Bayesian estimation of the GJR(1,1) model with student-t innovations and explanatory variables

Antoine Dubois Moustapha Ndiaye

January 2021

Contents

1 Introduction

Financial analysis has produced dozens of models intending to measure the relation between the previous and the later observations of a time series. Most of time, the frequentist approach is used to fit these models to data. Yet, during the last decade, Bayesian estimation methods have been developed. Many R packages, like bayesGARCH or MSGARCH, provide fast and accurate Bayesian estimation to the very general Markov-switching Glosten, Jagannathan and Runkle (GJR) model [Glosten et al., 1993]. However, to our knowledge, no package is able to consider the influence of exogenous variables on a time series of interest. For example, the price of an option can result from two effects. On the one hand, the price may be explained by a linear combination of its underlying asset. On the other side, the volatility of this price could be correlated with the volatility of the previous observations. Such an effect is called a GARCH effect or a cascade volatility effect.

Since the introduction of explanatory variables may improve the accuracy of a time series model, we propose a GJR model taking into account the influence of exogenous variables.

Firstly, we give an overview of Bayesian statistics. The benefits of this approach are discussed and a mathematical framework is defined. Also, we present the Metropolis sampler. It is a very popular tool for Bayesian models' estimation. Secondly, a generalised GJR model which introduces explanatory variables is defined. Thirdly, we present an R function which estimates the generalised GJR model using the Metropolis sampler. Incidentally, this function relies on stan language. Fourthly, we apply this code to real financial data. Ultimately, we conclude.

2 The Bayesian approach

Thanks to the increase of computational power, the Bayesian approach is now commonly used for time series analysis.

The Bayesian approach offers many benefits. Firstly, it provides an objective protocol for supplementing the information provided by the sample with a *priori* information on structural parameters. Secondly, the Bayesian method provides a natural and principled way of combining prior information with data, within a solid decision theoretical framework. Thus, past information about a parameter can be incorporated to compute a prior distribution which may be useful for any future statistical survey. When new observations are available, the previous posterior distribution can be used as a prior. Thirdly, inference is easy with the Bayesian approach.

Now, let us define a thorough framework to the Bayesian approach.

2.1 Prior, likelihood and posterior

First of all, this approach assumes that a, finite or infinite, sequence of observations $y := (y_1, \dots, y_t, \dots)$ is a realisation of the sequence of random variables $Y := (Y_1, \dots, Y_t, \dots)$. Moreover, the elements of this random sequence are supposed to be independent and identically distributed. For any i, the law of Y_i is $\mathscr{L}(\theta)$. This law depends on the parameter $\theta \in \mathbb{R}^q$. Furthermore, the probability of $Y = y$ given θ, denoted $\mathbb{P}_\theta(y)$, is called the likelihood of the model. Ultimately, θ is assumed to be the output of the random variable Θ whose law is $\mathscr{L}^p$. Incidentally, $\mathscr{L}^p$ is called a prior law.

The Bayesian approach aims to deduce θ when some y is observed given the prior law $\mathscr{L}^p$and the law of y given θ, $\mathscr{L}(\theta)$. To do so, the value of $\mathbb{P}(\Theta)$ and the value of $\mathbb{P}_\theta(y)$ are deduced from $\mathscr{L}(\theta)$ and $\mathscr{L}^p$ for any $\theta \in \mathbb{R}^q$. Afterwards, the posterior distribution $\mathbb{P}(\Theta|y)$ is computed using the Bayes formula for any $\theta \in \mathbb{R}^q$. This probability is defined by

$$\mathbb{P}(\Theta = \theta|y) = \frac{\mathbb{P}_\theta(y)\mathbb{P}(\Theta = \theta)}{\int \mathbb{P}_{\theta'}(y)\mathbb{P}(\Theta = \theta')\, d\theta'}$$

Furthermore, $\mathbb{P}(\Theta = \theta|y)$ is called the joint posterior distribution. In contrast, the posterior marginal distribution of θ_i is defined by

$$\mathbb{P}(\Theta_i = \theta_i|y) = \int \mathbb{P}(\Theta = \theta|y)\, d\theta_{-i}$$

Naturally, someone using the Bayesian approach intends to find the region where the parameter θ is the most likely to occur. The Metropolis algorithm is a sampler which returns a sequence of parameters θ which maximises the joint posterior distribution.

