BAYESIAN INFERENCE ON GARCH MODELS USING THE GIBBS SAMPLER *

Luc Bauwens[1] and Michel Lubrano[2]

July 1996

Abstract

This paper explains how the Gibbs sampler can be used to perform Bayesian inference on GARCH models. Although the Gibbs sampler is usually based on the *analytical* knowledge of the full conditional posterior densities, such knowledge is not available in regression models with GARCH errors. We show that the Gibbs sampler can be combined with a unidimensional deterministic integration rule applied to each coordinate of the posterior density. The full conditional densities are evaluated and inverted numerically to obtain random draws of the joint posterior. The method is shown to be feasible and competitive compared to importance sampling and the Metropolis-Hastings algorithm. It is applied to estimate an asymmetric GARCH model for the return on a stock exchange index, and to compute predictive option prices on the index.

*Support of the European Commission Human Capital and Mobility Program through the network "Econometric inference using simulation methods" is gratefully acknowledged. While remaining responsible for any error, the authors wish to thank Ana Justel and seminar participants at the University of Toronto, Erasmus University Rotterdam, and Humboldt-University-Berlin, for useful remarks and suggestions.

[1]CORE, Université catholique de Louvain; email: bauwens@core.ucl.ac.be.
[2]GREQAM, EHESS-CNRS; email: milu@ehess.cnrs-mrs.fr.

1 Introduction

Bayesian inference on ARCH models has been implemented using importance sampling - see Geweke (1989) or Kleibergen and van Dijk (1993) - and more recently using a Metropolis algorithm - see Geweke (1994). The drawback of this kind of methods is that they require a careful choice of a good approximation to the posterior density. Tuning is often necessary so that the method cannot be 'automatic'. Moreover, as they are purely based on a sampling approach, nice looking graphs for the marginals require a great number of draws.

The Gibbs sampler - see Geweke (1995) for a survey - uses analytical properties of the posterior density as much as possible. Gibbs sampling is simple when the vector of parameters, say θ, can be split into θ_1, θ_2 and the analytical expressions of both conditional posterior densities are known. Sampling alternatively in each conditional posterior density produces a Markov chain which converges to a draw of the joint posterior density. In some cases the conditional posterior density of $\theta_1|\theta_2$ is known, but that of $\theta_2|\theta_1$ is not. So direct sampling in $\varphi(\theta_1|\theta_2, y)$ is easy, but sampling in $\varphi(\theta_2|\theta_1, y)$ is more complicated. The traditional solution is then to use a rejection technique for sampling in the posterior density, for example the Metropolis-Hastings algorithm.

GARCH models correspond to a particular form of heteroskedasticity. A multiplicative heteroskedasticity model is written as:

$$\begin{cases} y_t = x_t'\beta + u_t \\ u_t|I_{t-1} \sim N(0, \sigma_t^2) \\ \sigma_t^2 = h(z_t, \gamma) \end{cases} \tag{1}$$

A Gibbs sampler can be combined with a rejection step in this model. As soon as z_t and γ are given, the posterior density of β is a Student density (assuming a flat prior). However, the posterior density of γ given β is not of a known form: a Metropolis algorithm for instance has to be used for sampling in this posterior. A GARCH model is complicated by the fact that $z_t = (y_{t-1} - x_{t-1}'\beta)^2$. So there is no separability between the regression parameters and the heteroskedastic parameters. The previous algorithm cannot be used.

The aim of this paper is to use the Griddy-Gibbs sampler of Ritter and Tanner (1992) in order to conduct Bayesian inference on GARCH models. In Section 2, we explain the algorithm and we illustrate its feasibility in GARCH models with Student errors. In the third section, we compare the efficiency of the method to importance sampling and to the Metropolis-Hastings algorithm. In Section 4, we discuss briefly the issues of persistence and asymmetry of shocks on the conditional variance. In Section 5 we apply an asymmetric model to a series of returns on a stock market index and we compute predictive option prices. The last section concludes.

2 Griddy-Gibbs sampler for GARCH models

We want to conduct Bayesian inference on a regression model with GARCH errors. To ease exposition, we consider a parsimonious model which is quite representative of GARCH models used in finance. It is a GARCH(1,1) model with Student errors. As reported in Bollerslev, Chou and Kroner (1992), the Student density, with its fat tails, is more able than the Normal to account for the excess kurtosis present in financial data.

$$\begin{cases} y_t = \mu + \rho y_{t-1} + u_t \\ u_t | I_{t-1} \sim \text{Student}(0, h_t, \nu) \\ h_t = \omega + \alpha u_{t-1}^2 + \beta h_{t-1} \end{cases} \tag{2}$$

The distribution of u_t, given the past information I_{t-1}, is Student with mean zero, and variance $h_t\nu/(\nu-2)$ (assuming $\nu > 2$). The u_t sequence is independent. The initial variance h_0 is treated as a known constant. The parameters of the variance equation are restricted to be positive.

Let θ denote the parameter vector $(\mu,\ \rho,\ \omega,\ \alpha,\ \beta,\ \nu)$. For a sample of T observations, the posterior density is

$$\varphi(\theta|y) \propto \varphi(\theta) \prod_{t=1}^{T} [\Gamma(\tfrac{\nu+1}{2})/\Gamma(\tfrac{\nu}{2})]\ (\nu h_t)^{-\frac{1}{2}}\ [1 + \frac{u_t^2}{\nu h_t}]^{-\frac{\nu+1}{2}} \tag{3}$$

The prior density $\varphi(\theta)$ can be chosen in any family of densities, provided it respects at least the positivity restrictions on ω, α, β and of course ν.

2.1 The usual Gibbs sampler

The Gibbs sampler produces a sequence $\{\theta^n\}_{n=1}^N$ of dependent draws from the posterior. Let us suppose that we can partition θ into θ_1 and θ_2 so that the posterior densities $\varphi(\theta_1|\theta_2, y)$ and $\varphi(\theta_2|\theta_1, y)$ have known analytical forms (e.g. Student densities) from which it is easy to draw random numbers. This ability to sample from conditional distributions is called conjugacy by Carlin and Gelfand (1991) and is central to the Gibbs sampler. The next draw θ^n of the joint posterior density is generated using the previous draw θ^{n-1} by the following cycle:

$$\begin{aligned} \theta_1^n &\sim \varphi(\theta_1|\theta_2^{n-1}, y) \\ \theta_2^n &\sim \varphi(\theta_2|\theta_1^n, y) \end{aligned} \tag{4}$$

($x^n \sim \varphi(x)$ means that x^n is a draw from the distribution $\varphi(x)$). The procedure is started using any value θ_2^a in the support of the posterior distribution. The influence of the starting conditions vanishes after a certain number of draws, say ns. If the algorithm converges, after ns draws it produces random draws of the posterior density of θ; the sample is however not independent because of the Markov chain structure.

It is rather difficult to devise a convergence criterion for Monte Carlo Markov chains. Very few have been proposed. Zellner and Min (1995) introduce three simple criteria which are valid only for the case of two blocks described in (4). A simple solution proposed by Yu and Mykland (1994) is the visual inspection of CUMSUM statistics. Robert (1995) reports that this criterion behaves relatively well. Suppose we have N draws of a Monte Carlo Markov chain (MCMC). A standardized version of the statistic proposed by Yu and Mykland is:

$$CS_t = \left(\frac{1}{t}\sum_{n=1}^{t}\theta^n - \mu_\theta\right)/\sigma_\theta \tag{5}$$

where μ_θ and σ_θ are the empirical mean and standard deviation of the N draws. If the MCMC converges, the graph of CS_t against t should converge smoothly to zero. On the contrary long and regular excursions away from zero are the indication of an absence of convergence. Standardizing the statistic is convenient if one wants to compare the convergence of different chains or parameters.

