Biometrika (2008), **95**, 2, *pp*. 325–333
doi: 10.1093/biomet/asn001
Advance Access publication 25 April 2008

Printed in Great Britain

Objective Bayesian analysis for the Student-*t* regression model

BY THAÍS C. O. FONSECA

Department of Statistics, University of Warwick, Coventry CV4 7AL, U.K.
T.C.O.Fonseca@warwick.ac.uk

MARCO A. R. FERREIRA

Department of Statistics, University of Missouri, Columbia, Missouri 65211-6100, U.S.A.
ferreiram@missouri.edu

AND HELIO S. MIGON

Department of Statistics, Federal University of Rio de Janeiro, CEP: 21945-970, Brazil
migon@im.ufrj.br

SUMMARY

We develop a Bayesian analysis based on two different Jeffreys priors for the Student-*t* regression model with unknown degrees of freedom. It is typically difficult to estimate the number of degrees of freedom: improper prior distributions may lead to improper posterior distributions, whereas proper prior distributions may dominate the analysis. We show that Bayesian analysis with either of the two considered Jeffreys priors provides a proper posterior distribution. Finally, we show that Bayesian estimators based on Jeffreys analysis compare favourably to other Bayesian estimators based on priors previously proposed in the literature.

Some key words: Heavy tail behaviour; Jeffreys prior; Robustness to outliers.

1. INTRODUCTION

Many datasets exhibit heavy tail behaviour that can be well modelled with Student-*t* errors (Fernández & Steel, 1998; Lange et al., 1989; Vrontos et al., 2000; Chib et al., 2002; Jacquier et al., 2004). Moreover, the use of the Student-*t* distribution for the error component reduces the influence of outliers and thus makes the statistical analysis more robust (Maronna, 1976; West, 1984; Lange et al., 1989). The degree of robustness of the analysis is directly related to the number of degrees of freedom, ν, of the Student-*t* distribution; smaller ν implies higher robustness. Unfortunately, the estimation of ν is not straightforward: the likelihood function tends to infinity as $\nu \to 0$ and improper prior distributions may lead to improper posterior distributions (Fernández & Steel, 1999). Moreover, even when the parameter space is restricted to the region in which the likelihood is bounded, the maximum likelihood estimator may not exist with positive probability. In addition, proper priors previously proposed in the literature may strongly influence the analysis and thus are inappropriate in applications where there is no prior information on ν. As an alternative, we develop here objective Bayesian analyses based on the Jeffreys-rule prior and on the independence Jeffreys prior for the linear regression model with independent Student-*t* errors with unknown ν. This allows a nonsubjective statistical analysis with adaptive robustness to outliers and with full account of the uncertainty.

2. DIFFICULTIES WITH MAXIMUM LIKELIHOOD ESTIMATION

Consider the linear regression model in which an n-vector of observations $y = (y_1, \ldots, y_n)'$ satisfies

$$y = X\beta + \varepsilon, \tag{1}$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ is the error vector and $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed according to the Student-t distribution with location zero, scale parameter σ and ν degrees of freedom. Here $X = [x_1, \ldots, x_n]'$ is the $n \times p$ matrix of explanatory variables and is taken to be of full-rank p. We denote the model parameters by $\theta = (\beta, \sigma, \nu) \in \mathbb{R}^p \times (0, \infty)^2$.

Thus, the likelihood function is given by

$$L(\beta, \sigma, \nu; y, x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)^n \nu^{n\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)^n \pi^{n/2}\sigma^n} \prod_{i=1}^{n} \left\{ \nu + \left(\frac{y_i - x_i'\beta}{\sigma}\right)^2 \right\}^{-(\nu+1)/2}. \tag{2}$$

Maximum likelihood estimation for the Student-t regression model is very problematic because the likelihood function is ill-behaved for ν close to zero and may be ill-behaved when $\nu \to \infty$. For any given β, let $s(\beta)$ be the number of observations for which $y_i = x_i'\beta$. In addition, let $\tilde{\beta} = \text{argmax}_\beta\ s(\beta)$ and $d_0 = s(\tilde{\beta})/\{n - s(\tilde{\beta})\}$. As a particular consequence of Theorem 5 of Fernández & Steel (1999), when $\nu < d_0$ the likelihood function in (2) goes to infinity as σ goes to zero. Even when the space of variation of ν is restricted to (d_0, ∞), the likelihood function supremum may be achieved when $\nu \to \infty$.

