta}$. The one-lag autocorrelations range from 0.82 for α_1^2 to 0.97 for β^2. We keep every fifth draw from the MCMC output for both models in order to diminish the autocorrelation in the chains. The two chains are then merged to get a final sample of length 10'000. Finally, we note that a three-state Markov-switching GJR model has also been estimated. However, post-processing the MCMC output has not allowed to find a clear identification constraint.

The posterior statistics for both models are reported in **Table 7.1**. In the case of the GJR model (upper panel), we note the high persistence for the conditional variance process, measured by $\overline{\alpha} + \beta$ where $\overline{\alpha} \doteq \frac{\alpha_1+\alpha_2}{2}$, as well as the presence of the leverage effect. The estimation of the probability $\mathbb{P}(\alpha_2 > \alpha_1 \mid \mathbf{y})$ is 0.999, supporting the asymmetric behavior of the conditional variance. The low value for the estimated degrees of freedom parameter indicates conditional leptokurtosis in the data set. In the MS-GJR case (lower panel), we note also the presence of the leverage effect in both states. A comparison of the scedastic function's parameters between regimes indicates similar 95% confidence intervals for the components of the vectors $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ while the difference for components of the $\boldsymbol{\alpha}_0$ vector is more pronounced. Indeed, for $i = 0, 1, 2$, the estimated probabilities $\mathbb{P}(\alpha_i^1 > \alpha_i^2 \mid \mathbf{y})$ are respectively 0.774, 0.397 and 0.543. As in the single-regime model, the posterior density for the degrees of freedom parameter indicates conditional leptokurtosis. We note however that the posterior mean and median are larger than for the GJR model. The posterior means for probabilities p_{11} and p_{22} are respectively 0.997 and 0.995 indicating infrequent mixing between states. Finally, the inefficiency factors (IF) reported in the last column of **Table 7.1** indicate that using 10'000 draws out of the MCMC sampler seems appropriate if we require that the Monte Carlo error in estimating the mean is smaller than one percent of the variation of the error due to the data. We recall that the IF are computed as the ratio of the squared numerical standard error (NSE) of the MCMC simulations and the variance estimate divided by the

number of iterations (*i.e.*, the variance of the sample mean from a hypothetical *iid* sampler). The NSE are estimated by the method of Andrews [1991], using a Parzen kernel and AR(1) pre-whitening as presented in Andrews and Monahan [1992]. As noted by Deschamps [2006], this ensures easy, optimal, and automatic bandwidth selection.

In **Fig. 7.3**, we display the marginal posterior densities for the MS-GJR model parameters. First, we note that the use of the constrained permutation sampler leads to marginal densities which are unimodal. Furthermore, we clearly notice that most of these densities are skewed. More precisely, the densities for the components of vector $\boldsymbol{\alpha}$ are right-skewed while components of $\boldsymbol{\beta}$ are left-skewed. In the case of parameters α_1^1 and α_1^2, the modes of the densities are close to the lower boundary of the parameter's space, suggesting that the parameters are close to zero. Finally, we can notice that the posterior densities for p_{11} and p_{22} are strongly left-skewed.

Some probabilistic statements on nonlinear functions of the parameters can be straightforwardly obtained by simulation from the joint posterior sample $\{\psi^{[j]}\}_{j=1}^{J}$. In particular, we can test the covariance stationarity condition and estimate the density of the unconditional variance when this condition is satisfied. Under the GJR specification, the process is covariance stationary if $\overline{\alpha} + \beta < 1$ where we recall that $\overline{\alpha} \doteq \frac{\alpha_1+\alpha_2}{2}$ for notational purposes. The estimated probability $\mathbb{P}(\overline{\alpha} + \beta < 1 \mid \mathbf{y})$ is one. Hence, the unconditional variance exists and is given by $\frac{\alpha_0}{1-\overline{\alpha}-\beta}$; the estimation of its posterior mean is 1.179 with a 95% confidence interval given by [1.173,1.189]. These estimations can be compared with the empirical variance of the process which is 1.136. In this case, the single-regime model slightly overestimates the variability of the underlying time series. For the Markov-switching model, our simulation study indicates that the process is covariance stationary in each state. The posterior mean of the unconditional variances is 0.56 in state 1 and 2.00 in state 2 with 95% confidence intervals respectively given by [0.557,0.563] and [1.992,2.012]. The unconditional variance of the process in state 1 is about four times lower than the one in state 2; we will therefore refer state 1 as the *low-volatility* regime and state 2 as the *high-volatility* regime. As found by Haas *et al.* [2004, Eq.11, p.500], the Markov-switching GARCH process is covariance stationary if $\xi(M) < 1$, where $\xi(M)$ denotes the largest eigenvalue in modulus of matrix M. This matrix is constructed from the model parameters and, in the case of the MS-GJR model, it is given by:

Table 7.1. Estimation results for the GJR model (upper panel) and MS-GJR model (lower panel).★

GJR model								
ψ	$\overline{\psi}$	$\psi_{0.5}$	$\psi_{0.025}$	$\psi_{0.975}$	min	max	NSE	IF
α_0	0.066	0.065	0.041	0.099	0.021	0.156	0.356	5.58
α_1	0.060	0.059	0.028	0.098	0.005	0.162	0.237	1.81
α_2	0.207	0.205	0.148	0.278	0.097	0.359	0.690	4.33
β	0.809	0.809	0.750	0.861	0.656	0.911	1.163	16.22
ν	8.083	7.954	6.258	10.580	4.871	13.930	34.643	9.79
MS-GJR model								
ψ	$\overline{\psi}$	$\psi_{0.5}$	$\psi_{0.025}$	$\psi_{0.975}$	min	max	NSE	IF
α_0^1	0.245	0.241	0.149	0.362	0.100	0.518	2.407	19.26
α_0^2	0.184	0.178	0.089	0.327	0.046	0.518	1.939	10.45
α_1^1	0.020	0.015	0.001	0.063	0.000	0.145	0.276	2.61
α_1^2	0.027	0.023	0.001	0.073	0.000	0.135	0.302	2.33
α_2^1	0.229	0.224	0.123	0.361	0.074	0.534	1.278	4.21
α_2^2	0.220	0.215	0.136	0.332	0.090	0.462	1.140	5.21
β^1	0.436	0.440	0.212	0.642	0.004	0.746	4.454	16.80
β^2	0.782	0.785	0.670	0.866	0.582	0.907	2.090	18.33
ν	9.459	9.264	7.051	12.880	5.881	23.740	55.931	13.45
p_{11}	0.997	0.997	0.992	0.999	0.982	1.000	0.022	1.23
p_{12}	0.003	0.003	0.001	0.008	0.001	0.018	0.022	1.23
p_{21}	0.005	0.004	0.001	0.011	0.001	0.023	0.027	1.13
p_{22}	0.995	0.996	0.989	0.999	0.978	1.000	0.027	1.13

★ $\overline{\psi}$: posterior mean; ψ_ϕ: estimated posterior quantile at probability ϕ; min: minimum value; max: maximum value; NSE: numerical standard error ($\times 10^3$); IF: inefficiency factor (*i.e.*, ratio of the squared numerical standard error and the variance of the sample mean from a hypothetical *iid* sampler). The posterior statistics are based on 10'000 draws from the constrained posterior sample.

$$M \doteq \begin{pmatrix} p_{11}(\overline{\alpha}^1 + \beta^1) & 0 & p_{21}(\overline{\alpha}^1 + \beta^1) & 0 \\ p_{11}\alpha_1^2 & p_{11}\beta^2 & p_{21}\alpha_1^2 & p_{21}\beta^2 \\ p_{12}\beta^1 & p_{12}\alpha_1^1 & p_{22}\beta^1 & p_{22}\alpha_1^1 \\ 0 & p_{12}(\overline{\alpha}^2 + \beta^2) & 0 & p_{22}(\overline{\alpha}^2 + \beta^2) \end{pmatrix} \tag{7.17}$$

where $\overline{\alpha}^k \doteq \frac{\alpha_1^k + \alpha_2^k}{2}$. Using the posterior sample we can thus estimate the density of $\xi(M)$ by substituting the values of the draws for the model parameters in the definition (7.17). In the upper part of **Fig. 7.4**, we present the posterior density for $\xi(M)$. As we can notice, none of the values exceed one in our simulation. Thus, the model is covariance stationary. Therefore, the unconditional variance of the MS-GJR process exists and is given by:

$$h_y \doteq (\operatorname{vec} P)' \times (I_4 - M)^{-1} \times (\boldsymbol{\pi} \otimes \boldsymbol{\alpha}_0) \tag{7.18}$$

where $\boldsymbol{\pi}$ is the 2×1 vector of ergodic probabilities of the Markov chain, I_4 is a 4×4 identity matrix, vec denotes the vectorization operator which stacks the columns of a matrix one underneath the other and $\otimes$ denotes the Kronecker product. Derivation of formula (7.18) can be found in Haas *et al.* [2004, p.501]. The posterior density of the unconditional variance is shown in the lower part of **Fig. 7.4**. Its posterior mean is 1.134 with a 95% confidence interval of [1.128,1.139]. In this case, the confidence band for the mean contains the empirical variance of 1.136 contrary to the one in the GJR model. This suggests that the Markov-switching model is more apt to reproduce the variability of the data.

Finally, since the states vector $\mathbf{s} \doteq (s_1 \cdots s_T)'$ is considered as a parameter in the MCMC procedure, the draws $\{\mathbf{s}^{[j]}\}_{j=1}^{J}$ can also be stored and used to make inference about the smoothed probabilities. Theses probabilities are estimated as the percentage of replications of s_t corresponding to regime k:

$$\mathbb{P}(s_t = k \mid \mathbf{y}) \approx \frac{1}{J} \sum_{j=1}^{J} \mathbb{I}_{\{s_t^{[j]} = k\}} \, .$$

In **Fig. 7.5**, we present the smoothed probabilities for the high-volatility regime (solid line, left axis) together with the in-sample daily log-returns (circles, right axis). The 95% confidence bands are shown in dashed lines but are almost indistinguishable from the point estimates. The beginning of year 1991 is associated with the high-volatility state. Then, from the second half of 1991 to 1997, the returns are clearly associated with the low-volatility regime, with the exception of 1994. From 1997 to 2000, the model remains in the high-volatility regime with a transition during the second semester 2000 to the low-volatility state.

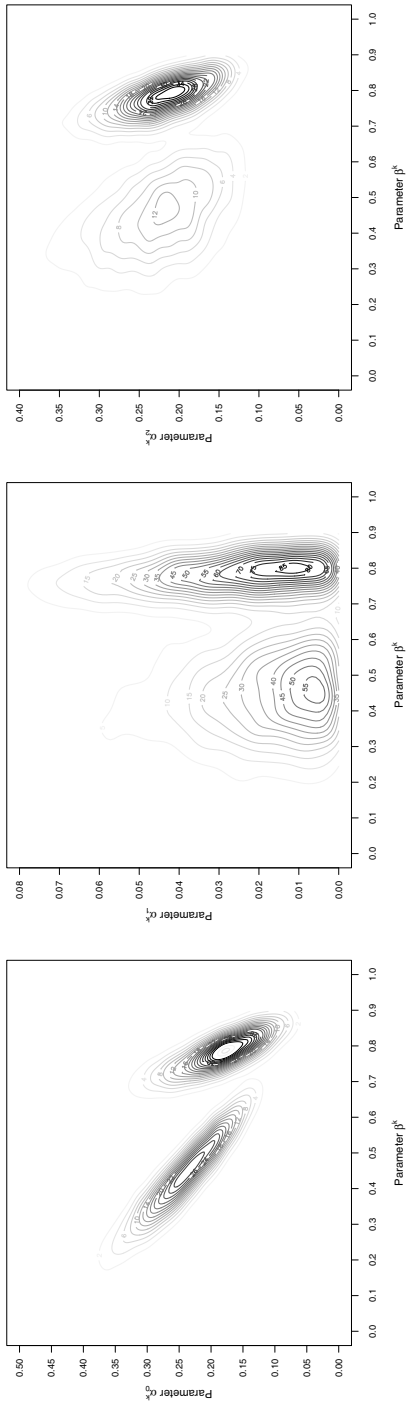

Fig. 7.2. Contour plots for (β^k, α_0^k), (β^k, α_1^k) and (β^k, α_2^k), respectively. The choice of k is arbitrary since all marginal densities contain the same information [see Frühwirth-Schnatter 2001b]. The graphs are based on 10'000 draws from the joint posterior sample.

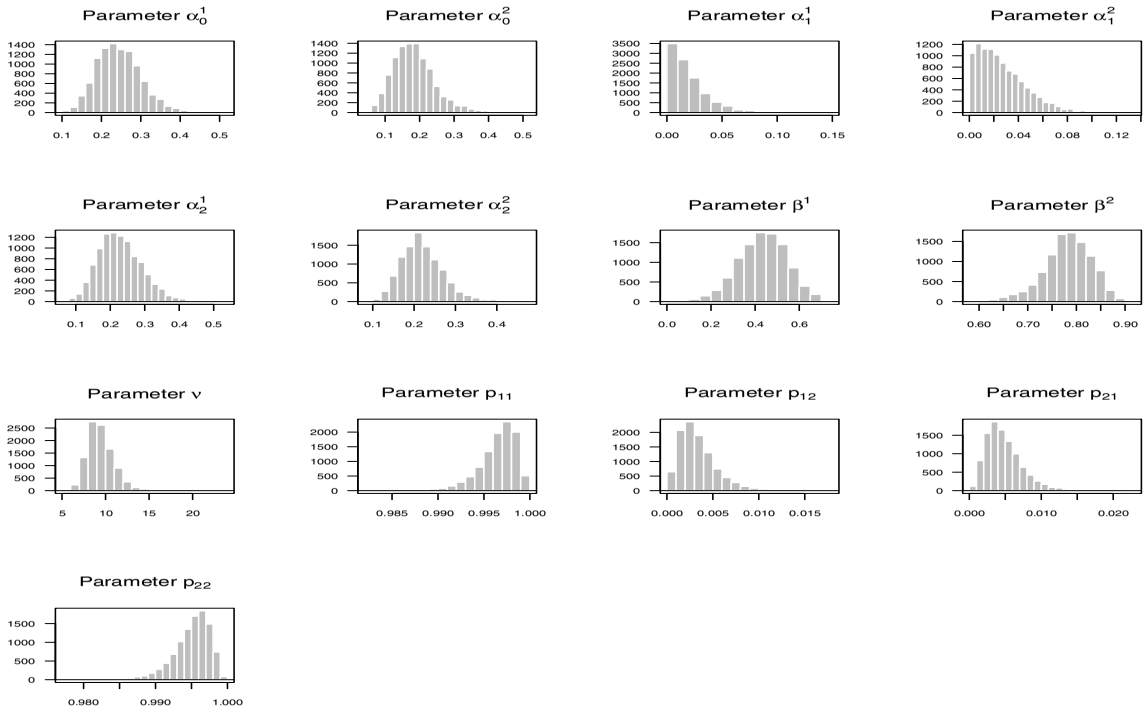

Fig. 7.3. Marginal posterior densities of the MS-GJR parameters. The histograms are based on 10'000 draws from the constrained posterior sample.

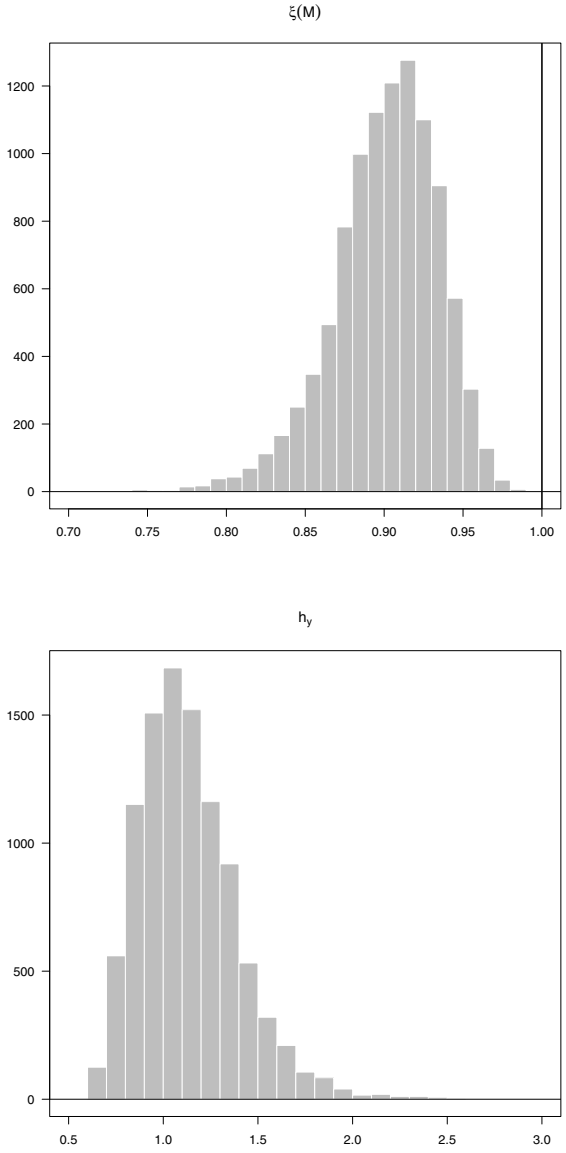

Fig. 7.4. Posterior densities of the covariance stationarity condition (upper graph) and the unconditional variance (lower graph) of the MS-GJR process. The histograms are based on 10'000 draws from the constrained posterior sample.

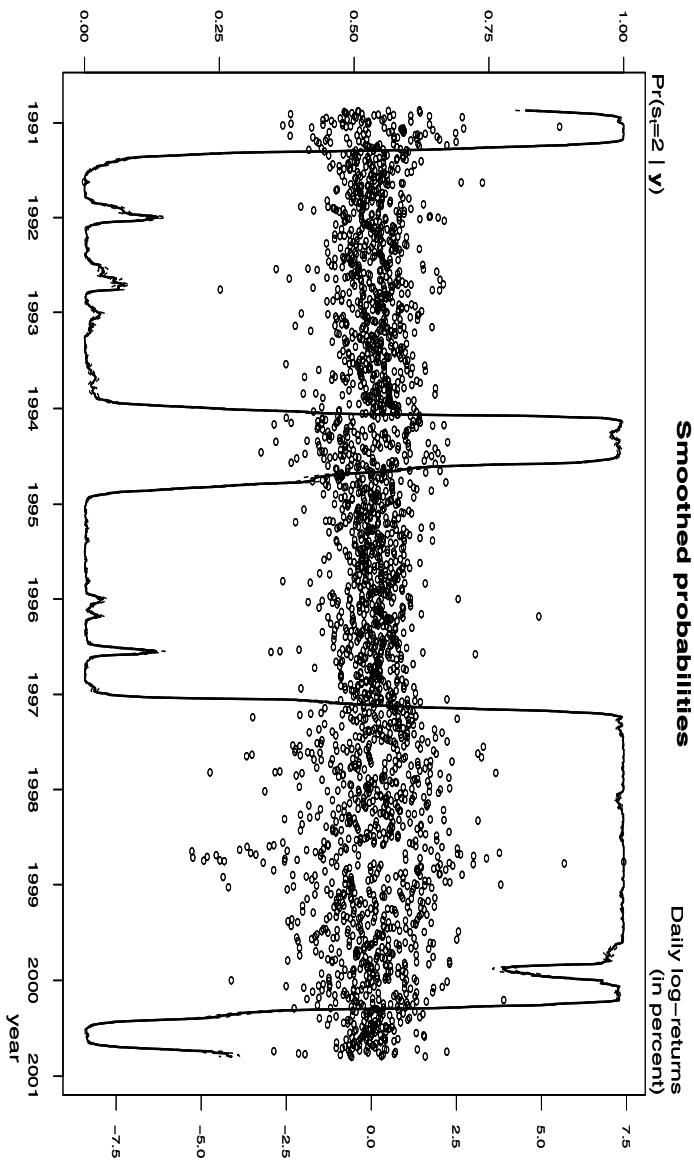

Fig. 7.5. Smoothed probabilities of the high-volatility state (solid line, left axis) together with the in-sample log-returns (circles, right axis). The 95% confidence bands are shown in dashed lines (they are almost indistinguishable from the point estimates).

7.4 In-sample performance analysis

7.4.1 Model diagnostics

We check for model misspecification by analyzing the predictive probabilities referred to as probability integral transforms or p-scores in the literature [see, *e.g.*, Diebold, Gunther, and Tay 1998, Kaufmann and Frühwirth-Schnatter 2002]. We make use of a simpler version of this method, as proposed by Kim, Shephard, and Chib [1998], which consists in conditioning on point estimates of ψ. To be meaningful, the point estimate has to be chosen when the identification is imposed. Hence, we consider the posterior mean $\overline{\psi}$ of the constrained posterior sample. Upon defining $\mathcal{F}_{t-1}$ as the information set up to time $t-1$, the (approximate) p-scores are defined as follows:

$$z_t \doteq \sum_{k=1}^{K} \mathbb{P}(Y_t \leqslant y_t \mid s_t = k, \overline{\psi}, \mathcal{F}_{t-1})\, \mathbb{P}(s_t = k \mid \overline{\psi}, \mathcal{F}_{t-1}) \; .$$

The probability $\mathbb{P}(Y_t \leqslant y_t \mid s_t = k, \overline{\psi}, \mathcal{F}_{t-1})$ can be estimated by the Student-t integral and the filtered probability $\mathbb{P}(s_t = k \mid \overline{\psi}, \mathcal{F}_{t-1})$ is obtained as a byproduct from the FFBS algorithm [see Chib 1996, p.83]. Under a correct specification, the p-scores should have independent uniform distributions asymptotically [see Rosenblatt 1952]. A further transformation through the Normal integral is often applied for convenience. In this case, we consider $u_t \doteq \Phi^{-1}(z_t)$ where $\Phi^{-1}(\bullet)$ denotes the inverse cumulative standard Normal function. If the model is correct, these *generalized residuals* $\{u_t\}$ should be independent standard Normal and common tests can be used to check these features. In particular, we test the presence of autocorrelation in the series $\{u_t\}$ and $\{u_t^2\}$ using a Wald test. We also report the results of a joint test for zero mean, unit variance, zero skewness, and the absence of excess kurtosis, employing the likelihood ratio framework proposed by Berkowitz [2001]. For precisions on the testing methodology, we refer the reader to Haas *et al.* [2004, p.516].

In the case of the GJR model, the Wald statistic for testing the joint nullity of autoregressive coefficients, up to lag 15, for u_t has a p-value of 0.0868 and for u_t^2, a p-value of 0.399. In the case of the MS-GJR model, the p-values are 0.0745 and 0.464, respectively. Therefore, both models seem adequate in removing the volatility clustering present in the data set. The likelihood ratio framework for testing the first four moments of the transformed residuals yields p-values of 0.125 for the GJR model and 0.0635 for the MS-GJR model. Overall, these results indicate no evidence of misspecification at the 5% significance level for both models.