There are two options to compute moments and marginal densities. In the first option, we can simply exploit the draws to compute sample moments and histograms. The estimator $\sum_{n=1}^{N}\theta_1^n/N$ of $\mathrm{E}(\theta_1|y)$ converges at the usual slow speed $1/\sqrt{N}$, as in all Monte Carlo techniques. The Monte Carlo variance of this estimator is however not equal to $\mathrm{Var}(\theta_1|y)/N$, as when the draws are independent (and from the exact posterior). When draws are dependent, the numerator of the variance of the estimator has to take account of the correlations between the draws (just as e.g. in ARMA models, the variance of the series is larger than the variance of the innovation). In any case, variance reduction techniques should be used, if possible. As discussed in Tierney (1994), conditioning is one possibility. So in the second option, instead of averaging the θ_1^n sequence to compute the posterior mean of θ_1, we average the conditional means, e.g.:

$$\mathrm{E}(\theta_1|y) = \frac{1}{N}\sum_{n=1}^{N}\mathrm{E}(\theta_1|\theta_2^n, y) \tag{6}$$

The Monte Carlo variance of this estimator is again not equal to $\mathrm{Var}[\mathrm{E}(\theta_1|\theta_2, y)]/N$, as in the case of an independent sample, where it is always smaller than $\mathrm{Var}(\theta_1|y)/N$. With a dependent sample, the numerator has to be adjusted for the correlation structure of the conditional expectation. To quote Tierney: "Whether this form of conditioning leads to a reduction in variance depends on the correlation structure". To compute the marginal variances, we cumulate the conditional variances and the outerproduct of the conditional means, as $\mathrm{Var}(\theta_1) = \mathrm{E}[\mathrm{Var}(\theta_1|\theta_2)] + \mathrm{Var}[\mathrm{E}(\theta_1|\theta_2)]$. For covariances, we can cumulate $\mathrm{E}(\theta_1|\theta_2)\cdot\theta_2'$ instead of $\theta_1\theta_2'$, exploiting $\mathrm{Cov}(\theta_1, \theta_2) = \mathrm{Cov}[\mathrm{E}(\theta_1|\theta_2), \theta_2]$. The graph of a marginal density requires a grid to be defined in advance. Each conditional, properly normalized, is evaluated over this grid, conditionally on the current draw of the other parameter. Conditionals are then averaged: suppose we want to compute $\varphi(\theta_{1*}|y)$, where θ_{1*} denotes one point of the grid of values of θ_1. Since

$$\varphi(\theta_{1*}|y) = \int \varphi(\theta_{1*}|\theta_2, y)\,\varphi(\theta_2|y)\,d\theta_2 \tag{7}$$

the latter quantity is estimated as

$$\sum_{n=1}^{N} \varphi(\theta_{1*}|\theta_2^n, y)/N \tag{8}$$

In our experiments, precise graphs were obtained with $N = 1000$, whereas five to ten times as many draws were required usually to get good-looking histograms with the first option.

2.2 The Griddy-Gibbs sampler

The previous simple algorithm cannot be applied to the GARCH model even if the error term is Normal. Let us write a very simplified ARCH model to demonstrate this:

$$\begin{cases} y_t = \mu + u_t \\ u_t|I_{t-1} \sim N(0, h_t) \\ h_t = 1 + \alpha\,(y_{t-1} - \mu)^2 \end{cases} \tag{9}$$

There are two parameters, μ for the regression and α for the ARCH. Let us assume a uniform prior. The conditional posterior density $\varphi(\mu|\alpha, y)$ has a kernel given by

$$\kappa(\mu|\alpha, y) = \prod_t [h_t(\alpha, \mu)]^{-1/2} \exp -\frac{1}{2}\frac{(y_t - \mu)^2}{h_t(\alpha, \mu)} \tag{10}$$

when α is a given fixed value. The conditional posterior density of α has exactly the same expression, but for given μ. If h_t were fixed, (10) would be a Normal density. But h_t is a function of both α and μ. So the conditional posterior density of μ contains h_t which is also a function of μ. Consequently, it cannot be a Normal or any other well known density from which random numbers could be easily generated. There is no property of conjugacy.

However, the kernel of $\varphi(\mu|\alpha, y)$ (respectively of $\varphi(\alpha|\mu, y)$), conditionally on a previous draw of the conditioning parameter, can be evaluated over a grid of points. One can then compute the corresponding distribution function using a deterministic integration rule - see e.g. Davis and Rabinowitz (1975). Afterwards, one can generate a draw of μ by inversion of the distribution at a random value sampled uniformly in $[0, 1]$. This technique (called the Griddy-Gibbs sampler, hereafter GGS) was applied by Ritter and Tanner (1992) to bivariate posterior densities.

It is possible to improve this algorithm so as to incorporate the above mentioned variance reduction techniques. They are especially important to get accurate graphs of the marginals for a relatively small number of draws. The complete algorithm works as follows for N draws:

- *step 1:* starting point of the chain at a value α^0 for α.

- *step 2:* loop starting at $n = 1$.

- *step 3:* compute $\kappa(\mu|\alpha^{n-1}, y)$ over the grid $(\mu_1, \mu_2, \cdots, \mu_G)$ to obtain the vector $G_\kappa = (\kappa_1, \kappa_2, \cdots, \kappa_G)$.

- *step 4:* by a deterministic integration rule using G points, compute the values $G_\Phi = (0, \Phi_2, \cdots, \Phi_G)$ where

$$\Phi_i = \int_{\mu_1}^{\mu_i} \kappa(\mu|\alpha^{n-1}, y)\, d\mu, \quad i = 2, \cdots, G$$

  Compute (and cumulate for the marginal) the normalized pdf values $G_\varphi = G_\kappa/\Phi_G$ of $\varphi(\mu|\alpha^{n-1}, y)$. Compute $\mathrm{E}(\mu|\alpha^{n-1}, y)$ and $\mathrm{Var}(\mu|\alpha^{n-1}, y)$ by the same type of integration rule as above, and store them in a table.

- *step 5:* generate $u \sim U[0, \Phi_G]$ and invert $\Phi(\mu|\alpha^{n-1}, y)$ by numerical interpolation to get a draw of $\mu|\alpha^{n-1}, y$, indexed μ^n. Store this draw in a table.

- *step 6:* repeat steps 3-6 for $\alpha^n|\mu^n, y$.

- *step 7:* increment n by 1 and go to step 3 unless $n > N$.

- *step 8:* compute the posterior moments of μ and α from the tables where conditional moments are stored (by averaging). Likewise, plot the marginal densities (cumulated G_φ/N). With the table containing $\{\mu^n, \alpha^n\}_{n=1}^N$, one can compute posterior moments (as sample averages) and draw a histogram of any function of the parameters.

Remarks:

- The choice of the grid of points has to be made carefully and constitutes the main difficulty in applying the method. One must have an idea of the bounds of integration. These may come from prior restrictions on the parameters, e.g. $0 < \alpha < 1$. However, even in such cases, we advise to restrict the integration to the subset of the parameter space where the value of the posterior density is big enough to contribute to the integrals. Otherwise, too many points of the grids will be practically useless, at the expense of precision in the inversion of the cdf. It is worth exploring the shape of the conditional densities by varying the conditioning values enough, before fixing the integration ranges.

- It is quite evident that a variable grid could be more efficient since more points should be used in the area where the posterior density varies a lot. A variable grid could be used, but it may be difficult to devise a way to adapt it at each iteration of the loop of draws (since the cdf varies because of the renewal of the conditioning values). An automatic way to adapt the grid is to use an iterative method of integration (but see the next remark).

- Our implementation of the algorithm requires that the grid of points of each parameter remains constant throughout the N draws. Otherwise, we would have to interpolate the value of the conditional pdf over a fixed grid (using the values of the pdf over the variable grid of the current cycle) in order to apply (8). This is not recommended if the functional evaluation of the posterior density is expensive (as in GARCH models). For this reason, iterative integration methods (like the iterative Simpson's method) are less attractive.

- We use the trapezoidal rule of integration (see Davis and Rabinowitz, 1975, p 40) over fixed grids of (usually 33) points, and linear interpolation between adjacent points to invert the cdf. The cost of the method being proportional to the number of points in the grids, a smaller number of points could be compensated by a more sophisticated interpolation method to invert the cdf.

2.3 Griddy Gibbs and GARCH models with Student errors

The applications of the Griddy-Gibbs sampler in Ritter and Tanner (1992) are interesting but mainly expository. If the posterior density is bivariate, usual deterministic integration rules can be used and are much more efficient. The method will be really useful if it can deal with higher dimensional posterior densities. The GARCH(1,1) model we presented in (2) has six parameters. Generalizing the Gibbs sampler to m parameters imply that we consider a more complex Markov chain. Let θ denote the vector of m parameters and let θ_{-k} denote the vector θ without its k-th element θ_k. The draw θ^n of the posterior density is generated using the previous draw θ^{n-1} by the following cycle:

$$
\begin{aligned}
\theta_1^n &\sim \varphi(\theta_1|\theta_{-1}^{n-1}, y) \\
\theta_2^n &\sim \varphi(\theta_2|\theta_1^n,\, \theta_3^{n-1},\, \cdots \theta_m^{n-1}, y) \\
&\cdots \\
\theta_k^n &\sim \varphi(\theta_k|\theta_1^n,\, \ldots,\, \theta_{k-1}^n, \theta_{k+1}^{n-1},\, \ldots\, \theta_m^{n-1}, y) \\
&\cdots \\
\theta_m^n &\sim \varphi(\theta_m|\theta_{-m}^n, y)
\end{aligned}
\tag{11}
$$

The procedure is started using any value θ_{-1}^a in the support of the posterior distribution.

We tried the algorithm on an artificial sample of 150 observations generated with $\omega = 0.1$, $\alpha = \beta = 0.4$, $\nu = 5$, $\rho = 0.5$, $\mu = 1$. We scaled the data, dividing y_t by its empirical standard deviation which was 1.104, so that ω has to be divided by $(1.104)^2 = 1.219$. We dropped the first 50 draws out of a chain of 1000 draws of the Markov chain. In Table 1, we report maximum likelihood estimates (with asymptotic standard errors), and posterior means (with standard deviations). Figures 1 show the marginal posterior densities obtained after 1000 draws. The ranges of the graphs are the intervals of integration of each parameter. This means that the prior is uniform over the product of these intervals.