For a fixed ν large enough, Maronna (1976) has shown that the conditional maximum likelihood estimator of (β, σ), denoted here by $(\hat{\beta}_\nu, \hat{\sigma}_\nu) = \text{argmax}_{(\beta,\sigma)} L(\beta, \sigma, \nu; y, x)$, exists and is unique. The profile likelihood function of ν is defined as

$$L^{\mathrm{P}}(\nu; y, x) = L(\hat{\beta}_\nu, \hat{\sigma}_\nu, \nu; y, x). \tag{3}$$

PROPOSITION 1. *The profile likelihood function given in* (3) *satisfies* $L^{\mathrm{P}}(\nu; y, x) = O(1)$, *as* $\nu \to \infty$, *with* $\lim_{\nu\to\infty} L^{\mathrm{P}}(\nu; y, x) = \prod_{i=1}^{n} \hat{\sigma}^{-1}\phi(\hat{z}_i)$, *where* $\phi(\cdot)$ *is the standard normal density,* $(\hat{\beta}, \hat{\sigma})$ *is the maximum likelihood estimator of* (β, σ) *under normality of the errors, and* $\hat{z}_i = (y_i - x_i'\hat{\beta})/\hat{\sigma}$, $i = 1, \ldots, n$, *are the standardized residuals under normality.*

Proof. This follows directly because $L(\beta, \sigma, \nu; y, x)$ converges to a Gaussian likelihood as $\nu \to \infty$. □

As the likelihood function converges to a positive constant as $\nu \to \infty$, the analysis is highly dependent on the rate of decay of the marginal prior density of ν. This makes truncated uniform priors such as the one used by Jacquier et al. (2004) very inappropriate, because estimates of ν and Bayes factors would be extremely sensitive to the place of truncation; see the discussion in Lavine & Wolpert (1995). In the extreme case when the marginal prior of ν is improper, as in the case of the one used by Vrontos et al. (2000), the posterior distribution may also be improper. Propriety of the posterior distributions of ν may be obtained with exponential priors for ν (Geweke, 1993; Fernández & Steel, 1999), but we show in § 4 that these priors may be too informative and have strong influence on the posterior inference. Liu (1995, p. 144) has presented the Jeffreys prior for ν for the case of independent and identically distributed multivariate t observations with known location vector and known scale matrix, and used an approximation of that as marginal prior of ν in the case of multivariate regression with Student-t errors. In contrast, we derive and use the Jeffreys-rule prior and the independence Jeffreys prior for the Student-t linear regression model.

From Proposition 1, we conclude that the estimation equation based on the profile likelihood function of ν has one root in the limit when ν goes to infinity. This may be the only root, in that it is possible that the profile likelihood function is a strictly increasing function of ν. In that case, the maximum likelihood estimator will not exist even if the parameter space is restricted to $(\beta, \sigma, \nu) \in \mathbb{R}^p \times (0, \infty) \times (d_0, \infty)$. The following theorem, proved in the Appendix, establishes a condition for the existence of the maximum likelihood estimator.

THEOREM 1. *Consider the Student-t regression model with likelihood function given in equation* (2) *and* $(\beta, \sigma, \nu) \in \mathbb{R}^p \times (0, \infty) \times (d_0, \infty)$. *If* $\sum (\hat{z}_i^2 - 1)^2 < 2n$ *then the profile likelihood function of* ν *is increasing as* $\nu \to \infty$.

We have computed the probability of an increasing profile likelihood function of ν as $\nu \to \infty$ for several values of ν and different sample sizes. The probability increases as ν increases, and decreases when the sample size increases. For example, the probability is about 0·15 when $n = 50$ and $\nu = 4$. Thus, a Bayesian analysis based on a truncated uniform prior on ν may strongly depend on where the prior is truncated.

3. JEFFREYS PRIORS

In this section, we derive the Jeffreys-rule prior and the independence Jeffreys prior for the parameters of the Student-t regression model. As will become clear in what follows, both of these priors belong to the class of improper prior distributions given by

$$\pi(\theta) \propto \frac{\pi(\nu)}{\sigma^a}, \tag{4}$$

where $a \in \mathbb{R}$ is a hyperparameter and $\pi(\nu)$ is the 'marginal' prior of ν. Any of these priors assumes the parameters are 'independent' a priori, and when $a = 1$ the 'marginal' prior of (β, σ) is the standard noninformative prior for location–scale parameters.

The Jeffreys-rule prior is given by $\pi(\theta) \propto \surd\{\det I(\theta)\}$, where $I(\theta)$ is the Fisher information matrix with (i, j) entry given by

$$\{I(\theta)\}_{ij} = E_{Y|\theta}\left[-\frac{\partial^2}{\partial\theta_i \partial\theta_j}\log\{L(\theta; y, x)\}\right], \qquad \theta_1 = \beta, \quad \theta_2 = \sigma, \quad \theta_3 = \nu,$$

where $L(\theta; y, x)$ is given by (2).