7.4.2 Deviance information criterion

In order to evaluate the goodness-of-fit of the models, we use first the Deviance information criterion (henceforth DIC) introduced by Spiegelhalter *et al.* [2002]. The DIC is not intended for identification of the correct model, but rather merely as a method of comparing a collection of alternative formulations (all of which may be incorrect) and determining the most appropriate. This criterion follows from an extension of the *Deviance* proposed by Dempster [1997]. A recent article by Berg, Meyer, and Yu [2004] has illustrated the potential advantages of this information criterion in determining the appropriate stochastic volatility model. This criterion presents an interesting alternative to the Bayes factor which is often difficult to calculate, especially for models that involve many random effects, large number of unknowns or improper priors.

Let us denote the model parameters by θ for the moment. Based on the posterior density of the Deviance $D(\theta) \doteq -2 \ln \mathcal{L}(\theta \mid \mathbf{y})$ where $\mathcal{L}(\theta \mid \mathbf{y})$ is the likelihood function, the DIC consists of two terms: a component that measures the goodness-of-fit and a penalty term for any increase in model complexity. The measure of fit is obtained by taking the posterior expectation of the Deviance:

$$\overline{D} \doteq \mathbb{E}_{\theta|\mathbf{y}}\big[D(\theta)\big] \tag{7.19}$$

where $E_{\theta|\mathbf{y}}[\bullet]$ denotes the expectation with respect to the joint posterior $p(\theta \mid \mathbf{y})$. Provided that $D(\theta)$ is available in closed form, $\overline{D}$ can easily be approximated using the posterior sample by estimating the sample mean of the simulated values of $D(\theta)$. The second component measures the complexity of the model using the *effective number of parameters*, denoted by p_D, and is defined as the difference between the posterior mean of the Deviance and the Deviance evaluated at a point estimate $\widetilde{\theta}$:

$$p_D \doteq \overline{D} - D(\widetilde{\theta}) \; . \tag{7.20}$$

A natural candidate for $\widetilde{\theta}$ is the posterior mean $\mathbb{E}_{\theta|\mathbf{y}}(\theta)$, as suggested by Spiegelhalter *et al.* [2002]. When the density is log-concave, this point estimate ensures a positive p_D due to Jensen's inequality. The DIC is then simply defined as $\text{DIC} \doteq \overline{D} + p_D$ and, given a set of models, the one with the smallest DIC has the best balance between goodness-of-fit and model complexity.

As noted in Celeux, Forbes, Robert, and Titterington [2006], the definition $\widetilde{\theta} \doteq \mathbb{E}_{\theta|\mathbf{y}}(\theta)$ is not appropriate in mixture models when no identification is imposed. Furthermore, when the state variable is discrete and considered as a parameter in θ, the posterior expectation usually fails to take one of the discrete values. To overcome these difficulties, we integrate out the state vector by

considering the *observed likelihood* instead [see Celeux *et al.* 2006, Sect.3.1] and make use of the constrained posterior sample in the estimation. In the context of MS-GARCH models, the observed likelihood, also referred to as the *marginal likelihood* in Kaufmann and Frühwirth-Schnatter [2002, p.457] is obtained as follows:

$$\mathcal{L}(\psi \mid \mathbf{y}) = \prod_{t=1}^{T} \left[\sum_{k=1}^{K} p(y_t \mid \psi, s_t = k, \mathcal{F}_{t-1}) \, \mathbb{P}(s_t = k \mid \psi, \mathcal{F}_{t-1}) \right] \tag{7.21}$$

where $p(y_t \mid \psi, s_t = k, \mathcal{F}_{t-1})$ can be estimated by the Student-t density and the filtered probability $\mathbb{P}(s_t = k \mid \psi, \mathcal{F}_{t-1})$ is obtained as a byproduct from the FFBS algorithm [see Chib 1996, p.83]. The DIC is then defined as the sum of components (7.19) and (7.20), which yields:

$$\text{DIC} \doteq 2 \Big\{ \ln \mathcal{L}\big(\overline{\psi} \mid \mathbf{y}\big) - 2\mathbb{E}_{\psi|\mathbf{y}}\big[\ln \mathcal{L}(\psi \mid \mathbf{y}) \big] \Big\}$$

where we recall that $\overline{\psi} \doteq \mathbb{E}_{\psi|\mathbf{y}}(\psi)$ with $\psi \doteq (\boldsymbol{\alpha}, \boldsymbol{\beta}, \nu, P)$.

In order to make statements about the goodness-of-fit of one model relative to another, it is important to consider the uncertainty in the DIC. While the confidence interval for $\overline{D}$ can be easily obtained from the MCMC output by using spectral methods as this is done for the posterior mean, the task is more tedious in the case of p_D and hence for the DIC itself. Approximation methods have been experimented in Zhu and Carlin [2000] but the *brute force* approach is still the most accurate. In this method, the variability of the DIC is estimated by running several MCMC chains and calculating the DIC's variance from the different runs. Obviously, this is extremely costly. A simpler alternative consists in running few MCMC runs and reporting the minimum and maximum DIC obtained. This gives however a crude idea of DIC's variability. In what follows, we make use of a methodology which estimates the whole distribution for the DIC based on a resampling technique. More precisely, from the joint posterior sample $\{\psi^{[j]}\}_{j=1}^{J}$, we generate randomly B new posterior samples of size J by using the block bootstrap technique and estimate DIC's components for these samples. By comparing the 95% confidence interval of the different DICs, we can find statistical evidence of differences in the fitting quality. With this methodology, the MCMC procedure does not need to be re-run which strongly diminishes the computing time. The choice of the block length is an important issue in the block bootstrap technique. For the block bootstrap to be effective, the length should be large enough so that it includes most of the dependence structure, but not too large so that the number of blocks becomes insufficient. In our analysis, we use the stationary bootstrap of Politis and Romano [1994]

and select the block length following the algorithm based on the spectral density estimation, as proposed by Politis and White [2004]. We apply the block length selection algorithm to each parameter's output. The maximum value is then defined as the optimal block length used for block bootstrapping the constrained posterior sample. This ad-hoc procedure allows to keep the autocorrelation in the chains as well as the cross-dependence structure between the parameters.

Results for the DIC and its components are reported in **Table 7.2**. They are based on 10'000 draws from the constrained posterior distribution. In squared brackets we give the 95% confidence interval obtained by the resampling technique using $B = 100$ replications. We keep every tenth draw from the joint posterior sample in the resampling technique in order to speed up the calculations and diminish the autocorrelation in the chains. For comparison purposes, we also consider the Bayesian information criterion introduced by Schwarz [1978] which is defined as follows:

$$\mathrm{BIC}(\psi) \doteq 2 \ln \mathcal{L}(\psi \mid \mathbf{y}) - n \ln T$$

where n is the number of parameters and T the number of observations. In our context, $T = 2'500$, $n = 5$ for the GJR model and $n = 11$ for the MS-GJR model (since parameters p_{12} and p_{21} are redundant due to the summability constraint). This criterion promotes model parsimony by penalizing models with increased model complexity (larger n) and sample size T. Hence, a model with the largest BIC is preferred. The computation of the Bayesian information criterion is based on the posterior mean $\mathbb{E}_{\psi|\mathbf{y}}\big[\mathrm{BIC}(\psi)\big]$ obtained over the 10'000 draws of the constrained posterior sample.

Table 7.2. Results of the DIC and BIC criteria.*

Model	DIC	$\overline{D}$	p_D	$\mathbb{E}_{\psi\|\mathbf{y}}$(BIC)
GJR	6770.4	6765.6	4.76	-6806.07
	[6769.9,6770.8]	[6765.3,6765.8]	[4.49,4.93]	(7.12)
MS-GJR	6713.3	6704.4	8.84	-6804.73
	[6712.6,6713.8]	[6793.9,6794.9]	[8.49,9.04]	(12.55)

* DIC: Deviance information criterion; $\overline{D}$: Deviance evaluated at the posterior mean $\overline{\psi}$ (see **Table 7.1**, p.127); p_D: effective number of parameters; $\mathbb{E}_{\psi|\mathbf{y}}$(BIC): posterior mean of BIC(ψ) obtained over the 10'000 draws of the constrained posterior sample; [•]: 95% confidence interval based on $B = 100$ replications of the constrained posterior sample; (•): numerical standard error ($\times 10^2$).

From **Table 7.2**, we can notice that both DIC and BIC criteria favor the MS-GJR model. Indeed, the DIC estimates based on the initial joint posterior sample is 6770.4 for the GJR model and 6713.3 for the MS-GJR model. Both 95% confidence intervals do not overlap which suggests significant improvement of the Markov-switching model. In the case of BIC, the differences between the criterion's values are less pronounced but still the Markov-switching model is favored compared to the single-regime model. If we consider now the estimations of p_D, we note that the estimated value is somewhat lower than five in the GJR model while about nine in the MS-GJR case. Hence, in the single-regime model, every parameter seems to be effective (or informative) when fitting the model to the data set. In the Markov-switching model however, about two third of the 13 parameters are effective. This is in line with the estimation results where it was shown that parameters α_1 and α_2 are almost identical across regimes. Furthermore, the 2×2 transition matrix only contains two free parameters due to the summability constraint. This suggests that the nine effective parameters of the MS-GJR model are α_0^1, α_0^2, α_1, α_2, β^1, β^2, ν, p_{11} and p_{22}.

Finally, we point out that we have also considered the posterior mode:

$$\widetilde{\psi} \doteq \arg\max_{\psi} \mathcal{L}(\psi \mid \mathbf{y})$$

in the definition of p_D, as suggested by Celeux *et al.* [2006, Sect.3.1]. It is argued that such a point estimate is more relevant since it depends on the posterior distribution of the whole parameter ψ, rather than on the marginal posterior distributions of its elements. The values of p_D obtained with this new definition are larger for both models with 95% confidence intervals respectively given by [5.17,5.66] and [10.06,11.12] for the single-regime and Markov-switching models. While the preferred model remains the MS-GJR, the interpretation of parameter p_D is now questionable in the GJR case since the value of p_D exceeds the total number of parameters.

7.4.3 Model likelihood

As a second criterion to discriminate between the models under study, we consider the *model likelihood* which may be expressed as follows:

$$p(\mathbf{y}) = \int \mathcal{L}(\psi \mid \mathbf{y}) p(\psi) \mathrm{d}\psi$$

where $\mathcal{L}(\psi \mid \mathbf{y})$ is the marginal likelihood given in (7.21) and $p(\psi)$ is the joint prior density on $\psi \doteq (\boldsymbol{\alpha}, \boldsymbol{\beta}, \nu, P)$. It is clear that the model likelihood is equal to the normalizing constant of the posterior density:

$$p(\psi \mid \mathbf{y}) = \frac{\mathcal{L}(\psi \mid \mathbf{y})p(\psi)}{p(\mathbf{y})} \ .$$

The estimation of $p(\mathbf{y})$ requires the integration over the whole set of parameters ψ, which is a difficult task in practice, especially for complex statistical models such as ours. A full investigation of the various approaches available to estimate the model likelihood for finite mixture models can be found in Frühwirth-Schnatter [2004]. In particular, the author documents that the *bridge sampling* technique using the MCMC output of the random permutation sampler and an *iid* sample from an *importance density* $q(\psi)$ which approximates the unconstrained posterior yields the best estimator of the model likelihood (*i.e.*, the estimator with the lowest variance). Moreover, the variance of the bridge sampling estimator depends on a ratio that is bounded regardless of the tail behaviour of the importance density. This renders the estimator robust and gives more flexibility in the choice of the importance density.

First, let us recall that the bridge sampling technique of Meng and Wong [1996] is based on the following result:

$$1 = \frac{\int a(\psi)p(\psi \mid \mathbf{y})q(\psi)\mathrm{d}\psi}{\int a(\psi)q(\psi)p(\psi \mid \mathbf{y})\mathrm{d}\psi} = \frac{\mathbb{E}_q\big[a(\psi)p(\psi \mid \mathbf{y})\big]}{\mathbb{E}_{\psi|\mathbf{y}}\big[a(\psi)q(\psi)\big]} \tag{7.22}$$

where $a(\psi)$ is an arbitrary function such that $\int a(\psi)p(\psi \mid \mathbf{y})q(\psi)\mathrm{d}\psi > 0$ and $\mathbb{E}_q$ denotes the expectation with respect to the importance density $q(\psi)$. Replacing $p(\psi \mid \mathbf{y})$ by $\frac{\mathcal{L}(\psi|\mathbf{y})p(\psi)}{p(\mathbf{y})}$ in expression (7.22) yields the key identity for bridge sampling:

$$p(\mathbf{y}) = \frac{\mathbb{E}_q\big[a(\psi)\mathcal{L}(\psi \mid \mathbf{y})p(\psi)\big]}{\mathbb{E}_{\psi|\mathbf{y}}\big[a(\psi)q(\psi)\big]} \ .$$

We can estimate the model likelihood for a given function $a(\psi)$ by replacing the expectations on the right-hand side of the latter expression by sample averages. More precisely, we use MCMC draws $\{\psi^{[m]}\}_{m=1}^{M}$ from the joint posterior $p(\psi \mid \mathbf{y})$ and *iid* draws $\{\psi^{[l]}\}_{l=1}^{L}$ from the importance sampling density $q(\psi)$ to get the following approximation:

$$p(\mathbf{y}) \approx \frac{\frac{1}{L}\sum_{l=1}^{L} a(\psi^{[l]})\mathcal{L}(\psi^{[l]} \mid \mathbf{y})p(\psi^{[l]})}{\frac{1}{M}\sum_{m=1}^{M} a(\psi^{[m]})q(\psi^{[m]})} \ . \tag{7.23}$$

Meng and Wong [1996] discuss an asymptotically optimal choice for $a(\psi)$, which minimizes the expected relative error of the $p(\mathbf{y})$ estimator for *iid* draws from $p(\psi \mid \mathbf{y})$ and $q(\psi)$. This function is given by:

$$a(\psi) \propto \frac{1}{Lq(\psi) + Mp(\psi \mid \mathbf{y})} \,.$$

This special case of bridge sampling estimator is referred to as the *optimal bridge sampling* estimator by Frühwirth-Schnatter [2001a] and will be used in what follows. As the optimal choice depends on the normalized posterior $p(\psi \mid \mathbf{y})$, Meng and Wong [1996] use an iterative procedure to estimate $p(\mathbf{y})$ as a limit of a sequence $\{p_t(\mathbf{y})\}$. Based on an estimate $p_{t-1}(\mathbf{y})$ of the normalizing constant, the posterior is normalized as follows:

$$p_{t-1}(\psi \mid \mathbf{y}) \doteq \frac{\mathcal{L}(\psi \mid \mathbf{y})p(\psi)}{p_{t-1}(\mathbf{y})}$$

and a new estimate $p_t(\mathbf{y})$ is computed using approximation (7.23). This leads to the following recursion:

$$p_t(\mathbf{y}) \doteq p_{t-1}(\mathbf{y}) \times \frac{\frac{1}{L}\sum_{l=1}^{L} \frac{p_{t-1}(\psi^{[l]}|\mathbf{y})}{Lq(\psi^{[l]})+Mp_{t-1}(\psi^{[l]}|\mathbf{y})}}{\frac{1}{M}\sum_{m=1}^{M} \frac{q(\psi^{[m]})}{Lq(\psi^{[m]})+Mp_{t-1}(\psi^{[m]}|\mathbf{y})}}$$

which can be initialized, *e.g.*, with the *reciprocal importance sampling* estimator of Gelfand and Dey [1994] given by:

$$p_0(\mathbf{y}) = \left[\frac{1}{M}\sum_{m=1}^{M} \frac{q(\psi^{[m]})}{\mathcal{L}(\psi^{[m]} \mid \mathbf{y})p(\psi^{[m]})}\right]^{-1} \,.$$

Note that this latter estimator is only based on MCMC draws from the joint posterior. Convergence of the bridge sampling technique is typically very fast in practice. In our case, the estimates converged after 3–4 iterations.

The remaining task consists in choosing an appropriate importance density to apply the bridge sampling technique. To that aim, we follow Kaufmann and Frühwirth-Schnatter [2002, pp.438–439] and Kaufmann and Scheicher [2006, pp.9–10]. The importance density is constructed in an unsupervised manner from the MCMC output of the random permutation sampler using a mixture of the proposal and conditional densities. Its construction is fully automatic and is easily incorporated in the MCMC sampler [see Frühwirth-Schnatter 2001a, p.39]. Formally, the importance density is defined as follows:

$$\begin{aligned} q(\psi) \doteq \Bigg[\frac{1}{R}\sum_{r=1}^{R} & q_\alpha(\boldsymbol{\alpha} \mid \boldsymbol{\alpha}^{[r]}, \boldsymbol{\beta}^{[r]}, \boldsymbol{\varpi}^{[r]}, \nu^{[r]}, \mathbf{s}^{[r]}, \mathbf{y}) \\ & \times q_\beta(\boldsymbol{\beta} \mid \boldsymbol{\alpha}^{[r]}, \boldsymbol{\beta}^{[r]}, \boldsymbol{\varpi}^{[r]}, \nu^{[r]}, \mathbf{s}^{[r]}, \mathbf{y}) \times p(P \mid \mathbf{s}^{[r]})\Bigg] \times q_\nu(\nu) \end{aligned} \tag{7.24}$$

where:

$$\boldsymbol{\alpha}^{[r]} \quad , \quad \boldsymbol{\beta}^{[r]} \quad , \quad \boldsymbol{\varpi}^{[r]} \quad , \quad \nu^{[r]} \quad , \quad \mathbf{s}^{[r]} \quad \text{for } r = 1, \ldots, R$$

are draws from the unconstrained posterior sample, $q_\alpha(\boldsymbol{\alpha} \mid \bullet)$ is the proposal density for parameter $\boldsymbol{\alpha}$ given in (7.13), $q_\beta(\boldsymbol{\beta} \mid \bullet)$ is the proposal density for parameter $\boldsymbol{\beta}$ given in (7.15) (the normalizing constants are easily obtained as the proposals are truncated multivariate Normal densities), $p(P \mid \bullet)$ is the product of Dirichlet posterior densities for the transition probabilities given in (7.9). For the degrees of freedom parameter ν, the optimized rejection technique of **Sect. 7.2.5** does not lead to a known expression for the marginal posterior on ν. To tackle this problem, we approximate the marginal posterior by using a truncated skewed Student-t density whose parameters are estimated by Maximum Likelihood from the posterior sample $\{\nu^{[j]}\}_{j=1}^{J}$. More precisely, the approximation may be written as follows:

$$q_\nu(\nu) \propto \mathcal{SS}(\nu \mid \widehat{\mu}, \widehat{\sigma}^2, \widehat{\tau}, \widehat{\gamma}) \mathbb{I}_{\{\nu > \delta\}}$$

where:

$$\begin{aligned} \mathcal{SS}(\nu \mid \mu, \sigma^2, \tau, \gamma) \doteq & \frac{2}{\gamma + \frac{1}{\gamma}} \frac{\Gamma\left(\frac{\tau+1}{2}\right)}{\Gamma\left(\frac{\tau}{2}\right) (\pi\tau\sigma^2)^{1/2}} \\ & \times \left[1 + \frac{(\nu-\mu)^2}{\tau\sigma^2} \left\{ \frac{1}{\gamma} \mathbb{I}_{\{\nu-\mu \geqslant 0\}} + \gamma^2 \mathbb{I}_{\{-\infty < \nu-\mu\}} \right\}\right]^{-\frac{\tau+1}{2}} \end{aligned} \tag{7.25}$$

is the skewed Student-t density as defined in Fernández and Steel [1998, Eq.13, p.363]. The parameters of the density defined in (7.25) are: the location parameter μ, the scale factor $\sigma^2 > 0$, the degrees of freedom parameter $\tau \geqslant 1$ and the asymmetry coefficient $\gamma > 0$. For $\gamma = 1$, the density coincides with the symmetric Student-t density. In cases where $\gamma > 1$, the density is right-skewed while it is left-skewed when $\gamma < 1$. Therefore, parametrization (7.25) allows for a wide range of asymmetric and heavy-tailed densities. Moreover, the normalizing constant for $q_\nu(\nu)$ is easily obtained by conventional quadrature methods.

Some comments are in order here. First, the generation of draws from the proposal densities $q_\alpha(\boldsymbol{\alpha} \mid \bullet)$ and $q_\beta(\boldsymbol{\beta} \mid \bullet)$ is achieved by the rejection technique. While we obtain good acceptance rates in our case, this method can become very inefficient if the mass of the density is close to the domain of truncation. For these cases, we would need a more sophisticated algorithm, as proposed in Philippe and Robert [2003], Robert [1995], to draw efficiently from a truncated

multivariate Normal distribution. Second, the density $q_\nu(\nu)$ is constructed in two steps. The parameters of the the skewed Student-t are first estimated by ML from the MCMC output and then the density is truncated to construct $q_\nu(\nu)$. An alternative approach would be to fit directly the truncated skewed Student-t density by ML. This is however not necessary in our case since the mass of the posterior on the degrees of freedom is far from the truncation domain. Finally, generating draws from $q_\nu(\nu)$ is achieved by the rejection technique. In cases where the boundary is close to the high probability mass, alternative approaches, such as the inversion technique, are required [see, *e.g.*, Geweke 1991].

As indicated previously, the parameters of the skewed Student-t density are estimated by ML using the posterior sample of ν. In the case of the MS-GJR model, we obtain the following ML estimates:

$$\widehat{\mu} = 9.49 \quad , \quad \widehat{\sigma}^2 = 1.50 \quad , \quad \widehat{\tau} = 16.67 \quad \text{and} \quad \widehat{\gamma} = 1.53 \; .$$

In the upper part of **Fig. 7.6**, we display the fitted truncated skewed Student-t density (in dashed line) together with the density of the posterior sample for ν (in solid line) obtained through Gaussian kernel density estimates [see Silverman 1986]. We can notice that the truncated skewed Student-t density approximates the marginal closely. In the lower part of the figure, we show the marginal posterior for parameter β^1 together with the importance density computed with $R = 1'000$. As the construction of the mixture (7.24) is based on averaging over proposal densities, where the state process is sampled from the unconstrained posterior with balanced label switching, the mixture importance density is multimodal. We also notice that the importance density provides a good approximation of the marginal posterior.

In **Table 7.3**, we report the natural logarithm of the model likelihoods obtained using the reciprocal sampling estimator (second column) and the bridge sampling estimator (last column) for $M = L = 1'000$ draws. From this table, we can notice that both estimators are higher for the MS-GJR model, indicating a better in-sample fit for the regime-switching specification. As an additional discrimination criterion, we compute the (transformed) Bayes factor in favor of the MS-GJR model [see Kass and Raftery 1995, Sect.3.2]. The estimated value is $2 \times \ln \mathrm{BF} = 2 \times (-3389.66 - (-3408.04)) = 36.76$, which strongly supports the in-sample evidence in favor of the regime-switching model.

