Evaluating the performance of degrees of freedom estimation in Asymmetric GARCH models with Student-t innovations

Thaís Cristina Oliveira da Fonseca*
Vinícius dos Santos Cerqueira†
Helio dos Santos Migon*
Cristian Andres Cruz Torres‡

Abstract
This work investigates the effects of using the independent Jeffreys prior for the degrees of freedom parameter of a Student-t model in the asymmetric generalised autoregressive conditional heteroskedasticity (GARCH) model. To capture asymmetry in the reaction to past shocks, smooth transition models are assumed for the variance. We adopt the fully Bayesian approach for inference, prediction and model selection. 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 the estimation of model parameters in the Student-t GARCH model is affected by small sample sizes and misspecification regarding the sampling distribution. An application to the Dow Jones stock market data illustrates the usefulness of the asymmetric GARCH model with Student-t errors.

Keywords: heavy tailed distributions, Bayesian inference, ill behaved likelihoods
JEL Codes: C11, C13, C18

1 Introduction
A popular choice for financial data analysis is the generalised autoregressive conditional heteroskedasticity (GARCH) approach (Bollerslev, 1986) which
models the variance as a function of past values and an error term assuming that the variance is independent of shocks in the mean. In this setup, the sampling distribution for the error term are usually modelled as Gaussian distributed. This is mainly due to mathematical convenience rather than being suitable for financial data (Bollerslev, Chou, & Kroner, 1992). Moreover, we are interested in models that can capture the stylised fact that the conditional variance can react asymmetrically to positive versus negative shocks or large versus small shocks. This may be accommodated by smooth transition models based on an asymmetric specification of the conditional variance. Thus, this paper investigates the sensitivity of both point estimation and model comparison resulting from assuming a Student-t distribution with unknown degrees of freedom for the shocks and how asymmetry estimation in smooth transition models is affected by relaxing the Gaussian assumption.

The GARCH approach allows for accessing directly the effect of mean changes in the dynamics of the conditional variance. This is an important issue in financial time series, as changes in the mean tend to have a relevant impact on the uncertainty of the process under study. However, the conditional variance may follow different regimes according to the size and signal of the shock. This happens because news could have an asymmetric impact on the economy and, for example, a large negative return might affect future volatility in a different way when compared to a positive return with the same size. Engle and Ng (1993) present a review regarding this issue. Awartani and Corradi (2005) discuss the importance of asymmetries in the prediction of an economic index. Glosten, Jagannathan, and Runkle (1993) assume that the conditional variance may react differently to positive and negative shocks. However, the change does not depend on the size of the shock leading to a abrupt change in the variance. Lubrano (2001) proposes a smooth transition model such that different specifications of the skedastic functions will take into account size and sign effects in the volatility.

In the context of tail behavior, 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, Creal, Koopman, and Lucas (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, Ferreira, and Migon (2008). For an illustration of this likelihood behaviour, Figure 1 presents the likelihood function for two datasets of size 150 for two parameters in the complete model we present in subsection 2.3. The first dataset has a well-behaved likelihood with a well-defined maximum in the true parameter values while the second dataset has an ill-behaved likelihood which goes to infinity as the parameters grow. 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 a 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 does not depend on the specification of

![Figure 1. Contour plots of the joint likelihood function for $(\gamma, \nu)$ for two illustrative datasets of size 150. The red “*” represents the true value of parameters. The parameters $\gamma$ and $\nu$ are the smooth transition and degrees of freedom parameters, respectively.](image-url)
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 the estimation of model parameters in the Student-t GARCH model are affected by small sample sizes, prior distributions and misspecification regarding the sampling distribution.

In section 2 we present the autoregressive moving average (ARMA) model with a GARCH component. We relax 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 $t \in Z_+$. For the mean term we assume an autoregressive moving average (ARMA) model and add a heteroskedasticity term (M) as suggested in Engle and Kroner (1995) to account for the effects of uncertainty on the mean function.

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

where $u_t$ are error terms with variance $h_t$ which are often modeled as Gaussian distributed; $\phi_1, \ldots, \phi_p$ are autoregression coefficients; $\theta_1, \ldots, \theta_q$ are moving average coefficients; and $\delta$ 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 $\epsilon_t$ and a positive mixing random variable $\omega_t$ as follows:

$$u_t = \epsilon_t \left( \frac{\nu - 2}{\nu} h_t \omega_t \right)^{1/2}, \quad t = p + 1, \ldots, N,$$

where $\epsilon_t \sim N(0, 1)$, $\omega_t \sim \text{IG}(\nu/2, \nu/2)$. Here $N(\mu, \sigma^2)$ denotes the Gaussian distribution with mean $\mu$ and variance $\sigma^2$ and $\text{IG}(a, b)$ denotes the Inverse Gamma distribution with mean $a/b$ and variance $a/b^2$. The parameter $\nu \in \mathbb{R}_+$ 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 $\nu \to \infty$ then $u_t \sim N(0, h_t)$, while for finite $\nu$ the distribution of $u_t$ will be Student-t with $\nu$ degrees of freedom. In the mixture representation, $\omega_t$ 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