The independence Jeffreys prior is based on assuming that the marginal priors for β and (σ, ν) are independent a priori, and separately computing priors for each of these groups of parameters by applying a Jeffreys-rule prior.

THEOREM 2. *Consider the Student-t regression model given in* (1). *Then the independence Jeffreys prior and the Jeffreys-rule prior for* (β, σ, ν), *to be denoted by* $\pi^{\mathrm{I}}(\beta, \sigma, \nu)$ *and* $\pi^{\mathrm{J}}(\beta, \sigma, \nu)$, *are of the form* (4) *with, respectively,*

$$a = 1, \qquad \pi^{\mathrm{I}}(\nu) \propto \left(\frac{\nu}{\nu + 3}\right)^{1/2} \left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu + 1}{2}\right) - \frac{2(\nu + 3)}{\nu(\nu + 1)^2}\right\}^{1/2}, \tag{5}$$

and

$$a = p + 1, \qquad \pi^{\mathrm{J}}(\nu) \propto \pi^{\mathrm{I}}(\nu) \left(\frac{\nu + 1}{\nu + 3} \right)^{p/2}, \tag{6}$$

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

Proof. See the Appendix. □

Corollary 1. *The marginal independence Jeffreys prior for* ν *given in* (5) *is a continuous function in* $[0, \infty)$ *and is such that* $\pi^{\mathrm{I}}(\nu) = O(\nu^{-1/2})$ *as* $\nu \to 0$ *and* $\pi^{\mathrm{I}}(\nu) = O(\nu^{-2})$ *as* $\nu \to \infty$.

Proof. See the Appendix. □

Corollary 2. *Provided that* $n \geqslant p + 1$, (i) *the independence Jeffreys prior* (5) *and the Jeffreys-rule prior* (6) *yield proper posterior densities, and* (ii) *the marginal posteriors* $\pi^{\mathrm{I}}(\nu | y, x)$ *and* $\pi^{\mathrm{J}}(\nu | y, x)$ *do not have any positive integer moments.*

Proof. As a consequence of Corollary 1, the marginal independence Jeffreys prior $\pi^{\mathrm{I}}(\nu)$ and the marginal Jeffreys-rule prior $\pi^{\mathrm{J}}(\nu)$ are probability measures on $(0, \infty)$. Thus, part (i) follows as a consequence of Theorem 1 of Fernández & Steel (1999, p. 156). Part (ii) follows from Corollary 1 and the fact that $L(\beta, \sigma, \nu; y, x)$ converges to a Gaussian likelihood as $\nu \to \infty$. □

4. Frequentist properties

This section presents frequentist properties of several estimators of ν based on the objective priors proposed here and based on priors previously proposed in the literature. Here we focus on the frequentist mean squared error and on the frequentist coverage of 95% credible intervals for two different sample sizes, $n = 30$ and $n = 100$.

We compare the results of the analyses based on the Jeffreys-rule and the independence Jeffreys prior with the analyses based on the prior proposed by Geweke (1993), which is of the form (4) with $a = 1$ and $\pi(\nu) = \lambda \exp(-\lambda\nu)$. Geweke (1993) suggested that λ should be chosen based on the prior information about the problem at hand, and in his analysis he used λ equal to 0·05, 0·1, 0·2, 0·33 or 1·0. Note that Geweke's prior includes as a particular case the prior used by Fernández & Steel (1999) when $\lambda = 0{\cdot}1$. The mean squared error is highly dependent on λ and on the true value of ν. Here we present results for the posterior medians resulting from the use of Geweke's prior with $\lambda = 0{\cdot}1$ and $\lambda = 1{\cdot}0$.

We computed the mean squared errors and frequentist coverage by Monte Carlo simulation with $p = 5$, $x_i' = (1, x_{1i}, x_{2i}, x_{3i}, x_{4i})$, $\sigma^2 = 1{\cdot}5$ and $\beta' = (2, 1, 0{\cdot}3, 0{\cdot}9, 1)$. Moreover, we assumed a joint Gaussian distribution for $(x_{1i}, x_{2i}, x_{3i}, x_{4i})$ with marginal variances equal to 1 and with two different correlation structures, namely independence or pairwise correlation equal to 0·5. As the results for the two correlation structures are qualitatively similar, here we show just the results for independent explanatory variables.

