Brazilian Journal of Probability and Statistics
2012, Vol. 26, No. 4, 327–343
DOI: 10.1214/11-BJPS142

Bayesian analysis based on the Jeffreys prior for the hyperbolic distribution

Thaís C. O. Fonseca[a], **Helio S. Migon**[a] **and Marco A. R. Ferreira**[b]

[a]*Universidade Federal do Rio de Janeiro*
[b]*University of Missouri*

Abstract. In this work, we develop Bayesian analysis based on the Jeffreys prior for the hyperbolic family of distributions. It is usually difficult to estimate the four parameters in this class: to be reliable the maximum likelihood estimator typically requires large sample sizes of the order of thousands of observations. Moreover, improper prior distributions may lead to improper posterior distributions, whereas proper prior distributions may dominate the analysis. Here, we show through a simulation study that Bayesian methods based on Jeffreys prior provide reliable point and interval estimators. Moreover, this simulation study shows that for the absolute loss function Bayesian estimators compare favorably to maximum likelihood estimators. Finally, we illustrate with an application to real data that our methodology allows for parameter estimation with remarkable good properties even for a small sample size.

1 Introduction

The hyperbolic is a flexible distribution for data that may have heavy tails and skewness. The hyperbolic distribution heavy tails result from the fact that its log-density is a hyperbola Barndorff-Nielsen (1977). Since its introduction by Barndorff-Nielsen (1977), the hyperbolic distribution has been used with success in many areas of application such as turbulence (Barndorff-Nielsen, 1979), biology (Blæsild, 1981) and finance (Eberlein et al., 1998; Prause, 1999; Bauer, 2000; Bingham and Kiesel, 2001). Even though the hyperbolic distribution allows for both skewness and heavy tails through easily interpretable parameters, the task of parameter estimation is not trivial. To be reliable, maximum likelihood estimation typically requires large sample sizes of the order of thousands of observations. In addition, improper prior distributions may lead to improper posterior distributions, whereas proper prior distributions may dominate the analysis. As a solution to the estimation problem, we derive here the Jeffreys prior for the hyperbolic distribution. In addition, we show through a simulation study that Bayesian methods based on this Jeffreys prior provide reliable point and interval estimators even for small datasets.

Key words and phrases. Asymmetry, heavy tails, normal-mean mixture, noninformative prior.
Received December 2010; accepted February 2011.

In the context of robustness, Barndorff-Nielsen (1977) pointed out that the hyperbolic distribution can be represented as a normal-mean mixture with a generalized inverse Gaussian (GIG) as mixing distribution. More specifically, let $Y|\mu, \beta, \sigma^2 \sim N(\mu + \beta\sigma^2, \sigma^2)$ and $\sigma^2 \sim \text{GIG}(1, \alpha^2 - \beta^2, \delta^2)$, where $\text{GIG}(\lambda, \phi, \gamma)$ has density given by

$$f(x; \lambda, \phi, \gamma) = \frac{(\phi/\gamma)^{\lambda/2}}{2K_\lambda(\sqrt{\phi\gamma})} x^{\lambda-1} \exp\left\{-\frac{1}{2}(\gamma x^{-1} + \phi x)\right\}, \tag{1.1}$$

where $\phi, \gamma \geq 0$, $\lambda \in \Re$ and K_λ is the modified Bessel function of third-order and index λ. For additional information on the GIG distribution, see Jørgensen (1982), Silva et al. (2006) and references therein. Integrating σ^2 out, we obtain that Y has a hyperbolic distribution with density given by

$$\begin{aligned} f_{\text{hyp}}(y; \alpha, \beta, \mu, \delta) = & \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} \\ & \times \exp\{-\alpha\sqrt{\delta^2 + (y-\mu)^2} + \beta(y-\mu)\}, \end{aligned} \tag{1.2}$$

where, $y \in \Re$ and α, β, μ e δ are parameters, satisfying $|\beta| < \alpha$, $\mu \in \Re$ and $\delta > 0$. The parameters α and β determine the shape, where β is responsible for skewness; δ and μ are scale and location parameters, respectively. Figure 1 presents the density function (1.2) for $\alpha = 2.5$, $\beta \in \{-2, 0, 2\}$, $\delta = 1$ and $\mu = 0$. When β is negative, we obtain positive asymmetry, when β is positive we obtain negative asymmetry and $\beta = 0$ implies a symmetric density function. We use the notation $\text{Hyp}(\alpha, \beta, \mu, \delta)$ to denote the hyperbolic distribution with parameters α, β, μ and δ.

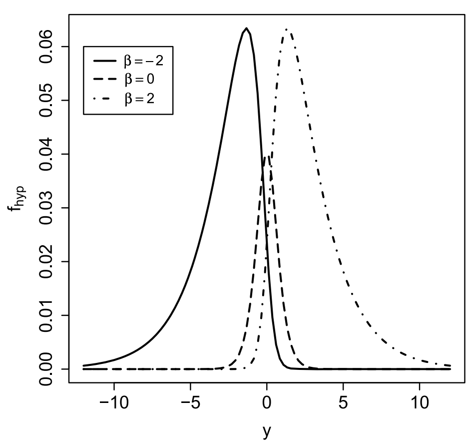


Figure 1 *Density function as presented in* (1.2) *for* $\alpha = 2.5$, $\beta \in \{-2, 0, 2\}$, $\delta = 1$ *and* $\mu = 0$.

All the moments of the Hyp($\alpha, \beta, \mu, \delta$) have explicit expressions and, in particular, the mean and the variance are

$$E(Y) = \mu + \frac{\beta\delta^2}{\rho}\frac{K_2(\rho)}{K_1(\rho)}, \tag{1.3}$$

$$\mathrm{Var}(Y) = \delta^2\left(\frac{K_2(\rho)}{\rho K_1(\rho)} + \frac{\beta^2\delta^2}{\rho^2}\left[\frac{K_3(\rho)}{K_1(\rho)} - \left(\frac{K_2(\rho)}{K_1(\rho)}\right)^2\right]\right), \tag{1.4}$$

where $\rho = \delta\sqrt{\alpha^2 - \beta^2}$. Note that when $\beta = 0$ the mean is simply μ. The mathematical properties of these univariate distributions are well-known (see Blæsild, 1981). Blæsild and Sørensen (1992) provide maximum likelihood methods to estimate parameters of this model. The HyperbolicDist package, within the R statistical environment (R Development Core Team, 2010), implements maximum likelihood estimation based on a number of numerical maximization methods.

