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

David, Ardia
University of Fribourg Switzerland

September 2006

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

David Ardia
<david.ardia@unifr.ch>
*University of Fribourg Switzerland*

First version: March 13, 2006
Last version: September, 2006

**Abstract:** In this article, we propose the Bayesian estimation of the parsimonious but effective GARCH(1,1) model with Normal innovations. We sample the parameters joint posterior distribution using the approach suggested by Nakatsuma [8]. As a first step, we fit the model to foreign exchange log-returns time series and compare the Maximum Likelihood and the Bayesian estimates. Next, we illustrate some appealing aspects of the Bayesian approach through interesting probabilistic statements made on the parameters.

JEL Classification: C11, C15, C22, C52.
Keywords: GARCH model, Bayesian estimation, Markov Chain Monte Carlo.

## 1 Introduction

Volatility plays a central role in financial risk management and lies at the heart of any model for pricing derivatives securities. Research on changing volatility using time series models has been active since the pioneer paper by Engle [4]. From there, ARCH and GARCH type models grew rapidly into a rich family of empirical models for volatility forecasting during the 80's. These stochastic processes, unconditionally non-Gaussian and possibly non-stationary, have been extensively applied to financial data in the econometric literature. Reasons for that lying in their ability to reproduce heteroscedasticity, volatility clustering and unconditional heavy-tailed distribution, both salient features for reproducing financial returns.

Whilst apparently simple by nature, we may encounter many difficulties when dealing with GARCH models; the model's parameters must be positive, and sometimes the model is required to be covariance stationary, which can complicate the optimization procedure. In addition, the finite sample evidence on the performance of GARCH Maximum Likelihood estimates and test statistics is still fairly limited. Reliable inference from the LM, Wald and LR test statistics generally require moderately large sample sizes of at least two hundred or more observations. However, most of these deficiencies break down when taking a Bayesian point of view. Any constraints on the

*published in *Student*, 5(3–4), September 2006, pp.283–298, ISSN: 1420-1011.

parameters can easily be integrated in the Markov Chain Monte Carlo (MCMC) procedure, exact inference in small samples is possible and non-nested models can be tested using Bayes factors. Furthermore, distributions of complex functions of the parameters can be obtained by simulation at low cost in contrast to the bootstrap approach.

In this article, we propose the Bayesian estimation of the parsimonious but effective GARCH(1,1) model with Normal innovations. We sample the parameters joint posterior distribution using the approach suggested by Nakatsuma [8]. As a first step, we fit the model to foreign exchange log-returns time series and compare the Maximum Likelihood and the Bayesian estimates. Next, we illustrate some appealing aspects of the Bayesian approach through interesting probabilistic statements made on the parameters.

The plan of this article is as follows: we set up the model in Section 2. The MCMC scheme is detailed in Section 3. The empirical results are presented in Section 4 while we give some illustrative applications of the Bayesian approach in Section 5. We conclude in Section 6.

## 2 The model and the priors

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

$$\begin{aligned} y_t &= \varepsilon_t h_t^{1/2} \quad \text{for} \quad t = 1, \ldots, T \\ \varepsilon_t &\overset{iid}{\sim} N(0,1) \\ h_t &:= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1} \end{aligned} \tag{2.1}$$

where $\alpha_0 > 0$, $\alpha_1 \geqslant 0$ and $\beta \geqslant 0$; $N(0,1)$ is the standard Normal distribution. In this setting, the conditional variance $h_t$ is a linear function of the squared past observation and the past variance. Positivity restrictions on the parameters ensure a positive conditional variance. In order to write the likelihood function, we define the following vectors: $\boldsymbol{y} := (y_1 \cdots y_T)'$, $\boldsymbol{\alpha} := (\alpha_0 \ \alpha_1)'$ and we regroup the parameters into $\Theta := (\boldsymbol{\alpha}, \beta)$. In addition, we define the $(T \times T)$ diagonal matrix $\Sigma := \Sigma(\Theta) = \text{diag}(\{h_t(\Theta)\}_{t=1}^T)$ where:

$$h_t(\Theta) = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1}(\Theta) \quad .$$