Figure 1 shows, for $n = 30$ and $n = 100$, the square root of the relative mean squared error, $\surd\{\mathrm{MSE}(\nu)\}/\nu$, for the posterior medians from the Geweke priors, with $\lambda = 0{\cdot}1$, $1{\cdot}0$, and the Jeffreys priors. As expected, the relative mean squared error is smaller for $n = 100$ than for $n = 30$. For both $n = 30$ and $n = 100$, the performance of the posterior medians based on the Geweke priors is not good if $1/\lambda$ is far from the true value of ν. In contrast, both the independence Jeffreys prior and the Jeffreys-rule prior provide estimators with stable relative mean squared error for a broad range of ν values.

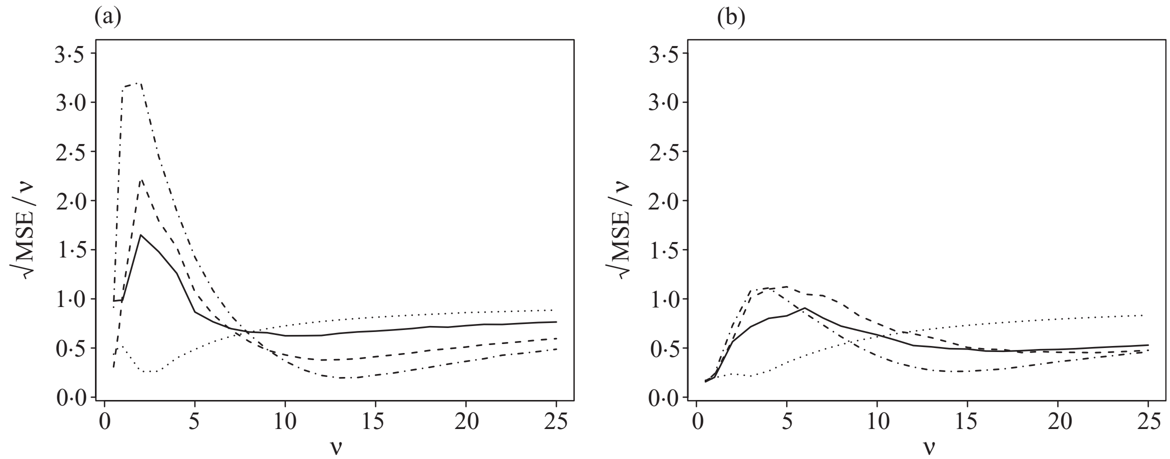

Fig. 1. Square root of relative mean squared error of estimators of ν based on Geweke priors ($\lambda = 0{\cdot}1$, dash-dotted; $\lambda = 1{\cdot}0$, dotted), based on Jeffreys-rule prior, solid, and independence Jeffreys prior, dashed, for (a) $n = 30$, (b) $n = 100$.

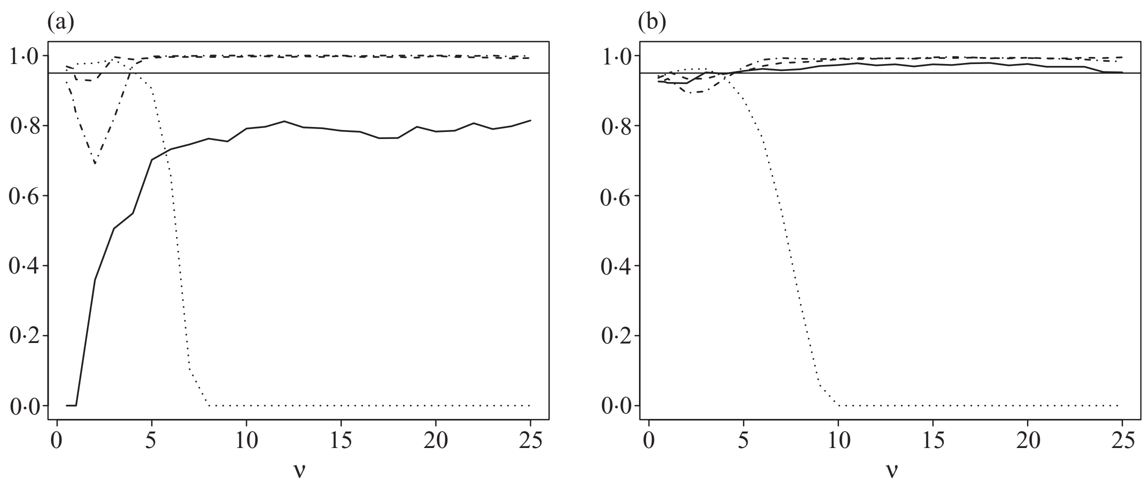

Fig. 2. Frequentist coverage of 95% credible intervals for ν based on Geweke priors ($\lambda = 0{\cdot}1$, dash-dotted; $\lambda = 1{\cdot}0$, dotted), based on Jeffreys-rule prior, solid, and independence Jeffreys prior, dashed, with nominal level, solid, for (a) $n = 30$, (b) $n = 100$.