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

2.2 GARCH specification

The variances $h_t$ are considered to be heterocedastic and given by the GARCH model which we denote by $\text{GARCH}(r, s)$,

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

with the restriction $\sum_{j=1}^{r} \alpha_j + \sum_{j=1}^{s} \beta_j < 1$, $\omega > 0$, $\alpha_i, \beta_j \geq 0$. For more details about stationarity of solutions based on moment conditions for the GARCH model see Ling and McAleer (2002). Consider observations $y_0 = (y_1, \ldots, y_p)'$ as known and set $u_p = u_{p-1} = \cdots = u_{p-q} = u_{p-q+1} = 0$. From equation (1) we may define $u_t, t = p + 1, \ldots, N$ recursively by

$$u_t = y_t - \sum_{j=1}^{p} \phi_j y_{t-j} - \sum_{j=1}^{q} \theta_j u_{t-j} - \delta \sqrt{h_t}. \quad (5)$$
From an inferential point of view, it is convenient to rewrite equation (1) as

\[ y = X\phi + A\theta + \tilde{H}\psi + u, \]  

(6)

where \( y = (y_{p+1}, \ldots, y_N)' \); \( \phi = (\phi_1, \ldots, \phi_p)' \); \( u = (u_{p+1}, \ldots, u_N)' \); \( \theta = (\theta_1, \ldots, \theta_q)' \); \( \psi = \delta 1 \); and the matrices,

\[ X = \begin{bmatrix} y_p & y_{p-1} & \cdots & y_1 \\ y_{p+1} & y_p & \cdots & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-1} & y_{N-2} & \cdots & y_{N-p} \end{bmatrix}, \]

\[ A = \begin{bmatrix} u_p & \cdots & u_{p-q+1} \\ u_{p+1} & \cdots & u_{p-q+2} \\ \vdots & \vdots & \vdots \\ u_{N-1} & \cdots & u_{N-q} \end{bmatrix}, \]

\[ \tilde{H} = \begin{bmatrix} \sqrt{h_{p+1}} & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \sqrt{h_N} \end{bmatrix}. \]

In the next section, we allow for asymmetric shock effects in order to consider more realistic set-ups for financial 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

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

(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. 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 \( \lambda \). On the contrary, if \(|u|\) is small then \( f_t \) will be close to 0 and \( h_t \) will be less affected by \( \lambda \). This reflects different effects of small and large past error terms in the variance \( h_t \). Consider the logistic function \( f_t(u, \gamma) = \frac{1 + \exp\{-\gamma u\}}{1 + \exp\{-\gamma u\}} \). If \(|u|\) is large, then \( f_t \) will tend to 1 and \( h_t \) will be affected by \( \lambda \). On the contrary, if \(|u|\) is small then \( f_t \) will be close to 0 and \( h_t \) will be less affected by \( \lambda \). This reflects different effects of small and large past error terms in the variance \( h_t \). Consider the logistic function \( f_t(u, \gamma) = \frac{1 + \exp\{-\gamma u\}}{1 + \exp\{-\gamma u\}} \). 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 \( u \to -\infty \) then \( f_t(u) \to 0 \) and the impact of \( f_t(u) \) in \( h_t \) is 0. As a result, when \( f_t(u_{t-1}, \gamma) = (1 + e^{-\gamma u_{t-1}})^{-1} \) 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 González-Rivera (1998) and Hagerud (1997). Note that the model specified for the conditional variance also impacts the mean through the \( M \) term in equation (1). It is often realistic to assume that the mean function is affected by the uncertainty, represented by the conditional variance \( h_t \). In this context, a flexible model for the conditional variance is crucial for correct uncertainty quantification. The smooth transition model plays the role of a general specification for the uncertainty so that if regime change is abrupt the threshold model is obtained as a limiting case. In case of a smooth change this can be captured by the parameter \( \gamma \) representing the velocity of transitions. As follows we consider two parameterisations to describe asymmetry:

1. If \( f_t(u_{t-1}, \gamma) = (1 - e^{-\gamma u_{t-1}^2}) \) then
   \[
f_t'(u_{t-1}, \gamma) = u_{t-1}^2 e^{-\gamma u_{t-1}^2},
   \]
   which allows for small and large effects.

2. If \( f_t(u_{t-1}, \gamma) = (1 + e^{-\gamma u_{t-1}})^{-1} \) then
   \[
f_t'(u_{t-1}, \gamma) = u_{t-1} e^{-\gamma u_{t-1}} (1 + e^{-\gamma u_{t-1}})^{-2},
   \]
   which allows for positive and negative effects.

The usual GARCH model without smooth transitions has symmetric impact of positive and negative shocks. The introduction of \( \gamma \) in the model produces an asymmetric effect of news. Furthermore, as \( \gamma \) increases the model becomes more asymmetric for both parameterisations. The effect of increasing \( \gamma \) is illustrated for the Logistic smooth model in Figure 2 which displays the impact of news \( u_{t-1} \).
on the variance $h_t$. If $\gamma$ is set to 0 then the symmetric GARCH model is obtained for both parametrisations. For both $\gamma \to \infty$ and $\gamma = 0$ the symmetric model is obtained and estimation problems occur. For instance, the first parameterisation $\gamma = 0$ implies $f_t(u_{t-1}, \gamma) = 0$ and $\lambda$ is nonidentifiable.