A final word about the robustness of these results is in order. It is indeed recognized that the model likelihood is sensitive to the choice of the prior density. We must therefore test whether an alternative joint prior specification would have modified the conclusion of our analysis. To answer this question, we modify the hyperparameters' values and run the sampler again. This time, we consider

Table 7.3. Results of the model likelihood estimators.*

Model	$\ln p_0(\mathbf{y})$	$\ln p(\mathbf{y})$
GJR	-3405.33 (2.979)	-3408.04 (2.644)
MS-GJR	-3386.14 (3.109)	-3389.66 (3.191)

* $\ln p_0(\mathbf{y})$: natural logarithm of the model likelihood estimate using reciprocal sampling; $\ln p(\mathbf{y})$: natural logarithm of the model likelihood estimate using bridge sampling; ($\bullet$) numerical standard error of the estimators ($\times 10^2$).

slightly more informative priors for the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by choosing diagonal covariance matrices whose variances are set to $\sigma^2_{\alpha_i} = \sigma^2_\beta = 1'000$ ($i = 0, 1, 2$). As an alternative prior on the degrees of freedom parameter, we choose $\lambda = 0.02$ and $\delta = 2$, which implies a prior mean of 52. Finally, the hyperparameters for the prior on the transition probabilities are set to $\eta_{ii} = 3$ and $\eta_{ij} = \eta_{ji} = 1$ for $i, j \in \{1, 2\}$. We recall that the hyperparameters of the initial joint prior were set to $\sigma^2_{\alpha_i} = \sigma^2_\beta = 10'000$, $\lambda = 0.01$, $\delta = 2$, $\eta_{ii} = 2$ and $\eta_{ij} = \eta_{ji} = 1$. In this case, the results are similar to those obtained previously. The natural logarithm of the bridge sampling estimator is -3402.11 for the GJR model and -3388.09 for the MS-GJR model, implying a (transformed) Bayes factor of 28.04. These results are in line with the conclusion of the previous section and confirm the better fit of the Markov-switching model.

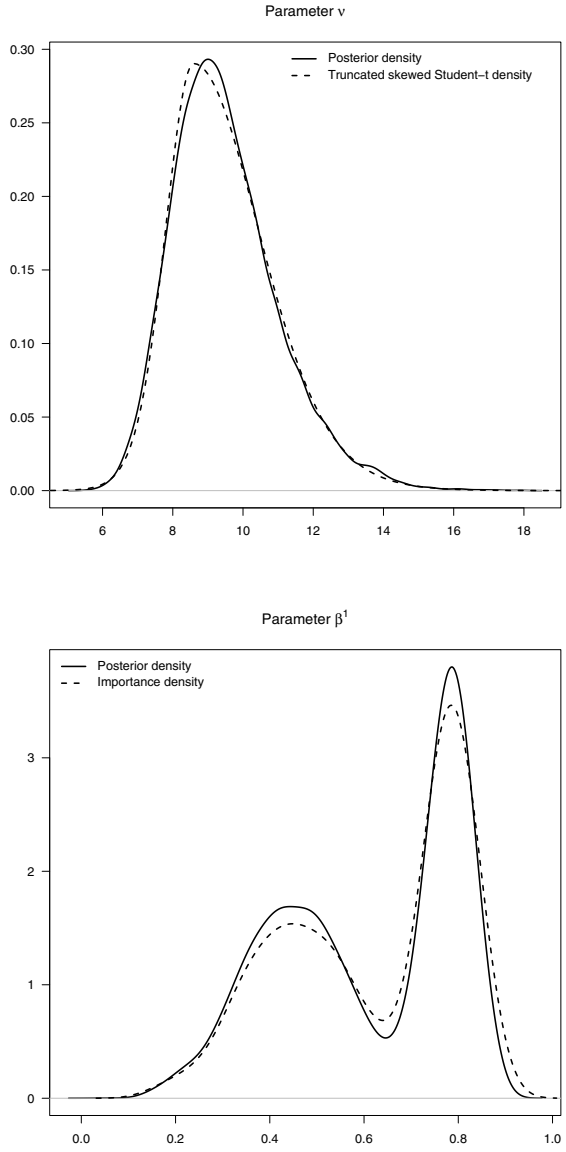

Fig. 7.6. Importance density (in dashed line) and marginal posterior density (in solid line) comparison. Gaussian kernel density estimates with bandwidth selected by the "Silverman's rule of thumb" criterion [see Silverman 1986, p.48]. Both graphs are based on 10'000 draws from the unconstrained posterior sample.

7.5 Forecasting performance analysis

In order to evaluate the ability of the competing models to predict the future behavior of the volatility process, we study the forecasted one-day ahead Value at Risk (henceforth VaR), which is a common tool to measure financial and market risks. The one-day ahead VaR at risk level $\phi \in (0, 1)$, denoted by VaR^{ϕ}, is estimated by calculating the ϕ^cth percentile of the one-day ahead predictive distribution, where $\phi^c \doteq (1 - \phi)$ for convenience. The predictive density is obtained by simulation from the joint posterior sample $\{\psi^{[j]}\}_{j=1}^{J}$ as follows:

$$
\begin{aligned}
s_{t+1}^{[j]} &\sim p(s_{t+1} \mid \psi^{[j]}, \mathcal{F}_t) \\
y_{t+1}^{[j]} &\sim p(y_{t+1} \mid \psi^{[j]}, s_{t+1}^{[j]}, \mathcal{F}_t)
\end{aligned}
$$

and VaR^{ϕ} is then simply estimated by calculating the ϕ^cth percentile of the empirical distribution $\{y_{t+1}^{[j]}\}_{j=1}^{J}$.

In order to simulate from the predictive density over the out-of-sample observation window, the posterior sample $\{\psi^{[j]}\}_{j=1}^{J}$ should be updated using the most recent information. Consequently, forecasting the one-day ahead VaR would necessitate the estimation of the joint posterior sample at each time point in the out-of-sample observation window. However, such an approach is computationally impractical for a large data set such as ours. Combination of MCMC and importance sampling to estimate efficiently this predictive density is proposed by Gerlach, Carter, and Kohn [1999]. Nevertheless, for the sake of simplicity, we will consider the same joint posterior sample, based on the in-sample data set, when forecasting the VaR.

In addition to the static GJR and MS-GJR models, we consider a GJR model estimated on rolling windows which is the standard practice in financial risk management. This methodology relies on the assumption that older data are not available or are irrelevant due to structural breaks, which are so complicated that they cannot be modeled. We refer the reader to **Sect. 6.4.1** for a detailed presentation of this procedure. For this approach, we use 750 log-returns to estimate the model and the next 50 log-returns are used as a forecasting window. Then, the estimation and forecasting windows are moved together by 50 days ahead, so that the forecasting windows do not overlap. In this manner, the estimation methodology fulfills the recommendations of the Basel Committee in the use of internal models [see Basel Committee on Banking Supervision 1996b]. When applied to our data set, this estimation design leads to the generation of 26 estimation windows for a total of $26 \times 50 = 1$'300 out-of-sample observations. In the case of the static GJR and MS-GJR models, the first 2'500 observations of our data set are used to estimate the models while the remaining 1'300 obser-

vations are used to test their predictive performance. For the three models, the VaR predictions are obtained for the same 1'300 out-of-sample daily log-returns.

To verify the accuracy of the VaR estimates for the analyzed models, we adopt the testing methodology proposed by Christoffersen [1998]. This approach is based on the study of the random sequence $\{V_t^\phi\}$ where:

$$V_t^\phi \doteq \begin{cases} 1 & \text{if} \quad y_{t+1} < \text{VaR}_t^\phi \\ 0 & \text{else} \ . \end{cases}$$

A sequence of VaR forecasts at risk level ϕ has correct conditional coverage if $\{V_t^\phi\}$ is an independent and identically distributed sequence of Bernoulli random variables with parameter ϕ^c. This hypothesis can be verified by testing jointly the independence on the series and the unconditional coverage of the VaR forecasts, *i.e.*, $\mathbb{E}(V_t^\phi) = \phi^c$, as proposed by Christoffersen [1998].

Forecasting results for the VaR are reported in **Table 7.4** for $\phi \in \{0.90, 0.95, 0.99\}$ which are typical risk levels used in financial risk management. The second and third columns give the expected and observed number of violations. The last three columns report the p-values for the tests of correct unconditional coverage (UC), independence (IND) and correct conditional coverage (CC). From this table, we first note that the observed number of violations for the MS-GJR model are closer to the expected values than for the static GJR model. Indeed, at the 1% significance level, the test of correct unconditional coverage is not rejected for the Markov-switching model while it is strongly rejected for the GJR model at risk level $\phi = 0.95$. The test of independence is not rejected for both models at the 1% significance level. We can notice that for risk level $\phi = 0.99$ this test is not applicable since no consecutive violations have been observed. The joint hypothesis of correct unconditional coverage and independent sequence is obtained via the test of correct conditional coverage. In the case of the MS-GJR model, p-values are close to 0.10 for risk levels $\phi = 0.9$ and $\phi = 0.95$ while it is 0.030 and 0.013 in the GJR case. We therefore reject the correct conditional coverage hypothesis for the static GJR model at the 5% significance level. These results indicate the better out-of-sample performance of the Markov-switching model compared to the static GJR model.

When comparing the MS-GJR model with the rolling GJR model, we can notice that both approaches perform equally well. Indeed, for both models, the test of independence is rejected at risk level $\phi = 0.90$ while the correct conditional coverage hypothesis is not rejected at the 5% significance level. Although the two models are successful in forecasting the conditional variance of the SMI log-returns, the MS-GJR model has two advantages over the rolling window

Table 7.4. Forecasting results of the VaR.[★]

GJR model (static approach)					
ϕ	$\mathbb{E}(V_t^{\phi})$	#	UC	IND	CC
0.99	13	14	0.783	NA	NA
0.95	65	89	0.004	0.624	0.013
0.90	130	143	0.236	0.018	0.030
GJR model (rolling windows approach)					
ϕ	$\mathbb{E}(V_t^{\phi})$	#	UC	IND	CC
0.99	13	15	0.586	NA	NA
0.95	65	73	0.318	0.547	0.506
0.90	130	126	0.710	0.032	0.093
MS-GJR model (static approach)					
ϕ	$\mathbb{E}(V_t^{\phi})$	#	UC	IND	CC
0.99	13	13	1.000	NA	NA
0.95	65	80	0.065	0.323	0.112
0.90	130	132	0.854	0.035	0.107

[★] ϕ: risk level; $\mathbb{E}(V_t^{\phi})$: expected number of violations; #: observed number of violations; UC: p-value for the correct unconditional coverage test; IND: p-value for the independence test; CC: p-value for the correct conditional coverage test; NA: not applicable.

approach. First, it is able to anticipate structural breaks in the conditional variance process. This is achieved through the estimation of the filtered probabilities $\mathbb{P}(s_t = k \mid \psi, \mathcal{F}_{t-1})$, as shown in **Fig. 7.7**. On the contrary, the rolling window methodology is merely an ad-hoc approach which is unable to forecast structural breaks. The updating frequency as well as the length of the rolling window are subjective quantities, albeit some ranges are recommended by regulators, so that different choices might lead to significant differences in the model's performance. Second, the MS-GJR model needs only to be estimated once. On the contrary, the parameters of the GJR model must be updated frequently to account for structural breaks in the time series and this can have practical consequences for risk management systems of financial institutions. This is a definite advantage of the regime-switching approach compared to the traditional rolling window methodology.

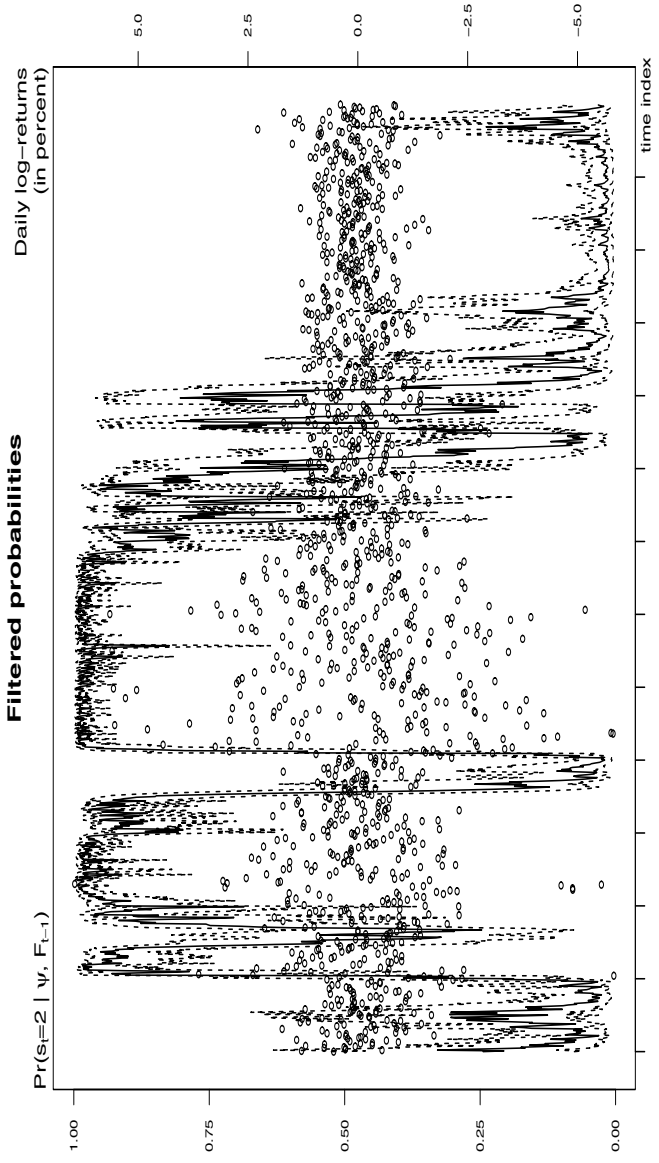

Fig. 7.7. Filtered probabilities of the high-volatility state (solid line, left axis) together with the out-of-sample log-returns (circles, right axis). The 95% confidence bands are shown in dashed lines.

7.6 One-day ahead VaR density

As emphasized in **Chap. 6**, the one-day ahead VaR risk measure can be expressed as a function of the model parameters when the underlying time series is described by a single-regime GARCH(1, 1) model. It turns out that this is also the case in the context of Markov-switching GARCH models. In effect, the one-day ahead VaR at risk level ϕ, estimated at time t, can be explicitly calculated for given ψ and future state s_{t+1} as follows:

$$\text{VaR}_t^{\phi}(\psi, s_{t+1}) \doteq \left[\varrho(\nu) \times \mathbf{e}'_{t+1}(s_{t+1})\mathbf{h}_{t+1}(\boldsymbol{\alpha}, \boldsymbol{\beta})\right]^{1/2} \times t_{\phi^c}(\nu) \tag{7.26}$$

where we recall that $\varrho(\nu) \doteq \frac{\nu-2}{\nu}$ and $t_{\phi^c}(\nu)$ denotes the ϕ^cth percentile of a Student-t distribution with ν degrees of freedom. Hence, the VaR risk measure can be simulated from the joint posterior sample $\{\psi^{[j]}\}_{j=1}^{J}$ by first generating $s_{t+1}^{[j]}$ from the filtered probability density $p(s_{t+1} \mid \psi^{[j]}, \mathcal{F}_t)$, and then inputting the joint draw $(\psi^{[j]}, s_{t+1}^{[j]})$ in expression (7.26).

The result of this procedure is shown in **Fig. 7.8** where we plot the one-day ahead VaR density of the MS-GJR model for two distinct time points in the out-of-sample observation window. We can notice that both densities are bimodal, which is a consequence of the Markov-switching nature of the conditional variance process. At time $t = 2$'501, the VaR density gives a higher probability to larger (in absolute value) VaR values. This suggests that, at that particular point in time, the probability of being in the high volatility state is higher than being in the low-volatility regime. At time $t = 3$'500, the bimodality of the density is slightly less pronounced. In this case, the VaR density puts more mass on smaller VaR values (in absolute value). This graph shows that the density of the VaR has a particular shape in the case of the MS-GJR model. In this context, it would be interesting to determine if the loss function of an agent, and therefore the location of his optimal Bayes estimate within the VaR density, would have any influence on the forecasting performance of the model.

In order to address this question, we consider different loss functions and determine the Bayes point estimates for the VaR by solving the optimization problem (6.10) of page 85. The loss functions we consider are the Linex with a parameter $a \in \{-3, 3\}$, the absolute error loss (AEL) as well as the squared error loss (SEL); the reader is referred to **Sect. 6.4.4** for further details. We recall however that the Linex function with a positive parameter could be attributed to a regulator or risk manager whose aim is to avoid systematic failure in risk measure estimation. On the contrary, a negative parameter could be attributed to a fund manager who seeks to save risk capital since it earns little or no return at all (see **Sect. 6.3.1** for details). The AEL and SEL correspond to the

perspective of an agent for whom under- and overestimation are equally serious. The SEL leads, however, to a larger penalty for larger deviations from the true value compared to the AEL function.

The VaR risk measure obtained with the different loss functions are then tested over the 1'300 out-of-sample observations. To test the adequacy of the point estimates to reproduce the true VaR, we rely on the forecasting methodology of Christoffersen [1998] as this was done in the preceding section. The results are reported in **Table 7.5** whose second column gives the observed number of violations and the third, fourth and fifth columns report the p-values for the tests of correct unconditional coverage (UC), independence (IND) and correct conditional coverage (CC), respectively. From this table, we note first that the observed number of violations is close to the expected value for the Linex function with parameter $a = 3$. In this case, the test of correct unconditional coverage, at the 5% significance level, is never rejected. On the contrary, the Linex function with parameters $a = -3$ leads to the rejection of the null for risk levels $\phi = 0.95$ and $\phi = 0.99$. The null hypothesis is also rejected for the AEL and SEL point estimates at risk level $\phi = 0.95$, where the estimates systematically underestimate (in absolute value) the true VaR. The joint hypothesis of correct unconditional coverage and independence is rejected at the 5% significance level for all functions, except the Linex with $a = 3$ and the SEL at risk level $\phi = 0.9$.

From what precedes, we can thus conclude that parameter uncertainty has to be taken seriously in the context of MS-GARCH models. In particular, the choice of a given point estimate within the VaR density has a significant impact on the forecasting performance of the model. A regulator (Linex $a = 3$) whose VaR point estimate are conservative, would conclude to a good performance of the model while a fund manager (Linex $a = -3$) would systematically underestimate (in absolute value) the true VaR.

Table 7.5. Forecasting results of the VaR point estimates for the MS-GJR model.★

$\phi = 0.90$, $\mathbb{E}(V_t^\phi) = 130$; Loss $\mathscr{L}$	#	UC	IND	CC
Linex ($a = 3$)	130	1.000	0.018	0.061
Linex ($a = -3$)	140	0.361	0.011	0.025
AEL[a]	133	0.782	0.011	0.039
SEL[b]	131	0.926	0.015	0.053
$\phi = 0.95$, $\mathbb{E}(V_t^\phi) = 65$; Loss $\mathscr{L}$	#	UC	IND	CC
Linex ($a = 3$)	71	0.452	0.270	0.410
Linex ($a = -3$)	87	0.008	0.171	0.011
AEL[a]	84	0.020	0.228	0.033
SEL[b]	83	0.028	0.249	0.046
$\phi = 0.99$, $\mathbb{E}(V_t^\phi) = 13$; Loss $\mathscr{L}$	#	UC	IND	CC
Linex ($a = 3$)	11	0.567	NA	NA
Linex ($a = -3$)	21	0.041	NA	NA
AEL[a]	17	0.287	NA	NA
SEL[b]	14	0.783	NA	NA

★ ϕ: risk level; $\mathbb{E}(V_t^\phi)$: expected number of violations; #: observed number of violations; UC: p-value for the correct uncoverage test; IND: p-value for the independence test; CC: p-value for the correct conditional coverage test; NA: not applicable.
[a] Absolute error loss function.
[b] Squared error loss function.

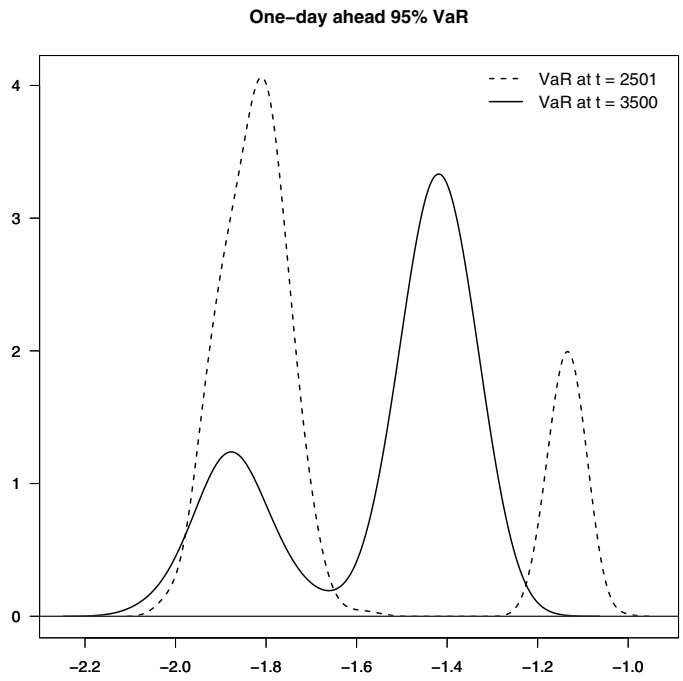

Fig. 7.8. Density of the one-day ahead VaR at risk level $\phi = 0.95$ for the MS-GJR model at two time points in the out-of-sample observation window. Gaussian kernel density estimates with bandwidth selected by the "Silverman's rule of thumb" criterion [see Silverman 1986, p.48]. Both graphs are based on 10'000 draws from the joint posterior density of the MS-GJR model parameters.

7.7 Maximum Likelihood estimation

We conclude this chapter with some comments regarding the Maximum Likelihood (henceforth ML) estimation of Markov-switching GARCH models. In this case, the estimation is handled as in Hamilton [1994, p.692], where the algorithm turns out to be a special case of the Expectation Maximization (henceforth EM) algorithm developed by Dempster, Laird, and Rubin [1977]. The classical ML approach cannot be applied directly, as the marginal likelihood where the latent process $\{s_t\}$ is integrated out, is not available in closed form. The estimation procedure is therefore decomposed into two stages. The first step consists in estimating the sequence of filtered probabilities $\{\mathbb{P}(s_t = k \mid \psi, \mathcal{F}_{t-1})\}_{t=1}^{T}$ for a fixed set of of parameters ψ. The second step maximizes the observed likelihood $\mathcal{L}(\psi \mid \mathbf{y})$ in expression (7.21) given this sequence of probabilities. The procedure is iterated until a given convergence criterion is satisfied. General results available for the EM algorithm indicate that the likelihood function increases in the number of iterations.

While apparently straightforward to handle, the ML estimation has practical drawbacks. Indeed, the EM algorithm guarantees a convergence to a local maximum of the likelihood, but not necessarily to the global optimum. As reported in Hamilton and Susmel [1994], many starting points are required to end up with a global maximum. Furthermore, the covariance matrix at the optimum can be extremely tedious to obtain and ad-hoc procedures are often required to get reliable results. *E.g.*, Hamilton and Susmel [1994] fix some transition probabilities to zero in order to determine the variance estimates for some model parameters. Finally, testing the null of K versus K' states is not possible within the ML framework since the regularity conditions for justifying the χ^2 approximation of the likelihood ratio statistic do not hold.