Despite the nice properties of the hyperbolic distribution, difficulties arise in the estimation of its parameters. More specifically, for some samples the likelihood function is maximized when a combination of the parameters goes to infinity. As a consequence, the MLE may not exist with positive probability. This probability of nonexistence of the MLE is higher for smaller samples. In addition, for any finite sample size the likelihood function does not vanish in the tails. As a result, Bayesian analysis based on improper priors may lead to useless improper posterior distributions.

The problem of the likelihood function not vanishing in the tails also occurs for many other classes of distributions. For example, this problem occurs for the skew-normal distribution (Azzalini, 1985, 2005) and the Student-t distribution (Zellner, 1976). In the context of Bayesian inference for these distributions, these problems have been solved through the use of noninformative priors. For the skew-normal distribution, Liseo and Loperfido (2006) have proposed a default Bayesian solution based on the reference prior for the parameters. For the Student-t distribution, Fonseca et al. (2008) have proposed a default Bayesian solution based on the Jeffreys prior. The proposals of Liseo and Loperfido (2006) and Fonseca et al. (2008) lead to valid proper posterior distributions. Here, we obtain a default Bayesian solution based on the Jeffreys prior for the hyperbolic distribution. As we show in Section 4, our Bayesian proposal yields estimation procedures with good frequentist properties.

The remainder of this paper is organized as follows. In Section 2, we discuss the MLE difficulties associated with the hyperbolic model. In Section 3, we derive the Jeffreys prior for the parameters of the hyperbolic distribution. Section 4 presents the frequentist properties for the Bayesian and maximum likelihood estimators. Section 5 presents an application to real data that shows that our Bayesian methodology allows for reliable parameter estimation even for small datasets. Final discussions and some extensions are given in Section 6.

2 The hyperbolic model and MLE difficulties

Consider a random sample $y = (y_1, \ldots, y_n)$ from the hyperbolic distribution with density function given by equation (1.2). Then the likelihood function is given by

$$
\begin{aligned}
L(\alpha, \beta, \mu, \delta; y) = {}& \left\{ \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} \right\}^n \\
& \times \exp\left\{ -\alpha \sum_{i=1}^{n} \sqrt{\delta^2 + (y_i - \mu)^2} + \beta \sum_{i=1}^{n} (y_i - \mu) \right\}.
\end{aligned}
\tag{2.1}
$$

We denote the model parameters by $\theta = (\alpha, \beta, \mu, \delta)'$.

Maximum likelihood estimation for the hyperbolic distribution is problematic since several models are limiting or particular cases. For example, the normal distribution $N(\mu, \sigma^2)$ is a limiting case when $\beta = 0$, $\alpha \to \infty$ and $\delta/\alpha \to \sigma^2$. In addition, the Laplace distribution is a limiting case when $\beta = 0$ and $\delta \to 0$.

Proposition 2.1. *The likelihood function given in equation (2.1) satisfies,*

$$
L(\alpha, \beta, \mu, \delta; y) = O(1), \qquad \text{as } \delta \to \infty, \frac{\delta}{\alpha} \to \sigma^2, \beta = 0, \tag{2.2}
$$

with $L(\alpha, \beta, \mu, \delta; y) \to \prod_{i=1}^{n} \phi(y_i; \mu, \sigma^2)$, *where* $\phi(\cdot; \mu, \sigma^2)$ *is the normal density with mean* μ *and variance* σ^2.

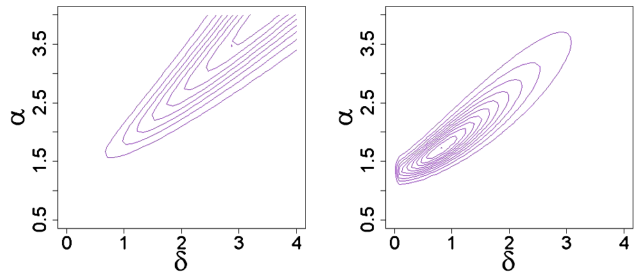
Proof. When $\beta \to 0$ we have that $f(y_i) \to \frac{1}{2\delta K_1(\delta\alpha)} \exp\{-\alpha\sqrt{\delta^2 + (y_i - \mu)^2}\}$. Moreover, for x large $K_1(x) \to \sqrt{\frac{\pi}{2x}} \exp\{-x\}$ (Abramowitz and Stegun, 1972, p. 378, equation (9.7.2)). Thus for $\delta \to \infty$ and $\frac{\delta}{\alpha} \to \sigma^2$ we find $f(y_i) \to \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\alpha\delta - \alpha\sqrt{\delta^2 + (y_i - \mu)^2}\}$. In addition, $\alpha\delta - \alpha\sqrt{\delta^2 + (y_i - \mu)^2} = \frac{-\alpha^2(y_i - \mu)^2}{(\alpha\delta + \alpha\sqrt{\delta^2 + (y_i - \mu)^2})}$ which converges to $\frac{-(y_i - \mu)^2}{2\sigma^2}$ as $\delta \to \infty$ and $\delta/\alpha \to \sigma^2$. □

Proposition 2.2. *The likelihood function given in equation (2.1) satisfies,*

$$
L(\alpha, \beta, \mu, \delta; y) = O(1), \qquad \text{as } \delta \to 0, \beta = 0, \tag{2.3}
$$

with $L(\alpha, \beta, \mu, \delta; y) \to \prod_{i=1}^{n} g(y_i; \alpha, \mu)$, *where* $g(\cdot; \alpha, \mu)$ *is the Laplace density with parameters* α *and* μ.

Proof. When $\beta \to 0$ we have that $f(y_i) \to \frac{1}{2\delta K_1(\delta\alpha)} \exp\{-\alpha\sqrt{\delta^2 + (y_i - \mu)^2}\}$. Moreover, for $x \to 0$, $K_1(x) \to x^{-1}$ (Abramowitz and Stegun, 1972, p. 375, equation (9.6.9)). Thus, for $\delta \to 0$ and α finite constant we find $f(y_i) \to \frac{\alpha}{2} \exp\{-\alpha|y_i - \mu|\}$. □



(a) Likelihood function. (b) Posterior density.