**Figure 2.** Impact curve for the Logistic smooth transition function with $\omega = 0.25$, $\alpha_1 = 3$ and $\lambda = 0.5$ and varying $\gamma$.

### 3 Likelihood issues and Bayesian Inference

If we consider latent variables $\omega_t \sim \text{IG}(\nu/2, \nu/2)$ and the mixture representation (2) then the likelihood function for model (1) is given by

$$L(\phi, \theta, \psi, \nu \mid X, A, \hat{H}, y, y_0) \propto |H|^{-1/2}\exp\left\{-\frac{1}{2}u' H^{-1} u\right\},$$

(8)

with $u = y - X\phi - A\theta - \hat{H}\psi$; $H = \text{diag}(\omega_{p+1}\frac{\nu-2}{\nu}h_{p+1}, \ldots, \omega_N\frac{\nu-2}{\nu}h_N)$. The estimation of the degree of freedom parameter $\nu$ is not straightforward. 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 $\nu$ is well behaved or
Evaluating the performance of degrees of freedom estimation in Asymmetric GARCH models with Student-t innovations

If the following condition is satisfied, the likelihood is not well behaved:

\[
\sum_{i=1}^{n} (\hat{u}_i^2 - 1)^2 < 2n,
\]

where \(n\) is the sample size and \(\hat{u}_i\) are standardised 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 \(\nu\) in the ARMA-GARCH set up. For the parameters \(\Phi, \theta, \psi, \alpha, \beta\) we consider flat prior distribution given by

\[
p(\phi, \theta, \psi, \alpha, \beta \mid \nu) \propto k.
\]

While for \(\nu\) we consider the independent Jeffreys prior given by

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

where \(\vartheta(a) = d \log \Gamma(a)/da\) and \(\vartheta'(a) = d^2\{\vartheta(a)\}/da\) 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 \(\nu\) than the maximum likelihood estimator. They also prove that the marginal posterior distribution of \(\nu\) 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 penalised 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 adequately 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. Geweke (1993) proposed the exponential prior for the degrees of freedom, however, the model depends on the specification of a hyperparameter which might bias the
estimation and provide poor coverage. This issue is investigated in Fonseca et al. (2008). 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 $\gamma$. 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 $\gamma$. Thus, the prior distribution of $\gamma$ needs to have tails that go fast enough to 0, allowing the posterior distribution to be an integrable function of $\gamma$. Lubrano (1998) suggested the prior given by

$$\pi(\gamma) = \left[1 + (\gamma - \gamma_0)^2\right]^{-1}, \quad \gamma, \gamma_0 > 0.$$  \hspace{1cm} (12)

Notice that for this proposal it is required to specify a hyperparameter $\gamma_0$. An alternative approach is presented by Goodwin, Holt, and Prestemon (2001) and Silvennoinen and Teräsvirta (2016) which consider a reparametrization $\gamma = \exp(\eta)$. However, alike discussed in Fonseca et al. (2008) for $\nu$, the flatness in likelihood of $\gamma$ is not corrected by parameter transformation and the non-identifiability problem persists.

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 $\alpha$, $\beta$ and $\nu$. The complete conditional distributions are presented in the Appendix A.
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 $p_m(\theta_m)$, $\theta_m \in \Theta_m$, $m = 1, \ldots, M$. 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 $(x,m,\theta_m)$ 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 $(m,\theta_m)$ after observing $x$. The posterior inference in the presence of uncertainties about the correct model involves the evaluation of the posterior $p(m|x)$, $m = 1, \ldots, M$ which depends on the marginal distribution $p_m(x)$ and the evaluation of $p_m(\theta_m|x)$ the posterior distribution of $\theta_m$. Let the Bayes Factor (Kass & Raftery, 1995) of model 1 with respect to model 2 be

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

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.

| $2\ln(B_{12})$ | $B_{12}$ | 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            |

Let us now consider the special case of two alternative models. A decision problem is completely specified by the triple $\{A, \Theta, \mathcal{X}\}$, where $A$ is the decision space, $\Theta$ the parameter space and $\mathcal{X}$ is the sample space. Let $L(\theta,a)$ be a loss function for decision $a \in A$ and $\theta \in \Theta$.