For comparison purposes, we estimate the MS-GJR model via the ML technique. The iterative procedure described previously has been run using 20 random starting values. In all cases, the optimizer has been trapped in a local maximum or even did not converge. The convergence has only been achieved by starting the ML optimizer at the posterior mean $\overline{\psi}$ (see **Table 7.1**, p.127) obtained with the Bayesian approach.

In **Fig. 7.9**, we display the marginal densities obtained via Gaussian kernel density estimates, for the model parameters obtained through the Bayesian approach (in solid lines) and the ML approach (in dashed lines). From these graphs, we note that the ML estimation leads to more peaked density estimates and therefore underestimates the parameter uncertainty. Furthermore, compared to the Bayesian approach, the ML approach underestimates the values of the components of vector $\boldsymbol{\alpha}$ whereas the components of $\boldsymbol{\beta}$ are overestimated.

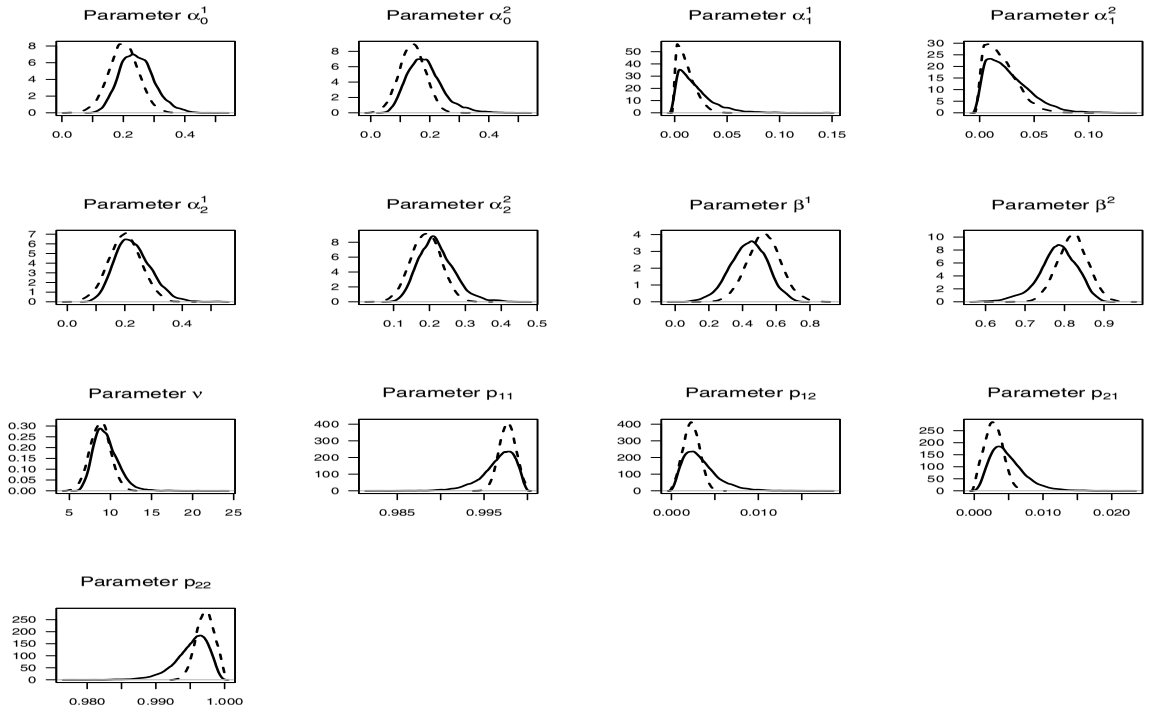

Fig. 7.9. Marginal posterior densities of the MS-GJR model parameters and comparison with the asymptotic Normal approximation. Results obtained via the Bayesian approach are given in solid lines while the ML estimates are shown in dashed lines. Gaussian kernel density estimates with bandwidth selected by the "Silverman's rule of thumb" criterion [see Silverman 1986, p.48]. The graphs are based on 10'000 draws from the constrained posterior sample.

8

Conclusion

Single-regime and regime-switching GARCH models are widespread and essential tools in financial econometrics and have, until recently, mainly been estimated using the Maximum Likelihood (henceforth ML) technique. However, the Bayesian estimation of these models has several advantages over the classical approach. First, computational methods based on Markov chain Monte Carlo (henceforth MCMC) procedures avoid the common problem of local maxima encountered in the ML estimation of these models. Second, the exploration of the joint posterior distribution gives a complete picture of the parameter uncertainty and this cannot be achieved via the classical approach. Third, exact distributions of nonlinear functions of the model parameters can be obtained at low cost by simulating from the joint posterior distribution. Fourth, constraints on the model parameters can be incorporated through appropriate prior specifications; in such a setting, imposing the constraint of covariance stationarity for the regime-switching GARCH model, for instance, is straightforward. Finally, discrimination between models can be achieved through the calculation of model likelihoods and Bayes factors. All these reasons strongly motivate the use of the Bayesian approach when estimating GARCH models.

The choice of the algorithm is the first issue when dealing with MCMC methods and it depends on the nature of the problem under study. In the case of GARCH models, due to the recursive nature of the conditional variance, the joint posterior and the full conditional densities are of unknown forms, whatever distributional assumptions are made on the model disturbances. Therefore, we cannot use the simple Gibbs sampler and need more elaborate estimation procedures. The sampling schemes adopted in this book are based on the approach of Nakatsuma [1998, 2000] which has the advantage of being fully automatic and thus avoids the time-consuming and difficult task of tuning a sampling algorithm. In addition, this approach is easy to extend to regime-switching GARCH

models. In this case, the parameters in each regime can be regrouped and updated by blocks which may enhance the sampler's efficiency.

This book presented in detail methodologies for the Bayesian estimation of single-regime and regime-switching GARCH models. It proposed empirical applications to real data sets and illustrated some interesting probabilistic statements on nonlinear functions of the model parameters made possible under the Bayesian framework.

The work was introduced in **Chap. 1** with a review of GARCH modeling and a presentation of the advantages of the Bayesian approach compared to the traditional ML technique. In **Chap. 2**, we proposed a short introduction to the Bayesian paradigm for inference and gave an overview of the basic MCMC algorithms used in the rest of the book.

In **Chap. 3**, we considered the Bayesian estimation of the parsimonious but effective GARCH$(1, 1)$ model with Normal innovations. We detailed the MCMC scheme based on the methodology of Nakatsuma [1998, 2000]. An empirical application to a foreign exchange rate time series was presented where we compared the Bayesian and the ML point estimates. In particular, we showed that even for a fairly large data set, the estimates and confidence intervals are different between the methods. Caution is therefore in order when applying the asymptotic Normal approximation for the model parameters in this case. We performed a sensitivity analysis to check the robustness of our results with respect to the choice of the priors and tested the residuals for misspecification. Finally, we compared the theoretical and sample autocorrelograms of the process and tested the covariance and strict stationarity conditions.

In **Chap. 4**, we analyzed the linear regression model with conditionally heteroscedastic errors which allowed the introduction of lagged dependent variables in the modeling; moreover, we considered the GJR$(1, 1)$ model to account for asymmetric responses to past shocks in the conditional error variance process. We fitted the model to the Standard and Poors 100 (henceforth S&P100) index log-returns and compared the Bayesian and the ML estimations. We performed a prior sensitivity analysis and tested the residuals for misspecification. Finally, we tested the covariance stationarity condition and illustrated the differences between the unconditional variance of the process obtained through the Bayesian approach and the delta method. In particular, we showed that the Bayesian framework leads to a more precise estimate.

In **Chap. 5**, we extended the linear regression model further with the introduction of Student-t-GJR$(1, 1)$ errors. An empirical application based on the S&P100 index log-returns was proposed with a comparison between the joint posterior and the asymptotic Normal approximation of the parameter estimates.

We performed a prior sensitivity analysis and tested the residuals for misspecification. Finally, we analyzed the conditional and unconditional kurtosis of the underlying time series.

In **Chap. 6**, we presented some financial applications of the Bayesian estimation of GARCH models. We introduced the concept of Value at Risk (henceforth VaR) risk measure and proposed a methodology to estimate the density of this quantity for different risk levels and time horizons. We reviewed some basics in decision theory and used this framework as a rational justification for choosing a point estimate of the VaR. We showed how agents facing different risk perspectives could select their optimal VaR point estimate and documented substantial differences in terms of regulatory capital between individuals. Finally, we extended our methodology to the Expected Shortfall (henceforth ES) risk measure.

In **Chap. 7**, we extended the single-regime GJR model to the regime-switching GJR model (henceforth MS-GJR); more precisely, we considered an asymmetric version of the Markov-switching GARCH$(1,1)$ specification of Haas *et al.* [2004]. We introduced a novel MCMC scheme which can be viewed as a multivariate extension of the sampler proposed by Nakatsuma [1998, 2000]. As an application, we fitted a single-regime and a Markov-switching GJR model to the Swiss Market Index log-returns. We used the random permutation sampler of Frühwirth-Schnatter [2001b] to find suitable identification constraints for the MS-GJR model and showed the presence of two distinct volatility regimes in the time series. By using the Deviance information criterion of Spiegelhalter *et al.* [2002] and by estimating the model likelihood using the bridge sampling technique of Meng and Wong [1996], we showed the in-sample superiority of the MS-GJR model. To test the predictive performance of the models, we ran a forecasting performance analysis based on the VaR. In particular, we compared the MS-GJR model to a single-regime GJR model estimated on rolling windows and concluded to the superiority of the MS-GJR specification. Finally, we proposed a methodology to depict the density of the one-day ahead VaR and presented a comparison with the traditional ML approach.

This book proposed two main contributions which are of practical relevance for both market participants and academics. First, we proposed a novel MCMC scheme to perform the Bayesian estimation of a Markov-switching model with Student-t innovations and asymmetric GJR specifications for the conditional variance in each regime. It allows to reproduce many stylized facts observed in financial time series, such as volatility clustering, conditional leptokurticity and Markov-switching dynamics. Furthermore, it helps to identify whether the leverage effect is different across regimes. Our multivariate extension of the ap-

proach proposed by Nakatsuma [1998, 2000] leads to a fast, fully automatic and efficient estimation procedure compared to alternative approaches such as the Griddy-Gibbs sampler. Practitioners who need to run the estimation frequently and/or for a large number of time series should find the procedure helpful. Second, we provided a manner to approximate the multi-day ahead VaR and ES densities when the underlying process is described by a GARCH model. Our methodology gives the possibility to determine the term structure of these risk measures and to characterize the uncertainty coming from the model parameters. In our empirical application, we documented that the choice of the model disturbances has a significant impact on the shape of both risk measures' densities and this effect gets larger as the time horizon increases. Moreover, the densities are strongly left-skewed, which implies substantial differences in risk capital allocation for agents facing different risk perspectives (*e.g.*, risk and fund managers).

Suggestions for further work

This study has raised many questions and suggests interesting further avenues of research.

First, in light of the results obtained in **Chap. 6**, additional work is required to assess the performance of multi-day ahead VaR models. This is essential for risk management purposes since the multi-day ahead VaR lies at the heart of the risk capital allocation's framework. The development of powerful methodologies for testing the VaR is the subject of current researches and we refer the reader to Berkowitz, Christoffersen, and Pelletier [2006], Kaufmann [2004], Seiler [2006], Zumbach [2006] for details. A natural extension of our analyses would consider the model uncertainty in addition to the parameter uncertainty. The Bayesian approach provides a natural framework for tackling this issue.

Second, regime-switching GARCH models might be compared to the class of stochastic volatility (henceforth SV) models [see, *e.g.*, Jacquier, Polson, and Rossi 1994, Kim *et al.* 1998]. While SV models are highly flexible (two different processes drive the dynamics of the underlying time series and the dynamics of the volatility), they are more difficult to estimate efficiently. Determining whether this additional flexibility results in a superior predictive ability would therefore be of interest.

Finally, in our study of the Markov-switching GJR model, we have considered a fixed transition matrix for the state process. Consequently, the expected persistence of the regimes is constant over time, which is questionable. In a more general formulation, we could allow the transition probabilities to change

over time depending on some observables [see, *e.g.*, Bauwens *et al.* 2006, Gray 1996]. This would allow determining whether some exogenous variables trigger the switching mechanism of the volatility process. The transition probabilities could also depend on the past level of the volatility. In this case, we could reproduce an additional feature of the volatility behavior, namely the fact that the probability of returning to a normal (*i.e.*, low or medium) volatility regime increases after a high upward jump in the volatility level [see, *e.g.*, Bauwens *et al.* 2006, Dueker 1997].

A

Recursive Transformations

In this appendix, we demonstrate the recursive transformations introduced in **Chaps. 3**, **4** and **5** which are used to express the function $z_t(\boldsymbol{\alpha})$ as a linear function of parameter $\boldsymbol{\alpha}$. The process for the conditional variance is based on observations $\{y_t\}$. Results can be straightforwardly extended when a linear component is included in the model by considering instead the process $\{u_t\}$ where $u_t \doteq y_t - \mathbf{x}'\boldsymbol{\gamma}$.

A.1 The GARCH(1, 1) model with Normal innovations

First, let us recall that, in the case of the GARCH$(1,1)$ process, the expression for the conditional variance of y_t is given by:

$$h_t \doteq \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1}$$

where $h_0 = y_0 \doteq 0$ for convenience. As shown in **Sect. 3.2**, the GARCH$(1,1)$ model with Normal innovations can be expressed as an ARMA$(1,1)$ model for the squared observations $\{y_t^2\}$ and approximated as follows:

$$y_t^2 = \alpha_0 + (\alpha_1 + \beta) y_{t-1}^2 - \beta z_{t-1} + z_t$$

where $\{z_t\}$ is a Martingale Difference process. Let us define $v_t \doteq y_t^2$ for notational purposes. The variable z_t can then be written as:

$$z_t = v_t - \alpha_0 - (\alpha_1 + \beta) v_{t-1} + \beta z_{t-1} \tag{A.1}$$

where $v_0 = z_0 = 0$.

Proposition A.1. *Upon defining the following recursive transformations:*

$$\begin{aligned} l_t^* &\doteq 1 + \beta\, l_{t-1}^* \\ v_t^* &\doteq v_{t-1} + \beta\, v_{t-1}^* \end{aligned} \tag{A.2}$$

where $l_0^* = v_0^* \doteq 0$, *expression* (A.1) *can be written as follows:*

$$z_t = v_t - (l_t^* \; v_t^*)\boldsymbol{\alpha} \tag{A.3}$$

where $\boldsymbol{\alpha} \doteq (\alpha_0 \; \alpha_1)'$. *The function* $z_t(\boldsymbol{\alpha})$ *in* (A.1) *can therefore be expressed as a linear function of the* 2×1 *vector* $\boldsymbol{\alpha}$.

Proof. By induction:

Beginning step:
For $t = 1$, we have:

$$v_1 - (l_1^* \; v_1^*)\boldsymbol{\alpha} \overset{(A.2)}{=} v_1 - (1 \; 0)\boldsymbol{\alpha} = v_1 - \alpha_0 \overset{(A.1)}{=} z_1 \; .$$

Assumption step:
Let us assume that expression (A.3) is satisfied for $t = k$.

Induction step:
For $t = k + 1$ we have:

$$\begin{aligned} v_{k+1} - (l_{k+1}^* \; v_{k+1}^*)\boldsymbol{\alpha} &\overset{(A.2)}{=} v_{k+1} - (1 + \beta l_k^* \quad v_k + \beta v_k^*)\boldsymbol{\alpha} \\ &= v_{k+1} - \alpha_0 - \alpha_1 v_k - \beta(\alpha_0 l_k^* + \alpha_1 v_k^*) \\ &= v_{k+1} - \alpha_0 - \alpha_1 v_k - \beta(l_k^* \; v_k^*)\boldsymbol{\alpha} \\ &\overset{(A.3)}{=} v_{k+1} - \alpha_0 - \alpha_1 v_k - \beta(v_k - z_k) \\ &= v_{k+1} - \alpha_0 - (\alpha_1 + \beta)v_k + \beta z_k \\ &\overset{(A.1)}{=} z_{k+1} \; . \end{aligned}$$

□

A.2 The GJR(1, 1) model with Normal innovations

First, let us recall that, in the case of the GJR(1, 1) model, the expression for the conditional variance of y_t is given by:

$$h_t \doteq \alpha_0 + (\alpha_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{y_{t-1} < 0\}}) y_{t-1}^2 + \beta h_{t-1}$$

where $h_0 = y_0 \doteq 0$ for convenience. As shown in **Sect. 4.2.2**, the GJR(1, 1) model with Normal innovations can be transformed for the squared observations $\{y_t^2\}$ and approximated as follows:

$$y_t^2 = \alpha_0 + (\alpha_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{y_{t-1} < 0\}} + \beta) y_{t-1}^2 - \beta z_{t-1} + z_t$$

where $\{z_t\}$ is a Martingale Difference process. Let us define $v_t \doteq y_t^2$ for notational purposes. Then, the variable z_t can be written as:

$$z_t = v_t - \alpha_0 - (\alpha_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{y_{t-1} < 0\}} + \beta) v_{t-1} + \beta z_{t-1} \tag{A.4}$$

where $v_0 = z_0 = 0$.

Proposition A.2. *Upon defining the following recursive transformations:*

$$\begin{aligned} l_t^* &\doteq 1 + \beta l_{t-1}^* \\ v_t^* &\doteq v_{t-1}\mathbb{I}_{\{y_{t-1}\geqslant 0\}} + \beta\, v_{t-1}^* \\ v_t^{**} &\doteq v_{t-1}\mathbb{I}_{\{y_{t-1}<0\}} + \beta\, v_{t-1}^{**} \end{aligned} \tag{A.5}$$

where $l_0^* = v_0^* = v_0^{**} \doteq 0$*, expression* (A.4) *can be written as follows:*

$$z_t = v_t - (l_t^*\ v_t^*\ v_t^{**})\boldsymbol{\alpha} \tag{A.6}$$

where $\boldsymbol{\alpha} \doteq (\alpha_0\ \alpha_1\ \alpha_2)'$*. The function* $z_t(\boldsymbol{\alpha})$ *in* (A.4) *can therefore be expressed as a linear function of the* 3×1 *vector* $\boldsymbol{\alpha}$*.*

Proof. By induction:

Beginning step:
For $t = 1$, we have:

$$v_1 - (l_1^*\ v_1^*\ v_1^{**})\boldsymbol{\alpha} \overset{(A.5)}{=} v_1 - (1\ 0\ 0)\boldsymbol{\alpha} = v_1 - \alpha_0 \overset{(A.4)}{=} z_1\ .$$

Assumption step:
Let us assume that expression (A.6) is satisfied for $t = k$.

Induction step:
For $t = k + 1$, we have:

$$\begin{aligned} & v_{k+1} - (l_{k+1}^*\ v_{k+1}^*\ v_{k+1}^{**})\boldsymbol{\alpha} \\ &\overset{(A.5)}{=} v_{k+1} - (1 + \beta l_k^* \quad v_k\mathbb{I}_{\{y_k\geqslant 0\}} + \beta v_k^* \quad v_k\mathbb{I}_{\{y_k<0\}} + \beta v_k^{**})\boldsymbol{\alpha} \\ &= v_{k+1} - \alpha_0 - \alpha_1 v_k\mathbb{I}_{\{y_k\geqslant 0\}} - \alpha_2 v_k\mathbb{I}_{\{y_k<0\}} \\ &\quad -\beta(\alpha_0 l_k^* + \alpha_1 v_k^* + \alpha_2 v_k^{**}) \\ &= v_{k+1} - \alpha_0 - \alpha_1 v_k\mathbb{I}_{\{y_k\geqslant 0\}} - \alpha_2 v_k\mathbb{I}_{\{y_k<0\}} - \beta(l_k^*\ v_k^*\ v_k^{**})\boldsymbol{\alpha} \\ &\overset{(A.6)}{=} v_{k+1} - \alpha_0 - \alpha_1 v_k\mathbb{I}_{\{y_k\geqslant 0\}} - \alpha_2 v_k\mathbb{I}_{\{y_k<0\}} - \beta(v_k - z_k) \\ &= v_{k+1} - \alpha_0 - (\alpha_1\mathbb{I}_{\{y_k\geqslant 0\}} + \alpha_2\mathbb{I}_{\{y_k<0\}} + \beta)v_k + \beta z_k \\ &\overset{(A.4)}{=} z_{k+1}\ . \end{aligned}$$

□

A.3 The GJR(1, 1) model with Student-t innovations

As shown in **Sect. 5.2.2**, the GJR(1, 1) model with Student-t innovations, denoted by $\{y_t\}$, can be transformed in a new sequence $\{v_t\}$, where $v_t \doteq \frac{y_t^2}{\tau_t}$ with $\tau_t \doteq \varpi_t\varrho$ and $\varrho \doteq \frac{\nu-2}{\nu}$. The process $\{v_t\}$ can then be approximated as follows:

$$v_t = \alpha_0 + (\alpha_1\mathbb{I}_{\{y_{t-1}\geqslant 0\}} + \alpha_2\mathbb{I}_{\{y_{t-1}<0\}})\tau_{t-1}v_{t-1} + \beta v_{t-1} - \beta z_{t-1} + z_t$$

where $\{z_t\}$ is a Martingale Difference process. From this expression, the variable z_t can be written as:

$$z_t = v_t - \alpha_0 - \big[(\alpha_1 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{y_{t-1} < 0\}})\tau_{t-1} + \beta\big] v_{t-1} + \beta z_{t-1} \tag{A.7}$$

where $v_0 = z_0 = 0$.

Proposition A.3. *Upon defining the following recursive transformations:*

$$\begin{aligned} l_t^* &\doteq 1 + \beta l_{t-1}^* \\ v_t^* &\doteq y_{t-1}^2 \mathbb{I}_{\{y_{t-1} \geqslant 0\}} + \beta\, v_{t-1}^* \\ v_t^{**} &\doteq y_{t-1}^2 \mathbb{I}_{\{y_{t-1} < 0\}} + \beta\, v_{t-1}^{**} \end{aligned} \tag{A.8}$$

where $l_0^* = v_0^* = v_0^{**} \doteq 0$*, expression* (A.7) *can be written as follows:*

$$z_t = v_t - (l_t^*\ v_t^*\ v_t^{**})\boldsymbol{\alpha} \tag{A.9}$$

where $\boldsymbol{\alpha} \doteq (\alpha_0\ \alpha_1\ \alpha_2)'$*. The function* $z_t(\boldsymbol{\alpha})$ *in* (A.7) *can therefore be expressed as a linear function of the* 3×1 *vector* $\boldsymbol{\alpha}$*.*

Proof. By induction:

Beginning step:
For $t = 1$, we have:

$$v_1 - (l_1^*\ v_1^*\ v_1^{**})\boldsymbol{\alpha} \overset{(A.8)}{=} v_1 - (1\ 0\ 0)\boldsymbol{\alpha} = v_1 - \alpha_0 \overset{(A.7)}{=} z_1\ .$$

Assumption step:
Let us assume that expression (A.9) is satisfied for $t = k$.