2.2 Metropolis algorithm

The algorithm of Metropolis generates a sequence of parameters $(\theta^j)_{j\in\{1,\dots,j\}}$. According to [Tsvetkov et al., 2013], the sequence generated by the Metropolis algorithm converges towards the best estimate of θ. In addition, the whole generated sequence is useful.

In practice, as proved by [Deschamps, 2006], the sequence $(\theta^j)_{j=1}^J$ provides information about the prior distribution. To reveal a piece of information, let us introduce ξ a function on $\mathbb{R}^q$ and $\bar{\xi} := \frac{1}{J}\sum_{j=1}^J \xi(\theta^j)$. Then, by the law of large numbers, $\bar{\xi}$ converges towards $\mathbb{E}(\xi(\Theta)|y)$.

Henceforth, the choice of ξ determines the piece of information to be revealed. For example, if we have $\xi = \mathbb{1}_C$ then $\bar{\xi}$ converges towards $\mathbb{P}(\Theta^{-1}(C)|y)$. Furthermore, if ξ is the identity function, then $\bar{\xi}$ converges towards the posterior expectation of Θ.

The Metropolis algorithm is a Bayesian convenient tool to evaluate time series models.

Algorithm 1 Metropolis algorithm

θ^0 ▷ Initial value of θ
$j \leftarrow 1$
while Stopping criteria not satisfied **do** ▷ Such a criteria may be complex
    θ^* generated by a random variable of density $q(\cdot|\theta^{j-1})$
    $p = min\left(\frac{\mathbb{P}(\theta^*|y)}{\mathbb{P}(\theta^{j-1}|y)}; 1\right)$
    $U \sim \mathscr{B}(p)$
    if $U = 1$ **then**
        $\theta^j = \theta^*$
    else $\theta^j = \theta^{j-1}$
    $j \leftarrow j + 1$
return $(\theta^j)_{j\in\{1,\dots,J\}}$ ▷ The sequence of approximations of θ

3 Student-t-GJR(1,1) Model

The generalised autoregressive conditional heteroskedasticity (GARCH) model is a general framework for understanding complex time series. In this model, the perturbations are not supposed independent nor identically distributed. Moreover, the generation of a perturbation depends on the square of the previous perturbations.

Authors have designed a plethora of GARCH model variants. In the Markov-switching Glosten, Jagannathan and Runkle (GJR) model, the volatility of the perturbations depends both on the sign and the square value of the previous perturbations.

Below, we define the GJR(1,1) model with Student-t innovations and explanatory variables. This model is detailed by Ardia [Ardia et al., 2008].

$$y_t = x_t^{'}\gamma + u_t \quad \text{for} \quad t \in \{1, \cdots, T\} \tag{1}$$
$$u_t = \varepsilon_t \sqrt{\rho h_t} \tag{2}$$
$$\varepsilon_t \sim \mathcal{T}(0, 1, \nu) \tag{3}$$
$$\rho = \frac{\nu - 2}{\nu} \tag{4}$$
$$h_t = \alpha_0 + (\alpha_1 \mathbb{1}_{\{u_{t-1}\geq 0\}} + \alpha_2 \mathbb{1}_{\{u_{t-1}<0\}}) u_{t-1}^2 + \beta h_{t-1} \tag{5}$$

where $\alpha_0, \alpha_1, \alpha_2, \beta \geq 0$, and $\nu > 2$. Also, y_t is a scalar dependent variable, x_t is a $m \times 1$ vector of exogenous or lagged dependent variables. Incidentally, h and y are initialised at $h_0 = y_0 = 0$ for convenience. Furthermore, γ is a $m \times 1$ vector of regression coefficients and $\mathcal{T}(0, 1, \nu)$ is the standard Student-t density with ν degrees of freedom. Finally, h_t is the conditional variance. It is defined as a piecewise linear combination of the squared past shocks u_{t-1} and the past conditional variance h_{t-1}.

On another note, the reader may question the reason why h_t depends on the square value of u_{t-1}.