From there, the likelihood function of $\Theta$ can be written as:

$$l(\Theta|\boldsymbol{y}) \propto (\det \Sigma)^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{y}'\Sigma^{-1}\boldsymbol{y}\right] \tag{2.2}$$

where, for convenience, we use the first observation as an initial condition and the initial variance is fixed to $\alpha_0$. This likelihood refers to the conditional likelihood of the GARCH process given in (2.1). We propose the following proper priors on the parameters $\boldsymbol{\alpha}$ and $\beta$ of the preceding model:

$$\begin{aligned} p(\boldsymbol{\alpha}) &\propto N_2(\boldsymbol{\alpha}|\boldsymbol{\mu}_\alpha, \Sigma_\alpha)\mathbb{I}_{[\boldsymbol{\alpha}>\mathbf{0}]} \\ p(\beta) &\propto N(\beta|\mu_\beta, \Sigma_\beta)\mathbb{I}_{[\beta>0]} \end{aligned}$$

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, and $N_d$ is the $d$-dimensional Normal distribution $(d > 1)$. In addition, we assume prior independence between $\boldsymbol{\alpha}$ and $\beta$

which implies that $p(\Theta) = p(\boldsymbol{\alpha})p(\beta)$. Then, we construct the joint posterior distribution via Bayes' rule:

$$p(\Theta|\boldsymbol{y}) \propto l(\Theta|\boldsymbol{y})p(\Theta) \quad . \tag{2.3}$$

# 3 Simulating the joint posterior

The recursive nature of the variance equation in (2.1) does not allow for conjugacy between the likelihood function and the prior distribution in (2.3). Therefore, we rely on the Metropolis-Hastings (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 [8]. We draw an initial value $\Theta^{[0]} := (\boldsymbol{\alpha}^{[0]}, \beta^{[0]})$ from the joint prior distribution and we generate iteratively $J$ passes for $\Theta$. A single pass is decomposed as follows:

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

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

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1} \\ \Leftrightarrow y_t^2 &= \alpha_0 + (\alpha_1 + \beta) y_{t-1}^2 - \beta w_{t-1} + w_t \end{aligned} \tag{3.1}$$

where $w_t$ can be written as:

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

By construction, $\{w_t\}$ is a Martingale Difference process with variance $2h_t^2$ since the conditional expectation of $w_t$ with respect to $\mathcal{F}_{t-1}$ (the natural filtration up to time $t-1$) is zero and the $\chi_1^2$ variable has a unit mean and a variance equal to 2. However, as noted by Nakatsuma [8], it is difficult to generate $\Theta$ directly from equation (3.1). Hence, we approximate $w_t$ 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:

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

By noting that $z_t$ and $h_t$ are both functions of $\Theta$, respectively given by:

$$\begin{aligned} z_t(\Theta) &= y_t^2 - \alpha_0 - (\alpha_1 + \beta) y_{t-1}^2 + \beta z_{t-1}(\Theta) \\ h_t(\Theta) &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1}(\Theta) \end{aligned} \tag{3.2}$$

and by defining the $(T \times T)$ diagonal matrix $\Lambda := \Lambda(\Theta) = \mathrm{diag}(\{2h_t^2(\Theta)\}_{t=1}^T)$ and the vector $\boldsymbol{z} := (z_1 \cdots z_T)'$ we can approximate the likelihood function of $\Theta$ from the auxiliary model as follows:

$$l(\Theta|\boldsymbol{y}) \propto (\det \Lambda)^{-1/2} \exp\left[ -\frac{1}{2} \boldsymbol{z}' \Lambda^{-1} \boldsymbol{z} \right] \quad . \tag{3.3}$$

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

### 3.1 Generating $\boldsymbol{\alpha}$: ARCH coefficients

Recursive transformations initially proposed by Chib and Greenberg [3] allow to express the function $z_t(\Theta)$ in (3.2) as a linear function of $\boldsymbol{\alpha}$. Let us define $v_t := y_t^2$ for notational convenience. The recursive transformations are defined as follows:

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

where the initial values $l_0^*$ and $v_0^*$ are set to zero. Let us regroup the terms within vectors: $\boldsymbol{v} := (v_1 \cdots v_T)'$, $c_t := (l_t^* \ v_t^*)$ and construct the $(T \times 2)$ matrix $C$ where the $t$th row is $c_t$. It turns out that $\boldsymbol{z} = \boldsymbol{v} - C\boldsymbol{\alpha}$ and that we can express the likelihood function of $\boldsymbol{\alpha}$ as follows:

$$l(\boldsymbol{\alpha}|\beta, \boldsymbol{y}) \propto (\det \Lambda)^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{v} - C\boldsymbol{\alpha})'\Lambda^{-1}(\boldsymbol{v} - C\boldsymbol{\alpha})\right] \quad .$$

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

$$\begin{aligned} q_\alpha(\widetilde{\boldsymbol{\alpha}}, \boldsymbol{\alpha}) &\propto N_2(\boldsymbol{\alpha}|\widehat{\boldsymbol{\mu}}_\alpha, \widehat{\Sigma}_\alpha)\mathbb{I}_{[\boldsymbol{\alpha}>\mathbf{0}]} \\ \widehat{\Sigma}_\alpha^{-1} &:= C'\widetilde{\Lambda}^{-1}C + \Sigma_\alpha^{-1} \\ \widehat{\boldsymbol{\mu}}_\alpha &:= \widehat{\Sigma}_\alpha(C'\widetilde{\Lambda}^{-1}\boldsymbol{v} + \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha) \end{aligned}$$

where the $(T \times T)$ diagonal matrix $\widetilde{\Lambda} := \text{diag}(\{2\widetilde{h}_t^2\}_{t=1}^T)$ with:

$$\widetilde{h}_t := \widetilde{\alpha}_0 + \widetilde{\alpha}_1 y_{t-1}^2 + \beta\widetilde{h}_{t-1} \quad .$$

The value $\widetilde{\boldsymbol{\alpha}}$ is the previous draw of $\boldsymbol{\alpha}$ in the M-H sampler. A candidate $\boldsymbol{\alpha}^\star$ is sampled from this proposal distribution and accepted with probability:

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

### 3.2 Generating $\beta$: GARCH coefficient

The function $z_t(\Theta)$ in (3.2) could be expressed, in the previous section, as a linear function of $\boldsymbol{\alpha}$ but can not be expressed as a linear function of $\beta$. To overcome the problem, we linearize $z_t(\beta)$ by the first order Taylor expansion at point $\widetilde{\beta}$, that is:

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

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

$$\begin{aligned} r_t &:= z_t(\widetilde{\beta}) + \widetilde{\beta}\nabla_t \\ \nabla_t &:= -\left.\frac{dz_t}{d\beta}\right|_{\beta=\widetilde{\beta}} \end{aligned}$$

where the terms $\nabla_t$ can be computed by the following recursion:

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

with the initial value $\nabla_0 := 0$. Then, we regroup these terms into vectors: $\boldsymbol{r} := (r_1 \cdots r_T)'$, $\boldsymbol{\nabla} := (\nabla_1 \cdots \nabla_T)'$ and we approximate the exponential in (3.3) by:

$$\exp\left[-\frac{1}{2}(\boldsymbol{r} - \beta\boldsymbol{\nabla})'\Lambda^{-1}(\boldsymbol{r} - \beta\boldsymbol{\nabla})\right] \quad .$$

The proposal distribution to sample $\beta$ is obtained by combining the approximated likelihood and the prior distribution by Bayes' update:

$$\begin{aligned} q_\beta(\widetilde{\beta}, \beta) &\propto N(\beta|\widehat{\boldsymbol{\mu}}_\beta, \widehat{\Sigma}_\beta)\mathbb{I}_{[\beta>0]} \\ \widehat{\Sigma}_\beta^{-1} &:= \boldsymbol{\nabla}'\widetilde{\Lambda}^{-1}\boldsymbol{\nabla} + \Sigma_\beta^{-1} \\ \widehat{\boldsymbol{\mu}}_\beta &:= \widehat{\Sigma}_\beta(\boldsymbol{\nabla}'\widetilde{\Lambda}^{-1}\boldsymbol{r} + \Sigma_\beta^{-1}\boldsymbol{\mu}_\beta) \end{aligned}$$

where the $(T \times T)$ diagonal matrix $\widetilde{\Lambda} := \text{diag}(\{2\widetilde{h}_t^2\}_{t=1}^T)$ with:

$$\widetilde{h}_t := \alpha_0 + \alpha_1 y_{t-1}^2 + \widetilde{\beta}\,\widetilde{h}_{t-1} \quad .$$

