da Fonseca, Thaís C. O.; Cerqueira, Vinícius dos Santos; Migon, Hélio dos Santos; Torres, Cristian A. C.

Working Paper

Full Bayesian inference for asymmetric Garch models with student-t innovations

Discussion Paper, No. 215

Provided in Cooperation with:
Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: da Fonseca, Thaís C. O.; Cerqueira, Vinícius dos Santos; Migon, Hélio dos Santos; Torres, Cristian A. C. (2016) : Full Bayesian inference for asymmetric Garch models with student-t innovations, Discussion Paper, No. 215, Institute for Applied Economic Research (ipea), Brasília

This Version is available at:
https://hdl.handle.net/10419/220303

215
DISCUSSION PAPER

FULL BAYESIAN INFERENCE FOR ASYMMETRIC GARCH MODELS WITH STUDENT-T INNOVATIONS

Thaís C. O. Fonseca
Vinícius S. Cerqueira
Helio S. Migon
Cristian A. C. Torres

ipea

215
DISCUSSION PAPER
Brasília, october 2016

FULL BAYESIAN INFERENCE FOR ASYMMETRIC GARCH MODELS WITH STUDENT-T INNOVATIONS

Thaís C. O. Fonseca[1]
Vinícius S. Cerqueira[2]
Helio S. Migon[3]
Cristian A. C. Torres[4]

1. Department of Statistics, Federal University of Rio de Janeiro (UFRJ), Brazil.
2. Researcher at Ipea, Brazil.
3. Department of Statistics, UFRJ, Brazil.
4. Department of Statistics, UFRJ, Brazil.

Federal Government of Brazil

Ministry of Planning, Development and Management
Minister, Interim Dyogo Henrique de Oliveira

ipea **Instituto de Pesquisa Econômica Aplicada**

A public foundation affiliated to the Ministry of Planning, Development and Management, Ipea provides technical and institutional support to government actions – enabling the formulation of numerous public policies and programs for Brazilian development – and makes research and studies conducted by its staff available to society.

President
Ernesto Lozardo

Director of Institutional Development
Juliano Cardoso Eleutério

Director of Studies and Policies of the State, Institutions and Democracy
João Alberto De Negri

Director of Macroeconomic Studies and Policies
Claudio Hamilton Matos dos Santos

Director of Regional, Urban and Environmental Studies and Policies
Alexandre Xavier Ywata de Carvalho

Director of Sectoral Studies and Policies, Innovation, Regulation and Infrastructure
Fernanda De Negri

Director of Social Studies and Policies
Lenita Maria Turchi

Director of International Studies, Political and Economic Relations
Alice Pessoa de Abreu

Chief of Staff
Márcio Simão

Chief Press and Communications Officer
Regina Alvarez

URL: http://www.ipea.gov.br
Ombudsman: http://www.ipea.gov.br/ouvidoria

DISCUSSION PAPER

A publication to disseminate the findings of research directly or indirectly conducted by the Institute for Applied Economic Research (Ipea). Due to their relevance, they provide information to specialists and encourage contributions.

Discussion paper / Institute for Applied Economic Research.- Brasília : Rio de Janeiro : Ipea, 1990-

ISSN 1415-4765

1. Brazil. 2. Economic Aspects. 3. Social Aspects. I. Institute for Applied Economic Research.

CDD 330.908

JEL: C11; C22.

CONTENTS

ABSTRACT

In this work, we consider modeling the past volatilities through an asymmetric generalised autoregressive conditional heteroskedasticity (Garch) model with heavy tailed sampling distributions. In particular, we consider the Student-t model with unknown degrees of freedom and indicate how it may be used adequately from a Bayesian point of view in the context of smooth transition models for the variance. We adopt the full Bayesian approach for inference, prediction and hypothesis testing. We discuss problems related to the estimation of degrees of freedom in the Student-t model and propose a solution based on independent Jeffreys priors, which correct problems in the likelihood function. A simulated study is presented to investigate how estimation of model parameters in the Student-t Garch model are affected by small sample sizes, prior distributions and mispecification regarding the sampling distribution. An application to the Dow Jones stock market data illustrates the usefulness of the asymmetric Garch model with Student-t erros. In this context, the Student-t model is preferable for prediction in the case of high volatility regimes.

Keywords: Student-t distribution; Garch model; Bayesian approach; Jeffreys prior.

1 INTRODUCTION

In this paper we consider modeling the future volatility using a generalised autoregressive conditional heteroskedasticity (Garch) model on the past error terms and volatilities. The Garch approach (Bollerslev, 1986) models the variance as a function of past values and an error term assuming that the variance is independent of shocks in the mean. The main advantage of these models is that they allow for accessing directly the effect of mean changes in the dynamics of the conditional variance. This is an important issue in finantial time series, as changes in the mean tend to have relevant impact in the uncertainty of the process under study. Several authors have considered Garch models for finantial time series. In particular, from a Bayesian perspective, Vrontos et al. (2000) proposed full Bayesian analysis for Garch models with Gaussian errors. An alternative approach is presented in Jacquier et al. (2004) known as stochastic volatility modeling which allow for correlated conditional variance and mean. However, it is difficult to directly evaluate the past effects of the mean in the conditional variance.

A stylized fact that needs to be captured by the models is that the conditional variance can react asymmetrically to positive versus negative shocks or large versus small shocks. That is, the conditional variance may follow different regimes according to the size and signal of the shock. News have asymmetric impact in the economy and, for example, a large negative return might affect the future volatility in a different way when compared to a positive return with the same size. Engle and Ng (1993) presents a review regarding this issue. Awartani and Corradi (2005) discuss the importance of asymmetries in the prediction of an economic index. This may be accommodated by smooth transition models based on asymmetric specification of the conditional variance model. Thus, different specifications of the skedastic functions will take into account size and sign effects in the volatility. In this work we consider smooth transition Garch models and discuss inferential issues related to the smooth transition function and Bayesian inference.

Regarding the sampling distribution for the error term, shocks are usually modeled as Gaussian distributed due mainly to mathematical convinience rather than being suitable for financial data. It is well known that financial time series exhibit havier tails than allowed by the usual Gaussian model (Bollerslev et al., 1992). In this work we relax the assumption of Gaussian errors and consider Student-t distributios for the error terms in the Garch model. Since Mandelbrot (1963) several authors have discussed the issue of fat tails in return datasets. Bollerslev (1987) introduced the Garch-t model

as a solution to the typical heavy tails of returns. Also in the context of robust analysis, Harvey and Chakravarty (2008) proposed a Beta-t-EGarch in which the volatilities depend on the score of a t distribution. Zhang et al. (2011) proposed to use Generalized Hyperbolic distributions to model volatilities to capture fat tails and skewness. Bauwens and Lubrano (2002) comment on how the introduction of Student-t errors in the Garch model may improve the fit to the data. However, the likelihood is ill behaved as discussed in Bauwens and Lubrano (1998) and Fonseca et al. (2008). Bauwens and Lubrano (1998) propose the use of Griddy-Gibbs sampler, which would not work in the cases where the likelihood is monotonic (Fonseca et al., 2008). Also Ardia (2008) describes Bayesian approach to Student-t Garch models using modified exponential prior distributions. This proposal would not work either for the situation where the likelihood is monotonic in the parameters. In this case, the choice of prior distributions may dominate the inference and posterior distributions will be similar to the prior selected. In this work, the degrees of freedom are estimated using the independent Jeffreys priors presented in Fonseca et al. (2008) which corrects the problems in the likelihood function for the Student-t model. Our proposal is a noninformative prior and do not depend on the specification of hyperparameters. This prior give the correct information regarding the curvature of likelihood functions and provide better results than the maximum likelihood estimator and informative priors. We investigate how estimation of model parameters in the Student-t Garch model are affected by small sample sizes, prior distributions and mispecification regarding the sampling distribution.