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 $\Theta_1$ and $\Theta_2$. The action space is defined by two components, $A = \{a_1,a_2\}$, with $a_i$ meaning that the hypothesis $H_i$ is accepted. Often
\( L(\theta, a_i) = 0 \) if \( \theta \in \Theta_i \) and \( K_j \) 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: \( \delta : \mathcal{X} \to \{0, 1\} \). As is well known from the decision making literature \( a_1 \succ a_2 \) if and only if \( \mathbb{E}[L(\theta, a_1)] < \mathbb{E}[L(\theta, a_2)] \), where the expectation is with respect to the posterior distributions. This is equivalent to accept \( H_1 \) if and only if \( \frac{P(H_1|x)}{P(H_2|x)} > \frac{k_2}{k_1} \), which is equivalent to

\[
B_{12} > \frac{k_2 P(H_2)}{k_1 P(H_1)}.
\]

(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 straightforward. It is needed to consider the samples obtained from the MCMC algorithm in order to numerically compute the predictive 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.

\[
\hat{p}_1(x) = \frac{dm}{1 - d} + \sum_{i=1}^{m} p(x | \theta^{(i)}) \left\{ d\hat{p}_1(x) + (1 - d)p(x | \theta^{(i)}) \right\}^{-1},
\]

(15)

where \( \theta^{(1)}, \ldots, \theta^{(m)} \) 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, Newton, Satagopan, and Krivitsky (2007). In this proposal, the outputs of the MCMC algorithm is used to calculate a sequence of log-likelihood values \( \{l_k : k = 1, \ldots, n\} \) and the posterior distribution of the log-likelihoods is given by

\[
l_{\text{max}} - l_k \sim \text{Gamma}(\alpha, \lambda^{-1}),
\]

(16)

where \( l_{\text{max}} \) is the maximum achievable likelihood, \( \alpha = d/2; \) \( d \) is the number of parameters in the model; and \( \lambda < 1 \). In practice, \( \lambda \) 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

\[
\log(p(x)) = l_{\text{max}} + \alpha \log(1 - \lambda).
\]

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

**5 Simulation study**

In this section we perform a Monte Carlo simulation study to evaluate the effects of model misspecification in the results of Bayesian hypothesis testing as described in section 4 and also to evaluate prediction performance of different models. For each scenario considered (sample size and sampling distribution) we simulate 100 datasets of size \(T\). In this study we vary the sample size \((T = 300, 600)\) and the sampling distribution (Gaussian and Student-t). For each dataset several measures of model performance were computed.

For model choice, a success is defined when the optimal selected decision 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 success by Monte Carlo simulation. We compute Mean Squared Errors for parameters in the GARCH model for all scenarios. We also evaluate the one-step-ahead variance in order to compare the Gaussian and Student-t models.

The model formulation considered for simulation is defined according to equations (1) and (7), that is,

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

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

Regarding the GARCH setup, Hansen and Lunde (2005) suggest the GARCH(1,1) as a parsimonious and competitive specification for modelling volatility. For the transition velocity \(\gamma\), we follow McAleer and Medeiros (2008) that shows evidence of slow velocities for financial series, thus we set \(\gamma = 5\). For all simulated
and real data applications this assumption was considered. To reflect realistic values for the other parameters, a small MCMC sample was obtained for a large Dow Jones log return series (03/01/2000 to 30/12/2008) with 4,769 observations. The estimated values of parameters were used to define the simulation study as follows. The parameters are set to be $\omega = 0.01$, $\alpha = 0.01$, $\beta = 0.91$, $\gamma = 5$, $\phi = -0.01$, $\theta = 0$, $\delta = 0$, $\lambda = 0.18$. Notice that $\delta$ is set to 0 in this simulation, however, in section 6 (Dow Jones data analysis) this parameter is estimated to allow for the data to dictate the effect of uncertainty in the mean. The sampling distribution considered for $u_t$ is either Gaussian with $u_t = \epsilon_t h_t$ or Student-t with

$$u_t = \epsilon_t \left( \frac{\nu - 2}{\nu} h_t \omega_t \right)^{1/2}, \quad t = 2, \ldots, N,$$

where $\epsilon_t \sim N(0, 1)$, $\omega_t \sim IG(\nu/2, \nu/2)$. For the degrees of freedom we considered datasets with $\nu = 2.5$ or $\nu = 5$.

Tables 2, 3 and 4 shows the Mean Square Error (MSE) for datasets simulated (n=300,600) 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 accommodate Gaussianity as a limiting case. On the other hand, for Student-t data with 2.5 degrees of freedom the Gaussian model has a poor performance when compared with the Student-t model as presented in Table 3 for both sample sizes. Notice that the Gaussian model result in very high MSE for the asymmetry parameter $\lambda$. In fact, the MSE for the Gaussian model is more than 5 times larger than for the Student-t model for $n = 300$. As expected, the same does not happen for $\nu = 5$ degrees of freedom, which performs more closely to the Gaussian model as shown in Table 4, with most of the smaller MSE obtained for the Student-t model.