Induction step:
For $t = k + 1$, we have:

$$\begin{aligned} & v_{k+1} - (l_{k+1}^*\ v_{k+1}^*\ v_{k+1}^{**})\boldsymbol{\alpha} \\ &\overset{(A.8)}{=} v_{k+1} - (1 + \beta l_k^* \quad y_k^2 \mathbb{I}_{\{y_k \geqslant 0\}} + \beta v_k^* \quad y_k^2 \mathbb{I}_{\{y_k < 0\}} + \beta v_k^{**})\boldsymbol{\alpha} \\ &= v_{k+1} - \alpha_0 - \alpha_1 y_k^2 \mathbb{I}_{\{y_k \geqslant 0\}} - \alpha_2 y_k^2 \mathbb{I}_{\{y_k < 0\}} \\ &\quad -\beta(\alpha_0 l_k^* + \alpha_1 v_k^* + \alpha_2 v_k^{**}) \\ &= v_{k+1} - \alpha_0 - \alpha_1 y_k^2 \mathbb{I}_{\{y_k \geqslant 0\}} - \alpha_2 y_k^2 \mathbb{I}_{\{y_k < 0\}} - \beta(l_k^*\ v_k^*\ v_k^{**})\boldsymbol{\alpha} \\ &\overset{(A.9)}{=} v_{k+1} - \alpha_0 - \alpha_1 y_k^2 \mathbb{I}_{\{y_k \geqslant 0\}} - \alpha_2 y_k^2 \mathbb{I}_{\{y_k < 0\}} - \beta(v_k - z_k) \\ &= v_{k+1} - \alpha_0 - \alpha_1 y_k^2 \mathbb{I}_{\{y_k \geqslant 0\}} - \alpha_2 y_k^2 \mathbb{I}_{\{y_k < 0\}} - \beta v_k + \beta z_k \\ &= v_{k+1} - \alpha_0 - \big[(\alpha_1 \mathbb{I}_{\{y_k \geqslant 0\}} + \alpha_2 \mathbb{I}_{\{y_k < 0\}})\tau_k + \beta\big] v_k + \beta z_k \\ &\overset{(A.7)}{=} z_{k+1}\ . \end{aligned}$$

□

B

Equivalent Specification

In this appendix, we demonstrate the equivalent specification introduced at the end of **Sect. 5.2**.

Proposition B.1. *The following model:*

$$\begin{aligned} &y_t = \varepsilon_t(\varpi_t h_t)^{1/2} \quad \textit{for } t = 1, \ldots, T \\ &\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1) \\ &\varpi_t \overset{iid}{\sim} \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu-2}{2}\right) \end{aligned} \tag{B.1}$$

where $h_t \doteq h_t(\boldsymbol{\alpha}, \beta)$ is a GARCH scedastic function, is equivalent to:

$$\begin{aligned} &y_t = \varepsilon_t h_t^{1/2} \quad \textit{for } t = 1, \ldots, T \\ &\varepsilon_t \overset{iid}{\sim} \mathcal{S}^*(0,1) \end{aligned}$$

where $\mathcal{S}^*(0,1)$ denotes the standardized Student-t density, i.e., its variance is one.

Proof. First, let us regroup the model parameters into $\psi \doteq (\boldsymbol{\alpha}, \beta, \nu)$ for notational purposes. In specification (B.1), the variables ϖ_t are independent and identically distributed from an Inverted Gamma density with parameters $\frac{\nu}{2}$ and $\frac{\nu-2}{2}$:

$$p(\varpi_t \mid \nu) = \left(\frac{\nu-2}{2}\right)^{\frac{\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} \varpi_t^{-\frac{\nu}{2}-1} \exp\left[-\frac{(\nu-2)}{2\varpi_t}\right]$$

and the joint density of the $T \times 1$ vector $\boldsymbol{\varpi} \doteq (\varpi_1 \cdots \varpi_T)'$ is therefore given by:

$$p(\boldsymbol{\varpi} \mid \nu) = \left(\frac{\nu-2}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \prod_{t=1}^{T} \varpi_t^{-\frac{\nu}{2}-1} \exp\left[-\frac{(\nu-2)}{2\varpi_t}\right] . \tag{B.2}$$

Based on the $T \times 1$ vector of observations $\mathbf{y} \doteq (y_1 \cdots y_T)'$, we can express the likelihood function of $(\psi, \boldsymbol{\varpi})$ as follows:

$$\mathcal{L}(\psi, \boldsymbol{\varpi} \mid \mathbf{y}) \propto \prod_{t=1}^{T} (\varpi_t h_t)^{-1/2} \exp\left[-\tfrac{1}{2}\frac{y_t^2}{\varpi_t h_t}\right]$$

and, by using the Bayes theorem, we obtain the following joint posterior:

$$\begin{aligned} p(\psi, \boldsymbol{\varpi} \mid \mathbf{y}) \propto & \left(\frac{\nu-2}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \left(\prod_{t=1}^{T} h_t^{-1/2}\right) \\ & \times \prod_{t=1}^{T} \omega_t^{-\frac{(\nu+3)}{2}} \exp\left[-\tfrac{1}{2}\frac{y_t^2}{\omega_t h_t}\right] . \end{aligned} \tag{B.3}$$

Now, by using the following result:

$$\int_0^\infty x^{-a/2} \exp\left[-\frac{b}{2x}\right] \mathrm{d}x = \Gamma\left(\frac{a-2}{2}\right) \times \left(\frac{2}{b}\right)^{\frac{a-2}{2}}$$

we can integrate (B.3) with respect to vector $\boldsymbol{\varpi}$ to get the following expression:

$$\begin{aligned} p(\psi \mid \mathbf{y}) \propto & \left(\frac{\nu-2}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^{T} \\ & \times 2^{\frac{T(\nu+1)}{2}} \prod_{t=1}^{T} h_t^{-1/2} b_t^{-\frac{(\nu+1)}{2}} \end{aligned} \tag{B.4}$$

where:

$$b_t \doteq \frac{y_t^2}{h_t} + (\nu-2) = (\nu-2) \times \left[1 + \frac{y_t^2}{(\nu-2)h_t}\right] .$$

Some simplifications of expression (B.4) yield:

$$\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu-2)^{1/2}}\right]^{T} \prod_{t=1}^{T} h_t^{-1/2} \left[1 + \frac{y_t^2}{(\nu-2)h_t}\right]^{-\frac{(\nu+1)}{2}}$$

which if proportional to the likelihood function of parameters ψ when observations $y_t \overset{iid}{\sim} \mathcal{S}^*(0, h_t)$ where $h_t \doteq h_t(\boldsymbol{\alpha}, \beta)$. □

The specification (B.1) gives an additional way of handling the Bayesian estimation of GARCH models with Student-t innovations. It has the appealing aspect that no additional scale factor $\varrho \doteq \frac{\nu-2}{\nu}$ needs to be included in the modeling. However, the simulation scheme presented in Deschamps [2006], Geweke [1993] needs to be slightly modified, as shown hereafter.

In our application, we aim to generate efficiently draws for the degrees of freedom parameter ν. The target density is obtained as follows:

$$
\begin{aligned}
p(\nu \mid \boldsymbol{\varpi}) &\propto p(\boldsymbol{\varpi} \mid \nu) p(\nu) \\
&= \left(\frac{\nu-2}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \left(\prod_{t=1}^{T} \varpi_t^{-\frac{(\nu+2)}{2}}\right) \\
&\times \exp\left[-\frac{(\nu-2)}{2}\sum_{t=1}^{T}\varpi_t^{-1}\right] \lambda \exp\left[-\lambda(\nu-\delta)\right] \mathbb{I}_{\{\nu>\delta\}}
\end{aligned}
$$

where we can express $\prod_{t=1}^{T} \varpi_t^{-\frac{(\nu+2)}{2}}$ as:

$$
\begin{aligned}
\prod_{t=1}^{T} \varpi_t^{-\frac{(\nu+2)}{2}} &= \prod_{t=1}^{T} \exp\left[\ln \varpi_t^{-\frac{(\nu+2)}{2}}\right] \\
&= \prod_{t=1}^{T} \exp\left[-\frac{(\nu+2)}{2} \ln \varpi_t\right] \\
&= \exp\left[-\frac{(\nu+2)}{2}\sum_{t=1}^{T} \ln \varpi_t\right] \\
&= \exp\left[-\frac{\nu}{2}\sum_{t=1}^{T} \ln \varpi_t - \sum_{t=1}^{T} \ln \varpi_t\right] \\
&\propto \exp\left[-\frac{\nu}{2}\sum_{t=1}^{T} \ln \varpi_t\right] .
\end{aligned}
$$

This allows to express the kernel of the target density as follows:

$$
k(\nu) \doteq \left(\frac{\nu-2}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \exp[-\varphi\nu] \mathbb{I}_{\{\nu>\delta\}}
$$

where:

$$
\varphi \doteq \frac{1}{2}\sum_{t=1}^{T}\left[\ln \varpi_t + \varpi_t^{-1}\right] + \lambda .
$$

Note that, since the function $\ln \varpi + \varpi^{-1}$ is minimized at $\varpi = 1$, we have that $\varphi \geqslant \frac{T}{2} + \lambda > \frac{T}{2}$.

Following Deschamps [2006], the sampling density is a translated Exponential with kernel density function given by:

$$
g(\nu; \mu, \delta) \doteq \mu \exp\left[-\mu(\nu-\delta)\right] \mathbb{I}_{\{\nu>\delta\}} \tag{B.5}
$$

where the parameter μ is chosen to maximize the acceptance probability. Following Geweke [1993], we can determine the value for this parameter. Given the usual regularity conditions, a necessary condition is that μ is part of a solution of the following system:

$$\frac{\partial}{\partial \nu}\big[\ln k(\nu) - \ln g(\nu;\mu,\delta)\big] = 0 \tag{B.6a}$$

$$\frac{\partial}{\partial \mu}\ln g(\nu;\mu,\delta) = 0 \ . \tag{B.6b}$$

Expliciting (B.6a) yields:

$$\frac{T}{2}\left[\ln\left(\frac{\nu-2}{2}\right) + \left(\frac{\nu}{\nu-2}\right) - \Psi\left(\frac{\nu}{2}\right)\right] - \varphi + \mu = 0 \tag{B.7}$$

where $\Psi(z) \doteq \frac{\mathrm{d}\ln\Gamma(z)}{\mathrm{d}z}$ denotes the Digamma function, while solving (B.6b) yields:

$$\nu = \frac{1}{\mu} + \delta = \frac{1+\mu\delta}{\mu} \ . \tag{B.8}$$

Furthermore, we note that in expression (B.7), the function:

$$\ln\left(\frac{\nu-2}{2}\right) + \left(\frac{\nu}{\nu-2}\right) - \Psi\left(\frac{\nu}{2}\right)$$

is monotone decreasing from ∞ to 1 on the $]2,\infty[$ interval. Hence, since $\varphi > \frac{T}{2}$, there exists an unique μ satisfying (B.7). Now, inserting (B.8) in expression (B.7) yields:

$$\frac{T}{2}\left[\ln\left(\frac{1+\mu(\delta-2)}{2\mu}\right) + \frac{1+\mu\delta}{1+\mu(\delta-2)} + \Psi\left(\frac{1+\mu\delta}{2\mu}\right)\right] + \mu - \varphi = 0$$

and solving for μ gives the optimal parameter $\bar{\mu}$ for the efficient sampling scheme. The value $\bar{\mu}$ can be found by standard iterative methods. Then, a candidate $\nu^\star$ is sampled from (B.5) with parameter $\bar{\mu}$ and accepted with probability:

$$p^\star \doteq \frac{k(\nu^\star)}{s(\bar{\mu},\delta)g(\nu^\star;\bar{\mu},\delta)}$$

where $s(\mu,\delta)$ is given by:

$$\begin{aligned} s(\mu,\delta) &\doteq k\left(\frac{1+\mu\delta}{\mu}\right)\left[g\left(\frac{1+\mu\delta}{\mu};\mu,\delta\right)\right]^{-1} \\ &= \left(\frac{1+\mu(\delta-2)}{2\mu}\right)^{\frac{T(1+\mu\delta)}{2\mu}}\left[\Gamma\left(\frac{1+\mu\delta}{2\mu}\right)\right]^{-T}\frac{\exp\left[1-\frac{\varphi(1+\mu\delta)}{\mu}\right]}{\mu} \ . \end{aligned}$$

Substituting for $k(\nu^\star)$, $s(\overline{\mu},\delta)$ and $g(\nu^\star;\overline{\mu},\delta)$ in the expression of the acceptance probability yields:

$$
\begin{aligned}
p^\star &= \left(\frac{\nu^\star - 2}{2}\right)^{\frac{T\nu^\star}{2}} \left[\Gamma\left(\frac{\nu^\star}{2}\right)\right]^{-T} \exp[-\varphi\nu^\star] \left(\frac{1+\bar{\mu}(\delta-2)}{2\bar{\mu}}\right)^{-\frac{T(1+\bar{\mu}\delta)}{2\bar{\mu}}} \\
&\times \left[\Gamma\left(\frac{1+\bar{\mu}\delta}{2\bar{\mu}}\right)\right]^{T} \bar{\mu} \exp\left[\frac{\varphi(1+\bar{\mu}\delta)}{\mu} - 1\right] \frac{\exp\left[\bar{\mu}(\nu^\star - \delta)\right]}{\bar{\mu}} \\
&= \left[\frac{\Gamma\left(\frac{1+\bar{\mu}\delta}{2\bar{\mu}}\right)}{\Gamma\left(\frac{\nu^\star}{2}\right)}\right]^{T} \left(\frac{\nu^\star - 2}{2}\right)^{\frac{T\nu^\star}{2}} \left(\frac{1+\bar{\mu}(\delta-2)}{2\bar{\mu}}\right)^{-\frac{T(1+\bar{\mu}\delta)}{2\bar{\mu}}} \\
&\times \exp\left[(\nu^\star - \delta)(\bar{\mu} - \varphi) + \frac{\varphi}{\bar{\mu}} - 1\right] .
\end{aligned}
$$

C

Conditional Moments

In this appendix, we demonstrate the propositions for the conditional moments of the cumulative return $y_{t,s} \doteq \sum_{i=1}^{s} y_{t+i}$ used in **Sect. 6.2.2**. We consider the case where the process $\{y_t\}$ is described by a GARCH$(1,1)$ model for ease of exposition but the methodology can be extended, upon modifications, to higher order GARCH models as well as asymmetric specifications. We recall that the scedastic function of the GARCH$(1,1)$ model is given by:

$$h_t \doteq \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1}$$

where $h_0 = y_0 \doteq 0$ for convenience. For the model disturbances, we consider standardized Normal and Student-t innovations. We define $\mathbb{E}_t(\bullet) \doteq \mathbb{E}(\bullet \mid \mathcal{F}_t)$ and suppress the dependence of the model parameters for notational purposes. The GARCH$(1,1)$ parameters $\boldsymbol{\alpha} \doteq (\alpha_0 \ \alpha_1)'$ and β are regrouped into $\psi \doteq (\boldsymbol{\alpha}, \beta)$. In the case of Student-t innovations, $\psi \doteq (\boldsymbol{\alpha}, \beta, \nu)$. Moreover, the p-th conditional moment of $y_{t,s}$ is denoted by κ_p.

The following properties will be used henceforth:

A. the errors ε_t are *iid* (*i.e.*, independent and identically distributed);

B. $\mathbb{E}_t(\varepsilon_{t+i}) = 0$ for $i \geqslant 1$ (*i.e.*, centered distribution);

C. $\mathbb{E}_t(\varepsilon_{t+i}^2) = 1$ for $i \geqslant 1$ (*i.e.*, unit variance);

D. $\mathbb{E}_t(\varepsilon_{t+i}^3) = 0$ for $i \geqslant 1$ (*i.e.*, symmetric distribution);

E. the conditional variance h_t is known given $\mathcal{F}_{t-1}$, *i.e.*, it is predictable with respect to the natural filtration of the process $\{y_t\}$.

To keep the calculations similar for both Normal and Student-t innovations, we assume unit variance for the disturbances ε_t as emphasized in property C. This has an implication for the fourth conditional moment of the innovations $\kappa_\varepsilon \doteq \mathbb{E}_t(\varepsilon_{t+i}^4)$ for $i \geqslant 1$. Indeed, in the Normal case, $\kappa_\varepsilon = 3$ while in the normalized Student-t case, $\kappa_\varepsilon = 3(\nu-2)/(\nu-4)$. In comparison, the fourth moment of the usual Student-t is $3\nu^2/(\nu-4)(\nu-2)$.

Proposition C.1 (First conditional moment). *For horizon $s \geqslant 1$, the value of the first conditional moment κ_1 is zero.*

Proof.

$$\kappa_1 \doteq \mathbb{E}_t(y_{t,s}) = \mathbb{E}_t\left(\sum_{i=1}^{s} y_{t+i}\right) = \sum_{i=1}^{s} \mathbb{E}_t(y_{t+i}) = 0$$

since for $1 \leqslant i \leqslant s$ we have:

$$\mathbb{E}_t(y_{t+i}) = \mathbb{E}_t(\varepsilon_{t+i} h_{t+i}^{1/2}) \overset{\text{A}}{=} \mathbb{E}_t(\varepsilon_{t+i})\mathbb{E}_t(h_{t+i}^{1/2}) \overset{\text{B}}{=} 0 \,.$$

□

Proposition C.2 (Second conditional moment). *For horizon $s \geqslant 2$, the value of the second conditional moment κ_2 is:*

$$\kappa_2 = \sum_{i=1}^{s} \mathbb{E}_t(h_{t+i})$$

where $\mathbb{E}_t(h_{t+i}) = \alpha_0 + \rho_1 \mathbb{E}_t(h_{t+i-1})$ with $\rho_1 \doteq (\alpha_1 + \beta)$.

Proof. For horizon $s \geqslant 2$, the second power of the cumulative log-returns $y_{t,s}$ is given by:

$$\begin{aligned} y_{t+s}^2 &= \sum_{\substack{i_1, \ldots, i_s \\ i_1 + \ldots + i_s = 2}} \frac{2\,!}{i_1 ! \ldots i_s !} \times y_{t+1}^{i_1} \cdots y_{t+s}^{i_s} \\ &= \sum_{i=1}^{s} y_{t+i}^2 + 2 \sum_{\substack{1 \leqslant i,j \leqslant s \\ i<j}} y_{t+i}\, y_{t+j} \end{aligned}$$

and the second conditional moment κ_2 is obtained by taking the conditional expectation as follows:

$$\kappa_2 = \sum_{i=1}^{s} \mathbb{E}_t(y_{t+i}^2) + 2 \sum_{\substack{1 \leqslant i,j \leqslant s \\ i<j}} \mathbb{E}_t(y_{t+i}\, y_{t+j}) \,. \tag{C.1}$$

Let us consider the first term in expression (C.1). For $1 \leqslant i \leqslant s$, we have:

$$\mathbb{E}_t(y_{t+i}^2) = \mathbb{E}_t(\varepsilon_{t+i}^2 h_{t+i}) \overset{\text{A}}{=} \mathbb{E}_t(\varepsilon_{t+i}^2)\mathbb{E}_t(h_{t+i}) \overset{\text{C}}{=} \mathbb{E}_t(h_{t+i}) \,.$$

For the second term in expression (C.1), since $i < j$, we have:

$$\mathbb{E}_t(y_{t+i}\, y_{t+j}) = \mathbb{E}_t(y_{t+i}\, \varepsilon_{t+j} h_{t+j}^{1/2}) \overset{\text{A}}{=} \mathbb{E}_t(y_{t+i} h_{t+j}^{1/2})\mathbb{E}_t(\varepsilon_{t+j}) \overset{\text{B}}{=} 0 \,.$$

Hence, expression (C.1) simplifies to $\kappa_2 = \sum_{i=1}^{s} \mathbb{E}_t(h_{t+i})$ where $\mathbb{E}_t(h_{t+i})$ can be expressed recursively as follows:

$$\mathbb{E}_t(h_{t+i}) = \alpha_0 + \rho_1 \mathbb{E}_t(h_{t+i-1})$$

with $\mathbb{E}_t(h_{t+1}) = h_{t+1}$ since this value is known given $\mathcal{F}_t$. The parameter ρ_1 is a function of the model parameters ψ. Its expression is found by noting first that the conditional variance at time t can be written as follows:

$$\begin{aligned} h_{t+i} &= \alpha_0 + \alpha_1 y_{t+i-1}^2 + \beta h_{t+i-1} \\ &= \alpha_0 + (\alpha_1 \varepsilon_{t+i-1}^2 + \beta)\, h_{t+i-1} \\ &= \alpha_0 + \varphi_{t+i-1} h_{t+i-1} \end{aligned} \tag{C.2}$$

where $\varphi_t \doteq (\alpha_1 \varepsilon_t^2 + \beta)$. By taking the conditional expectation of (C.2), we obtain:

$$\begin{aligned} \mathbb{E}_t(h_{t+i}) &= \alpha_0 + \mathbb{E}_t(\varphi_{t+i-1} h_{t+i-1}) \\ &\overset{\text{A}}{=} \alpha_0 + \mathbb{E}_t(\varphi_{t+i-1}) \mathbb{E}_t(h_{t+i-1}) \\ &= \alpha_0 + \rho_1 \mathbb{E}_t(h_{t+i-1}) \end{aligned}$$

where $\rho_1 \doteq \mathbb{E}_t(\varphi_{t+i-1}) = \mathbb{E}_t(\alpha_1 \varepsilon_{t+i-1}^2 + \beta) = \alpha_1\, \mathbb{E}_t(\varepsilon_{t+i-1}^2) + \beta \overset{\text{C}}{=} \alpha_1 + \beta$. □

Proposition C.3 (Third conditional moment). *For horizon $s \geqslant 3$, the value of the third conditional moment κ_3 is zero.*

Proof. For $s \geqslant 3$, the third power of the cumulative log-returns $y_{t,s}$ is given by:

$$\begin{aligned} y_{t,s}^3 &= \sum_{\substack{i_1, \ldots, i_s \\ i_1 + \ldots + i_s = 3}} \frac{3\,!}{i_1! \ldots i_s!} \times y_{t+1}^{i_1} \cdots y_{t+s}^{i_s} \\ &= \sum_{i=1}^{s} y_{t+i}^3 + 3 \sum_{\substack{1 \leqslant i,j \leqslant s \\ i \neq j}} y_{t+i}^2\, y_{t+j} + 6 \sum_{\substack{1 \leqslant i,j,k \leqslant s \\ i<j<k}} y_{t+i}\, y_{t+j}\, y_{t+k} \end{aligned}$$

and the third conditional moment κ_3 is obtained by taking the conditional expectation as follows:

$$\kappa_3 = \sum_{i=1}^{s} \mathbb{E}_t(y_{t+i}^3) + 3 \sum_{\substack{1 \leqslant i,j \leqslant s \\ i \neq j}} \mathbb{E}_t(y_{t+i}^2\, y_{t+j}) + 6 \sum_{\substack{1 \leqslant i,j,k \leqslant s \\ i<j<k}} \mathbb{E}_t(y_{t+i}\, y_{t+j}\, y_{t+k})\,. \tag{C.3}$$