In section 2, we present the autoregressive moving average (Arma) model with a Garch component. We flexibilize the Gaussian assumption and consider Student-t error terms. We discuss the main issues related to the likelihood function and estimation of parameters such as the degree of freedom which is not usually well estimated in the literature. We present the prior considered to correct the problems with the Student-t likelihood and simulated examples which illustrate the effects of model misspecification. Section 3 presents the asymmetric Garch model and the proposed prior distribution for the parameters of interest. The likelihood issues are discussed in the context of asymmetric models. Section 4 presents a simulation study to evaluate the performance of Bayesian estimators and Bayesian model selection. An application to the Dow Jones returns is presented in section 5. Section 6 concludes the work with main results and future developments of the proposed models.

2 ARMA-GARCH-M MODELS

Consider a univariate time series y_t indexed in discrete time $t \in N_+$. For the mean term we assume an autoregressive moving average (Arma) model and add a heteroskedasticity term (M).

$$y_t = \mu + \sum_{j=1}^{p} \varphi_j y_{t-j} + \sum_{j=1}^{q} \theta_j u_{t-j} + \delta\sqrt{h_t} + u_t, \tag{1}$$

where u_t are error terms with variance h_t which are often modeled as Gaussian distributed, $\varphi_1,\ldots, \varphi_p$ are autoregression coefficcients, $\theta_1,\ldots, \theta_q$ are moving average coefficcients and δ is a parameter allowing for a direct effect of the variance term in the mean. This model is denoted as Arma (p, q) -M (1). As follows we consider Student-t error terms and exploit the mixture model representation which could be used to capture extreme observations in financial and economic time series as follows.

2.1 Student-t innovations with unknown degrees of freedom

Define the error term u_t as a function of a white noise ε_t and a positive mixing random variable ω_t as follows.

$$u_t = \varepsilon_t \left(\frac{v-2}{v} h_t \omega_t\right)^{1/2}, \; t = p+1,\ldots, N, \tag{2}$$

where $\varepsilon_t \sim N(0, 1)$, $\omega_t \sim IG(v/2, v/2)$. Here $N(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2 and $IG(a, b)$ denotes the Inverse Gamma distribution with mean a/b and variance a/b^2. The parameter $v \in \Re_+$ is responsible for the heavy tail of the sampling distribution and is considered to be an unknown constant. The mixture representation is obtained by considering the observation equation (1) and the error term specification (2). Notice that as $v \to \infty$ then $u_t \sim N(0, h_t)$, while for finite v the distribution of ut will be Student-t with v degrees of freedom. In the mixture representation, ω_t is responsible for inflating the variance h_t. This is an important modeling tool in the identification of periods of larger volatility in the series. The marginal density of u_t is the Student-t model given by

$$f(u_t; \nu, h_t) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} ((v-2)h_t)^{-1/2} \left(1 + \frac{u_t^2}{(v-2)\, h_t}\right)^{-(v+1)/2}. \tag{3}$$

2.2 Garch specification

The variances h_t are considered to be heterocedastic and given by the Garch model which we denote by Garch(r, s).

$$h_t = \omega + \sum_{j=1}^{r} \beta_j h_{t-j} + \sum_{j=1}^{s} \alpha_j u_{t-j}^2 , \tag{4}$$

with the restriction $\sum_{j=1}^{r} \alpha_j + \sum_{j=1}^{s} \beta_j < 1$, $\omega > 0$, $\alpha_i, \beta_j \geq 0$. Without loss of generality consider $\mu = 0$. Consider observations $\mathbf{y_0} = (y_1,..., y_p)'$ as known and set $u_p = u_{p-1} = ... = u_{p-q} = u_{p-q+1} = 0$. From equation (1) we may define u_t, $t = p + 1,..., N$ recursively by

$$u_t = y_t - \sum_{j=1}^{p} \varphi_j \, y_{t-j} - \sum_{j=1}^{q} \theta_j u_{t-j} - \delta \sqrt{h_t}. \tag{5}$$

From an inferential point of view, it is convinient to rewrite Equation (1) as

$$y = X\varphi + A\theta + \tilde{H}\psi + u \tag{6}$$

where $\mathbf{y} = (y_{p+1},..., y_N)'$, $\varphi = (\varphi_1,..., \varphi_p)'$, $u = (u_{p+1},..., u_N)'$, $\theta = (\theta_1,..., \theta_q)'$, $\psi = (\psi_1,..., \psi_k)'$ and

$$X = \begin{pmatrix} y_p & y_{p-1} & \dots & y_1 \\ y_{p+1} & y_p & \dots & y_2 \\ \vdots & \vdots & \vdots & \vdots \\ y_{N-1} & y_{N-2} & \dots & y_{N-p} \end{pmatrix}, \; A = \begin{pmatrix} u_p & \dots & u_{p-q+1} \\ u_{p+1} & \dots & u_{p-q+2} \\ \vdots & \vdots & \vdots \\ u_{N-1} & \dots & u_{N-q} \end{pmatrix}, \; \tilde{H} = \begin{pmatrix} h_p & \dots & h_{p-k+1} \\ h_{p+1} & \dots & h_{p-k+2} \\ \vdots & \vdots & \vdots \\ h_{N-1} & \dots & h_{N-k} \end{pmatrix}.$$

In the next section, we allow for asymmetric shock effects in order to consider more realistic set-ups for finantial time series analysis. However, the likelihood function for this model is ill-behaved. This issue will be discussed and a solution will be proposed in section 3. As follows we present the asymmetric Student-t Garch model.

2.3 Asymmetric Garch model

Consider the observation equation as specified in (1). We now modify the variance equation (4) to accommodate asymmetric shocks in the volatility modelling. The asymmetric Garch model for the conditional variance term is given by

$$h_t = \omega + \lambda \, u_{t-1}^2 f_t(u_{t-1}, \gamma) + \sum_{j=1}^{r} \beta_j h_{t-j} + \sum_{j=1}^{s} \alpha_j u_{t-j}^2 \tag{7}$$

where $\omega, \lambda > 0$ are unknown parameters. The functions defining asymmetric volatilities may be specified to accommodate small and large (or positive and negative) effects as follows.

1. If $f_t(u_{t-1}, \gamma) = (1 - e^{-\gamma u^2_{t-1}})$ then $f'_t(u_{t-1}, \gamma) = u^2_{t-1}\, e^{-\gamma u^2_{t-1}}$. This function allows for mall and large effects.
2. If $f_t(u_{t-1}, \gamma) = (1 + e^{-\gamma u_{t-1}})$ then $f'_t(u_{t-1}, \gamma) = u_{t-1}\, e^{-\gamma u_{t-1}}\, (1+ e^{-\gamma u_{t-1}})^{-2}$. This functions allows for positive and negative effects.