Figure 2 shows, for $n = 30$ and $n = 100$, the frequentist coverage of 95% credible intervals for ν based on the Geweke priors, with $\lambda = 0{\cdot}1$, $1{\cdot}0$, and the Jeffreys priors. Results are fairly different depending on sample size. For $n = 30$, Jeffreys-rule-prior-based analysis performs poorly and the Geweke prior with $\lambda = 1{\cdot}0$ leads to good performance for small ν but disastrous performance otherwise. For the Geweke prior with $\lambda = 0{\cdot}1$, the frequentist coverage is much smaller than the nominal level for small ν and it is undesirably close to 1 for $\nu > 6$. For the independence Jeffreys prior, the frequentist coverage is conservative but better than the Geweke-prior-based analysis. For $n = 100$, the analysis based on the Geweke prior with $\lambda = 1{\cdot}0$ again performs very poorly for $\nu > 6$, and the analysis with $\lambda = 0{\cdot}1$ has frequentist coverage smaller than the nominal level for small ν and larger than nominal for larger ν. For $n = 100$, the Jeffreys-rule prior leads to frequentist coverage smaller than nominal for small values of ν and slightly conservative credible intervals for larger values of ν, whereas the independence Jeffreys prior leads to frequentist coverage close to nominal for small values of ν and larger than nominal for larger values of ν.

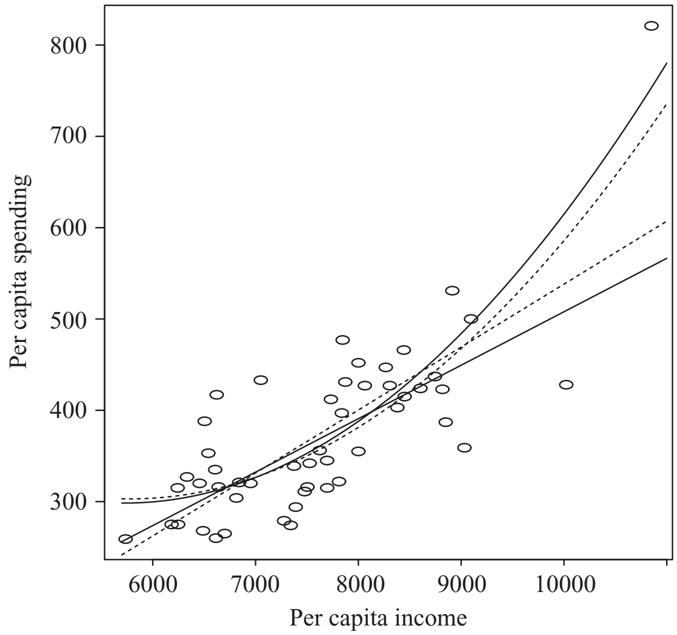

Fig. 3. School spending. Scatterplot of per capita income against per capita spending and linear and quadratic fits under Gaussian errors, dashed, and Student-t errors, solid.

We have also analysed frequentist properties of estimators of σ and β. For β, the Jeffreys-rule prior is inferior in terms of both mean squared error and coverage of intervals. In terms of mean squared error, Geweke priors provide slightly better performance than the independence Jeffreys prior in the range of values considered for ν. In terms of coverage, the three priors provide frequentist coverage close to nominal. For σ, the independence Jeffreys prior provides better results, with conclusions qualitatively similar to the conclusions above for ν.

5. Application

We illustrate our approach with the well-known dataset on per capita income and per capita spending in public schools by state in the United States in 1979. This dataset has been previously analysed by Greene (1997) and Cribari-Neto et al. (2000) using heuristic approaches to the so-called heteroscedasticity-of-unknown-form problem. As the Student-t distribution belongs to the class of scale mixtures of normals with a gamma mixing distribution (West, 1984), Student-t errors may be interpreted as Gaussian errors with varying variances and thus may be used as parametric alternatives to the heteroscedasticity-of-unknown-form problem. As in those previous analyses, we take per capita spending as the dependent variable and per capita income and its square as the regressor variables.

Figure 3 shows the scatterplot of the dependent variable y against the explanatory variable x as well as linear and quadratic fits. It is clear from the figure that there is a high-leverage outlying observation corresponding to Alaska. Standard analyses with Gaussian errors select the quadratic or the linear model as the best model depending on the inclusion or exclusion of Alaska, respectively.