Let us consider the first term in expression (C.3). For $1 \leqslant i \leqslant s$, we have:

$$\mathbb{E}_t(y_{t+i}^3) = \mathbb{E}_t(\varepsilon_{t+i}^3 h_{t+i}^{3/2}) \overset{\text{A}}{=} \mathbb{E}_t(\varepsilon_{t+i}^3) \mathbb{E}_t(h_{t+i}^{3/2}) \overset{\text{D}}{=} 0\,.$$

For the second term, we need to distinguish two cases. First, when $i < j$, we obtain:

$$\mathbb{E}_t(y_{t+i}^2\, y_{t+j}) = \mathbb{E}_t(y_{t+i}^2\, \varepsilon_{t+j} h_{t+j}^{1/2}) \overset{\text{A}}{=} \mathbb{E}_t(y_{t+i}^2 h_{t+j}^{1/2}) \mathbb{E}_t(\varepsilon_{t+j}) \overset{\text{B}}{=} 0\,.$$

In the case where $j < i$, we have:

$$\begin{aligned}
\mathbb{E}_t(y_{t+i}^2\, y_{t+j}) &= \mathbb{E}_t(\varepsilon_{t+i}^2 h_{t+i}\, y_{t+j}) \\
&\overset{\text{A}}{=} \mathbb{E}_t(\varepsilon_{t+i}^2)\mathbb{E}_t(h_{t+i} y_{t+j}) \\
&\overset{\text{C}}{=} \mathbb{E}_t(h_{t+i} y_{t+j}) \\
&= \mathbb{E}_t(h_{t+i}\varepsilon_{t+j} h_{t+j}^{1/2}) \\
&= \mathbb{E}_t\big(\mathbb{E}_{t+j-1}[h_{t+i}\varepsilon_{t+j} h_{t+j}^{1/2}]\big) \\
&\overset{\text{E}}{=} \mathbb{E}_t\big(h_{t+j}^{1/2}\, \mathbb{E}_{t+j-1}[h_{t+i}\varepsilon_{t+j}]\big) \qquad \text{(C.4)}
\end{aligned}$$

where the conditional expectation $\mathbb{E}_{t+j-1}[\bullet]$ in (C.4) can be expressed as:

$$\begin{aligned}
\mathbb{E}_{t+j-1}[\varepsilon_{t+j} h_{t+i}] &\overset{\text{(C.2)}}{=} \mathbb{E}_{t+j-1}\big[\varepsilon_{t+j}(\alpha_0 + \varphi_{t+i-1} h_{t+i-1})\big] \\
&\overset{\text{B}}{=} \mathbb{E}_{t+j-1}[\varepsilon_{t+j}\varphi_{t+i-1} h_{t+i-1}] \\
&\overset{\text{A}}{=} \mathbb{E}_{t+j-1}[\varepsilon_{t+j} h_{t+i-1}]\mathbb{E}_{t+j-1}[\varphi_{t+i-1}] \\
&= \rho_1 \mathbb{E}_{t+j-1}[\varepsilon_{t+j} h_{t+i-1}] \\
&\vdots \\
&= \rho_1^{i-j}\mathbb{E}_{t+j-1}[\varepsilon_{t+j}\varphi_{t+j} h_{t+j}] \\
&\overset{\text{E}}{=} \rho_1^{i-j}\, h_{t+j}\mathbb{E}_{t+j-1}[\varepsilon_{t+j}\varphi_{t+j}] \\
&= 0 \ . \qquad \text{(C.5)}
\end{aligned}$$

The last equality follows from:

$$\begin{aligned}
\mathbb{E}_{t+j-1}\big[\varepsilon_{t+j}\varphi_{t+j}\big] &= \mathbb{E}_{t+j-1}\big[\varepsilon_{t+j}(\alpha_1\varepsilon_{t+j}^2 + \beta)\big] \\
&= \alpha_1\, \mathbb{E}_{t+j-1}[\varepsilon_{t+j}^3] + \beta\, \mathbb{E}_{t+j-1}[\varepsilon_{t+j}] \\
&\overset{\text{B,D}}{=} 0 \ .
\end{aligned}$$

Finally, let us consider the last term in expression (C.3). Since $i < j < k$, we have:

$$\mathbb{E}_t(y_{t+i}\, y_{t+j}\, y_{t+k}) \overset{\text{A}}{=} \mathbb{E}_t(y_{t+i}\, y_{t+j} h_{t+k}^{1/2})\mathbb{E}_t(\varepsilon_{t+k}) \overset{\text{B}}{=} 0 \ .$$

Each term in expression (C.3) vanishes so that the third conditional moment is zero. □

Proposition C.4 (Fourth conditional moment). *For horizon $s \geqslant 4$, the value of the fourth conditional moment κ_4 is:*

$$\kappa_4 = \kappa_\varepsilon \sum_{i=1}^{s} \mathbb{E}_t(h_{t+i}^2) + 6 \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} \mathbb{E}_t(y_{t+i}^2\, y_{t+j}^2)$$

where:

$$\mathbb{E}_t(h_{t+i}^2) = \alpha_0^2 + \tau_1\,\mathbb{E}_t(h_{t+i-1}) + \tau_2\,\mathbb{E}_t(h_{t+i-1}^2)$$
$$\mathbb{E}_t(y_{t+i}^2\; y_{t+j}^2) = \alpha_0\left(\frac{1-\rho_1^{j-i}}{1-\rho_1}\right)\mathbb{E}_t(h_{t+i}) + \rho_1^{j-i-1}\,\rho_2\,\mathbb{E}_t(h_{t+i}^2)$$

with $\rho_1 \doteq (\alpha_1+\beta)$, $\rho_2 \doteq (\kappa_\varepsilon\alpha_1+\beta)$, $\tau_1 \doteq 2\alpha_0\rho_1$ *and* $\tau_2 \doteq \kappa_\varepsilon\alpha_1^2 + \beta(2\alpha_1+\beta)$.

Proof. For $s \geqslant 4$, the fourth power of the cumulative log-returns $y_{t,s}$ is given by:

$$\begin{aligned}
y_{t,s}^4 &= \sum_{\substack{i_1,\,\ldots\,,i_s \\ i_1+\ldots+i_s=4}} \frac{4\,!}{i_1!\ldots i_s!} \times y_{t+1}^{i_1}\cdots y_{t+s}^{i_s} \\
&= \sum_{i=1}^{s} y_{t+i}^4 + 4\sum_{\substack{1\leqslant i,j\leqslant s \\ i\neq j}} y_{t+i}^3\,y_{t+j} + 6\sum_{\substack{1\leqslant i,j\leqslant s \\ i<j}} y_{t+i}^2\,y_{t+j}^2 \\
&+ 6\sum_{\substack{1\leqslant i,j,k\leqslant s \\ i\neq j\neq k}} y_{t+i}^2\,y_{t+j}\,y_{t+k} \\
&+ 24\sum_{\substack{1\leqslant i,j,k,l\leqslant s \\ i<j<k<l}} y_{t+i}\,y_{t+j}\,y_{t+k}\,y_{t+l}\;.
\end{aligned}$$

and the fourth conditional moment κ_4 is obtained by taking the conditional expectation as follows:

$$\begin{aligned}
\kappa_4 = &\sum_{i=1}^{s} \mathbb{E}_t(y_{t+i}^4) + 4\sum_{\substack{1\leqslant i,j\leqslant s \\ i\neq j}} \mathbb{E}_t(y_{t+i}^3\,y_{t+j}) + 6\sum_{\substack{1\leqslant i,j\leqslant s \\ i<j}} \mathbb{E}_t(y_{t+i}^2\,y_{t+j}^2) \\
&+ 6\sum_{\substack{1\leqslant i,j,k\leqslant s \\ i\neq j\neq k}} \mathbb{E}_t(y_{t+i}^2\,y_{t+j}\,y_{t+k}) \\
&+ 24\sum_{\substack{1\leqslant i,j,k,l\leqslant s \\ i<j<k<l}} \mathbb{E}_t(y_{t+i}\,y_{t+j}\,y_{t+k}\,y_{t+l})
\end{aligned} \tag{C.6}$$

Let us first consider the terms in (C.6) which vanish. For the second term in expression (C.6), when $i<j$, we have:

$$\mathbb{E}_t(y_{t+i}^3\,y_{t+j}) = \mathbb{E}_t(y_{t+i}^3\,\varepsilon_{t+j}h_{t+j}^{1/2}) \overset{\text{A}}{=} \mathbb{E}_t(y_{t+i}^3\,h_{t+j}^{1/2})\mathbb{E}_t(\varepsilon_{t+j}) \overset{\text{B}}{=} 0$$

while in the case where $j<i$, we obtain:

$$\mathbb{E}_t(y_{t+i}^3\,y_{t+j}) = \mathbb{E}_t(\varepsilon_{t+i}^3 h_{t+i}^{3/2}\,y_{t+j}) \overset{\text{A}}{=} \mathbb{E}_t(\varepsilon_{t+i}^3)\mathbb{E}_t(h_{t+i}^{3/2}\,y_{t+j}) \overset{\text{D}}{=} 0\;.$$

Let us now consider the fourth term in expression (C.6). Assuming that $i < j < k$ yields:

$$\mathbb{E}_t(y_{t+i}^2\, y_{t+j}\, y_{t+k}) \overset{\text{A}}{=} \mathbb{E}_t(y_{t+i}^2\, y_{t+j}\, h_{t+k}^{1/2})\mathbb{E}_t(\varepsilon_{t+k}) \overset{\text{B}}{=} 0\ .$$

The same holds for $i < k < j$, $j < i < k$ and $k < i < j$. It remains to consider the cases where i is the greatest integer, *i.e.*, $k < j < i$ and $j < k < i$. By assuming (without loss of generality) that $j < k < i$ we have:

$$\begin{aligned}
\mathbb{E}_t(y_{t+i}^2\, y_{t+j}\, y_{t+k}) &= \mathbb{E}_t(\varepsilon_{t+i}^2 h_{t+i}\, y_{t+j}\, y_{t+k}) \\
&\overset{\text{A}}{=} \mathbb{E}_t(\varepsilon_{t+i}^2)\mathbb{E}_t(h_{t+i}\, y_{t+j}\, y_{t+k}) \\
&\overset{\text{C}}{=} \mathbb{E}_t(h_{t+i}\, y_{t+j}\, y_{t+k}) \\
&= \mathbb{E}_t\big(\mathbb{E}_{t+j-1}[h_{t+i}\, \varepsilon_{t+j} h_{t+j}^{1/2}\, y_{t+k}]\big) \\
&\overset{\text{F}}{=} \mathbb{E}_t\big(h_{t+j}^{1/2}\, \mathbb{E}_{t+j-1}[h_{t+i}\, \varepsilon_{t+j}\, y_{t+k}]\big) \\
&= \mathbb{E}_t\Big(h_{t+j}^{1/2}\, \mathbb{E}_{t+j-1}\big[\mathbb{E}_{t+k-1}\{h_{t+i}\, \varepsilon_{t+j}\, \varepsilon_{t+k} h_{t+k}^{1/2}\}\big]\Big) \\
&\overset{\text{F}}{=} \mathbb{E}_t\Big(h_{t+j}^{1/2}\, \mathbb{E}_{t+j-1}\big[h_{t+k}^{1/2}\, \varepsilon_{t+j}\, \mathbb{E}_{t+k-1}\{h_{t+i}\, \varepsilon_{t+k}\}\big]\Big) \\
&\overset{(\text{C.5})}{=} 0\ .
\end{aligned}$$

Finally, let us consider the last term in expression (C.6). By assuming (without loss of generality) that $i < j < k < l$, we have:

$$\mathbb{E}_t(y_{t+i}\, y_{t+j}\, y_{t+k}\, y_{t+l}) \overset{\text{A}}{=} \mathbb{E}_t(y_{t+i}\, y_{t+j}\, y_{t+k}\, h_{t+l}^{1/2})\mathbb{E}_t(\varepsilon_{t+l}) \overset{\text{B}}{=} 0\ .$$

With these intermediate results, expression (C.6) can be simplified as follows:

$$\begin{aligned}
\kappa_4 &= \sum_{i=1}^{s} \mathbb{E}_t(y_{t+i}^4) + 6 \sum_{\substack{1 \leqslant i,j \leqslant s \\ i<j}} \mathbb{E}_t(y_{t+i}^2\, y_{t+j}^2) \\
&= \sum_{i=1}^{s} \mathbb{E}_t(y_{t+i}^4) + 6 \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} \mathbb{E}_t(y_{t+i}^2\, y_{t+j}^2)\ . \qquad \text{(C.7)}
\end{aligned}$$

Let us explicit the first term in expression (C.7). We have:

$$\mathbb{E}_t(y_{t+i}^4) = \mathbb{E}_t(\varepsilon_{t+i}^4 h_{t+i}^2) \overset{\text{A}}{=} \mathbb{E}_t(\varepsilon_{t+i}^4)\mathbb{E}_t(h_{t+i}^2) = \kappa_\varepsilon\, \mathbb{E}_t(h_{t+i}^2)\ .$$

The term $\mathbb{E}_t(h_{t+i}^2)$ can be estimated by recursion as follows:

$$\begin{aligned}
\mathbb{E}_t(h_{t+i}^2) &\overset{(C.2)}{=} \mathbb{E}_t\big([\alpha_0 + \varphi_{t+i-1} h_{t+i-1}]^2\big) \\
&= \alpha_0^2 + 2\alpha_0\, \mathbb{E}_t(\varphi_{t+i-1} h_{t+i-1}) + \mathbb{E}_t(\varphi_{t+i-1}^2 h_{t+i-1}^2) \\
&\overset{A}{=} \alpha_0^2 + 2\alpha_0\, \mathbb{E}_t(\varphi_{t+i-1})\mathbb{E}_t(h_{t+i-1}) + \mathbb{E}_t(\varphi_{t+i-1}^2)\mathbb{E}_t(h_{t+i-1}^2) \\
&= \alpha_0^2 + \tau_1\, \mathbb{E}_t(h_{t+i-1}) + \tau_2\, \mathbb{E}_t(h_{t+i-1}^2) \qquad \text{(C.8)}
\end{aligned}$$

where $\tau_1 \doteq 2\alpha_0\rho_1$, $\rho_1 \doteq \mathbb{E}_t(\varphi_{t+i-1})$ and $\tau_2 \doteq \mathbb{E}_t(\varphi_{t+i-1}^2)$. Expression (C.8) can be computed recursively since h_{t+1} is known given $\mathcal{F}_t$. Furthermore, we can explicit the term τ_2 as follows:

$$\begin{aligned}
\tau_2 &\doteq \mathbb{E}_t(\varphi_{t+i-1}^2) \\
&= \mathbb{E}_t(\alpha_1^2 \varepsilon_{t+i-1}^4 + 2\alpha_1\beta\varepsilon_{t+i-1}^2 + \beta^2) \\
&= \alpha_1^2\, \mathbb{E}_t(\varepsilon_{t+i-1}^4) + 2\alpha_1\beta\, \mathbb{E}_t(\varepsilon_{t+i-1}^2) + \beta^2 \\
&\overset{C,E}{=} \kappa_\varepsilon \alpha_1^2 + 2\alpha_1\beta + \beta^2 \\
&= \kappa_\varepsilon \alpha_1^2 + \beta(2\alpha_1 + \beta)\ .
\end{aligned}$$

It remains to consider the last term in expression (C.6). Since $i < j$, we have:

$$\begin{aligned}
\mathbb{E}_t(y_{t+i}^2\, y_{t+j}^2) &= \mathbb{E}_t(y_{t+i}^2\, \varepsilon_{t+j}^2 h_{t+j}) \\
&\overset{A}{=} \mathbb{E}_t(y_{t+i}^2\, h_{t+j})\mathbb{E}_t(\varepsilon_{t+j}^2) \\
&\overset{C}{=} \mathbb{E}_t(y_{t+i}^2\, h_{t+j}) \\
&= \mathbb{E}_t\big(\mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2 h_{t+i}\, h_{t+j}]\big) \\
&\overset{F}{=} \mathbb{E}_t\big(h_{t+i}\, \mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2\, h_{t+j}]\big) \qquad \text{(C.9)}
\end{aligned}$$

where the conditional expectation $\mathbb{E}_{t+i-1}[\bullet]$ in the last equality can be developed as follows:

$$
\begin{aligned}
\mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2\, h_{t+j}] &\overset{(C.2)}{=} \mathbb{E}_{t+i-1}\big[\varepsilon_{t+i}^2\,(\alpha_0+\varphi_{t+j-1}h_{t+j-1})\big] \\
&= \alpha_0\,\mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2] + \mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2\varphi_{t+j-1}h_{t+j-1}] \\
&\overset{A,C}{=} \alpha_0 + \mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2 h_{t+j-1}]\mathbb{E}_{t+i-1}[\varphi_{t+j-1}] \\
&= \alpha_0 + \rho_1\mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2 h_{t+j-1}] \\
&\qquad\vdots \\
&= \alpha_0(1+\ldots+\rho_1^{j-i-2}) + \rho_1^{j-i-1}\mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2 h_{t+i+1}] \\
&= \alpha_0\sum_{k=0}^{j-i-2}\rho_1^k + \rho_1^{j-i-1}\mathbb{E}_{t+i-1}\big[\varepsilon_{t+i}^2\,(\alpha_0+\varphi_{t+i}h_{t+i})\big] \\
&\overset{C}{=} \alpha_0\sum_{k=0}^{j-i-2}\rho_1^k + \alpha_0\rho_1^{j-i-1} + \rho_1^{j-i-1}\mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2\varphi_{t+i}h_{t+i}] \\
&\overset{E}{=} \alpha_0\sum_{k=0}^{j-i-1}\rho_1^k + \rho_1^{j-i-1}\,h_{t+i}\,\mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2\varphi_{t+i}] \\
&= \alpha_0\left(\frac{1-\rho_1^{j-i}}{1-\rho_1}\right) + \rho_1^{j-i-1}\rho_2\,h_{t+i} \qquad \text{(C.10)}
\end{aligned}
$$

with $\rho_2 \doteq \mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2\varphi_{t+i}]$. The parameter ρ_2 is a function of the model parameters:

$$
\begin{aligned}
\rho_2 &\doteq \mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2\varphi_{t+i}] \\
&= \mathbb{E}_{t+i-1}\big[\varepsilon_{t+i}^2(\alpha_1\varepsilon_{t+i}^2+\beta)\big] \\
&= \alpha_1\,\mathbb{E}_{t+i-1}[\varepsilon_{t+i}^4] + \beta\,\mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2] \\
&\overset{C,E}{=} \kappa_\varepsilon\alpha_1+\beta\,.
\end{aligned}
$$

Replacing (C.10) in expression (C.9) yields:

$$
\begin{aligned}
\mathbb{E}_t(y_{t+i}^2\,y_{t+j}^2) &= \mathbb{E}_t\big(h_{t+i}\,\mathbb{E}_{t+i-1}[\varepsilon_{t+i}^2\,h_{t+j}]\big) \\
&= \alpha_0\left(\frac{1-\rho_1^{j-i}}{1-\rho_1}\right)\mathbb{E}_t(h_{t+i}) + \rho_1^{j-i-1}\rho_2\,\mathbb{E}_t(h_{t+i}^2)\,.
\end{aligned}
$$

which can be computed recursively since h_{t+1} is known given $\mathcal{F}_t$. □

Computational Details

The algorithms have been written in the `R` language, version 2.4.1 [see R Development Core Team 2007], with some subroutines implemented in `C` in order to speed up the simulation procedure; this is required in the case of GARCH models due to the recursive nature of the conditional process which drastically slows down the computations with high level interpreted languages. Moreover, the validity of the algorithms as well as the correctness of the computer code were verified by a variant of the method proposed by Geweke [2004]. We refer the reader to the end of **Sect. 3.2.2** for further details. The `R` packages **boa** 0-1.3, **mvtnorm** 0-7.5, **coda** 0.10-7, **sandwich** 2.0-0, **lmtest** 0.9-18, **MASS** 7.2-32 have also been used to perform specific tasks. The `R` program itself and packages are available from CRAN at `http://CRAN.R-project.org`.

Regarding the ML estimation of the models, the likelihood functions were maximized using the `R` function `nlminb` which performs unconstrained and constrained optimization using PORT routines. In some cases, the procedure was initialized using the `DEoptim` function provided by the `R` package **DEoptim** 1.01-2 [see Ardia 2007b]. This function performs a global (robust) optimization based on the *Differential Evolution* algorithm. The reader is referred to Price, Storn, and Lampinen [2006] for further details.

Finally, we note that the `R` package **bayesGARCH** will soon be available from CRAN [see Ardia 2007a]. This package allows the Bayesian estimation of the GARCH$(1, 1)$ model with Student-t innovations. The underlying algorithm is based on Nakatsuma [1998, 2000] for the generation of the scedastic function's parameters. The generation of the degrees of freedom parameter is achieved following Deschamps [2006], Geweke [1993]. By using a translated Exponential as a prior on the degrees of freedom parameter, Normal innovations can be obtained as a special case of the sampler. Moreover, the function `addPriorConditions` allows to add any constraints on the model parameters in the M-H sampler.

Abbreviations and Notations

Most of the notation as well as dimensions of vectors and matrices are clearly defined in the text where it is used. Occasionally, a mathematical symbol has been assigned to a different object.