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 ($n = 300$) and Gaussian data, in this case both models are similar and the model generating the data is selected 78% of the times. In summary, we recommend the use of the Bayes factor for model choice for the ARMA-GARCH with Student-t or Gaussian errors.
Evaluating the performance of degrees of freedom estimation in Asymmetric GARCH models with Student-t innovations

Table 2. Mean Square Error (MSE) for $\alpha$, $\beta$, $\lambda$ and $\gamma$ obtained in the simulated study, $n = 300$ and $n = 600$.

| Model | Gaussian | Student-t | Gaussian | Student-t |
|-------|----------|-----------|----------|-----------|
| $\alpha$ | 0.0011 | 0.0015 | 0.0003 | 0.0005 |
| $\beta$ | 0.0085 | 0.0200 | 0.0009 | 0.0008 |
| $\lambda$ | 0.0070 | 0.0120 | 0.0015 | 0.0017 |
| $\gamma$ | 9.2795 | 10.2858 | 6.0499 | 6.2145 |

Table 3. Mean Square Error (MSE) for $\alpha$, $\beta$, $\lambda$ and $\gamma$ obtained in the simulated study, $n = 300$ and $n = 600$.

| Model | Gaussian | Student-t (\(\nu = 2.5\)) | Gaussian | Student-t (\(\nu = 2.5\)) |
|-------|----------|----------------------------|----------|----------------------------|
| $\alpha$ | 0.0102 | 0.0111 | 0.0020 | 0.0018 |
| $\beta$ | 0.2629 | 0.1683 | 0.0744 | 0.0265 |
| $\lambda$ | 1.3345 | 0.2340 | 0.7575 | 0.1132 |
| $\gamma$ | 12.3270 | 12.8239 | 11.7154 | 8.1966 |

Table 4. Mean Square Error (MSE) for $\alpha$, $\beta$, $\lambda$ and $\gamma$ obtained in the simulated study, $n = 300$ and $n = 600$.

| Model | Gaussian | Student-t (\(\nu = 5\)) | Gaussian | Student-t (\(\nu = 5\)) |
|-------|----------|-----------------------------|----------|-----------------------------|
| $\alpha$ | 0.0055 | 0.0063 | 0.0003 | 0.0004 |
| $\beta$ | 0.1117 | 0.0759 | 0.0022 | 0.0011 |
| $\lambda$ | 0.1263 | 0.1131 | 0.0059 | 0.0045 |
| $\gamma$ | 11.4943 | 11.0689 | 6.5714 | 6.2563 |
In terms of predictive variance (Table 6), the Gaussian model presented smaller MSE for the Gaussian data, while the Student-t model had smaller MSE for all Student-t scenarios. For small sample size \((n = 300)\) and Student-t data \((\nu = 2.5)\), the Gaussian model presented very large MSE, indicating poor variance prediction.

Table 5. Results for the hypothesis testing procedure based on Bayes Factors. Proportion of right decisions obtained in the simulated study, \(n = 300\) and \(n = 600\).

| Model       | Data \((n = 300)\) | Data \((n = 600)\) |
|-------------|---------------------|---------------------|
|             | Gaussian            | Student-t \((\nu = 2.5)\) | Gaussian            | Student-t \((\nu = 5)\) |
| Gaussian    | 0.78                | –                    | 0.99                | –                    |
| Student-t \((\nu = 2.5)\) | –                    | 0.94                | –                    | 0.90                |
| Student-t \((\nu = 5)\) | –                    | –                    | –                    | 1.00                |

Table 6. Mean Square Error (MSE) for the predictive variance obtained in the simulated study in the one-step-ahead prediction, \(n = 300\) e \(n = 600\).

| Model       | \(n = 300\)  | \(n = 600\)  |
|-------------|--------------|--------------|
|             | Gaussian     | Student-t    | Gaussian     | Student-t    |
| Data        |              |              |              |              |
| Gaussian    | 374.02       | 670.39       | 1246.49      | 2605.07      |
| Student-t \((\nu = 2.5)\) | 2769.58    | 572.26       | 14.63        | 0.01         |
| Student-t \((\nu = 5)\) | 1744.09    | 54.44        | 718.00       | 9.71         |

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 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 \(r_t = 100\ln(DJ_t/DJ_{t-1})\). We present results from 02/01/2007 to 30/12/2008,
with the first period going up to 01/06/2008 used for estimation which results in 352 observations and the final part (151 observations) used for predictive performance evaluation. We successively 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.

Correlation in the series was initially investigated computing the empirical autocorrelation (acf) and partial autocorrelation (pacf) functions as presented in Appendix B. The plots indicate that autocorrelation must be accounted for in the modeling. Heteroskedasticity was also investigated by analysing the acf and pacf functions for the squared returns.

In this application, we fitted two asymmetric models: the logistic and exponential as presented in subsection 2.3. Thus, we have 4 model formulations: Gaussian GARCH with logistic asymmetry, Gaussian GARCH with exponential asymmetry, Student-t GARCH with logistic asymmetry and Student-t GARCH with exponential asymmetry. In addition, we compute prediction based on a hybrid approach which considers the results from hypothesis tests as presented in section 4. The asymmetric model of González-Rivera (1998) was also tested but the performance was worse than the Logistic and Exponential models and results are omitted.

In the MCMC algorithm we considered ten thousand iterations and the acceptance rate was tuned 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.