Consider the function $f_t(u, \gamma) = 1 - \exp\{-\gamma u^2\}$. If $|u|$ is large, then f_t will tend to 1 and h_t will be affected by λ. On the contrary, if $|u|$ is small then f_t will be close to 0 and h_t will be less affected by λ. This reflects different effects of small and large past error terms in the variance h_t. Consider the logistic function $f_t(u, \gamma) = (1 + \exp\{-\gamma u\})^{-1}$, if $u \to 0$ then $f_t(u) \to 1/2$. In this scenarium the total impact of f_t (u) is $\lambda/2$. If $u \to \infty$ then $f_t(u) \to 1$ and the effect in h_t is $\lambda u^2 \to \infty$. On the other hand, if $u \to -\infty$ $f_t(u) \to 0$ and the impact in $h_t \in [0, \lambda u^2]$. As a result, when $f_t(u_{t-1}, y) = (1 + e^{-\gamma u_{t-1}})^{-1}$ the asymmetric effects generated from positive and negative shocks depend also on the shock size. If the shocks are large their effect in the variance will be significantly asymmetric. On the contrary, if the sizes are small then the model will have small asymmetry. For more details on this issue see Gonzalez-Rivera (1998) and Hagerud (1997).

Furthermore, from an inferential point of view in smooth transition models if we set $\gamma \to \infty$ then $lim_{\gamma} \to \infty\, f_t(u, \gamma) = 1$, which implies that the likelihood function for the asymmetric Garch model tends to a constant given by the likelihood function of the symmetric Garch model. This condition may lead to ill behaved likelihood functions. That is, there is a positive probability that the likelihood function is a increasing function of γ. Theorem 1 of Lubrano (1998) states this problem of the likelihood and suggest an informative prior to correct this problem. We will consider this correction in this work and focus on the degrees of freedom estimation.

3 LIKELIHOOD ISSUES AND BAYESIAN INFERENCE

If we consider latent variables $\omega_t \sim GI(v/2, v/2)$ and the mixture representation (2) then the likelihood function for model (1) is given by

$$L(\phi, \theta, \psi, v \mid \mathrm{X}, \mathrm{A}, \overline{\mathrm{H}}, \mathrm{y}, \mathrm{y}_0) \propto |H|^{-1/2} \exp\left\{-\frac{1}{2} z' H^{-1} z\right\}. \tag{8}$$

with $z = \mathrm{y} - X\varphi - A\theta - \tilde{H}\psi$, $H = diag(\omega_{p+1}\frac{v-2}{v} h_{p+1}, \ldots, \omega_N \frac{v-2}{v} h_N)$. The estimation of the degree of freedom parameter v is not straighforward. As discussed in Fonseca et al. (2008)

the likelihood function is ill behaved and the use of naive noninformative priors such as the uniform may lead to improper posterior distributions for the parameters of interest. As showed in the paper, there is a positive probability that the maximum likelihood estimator does not exist for some data sets. This is not an issue related to the frequentist approach but an intrinsic problem of the likelihood. The following test may be applied to a given data set in order to test whether the likelihood of ν is well behaved or does not have a maximum. If the following condition is satisfied, the likelihood is not well behaved.

$$\sum_{i=1}^{n}(\hat{z}^2 - 1)^2 < 2n, \tag{9}$$

where n is the sample size and $\hat{z}_i$ are standardized residuals under normality. In this paper, we consider the correction for the likelihood proposed by Fonseca et al. (2008), that is, we consider Jeffreys prior for the degree of freedom ν in the Arma-Garch set up. For the parameters Φ, θ, ψ, α, β we consider flat prior distribution given by

$$p(\Phi, \theta, \psi, \alpha, \beta \mid \nu) \propto k. \tag{10}$$

While for ν we consider the independent Jeffreys prior given by

$$p(\nu) \propto \left(\frac{\nu}{\nu+3}\right)^{1/2}\left\{\vartheta'(\nu/2) - \vartheta'((\nu+1)/2) - \frac{2(\nu+3)}{\nu(\nu+1)}\right\}^{1/2}, \tag{11}$$

where $\vartheta(a) = d\,log\Gamma(a)/da$ and $\vartheta'(a) = d\{\vartheta(a)\}/da$ are the digamma and trigamma functions, respectively.

Fonseca et al. (2008) present a simulated study showing that the posterior median is a better estimator for ν than the maximum likelihood estimator. They also prove that the marginal posterior distribution of ν is proper. In the frequentist context this may be seen as a problem of bias reduction of maximum likelihood estimates as presented in Firth (1993). Thus the likelihood are penalized by a function, the Jeffreys invariant prior, which is responsible for correcting the estimation problems.

Other authors also reported results related to the Student-t likelihood being ill behaved such as Bauwens and Lubrano (1998). Their theorem 1 is about the flatness of the likelihood function for the Student-t model in the Garch model. However, the suggested priors for the degrees of freedom would not work in cases where the likelihood has

not a maximum as the posterior will have the same behaviour as the prior distribution. Thus sufficient prior information would be needed to well estimate this parameter. Thus, this solution will not work in situations where there is no prior information. The authors propose a Griddy-Gibbs sampler to solve the inferential problem. The method will find an estimate for the degrees of freedom, however, this will tend to the limit of the grid whenever the likelihood is too much ill behaved. Villa and Walker (2014) recently proposed a reference prior distribution for the degrees of freedom parameter. However, their proposal considered a discrete set of possible values for the degrees of freedom and truncated the upper limit of this set so that the Gaussian case is no considered.

The issue of likelihood functions tending to a constant will always come up whenever there are limiting cases or distributions in the model. Indeed, this is the case of the asymmetric model used here for the volatility. Regarding the asymmetric model (7), consider the case where $\gamma \to \infty$. Then $\lim_{\gamma \to \infty} f_t(u, \gamma) = 1$, which implies that the likelihood function for the asymmetric Garch model tends to a constant given by the likelihood function of the symmetric Garch model. This condition may lead to ill behaved likelihood functions. That is, there is a positive probability that the likelihood function is a increasing function of γ. Theorem 1 of Lubrano (1998) states this problem of the likelihood. In other words, the likelihood for the asymmetric model is also ill behaved, and may be increasing in γ. Thus, the prior distribution of γ needs to have tails that go fast enought to 0, allowing the posterior distribution to be an integrable function of γ . An alternative is to consider the prior proposed by Lubrano (1998) which is given by

$$\pi(\gamma) = (1 + (\gamma - \gamma_0)^2)^{-1}, \tag{12}$$

$\gamma, \gamma_0 > 0$. Notice that for this proposal it is required to specify a hyperparameter γ_0. The focus of this paper is the estimation of the degrees of freedom in the Student-t model for the error term. Thus we consider a proper prior for the asymmetry parameter as suggested by Lubrano (1998) but an actual derivation of Jeffreys noninformative priors for this case will be considered in a future work.

Having defined all the required prior distributions, posterior analysis may be obtained by Gibbs sampling algorithm generating from the complete conditional distributions with Metropolis steps for α, β and ν. The complete conditional distributions are presented in the Appendix.

