

Objective Bayesian Analysis for the Student- t Linear Regression

Daojiang He^{*,||}, Dongchu Sun^{†,‡,¶,**}, and Lei He[§]

Abstract. In this paper, objective Bayesian analysis for the Student- t linear regression model with unknown degrees of freedom is studied. The reference priors under all the possible group orderings for the parameters in the model are derived. The posterior propriety under each reference prior is validated by considering a larger class of priors. Simulation studies are carried out to investigate the frequentist properties of Bayesian estimators based on the reference priors. Finally, the Bayesian approach is applied to two real data sets.

Keywords: scale mixture of normals, reference prior, independent Jeffreys prior.

MSC2020 subject classifications: Primary 62F15; secondary 62J05.

1 Introduction

In traditional linear regression models, error terms are commonly assumed to follow a normal distribution. However, when the data have thicker tails than the normal distribution, the Student- t distribution represents an attractive alternative to model this behavior. Moreover, the Student- t regression model can significantly reduce the influence of outliers, leading to a more robust analysis. See, for example, West (1984) and Lange et al. (1989). In particular, the degree of freedom of the t distribution, say ν , determines the degree of robustness of analysis. Specifically, the smaller the number of ν is, the more robust the analysis tends to be. Thus, the problem of estimating the parameter ν has attracted much attention in the literature. There are some frequentist approaches to deal with this problem, see Zellner (1976), Singh (1988), Liu and Rubin (1995), Wang and Ip (2003), among others.

In this paper, we are concerned with Bayesian inferences, in which case the choice of the prior for the parameter ν becomes very challenging. When the number of degrees of freedom is considered as discrete, Jacquier et al. (2004) proposed a truncated uniform

*Department of Statistics, Anhui Normal University, Wuhu 241003, China, djheahnu@163.com

†Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68583, USA, dsun9@unl.edu

‡School of Statistics, East China Normal University, Shanghai 200062, China, dsun9@unl.edu

§Department of Statistics, Anhui Normal University, Wuhu 241003, China, lhstat@163.com

¶Corresponding author.

||The work was supported by the National Natural Science Foundation of China (Grant No. 11201005) and the Humanities and Social Sciences Foundation of Ministry of Education, China (Grant No. 17YJC910003).

**The work was supported by the Chinese 111 Project B14019, and the National Natural Science Foundation of China (Grant No. 11671146).

prior to overcome improper posteriors, which is

$$\pi(\nu) \propto 1, \quad 3 \leq \nu \leq 40.$$

However, as commented in Fonseca et al. (2008), this type of prior is not appropriate where there is no prior information on ν , in that the estimate of ν is sensitive to the chosen truncation. In order to find an objective criterion to derive the prior for discrete degrees of freedom, Villa and Walker (2014) introduced a new approach to define objective priors based on loss functions rather than probabilities.

When the number of degrees of freedom is considered as continuous, Geweke (1993) proposed an exponential prior for ν , which is of the form

$$\pi(\nu) = \lambda \exp\{-\lambda\nu\}, \quad \nu > 0.$$

The hyperparameter λ was suggested to be chosen based on the prior information about the problem at hand. However, estimating the number of degrees of freedom in this case depends strongly on the value of λ . In order to make the prior more objective, Fonseca et al. (2008) introduced two objective priors based on formal rules: the independence Jeffreys prior

$$\pi_{IJ}(\nu) \propto \left(\frac{\nu}{\nu+3}\right)^{1/2} \left\{ \psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{1/2}, \quad \nu > 0,$$

and the Jeffreys-rule prior

$$\pi_J(\nu) \propto \pi_{IJ}(\nu) \left(\frac{\nu+1}{\nu+3}\right)^{p/2}, \quad \nu > 0,$$

where $\psi(a) = \frac{d}{da}\{\log \Gamma(a)\}$ and $\psi'(a) = \frac{d}{da}\{\psi(a)\}$ are the digamma and trigamma functions, respectively. It is shown in Fonseca et al. (2008) that the posterior under the prior $\pi_{IJ}(\nu)$ is proper, while it is corrected in Vallejos and Steel (2013) that the posterior under the prior $\pi_J(\nu)$ is actually improper.