Firstly, the variance of u_{t-1} is defined by

$$\mathbb{V}(u_{t-1}) = \rho h_{t-1} \mathbb{V}(\epsilon_{t-1}) = h_{t-1}$$

since $\nu > 2$ and $\mathbb{V}(\epsilon_{t-1}) = \frac{\nu}{\nu-2} = \frac{1}{\rho}$. Secondly, the expectation of u_{t-1} is

$$\mathbb{E}(u_{t-1}) = \rho h_{t-1} \mathbb{E}(\epsilon_{t-1}) = 0$$

since $\nu > 1$. Thereby, we have $\mathbb{V}(u_{t-1}) = \mathbb{E}(u_{t-1}^2)$.

To sum up, h_t depends on the square value of u_{t-1} because u_{t-1}^2 is the empirical variance of U_{t-1}.

The GJR differs from the standard GARCH model. Firstly, this variant assumes that the distribution of the perturbation only depends on the last value of the noise. Secondly, the explained variables Y_t are partially explained by a linear combination of the dependent variables X_t and γ. Finally, the weight associated with the square of the last shock differs according to its sign. If the previous perturbation is positive, the weight is α_1, and α_2 otherwise. These characteristics distinguish the GJR model from the GARCH model. To conclude, the GJR model is a generalised GARCH model.

3.1 Likelihood

As stated in equation (3), the innovations are t-student distributed. In consequence, y_t is t-student distributed with localisation $x_t \cdot \gamma$, scale $\sqrt{\rho h_t}$ and ν degrees of freedom. In consequence, the whole time series Y follows the multivariate t-student distribution

$$Y \sim \mathcal{T}\left(X\gamma, \Sigma, \nu\right)$$

where Σ is the diagonal covariance matrix defined by $\Sigma = \text{diag}(\{\rho h_t(\gamma, \alpha, \beta)\}_{t=1}^T)$, with $\alpha := (\alpha_0, \alpha_1, \alpha_2)$.

3.2 Prior laws

The choice of the prior laws of α, β, γ and ν is crucial for the accuracy of the model fitting. In consequence, we rely on the expertise of [Ardia et al., 2008].

Firstly, α follows a three-dimensional normal law truncated at zero. The expectation and covariance of α are the hyperparameters μ_α and Σ_α.

$$p(\alpha) \propto \mathcal{N}_3\left(\alpha \,|\, \mu_\alpha, \Sigma_\alpha\right) \mathbb{1}_{\alpha \geq 0}$$

Secondly, β follows a normal law truncated at zero. Its expectation and variance are given by the hyperparameters μ_β and Σ_β.

$$p(\beta) \propto \mathcal{N}\left(\beta \,|\mu_\beta, \Sigma_\beta\right) \mathbb{1}_{\beta \geq 0}$$

Thirdly, γ follows a multivariate normal law. Naturally, the dimension of this distribution is equal to the dimension of X_t. Its expectation and variance are the hyperparameters μ_γ and Σ_γ.

$$p(\gamma) = \mathcal{N}_m\left(\gamma|\mu_\gamma, \Sigma_\gamma\right)$$

Fourthly, ν follows a translated exponential law with parameters $\lambda > 0$ and $\delta \geq 2$.

$$p(\nu) = \lambda \exp\left[-\lambda(\nu - \delta)\right] \mathbb{1}_{\nu>\delta}$$

The translation is important. Indeed, a t-student distribution has infinite variance when $1 < \nu \leq 2$ and is even undefined when $\nu \leq 1$. In consequence, a small ν makes Y to have infinite or undefined variance. In such a case, no sampler would generate convergent series within a reasonable amount of time.

4 The code

During the last few years, stan has demonstrated that it was a reliable, accurate and fast language for Bayesian statistics modelling. Moreover, the Metropolis algorithm is already implemented in stan which avoids user's mistakes. For these reasons, we coded a normal-GJR(1,1) and a student-t-GJR(1,1) model with explanatory variables in this language.

Furthermore, the function **GJR11**, coded in R, described in Appendix B evaluates by default the model with normal innovations. Yet, by defining **student=TRUE**, the function evaluates the GJR model with student-t innovations.

Ultimately, the function **GJR11** returns chains produced by Metropolis algorithm. In addition, using the function **summary** the user can behold a table displaying the average, the quantiles and the standard deviation of each parameter. Also, this table shows the size of the effective samples and the $\hat{R}$ of each parameter[1].

5 Application

Financial data often describes Markov-switching effects. However, programs carrying out a Bayesian estimation of the GJR(1,1) model taking into account explanatory variables are uncommon. The function **GJR11** provides a Bayesian estimation of a GJR(1,1) model with as many explanatory variables as desired.