4 BAYESIAN MODEL SELECTION

In this work model uncertainty in the sampling distribution of the data is considered through Bayes Factor computations and Bayesian hypothesis testing. As follows the main tools used are described. Consider M competing models, so that we have M likelihood and prior distributions, denoted respectively by $p_m(x|\theta_m)$ and $p_m(\theta_m)$, $\theta_m \in \Theta_m$, m = 1,..., M. Let us introduce a discrete prior distribution over the set of M alternative models, denoted by $p(m) = \pi_m$, with $\sum_1^M \pi_m = 1$. The joint distribution of (x, m, θ_m) is given by $p(x, m, \theta_m) = p_m(x|\theta_m)p_m(\theta_m)p(m)$. Using Bayes theorem we immediately obtain the posterior distribution $p(m, \theta_m|x)$ and the marginal distribution $p(m|x)$ which encapsulates the uncertainty about the unknowns (m, θ_m) after observing x. The posterior inference in the presence of uncertainties about the correct model involves the evaluation of the posterior $p(m|x)$, m = 1,... M which depends on the marginal distribution $p_m(x)$ and the evaluation of $pm(\theta_m|x)$ the posterior distribution of θ_m. Let the Bayes Factor (Kass and Raftery, 1995) of model 1 with respect to model 2 be

$$B_{12} = \frac{p(x|m_1)}{p(x|m_2)}. \tag{13}$$

In the model choice problem one may consider the following benchmarks to decide between models. The guideline provided in Kass and Raftery (1995) for interpretation of the Bayes factor is presented in Table 1.

Let us now consider the special case of just two alternative models. This is known in the literature as the hypothesis test problem. A decision problem is completely specified by the triple {A, Θ, X}, where A is the decision space, Θ the parameter space and X is the sample space. Let $LA \times \Theta \longrightarrow \{0, k_i\}$ be a loss function.

Model m_1 is defined by $\theta \in \Theta_1$ and the alternative model m_2 by $\theta \in \Theta_2$, which are denoted by H_i, i = 1, 2. The parameter space is partitioned in two disjoint components Θ_1 and Θ_2. The action space is defined by two components, $A = \{a_1, a_2\}$, meaning that a_i - the hypothesis H_i is the true one and so must be accepted. Often $L(\theta, a_i) = 0$ if $\theta \in \Theta_i$ and K_i if $\theta \in \Theta_j$, $j \neq i$. Actually a hypothesis test is a decision rule defined on the sample space and assuming values $\{a_1, a_2\}$, that is: $\delta : X \longrightarrow \{0, 1\}$. As is well known from the decision making literature $a_1 \succ a_2$ if and only if $E[L(\theta, a_1) < E[L(\theta, a_2)$, where the expectation is with respect to the posterior distributions. This is equivalent to accept H_1 if and only if $\text{if} = \frac{P(H_1|x)}{P(H_2|x)} > \frac{k_2}{k_1}$, which is equivalent to

$$B_{12} > \frac{k_2 P(H_2)}{k_1 P(H_1)} \tag{14}$$

This will be used in the paper to chose between models of interest (e.g. Gaussian versus Student-t, asymmetric versus symmetric). Some drawbacks with the Bayesian hypothesis testing include the treatment of precise hypothesis or point null and one side hypothesis. Also the choice of the prior distribution are influential in the final results.

The computation of the predictive distribution $p(x|m_i)$ is not straighfoward. It is needed to consider the samples obtained from the MCMC algorithm in order to numericaly compute the precidtive distribution for each model of interest. For a given model m_i Newton and Raftery (1994) proposed the following estimator based on samples from the posterior distribution.

$$\hat{p}_1(x) = \frac{\frac{dm}{1-d} + \sum_{i=1}^{m} p(x|\theta^{(i)})\{d\hat{p}_1(x) + (1-d)p(z|\theta^{(i)})\}^{-1}}{\frac{dm}{(1-d)\,\hat{p}_1(x)} + \sum_{i=1}^{m} \{d\hat{p}_1(x) + (1-d)p(x|\theta^{(i)})\}^{-1}}, \tag{15}$$

where $\theta^{(1)}$,..., $\theta^{(m)}$ are generated from the posterior distribution $p(\theta\,|x)$. This estimator performs well for d as small as 0.01. An alternative proposal is the shifted gamma estimator proposed by Raftery et al. (2007). In this proposal, the outputs of the MCMC algorithm is used to calculate a sequence of loglikelihood values $\{l_k : k = 1,..., n\}$ and the posterior distribution of the loglikelihoods is given by

$$l_{max} - l_k \sim \text{Gamma}(\alpha, \lambda^{-1}), \tag{16}$$

where l_{max} is the maximum achievable likelihood, $\alpha = d/2$, d is the number of parameters in the model and $\lambda < 1$. In practice, λ is not much less than 1. Combining the harmonic mean identity $\frac{1}{p(x)} = E\left\{\frac{1}{p(x|\theta)}\right\}$ with (16) results in

$$log(p(x)) = l_{max} + \alpha log(1 - \lambda) \tag{17}$$

In general, l_{max} is not known, thus $\hat{l}_{max} = \max\{\bar{l} + s_l^2, l_k\}$ is used, where $\bar{l} + s_l^2$ is the moment estimator of l_{max}, $\bar{l}$ and s_l^2 are the sample mean and variance of the $l'_k s$.

5 SIMULATION STUDY

In this section we perform a simulation study to evaluate the effects of model mispecification in the results of Bayesian hypothesis testing as described in section 4. We define a sucess when the optimal decision selected in the Bayesian test coincides with the true model. The optimal decision is to choose the model which has Bayes Factor greater than 3 as described in table 1. We compute the rate of sucess by Monte Carlo simulation considering 100 replicates of each data scenarium. In this simulation study we vary the sample size (T = 150, 500) and the sampling distribution (Gaussian and student-t). We also evaluate the one-step-ahead variance in order to compare the Gaussian and Student-t models. The model considered for simulation is defined according to equations (1) and (7), that is,

$$y_t = \phi y_{t-1} + \theta u_{t-1} + \delta\sqrt{h_t} + u_t$$

$$h_t = \omega + \lambda u_{t-1}^2 f_t(u_{t-1}, \gamma) + \beta h_{t-1} + \alpha u_{t-1}^2$$

The parameters are set to be $\omega = 0.25$, $\alpha = 0.5$, $\beta = 0.1$, $\gamma = 5$, $\varphi = 0.8$, $\theta = 0.1$, $\delta = 0$, $\lambda = 1$. The sampling distribution considered for u_t is either Gaussian with $u_t = \varepsilon_t h_t$ or Student-t with

$$u_t = \varepsilon_t \left(\frac{v-2}{v} h_t \omega_t\right)^{1/2}, \; t = 2,..., N, \tag{18}$$

where $\varepsilon_t \sim N(0, 1)$, $\omega_t \sim IG(v/2, v/2)$. For the degrees of freedom we considered datasets with v = 3 ou v = 6. Tables 2, 3 and 4 shows the Mean Square Error (MSE) for datasets simulated (n = 150, 500) using the Gaussian and the Student-t models with asymmetric volatility. In the case of Gaussian data both models (Gaussian and Student-t) have similar behaviours in the estimation of all parameters ($\alpha, \beta, \lambda, \gamma$) as presented in table 2. Indeed, this is a good property of Student-t models as it is able to acommodate gaussianity as a limiting case. On the other hand, for Student-t data with 3 degrees of freedom the Gaussian model has a poor performance when compared with the Student-t model as presented in table 3. Notice that the Gaussian model give very high MSE for the asymmetry parameter λ. In fact, the MSE for the Gaussian model is more than fourteen times larger than for the Student-t model. As expected, the same does not happen for v = 6 degrees of freedom, which is closer to the Gaussian case as shown in table 4. In this scenario both models give similar results.