Posterior means of μ and ρ are close to the ML results, and the posterior standard deviations are a little larger. Posterior moments for the GARCH parameters differ a little more from the ML results, because of the skewness of the posterior densities. In the case of β, the skewness is caused by the positivity restriction (the ML standard error is larger than

Figure 1: Posterior densities (GARCH model with Student errors)

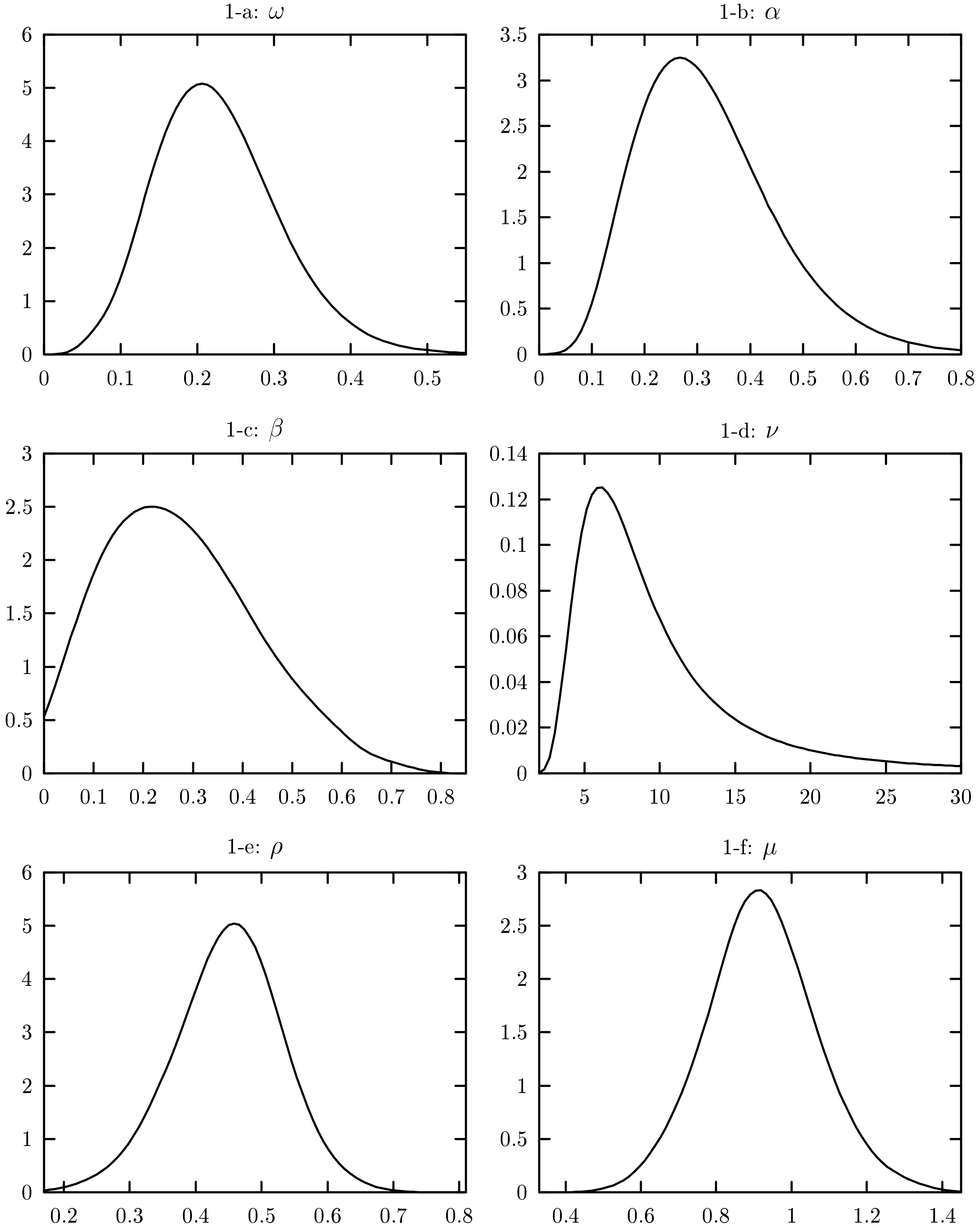


Table 1: ML and posterior results with Griddy Gibbs

parameter	ML	GGS (1)	GGS (2)	GGS (2)
ω	0.175 [0.079]	0.222 [0.079]	0.226 [0.082]	0.225 [0.082]
α	0.262 [0.135]	0.317 [0.132]	0.319 [0.130]	0.313 [0.130]
β	0.328 [0.240]	0.275 [0.154]	0.280 [0.154]	0.282 [0.157]
ν	6.093 [2.017]	9.540 [4.926]	9.519 [5.149]	9.736 [5.244]
ρ	0.472 [0.077]	0.450 [0.085]	0.450 [0.085]	0.458 [0.083]
μ	0.878 [0.139]	0.923 [0.155]	0.919 [0.152]	0.907 [0.149]
N		1000	1000	10000
time (3)		7'	7'	64'

(1) Posterior means [standard deviations] computed as sample moments of draws.
(2) Results computed by averaging conditional moments. Same draws as in (1).
(3) CPU time in minutes on a Pentium 133 (with GAUSS).

the posterior standard deviation because the ML procedure does not take the restriction into account). The most striking difference concerns the degrees of freedom parameter ν. Its marginal posterior is very skewed probably because the data do not provide enough evidence in favour of the low degrees of freedom of the DGP, the sample size being relatively small. Overall, the method is successful in dealing with the skewness of the posterior which appears especially for the GARCH parameters.[1] This is due to the fact that integration is done on a grid so that every direction can be explored in detail.

The algorithm seems to have converged as multiplying by ten the number of draws does not change significantly the results. Figures 2 display the CUMSUM evolution of the Monte Carlo estimates of the posterior mean of each parameter. Each figure contains the CUMSUM plots for the two types of estimates, one based on the sample mean of the draws (option 1), the other on the sample mean of the conditional means (option 2). Convergence occurs more quickly for some parameters (ν) than for others (ρ, μ); it seems that convergence is slower for parameters that are highly correlated[2] . A value of 0.05 for a CUMSUM means that the estimate of the posterior expectation diverges from the final estimate by 5 percent in units of the final estimate of the posterior standard deviation; so a divergence of even 25 percent is not a bad figure. In no case is the CUMSUM greater than 0.25, 0.16, and 0.1 (in absolute value) after 100, 200 and 1000 draws, respectively. As the method is relatively greedy in computational time, the question is of course to determine the minimal number of draws

[1] We have also applied the method to a larger artificial sample of 300 observations. The results were qualitatively similar, though with less skewness in the posterior of ν. See also Section 5 for applications to real data.

[2] In the posterior correlation matrix, most correlation coefficients are lower than 0.1; the three large correlations are between ρ and μ (-0.90), ω and β (-0.62), and α and β (- 0.50). When parameters are highly correlated, the draws tend to cluster, so that it takes a lot of draws to explore the posterior, and convergence is slower. This is revealed also by the autocorrelograms of the draws: the first autocorrelations are in general larger (and the autocorrelations go more slowly to zero) for highly correlated parameters than for weakly correlated ones (see Figures 3).

Figure 2: CUMSUM plots for posterior means (GARCH model with Student errors)