Figure 2 *Contour plots of the* (a) *likelihood function and* (b) *posterior density for* (α, δ), *holding* β *and* μ *at their true values*, *for a dataset of size* $n = 100$ *simulated from model* (1.2) *with* $\alpha = 2$, $\beta = 0$, $\delta = 1$ *and* $\mu = 2$.

To illustrate the consequences of Proposition 2.1 in terms of estimation, we have simulated a dataset from the hyperbolic distribution with parameters $\alpha = 2$, $\beta = 0$, $\mu = 2$ and $\delta = 1$, and with sample size $n = 100$. Figure 2(a) shows a contour plot of the likelihood function for α and δ holding β and μ at their true values. Unfortunately, the maximum of this likelihood function is located far from the true values of α and δ. We have used the function hyperbFit(·) of the R-package HyperbolicDist to compute the MLE of the parameters using three optimization methods: Newton, quasi-Newton, and Nelder–Mead. Table 1 shows the results, that are far from the true values of the parameters. The problem is not with the optimization methods, but with the bad behavior of the likelihood function. Other authors, for example, Barndorff-Nielsen and Blæsild (1981) and Eberlein and Keller (1995) have also noticed this problematic behavior of the likelihood function. In the extreme case, the likelihood may be maximized when a combination of the parameters goes to infinity and then the MLE may not exist.

To shed light on how the likelihood function problematic behavior depends on sample size, we have computed the probability of nonexistence of the MLE of α for the symmetric ($\beta = 0$) and asymmetric ($\beta = 0.1\alpha$) cases. For both cases, we consider sample sizes $n \in \{30, 50, 100, 200, 2{,}000\}$, and parameter values $\delta = \alpha$, $\mu = 0$, and $\alpha \in \{0.5, 1, 2, 3, 5\}$. Tables 2 and 3 present the results for the symmetric and

Table 1 *Maximum likelihood estimates for* $\theta = (\alpha, \beta, \mu, \delta)$ *obtained using* 100 *observations simulated from the* $\mathrm{Hyp}(\alpha, \beta, \mu, \delta)$ *with* $\alpha = 2$, $\beta = 0$, $\mu = 2$ *and* $\delta = 1$

θ	Newton	Q-Newton	Nelder–Mead
$\alpha = 2$	27.4663	21.1749	27.4605
$\beta = 0$	−0.4910	−13.4426	0.1944
$\mu = 2$	2.4747	9.5139	1.8321
$\delta = 1$	25.6837	9.0368	25.6855

Table 2 *Probability of nonexistence of the maximum likelihood estimator of* α. *The sample with size n was generated from* $\text{Hyp}(\alpha, \beta, \mu, \delta)$ *with* $\beta = 0$, $\mu = 0$ *and* $\delta = \alpha$

		α				
		0.5	1	2	3	5
n	30	0.214	0.347	0.524	0.585	0.647
	50	0.085	0.168	0.416	0.527	0.601
	100	0.022	0.061	0.290	0.417	0.547
	200	0.000	0.006	0.159	0.301	0.463
	2,000	0.000	0.000	0.000	0.015	0.182

asymmetric cases, respectively. Both cases lead to similar behavior with respect to the probability of nonexistence of the MLE of α. More specifically, this probability increases as α increases and decreases as the sample size n increases. Therefore, it seems safe to use maximum likelihood estimation for datasets of size $n = 2{,}000$ or larger, but for smaller sample sizes the MLE does not seem to be adequate.

In a Bayesian context, the bad behavior of the likelihood function is also an issue and, as a consequence, the choice of the prior distribution for the parameters is extremely important. As a solution for the inference problems, we propose the use of the Jeffreys prior as a calibration for the likelihood function. Figure 2(b) shows, for the same simulated dataset of Figure 2(a), the resulting posterior density for (α, δ) using the Jeffreys prior that we derive in the next section. The Jeffreys prior corrects the bad behavior of the likelihood function and leads to a posterior density located close to the true parameters values.

3 Jeffreys prior

In this section, we derive the Jeffreys prior for the parameters of the hyperbolic distribution. As shown in Firth (1993), the Jeffreys prior works as a calibration tool

Table 3 *Probability of nonexistence of the maximum likelihood estimator of* α. *The sample with size n was generated from* $\text{Hyp}(\alpha, \beta, \mu, \delta)$ *with* $\beta = 0.1\alpha$, $\mu = 0$ *and* $\delta = \alpha$

		α				
		0.5	1	2	3	5
n	30	0.199	0.306	0.517	0.558	0.597
	50	0.091	0.169	0.409	0.505	0.535
	100	0.016	0.057	0.269	0.423	0.524
	200	0.001	0.010	0.144	0.309	0.472
	2,000	0.000	0.000	0.001	0.016	0.237

for the information provided by the likelihood function. In particular, inference for small samples is possible using this approach as the Jeffreys prior compensates for the fact that the likelihood function in (2.1) does not vanish in the tails. As we shall see, this implies very different inferences obtained using the frequentist and default Bayesian estimation approaches for the parameters in this model, with results being in favor of the Bayesian approach.

Theorem 3.1. *The Jeffreys prior associated with model* (2.1) *is*

$$P^J(\theta) \propto |I(\theta)|^{1/2}, \tag{3.1}$$

where the elements of the the Fisher expected information matrix $I(\theta)$ *are given by*

$$\begin{aligned}
I_{11}(\theta) &= \frac{\alpha^2\delta^4 S_1 - \rho\delta^2 R_1}{\rho^2} - \frac{1}{\alpha^2},\\
I_{12}(\theta) &= -\frac{\alpha\beta\delta^4 S_1}{\rho^2},\\
I_{13}(\theta) &= -\frac{\beta}{\alpha},\\
I_{14}(\theta) &= \alpha\delta S_1 + \frac{\rho R_1}{\alpha\delta} - \frac{2\alpha\delta R_1}{\rho} - \frac{2}{\alpha\delta},\\
I_{22}(\theta) &= \frac{\rho\delta^2 R_1 + \beta^2\delta^4 S_1}{\rho^2},\\
I_{23}(\theta) &= 1,\\
I_{24}(\theta) &= \frac{2\beta\delta R_1}{\rho} - \beta\delta S_1,\\
I_{33}(\theta) &= \alpha^4(\varphi_2 - 2\mu\varphi_1 + \mu^2\varphi_0),\\
I_{34}(\theta) &= \alpha^4(\mu\delta\varphi_0 - \delta\varphi_1) + \frac{\beta\rho R_1 - 2\beta}{\delta},\\
I_{44}(\theta) &= \delta^2\alpha^4\varphi_0 - \frac{\rho^2 R_1^2}{\delta^2} + \frac{4\rho R_1 - 4}{\delta^2},
\end{aligned}$$