Table 5 presents the proportion of right decisions regarding the model which was used to simulate the data. The hypothesis testing procedure led to the correct decisions (Gaussian against Student-t model) as expected for most of scenarios. Gaussian model is selected for Gaussian data and Student-t model is selected for Student-t data. Except for small samples (n = 150) and ν = 6, which is a scenario with quite light tails, in this case both models are similar. In summary, we recomend the use of the Bayes factor for model choice for the Arma-Garch with Student-t or Gaussian errors.

In terms of predictive variance the Gaussian model presented smaller MSE in general, with the Student-t model having a smaller MSE in scenarios with high values of y_{T+1} as presented in Table 6. These scenarios are the ones with very large MSE values.

6 APPLICATION

In this section we present an application of the proposed model to study the dynamics in the daily Dow Jones returns. Our main goal in this application is to highlight the predictive advantages of the Student-t model when compared to the Gaussian model in periods of high volatility of the series. In this context, we analysed the daily Dow Jones index of the New York stock market and is given by $r_t = 100ln(DJ_t/DJ_t-1)$. We present results from 02/01/2007 to 31/12/2008, with the first period going up to 30/05/2008 used for estimation which results in 354 observations and the final part used for predictive performance evaluation. We sucessively update the data with one observation after prediction of this point resulting in a total of 503 observations. The period selected for prediction has two different volatility regimes. The first period from 02/06/2008 to 12/09/2008 is before the bankruptcy of Lehman Brothers Bank. This event and the Estate Market crisis resulted in large volatility in the american stock market from 15/09/2008. This instability decreased from December 2008. Thus, the data selected for prediction will allow comparison of predictive performance of the models for different kinds of volatility.

In this application we fitted three kinds of asymmetric models: the logistic and exponential as presented in subsection 2.3 and a modified logistic model which we call logistic model 2, which is defined as follows. As pointed out by Gonzalez-Rivera (1998) in the finantial market good news are associated with positive shocks which tend to result in small volatility while bad news tend to result in negative shocks which will

usually be associated with large volatilities. In this regard, we consider also the logistc function 2 defined by $f_t(u) = (1 + \exp\{\gamma u\})^{-1}$. Thus, we have six model formulations: Gaussian Garch with logistic asymmetry types 1 and 2, Gaussian Garch with exponencial asymmetry, Student-t Garch with logistic asymmetry types 1 and 2 and Student-t Garch with exponencial asymmetry. In addition, we compute prediction based on a hybrid approach which considers the results from hypothesis tests as presented in section 4. We fitted a Arma(3,0) for the mean which resulted in a stationary time series and a Garch(1,1) for the variance modeling. In the MCMC algorithm we considered ten thousand iteractions and the acceptance rate was tunned to be in the range 0.2 to 0.4. We follow Awartani and Corradi (2005) and considered the squared observed returns as proxy for the variance of the process in the computation of Mean Squared Errors (MSE) for the four competing models.

Graph 1 present the variance proxy evolution through time and the predictions. Table 7 to table 9 present a summary of main MSE results for this application. In general, the asymmetric logistic model 2 presented a better predictive performance. This result was expected for daily returns as negative shocks in the mean tend to have greater impact in the volatilities than positive shocks with the same size. This behaviour is well modeled by a logistic function 2. In the context of model selection, the majority of the testes indicate the Student-t model as the best fit. Although the Gaussian model had the best predictive performance in general, the Student-t model was more competitive in the larger volatility periods. The best Gaussian model with logistic asymmetry 1 against the best Student-t model (which has exponential asymmetry) for the period from 02/06/2008 to 12/09/2008 was 0.915, that is, the best Gaussian model had a 8,5% smaller MSE. While in the period from 15/09/2008 to 31/12/2008 the MSE ratio for the best Gaussian model with logistic asymmetry 2 against the best Student-t model (which has logistic asymmetry 2) was 0.995, that is, the best Gaussian model had a 0,54% smaller MSE.

In order to analyse closely the predictive performance for the Gaussian and Student-t models in the two different regimes we define the MSE ratio for the Gaussian against Student-t model for the last five observed days given by

$$I_{t+5} = \frac{\sum_{n=1}^{5} (\hat{h}_{t+n}^{G} - h_{t+n})^2}{\sum_{n=1}^{5} (\hat{h}_{t+n}^{ST} - h_{t+n})^2} \tag{19}$$

with h_{t+n} the conditional variance (squared observed returns), $\hat{h}^{G}_{t+n}$ and $\hat{h}^{ST}_{t+n}$ the one step ahead prediction for the conditional variance in the Gaussian and Student-T models, respectively.

Graph 2 presents the MSE ratio (Gaussian versus Student-t) for I_{t+5} across time for the logistic, logistic 2 and exponential asymmetric models, respectively. The evolution in panel (a) and (b) indicate that for the logistic model and logistic model 2, periods of time with high volatility tend to have I_{t+5} greater than 1, which suggests that the Student-t model would be more indicated for these periods. For the exponential asymmetric model (panel (c)) this difference is less evident, although there are periods in which the Student-t model would be more recommended. Indeed, the correlation between I_{t+5} and the mean squared returns in the five past days is 0,54 for the logistic model 1, 0,42 for the logistic model 2 and it is 0,29 for the exponential model.

7 CONCLUSIONS

This work shows how Student-t smooth transition models can be used adequately from a Bayesian point of view. Our main interest in this paper was to investigate the prediction and estimation performances of Student-t sampling distributions based on the independent Jeffreys prior assumption for the degrees of freedom parameter in the context of smooth transition models. The degrees of freedom in the Student-t model are difficult to estimate and it was considered independent Jeffreys prior to solve this estimation problem. The likelihood tends to a constant with positive probability. This behaviour is intrisic to the likelihood and not corrected by usual parametric priors such as exponential and gamma. This is due to the existance of limiting cases (Gaussian and symmetric models) for the sampling distributions assumed for the data. Thus, we suggested the use of independent Jeffreys priors, which are proved to lead to proper posteriors and have nice frequentist properties (Fonseca et al., 2008).

The proposed prior gave positive results as presented in the simulated study. For model selection we use Bayesian Hypothesis testing based on the numerical computation of predictive distributions. The Bayesian test based on Bayes factors was effective in the decision between Gaussian and Student-t models. Furthermore, the simulated study indicate that some factors such as parameter estimation and prediction one step ahead may increase competitivity of Student-t models. For instance, this is the case when data is simulated from Student-t models with relatively heavy tails ($\nu = 3$).