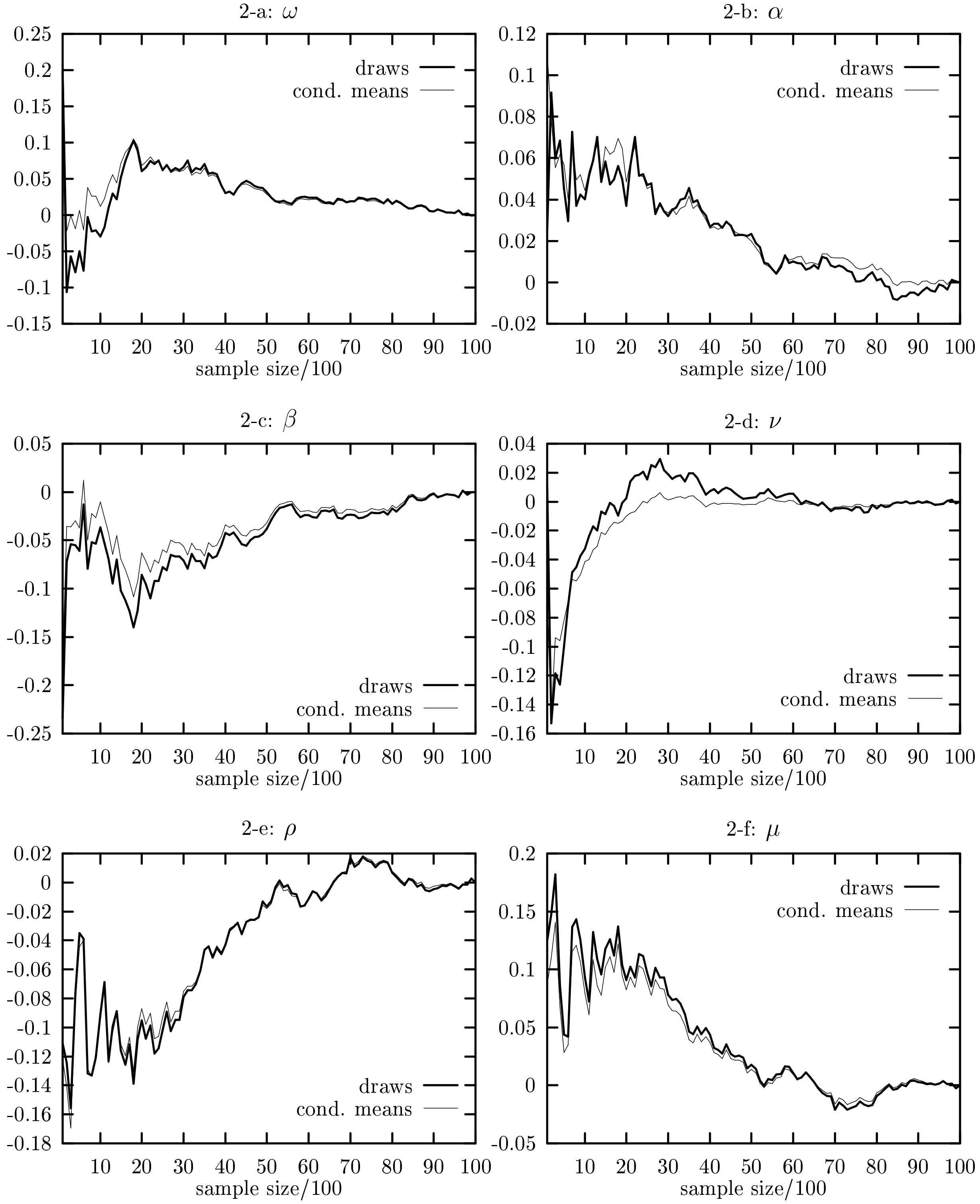
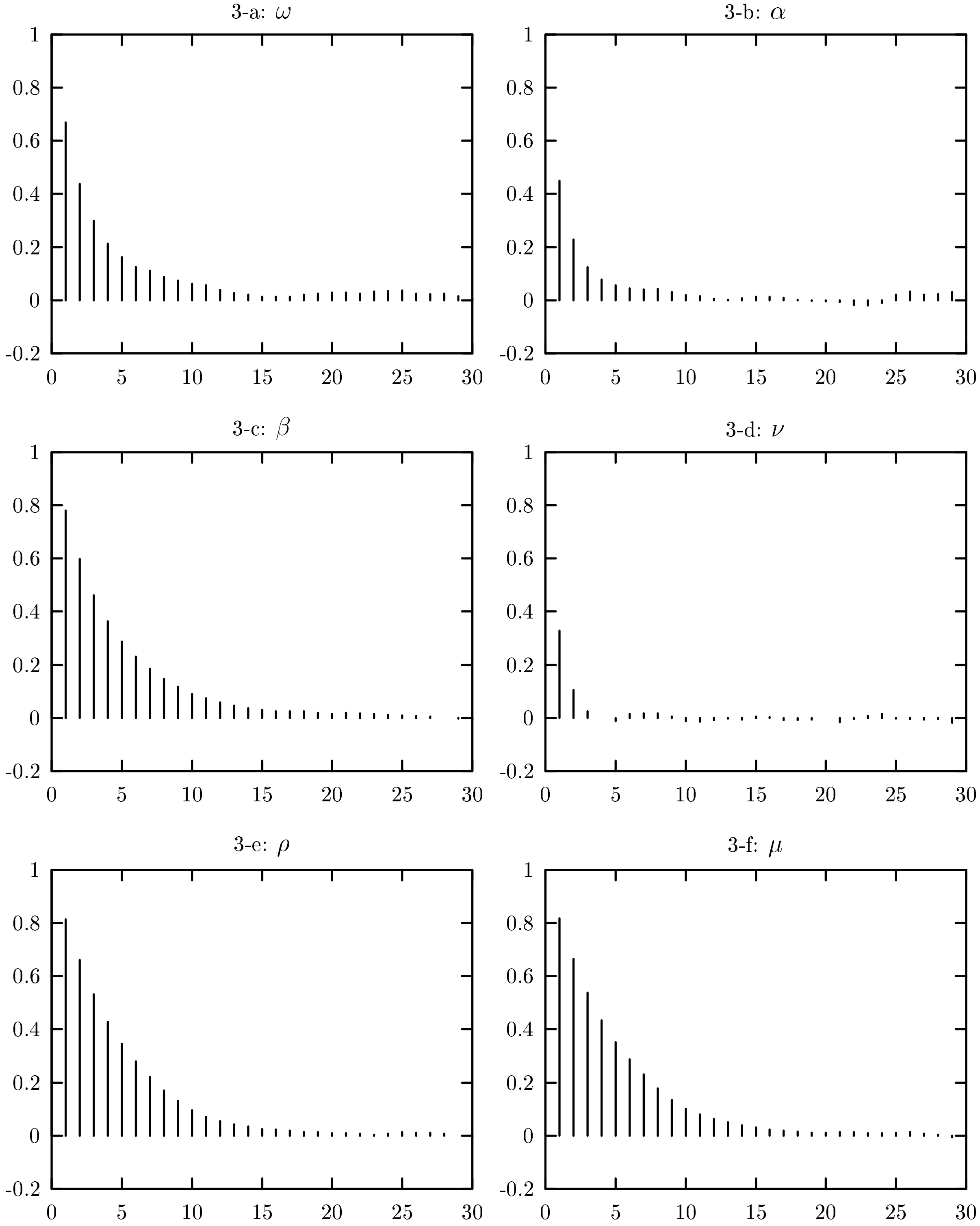


Figure 3: Autocorrelograms of the draws (GARCH model with Student errors)



necessary to achieve a reasonable precision. With 1000 draws (and 50 additional drops), one gets very good results and this value should be kept for final results. But provisional results can be obtained with as few as 100 draws with 5 initial drops, see Table 2. In this case the order of the parameters in the cycle described by (11) may be important as changing this order induced differences in the posterior means ranging from 3% to 13%. The best order is that which produces the smoothest graphs for the marginals. It seems that it is obtained when the parameters which are highly correlated are considered last in the cycles (here ρ and μ). It is also likey that the order matters in conjunction with the selection of the initial values corresponding to the most correlated parameters.

In Table 1 (and for other sample sizes in Figures 2), we see also that the posterior moments computed as sample moments of the draws are close to the results computed by averaging conditional moments. The largest difference is 2 percent for the means, and 4.5 percent for the standard deviations. If we compare both sets of results with the results after 10000 draws[3] , the results based on the second option are a little bit more precise, especially for the posterior standard deviations. We also looked at the autocorrelograms of the draws (Figures 3) and of the conditional expectations: they are very similar, but it is not the case that the autocorrelations of the conditional expectations are always smaller than the autocorrelations of the draws. This is mentioned by Tierney (1994) as a sufficient condition for the Monte Carlo variance of the estimator of the mean to be smaller with option 2 than with option 1. Finally, the difference in computing time between the two options is negligible.

The main cost of the method is of course the evaluation of the posterior density kernel (3): with 33 point grids[4] for 6 parameters, and 1 050 draws, the algorithm requires 207 900 functional evaluations. So it can be greedy in computational time. Consequently it is worth comparing it with other methods such as importance sampling [Kleibergen and van Dijk (1993)] and the Metropolis-Hastings algorithm [Geweke (1994)].

3 Comparisons with other methods

3.1 Importance sampling

The importance sampling (IS) algorithm uses an approximation to the posterior density to generate random draws. This approximation is called the importance function. Bayesian computations imply that we evaluate integrals of the form:

$$\mathrm{E}[g(\theta)] = \frac{\int g(\theta)\,\varphi(\theta|y)\,d\theta}{\int \varphi(\theta|y)d\theta} \tag{12}$$

[3]With 10000 draws, results with option 1 differ by less than 1.1 percent from results with option 2.