where $\theta = (\alpha, \beta, \mu, \delta)$, $R_1 = R_1(\rho) = \frac{K_2(\rho)}{K_1(\rho)}$, $S_1 = S_1(\rho) = \frac{K_3(\rho)}{K_1(\rho)} - R_1^2(\rho)$ *and* $\varphi_k(\alpha, \beta, \mu, \delta) = \int_{-\infty}^{\infty} \frac{y^k}{\alpha^2\delta^2 + \alpha^2(y-\mu)^2} f(y|\theta)\, dy$.

Proof. Define $\vartheta = \alpha\sqrt{\delta^2 + (y-\mu)^2}$.

Using the property that $E[\frac{\partial}{\partial\theta_i}\log L(\theta; y)] = 0$, $i = 1, 2, 3, 4$, the first derivatives of the log likelihood are given by

(a) $\frac{\partial}{\partial\alpha}\log L(\theta; y) = -\frac{1}{\alpha}\{\vartheta - E[\vartheta]\}$,
(b) $\frac{\partial}{\partial\beta}\log L(\theta; y) = y - E[y]$,
(c) $\frac{\partial}{\partial\delta}\log L(\theta; y) = -\alpha^2\delta\{\frac{1}{\vartheta} - E[\frac{1}{\vartheta}]\}$,
(d) $\frac{\partial}{\partial\mu}\log L(\theta; y) = \alpha^2\delta\{\frac{y-\mu}{\vartheta} - E[\frac{(y-\mu)}{\vartheta}]\}$.

The Fisher information matrix is given by

$$I_{ij} = E\left\{\left(\frac{\partial}{\partial\theta_i}\log L(\theta; y)\right)\left(\frac{\partial}{\partial\theta_j}\log L(\theta; y)\right)\right\} \qquad \text{for } i, j = 1, \dots, 4.$$

Thus, $I_{11} = E[(\frac{\partial}{\partial\alpha}\log L(\theta; y))^2] = \frac{1}{\alpha^2}\operatorname{Var}(\vartheta)$, which is obtained from the result $E[(y-\mu)^2] = \frac{\delta^2 R_1}{\rho} + \frac{\beta^2\delta^4 S_1}{\rho^2}$ from (1.3) and (1.4).

$I_{33} = E[(\frac{\partial}{\partial\mu}\log L(\theta; y))^2] = \alpha^4(\varphi_2 - 2\mu\varphi_1 + \mu^2\varphi_0)$.

$I_{44} = E[(\frac{\partial}{\partial\delta}\log L(\theta; y))^2] = \alpha^4\delta^2\operatorname{Var}(\frac{1}{\vartheta}) = \alpha^4\delta^2\{\varphi_0 - E^2[\frac{1}{\vartheta}]\}$, which follows from the expectation of (c).

$I_{13} = E[(\frac{\partial}{\partial\alpha}\log L(\theta; y))(\frac{\partial}{\partial\mu}\log L(\theta; y))] = -\alpha\operatorname{Cov}(\vartheta, \frac{y-\mu}{\vartheta})$, which follows from the expectation of $y-\mu$ in (1.3), (a) and (d).

$I_{14} = E[(\frac{\partial}{\partial\alpha}\log L(\theta; y))(\frac{\partial}{\partial\delta}\log L(\theta; y))] = \alpha\delta\operatorname{Cov}(\vartheta, \frac{1}{\vartheta})$, which follows from the expectation of (a) and (c).

$$\begin{aligned} I_{34} &= E\left[\left(\frac{\partial}{\partial\mu}\log L(\theta; y)\right)\left(\frac{\partial}{\partial\delta}\log L(\theta; y)\right)\right] = -\alpha^4\delta\operatorname{Cov}\left(\frac{1}{\vartheta}, \frac{y-\mu}{\vartheta}\right) \\ &= -\alpha^4\delta\left\{\varphi_1 - \mu\varphi_0 - E\left[\frac{1}{\vartheta}\right]E\left[\frac{y-\mu}{\vartheta}\right]\right\}, \end{aligned}$$

which follows from the expectation of (c) and (d).

As they do not depend on y, the terms I_{12}, I_{22}, I_{23}, and I_{24} are computed directly using $I_{ij} = E[-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log L(\theta; y)]$. □

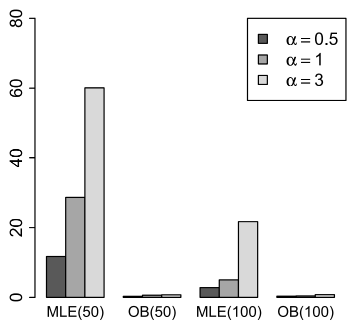
4 Frequentist properties

This section presents frequentist properties of the maximum likelihood estimator (MLE) and of the posterior median (Bayesian estimator) of θ based on the Jeffreys prior proposed here. We focus on the bias, the frequentist mean squared error (MSE) and the frequentist coverage of 95% credible intervals for two different sample sizes, $n = 50$ and $n = 100$. We have computed the bias, the MSE and the frequentist coverage via Monte Carlo simulation. More specifically, for each true value of $\alpha \in \{0.5, 1, 3\}$, $\beta \in \{0, 0.1\alpha\}$ and $\delta \in \{0.5, 3\}$ and each sample size

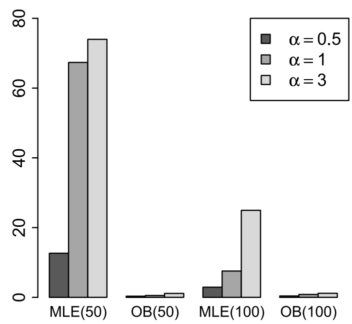
$n \in \{50, 100\}$ we have simulated 1,000 samples, computed the two estimates and the credibility interval for each sample and then estimated the bias, the MSE of each estimator and the frequentist coverage. The results obtained when $\beta = 0.1\alpha$ are very similar to the results when $\beta = 0$ and are not presented here. In all simulations, we have assumed $\mu = 2$. The MLE was obtained using the Nelder–Mead method in the R-package HyperbolicDist.