For the Dow Jones application, in general the Gaussian model had the best predictive performance with the Student-t model being preferable in large volatility periods, specially for the logistic asymmetric model 2. This is crucial to correctly estimating the uncertainty in periods of large volatility in the market.

In this work the informative prior proposed by Lubrano (1998) was considered for the smooth transition parameter. However, this prior depends on hyperparameters, thus in a future research it will be considered the development of new reference priors for this problem which would not depend on hyperparameter specification.

REFERENCES

ARDIA, D. **Financial risk management with bayesian estimation of Garch models:** theory and applications. Springer-Verlag, 2008. (Lecture Notes in Economics and Mathematical Systems, v. 612).

AWARTANI, B. M.; CORRADI, V. Predicting the volatility of the S&P-500 stock index via Garch models: the role of asymmetries. **International Journal of Forecasting**, v. 21, p. 167-184, 2005.

BAUWENS, L.; LUBRANO, M. Bayesian inference on Garch models using the gibbs sampler. **Econometrics Journal**, v. 1, p. 23-46, 1998.

_____. Bayesian option pricing using asymmetric garch models. **Journal of Empirical Finance**, v. 9, n. 3, p. 321-342, 2002.

BOLLERSLEV, T. Generalized autoregressive conditional heteroscedasticity. **Journal of econometrics**, v. 37, p. 307-327, 1986.

_____. A conditionally heteroskedastic time series model for speculative prices and rates of return. **The Review of Economics and Statistics**, v. 69, n. 3, p. 542-547, 1987.

BOLLERSLEV, T.; CHOU, R. Y.; K. KRONER, F. Arch modeling in finance. **Journal of econometrics**, v. 52, p. 5-59, 1992.

ENGLE, R. F.; NG, V. K. Measuring and testing the impact of news on volatility. **Journal of Finance**, v. 48, n.1, p. 1749-1778, 1993.

FIRTH, D. Bias reduction of maximum likelihood estimates. **Biometrika**, v. 80, p. 27-38, 1993.

FONSECA, T. C. O.; FERREIRA, M. A. R.; MIGON, H. S. Objective bayesian analysis for the student-t regression model. **Biometrika**, v. 95, p. 325-333, 2008.

GONZALEZ-RIVERA, G. Smooth transition Garch models. **Studies in Nonlinear Dynamics and Econometrics**, v. 3, p. 61-78. 1998.

HAGERUD, G. E. **A new non-linear Garch model**. 1997. Thesis (Ph.D) – Stockholm School of Economics, 1997.

HARVEY, A.; CHAKRAVARTY, T. **Beta-t-(e)-garch**. Technical report, University of Cambridge, 2008.

JACQUIERA, E.; POLSON, N. G.; ROSSI, P. E. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. **Journal of Econometrics**, v. 122, p. 185-212. 2004.

KASS, R.; RAFTERY, A. E. Bayes factor. **Journal of the American Statistical Association**, 90, p. 773-795, 1995.

LUBRANO, M. **Smooth transition Garch models**: a bayesian perspective. Université catholique de Louvain, Center for Operations Research and Econometrics. 1998. (Core discussion papers, n. 1998066).

MANDELBROT, B. The variation of certain speculative prices. **Journal of Business**, v. 36, p. 394-419, 1963.

NEWTON, M. A.; RAFTERY, A. E. Approximate bayesian inference with the weighted likelihood bootstrap. **Journal of the Royal Statistical Society Series B**, v. 56, p. 3-48, 1994.

RAFTERY, A. E.; et al. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *In*: **Bayesian Statistics**. Oxford University Press, v. 8 p. 371-416, 2007.

VILLA, C. WALKER, S. G. Objective prior for the number of degrees of freedom of a t distribution. **Bayesian Analysis**, v. 9, p. 197-220, 2014.

VRONTOS, I. D.; DELLAPORTAS, P.; POLITIS, D. N. Full bayesian inference for Garch and egarch models. **Journal of business and economic statistics**, v. 18, p. 187-198, 2000.

ZHANG, X. et al. **Modeling dynamic volatilities and correlations under skewness and fat tails**. Technical report, Tinbergen Institute, 2011.

APPENDIX

This appendix presents the complete conditional distributions used in Markov Chain Monte Carlo sampling of model parameters for the Student-t Garch model with Independent Jeffreys prior for the degrees of freedom. The code where this algorithm is implemented is available for general R package users.

- Parameter Φ:

$\Phi|\theta, \psi, \alpha, \beta, \mathbf{X}, \mathbf{y}, \mathbf{y_0} \sim N(\hat{\Phi}, (X'H^{-1}X)^{-1})$,

where $\hat{\Phi} = (X'H^{-1}X)^{-1}(X'H^{-1}\mathbf{y} - X'H^{-1}A\theta - X'H^{-1}\widetilde{H}\psi)$.

- Parameter θ:

$\theta|\Phi, \psi, \alpha, \beta, X, y, y_0 \sim N(\mu_\theta, V_\theta)$,

where $V_\theta = (A'H^{-1}A)^{-1}$ and $\mu_\theta = (A'H^{-1}A)^{-1}(A'H^{-1}\mathbf{y} - A'H^{-1}X\Phi - A'H^{-1}\widetilde{H}\psi)$.

- Parameter ψ:

$\psi|\Phi, \theta, \alpha, \beta, \mathbf{X}, \mathbf{y}, \mathbf{y_0} \sim N(\mu_\psi, V_\psi)$,

where $V_\psi = (\widetilde{H}'H^{-1}\widetilde{H})^{-1}$ and $\mu_\psi = (\widetilde{H}'H^{-1}\widetilde{H})^{-1}(\widetilde{H}'H^{-1}y - \widetilde{H}'H^{-1}X\Phi - \widetilde{H}'H^{-1}A\theta)$.

- Parameter α and β:

$$p(\alpha, \beta|...) = \prod_{j=1}^{N-p} \phi(y_t|\mu_t, h_t),$$

where $\mu_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{q} \theta_j u_{t-j} + \sum_{j=1}^{k} \psi_j h_{t-j}$ and $\phi(.)$ is the Gaussian density function. We define independent proposal distributions given by

$$\alpha_i \sim N(\hat{\alpha}_i, \Sigma\,\alpha_i), \beta_i \sim N(\hat{\beta}_i, \hat{\Sigma}_{\beta i}), \beta_i \sim N(\hat{\beta}_i, \hat{\Sigma}_{\beta i})$$

- Parameter ν:

$$p(\nu|\Phi, \theta, \psi, \alpha, \beta, \mathbf{X}, \mathbf{y}, \mathbf{y_0}) \propto |H|^{-1/2}\exp\left\{-\frac{1}{2}z'H^{-1}z\right\}\frac{(\nu/2)^{(N-p)\nu/2}}{\Gamma(\nu/2)^{N-p}} \times \left(\prod_{t=1}^{N-p}\omega_t\right)^{-\nu/2-1}\exp\left\{-\frac{1}{2}\sum_{t=1}^{N-p}\frac{\nu}{\omega_t}\right\}p(\nu),$$

where $z = (y - X\phi - A\theta - \widetilde{H}\psi)$.

TABLE 1
Guideline for interpretation of the Bayes factor B_{12} which is the evidence in favour of model M_1 versus model M_2

$2ln(B_{12})$	B_{12}	Evidence against M_2
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very strong

Elaborated by the authors.