Abbreviations	
ACF	Autocorrelation function
AEL	Absolute error loss
ARCH	AutoRegressive Conditional Heteroscedasticity
ARDS	Adaptive Radial-Based Direction Sampling
AR	AutoRegressive
ARMA	AutoRegressive Moving Average
BIC	Bayesian information criterion
BF	Bayes factor
BUGS	Bayesian analysis Using Gibbs Software
cont.	Continued
CC	Conditional coverage
CSC	Covariance stationarity condition
DEM/GBP	Deutschmark vs British Pounds
DIC	Deviance information criterion
EM	Expectation Maximization
ES	Expected Shortfall
FFBS	Forward Filtering Backward Sampling
GARCH	Generalized ARCH
GJR	Asymmetric GARCH
IF	Inefficiency factor
IND	Independence
Linex	Linex loss
MCMC	Markov Chain Monte Carlo
M-H	Metropolis-Hastings

continued on the next page

Abbreviations (*cont.*)	
ML	Maximum Likelihood
MS-GARCH	Markov-switching GARCH
MS-GJR	Markov-switching GJR
NA	Not applicable
NSE	Numerical standard error
P&L	Profit and loss
SEL	Squared error loss
SMI	Swiss Market Index
S&P100	Standard & Poors 100
S&P500	Standard & Poors 500
SSC	Strict stationarity condition
UC	Unconditional Coverage
VaR	Value at Risk
VIX	Volatility index of the S&P100 index

Densities	
$\mathcal{N}(0,1)$	Univariate standard Normal density
$\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$	d-dimensional Normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ
$\mathcal{S}(\mu, \sigma^2, \nu)$	Univariate standard Student-t density with mean μ, scale parameter σ^2 and degrees of freedom ν
$\mathcal{SS}(\mu, \sigma^2, \tau, \gamma)$	Univariate skewed Student-t density with mean μ, scale parameter σ^2, degrees of freedom τ and asymmetry parameter γ
$\mathcal{IG}(a, b)$	Inverted Gamma density with shape parameter a and rate parameter b
$\mathcal{D}(\boldsymbol{\eta})$	Dirichlet density with parameter $\boldsymbol{\eta}$
χ^2_k	Chi-squared density with k degrees of freedom

Notations used throughout the book	
$\sim$	"is generated from"
$\propto$	"is proportional to"
$\succ$	"is preferred over"
$\approx$	"is approximately equal to"
$\doteq$	"is defined to"
$\equiv$	"is equivalent to"
$\in$	"belongs to"
iid	Independent and identically distributed
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^*$	Set of non-zero real numbers
$\mathbb{R}^+$	Set of real positive numbers

continued on the next page

Notations used throughout the book (*cont.*)

$\mathbb{N}^*$	Set of non-zero natural numbers
$\bullet$	A scalar, a vector, a matrix, a parameter or a set of parameters (depending on the context)
$\exp[\bullet]$	Exponential
$\ln[\bullet]$	Natural logarithm
$\bullet!$	Factorial
$\lvert \bullet \rvert$	Absolute value
$\sum$	Summation operator
$\prod$	Product operator
$\times$	Multiplication operator or Cartesian product
$\int$	Integral
$\bullet_{\neq i}$	Vector without its ith component
$\{\bullet\}$	Sequence (or set) of variables (or observations)
$\max\{\bullet\}$	Maximum value of a set
$\min\{\bullet\}$	Minimum value of a set
$\inf\{\bullet\}$	Infimum of a set
$\#\{\bullet\}$	Number of elements in a set
$\det(\bullet)$	Determinant of a square matrix
$\bullet^{-1}$	Inverse of a number or a square matrix
$\Gamma(\bullet)$	Gamma function
$\Psi(\bullet)$	Digamma function
I_d	Identity matrix of size d
$\mathcal{F}_t$	Information set up to time t
$I(d)$	Integrated of order d
$\mathbb{I}_{\{\bullet\}}$	Indicator function
$\bullet'$	Transpose of a vector or a matrix
$\bullet_{\text{MLE}}$	Maximum Likelihood estimate
$\overline{\bullet}$	Sample average
$\bullet_\phi$	ϕth percentile of a sample
$\mathcal{L}(\bullet)$	Likelihood function
$\mathcal{L}(\bullet \mid \mathbf{y})$	Marginal (observed) likelihood function
$f_\bullet(\bullet)$	Density function
$F_\bullet(\bullet)$	Distribution function
$p(\bullet \mid \mathbf{y})$	Posterior density
$\mathbb{P}(\bullet)$	Probability
$\mathbb{P}(\bullet \mid \bullet)$	Conditional probability
$\mathbb{E}(\bullet)$	Expectation
$\mathbb{E}(\bullet \mid \bullet), \mathbb{E}_\bullet(\bullet)$	Conditional expectations
$\widehat{R}$	Potential scale reduction factor
$\mathbf{0}$	Vector of zeros
α_0 α_1 (α_2)	ARCH parameters
$\boldsymbol{\alpha}$	Vector of ARCH parameters
β	GARCH parameter
$\boldsymbol{\gamma}$	Vector of linear regression's parameters

continued on the next page

Notations used throughout the book (*cont.*)	
ν	Degrees of freedom parameter of the Student-t density
ϖ_t	Latent scale variable at time t
$\boldsymbol{\varpi}$	Vector of latent scale variables
$\theta\ \psi\ \Theta$	Sets (or vectors) of the model parameters
$\overline{\alpha}$	$\overline{\alpha} \doteq (\alpha_1 + \alpha_2)/2$
$\triangle\alpha$	Leverage effect coefficient: $\triangle\alpha \doteq (\alpha_2 - \alpha_1)$
T	Length of the underlying time series (number of observations)
y_t	Dependent variable at time t
$\mathbf{y}$	Vector containing the observations of y_t
ε_t	Model innovation at time t
$\widehat{\varepsilon}_t$	Model residual at time t
u_t	Linear regression's error at time t
$\mathbf{u}$	Vector of linear regression's errors
m	Number of exogenous or lagged dependent variables
$\mathbf{x}_t$	Vector of exogenous or lagged dependent variables at time t
X	Matrix whose tth row is $\mathbf{x}_t'$
h_t	Conditional variance at time t
h_y	Unconditional variance of the underlying process
$\Sigma\ \Lambda$	Diagonal matrices of conditional variances
ϱ	A scaling factor: $\varrho \doteq \frac{\nu-2}{\nu}$
J	Number of draws in the posterior sample
$q_\bullet(\bullet)$	Proposal density
$\widetilde{\bullet}$	Previous draw in the sampler
$\bullet^{[j]}$	jth draw in the sampler
$\bullet^\star$	New draw in the sampler
$\mu_\bullet\ \Sigma_\bullet$	Hyperparameters of the (truncated) Normal priors
$\widehat{\boldsymbol{\mu}}_\bullet\ \widehat{\Sigma}_\bullet$	Parameters of the (truncated) Normal proposals
$\lambda\ \delta$	Hyperparameters of the translated Exponential prior
κ_ε	Kurtosis of the innovations
κ_y	Unconditional kurtosis of the underlying process
$l_t^*\ v_t^*\ v_t^{**}$	Recursive transformations
$\mathbf{c}_t$	Vector of recursive transformations
C	Matrix whose tth row is $\mathbf{c}_t'$
V_t^ϕ	Indicator variable at time t for the violation of the one-day ahead VaR at risk level ϕ

Notations specific to Chap. 6	
$\phi\ \phi^c$	Risk level of the VaR and $\phi^c \doteq 1 - \phi$
s	Time horizon (in days)
$y_{t,s}$	Cumulated return over a s-day ahead horizon at time t

continued on the next page

Notations specific to Chap. 6 (*cont.*)

$y_{t:s}$	Vector of s-day ahead log-returns at time t
$\kappa_p(\bullet)$	pth conditional moment
z_ϕ	ϕth percentile of the $\mathcal{N}(0,1)$ density
$t_\phi(\nu)$	ϕth percentile of the $\mathcal{S}(0,1,\nu)$ density
$\mathscr{L}$	Loss function
ω	True value of the forecast (or one-dimensional parameter)
$\widehat{\omega}$	Point estimate of ω
$\widehat{\omega}_{\mathscr{L}}$	Optimal (Bayes) point estimate under loss function $\mathscr{L}$
$R_{\mathscr{L}}(\bullet \mid \mathbf{y})$	Posterior risk function under loss $\mathscr{L}$
a	Parameter of the Linex loss function
a_1 a_2 q	Parameters of the Monomial loss function
Δ	Statistical error: $\Delta \doteq (\widehat{\omega} - \omega)$
N	Number of out-of-sample observations
$\widehat{\mathrm{VaR}}$	VaR point estimate
VaR_t^ϕ	One-day ahead VaR at time t for risk level ϕ
$\mathrm{VaR}_{t,s}^\phi$	s-day ahead VaR at time t for risk level ϕ
$\widehat{\mathrm{VaR}}_{\mathscr{L},t,s}^{\phi}$	s-day ahead VaR point estimate at time t for risk level ϕ and loss function $\mathscr{L}$
$\widehat{\mathrm{RC}}_{\mathscr{L},t}$	Regulatory capital point estimate at time t for loss function $\mathscr{L}$
$V_{t,s}^\phi$	Indicator variable at time t for the violation of the s-day ahead VaR at risk level ϕ
ES_t^ϕ	One-day ahead ES at time t for risk level ϕ
$\mathrm{ES}_{t,s}^\phi$	s-day ahead ES at time t for risk level ϕ

Notations specific to Chap. 7

$\otimes$	Kronecker product
$\odot$	Hadamard product, *i.e.*, element-by-element multiplication
$\mathrm{vec}(\bullet)$	Vectorization (column stacking) of a matrix
$\mathrm{tr}(\bullet)$	Trace of a square matrix
K	Number of regimes
α_0^k α_1^k α_2^k	ARCH parameters in state k
$\boldsymbol{\alpha}_0$ $\boldsymbol{\alpha}_1$ $\boldsymbol{\alpha}_2$	Vectors of ARCH parameters
$\boldsymbol{\alpha}$	Vector containing the vectors of ARCH parameters
β^k	GARCH parameter in state k
$\boldsymbol{\beta}$	Vector of GARCH parameters
$\mathbf{e}_t$	s_tth column of the matrix I_K
$\boldsymbol{\iota}_K$	Vector of ones of size K
$\mathbf{h}_t$	Vector of conditional variances at time t
$\boldsymbol{\eta}_i$	Parameter's vector of the ith Dirichlet density

continued on the next page

Notations specific to Chap. 7 (*cont.*)	
$[\bullet]_i$	ith component of a vector
$\boldsymbol{\pi}$	Vector of ergodic probabilities
P	Matrix of transition probabilities
s_t	Discrete state variable estimated at time t
$\mathbf{s}$	Vector of state variables
z_t	Approximate p-score at time t
u_t	Generalized residual at time t
$D(\bullet)$	Deviance function
p_D	Effective number of parameters
B	Number of replications for the block bootstrap
$q(\psi)$	Importance density for parameter ψ
$p(\mathbf{y})$	Model likelihood
$p_t(\mathbf{y})$	tth estimate of the model likelihood in the bridge sampling
$p_0(\mathbf{y})$	Reciprocal sampling estimator of $p(\mathbf{y})$

List of Tables

List of Figures

References

Andrews DWK (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica*, **59**(3), 817–858. Cited on pages 14, 26, and 126.

Andrews DWK, Monahan JC (1992). "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator." *Econometrica*, **60**(4), 953–966. Cited on pages 14, 26, and 126.

Ardia D (2007a). *bayesGARCH: Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations in R*. R Foundation for Statistical Computing, Vienna, Austria. In preparation, URL `http://stat.ethz.ch/CRAN/doc/packages/bayesGARCH.PDF`. Cited on page 179.

Ardia D (2007b). *DEoptim: Differential Evolution Optimization*. R Foundation for Statistical Computing, Vienna, Austria. Version 1.1-8, URL `http://stat.ethz.ch/CRAN/doc/packages/DEoptim.PDF`. Cited on page 179.

Artzner P, Delbaen F, Eber JM, Heath D (1999). "Coherent Measures of Risk." *Quantitative Finance*, **9**(3), 203–228. Cited on page 104.

Ausín MC, Galeano P (2007). "Bayesian Estimation of the Gaussian Mixture GARCH Model." *Computational Statistics and Data Analysis*, **51**(5), 2636–2652. doi:10.1016/j.csda.2006.01.006. Cited on page 4.

Bams D, Lehnert T, Wolff CCP (2005). "An Evaluation Framework for Alternative VaR Models." *Journal of International Money and Finance*, **24**(6), 944–958. doi:10.1016/j.jimonfin.2005.05.004. Cited on page 75.

Basel Committee on Banking Supervision (1995). *An Internal Model-Based Approach to Market Risk Capital Requirements*. The Bank for International Settlements, Basel, Switzerland. URL `http://www.bis.org/publ/bcbs17.htm`. Cited on page 73.

Basel Committee on Banking Supervision (1996a). *Amendment to the Capital Accord to Incorporate Market Risk*. The Bank for International Settlements, Basel, Switzerland. URL `http://www.bis.org/publ/bcbs24.htm`. Cited on page 96.

Basel Committee on Banking Supervision (1996b). *Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Approach to Market Risk Capital Requirement*. The Bank for International Settlements,

Basel, Switzerland. URL http://www.bis.org/publ/bcbs22.htm. Cited on pages 93, 100, and 144.

Bauwens L, Bos CS, van Dijk HK, van Oest RD (2004). "Adaptive Radial-Based Direction Sampling: Some Flexible and Robust Monte Carlo Integration Methods." *Journal of Econometrics*, **123**(2), 201–225. doi:10.1016/j.jeconom.2003.12.002. Special issue on recent advances in Bayesian econometrics. Cited on pages 4 and 5.

Bauwens L, Lubrano M (1998). "Bayesian Inference on GARCH Models Using the Gibbs Sampler." *The Econometrics Journal*, **1**(1), C23–C46. doi:10.1111/1368-423X.11003. Cited on pages 4 and 5.

Bauwens L, Preminger A, Rombouts JVK (2006). "Regime Switching GARCH Models." *Discussion paper 2006/11*, CORE and Department of Economics, Université catholique de Louvain. Cited on pages 4 and 159.

Bauwens L, Rombouts JVK (2007). "Bayesian Inference for the Mixed Conditional Heteroskedasticity Model." *The Econometrics Journal*, **10**(2), 408–425. doi:10.1111/j.1368-423X.2007.00213.x. Cited on page 4.

Berg A, Meyer R, Yu J (2004). "Deviance Information Criterion for Comparing Stochastic Volatility Models." *Journal of Business and Economic Statistics*, **22**(1), 107–120. Cited on page 134.

Berkowitz J (2001). "Testing Density Forecasts, with Applications to Risk Management." *Journal of Business and Economic Statistics*, **19**(4), 465–474. Cited on page 133.

Berkowitz J, Christoffersen PF, Pelletier D (2006). "Evaluating Value-at-Risk Model with Desk-Level Data." *Working Paper Series 010*, Department of Economics, North Carolina State University. URL http://ideas.repec.org/p/ncs/wpaper/010.html. Cited on page 158.

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. Cited on pages 2 and 39.

Bollerslev T (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, **31**(3), 307–327. doi:10.1016/0304-4076(86)90063-1. Cited on pages 1, 3, and 36.

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. Cited on page 2.

Bollerslev T, Engle RF, Nelson DB (1994). "ARCH Models." In RF Engle, DL McFadden (eds.), "Handbook of Econometrics," volume 4, chapter 49, pp. 2959–3038. North Holland, Amsterdam, NL, first edition. ISBN 0444887660. doi:10.1016/S1573-4412(05)80019-4. Cited on page 2.

Bollerslev T, Ghysels E (1996). "Periodic Autoregressive Conditional Heteroscedasticity." *Journal of Business and Economic Statistics*, **14**(2), 139–151. Cited on page 22.

Campbell JY, Lo AW, MacKinlay AC (1996). *The Econometrics of Financial Markets*. Princeton University Press, Princeton, USA, first edition. ISBN 0691043019. Cited on page 46.

Casella G, George EI (1992). "Explaining the Gibbs Sampler." *The American Statistician*, **46**(3), 167–174. Cited on page 11.

Celeux G, Forbes F, Robert CP, Titterington M (2006). "Deviance Information Criterion for Missing Data Models." *Bayesian Analysis*, **1**(4), 651–706. With discussion and rejoinder. Cited on pages 134, 135, and 137.

Chib S (1996). "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models." *Journal of Econometrics*, **75**(1), 79–97. doi:10.1016/0304-4076(95)01770-4. Cited on pages 116, 117, 133, and 135.

Chib S, Greenberg E (1994). "Bayes Inference in Regression Models with ARMA(p,q) Errors." *Journal of Econometrics*, **64**(1–2), 183–206. doi:10.1016/0304-4076(94)90063-9. Cited on page 20.

Chib S, Greenberg E (1995). "Understanding the Metropolis-Hasting Algorithm." *The American Statistician*, **49**(4), 327–335. Cited on page 13.

Chib S, Greenberg E (1996). "Markov Chain Monte Carlo Simulation Methods in Econometrics." *Econometric Theory*, **12**(3), 409–431. Cited on pages 9 and 11.

Christoffersen PF (1998). "Evaluating Interval Forecasts." *International Economic Review*, **39**(4), 841–862. Cited on pages 74, 102, 103, 104, 145, and 149.

Christoffersen PF, Diebold FX (1996). "Further Results on Forecasting and Model Selection Under Asymmetric Loss." *Journal of Applied Econometrics*, **11**(5), 561–571. doi:10.1002/(SICI)1099-1255(199609)11:5<561::AID-JAE406>3.0.CO;2-S. Special issue in econometric forecasting. Cited on page 86.

Christoffersen PF, Diebold FX (1997). "Optimal Prediction under Asymmetric Loss." *Econometric Theory*, **13**(6), 808–817. Cited on page 86.

Christoffersen PF, Gonçalves S (2004). "Estimation Risk in Financial Risk Management." *CIRANO Working Papers 2004s–15*, CIRANO. URL `http://ideas.repec.org/p/cir/cirwor/2004s-15.html`. Cited on page 75.

Cornish EA, Fisher RA (1937). "Moments and Cumulants in the Specification of Distributions." *Revue de l'Institut International de Statistique*, **4**, 1–14. Reprinted in: Contributions to Mathematical Statistics, Fisher, R. A. (Ed.), New York: Whiley, 1950. Cited on pages 76 and 79.

Cowles MK, Carlin BP (1996). "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review." *Journal of the American Statistical Association*, **91**(434), 883–904. Cited on page 14.

Dempster AP (1997). "The Direct Use of Likelihood for Significance Testing." *Statistics and Computing*, **7**(4), 247–252. doi:10.1023/A:1018598421607. Cited on page 134.

Dempster AP, Laird NM, Rubin DB (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society*, **39**(1), 1–38. Cited on page 152.

Deschamps PJ (2006). "A Flexible Prior Distribution for Markov Switching Autoregressions with Student-t Errors." *Journal of Econometrics*, **133**(1), 153–190. doi:10.1016/j.jeconom.2005.03.012. Cited on pages 14, 26, 58, 63, 115, 126, 166, 167, and 179.

Diebold FX, Gunther TA, Tay AS (1998). "Evaluating Density Forecasts with Applications to Financial Risk Management." *International Economic Review*, **39**(4), 863–883. Cited on page 133.

Diebold FX, Mariano RS (1995). "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, **13**(3), 253–263. Cited on page 102.

Diebolt J, Robert CP (1994). "Estimation of Finite Mixture Distributions through Bayesian Sampling." *Journal of the Royal Statistical Society*, **56**(2), 363–375. Cited on page 114.

Dueker MJ (1997). "Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility." *Journal of Business and Economic Statistics*, **15**(1), 26–34. Cited on pages 2, 109, and 159.

Engle RF (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, **50**(4), 987–1008. Cited on page 1.

Engle RF (2004). "Risk and Volatility: Econometric Models and Financial Practice." *The American Economic Review*, **94**(3), 405–420. doi:10.1257/0002828041464597. Cited on pages 2 and 83.

Fernández C, Steel MF (1998). "On Bayesian Modeling of Fat Tails and Skewness." *Journal of the American Statistical Association*, **93**(441), 359–371. Cited on page 140.

Frühwirth-Schnatter S (2001a). "Fully Bayesian Analysis of Switching Gaussian State Space Models." *Annals of the Institute of Statistical Mathematics*, **53**(1), 31–49. Special issue on nonlinear non-Gaussian models and related filtering methods. Cited on page 139.

Frühwirth-Schnatter S (2001b). "Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models." *Journal of the American Statistical Association*, **96**(453), 194–209. Cited on pages 7, 113, 116, 125, 129, and 157.

Frühwirth-Schnatter S (2004). "Estimating Marginal Likelihoods for Mixture and Markov Switching Models Using Bridge Sampling Techniques." *The Econometrics Journal*, **7**(1), 143–167. doi:10.1111/j.1368-423X.2004.00125.x. Cited on page 138.

Frühwirth-Schnatter S (2006). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. Springer Verlag, New York, USA, first edition. ISBN 0387329099. Cited on pages 3, 117, and 118.

Gallant AR, Tauchen G (1989). "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications." *Econometrica*, **57**(5), 1091–1120. Cited on page 2.

Gelfand AE, Dey DK (1994). "Bayesian Model Choice: Asymptotics and Exact Calculations." *Journal of the Royal Statistical Society*, **56**(3), 501–514. Cited on page 139.

Gelfand AE, Smith AFM (1990). "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association*, **85**(410), 398–409. Cited on page 11.

Gelman A (1995). "Inference and Monitoring Convergence." In WR Gilks, S Richardson, DJ Spiegelhalter (eds.), "Markov Chain Monte Carlo in Prac-

tice," chapter 8, pp. 131–143. Chapman and Hall, London, UK, first edition. ISBN 0412055511. Cited on page 14.

Gelman A, Rubin DB (1992). "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science*, **7**(4), 457–472. Cited on pages 14, 24, 25, 27, 47, 65, 94, and 123.

Geman S, Geman D (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741. Cited on page 11.

Gerlach RH, Carter C, Kohn R (1999). "Diagnostics for Time Series Analysis." *Journal of Time Series Analysis*, **20**(3), 309–330. doi:10.1111/1467-9892.00139. Cited on page 144.

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. Cited on pages 3 and 4.

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

Geweke JF (1991). "Efficient Simulation From the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities." In EM Keramidas, SM Kaufman (eds.), "Computing Science and Statistics: The 23rd Symposium on the Interface," pp. 571–578. Seattle, Washington, USA. Special issue on critical applications of scientific computing. Cited on page 141.

Geweke JF (1992). "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments." In JO Berger, JM Bernardo, AP Dawid, AFM Smith (eds.), "Bayesian Statistics," volume 4, pp. 169–194. Oxford University Press, Oxford, UK. ISBN 0198522665. Fourth Valencia international meeting. Cited on page 14.

Geweke JF (1993). "Bayesian Treatment of the Independent Student-t Linear Model." *Journal of Applied Econometrics*, **8**(S1), S19–S40. doi:10.1002/jae.3950080504. Special issue on econometric inference using simulation techniques. Cited on pages 6, 55, 57, 64, 166, 167, and 179.

Geweke JF (1999). "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication." *Econometric Reviews*, **18**(1), 1–73. doi:10.1080/07474939908800428. Cited on page 30.

Geweke JF (2004). "Getting it Right: Joint Distribution Tests of Posterior Simulators." *Journal of the American Statistical Association*, **99**(467), 799–804. Cited on pages 22 and 179.

Gilks WR, Wild P (1992). "Adaptive Rejection Sampling for Gibbs Sampling." *Applied Statistics*, **41**(2), 337–348. Cited on page 12.

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. Cited on pages 2, 6, and 39.

Gray SF (1996). "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process." *Journal of Financial Economics*, **42**(1), 27–62. doi:10.1016/0304-405X(96)00875-6. Cited on pages 109, 117, and 159.

Guidolin M, Timmermann A (2006). "Term Stucture of Risk under Alternative Econometric Specifications." *Journal of Econometrics*, **131**(1–2), 285–308. doi:10.1016/j.jeconom.2005.01.033. Cited on pages 91 and 92.