We have implemented a Metropolis–Hastings algorithm for the computation of the posterior median and the credible intervals. We sample in blocks (β, μ) and (α, δ) to ameliorate the problems caused by the high posterior correlation of the parameters and to speed up the convergence of the algorithm. In addition, the proposal distributions are $\mu^{(\text{prop})} \sim N(\mu^{(k)}, d_1^2)$, $\beta^{(\text{prop})} \sim \text{TN}_{(-\alpha^{(k)},\alpha^{(k)})}(\beta^{(k)}, d_2^2)$, $\alpha^{(\text{prop})} \sim \text{TN}_{(|\beta^{(k+1)}|,\infty)}(\alpha^{(k)}, d_3^2)$ and $\log(\delta^{(\text{prop})}) \sim N(\log(\delta^{(k)}), d_4^2)$, where $\text{TN}_A(\mu, \sigma^2)$ denotes the Gaussian distribution with location μ and scale σ^2 truncated to the region A. We have tuned the variances $d_1^2, \ldots, d_4^2$ to obtain an acceptance rate around 30%.

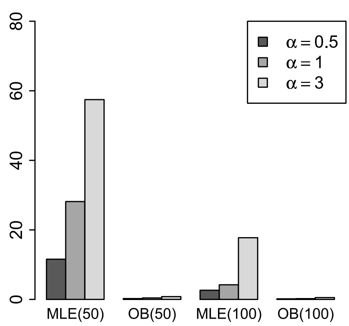
Figure 3 shows, for the four parameters in the model, the square root of the MSE for each combination of (α, δ, n). As expected, the MSE decreases as the



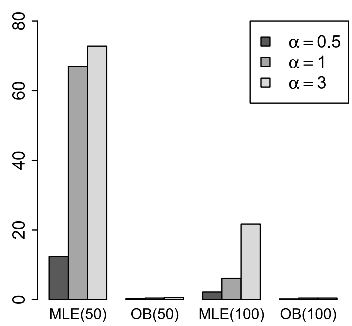
(a) $\sqrt{\text{MSE}(\hat{\alpha})}$, $\delta = 0.5$.



(b) $\sqrt{\text{MSE}(\hat{\alpha})}$, $\delta = 3$.

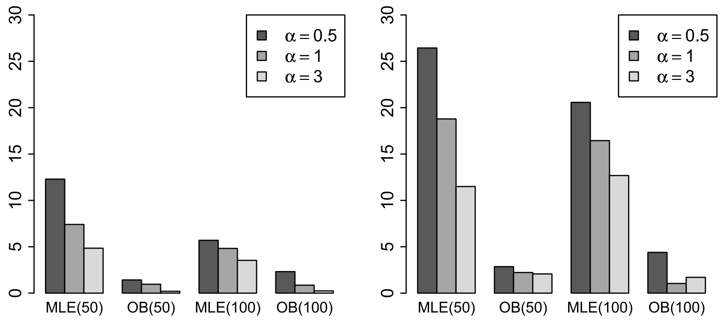


(c) $\sqrt{\text{MSE}(\hat{\beta})}$, $\delta = 0.5$.

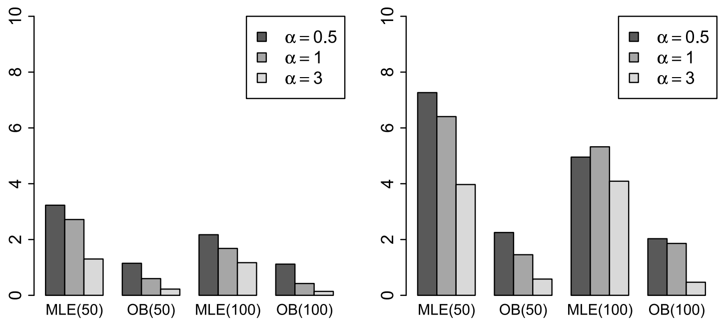


(d) $\sqrt{\text{MSE}(\hat{\beta})}$, $\delta = 3$.

Figure 3 *Square root of the mean squared error for the maximum likelihood* (*MLE*) *and Bayesian* (*OB*) *estimators of* α, β, δ *and* μ *when* $\beta = 0$.



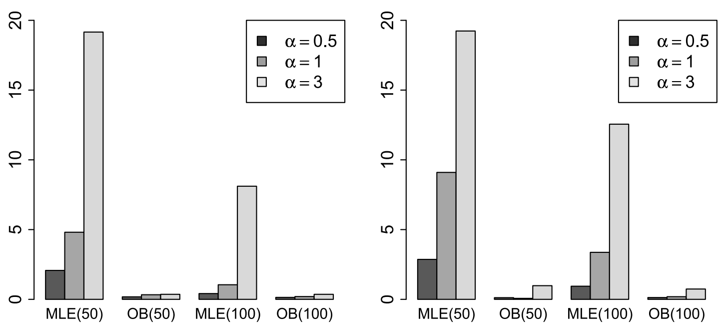
(e) $\sqrt{\mathrm{MSE}(\hat{\delta})}$, $\delta = 0.5$. (f) $\sqrt{\mathrm{MSE}(\hat{\delta})}$, $\delta = 3$.



(g) $\sqrt{\mathrm{MSE}(\hat{\mu})}$, $\delta = 0.5$. (h) $\sqrt{\mathrm{MSE}(\hat{\mu})}$, $\delta = 3$.