A candidate $\beta^\star$ is sampled from this proposal distribution and accepted with probability:

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

We end this section with some comments regarding the implementation of the MCMC scheme. The program is written in the `R` language with some subroutines written 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 [7]. We sample $\Theta$ from a proper joint prior and generate some passes of the M-H algorithm; at each pass, we simulate the dependent variable $\boldsymbol{y}$ from the full conditional $p(\boldsymbol{y}|\Theta)$ (given by the conditional likelihood). This way, we draw a sample from the joint distribution $p(\boldsymbol{y}, \Theta)$. If the algorithm was correct, the resulting replications of $\Theta$ should reproduce the prior. The Kolmogorov-Smirnov comparison test does not reject this hypothesis at the 1% level.

# 4 Empirical analysis

We apply our Bayesian estimation to daily observations of the Deutschmark vs British Pound foreign exchange log-returns. The sample period is from January 3, 1985, to December 31, 1991, for a total of 1974 observations. The nominal returns are expressed in percent. This data set has been promoted as an informal benchmark for GARCH time series software validation and is available from the Journal of Business and Economic Statistics (JBES). From this time series, the first 750 observations, which is somewhat less than three financial years, are used to illustrate the Bayesian approach. The number of data is large enough to perform classical Maximum Likelihood (ML) estimation and apply asymptotic justifications. Hence, we have an interesting point of view from which to compare classical and Bayesian approaches.

The observation window excerpt from our data set is plotted on the left-hand side of Figure 1. We test for autocorrelation in the times series by testing the joint nullity of autoregression coefficients for $\{y_t\}$. We estimate the regression with autoregression coefficients up to lag 20 and compute the covariance matrix using the White estimate.

The $p$-valueof the Wald test is 0.377 which does not support the presence of autocorrelation. However, from Figure 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 on the right-hand side of Figure 1 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 absence of autocorrelation in the squares. As an additional data

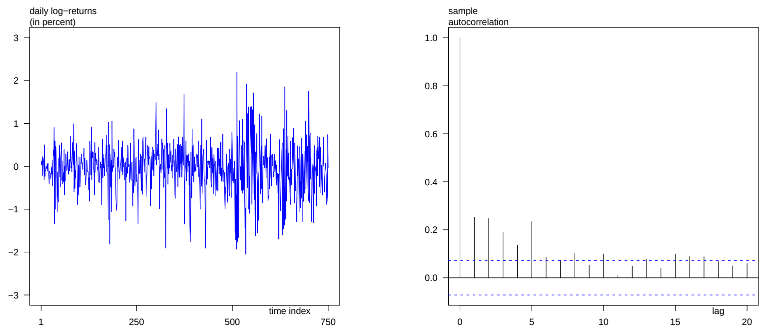


**Figure 1:** Daily log-returns (left) and sample autocorrelogram (right)

analysis, we test for unit root using the Phillips and Perron [10] test. The test strongly rejects the $I(1)$ hypothesis. From this preliminary 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.

## 4.1 Model estimation

We fit the parsimonious GARCH(1,1) model to the data for this observation window. As a prior distribution for the Bayesian estimation we choose a truncated tridimensional Normal distribution with a zero mean vector and a diagonal covariance matrix. The variances are set to 10'000 so we do not introduce tight prior information into our estimation. We run two chains for 10'000 passes each. We emphasize the fact that only positivity constraints are implemented in the M-H algorithm; no stationarity conditions are imposed in the simulation process.

In Figure 2, the running mean is plotted over iterations. For all parameters, we notice a convergence of the two chains toward a constant value after something like 5'000 iterations. The diagnostic test by Gelman and Rubin [5] does not reject convergence of the chain after 5'000 passes (values ranging from 1.04 to 1.05 for the 97.5th percentile of the potential scale reduction factor). The one lag autocorrelations in the chains range from 0.75 for $\alpha_1$ to 0.95 for $\beta$ which is reasonable. The sampling algorithm allows to reach very high acceptance rates ranging from 89% for $\boldsymbol{\alpha}$ to 95% for $\beta$. From the overall MCMC output, we discard the first 5'000 draws as a *burn in* period and merge the two chains to get a final sample's length of 10'000. In addition, we estimate the model by the usual Maximum Likelihood technique for comparison purposes.