Haas M, Mittnik S, Paolella MS (2004). "A New Approach to Markov-Switching GARCH Models." *Journal of Financial Econometrics*, **2**(4), 493–530. doi:10.1093/jjfinec/nbh020. Cited on pages 2, 7, 110, 112, 126, 128, 133, and 157.

Hamilton JD (1994). *Time Series Analysis*. Princeton University Press, Princeton, USA, first edition. ISBN 0691042896. Cited on pages 114 and 152.

Hamilton JD, Susmel R (1994). "Autoregressive Conditional Heteroskedasticity and Changes in Regime." *Journal of Econometrics*, **64**(1–2), 307–333. doi:10.1016/0304-4076(94)90067-1. Cited on pages 2, 3, 109, and 152.

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. Cited on page 10.

He C, Teräsvirta T (1999). "Properties of Moments of a Family of GARCH Processes." *Journal of Econometrics*, **92**(1), 173–192. doi:10.1016/S0304-4076(98)00089-X. Cited on page 71.

Highfield RA, Zellner A (1988). "Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions." *Journal of Econometrics*, **37**(2), 195–209. doi:10.1016/0304-4076(88)90002-4. Cited on page 83.

Hwang S, Knight J, Satchell SE (1999). "Forecasting Volatility Using Linex Loss Functions." *Working paper 99-07*, Financial Econometrics Research Center, Warwick Business School. Cited on page 86.

Hwang S, Knight J, Satchell SE (2001). "Forecasting Nonlinear Functions of Returns Using Linex Loss Functions." *Annals of Economics and Finance*, **2**(1), 187–213. Cited on page 86.

Jacquier E, Polson NG, Rossi PE (1994). "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business and Economic Statistics*, **12**(4), 371–389. Cited on page 158.

Jaschke SR (2002). "The Cornish-Fisher Expansion in the Context of Delta-Gamma-Normal Approximations." *Journal of Risk*, **4**(4), 33–52. Cited on page 80.

Kass RE, Carlin BP, Gelman A, Neal RM (1998). "Markov Chain Monte Carlo in Practice: A Roundtable Discussion." *The American Statistician*, **52**(2), 93–100. Cited on page 13.

Kass RE, Raftery AE (1995). "Bayes Factors." *Journal of the American Statistical Association*, **90**(430), 773–795. Cited on pages 31 and 141.

Kaufmann R (2004). *Long-Term Risk Management*. Phd thesis, Swiss Federal Institute of Technology Zürich. Dissertation ETH No. 15595. Cited on page 158.

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. Cited on pages 4, 117, 133, 135, and 139.

Kaufmann S, Scheicher M (2006). "A Switching ARCH Model for the German DAX Index." *Studies in Nonlinear Dynamics and Econometrics*, **10**(4), 1–35. Article nr. 3, URL `http://www.bepress.com/snde/vol10/iss4/art3/`. Cited on pages 4 and 139.

Kim S, Shephard N, Chib S (1998). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies*, **65**(3), 361–393. doi:10.1111/1467-937X.00050. Cited on pages 133 and 158.

Klaassen F (2002). "Improving GARCH Volatility Forecasts with Regime-Switching GARCH." *Empirical Economics*, **27**(2), 363–394. doi:10.1007/s001810100100. Cited on pages 2, 109, 112, and 117.

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. Cited on page 4.

Knight J, Satchell S, Wang G (2003). "Value at Risk Linear Exponent Forecasts -VARLINEX-." *Quantitative Finance*, **3**, 332–344. Cited on pages 86 and 97.

Koop G (2003). *Bayesian Econometrics*. Wiley-Interscience, London, UK, first edition. ISBN 0470845678. Cited on page 9.

Kupiec PH (1995). "Techniques for Verifying the Accuracy of Risk Measurement Models." *Journal of Derivatives*, **3**, 73–84. Cited on page 74.

Lamoureux CG, Lastrapes WD (1990). "Persistence in Variance, Structural Change, and the GARCH Model." *Journal of Business and Economic Statistics*, **8**(2), 225–243. Cited on pages 2 and 109.

Lee SW, Hansen BE (1994). "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator." *Econometric Theory*, **10**(1), 29–52. Cited on page 2.

Ljung GM, Box GEP (1978). "On a Measure of Lack of Fit in Time Series Models." *Biometrika*, **65**(2), 297–303. doi:10.1093/biomet/65.2.297. Cited on pages 32 and 53.

Marcucci J (2005). "Forecasting Stock Market Volatility with Regime-Switching GARCH Models." *Studies in Nonlinear Dynamics and Econometrics*, **9**(4), 1–53. Article nr. 6, URL `http://www.bepress.com/snde/vol9/iss4/art6/`. Cited on pages 2 and 109.

McNeil AJ, Frey R (2000). "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach." *Journal of Empirical Finance*, **7**(3–4), 271–300. doi:10.1016/S0927-5398(00)00012-8. Cited on page 74.

Meng XL, Wong WH (1996). "Simulating Ratios of Normalizing Constants via a Simple Identity: A Theoretical Exploration." *Statistica Sinica*, **6**, 831–860. Cited on pages 7, 138, 139, and 157.

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. Cited on pages 10 and 13.

Miazhynskaia T, Aussenegg W (2006). "Uncertainty in Value at Risk Estimates under Parametric and Non-Parametric Modeling." *Financial Markets and Portfolio Management*, **20**(3), 243–264. doi:10.1007/s11408-006-0020-8. Cited on pages 75 and 92.

Müller P, Pole A (1998). "Monte Carlo Posterior Integration in GARCH Models." *Sankhya: The Indian Journal of Statistics*, **60**, 127–144. Cited on page 4.

Nakatsuma T (1998). "A Markov-Chain Sampling Algorithm for GARCH Models." *Studies in Nonlinear Dynamics and Econometrics*, **3**(2), 107–117. Algorithm nr.1, URL `http://www.bepress.com/snde/vol3/iss2/algorithm1/`. Cited on pages 4, 5, 6, 7, 17, 18, 19, 116, 155, 156, 157, 158, and 179.

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. Cited on pages 4, 5, 6, 7, 17, 18, 19, 116, 155, 156, 157, 158, and 179.

Nelson DB (1990). "Stationarity and Persistence in the GARCH(1,1) Model." *Econometric Theory*, **6**(3), 318–334. Cited on page 36.

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

Newey WK, West KD (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, **55**(3), 703–708. Cited on page 14.

Pagan AR, Schwert GW (1990). "Alternative Models for Conditional Stock Variability." *Journal of Econometrics*, **45**(1–2), 267–290. doi:10.1016/0304-4076(90)90101-X. Cited on page 2.

Philippe A, Robert CP (2003). "Perfect Simulation of Positive Gaussian Distributions." *Computer Science and Mathematics and Statistics*, **13**(2), 179–186. doi:10.1023/A:1023264710933. Cited on page 140.

Phillips PCB, Perron P (1988). "Testing for a Unit Root in Time Series Regression." *Biometrika*, **75**(2), 335–346. doi:10.1093/biomet/75.2.335. Cited on pages 22 and 123.

Politis DN, Romano JP (1994). "The Stationary Bootstrap." *Journal of the American Statistical Association*, **89**(428), 1303–1313. Cited on page 135.

Politis DN, White H (2004). "Automatic Block-Length Selection for the Dependent Bootstrap." *Econometric Reviews*, **23**(1), 53–70. doi:10.1081/ETC-120028836. Cited on page 136.

Price KV, Storn RM, Lampinen JA (2006). *Differential Evolution: A Practical Approach to Global Optimization*. Springer Verlag, Berlin, Germany. ISBN 3540209506. Cited on page 179.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Version 2.4.1, URL `http://www.R-project.org/`. Cited on pages 21 and 179.

Ripley B (1987). *Stochastic Simulation*. Probability and Mathematical Statistics. Whiley, New York, USA, first edition. ISBN 0471818844. Cited on page 12.

RiskMetrics Group (1996). *RiskMetrics Technical Document*. J. P. Morgan/Reuters, fourth edition. URL `http://www.riskmetrics.com/rmcovv.html`. Cited on page 96.

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. Cited on pages 5 and 12.

Robert CP (1995). "Simulation of Truncated Normal Variables." *Statistics and Computing*, **5**(2), 121–125. doi:10.1007/BF00143942. Cited on page 140.

Roberts GO, Smith AFM (1994). "Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithm." *Stochastic Processes and their Applications*, **49**(2), 207–216. doi:10.1016/0304-4149(94)90134-1. Cited on pages 11 and 12.

Rosenblatt M (1952). "Remarks on a Multivariate Transformation." *Annals of Mathematical Statistics*, **23**, 470–472. Cited on page 133.

Schwarz G (1978). "Estimating the Dimension of a Model." *The Annals of Statistics*, **6**(2), 461–464. Cited on page 136.

Seiler D (2006). *Backtesting Multiple-Period Forecasting Models*. Master thesis, Swiss Federal Institute of Technology Zürich. Cited on page 158.

Silverman BW (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, USA, first edition. ISBN 0412246201. Cited on pages 15, 37, 141, 143, 151, and 153.

Smith AFM, Roberts GO (1993). "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods." *Journal of the Royal Statistical Society*, **55**(1), 3–24. Cited on pages 9 and 13.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002). "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society*, **64**(4), 583–639. doi:10.1111/1467-9868.00353. Cited on pages 7, 134, and 157.

Spiegelhalter DJ, Thomas A, Best NG, Gilks WR (1995). *BUGS: Bayesian Inference Using Gibbs Sampling*. The BUGS project. Version 0.50. Cited on page 5.

Spiegelhalter DJ, Thomas A, Best NG, Lunn D (2007). *WinBUGS User's manual*. The BUGS project. Version 1.4.1, URL `http://www.mrc-bsu.cam.ac.uk/bugs/`. Cited on page 5.

Tanner MA, Wong WH (1987). "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association*, **82**(398), 528–540. Cited on page 11.

Thompson RD, Basu AP (1995). "Asymmetric Loss Functions for Estimating System Reliability." In DA Berry, KM Chaloner, JF Geweke (eds.), "Bayesian Analysis in Statistics and Econometrics," Probability and Statistics, pp. 471–482. John Wiley and Sons Ltd., New York, USA, first edition. ISBN 0471118567. Essays in honor of Arnold Zellner. Cited on page 90.

Tierney L (1994). "Markov Chains for Exploring Posterior Distributions." *The Annals of Statistics*, **22**(4), 1701–1762. With discussion. Cited on pages 9, 10, 11, 12, 13, and 14.

Varian HR (1974). "A Bayesian Approach to Real Estate Assessment." In SE Fienberg, A Zellner (eds.), "Studies in Bayesian Econometrics and Statistics," Number 86 in Contributions to economic analysis, pp. 195–208. North Holland, Amsterdam, NL, first edition. Cited on page 86.

Vrontos ID, Dellaportas P, Politis DN (2000). "Full Bayesian Inference for GARCH and EGARCH Models." *Journal of Business and Economic Statistics*, **18**(2), 187–198. Cited on page 4.

Zellner A (1986). "Bayesian Estimation and Prediction Using Asymmetric Loss Functions." *Journal of the American Statistical Association*, **81**(394), 446–451. Cited on page 87.

Zhu L, Carlin BP (2000). "Comparing Hierarchical Models for Spatio-Temporally Misaligned Data Using the Deviance Information Criterion." *Statistics in Medicine*, **19**, 2265–2278. doi:10.1002/1097-0258(20000915/30)19:17/18<2265::AID-SIM568>3.0.CO;2-6. Cited on page 135.

Zumbach G (2006). "Backtesing Risk Methodologies from one Day to one Year." *The Journal of Risk*, **9**(2), 0–36. Cited on pages 104 and 158.

Index

Lecture Notes in Economics and Mathematical Systems

For information about Vols. 1–524
please contact your bookseller or Springer-Verlag

Vol. 525: S. Spinler, Capacity Reservation for Capital-Intensive Technologies. XVI, 139 pages. 2003.

Vol. 526: C. F. Daganzo, A Theory of Supply Chains. VIII, 123 pages. 2003.

Vol. 527: C. E. Metz, Information Dissemination in Currency Crises. XI, 231 pages. 2003.

Vol. 528: R. Stolletz, Performance Analysis and Optimization of Inbound Call Centers. X, 219 pages. 2003.

Vol. 529: W. Krabs, S. W. Pickl, Analysis, Controllability and Optimization of Time-Discrete Systems and Dynamical Games. XII, 187 pages. 2003.

Vol. 530: R. Wapler, Unemployment, Market Structure and Growth. XXVII, 207 pages. 2003.

Vol. 531: M. Gallegati, A. Kirman, M. Marsili (Eds.), The Complex Dynamics of Economic Interaction. XV, 402 pages, 2004.

Vol. 532: K. Marti, Y. Ermoliev, G. Pflug (Eds.), Dynamic Stochastic Optimization. VIII, 336 pages. 2004.

Vol. 533: G. Dudek, Collaborative Planning in Supply Chains. X, 234 pages. 2004.

Vol. 534: M. Runkel, Environmental and Resource Policy for Consumer Durables. X, 197 pages. 2004.

Vol. 535: X. Gandibleux, M. Sevaux, K. Sörensen, V. T'kindt (Eds.), Metaheuristics for Multiobjective Optimisation. IX, 249 pages. 2004.

Vol. 536: R. Brüggemann, Model Reduction Methods for Vector Autoregressive Processes. X, 218 pages. 2004.

Vol. 537: A. Esser, Pricing in (In)Complete Markets. XI, 122 pages, 2004.

Vol. 538: S. Kokot, The Econometrics of Sequential Trade Models. XI, 193 pages. 2004.

Vol. 539: N. Hautsch, Modelling Irregularly Spaced Financial Data. XII, 291 pages. 2004.

Vol. 540: H. Kraft, Optimal Portfolios with Stochastic Interest Rates and Defaultable Assets. X, 173 pages. 2004.

Vol. 541: G.-y. Chen, X. Huang, X. Yang, Vector Optimization. X, 306 pages. 2005.

Vol. 542: J. Lingens, Union Wage Bargaining and Economic Growth. XIII, 199 pages. 2004.

Vol. 543: C. Benkert, Default Risk in Bond and Credit Derivatives Markets. IX, 135 pages. 2004.

Vol. 544: B. Fleischmann, A. Klose, Distribution Logistics. X, 284 pages. 2004.

Vol. 545: R. Hafner, Stochastic Implied Volatility. XI, 229 pages. 2004.

Vol. 546: D. Quadt, Lot-Sizing and Scheduling for Flexible Flow Lines. XVIII, 227 pages. 2004.

Vol. 547: M. Wildi, Signal Extraction. XI, 279 pages. 2005.

Vol. 548: D. Kuhn, Generalized Bounds for Convex Multistage Stochastic Programs. XI, 190 pages. 2005.

Vol. 549: G. N. Krieg, Kanban-Controlled Manufacturing Systems. IX, 236 pages. 2005.

Vol. 550: T. Lux, S. Reitz, E. Samanidou, Nonlinear Dynamics and Heterogeneous Interacting Agents. XIII, 327 pages. 2005.

Vol. 551: J. Leskow, M. Puchet Anyul, L. F. Punzo, New Tools of Economic Dynamics. XIX, 392 pages. 2005.

Vol. 552: C. Suerie, Time Continuity in Discrete Time Models. XVIII, 229 pages. 2005.

Vol. 553: B. Mönch, Strategic Trading in Illiquid Markets. XIII, 116 pages. 2005.

Vol. 554: R. Foellmi, Consumption Structure and Macro-economics. IX, 152 pages. 2005.

Vol. 555: J. Wenzelburger, Learning in Economic Systems with Expectations Feedback (planned) 2005.

Vol. 556: R. Branzei, D. Dimitrov, S. Tijs, Models in Cooperative Game Theory. VIII, 135 pages. 2005.

Vol. 557: S. Barbaro, Equity and Efficiency Considerations of Public Higer Education. XII, 128 pages. 2005.

Vol. 558: M. Faliva, M. G. Zoia, Topics in Dynamic Model Analysis. X, 144 pages. 2005.

Vol. 559: M. Schulmerich, Real Options Valuation. XVI, 357 pages. 2005.

Vol. 560: A. von Schemde, Index and Stability in Bimatrix Games. X, 151 pages. 2005.

Vol. 561: H. Bobzin, Principles of Network Economics. XX, 390 pages. 2006.

Vol. 562: T. Langenberg, Standardization and Expectations. IX, 132 pages. 2006.

Vol. 563: A. Seeger (Ed.), Recent Advances in Optimization. XI, 455 pages. 2006.

Vol. 564: P. Mathieu, B. Beaufils, O. Brandouy (Eds.), Artificial Economics. XIII, 237 pages. 2005.

Vol. 565: W. Lemke, Term Structure Modeling and Estimation in a State Space Framework. IX, 224 pages. 2006.

Vol. 566: M. Genser, A Structural Framework for the Pricing of Corporate Securities. XIX, 176 pages. 2006.

Vol. 567: A. Namatame, T. Kaizouji, Y. Aruga (Eds.), The Complex Networks of Economic Interactions. XI, 343 pages. 2006.

Vol. 568: M. Caliendo, Microeconometric Evaluation of Labour Market Policies. XVII, 258 pages. 2006.

Vol. 569: L. Neubecker, Strategic Competition in Oligopolies with Fluctuating Demand. IX, 233 pages, 2006.

Vol. 570: J. Woo, The Political Economy of Fiscal Policy. X, 169 pages. 2006.

Vol. 571: T. Herwig, Market-Conform Valuation of Options. VIII, 104 pages. 2006.

Vol. 572: M. F. Jäkel, Pensionomics. XII, 316 pages. 2006

Vol. 573: J. Emami Namini, International Trade and Multinational Activity. X, 159 pages, 2006.

Vol. 574: R. Kleber, Dynamic Inventory Management in Reverse Logistics. XII, 181 pages, 2006.

Vol. 575: R. Hellermann, Capacity Options for Revenue Management. XV, 199 pages, 2006.

Vol. 576: J. Zajac, Economics Dynamics, Information and Equilibnum. X, 284 pages, 2006.

Vol. 577: K. Rudolph, Bargaining Power Effects in Financial Contracting. XVIII, 330 pages, 2006.

Vol. 578: J. Kühn, Optimal Risk-Return Trade-Offs of Commercial Banks. IX, 149 pages, 2006.

Vol. 579: D. Sondermann, Introduction to Stochastic Calculus for Finance. X, 136 pages, 2006.

Vol. 580: S. Seifert, Posted Price Offers in Internet Auction Markets. IX, 186 pages, 2006.

Vol. 581: K. Marti; Y. Ermoliev; M. Makowsk; G. Pflug (Eds.), Coping with Uncertainty. XIII, 330 pages, 2006.

Vol. 582: J. Andritzky, Sovereign Default Risks Valuation: Implications of Debt Crises and Bond Restructurings. VIII, 251 pages, 2006.

Vol. 583: I.V. Konnov, D.T. Luc, A.M. Rubinov† (Eds.), Generalized Convexity and Related Topics. IX, 469 pages, 2006.

Vol. 584: C. Bruun, Adances in Artificial Economics: The Economy as a Complex Dynamic System. XVI, 296 pages, 2006.

Vol. 585: R. Pope, J. Leitner, U. Leopold-Wildburger, The Knowledge Ahead Approach to Risk. XVI, 218 pages, 2007 (planned).

Vol. 586: B.Lebreton, Strategic Closed-Loop Supply Chain Management. X, 150 pages, 2007 (planned).

Vol. 587: P. N. Baecker, Real Options and Intellectual Property: Capital Budgeting Under Imperfect Patent Protection. X, 276 pages , 2007.

Vol. 588: D. Grundel, R. Murphey, P. Panos , O. Prokopyev (Eds.), Cooperative Systems: Control and Optimization. IX, 401 pages , 2007.

Vol. 589: M. Schwind, Dynamic Pricing and Automated Resource Allocation for Information Services: Reinforcement Learning and Combinatorial Auctions. XII, 293 pages , 2007.

Vol. 590: S. H. Oda, Developments on Experimental Economics: New Approaches to Solving Real-World Problems. XVI, 262 pages, 2007.

Vol. 591: M. Lehmann-Waffenschmidt, Economic Evolution and Equilibrium: Bridging the Gap. VIII, 272 pages, 2007.

Vol. 592: A. C.-L. Chian, Complex Systems Approach to Economic Dynamics. X, 95 pages, 2007.

Vol. 593: J. Rubart, The Employment Effects of Technological Change: Heterogenous Labor, Wage Inequality and Unemployment. XII, 209 pages, 2007.

Vol. 594: R. Hübner, Strategic Supply Chain Management in Process Industries: An Application to Specialty Chemicals Production Network Design. XII, 243 pages, 2007.

Vol. 595: H. Gimpel, Preferences in Negotiations: The Attachment Effect. XIV, 268 pages, 2007.

Vol. 596: M. Müller-Bungart, Revenue Management with Flexible Products: Models and Methods for the Broadcasting Industry. XXI, 297 pages, 2007.

Vol. 597: C. Barz, Risk-Averse Capacity Control in Revenue Management. XIV, 163 pages, 2007.

Vol. 598: A. Ule, Partner Choice and Cooperation in Networks: Theory and Experimental Evidence. Approx. 200 pages, 2007.

Vol. 599: A. Consiglio, Artificial Markets Modeling: Methods and Applications. XV, 277 pages, 2007.

Vol. 600: M. Hickman, P. Mirchandani, S. Voss (Eds.): Computer-Aided Scheduling of Public Transport. Approx. 424 pages, 2007.

Vol. 601: D. Radulescu, CGE Models and Capital Income Tax Reforms: The Case of a Dual Income Tax for Germany. XVI, 168 pages, 2007.

Vol. 602: N. Ehrentreich, Agent-Based Modeling: The Santa Fe Institute Artificial Stock Market Model Revisited. XVI, 225 pages, 2007.

Vol. 603: D. Briskorn, Sports Leagues Scheduling: Models, Combinatorial Properties, and Optimization Algorithms. XII, 164 pages, 2008.

Vol. 604: D. Brown, F. Kubler, Computational Aspects of General Equilibrium Theory: Refutable Theories of Value. XII, 202 pages, 2008.

Vol. 605: M. Puhle, Bond Portfolio Optimization. XIV, 137 pages, 2008.

Vol. 606: S. von Widekind, Evolution of Non-Expected Utility Preferences. X, 130 pages, 2008.

Vol. 607: M. Bouziane, Pricing Interest Rate Derivatives: A Fourier-Transform Based Approach. XII, 191 pages, 2008.

Vol. 608: P. Nicola, Experimenting with Dynamic Macromodels: Growth and Cycles. XIII, 241 pages, 2008.

Vol. 609: X. Fang, K.K. Lai, S. Wang, Fuzzy Portfolio Optimization: Theory and Models. IX, 173 pages, 2008.

Vol. 610: M. Hillebrand, Pension Systems, Demographic Change, and the Stock Market. X, 176 pages, 2008 .

Vol. 611: R. Brosch, Portfolios of Real Options. XVI, 174 pages, 2008.

Vol. 612: D. Ardia, Financial Risk Management with Bayesian Estimation of GARCH Models. XII, 203 pages, 2008.