[4]We also computed the results using grids of 17 and 65 points. There is no systematic difference regarding the posterior means. Posterior standard deviations are increasing when the number of points of the grid decreases: e.g. for the parameter ν, they are 4.88, 4.93 and 5.11 (for $N = 1000$ and sample moments). Results with 33 point grids seem accurate enough. The method of integration (Simpson or Trapezoidal) had strictly no influence on the final results for this numerical example with 33 point grids.

where $\varphi(\theta|y)$ is a kernel of the posterior density and $g(.)$ is an integrable function. These integrals can be expressed as expectations with respect to the importance function, say $I(\theta)$. For instance:

$$\int \varphi(\theta|y)\, d\theta = \int I(\theta)\, \varphi(\theta|y)/I(\theta)\, d\theta = \mathrm{E}_{I(\theta)} \frac{\varphi(\theta|y)}{I(\theta)} \tag{13}$$

which can then be approximated by the sample mean:

$$\mathrm{E}_{I(\theta)} \frac{\varphi(\theta|y)}{I(\theta)} \approx \frac{1}{N} \sum_{i=1}^{N} \frac{\varphi(\theta_i|y)}{I(\theta_i)} \tag{14}$$

The estimation converges provided the weight function $\varphi(\theta|y)/I(\theta)$ is bounded. The smaller the variation coefficient of this ratio, the better the precision. The crux of the method is to find a good importance function. Most of the time the choice is based on the Student density, which is relatively easy to simulate. This density is rather flexible as the degrees of freedom parameter governs the thickness of its tails, but it is a symmetric density (no skewness). If m is the location parameter, M the scale matrix and λ the degrees of freedom, a usual way to calibrate this density is to fix $m = \hat{\theta}$, the posterior mode, and $M = [c^2 H^{-1}(\lambda - 2)/\lambda]^{-1}$, where H is the Hessian of the log posterior evaluated at its mode and c is a tuning constant (e.g. 1.2 to inflate the standard deviations). The choice of λ is experimental, unless one has specific information on the tails of the posterior, e.g. through conditions on existence of moments. A sometimes useful practice is to get a first evaluation of $\mathrm{E}(\theta|y)$ and $\mathrm{Var}(\theta|y)$ with a moderate number of draws ($N_1 = 1000$ to 5000), to use this evaluation to recalibrate the importance function and then to launch a run with a greater number of draws ($N_2 = 5000$ to 50000).

The precision which can be obtained by the IS method can be measured as follows. An integral is evaluated by a sample mean which we call m. For N large enough, by application of a central limit theorem, the estimator of the integral is distributed as a Normal density with mean equal to m and variance s^2/N (provided the sample is I.I.D.), where s^2 is the variance of the function to estimate. With 95% level of confidence, the relative error of the integral is lower than:

$$\frac{1.96 \times 2 \times s}{\sqrt{N} \times m} \tag{15}$$

For the integral (13), the relative error is obtained by replacing s/m by the variation coefficient of the weight function. It is also possible to compute the relative precision obtained for the expectation of each component of θ. This expectation is obtained as the ratio of two integrals given in (12). The variance of this ratio can be approximated using formula (27.7.3) in Cramer (1946) as suggested by Kloek and van Dijk (1978). This method is not directly applicable to Markov chain Monte Carlo methods as the draws are not independent in this case (so that the variance of the estimator is difficult to estimate).

We launched an experiment with $N_1 = 5000$, $N_2 = 15000$, $\lambda = 15$ and $c = 1.2$. A Student importance function is not totally adequate for the GARCH model as the parameters of the GARCH, e.g. ω, α, β and ν have to be positive. Negative draws must be rejected. We got 20% of rejection in the first round and 5% in the second. The overall quality of the importance function is not too bad, although it is not very good, since the variation coefficient of the weight function is equal to 2.65 in the first round and 1.78 in the second round (a variation coefficient below 1 is a good benchmark). Within a confidence interval of 95%, this corresponds to a 6% relative error. Decreasing the value of λ from 15 to 5 deteriorated greatly the quality of the importance function as for this case the coefficients of variation jumped to 4.29. Table 2 compares the posterior means and standard deviations obtained with GGS to those obtained with IS on the same artificial sample as in Section 2. Results are fairly comparable. They should not be exactly the same because IS does not truncate the parameter space. Computational time is moderate when one is interested only in posterior moments (5000 + 15000 draws were enough in this case). But they are not at all sufficient to get smooth graphs of the marginals. With only 100 draws, the GGS is, in this example, competitive.

Table 2: ML and posterior results by different methods

parameter	MLE	GGS	IS	Metropolis
ω	0.175 [0.079]	0.216 [0.077]	0.221 [0.085]	0.192 [0.069]
α	0.262 [0.135]	0.304 [0.125]	0.318 [0.135]	0.286 [0.115]
β	0.328 [0.240]	0.284 [0.140]	0.282 [0.157]	0.308 [0.168]
ν	6.093 [2.017]	9.603 [5.089]	8.688 [3.561]	6.779 [2.168]
ρ	0.472 [0.077]	0.439 [0.078]	0.468 [0.079]	0.468 [0.084]
μ	0.878 [0.139]	0.940 [0.146]	0.896 [0.142]	0.885 [0.151]
N		100	5000 + 15000	20000
error (1)		2.79	2.92	11.2
time (2)		61”	85”	95”

(1) Average absolute errors of the estimates of the means (in percent), with respect to the GGS results for $N = 10000$ in Table 1.
(2) Computational time with GAUSS on a Pentium 133Mh.

3.2 Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm was proposed by Metropolis *et al* (1953) and extended by Hastings (1970). It is a Markov chain algorithm which, when convergence is achieved, draws from the exact posterior density. It is very much related to the importance sampling method as it uses an approximation to the posterior to draw random numbers. But there is a rejection mechanism (instead of a weighting) to decide if a draw belongs or not to the exact posterior density. Given a previous draw θ^{n-1}, a new draw θ^n is accepted with probability p given by:

$$p = \text{Min}\left[\frac{\varphi(\theta^n|y)/I(\theta^n)}{\varphi(\theta^{n-1}|y)/I(\theta^{n-1})}, 1\right] \tag{16}$$

otherwise the previous draw is kept.

We launched a Metropolis chain of 20000 draws with the same Student density as in the first round of IS. To respect the positivity constraints, 5356 draws had to be discarded. In the Metropolis chain, 6525 draws were 'rejected', which means that in 32% of the draws, the previous draw was kept. This does not seem too bad. Posterior results obtained by this method are also in Table 2. The striking fact is that the Metropolis chain did not yield accurate results, even for the means. This is especially the case for the degrees of freedom parameter. Consequently, trying a second round with a recalibrated importance function was useless. We also tried, without success, to calibrate differently the importance function. Changing the values of λ did not improve the results. The most puzzling fact is that when we increased the tuning constant c of the scale matrix, we had the impression that we could get any arbitrary results.

These bad results seem to be specific to the particular artificial sample considered here. Its characteristics are first a small number of observations (150) and second that the unconditional variance of the series does not exist, as $1 - \beta - \alpha\nu/(\nu - 2) = -0.07$ (see Section 4.1). With more observations (500) and slightly different parameters which insure the existence of the unconditional variance of the series ($\nu = 4$, $\alpha = \beta = 0.3$), the Metropolis-Hastings algorithm seems to converge for the posterior means even if the results for the posterior variances are not so precise.

3.3 Partial conclusions

Our experience is thus not so favourable to the Metropolis-Hastings algorithm as Geweke's (1994) (who uses normal errors). The old fashioned importance sampling algorithm provided as accurate posterior moments as the griddy-Gibbs sampler, but less precise graphs of the marginal densities. Although GGS is more costly in computer time, it seems easier to implement, since in our experiments, the calibration of a Student importance function was not obvious.

4 Persistence and asymmetry of shocks

In this section, we discuss extensions of the basic GARCH-Student model used in Section 2 needed for the empirical application presented in the next section. These extensions concern the possible persistence of shocks on future volatilities, and the possibility of a different impact of a shock on the next variance according to the sign of the shock.

4.1 Persistence

The conditional variance can be expressed using the standardized shock $z_t = u_t/\sqrt{h_t}$:

$$h_t = \omega + h_{t-1}(\alpha z_{t-1}^2 + \beta) \tag{17}$$

Using the recursion defined in Nelson (1990), we can express the series of the h_t as

$$h_t = h_0 \prod_{i=1}^{t} (\alpha z_{t-i}^2 + \beta) + \omega[1 + \sum_{k=1}^{t-1} \prod_{i=1}^{k} (\alpha z_{t-i}^2 + \beta)] \tag{18}$$

We can define the persistence of an initial shock z_0 on the conditional variance h_t as:

$$\lim_{t\to\infty} \mathrm{E}\frac{\partial h_t}{\partial z_0^2} \tag{19}$$

This corresponds to a particular case of formula (36) in Nelson (1990) (persistence in L^1). Assuming that h_0 is independent of the sequence $\{z_t\}$ (which is itself an independent sequence), from (18), the expected derivative is equal to:

$$\begin{aligned} \mathrm{E}\frac{\partial h_t}{\partial z_0^2} &= \mathrm{E}(h_0)\mathrm{E}\left(\frac{\partial \prod_{i=1}^{t}(\alpha z_{t-i}^2+\beta)}{\partial z_0^2}\right) \\ &= \mathrm{E}(h_0)\,\mathrm{E}\left(\frac{\partial(\alpha z_0^2+\beta)}{\partial z_0^2}\right)\,\mathrm{E}\left(\Pi_{i=1}^{t-1}(\alpha z_{t-i}^2+\beta)\right) \\ &= \mathrm{E}(h_0)\,\alpha\left[\alpha \mathrm{E}(z^2)+\beta\right]^{t-1} \end{aligned} \tag{20}$$

We have persistence if (20) converges to a finite and non zero quantity for $t \to \infty$. This condition is verified for $\alpha\mathrm{E}(z^2)+\beta = 1$. Persistence is then measured by $\alpha\,\mathrm{E}(h_0)$, assuming that $\mathrm{E}(h_0)$ is finite.