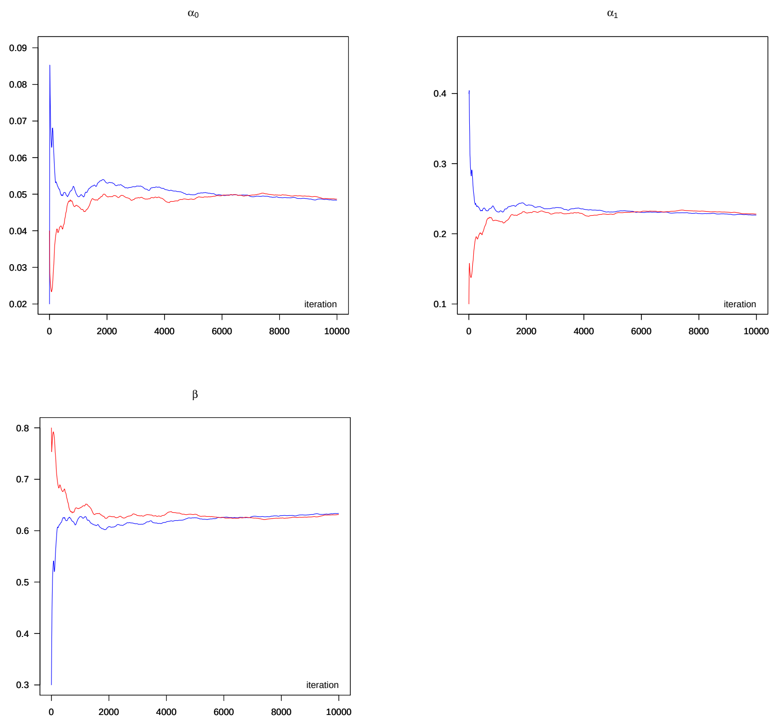


**Figure 2:** Running means of the chains over iterations (up to 10'000). The sampler generates $\boldsymbol{\alpha}$ and $\beta$ from the candidate distribution derived in Section 3. The acceptance rate ranges from 89% for $\boldsymbol{\alpha}$ to 95% for $\beta$. The autocorrelations range from 0.75 for $\alpha_1$ to 0.95 for $\beta$. The convergence diagnostic indicates convergence of the chains from iteration 5'000. The 97.5th percentile of the potential shrink factor ranges from 1.04 to 1.05.

The posterior statistics as well as the ML results are presented in Table 1. First, we note that even if the number of observations is large, the ML estimates and the Bayesian posterior means are different; the ML estimation is lower for $\boldsymbol{\alpha}$ and higher for $\beta$. 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 trough Figure 3 where the posterior distributions for the parameters are displayed. Indeed, we clearly notice the asymmetric shape of the histograms for $\alpha_0$ and $\alpha_1$. The skewness values are 0.46 and 0.39, both significantly different from zero. 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 $\beta$, the skewness is -0.09, also significant; in this case, the Maximum Likelihood overestimate the left boundary of the 95% confidence band. Furthermore, as shown in the bottom right-hand side of Figure 3, the joint distribution for the parameters $\alpha_0$ and $\beta$ is slightly different from the ellipsoid

| | $\Theta_{\text{MLE}}$ | $\overline{\Theta}$ | $\Theta_{0.5}$ | $\Theta_{0.025}$ | $\Theta_{0.975}$ | min | max | IF |
|---|---|---|---|---|---|---|---|---|
| $\alpha_0$ | 0.039 | 0.048 | 0.047 | 0.022 | 0.080 | 0.011 | 0.119 | 9.79 |
| | [0.014,0.064] | (0.448) | | | | | | |
| $\alpha_1$ | 0.198 | 0.226 | 0.223 | 0.128 | 0.337 | 0.083 | 0.499 | 5.85 |
| | [0.102,0.294] | (1.284) | | | | | | |
| $\beta$ | 0.686 | 0.636 | 0.636 | 0.476 | 0.795 | 0.338 | 0.849 | 40.79 |
| | [0.538,0.833] | (5.021) | | | | | | |