To illustrate how an analyst could use the student-t-GJR(1,1) model with explanatory variables, we intend to explain the relation between the rate exchange of the Euro to the United States dollar, the exchange rate of the Great Britain pound sterling and the Japanese Yen to the United States dollar[2] from

[1] $\hat{R}$ is an index comparing Markov chains with different initialisation values. Two chains are considered to have converged towards the same distribution when $\hat{R}$ belongs to $[1; 1.1[$.

[2] The data comes from the website https://www.macrotrends.net

the 20^{th} to 30^{th} of July 2020.

Before running the function **GJR11**, we compute the auto-correlation of the data. We obtain Figure 1, in Appendix A.1. In this figure, we notice that there is no auto-correlation between the three different time series. In consequence, the relation between the three exchange rates may admit GARCH and Markov-switching effects.

After running twelve minutes the function **GJR11** and defining as 100 000 the number iterations of the Metropolis algorithm, we obtain the fourth-decimal-rounded estimations of α, β, γ and ν described in Table 1. This cumbersome table is in Appendix A.2.

In Table 1, we first notice that the $\hat{R}$ are equal to 1 with an accuracy of four decimals. Thereby, the chains are considered to have converged towards the same stationary distribution. Secondly, while the standard derivation of γ_1 is small (0.0555), that of γ_2 is wide (7.3514). Thirdly, α_0 is almost surely equal to 0.

Fourthly, α_1 and α_2 are unlikely to be nought. In consequence, the relation between the three different exchange rates display GARCH effects. Furthermore, α_1 and $\alpha_2$have the same confidence interval with 95% certainty with an accuracy of 2 decimals. They both are $[0.59; 1.39]$. This similarity would suggest that the relation between the three exchange rates does not admit Markov-switching effects. However, the average absolute difference between α_1 and α_2 is equal to 0.23. Thus the average distance between α_1 and α_2 is different from 0. Thereby, the relation may admit marginal Markov-switching effects.

The reader may take a look at Figure 4 to 8 in Appendix A.3 to better visualise the distribution of the model's parameters. Incidentally, Figure 9 provides an overview of the log-posterior distribution.

6 Conclusion

The Bayesian approach is a very convenient way to estimate time series models. In addition, this approach offers better accuracy when data is scanty. Furthermore, Bayesian estimator of the Markov-switching Glosten, Jagannathan and Runkle model which takes into account explanatory variables is tractable. Moreover, the programming of this estimator is easy and computationally efficient when combining R and stan languages.

With our code, an analyst chooses first the prior laws of each parameter as well as several time series of the same length. Then, he defines the series to be explained and those to be explanatory. Nevertheless, a careful analyst should measure the bias introduced by the choice of the prior laws. To do so, we recommend the computation of Bayesian coefficients.

References

[Ardia et al., 2008] Ardia, D. et al. (2008). *Financial risk management with Bayesian estimation of GARCH models*, volume 18. Springer.

[Deschamps, 2006] Deschamps, P. J. (2006). A flexible prior distribution for markov switching autoregressions with student-t errors. *Journal of Econometrics*, 133(1):153–190.

[Glosten et al., 1993] Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801.

[Tsvetkov et al., 2013] Tsvetkov, D., Hristov, L., and Angelova-Slavova, R. (2013). On the convergence of the metropolis-hastings markov chains. *arXiv preprint arXiv:1302.0654*.

Appendix

A Bayesian estimation for the Student-t-GJR(1,1) model with explanatory variables