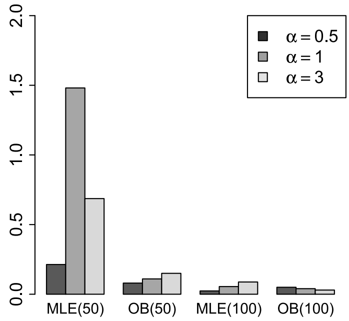
Figure 3 (*Continued*).

sample size increases. Moreover, the MLEs of α and β have MSEs two orders of magnitude larger than those of the competing Bayesian estimators. In addition, the MLEs of δ and μ have much larger MSEs than the competing Bayesian estimators. Figure 4 shows the absolute value of the bias of the maximum likelihood and

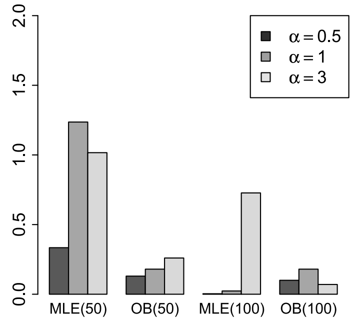


(a) $|\,\mathrm{Bias}(\hat{\alpha})|$, $\delta = 0.5$. (b) $|\,\mathrm{Bias}(\hat{\alpha})|$, $\delta = 3$.

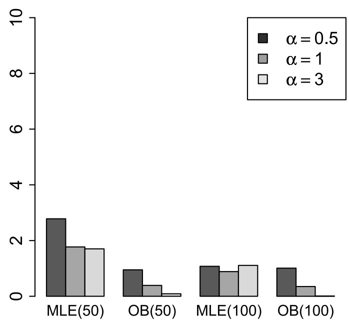
Figure 4 *Absolute value of the bias for the maximum likelihood* (*MLE*) *and Bayesian estimators* (*OB*) *of* α, β, δ *and* μ *when* $\beta = 0$.



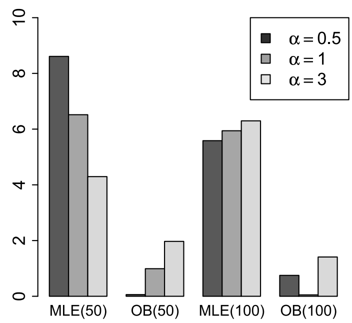
(c) $|\,\mathrm{Bias}(\hat{\beta})|$, $\delta = 0.5$.



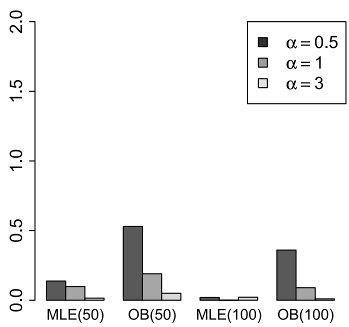
(d) $|\,\mathrm{Bias}(\hat{\beta})|$, $\delta = 3$.



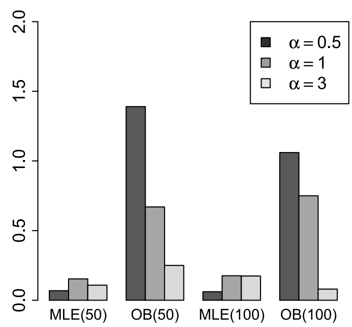
(e) $|\,\mathrm{Bias}(\hat{\delta})|$, $\delta = 0.5$.



(f) $|\,\mathrm{Bias}(\hat{\delta})|$, $\delta = 3$.



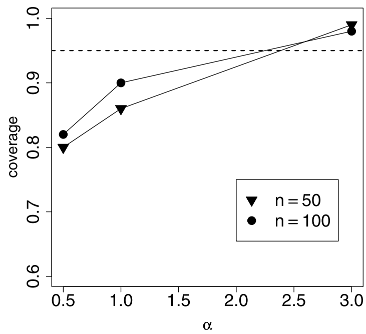
(g) $|\,\mathrm{Bias}(\hat{\mu})|$, $\delta = 0.5$.



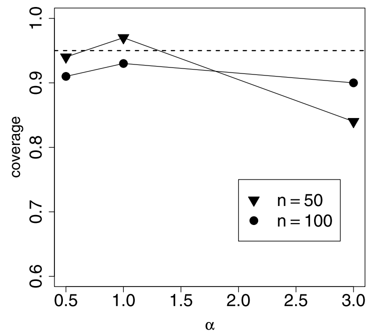
(h) $|\,\mathrm{Bias}(\hat{\mu})|$, $\delta = 3$.

Figure 4 (*Continued*).

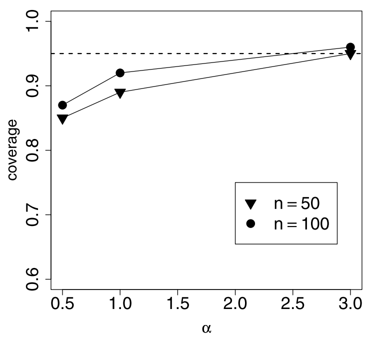
Bayesian estimators. The MLE of μ has smaller bias, but the MLEs of α, β and δ have much larger bias than the corresponding Bayesian estimators. It is important to note that the comparison of bias is not as important as the comparison of MSE. Specifically, the MSE already accounts for bias and variance of the estimator. As we mention above, when compared with the MLEs, the MSE of the Bayesian estimators is smaller for all parameters. Hence, even when the Bayesian estimator has a larger bias as in the case of μ, it has a much smaller variance than the MLE which leads to a smaller MSE. Figure 5 shows the frequentist coverage, as a function of the true value of α, of 95% credible intervals for α, β, δ and μ. The performance of the credible intervals decreases for larger δ. Finally, the frequentist coverage becomes closer to the nominal level as the sample size increases.



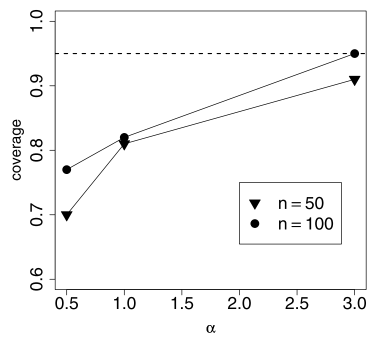
(a) Coverage($\hat{\alpha}$), $\delta = 0.5$.



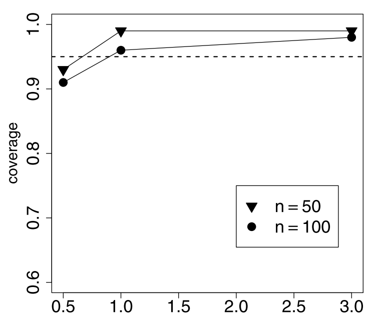
(b) Coverage($\hat{\alpha}$), $\delta = 3$.