**Table 1:** Estimation results for the GARCH(1,1) model. $\Theta_{\text{MLE}}$: Maximum Likelihood estimate. $\overline{\Theta}$: posterior mean. $\Theta_\phi$: estimated posterior quantile at probability $\phi$. IF: inefficiency factor (ratio of the variance of mean relative to a $iid$ sequence). [•]: Maximum Likelihood 95% confidence interval. (•): numerical standard error ($\times 10^3$).

we should obtain under the multivariate Normal assumption. Therefore, these results warn us against the abusive use of asymptotic justifications; in the present case, even 750 observations do not suffice to assume the asymptotic Normal distribution for the parameters.

Finally, the last column of Table 1 gives the inefficiency factors for the different parameters. Their values are computed as the ratio of the squared numerical standard error 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 numerical standard errors are estimated by the method of Andrews [1], using a Parzen kernel and AR(1) pre-whitening as presented in Andrews and Monahan [2]. This ensures easy, optimal, and automatic bandwidth selection. 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 one percentage of the variation of the error due to the data. The larger inefficiency factor reported for $\beta$ is reflected in a larger autocorrelation in the simulated values.

## 4.2 Sensitivity analysis

The Bayesian approach is often criticized by the fact that the prior distribution for the parameters can have a significant impact to the posterior distribution, and as a consequence, bias the results. It is therefore important to determine the extend of this impact trough a sensitivity analysis. As noted by Geweke [6, Section 2], 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.

We test the sensitivity by considering four alternative prior distributions; either by modifying the mean or/and increasing the variances relative to our initial prior. The Bayes factors are then estimated as explained in the previous paragraph and ranked with the Jeffrey's scale of evidence. In all cases, we conclude to a weak evidence for our initial specification relative to the alternative prior, indicating that our initial prior is vague enough. The results are not shown to save space but can be obtained from the author upon request.

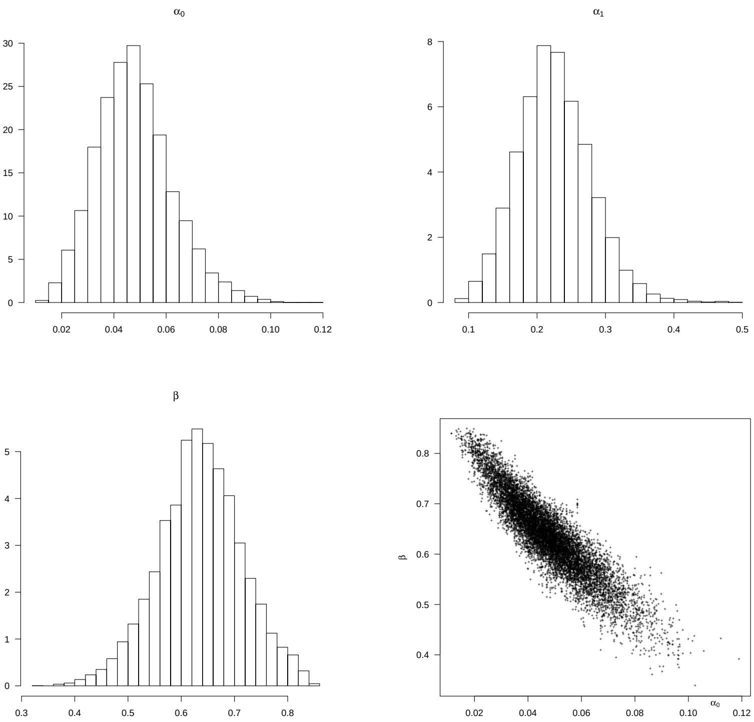


**Figure 3:** Posterior distributions for the GARCH(1,1) parameters based on 10'000 draws. In the lower-right graphic, we present a scatter plot of posterior $(\alpha_0, \beta)$

## 4.3 Model diagnostic

We check for model misspecification by testing the standardized residuals.[1] They are defined by:

$$\widehat{\varepsilon}_t := y_t \widehat{h}_t^{-1/2}$$

for $t = 1, \ldots, 750$ where $\widehat{h}_t$ is the conditional variance computed with $\Theta_{0.5}$ (the median of the posterior sample). If the statistical assumptions in (2.1) are satisfied, these residuals should be independent and Normally distributed.