In the context of model fitting, Table 7 presents the Bayes Factor comparison for the period 02/06/2008 to 12/09/2008 and the whole period 02/09/2008 to 31/12/2008. The criteria indicate that the Exponential model is preferable to the Logistic model in the first period. However, the Logistic model is preferred when we account for the second period and consider the whole period 02/09/2008 to 31/12/2008. Regarding Gaussianity, the tests indicate the Student-t model as the best fit for both asymmetry function and periods. Note that the larger preference of Student-t model is observed when we include the more volatile
period in the analysis. If all data is considered, including the final period from 15/09/2008 to 31/12/2008 the Student-t model is strongly preferred when compared to the Gaussian model. This is due to the presence of large volatility in the second period. Overall the model selected for the complete data is the Student-t model with Logistic function.

The best model configuration for the complete period (Student-t with Logistic function) was also analysed regarding the ARMA specification. Three models were compared based on Bayes Factor with the ARMA(3,0) providing the best fit (Table 8).

As follows the main posterior summaries are presented for the best model. The posterior median for the degree of freedom parameter was 5.44 indicating that the tails do not behave as the Gaussian model and heavier tails are necessary for this application. Regarding the mean parameters, the posterior medians are $\hat{\mu} = 0.13$, $\hat{\phi}_1 = -0.11$, $\hat{\phi}_2 = -0.04$ and $\hat{\phi}_3 = 0.03$. The parameter $\delta$ was allowed to be different from 0 a priori and the posterior median was $\hat{\delta} = -0.07$. The parameter $\gamma$ defining the smooth transition function is estimated as $\hat{\gamma} = 1.27$ which is not close to 0 and not too large to indicate a limiting behaviour. The parameter $\lambda$ is well identified as $\hat{\lambda} = 0.20$. Table C1 with posterior summaries (median and credible intervals) is presented in Appendix C.

Next, we investigate the predictive performance for the proposed model assuming the ARMA(3,0) for the mean and a GARCH(1,1) for the variance.

**Table 7.** Model selection test for the logistic and exponential asymmetry functions. The values are $2\ln(B_{12})$ in favor of the Student-t model against the Gaussian model.

| Time period                  | Logistic | Exponential |
|------------------------------|----------|-------------|
| 02/06/2008–12/09/2008        | 15.11    | 21.84       |
| 02/06/2008–31/12/2008        | 29.56    | 23.39       |

**Table 8.** Model selection test for the ARMA component based on Bayes factor.

| Models                  | $2\ln(B_{12})$   | Evidence                           |
|-------------------------|------------------|------------------------------------|
| ARMA(1,0) × ARMA(3,0)   | 13.61            | Very strong evidence against ARMA(1,0) |
| ARMA(2,0) × ARMA(3,0)   | 5.36             | Strong evidence against ARMA(2,0)   |
| ARMA(1,1) × ARMA(3,0)   | 19.48            | Very strong evidence against ARMA(1,1) |
Figure 3 presents the variance proxy evolution through time and the predictions. Note that the Exponential model predicted a larger volatility from 09/2008 to 12/2008. However, this did not result in better predictive performance as the root MSE for this period was 523.27 (Logistic) and 817.82 (Exponential) for the Student-t model.

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

\[
I_{t+5} = \frac{\sum_{n=1}^{5}(\hat{h}^G_{t+n} - h_{t+n})^2}{\sum_{n=1}^{5}(\hat{h}^{ST}_{t+n} - h_{t+n})^2},
\]

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. Figure 4 presents the MSE ratio (Gaussian versus Student-t) for \( I_{t+5} \) across time for the Logistic and Exponential models, respectively. The evolution in panel (a) indicates that for the Logistic model, periods with high volatility tend to have \( I_{t+5} \) greater than 1, suggesting that the Student-t model would be more indicated. For the Exponential model (panel (b)) this difference is less evident, although there are periods in which the Student-t model would be recommended. Indeed, the correlation between \( I_{t+5} \) and the squared returns is 0.54 for the Logistic model and it is 0.29 for the Exponential model. This confirms that large instability (large squared returns) are positively correlated to the better predictive performance of Student-t models (\( I_{t+5} > 1 \)). Figure 5 illustrates this relation for the Logistic model.

7 Conclusions

This work investigates the Student-t smooth transition models from a Bayesian point of view. Our main interest in this paper was to evaluate 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 intrinsic to the likelihood and not corrected by usual parametric priors such as exponential and gamma. This is
Figure 3. Squared returns (light grey full line), Gaussian (grey dashed line) and Student-t (black full line) predictions with asymmetric volatility model.

Figure 4. Mean squared error ratio between the Gaussian and Student-t models for the past five days. Values greater than one indicate the Student-t model is preferred for prediction based on MSE error.
Evaluating the performance of degrees of freedom estimation in Asymmetric GARCH models with Student-t innovations

Figure 5. The squared returns ($r_t^2$) versus the mean squared error ratio between the Gaussian and Student-t models for the past five days ($I_{t+5}$) for the Logistic asymmetry function and Student-t model and linear regression fit.

due to the existence 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 indicates that some factors such as parameter estimation and prediction one step ahead may increase competitively of Student-t models. For instance, this is the case when data is simulated from Student-t models with relatively heavy tails ($\nu = 2.5$).