Table 1 displays posterior summaries based on the independence Jeffreys prior for both linear and quadratic models. The posterior median for ν is fairly low, equal to 4·46 and 4·82 for the linear and quadratic models, respectively. Analyses for both models based on the Jeffreys prior,

Table 1. *School spending. Student-t linear and quadratic models. Posterior summaries based on the independence Jeffreys prior*

Parameter	Linear model		Quadratic model	
	Median	95% C.I.	Median	95% C.I.*
ν	4·46	(1·81, 16·53)	4·82	(1·72, 24·92)
σ	46·7	(33·7, 62·8)	46·1	(32·2, 62·7)
β_1	−75·3	(−210·8, 53·0)	899·7	(−177·5, 1596·8)
β_2	583·2	(409·0, 761·5)	−2077·0	(−3865·7, 875·8)
β_3	–	–	1789·7	(−181·8, 2909·2)

*C.I., credible interval.

Table 2. *School spending. Fractional Bayes factor of each model against the linear model with Gaussian errors*

	Gaussian errors	Student-*t* errors
Linear model	1·00	$1{\cdot}748112 \times 10^{46}$
Quadratic model	17·01	$2{\cdot}002119 \times 10^{43}$

not shown, lead to similar results for ν, and provide smaller estimates for σ, with similar results for the regression coefficients. Analyses based on Geweke's prior with $\lambda = 0{\cdot}1$, not shown, provide larger ν estimates, with the posterior median equal to 6·35 and 8·07 for the linear and quadratic models, respectively. As a consequence of the larger ν, when compared with both Jeffreys priors, the Geweke prior with $\lambda = 0{\cdot}1$ provides estimates of the regression coefficients that are closer to the estimates based on Gaussian errors. Moreover, Geweke's prior with $\lambda = 1$ leads to the opposite behaviour, with regression coefficient estimates farther away from the estimates for the Gaussian model. In summary, the analysis based on Geweke's prior is strongly dependent on the value of λ.

In order to compare different models, we use Bayes factors. The impropriety of the 'marginal' prior for β prevents the use of standard Bayes factors to compare linear and quadratic models. Several adjusted Bayes factors have been proposed, such as the intrinsic Bayes factor in its several versions (Berger & Pericchi, 1996) and the fractional Bayes factor (O'Hagan, 1995). When a Markov chain Monte Carlo scheme is available for posterior exploration, the fractional Bayes factor is much easier to implement and faster to compute than the intrinsic Bayes factor. For these reasons, here we use the fractional Bayes factor for model comparison. Table 2 shows, for robust training fraction 7/50, the fractional Bayes factors for linear and quadratic models under Gaussian and Student-*t* errors against the linear model with Gaussian errors. Our Bayesian framework clearly points to the Student-*t* linear model as the best.

6. Discussion

This model-based framework can be easily extended to more complex situations with the substitution of Gaussian by Student-*t* errors (Jacquier et al., 2004). Objective priors would then have to be derived for the specific model at hand and the derivation may be difficult. In those cases, the marginal priors for ν developed here may be a good option provided that the posterior distribution is proper.

Acknowledgement

The authors gratefully acknowledge an anonymous referee, an associate editor and Professor D. M. Titterington for their helpful comments and suggestions that considerably improved the article. This work was partially supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil.

Appendix

Technical details

First, we state some auxiliary facts.

Lemma A1. *Consider the regression model* $y \sim t(\nu, x'\beta, \sigma^2)$. *Define*

$$c(\nu, \sigma^2) = \frac{\Gamma\{(\nu+1)/2\}\nu^{\nu/2}}{\Gamma(\nu/2)\sqrt{(\pi\sigma^2)}},$$

$z = (y - x'\beta)/\sigma$, $a_k = E\{(\nu + z^2)^{-k}\}$, $b_q = E\{(y - x'\beta)^q(\nu + z^2)^{-1}\}$ *and* $c_q = E\{(y - x'\beta)^q(\nu + z^2)^{-2}\}$. *Then the following hold.*

(i) $E(z) = 0$, *for* $\nu > 1$; $E(z^2) = \nu/(\nu - 2)$, *for* $\nu > 2$; $E(z^3) = 0$, *for* $\nu > 3$; $E(z^4) = 3\nu^2/\{(\nu - 2)(\nu - 4)\}$, *for* $\nu > 4$.

(ii) $a_k = E\{(\nu + z^2)^{-k}\} = \int_{-\infty}^{\infty} c(\nu, \sigma^2)(\nu + z^2)^{-(\nu+2k+1)/2}\, dy = \{(\nu + 2k)/\nu\}^{(\nu+2k+1)/2} \times c(\nu, \sigma^2)/c\{\nu + 2k, \nu\sigma^2/(\nu + 2k)\}$. *In particular,* $a_1 = (\nu + 1)^{-1}$ *and* $a_2 = (\nu + 2)/\{\nu(\nu + 1)(\nu + 3)\}$.