On the left-hand side of Figure 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. The test does not reject the null at the 5% level ($p$-value= 0.652). This is also true for the squared residuals ($p$-value= 0.961). Hence, 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 on the right-hand side of Figure 4. The distribution is almost Normal at its center whereas the tails are slightly fatter, especially the left one. The Kolmogorov-Smirnov Normality

[1] An alternative would be to test the predictive performance of the model over an out-of-the-sample window (i.e. specification test). This approach is however not pursued in this article since we consider a single model and we focus on the estimation instead of the forecasting performance. A misspecification test is simpler and certainly sufficient in this context.

test rejects the null hypothesis at the 5% 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 is recurrent with financial data and heavy tails distributions for the innovations are sometimes useful to overcome this problem.

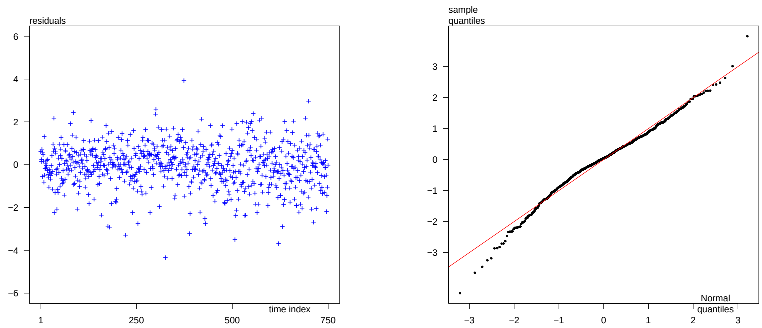


**Figure 4:** Residuals (left) and quantile-quantile plot (right)

# 5 Illustrative applications

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

## 5.1 Persistence

As pointed out in Section 3, 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 $-\beta$. The autocorrelation function (ACF) comes from the standard formulae for the ARMA(1,1) model. It is recursively given by $\rho_i := (\alpha_1 + \beta) \cdot \rho_{i-1}$ for $(i > 2)$ with the first order autocorrelation given by:

$$\rho_1 := \alpha_1 \frac{1 - \beta^2 - \alpha_1 \beta}{1 - \beta^2 - 2\alpha_1 \beta} \quad .$$

The term $(\alpha_1 + \beta)$ is the degree of persistence in the autocorrelation of the squares. It 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 persistence and ACF, we simply use the posterior sample $\Theta^{[j]}$ and generate $(\alpha^{[j]} + \beta^{[j]})$ and $\rho_i^{[j]}$ for $j = 1, \ldots, 10'000$ and $i = 1, \ldots, 20$. The graphic of the posterior distribution of the persistence $(\alpha_1 + \beta)$ is plotted on the left-hand side of Figure 5. The histogram is slightly 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. On the right-hand side of Figure 5 we display the posterior ACF with its 95% and 99% confidence bands together with the sample autocorrelations. 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.

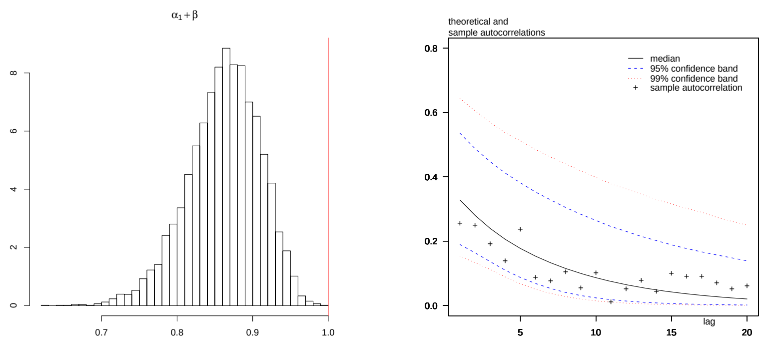


**Figure 5:** Posterior distribution for the persistence (left) and posterior autocorrelogram (right). The solid line is the posterior median, the dashed lines the 95% confidence bands and the dotted lines the 99% confidence bands. The cross symbols are values of the sample autocorrelation of the squared log-returns up to lag 20.