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, especially for the logistic asymmetric model 2. This is crucial to correctly estimate the uncertainty in periods of large volatility in the market. Note that even though the Gaussian model is a better predictive model for certain periods (according to MSE), the Student-t model can account for Gaussian kurtosis by estimating a large value for the degree of freedom. Ideally, the model
kurtosis should change over time to account for periods with lighter or fatter tails. This is a topic of ongoing research.

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 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. (2008). Financial risk management with Bayesian estimation of GARCH models: Theory and applications. Springer-Verlag. http://dx.doi.org/10.1007/978-3-540-78657-3

Awartani, B. M. A., & Corradi, V. (2005). Predicting the volatility of the S&P-500 index via GARCH models: The role of asymmetries. International Journal of Forecasting, 21(1), 167–184. http://dx.doi.org/10.1016/j.ijforecast.2004.08.003

Bauwens, L., & Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler. Econometrics Journal, 1(1), 23–46. http://dx.doi.org/10.1111/1368-423X.11003

Bauwens, L., & Lubrano, M. (2002). Bayesian option pricing using asymmetric GARCH models. Journal of Empirical Finance, 9(3), 321–342. http://dx.doi.org/10.1016/S0927-5398(01)00058-5

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327. http://dx.doi.org/10.1016/0304-4076(86)90063-1

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. The Review of Economics and Statistics, 69(3), 542–547. http://dx.doi.org/10.2307/1925546

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance. Journal of Econometrics, 52(1-2), 5–59. http://dx.doi.org/10.1016/0304-4076(92)90064-X

Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. Econometric Theory, 11(1), 122–150. http://dx.doi.org/10.1017/S0266466600009063

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. Journal of Finance, 48(5), 1749–1778. http://dx.doi.org/10.1111/j.1540-6261.1993.tb05127.x

Firth, D. (1993). Bias reduction of maximum likelihood estimates. Biometrika, 80, 27–38. http://dx.doi.org/10.1093/biomet/80.1.27
Evaluating the performance of degrees of freedom estimation in Asymmetric GARCH models with Student-t innovations

Fonseca, T. C. O., Ferreira, M. A. R., & Migon, H. S. (2008). Objective Bayesian analysis for the Student-t regression model. Biometrika, 95(2), 325–333. http://dx.doi.org/10.1093/biomet/asn001

Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. Journal of Applied Econometrics, 8(S1), 519–540. http://dx.doi.org/10.1002/jae.3950080504

Glosten, L. W., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance, 48(?), 1779–1801. http://dx.doi.org/?

González-Rivera, G. (1998). Smooth-transition GARCH models. Studies in Nonlinear Dynamics & Econometrics, 3(2), 61–78. http://dx.doi.org/10.2202/1558-3708.1041

Goodwin, B. K., Holt, M. T., & Prestemon, J. P. (2001). North American oriented strand board markets, arbitrage activity, and market price dynamics: A smooth transition approach. American Journal of Agricultural Economics, 93(4), 993–. http://dx.doi.org/10.1093/ajae/aar024

Hagerud, G. (1997). A new non-linear GARCH model (Akademisk Avhandling). Stockholm School of Economics, Stockholm, Sweden.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? Journal of Applied Econometrics, 10(7), 873–889. http://dx.doi.org/10.1002/jae.800

Harvey, A., & Chakravarty, T. (2008, September). Beta-t-(E)-GARCH. https://www.repository.cam.ac.uk/bitstream/handle/1810/229418/0840.pdf

Kass, R. E., & Raftery, A. E. (1995). Bayes factor. Journal of the American Statistical Association, 90(430), 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572

Ling, S., & McAleer, M. (2002). Stationarity and the existence of moments of a family of GARCH processes. Journal of Econometrics, 106(1), 109–117. http://dx.doi.org/10.1016/S0304-4076(01)00090-2

Lubrano, M. (1998). Smooth transition GARCH models: A Bayesian perspective (CORE Discussion Paper No. 1998/66). Universite Catholique de Louvain, Center for Operations Research and Econometrics. http://hdl.handle.net/2078.1/3968

Lubrano, M. (2001). Smooth transition GARCH models: A Bayesian perspective. Recherches Économiques de Louvain, 67(3), 257–287. http://dx.doi.org/10.1017/S0770451800004346

Mandelbrot, B. (1963). The variation of certain speculative prices. Journal of Business, 36(4), 394–419. http://dx.doi.org/10.1086/294632

McAleer, M., & Medeiros, M. C. (2008). A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries.
Newton, M. A., & Raftery, A. E. (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1), 3–48. [http://dx.doi.org/10.1111/j.2517-6161.1994.tb01956.x](http://dx.doi.org/10.1111/j.2517-6161.1994.tb01956.x)

Raftery, A. E., Newton, M. A., Satagopan, J. M., & Krivitsky, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Bayesian Statistics*, 8, 371–416.