TABLE 2
Mean Square Error (MSE) for α, β , λ and γ obtained in the simulated study, n = 150 e n = 500

Model	Gaussian data (ν = 3)			
	n = 150		n = 500	
	Gaussian	Student-t	Gaussian	Student-t
α	0.044	0.046	0.023	0.024
β	0.003	0.026	0.001	0.003
λ	0.197	0.386	0.113	0.142
γ	13.006	13.646	9.718	10.225

Elaborated by the authors.

TABLE 3
Mean Square Error (MSE) for α, β , λ and γ obtained in the simulated study, n = 150 e n = 500

Model	Student-t data (ν = 3)			
	n = 150		n = 500	
	Gaussian	Student-t	Gaussian	Student-t
α	0.081	0.067	0.064	0.041
β	0.007	0.027	0.008	0.003
λ	1.414	0.739	5.903	0.403
γ	12.319	13.291	11.193	12.523

Elaborated by the authors.

TABLE 4
Mean Square Error (MSE) for α, β , λ and γ obtained in the simulated study, n = 150 e n = 500

Model	Student-t data (ν = 6)			
	n = 150		n = 500	
	Gaussian	Student-t	Gaussian	Student-t
α	0.048	0.051	0.038	0.037
β	0.004	0.026	0.003	0.005
λ	0.531	0.531	0.181	0.156
γ	13.016	13.643	11.107	11.528

Elaborated by the authors.

TABLE 5

Results for the hypothesis testing procedure based on Bayes Factors

Proportion of right decisions obtained in the simulated study, n = 150 and n = 500.

		Model	
		Gaussian	Student-t
Data (n = 150)	Gaussian	0.86	-
	Student-t ($\nu = 3$)	-	0.89
	Student-t ($\nu = 6$)	-	0.45
Data (n = 500)	Gaussian	1.00	-
	Student-t ($\nu = 3$)	-	0.99
	Student-t ($\nu = 6$)	-	0.93

Elaborated by the authors.

TABLE 6

Mean Square Error (MSE) for the predictive variance obtained in the simulated study in the one-step-ahead prediction, n = 150 e n = 500

		Model			
		n = 150		n = 500	
		Gaussian	Student-t	Gaussian	Student-t
Data	Gaussian	7.4046	16.9097	830.23	791.66
	Student-t ($\nu = 3$)	15.26	20.70	4.93	3.63
	Student-t ($\nu = 6$)	11650.43	11186.32	9.04	9.34

Elaborated by the authors.

TABLE 7

Square root of MSE for the Garch model with logistic asymmetry 1 and model selection test

Time period	Gaussian model	Student-t model	Model selection
02/06/2008 - 12/09/2008	7.14	13.16	12.32
15/09/2008 - 31/12/2008	499.90	523.27	523.27
02/06/2008 - 31/12/2008	258.48	273.35	272.94

Elaborated by the authors.

TABLE 8

Square root of MSE for the Garch model with logistic asymmetry 2 and model selection test

Time period	Gaussian model	Student-t model	Model selection
02/06/2008 - 12/09/2008	7.28	8.62	9.06
15/09/2008 - 31/12/2008	444.86	447.30	449.78
02/06/2008 - 31/12/2008	230.47	232.38	233.86

Elaborated by the authors.

TABLE 9

Square root of MSE for the Garch model with exponential and model selection test

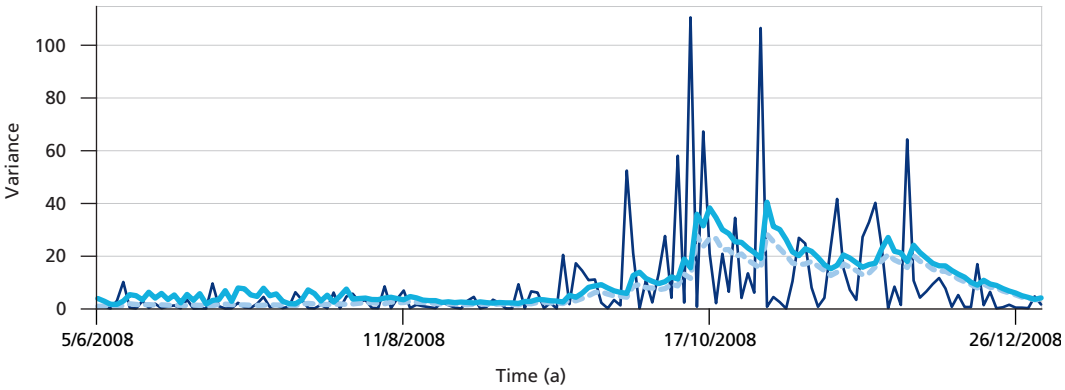
Time period	Gaussian model	Student-t model	Model selection
02/06/2008 - 12/09/2008	7.17	7.80	8.41
15/09/2008 - 31/12/2008	727.16	817.82	820.75
02/06/2008 - 31/12/2008	374.41	420.96	422.76

Elaborated by the authors.

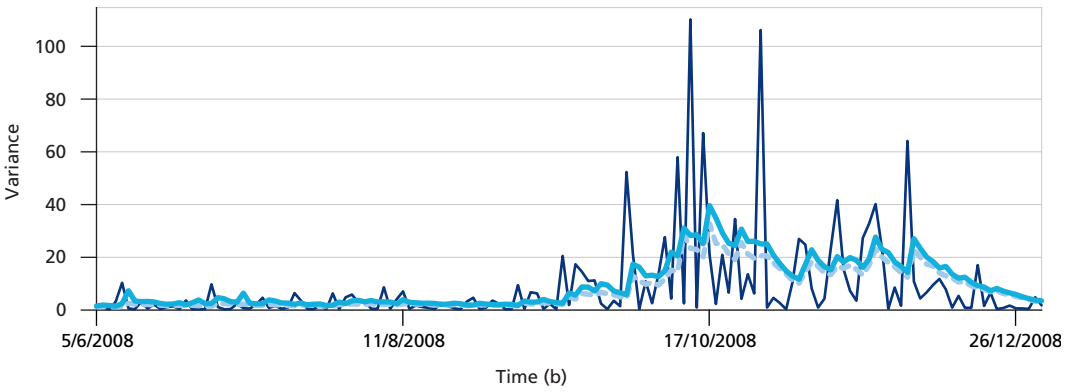
GRAPH 1

Squared returns (blue line), Gaussian (red line) and Student-t (black line) predictions with asymmetric volatility model