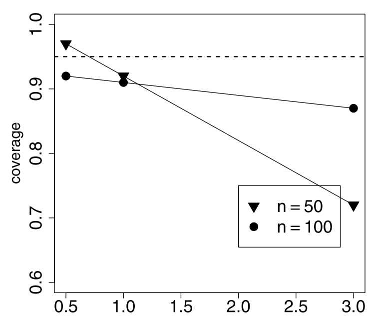
(c) Coverage($\hat{\beta}$), $\delta = 0.5$.



(d) Coverage($\hat{\beta}$), $\delta = 3$.

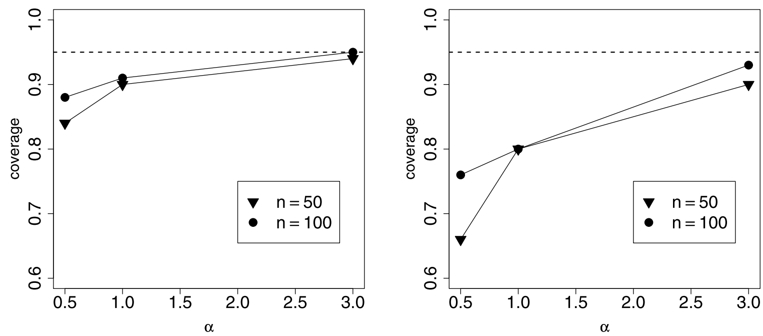


(e) Coverage($\hat{\delta}$), $\delta = 0.5$.



(f) Coverage($\hat{\delta}$), $\delta = 3$.

Figure 5 *Frequentist coverage, as a function of the true value of* α, *of* 95% *credible intervals for* α, β, δ *and* μ *when* $\beta = 0$.



(g) Coverage($\hat{\mu}$), $\delta = 0.5$. (h) Coverage($\hat{\mu}$), $\delta = 3$.

Figure 5 (*Continued*).

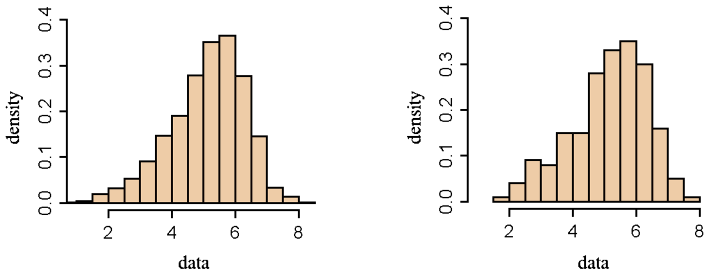
The use of the Jeffreys prior amends the bad behavior of the likelihood function, and as a result the Bayesian estimator has much better frequentist properties than the MLE for small sample sizes.

5 Illustrative example

The dataset used in this example corresponds to the size of gravels collected from a sandbar in the Mamquam River, British Columbia, Canada. This dataset is available with the HyperbolicDist package. Gravel sizes are determined by passing clasts through templates of particular sizes. This gives a range in which the size of each clast lies. Sizes (in mm) are then converted into psi units by taking the base 2 logarithm of the size. The midpoints specified are the midpoints of the psi unit ranges, and the counts give the number of observations in each size range. The classes are of length 0.5 psi units. There are 3,574 observations as described in Rice and Church (1996). Table 4 shows the maximum likelihood estimates of the model parameters obtained using the following methods implemented in *The HyperbolicDist Package*: Quasi-Newton, Nelder and Mead and Newton Raphson. Below, we use these MLE estimates based on the large sample of 3,574 observations as benchmark.

Table 4 *Maximum likelihood estimates in the model* Hyp$(\alpha, \beta, \delta, \mu)$ *using the complete Mamquam river dataset with* $n = 3{,}574$ *observations*

θ	Newton	Q-Newton	Nelder–Mead
α	5.619	5.402	5.618
β	−3.908	−3.706	−3.907
δ	2.340	2.325	2.340
μ	7.754	7.682	7.754



(a) Complete dataset. (b) Subsample of 200 observations.

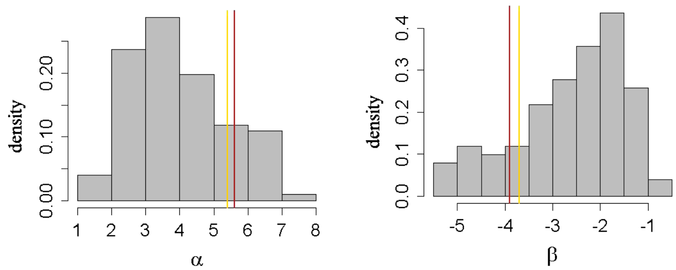
Figure 6 *Size of small stones at Mamquam river.*

It is worth pointing out that the estimates obtained via the three alternative optimization methods are very similar, mainly due to the fact that the dataset is very large. A relevant question is: what would have happened if these methods were applied to a small dataset? To answer this question, we next present some results using a subset of the original data. Figure 6 shows a histogram of a subsample of size $n = 200$ taken from the original dataset. The sampling was done by dividing the domain of the original data in subintervals and by sampling observations in each subinterval with probability equal to the relative frequency observed in each subinterval.

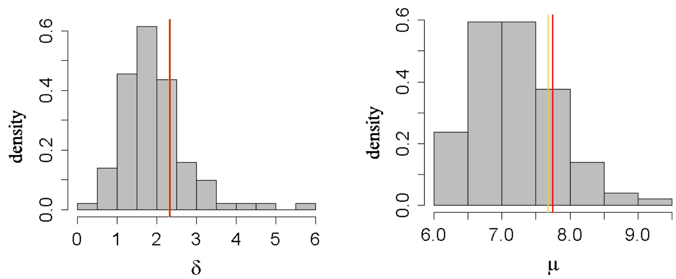
Table 5 presents summaries of the posterior distribution and maximum likelihood estimates for the model Hyp($\alpha, \beta, \delta, \mu$), based on the subsample. Note that, using the Jeffreys prior, point and interval estimates are reasonably similar to the benchmark estimates obtained using the complete dataset. Whereas using maximum likelihood approach the estimates obtained for α and β have very large absolute values. In addition, the Bayesian 95% credible intervals contain all the benchmark estimates. This can be seen in Figure 7, that shows histograms of samples from the marginal posterior distributions for the parameters of interest. On the other hand, Table 5 shows that maximum likelihood estimation breaks down when we use only the subsampled 200 observations. This example illustrates the superiority of the proposed Bayesian approach in the estimation of parameters of the hyperbolic distribution when the sample size is small.