If $z \sim N(0,1)$, persistence occurs if there is a unit root in the GARCH process ($\alpha+\beta=1$). This corresponds to the IGARCH process studied by Engle and Bollerslev (1986) who were the first to underline the similarity of GARCH processes with a unit root (IGARCH) and processes that are integrated in the mean. In a process integrated in the mean, a shock in the current period affects the future of the variable indefinitely. In an IGARCH process, a certain type of future of h is affected indefinitely. The condition for integration when $z \sim t(0,1,\nu)$ is slightly different:

$$\alpha\frac{\nu}{\nu-2}+\beta = 1 \qquad (\text{if } \nu > 2) \tag{21}$$

As underlined by Nelson (1990), this condition for persistence corresponds to the absence of weak stationarity (because the unconditional variance of y_t or the unconditional expectation of h_t does not exist). While not being weakly stationary, the process generating the y_t can still be strongly stationary since the condition for strong stationarity, which is $\mathrm{E}\log(\alpha z_t^2+\beta) < 0$, is less stringent than the condition for weak stationarity. The condition for strong stationarity implies that the posterior expectation of h_t does not explode as t tends to infinity.

4.2 Asymmetry

The GARCH model studied in the previous section assumes that a shock influences the conditional variance h_t only by its squared magnitude, not by its sign. This may be somewhat at odds with the empirical behaviour of stock markets - see e.g. the survey by Bollerslev, Chou and Kroner (1992) for a discussion and references. Several specifications have been proposed to introduce asymmetry, such as the EGARCH model of Nelson (1991) or the threshold ARCH model of Zakoian (1994). In the sequel we choose a particular type of threshold GARCH model given by:

$$h_t = \omega + \alpha^+ u_{t-1}^{2+} + \alpha^- u_{t-1}^{2-} + \beta h_{t-1} \tag{22}$$

where $u_t^{2+} = u_t^2$ if $u_t > 0$ and 0 otherwise, and u_t^{2-} is defined likewise for negative u_t. This is a slightly different version of the model used by Zakoian (1994) as he writes the GARCH equation for the standard deviation instead of the variance.

In the case of asymmetry of shocks, to measure persistence, we have to distinguish positive and negative shocks. The derivative in (20) can be computed for positive and for negative z_0. This gives:

$$\begin{aligned} \mathrm{E}\frac{\partial h_t}{\partial z_0^{2+}} &= \mathrm{E}(h_0)\,\mathrm{E}\left(\frac{\partial(\alpha^+ z_0^{2+} + \alpha^- z_0^{2-} + \beta)}{\partial z_0^{2+}}\right)\mathrm{E}\left(\prod_{i=1}^{t-1}(\alpha^+ z_{t-i}^{2+} + \alpha^- z_{t-i}^{2-} + \beta)\right) \\ &= \mathrm{E}(h_0)\,\alpha^+\left[\alpha^+\mathrm{E}(z^2|z>0) + \alpha^-\mathrm{E}(z^2|z<0) + \beta\right]^{t-1} \\ &= \mathrm{E}(h_0)\,\alpha^+\left[(\alpha^+ + \alpha^-)\mathrm{E}(z^2)/2 + \beta\right]^{t-1} \end{aligned} \tag{23}$$

(where the last equality follows from the symmetry of the distribution of z). Weak nonstationarity is obtained for $(\alpha^+ + \alpha^-)\mathrm{E}(z^2)/2 + \beta = 1$. The measure of persistence is $\mathrm{E}(h_0)\,\alpha^+$ for positive shocks and $\mathrm{E}(h_0)\,\alpha^-$ for negative shocks.

5 Application to the Brussels stock market

We estimated an asymmetric GARCH model on stock returns. The precise specification is

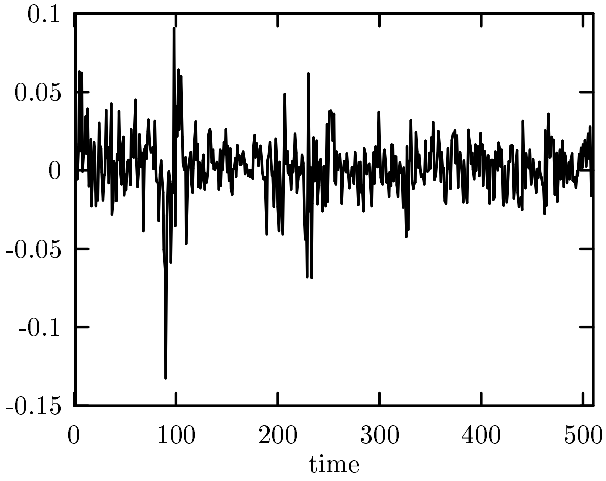
$$\begin{cases} y_t = \mu + \rho y_{t-1} + u_t \\ u_t|I_{t-1} \sim \text{Student}(0, h_t, \nu) \\ h_t = \omega + \alpha^+ u_{t-1}^{2+} + \alpha^- u_{t-1}^{2-} + \beta h_{t-1} \end{cases} \tag{24}$$

(see Section 4.2 for the definition of u_t^{2+} and u_t^{2-}).

We use weekly data on the spot market index of shares of Belgian firms at the Brussels stock exchange, for the period 03/01/86 to 26/01/96. The data consist of closing quotes on Friday, providing 508 observations. The dependent variable is the index return, measured as the first difference of the logarithm of the index at time t. A time series plot of the data is provided on Figure 4. The mean return is equal to 0.002246 and the standard deviation to 0.01909. The index exhibits an overall upward trend over the sample period.

5.1 Posterior results

Figure 4: Brussels spot market index return (3/1/86 - 26/1/96)



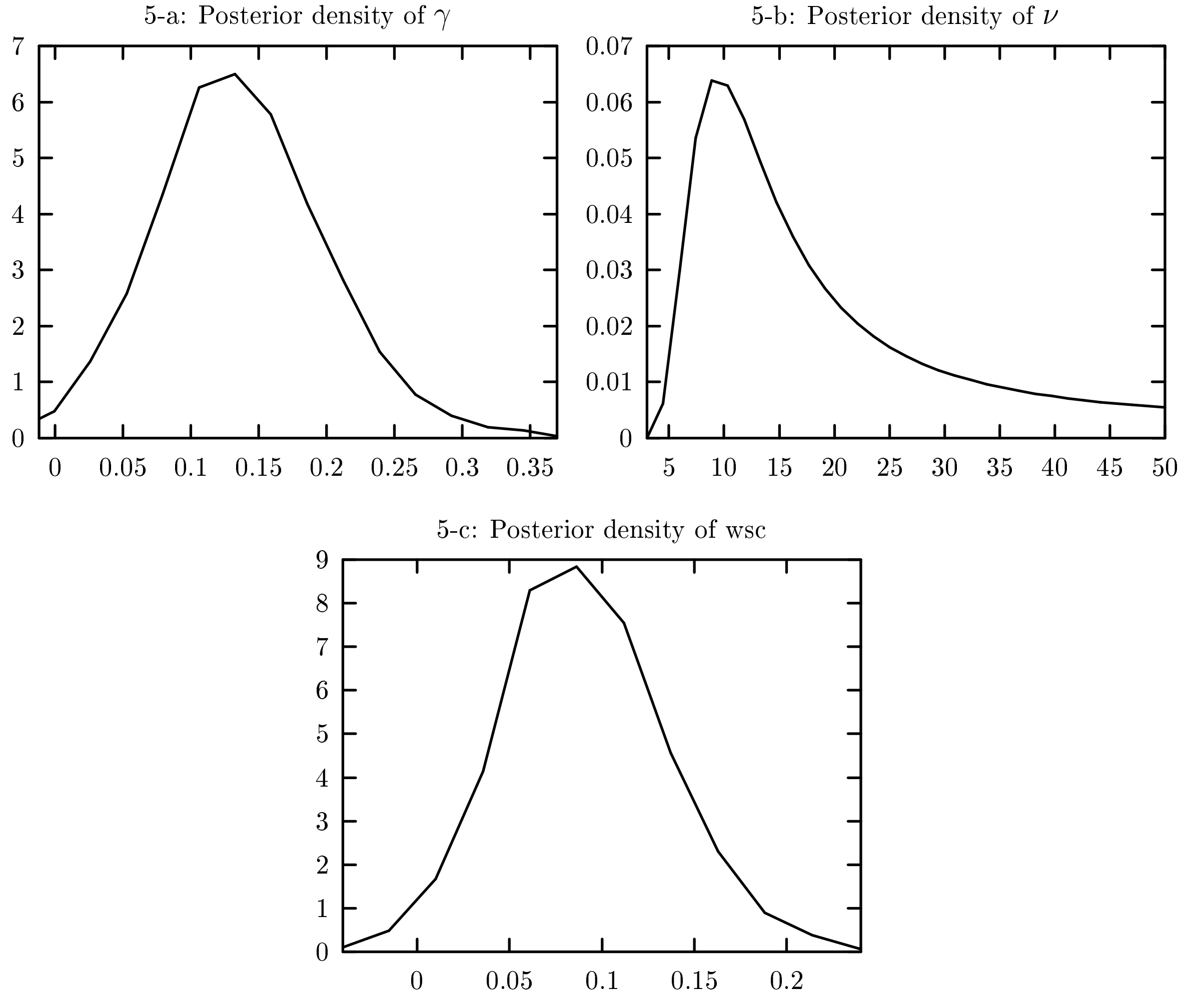
All the results were computed with the Griddy-Gibbs sampler using 1000 draws plus 50 initial dropped, unless otherwise stated. A uniform prior on finite ranges was used for each parameter, such that no significant truncation of the posterior occured. Other non-informative prior measures could be considered, which reflect other forms of ignorance (according to relevant criteria). For example, instead of a uniform prior on the degrees of freedom parameter, one could consider a uniform prior on its logarithm, which gives more weight to small values than to big values. Jeffreys' invariant prior would be another possibility. We just stress that the Griddy-Gibbs algorithm can deal with any form of prior. Table 3 presents the results with our prior. They call for the following comments:

1) Negative shocks have a stronger impact than positive shocks in all specifications. When a symmetric model is estimated (GARCH-t column in Table 3), the posterior mean of α ($= \alpha^+$) is approximately equal to the average of the posterior means of α^+ and α^- in the corresponding asymmetric model (see GARCH-A-t column). The difference $\gamma = \alpha^- - \alpha^+$ has a posterior mean equal to 0.13 with a standard deviation equal to 0.09. The posterior odds ratio for $\gamma > 0$ against $\gamma = 0$ is approximately equal to four. The posterior density of γ is reported in Figure 5-a.

2) The results for the asymmetric GARCH with Student errors do not differ markedly from the results with normal errors. The degrees of freedom parameter has a posterior mean equal to 18, and an extremely long tail to the right (see Figure 5-b), indicating that a normal distribution may be a good enough approximation for this data set.[5] We notice that for the symmetric model (GARCH-t) the degrees of freedom are less high: the underestimation of the effect of negative shocks is partially compensated by additional kurtosis.

[5]We have also estimated the model for individual stocks and found that the degrees of freedom parameter was close to 5 in many cases.

Figure 5: Posterior densities (GARCH-A-t)



3) The weak stationarity condition for the Student model requires that $wsc = 1 - (\alpha^+ + \alpha^-)\nu/[2(\nu-2)] - \beta$ be positive. Its posterior probability is quite close to 1 (see also the density on Figure 5-c). Nevertheless, we estimated the asymmetric IGARCH model (see column IGARCH-A-t of Table 3), where the parameter β is constrained to be equal to $1-(\alpha^+ + \alpha^-)\nu/[2(\nu-2)]$. The constraint is quite easy to impose in the Griddy-Gibbs (it suffices to replace β by its expression in the evaluation of the posterior density).

4) The returns exhibit positive autocorrelation of the order of 0.20 (the posterior mean of ρ), thus ruling out weak efficiency of the market.

5.2 Predictive option prices

The cash price (C^+_{T+s}) of a European call option s units of time before maturity is evaluated by [see Ghysels, Harvey and Renault (1995)]:

$$C^+_{T+s} = e^{-rs}\,\mathrm{E}(S_{T+s} - K)^+ \tag{25}$$

where $(S_{T+s} - K)^+ = \max(S_{T+s} - K, 0)$, S_{T+s} is the price of the underlying security at maturity (assumed without dividend), K is the exercise price ("strike"), and r is the interest rate (assumed known, though for practical applications, it could be modelled). We choose to normalise C^+_{T+s} as

$$C_{T+s} = \frac{C^+_{T+s}}{Ke^{-rs}} = \mathrm{E}\left(\frac{S_{T+s}}{K} - 1\right)^+ \tag{26}$$

and to compute the expectation using the predictive distribution of S_{T+s} (short of knowing the "risk-neutral" probability distribution that should be used). The predictive option price C_{T+s} is

$$\begin{aligned}
\mathrm{E}[(\frac{S_{T+s}}{K} - 1)^+|y] &= \int (S_{T+s}/K - 1)^+\, f(S_{T+s}|y)\, dS_{T+s} \\
&= \int (S_{T+s}/K - 1)^+ \left[\int f(S_{T+s}|\theta, y)\,\varphi(\theta|y)\, d\theta\right] dS_{T+s} \\
&= \int \left[\int (S_{T+s}/K - 1)^+\, f(S_{T+s}|\theta, y)\, dS_{T+s}\right] \varphi(\theta|y)\, d\theta
\end{aligned} \tag{27}$$

If we can simulate R realisations of the distribution of $S_{T+s}|\theta, y$, the inner integral of the last line of the above formula can be approximated by its empirical counterpart $R^{-1}\sum_{r=1}^{R}(S_{T+s}/K - 1)^+$ (given θ). Then having N realisations of the posterior distribution of θ (e.g. a Gibbs sample), we compute the simple mean of the N estimates of the inner integral to estimate the outside integral. To simulate the distribution of $S_{T+s}|\theta, y$, we use the econometric model where $\Delta \ln S_t = y_t$. The future price S_{T+s} is related to the last observed prics S_T by

$$S_{T+s} = S_T\, \exp \sum_{i=1}^{h} y_{T+i} \tag{28}$$

so that we have to simulate the future returns y_{T+i}'s. If $s = 1$, we just have to simulate y_{T+1} (the first post-sample return). If the econometric model is $y_t \sim$ Student$(0, h_t, \nu)$ with $h_t = \omega + \alpha y^2_{t-1} + \beta h_{t-1}$, then given θ and y (the observed sample), we know y_T, so that we can compute h_{T+1} and generate y_{T+1} from its Student distribution. Afterwards, y_{T+2} can be generated from its Student distribution, with h_{T+2} computed using the previously generated y_{T+1} and computed h_{T+1}.

As an alternative to the evaluation formula (26), we can use the Black-Scholes formula - see Black and Scholes (1973) - which we write in percentage of the discounted strike:

$$BS_{T+s} = \frac{S_{T+s}}{Ke^{-rs}}\, N(d_1) - N(d_2) \tag{29}$$

where $N(.)$ is the cumulative Normal distribution function. Its arguments are given by:

$$
\begin{aligned}
d_1 &= \frac{\ln(S_{T+s}/Ke^{-rs})}{\sigma\sqrt{s}} + \frac{1}{2}\sigma\sqrt{s} \\
d_2 &= \frac{\ln(S_{T+s}/Ke^{-rs})}{\sigma\sqrt{s}} - \frac{1}{2}\sigma\sqrt{s}
\end{aligned}
\tag{30}
$$

They depend on the volatility parameter σ which is the standard deviation (per time unit) of the return of the underlying asset. Since σ is unknown, it must be predicted to apply (29). So we can apply the Black-Scholes formula with the volatility predicted using the GARCH model on the return. This is known as the 'plug-in' method - see Noh, Engle and Kane (1994), who find that it may be a profitable trading rule, although the Black-Scholes formula is only applicable in the context of *deterministic* volatility and other assumptions not respected by the GARCH model. When the time to maturity, s, is larger than one, σ^2 is the average of the s future variances v_{T+i}:

$$
\sigma^2 = \frac{1}{s}\sum_{i=1}^{s} v_{T+i} \tag{31}
$$

With a GARCH-Student model like (24), $v_{T+i} = \nu h_{T+i}/(\nu - 2)$, provided ν is larger than two. Therefore σ^2 is a function of the GARCH parameters and we can compute the predictive mean of BS_{T+s} for any given s and S_{T+s}/Ke^{-rs}, as soon as we have a Gibbs (or any other) sample of the posterior distribution of the GARCH parameters. Actually, the computation of the predictive mean of BS_{T+s} is be done by simulation of the predictive density of BS_{T+s}. In the same way as in the last line of (27), we write that

$$
BS^e_{T+s} = \mathrm{E}(BS_{T+s}|y) = \mathrm{E}_{\theta|y}\left[\mathrm{E}(BS_{T+s}|\theta, y)\right]. \tag{32}
$$

As explained above, BS_{T+s} is a function of the future conditional variances $h^* = (h_{T+1}, \; \cdots \; h_{T+s})$, and therefore a function of the future standardized innovations $z^* = (z_{T+1}, \; \cdots \; z_{T+s})$, see (17):

$$
BS_{T+s} = BS_{T+s}(\sigma) = BS_{T+s}[\sigma(h^*)] = BS_{T+s}\{\sigma[h^*(z^*)]\} \tag{33}
$$

Since $z_i = u_i/\sqrt{h_i}$ has the known distribution $z_i|\theta \sim \mathrm{Student}(0, 1, \nu)$ (given the past), we can easily simulate z^* and compute $BS_{T+s}(z^*)$.