## 5.2 Stationarity

In the case of the GARCH(1,1) process, Nelson [9] gave the conditions for covariance stationarity (CSC) and strict stationarity (SSC). These conditions are given by:

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

where the error term $\varepsilon_t$ is Normally distributed. As pointed out in Section 4, the covariance stationary condition has not been imposed in the M-H algorithm. The joint posterior sample can be used to estimate the posterior distribution of these functions:

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

for $j = 1, \ldots, 10'000$, where $\eta^{[k]}$ is a draw from a standard Normal distribution and $K$ is set large enough (in our application we choose $K = 1'000$). In Figure 6 we present the posterior distributions for $CSC$ and $SSC$. None of these values exceed zero in our simulation study. The estimated model is therefore covariance stationary and strictly stationary. Other probabilistic statements on interesting functions can be

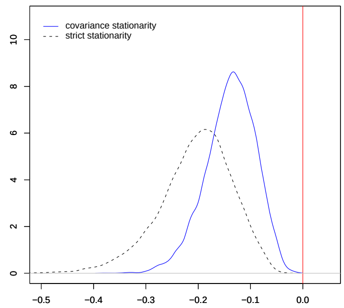


**Figure 6:** $CSC$ and $SSC$ posterior distributions. Gaussian kernel density estimates with bandwidth selected by the 'Silverman's rule of thumb' criterion (Silverman [11, page 48]).

obtained using the joint posterior sample. For example, the posterior median is 0.341 for the marginal variance and 4.54 for the marginal kurtosis. They approximately correspond to the sample estimations of 0.323 and 4.63.

# 6 Conclusion

This paper has proposed the estimation of the Bayesian GARCH(1,1) model with Normal innovations. The MCMC scheme has been derived in order to simulate the joint posterior distribution for the model's parameters. The GARCH(1,1) model has been applied to foreign exchange log-returns time series and comparison with the traditional Maximum Likelihood has been performed. It has been shown that even if the sample size is fairly large (in our case 750 observations), point estimates differ slightly between the two approaches. In addition, the posterior distribution for some parameters is skewed which warn us against the abusive use of the Normal approximation. A sensitivity analysis has been performed in order to robustify the estimation results. Finally, we have illustrated some appealing aspects of the Bayesian approach through interesting probabilistic statements made on the parameters.

As a final comment, we note that some financial models might use GARCH parameters as input quantities. This is the case for instance with the Black-Scholes formula in options pricing, which is a function of the marginal variance of the underlying financial asset. Under GARCH(1,1) dynamics, this marginal variance is $\alpha_0/(1-\alpha_1-\beta)$ if the process is covariance stationary, and this value can be simulated to obtain a full distribution. Then, the posterior distribution of the marginal variance could be used to simulate the option price's distribution. Furthermore, subjective constraints on the parameters could be integrated in the MCMC procedure; for instance, an option trader could set prior lower or/and upper boundaries for the unconditional variance and then run the estimation process to estimate the GARCH parameters.

# References

[1] D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, May 1991.

[2] D. W. K. Andrews and J. C. Monahan. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4):953–966, July 1992.

[3] Siddhartha Chib and Edward Greenberg. Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics*, 64:183–206, 1994.

[4] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1008, 1982.

[5] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

[6] John F. Geweke. Simulation methods for model criticism and robustness analysis. In J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 6, pages 275–299. Oxford: Oxford University Press, 1999.

[7] John F. Geweke. Getting it right: Joint distribution tests of posterior simulators. *Journal of the American Statistical Association*, 99(467):799–804, September 2004.

[8] Teruo Nakatsuma. A markov-chain sampling algorithm for GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, 3(2):107–117, 1998.

[9] Daniel B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991.

[10] P. C. B. Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75:335–346, 1988.

[11] B. W. Silverman. *Density estimation for statistics and data analysis*. London:Chapman and Hall, first edition, 1986.