Table 5 *Standard deviation* (SD[$\theta|y$]), *median* (MD[$\theta|y$]) *and quantiles of* 2.5% (q0.025) *and* 97.5% (q0.975) *of the posterior distribution and maximum likelihood estimates for the model* Hyp($\alpha, \beta, \delta, \mu$), *using a subsample of* 200 *observations from the Mamquam river dataset*

θ	SD[$\theta\|y$]	MD[$\theta\|y$]	q0.025	q0.975	Newton	Q-Newton	Nelder–Mead
α	1.3364	3.7802	1.9358	6.7346	31446.898	44.691	63.980
β	1.126	−2.4233	−5.0103	−1.0207	−31445.088	−42.916	−62.186
δ	0.8305	1.7831	0.8251	3.8229	0.048	1.201	1.032
μ	0.5989	7.1576	6.2404	8.5026	10.071	9.694	9.827



(a) Marginal posterior of α. (b) Marginal posterior of β.



(c) Marginal posterior of δ. (d) Marginal posterior of μ.

Figure 7 *Histograms of the marginal posterior distribution for the parameters in the model obtained using a subsample of size* 200. *The vertical lines represent the estimates obtained for the complete dataset using the Quasi-Newton, Nelder–Mead and Newton methods.*

6 Discussion

In this paper, we have developed Bayesian analysis for the hyperbolic family of distributions using the noninformative Jeffreys prior. We have shown that our proposed methodology solves the estimation problems associated with the bad behavior of the likelihood function. Moreover, a simulation study has shown that, when compared to the maximum likelihood estimator, our Bayesian estimator is superior.

Barndorff-Nielsen and Blæsild (1981) raised three relevant questions that we answer in our paper. Their first question was “For what sample size is it reasonable to consider a 4 parameter distribution?” For the hyperbolic distribution, the answer depends on which estimation method is used. Whereas the MLE needs a sample size of the order of thousands to be reliable, our Bayesian approach based on the Jeffreys prior provides reliable results with sample size as small as 50. Their second question was “Which parametrization of the distribution gives the most tractable form of the loglikelihood?” The second question suggests reparametrization in order to achieve tractability of the likelihood function. This is not an issue in our proposed Bayesian analysis because the Jeffreys prior is invariant under reparametrization. Their third question was “Which numerical procedure is optimal?” In the case of small samples, no numerical procedure will save the MLE; the problem is not the numerical optimization procedure but the bad behavior of the likelihood function. Conversely, for the Bayesian analysis we propose, the MCMC algorithm which we briefly describe in Section 4 works well even for small samples.

Acknowledgments

The work of Fonseca was supported in part by CAPES and CRISM grants. The work of Migon was supported in part by CNPq and CAPES grants. The work of Ferreira was supported in part by National Science Foundation Grant DMS-09-07064. The authors gratefully acknowledge an anonymous referee and the associate editor for their helpful comments and suggestions that improved considerably the article.

References

Abramowitz, M. and Stegun, A. S. (1972). *Handbook of Mathematical Functions*. New York: Dover Publications.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178. MR0808153

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics* **32**, 159–188. MR2188669

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particles size. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **353**, 401–419.

Barndorff-Nielsen, O. (1979). Models for non-Gaussian variation, with applications to turbulence. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **368**, 501–520. MR0554672

Barndorff-Nielsen O. and Blæsild, P. (1981). Hyperbolic distributions and ramifications: contributions to theory and application. In *Statistical Distributions in Scientific Work* (C. Taillie, G. P. Patil and B. A. Baldessari, eds.) **4**, 19–44. Dordrecht: Reidel. MR0656143

Bauer, C. (2000). Value at risk using hyperbolic distributions. *Journal of Economics and Business* **52**, 455–467.

Bingham N. H. and Kiesel, R. (2001). Modelling asset returns with hyperbolic distributions. In *Asset Return Distributions* (J. Knight and S. Satchell, eds.) **1**, 1–20. Oxford: Butterworth-Heinemann.

Blæsild, P. (1981). The two dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data. *Biometrika* **68**, 251–263. MR0614962

Blæsild, P. and Sørensen, M. (1992). Hyp—A computer program for analysing data by means of the hyperbolic distribution. Research Report 248, Dept. Theoretical Statistics, Aarhus Univ., Aarhus, Denmark.

Eberlein E. and Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli* **1**, 281–299.

Eberlein E., Keller, U. and Prause, K. (1998). New insights into smile, mispricing and value at risk: The hyperbolic model. *Journal of Business* **71**, 371–406.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38. MR1225212

Fonseca, T. C. O., Ferreira, M. A. R. and Migon, H. S. (2008). Objective Bayesian analysis for the Student-t regression model. *Biometrika* **95**, 325–333. MR2521587

Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. *Lecture Notes in Statistics* **9**. Heidelberg: Springer. MR0648107

Liseo, B. and Loperfido, N. (2006). A note on the reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* **136**, 373–389. MR2211345

Prause, K. (1999). The generalized hyperbolic model: estimation, financial derivatives, and risk measures. Ph.D. thesis, Univ. Freiburg.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rice, S. and Church, M. (1996). Sampling superficial gravels: The precision of size distribution percentile estimates. *Journal of Sedimentary Research* **66**, 654–665.

Silva, R. S., Lopes, H. F. and Migon, H. S. (2006). The extended generalized inverse Gaussian distribution for log-linear and stochastic volatility models. *Brazilian Journal of Probability and Statistics* **20**, 67–91. MR2340185

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term. *Journal of the American Statistical Association* **71**, 400–405. MR0405699

T. C. O Fonseca
H. S. Migon
Departamento de Métodos Estatísticos
Universidade Federal do Rio de Janeiro
Centro de Tecnologia
Av. Athos da Silveira Ramos
Bloco C Sala C114 D
CEP 21941-909 Rio de Janeiro
RJ
Brasil
E-mail: thais@im.ufrj.br
migon@im.ufrj.br

M. A. R. Ferreira
Department of Statistics
University of Missouri, Columbia
146 Middlebush Hall
Columbia, MO 65211-6100
USA
E-mail: ferreiram@missouri.edu