We can now detail our algorithm for computing the predictive means C_{T+s} and BS^e_{T+s}. For each draw of θ provided by the GGS, we perform the following steps:

- *step 1:* compute u_T and h_T for the given θ (period T is the last one of the observed sample).
- *step 2:* compute $h_{T+1} = \omega + \alpha\, u_T^2 + \beta h_T$.
- *step 3.1:* $r = 1$; do while $r \leq R$.
- *step 3.2:* simulate $z_{T+1,r} \sim \mathrm{Student}(0, 1, \nu)$, and compute $y_{T+1,r} = h_{T+1}^{1/2} z_{T+1,r}$.

- *step 3.3.1:* $i = 2$; do while $i \leq s$.

- *step 3.3.2:* compute $h_{T+i,r} = \omega + h_{T+i-1,r}\,(\alpha\, z^2_{T+i-1,r} + \beta)$ (for $h_{T+1,r}$ always use h_{T+1} from step 2).

- *step 3.3.3:* simulate $z_{T+i,r} \sim \text{Student}(0, 1, \nu)$.

- *step 3.3.4:* compute $y_{T+i,r} = h^{1/2}_{T+i,r} z_{T+i,r}$.

- *step 3.3.5:* increment i by 1 and go to step 3.3.2 unless $i > s$.

- *step 3.4:* for Black-Scholes evaluation: compute $\sigma^2_r = s^{-1} \sum_{i=1}^{s} \nu h_{T+i,r}/(\nu - 2)$ and the implied $BS_{T+s,r}$; for direct evaluation, compute $S_{T+s,r} = S_T.\exp \sum_{i=1}^{s} y_{T+i,r}$.

- *step 3.5:* increment r by 1 and go to step 3.2 unless $r > R$.

- *step 4:* compute $\overline{BS}_{T+s} = \sum_{r=1}^{R} BS_{T+s,r}/R$ and $\overline{C}_{T+s} = \sum_{r=1}^{R} (S_{T+s,r}/K - 1)^+/R$.

With N values of θ, we get N values of $\overline{BS}_{T+s}$ and of $\overline{C}_{T+s}$ and we average them to get the estimates of the corresponding predictive means. A GAUSS program is available from the authors.

For the Brussels spot index, we report in Table 4, the predictive option prices of "at the money" options for different maturities (one week: $s = 1$; 3 months: $s = 13$; and 6 months: $s = 26$) during the post-sample period. For the Black and Scholes evaluation formula, we have set $S_{T+s}/Ke^{-rs} = 1$, in which case (29) simplifies into $BS_{T+s} = 2N(\sigma\sqrt{s}/2) - 1$. For the direct evaluation formula (26), we have set the strike at the last observed value of the stock index, in which case $S_{T+s}/K = \exp \sum_{i=1}^{s} y_{T+i}$. For predicting future returns, we have assumed that y_t has zero mean, instead of $\mu + \rho y_{t-1}$ (see steps 3.2 and 3.3.4), because the latter specification does not make sense for long-run predictions.

The results of the two evaluations are almost identical whether we fix R equal to 1 or to a large value, like 50. More important is the finding that the Black and Scholes evaluation hardly differs from the exact evaluation for at the money options using a GARCH model.

6 Conclusion

Drèze (1977) stated that for the resolution of a statistical problem in a Bayesian framework, "it has become acceptable to quit upon obtaining analytically a marginal posterior density for one or two parameters of interest, and declaring victoriously that it could be analyzed by numerical integration techniques". But, as he mentioned in a footnote, usual quadrature rules would require about sixteen years of computing time when the dimension of the integral is equal to nine.[6] The paradox of our paper is that the integration method

[6] He assumed that it takes one millisecond to compute the posterior, and that a grid of twenty points is used for each coordinate (leading to 20^9 function evalutions). Even if nowadays microcomputers are one hundred times faster than computers we used twenty years ago, the computing time would be sixty days.

we use for high dimensions is based on quadrature rules, but that it does not suffer from the usual curse of dimensionality which affects those methods. The Griddy-Gibbs sampler evaluates a $k-$dimensional integral by splitting it into a series of k unidimensional integrals and finding an accurate result by averaging those sets of k univariate integrals. On powerful microcomputers, the route is now opened for analyzing complex statistical models requiring numerical integration in a dimension not exceeding say fifteen.

The method has been applied to estimate GARCH models for the Brussels stock market. Asymmetric effects of a shock on the next variance are found, but as these effects do not seem to be persistent, it makes sense to compute predictive option prices.

There are many econometric models which lend themselves to using the Griddy-Gibbs sampler. In many cases, it is not necessary to apply it to the complete parameter vector as in GARCH models. As an example, consider the heteroskedastic regression model (1): we can apply the Griddy-Gibbs sampler to the posterior density of γ, and marginalize the posterior conditional density and moments of β. Other examples are ARMA models, and simultaneous equation models.

References

Black, F. and Scholes, M. (1973). The pricing of options and coporate liabilities. *Journal of Political Economy* 81, 637-659.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modeling in finance. *Journal of Econometrics* 52, 5-59.

Carlin, B.P. and Gelfand, A.E. (1991). An iterative Monte Carlo method for nonconjugate Bayesian analysis. *Statistics and Computing* 1, 119-128.

Cramer, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Davis, P.J. and Rabinowitz, P. (1975). *Methods of Numerical Integration*. New-York: Academic Press.

Drèze, J.H. (1977). Bayesian regression analysis using poly-t densities. *Journal of Econometrics*, 6, 329-354.

Engle, R.F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews* 5, 1-50, 81-87.

Geweke, J. (1989). Exact predictive densities in linear models with ARCH disturbances. *Journal of Econometrics* 40, 63-86.

Geweke, J. (1994). Bayesian comparison of econometric models. *Working Paper 532*. Research Department, Federal Reserve Bank of Minneapolis.

Geweke, J. (1995). Posterior simulators in econometrics. *Working Paper 555*. Research Department, Federal Reserve Bank of Minneapolis.

Ghysels, E., Harvey, A. and Renault, E. (1995). Stochastic Volatility. Forthcoming in *Handbook of Statistics*, Vol. 14. CORE DP 9569, Louvain- La-Neuve.

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

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

Kloek, T. and van Dijk, H.K. (1978). Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. *Econometrica* 46, 1-19.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21, 1087-1091.

Nelson D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318-334.

Nelson D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347-370.

Noh, J., Engle, R.F. and Kane, A. (1994). Forecasting volatility and option prices of the S & P 500 index. *Journal of Derivatives* ?, 17-30.

Robert, C.P. (1995). A pathological MCMC algorithm and its use as a benchmark for convergence assessment techniques. *Document de Travail CREST 9555.* INSEE, Paris.

Ritter, C. and Tanner, M.A. (1992). Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association* 87, 861-868.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics* 22, 1701-1762.

Yu, B. and Mykland, P. (1994). Looking at Markov samplers through CUMSUM path plots: a simple diagnostic idea. *Technical Report 413.* Department of Statistics, University of California at Berkeley.

Zakoian, J-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dyanmics and Control* 18, 931-955.

Zellner, A. and Min, C. (1995). Gibbs Sampler Convergence Criteria. *Journal of the American Statistical Association* 90, 921-927.

Table 3: Posterior results for Brussels stock returns

Parameter	GARCH-A-N	GARCH-A-t	IGARCH-A-t	GARCH-t
μ	0.0801	0.0876	0.0765	0.105
	[0.0362]	[0.0356]	[0.0360]	[0.0344]
ρ	0.206	0.200	0.203	0.197
	[0.0470]	[0.0466]	[0.0469]	[0.0472]
ω	0.0829	0.0695	0.0414	0.0819
	[0.0266]	[0.0244]	[0.0156]	[0.0299]
α^{+}	0.116	0.0972	0.134	0.176
	[0.0455]	[0.0442]	[0.0500]	[0.0474]
α^{-}	0.279	0.231	0.294	-
	[0.0630]	[0.0624]	[0.0686]	
β	0.715	0.718	-	0.670
	[0.0522]	[0.0620]		[0.0719]
ν	∞	18.01	13.01	13.47
		[10.75]	[7.075]	[8.619]
wsc	0.0907	0.0886	-	0.110
	[0.0392]	[0.0426]		[0.0564]
P(wsc> 0)	0.993	0.984	0	0.980

A = asymmetric; N = Normal; t = Student; wsc = weak stationarity condition (see text).

Table 4: Predictive call option prices on Brussels index (in percent)

Maturity	C_t	BS_t^e
1 week	0.717	0.731
3 months	2.580	2.533
6 months	3.684	3.664