(iii) $b_q = E\{(y - x'\beta)^q(\nu + z^2)^{-1}\} = \int_{-\infty}^{\infty}(y - x')^q c(\nu, \sigma^2)(\nu + z^2)^{-(\nu+3)/2}dy = a_1 E\{(y - x'\beta)^q\}$, *where* $y \sim t\{\nu + 2, x'\beta, \nu\sigma^2/(\nu + 2)\}$.

(iv) $c_q = E\{(y - x'\beta)^q(\nu + z^2)^{-2}\} = \int_{-\infty}^{\infty}(y - x'\beta)^q c(\nu, \sigma^2)(\nu + z^2)^{-(\nu+5)/2}dy = a_2 E\{(y - x'\beta)^q\}$, *where* $y \sim t\{\nu + 4, x'\beta, \nu\sigma^2/(\nu + 4)\}$.

Proof of Theorem 1. Define $\hat{z}_{\nu i} = (y_i - x_i'\hat{\beta}_\nu)/\hat{\sigma}_\nu$. From equation (3), the first derivative of the log profile likelihood function of ν is

$$\frac{d}{d\nu}\, l^{\mathrm{P}}(\nu; y, x) = \frac{n}{2}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right\} + \frac{1}{2}\sum_{i=1}^{n}\left\{-\log\left(\frac{\nu + \hat{z}_{i\nu}^2}{\nu}\right) + \frac{\hat{z}_{i\nu}^2 - 1}{\nu + \hat{z}_{i\nu}^2}\right\}.$$

We consider the behaviour of the above expression for large ν. In this case, $\hat{z}_{\nu i} \doteq \hat{z}_i = (y_i - x_i'\hat{\beta})/\hat{\sigma}$. Moreover, using Stirling's asymptotic formula, $\Gamma(a) = e^{-a}a^{a-1/2}\sqrt{(2\pi)}$ (Abramowitz & Stegun, 1986, p. 257), we obtain $\psi(a) \doteq \log(a) - (2a)^{-1}$. In addition, using the second-order Taylor expansion of $\log(a)$ around $a = 1$, we have $\log(a) \doteq (a - 1) - (a - 1)^2/2$. Thus

$$\frac{d}{d\nu}\, l^{\mathrm{P}}(\nu; y, x) \doteq \frac{1}{2}\sum_{i=1}^{n}\left\{\frac{1}{\nu(\nu+1)} - \frac{\left(\hat{z}_i^2 - 1\right)^2}{2\left(\hat{z}_i^2 + \nu\right)^2}\right\} \doteq \frac{1}{2\nu^2}\left\{n - \frac{1}{2}\sum_{i=1}^{n}\left(\hat{z}_i^2 - 1\right)^2\right\}.$$

Therefore, if $\sum_{i=1}^{n}(\hat{z}_i^2 - 1)^2 < 2n$ then $(d/d\nu)\, l^{\mathrm{P}}(\nu; y, x) > 0$ when $\nu \to \infty$ and the profile likelihood function is increasing. □

Proof of Theorem 2. Direct differentiation of the loglikelihood function and use of the auxiliary facts in Lemma A1 gives the Fisher information matrix as

$$I(\theta) = \begin{bmatrix} \frac{1}{\sigma^2}\frac{\nu+1}{\nu+3}\sum_{i=1}^{n} x_i x_i' & 0 & 0 \\ 0 & \frac{2n}{\sigma^2}\frac{\nu}{\nu+3} & -\frac{2n}{\sigma}\frac{1}{(\nu+1)(\nu+3)} \\ 0 & -\frac{2n}{\sigma}\frac{1}{(\nu+1)(\nu+3)} & \frac{n}{4}\left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right\} \end{bmatrix}.$$

Independence Jeffreys Case. In this case, $\pi^{\mathrm{I}}(\beta, \sigma, \nu) = \pi^{\mathrm{I}}(\nu, \sigma)\pi^{\mathrm{I}}(\beta)$, where each marginal prior is computed under the assumption that the other parameters are known in the sampling distribution. Thus,

$\pi^{\mathrm{I}}(\beta) \propto \surd[\det\{I(\theta)\}_{11}] \propto 1$ and

$$\pi^{\mathrm{I}}(\nu,\sigma) \propto \surd\left[\{I(\theta)\}_{22}\{I(\theta)\}_{33} - \{I(\theta)\}_{23}^2\right]$$
$$\propto \sigma^{-1}\left(\frac{\nu}{\nu+3}\right)^{1/2}\left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}^{1/2}.$$