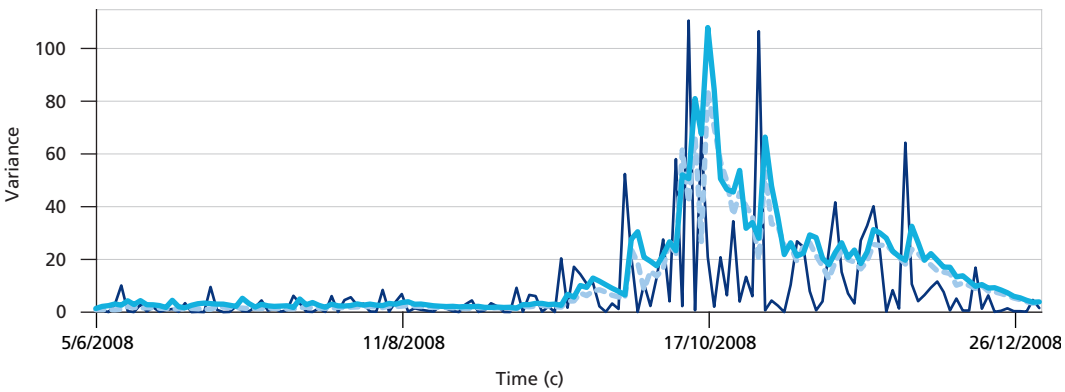
1A – Logistic asymmetric model



1B – Logistic 2 asymmetric model



1C – Exponential asymmetric model

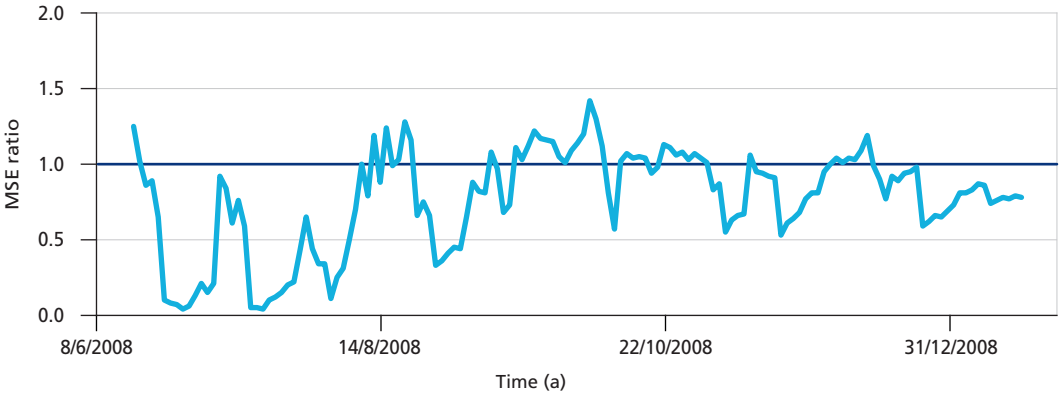


Elaborated by the authors.

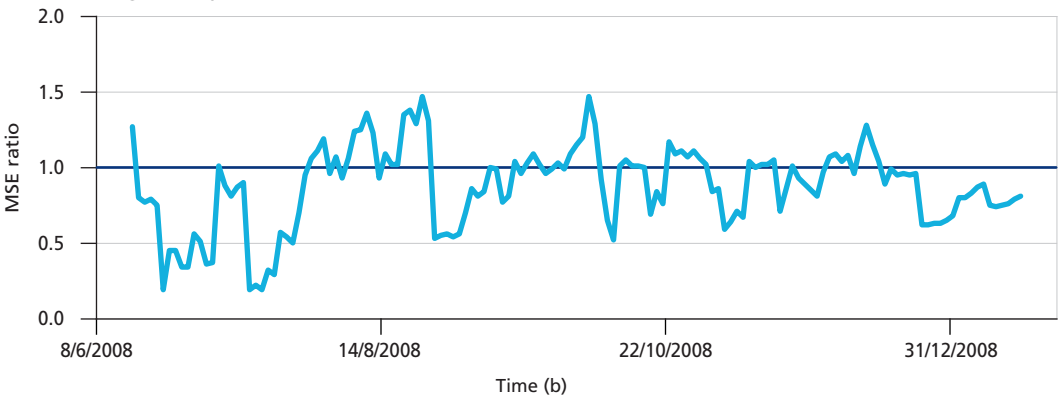
GRAPH 2

Mean squared error ratio between the Gaussian and Student-t models with asymmetric volatility for the past five days

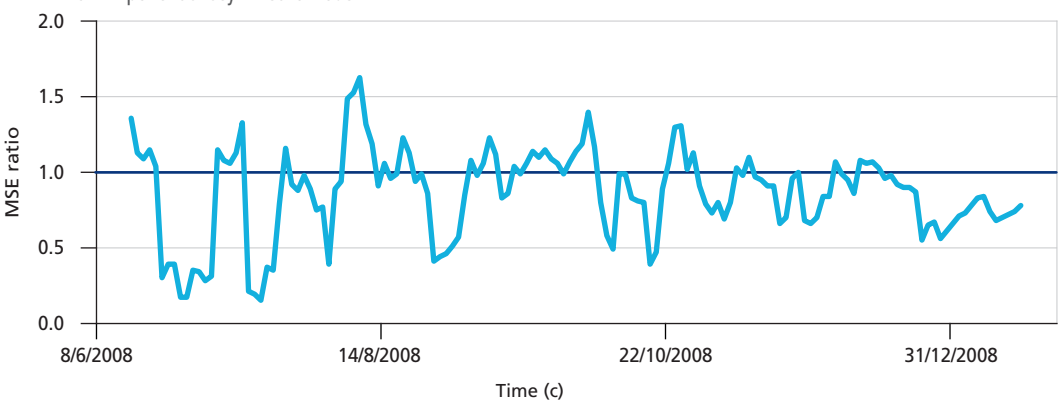
2A – Logistic asymmetric model



2B – Logistic 2 asymmetric model



2C – Exponential asymmetric model



Elaborated by the authors.

Ipea – Institute for Applied Economic Research

PUBLISHING DEPARTMENT

Coordination
Cláudio Passos de Oliveira

Supervision
Everson da Silva Moura
Reginaldo da Silva Domingos

Proofreading
Clícia Silveira Rodrigues
Idalina Barbara de Castro
Leonardo Moreira Vallejo
Marcelo Araujo de Sales Aguiar
Marco Aurélio Dias Pires
Olavo Mesquita de Carvalho
Regina Marta de Aguiar
Alessandra Farias da Silva (intern)
Lilian de Lima Gonçalves (intern)
Luiz Gustavo Campos de Araújo Souza (intern)
Paulo Ubiratan Araujo Sobrinho (intern)
Pedro Henrique Ximendes Aragão (intern)
Thayles Moura dos Santos (intern)

Typesetting
Bernar José Vieira
Cristiano Ferreira de Araújo
Daniella Silva Nogueira
Danilo Leite de Macedo Tavares
Jeovah Herculano Szervinsk Junior
Leonardo Hideki Higa

Cover design
Luís Cláudio Cardoso da Silva

Graphic design
Renato Rodrigues Bueno

The manuscripts in languages other than Portuguese published herein have not been proofread.

Ipea Bookstore

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo
70076-900 – Brasília – DF – Brazil
Tel.: + 55 (61) 2026 5336

Email: livraria@ipea.gov.br

Composed in Adobe Garamond 11/13.2 (text)
Frutiger 47 (headings, graphs and tables)
Brasília – DF – Brazil

Ipea's mission
Enhance public policies that are essential to Brazilian development by producing and disseminating knowledge and by advising the state in its strategic decisions.

ISSN 1415-4765

ipea Institute for Applied Economic Research

MINISTRY OF PLANNING, DEVELOPMENT AND MANAGEMENT

BRASIL BRAZILIAN GOVERNMENT