A.1 Autocorrelation between time series

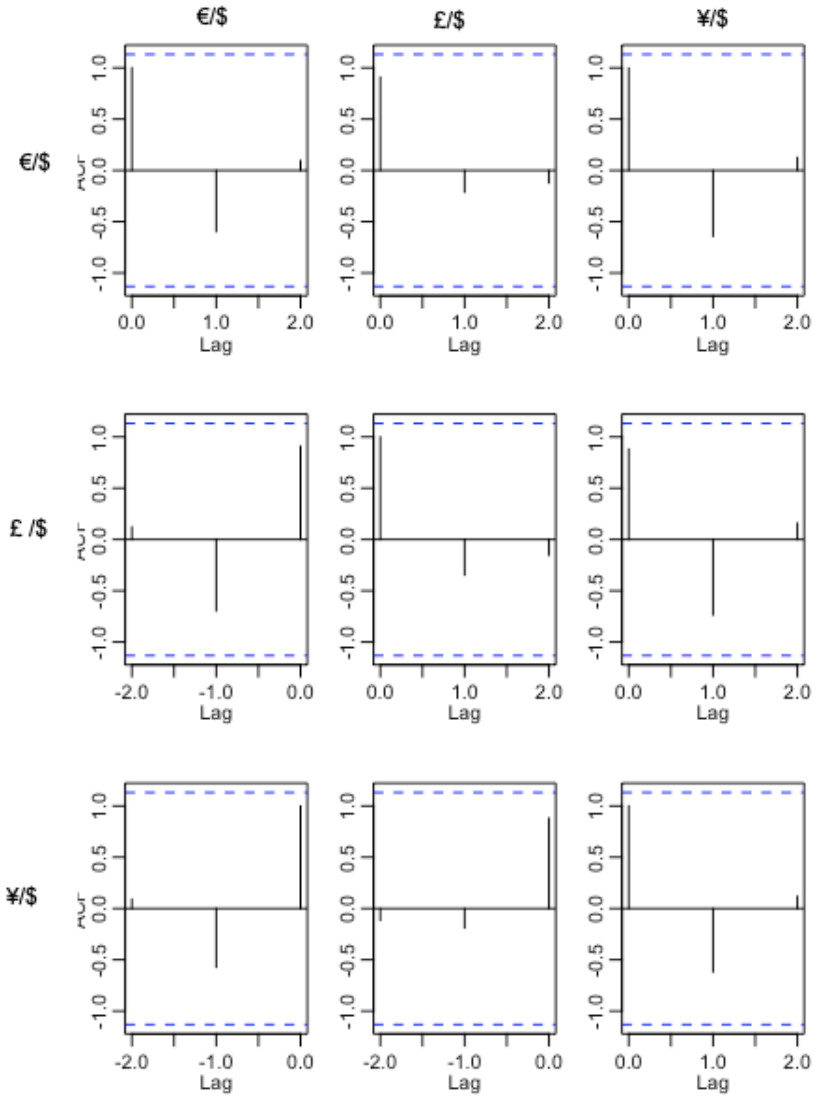

Figure 1: Correlation between the three different time series for different gaps

A.2 Table of the estimated parameters

Table 1: Estimation results for the Student-t-GJR(1,1) model with explanatory variables

Variable	$mean$	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	$\hat{R}$
γ_1	1.0624	0.0002	0.0555	0.9542	1.0246	1.0623	1.0999	1.1713	84242.80	1.0000
γ_2	−20.8665	0.0253	7.3514	−35.2903	−25.8310	−20.8643	−15.8642	−6.5193	84280.41	1.0000
α_0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	105870.71	1.0000
α_1	0.9907	0.0005	0.1999	0.5974	0.8554	0.9907	1.1263	1.3826	145496.51	1.0000
α_2	0.9844	0.0005	0.2009	0.5881	0.8493	0.9850	1.1204	1.3772	138034.53	1.0000
β	0.8933	0.0001	0.0497	0.7956	0.8599	0.8933	0.9267	0.9910	156971.16	1.0000
ν	3.4257	0.0010	0.4412	3.0104	3.1175	3.2865	3.5854	4.6239	206514.15	1.0000