Jeffreys-Rule Case. In this case, $\pi^{\mathrm{J}}(\beta,\sigma,\nu) \propto \surd\{\det I(\theta)\} = \surd[\{I(\theta)\}_{22}\{I(\theta)\}_{33} - \{I(\theta)\}_{23}^2] \times \surd[\det\{I(\theta)\}_{11}]$, where $\det\{I(\theta)\}_{11} = (\sigma)^{-2p}\{(\nu+1)/(\nu+3)\}^p \det(\sum_{i=1}^n x_i x_i')$. Therefore, the Jeffreys-rule prior is

$$\pi^{\mathrm{J}}(\beta,\sigma,\nu) \propto \sigma^{-p}\left(\frac{\nu+1}{\nu+3}\right)^{p/2}\pi^{\mathrm{I}}(\nu,\sigma). \qquad \square$$

Proof of Corollary 1. The continuity of $\pi^{\mathrm{I}}(\nu)$ follows directly from the continuity of the trigamma function on $(0,\infty)$. Let us consider $\{\pi^{\mathrm{I}}(\nu)\}^2$:

$$\left(\frac{\nu}{\nu+3}\right)\left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}.$$

When $\nu \to 0$, the first factor $\nu/(\nu+3)$ behaves as $O(\nu)$. Moreover, when $\nu \to 0$, $\psi'(\nu/2) = O(\nu^{-2})$ dominates the second term of the expression above. Thus, for small ν, $\{\pi^{\mathrm{I}}(\nu)\}^2 = O(\nu^{-1})$. Therefore, $\pi^{\mathrm{I}}(\nu) = O(\nu^{-1/2})$ as $\nu \to 0$. When $\nu \to \infty$, the factor $\nu/(\nu+3)$ tends to 1. Moreover, by Stirling's asymptotic formula, $\psi'(a) \simeq a^{-1} + (2a^2)^{-1}$ for large a. Thus, for large ν,

$$\{\pi^{\mathrm{I}}(\nu)\}^2 \simeq \frac{2\nu^2+6\nu+2}{\nu^2(\nu+1)^2} - \frac{2(\nu+3)}{\nu(\nu+1)^2} = \frac{2}{\nu^2(\nu+1)^2} = O(\nu^{-4}).$$

Therefore, $\pi^{\mathrm{I}}(\nu) = O(\nu^{-2})$ as $\nu \to \infty$. $\square$

References

Abramowitz, M. & Stegun, I. A. (1986). *Handbook of Mathematical Functions*. New York: Dover.

Berger, J. & Pericchi, L. (1996). The intrinsic Bayes factor for model selection and prediction. *J. Am. Statist. Assoc.* **91**, 109–22.

Chib, S., Nardari, F. & Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *J. Economet.* **108**, 281–316.

Cribari-Neto, F., Ferrari, S. L. P. & Cordeiro, G. M. (2000). Improved heteroscedasticity-consistent covariance matrix estimators. *Biometrika* **87**, 907–18.

Fernández, C. & Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *J. Am. Statist. Assoc.* **93**, 359–71.

Fernández, C. & Steel, M. F. J. (1999). Multivariate Student-*t* regression models: Pitfalls and inference. *Biometrika* **86**, 153–67.

Geweke, J. (1993). Bayesian treatment of the independent Student-*t* linear model. *J. Appl. Economet.* **8**, 519–40.

Greene, W. H. (1997). *Econometric Analysis*. Upper Saddle River, NJ: Prentice-Hall.

Jacquier, E., Polson, N. G. & Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *J. Economet.* **122**, 185–212.

Lange, K. L., Little, R. J. A. & Taylor, J. M. G. (1989). Robust statistical modeling using the *t* distribution. *J. Am. Statist. Assoc.* **84**, 881–96.

Lavine, M. & Wolpert, R. (1995). Discussion of a paper by A. O'Hagan. *J. R. Statist. Soc.* B **57**, 132.

Liu, C. H. (1995). Missing data imputation using the multivariate *t* distribution. *J. Mult. Anal.* **53**, 139–58.

Maronna, R. A. (1976). Robust *m*-estimators of multivariate location and scatter. *Ann. Statist.* **4**, 51–67.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison (with Discussion). *J. R. Statist. Soc.* B **57**, 99–138.

Vrontos, I. D., Dellaportas, P. & Politis, D. N. (2000). Full Bayesian inference for GARCH and EGARCH models. *J. Bus. Econ. Statist.* **18**, 187–98.

West, M. (1984). Outlier models and prior distributions in Bayesian linear regression. *J. R. Statist. Soc.* B **46**, 431–9.

[*Received May* 2006. *Revised November* 2007]