Silvennoinen, A., & Teräsvirta, T. (2016). Testing constancy of unconditional variance in volatility models by misspecification and specification tests. *Studies in Nonlinear Dynamics & Econometrics*, 20(4), 347–364. [http://dx.doi.org/10.1515/snde-2015-0033](http://dx.doi.org/10.1515/snde-2015-0033)

Villa, C., & Walker, S. G. (2014). Objective prior for the number of degrees of freedom of a t distribution. *Bayesian Analysis*, 9(1), 197–220. [http://dx.doi.org/10.1214/13-BA854](http://dx.doi.org/10.1214/13-BA854)

Zhang, X., Creal, D., Koopman, S. J., & Lucas, A. (2011, May). *Modeling dynamic volatilities and correlations under skewness and fat tails* (Discussion Paper No. TI 11-078/2/DSF22). Amsterdam, The Netherlands: Tinbergen Institute, Duisenberg School of Finance.

**Appendix A**

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 $\phi$:**

  \[ \phi \mid \theta, \psi, \alpha, \beta, X, y, y_0 \sim N\left(\hat{\phi}, (X' H^{-1} X)^{-1}\right), \]

  where \( \hat{\phi} = (X' H^{-1} X)^{-1}(X' H^{-1} y - X' H^{-1} A \theta - X' H^{-1} \tilde{H} \psi) \).

- **Parameter $\theta$:**

  \[ \theta \mid \phi, \psi, \alpha, \beta, X, y, y_0 \sim N(\mu_\theta, V_\theta), \]

  where \( V_\theta = (A' H^{-1} A)^{-1} \) and

  \[ \mu_\theta = (A' H^{-1} A)^{-1}(A' H^{-1} y - A' H^{-1} X \Phi - A' H^{-1} \tilde{H} \psi). \]
Evaluating the performance of degrees of freedom estimation in Asymmetric GARCH models with Student-t innovations

- Parameter $\psi$:
  $$\psi \mid \phi, \theta, \alpha, \beta, X, y, y_0 \sim \mathcal{N}(\mu_\psi, V_\psi),$$
  where $V_\psi = (\hat{H}'H^{-1}\hat{H})^{-1}$ and
  $$\mu_\psi = (\hat{H}'H^{-1}\hat{H})^{-1}(\hat{H}'H^{-1}y - \hat{H}'H^{-1}X\phi - \hat{H}'H^{-1}A\theta).$$

- Parameter $ff$ and $fi$:
  $$p(\alpha, \beta \mid \phi, \theta, \psi, X, y, y_0) = \prod_{j=1}^{N-p} g(y_t \mid \mu_t, h_t),$$
  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 $g(\cdot \mid \mu, \sigma^2)$ is the Gaussian density function with mean $\mu$ and variance $\sigma^2$. We define independent proposal distributions given by
  $$\alpha_i \sim \mathcal{N}(\hat{\alpha}_i, \hat{\Sigma}_\alpha), \quad \beta_i \sim \mathcal{N}(\hat{\beta}_i, \hat{\Sigma}_\beta).$$

- Parameter $\nu$:
  $$p(\nu \mid \phi, \theta, \psi, \alpha, \beta, X, y, y_0) \propto |H|^{-1/2}\exp\left\{-\frac{1}{2}u'H^{-1}u\right\}\frac{1}{\Gamma(\nu/2)^{N-p}}\frac{(\nu/2)^{(N-p)\nu/2}}{\Gamma(\nu/2)\nu^{N-p}}$$
  $$\times \left(\prod_{t=1}^{N-p} \omega_t\right)^{-\nu/2-1} \exp\left\{-\frac{1}{2} \sum_{t=1}^{N-p} \frac{\nu}{\omega_t}\right\} p(\nu),$$
  where $u = (y - X\phi - A\theta - \hat{H}\psi)$. 

Appendix B

Autocorrelation and partial autocorrelation functions for the Dow Jones returns.

Figure B1. Autocorrelation (ACF) and partial autocorrelation (PACF) functions for the returns (a, b) and squared returns (c, d).
Appendix C

As follows, we present posterior summaries for the Student-t model with Logistic asymmetry function.

Table C1. Posterior summaries: median and 95% credible intervals.

| Parameter | Median | 95% CI     |
|-----------|--------|------------|
| $\mu$     | 0.13   | $(-0.04; 0.28)$ |
| $\phi_1$  | $-0.11$ | $(-0.20; -0.02)$ |
| $\phi_2$  | $-0.04$ | $(-0.14; 0.05)$ |
| $\phi_3$  | 0.03   | $(-0.06; 0.12)$ |
| $\delta$  | $-0.07$ | $(-0.25; 0.11)$ |
| $\omega$  | 0.05   | $(-0.02; 0.12)$ |
| $\lambda$ | 0.20   | $(0.07; 0.42)$ |
| $\gamma$  | 1.27   | $(0.23; 3.44)$ |
| $\beta$   | 0.85   | $(0.77; 0.90)$ |
| $\alpha$  | 0.04   | $(0.01; 0.18)$ |
| $\nu$     | 5.44   | $(3.03; 12.74)$ |