A.3 Posterior distribution and likelihood

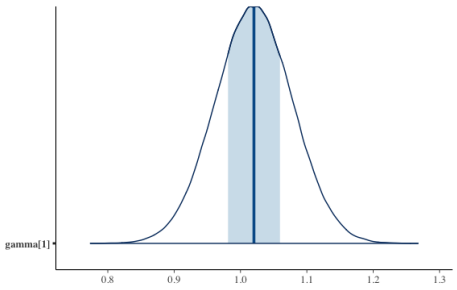

Figure 2: Posterior distribution of γ_1

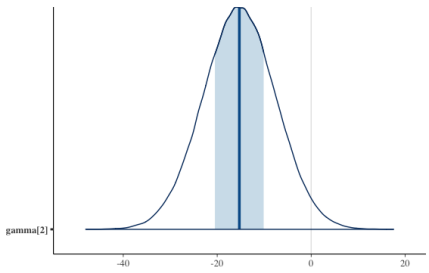

Figure 3: Posterior distribution of γ_2

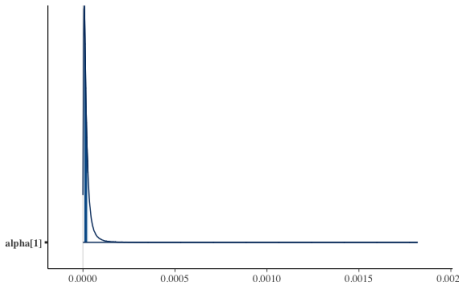

Figure 4: Posterior distribution of α_0

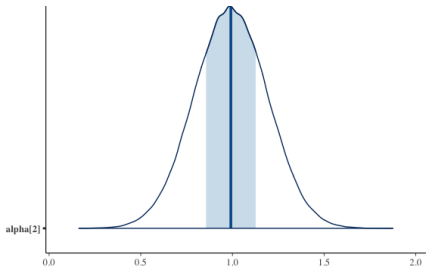

Figure 5: Posterior distribution of α_1

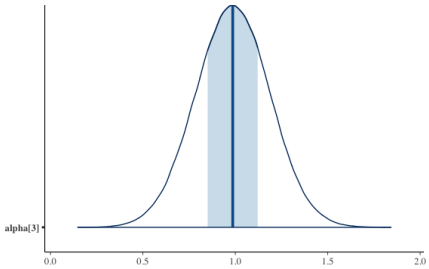

Figure 6: Posterior distribution of α_2

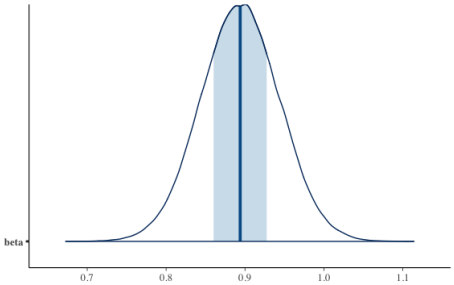

Figure 7: Posterior distribution of β

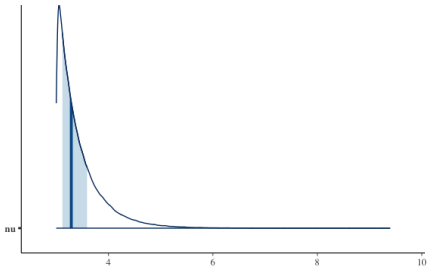

Figure 8: Posterior distribution of ν

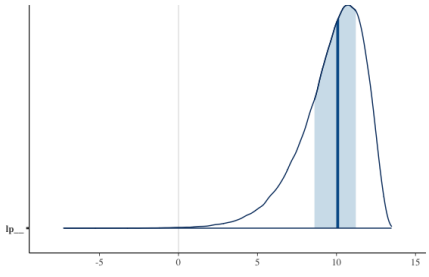

Figure 9: Posterior log-likelihood

B Fit the normal-GJR(1,1) or student-t-GJR(1,1) model with GJR11

GJR11(**y**, **X**,
mu_alpha=rep(1,3),**sigma_alpha**=diag(x=rep(0.01,3,nrow=3,ncol=3)),
mu_beta=1, **sigma_beta**=0.01,
mu_gamma=rep(1,ncol(X)), **sigma_gamma**=diag(x=rep(0.01,ncol(X)),
row=ncol(X), ncol=ncol(X)),
lambda=1, **delta**=0, **student**=FALSE,
iter=3000, **chains**=1, **cores**=1, **init**=0.1)

Arguments

- **y** The response variable vector.
- **X** The explanatory variables matrix.
- **mu_alpha** The mean of α.
- **sigma_alpha** The covariance matrix of α.
- **mu_alpha** The mean of β.
- **sigma_alpha** The standard deviation of β.
- **mu_alpha** The mean of γ.
- **sigma_alpha** The covariance matrix of γ.
- **lambda** The mean of ν.
- **delta** The translation coefficient of ν. It should exceed 2.
- **student** TRUE or FALSE: flag indicating whether to evaluate the student-t-GJR(1,1) or the normal-GJR(1,1) model.
- **iter** The number of iterations of the MCMC algorithm.
- **chains** The number of chains generated by the MCMC algorithm.
- **cores** The number of cores to use when executing the Markov chains in parallel.
- **init** Initial values for all or some parameters.