In addition to the Jeffreys priors, another kind of important objective priors are the reference priors proposed in Bernardo (1979) and Berger and Bernardo (1992). A formal definition for one block of parameters is given in Berger et al. (2009). Following this, Wang and Yang (2016) showed that there are only two type of reference priors deriving from all six one-at-a-time reference priors of (β, σ, ν) for the linear model with Student- t errors with unknown degrees of freedom ν . They came to the conclusion that the posteriors under the reference priors are all improper. In this paper, we systematically investigate the reference priors for all the possible group orderings of the parameters. By considering a large class of priors, we obtain the posterior propriety for each reference prior. As a byproduct, we draw a different conclusion with respect to the posterior propriety for the priors in Wang and Yang (2016). In addition, we compare all reference priors from frequentist coverage and mean squared error loss.

The rest of the paper is organized as follows. In Section 2, the reference priors under different group orderings for the parameters in the Student- t linear regression model

are derived. In Section 3, the posterior propriety for each reference prior is validated by considering a larger class of priors. In Section 4, the frequentist properties of Bayesian estimators of ν based on the objective priors are presented. In Section 5, the proposed Bayesian approach is applied to two real data sets. Some concluding remarks are given in Section 6.

2 The model and priors

Consider the following linear regression model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.1)$$

where $\mathbf{y} = (y_1, \dots, y_n)'$ is the $n \times 1$ vector of response variables, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]'$ is the $n \times p$ matrix of explanatory variables with full column rank, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)'$ is the error vector and ϵ_i 's are independent and identically distributed according to the Student- t distribution with location zero, scale parameter σ and degrees of freedom ν . The parameter space is $\Theta = \{(\boldsymbol{\beta}, \sigma, \nu) : \boldsymbol{\beta} \in \mathbf{R}^p, \sigma, \nu \in \mathbf{R}^+\}$.

Thus, the likelihood function of $(\boldsymbol{\beta}, \sigma, \nu)$ for model (2.1) is given by

$$L(\boldsymbol{\beta}, \sigma, \nu) = \frac{[\Gamma(\frac{\nu+1}{2})]^n \nu^{n\nu/2}}{[\Gamma(\frac{\nu}{2})]^n \pi^{n/2} \sigma^n} \prod_{i=1}^n \left\{ \nu + \left(\frac{y_i - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right)^2 \right\}^{-(\nu+1)/2}.$$

It follows from Fonseca et al. (2008) that the Fisher information matrix of the parameter $\theta = (\boldsymbol{\beta}, \sigma, \nu)$ is

$$\mathbf{I}(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} \frac{\nu+1}{\nu+3} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \frac{2n}{\sigma^2} \frac{\nu}{\nu+3} & \frac{-2n}{\sigma} \frac{1}{(\nu+1)(\nu+3)} \\ \mathbf{0} & \frac{-2n}{\sigma} \frac{1}{(\nu+1)(\nu+3)} & \frac{n}{4} \delta_1(\nu) \end{pmatrix},$$

where

$$\delta_1(\nu) = \psi' \left(\frac{\nu}{2} \right) - \psi' \left(\frac{\nu+1}{2} \right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}.$$

If we treat $\boldsymbol{\beta}$ and (σ, ν) to be independent, the so-called independent Jeffereys prior of $(\boldsymbol{\beta}, \sigma, \nu)$ is of the form,

$$\pi_{IJ}(\boldsymbol{\beta}, \sigma, \nu) \propto \frac{1}{\sigma} \left(\frac{\nu}{\nu+3} \right)^{1/2} \{\delta_2(\nu)\}^{1/2},$$

where

$$\delta_2(\nu) = \psi' \left(\frac{\nu}{2} \right) - \psi' \left(\frac{\nu+1}{2} \right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}.$$

The usual Jeffereys-rule prior of $(\boldsymbol{\beta}, \sigma, \nu)$ is of the form,

$$\pi_J(\boldsymbol{\beta}, \sigma, \nu) \propto \frac{1}{\sigma^{p+1}} \left(\frac{\nu+1}{\nu+3} \right)^{p/2} \pi_{IJ}(\boldsymbol{\beta}, \sigma, \nu).$$

Fonseca et al. (2008) showed that $\pi_J(\boldsymbol{\beta}, \sigma, \nu)$ and $\pi_{IJ}(\boldsymbol{\beta}, \sigma, \nu)$ both result in proper posterior. However, Vallejos and Steel (2013) and Fonseca et al. (2014) subsequently corrected that $\pi_J(\boldsymbol{\beta}, \sigma, \nu)$, in fact, results in improper posterior.

Recently, Wang and Yang (2016) derived all one-at-a-time reference priors for six group orderings of $(\boldsymbol{\beta}, \sigma, \nu)$. Interestingly, there are only two different one-at-a-time reference priors. For the orderings of $\{\boldsymbol{\beta}, \sigma, \nu\}$, $\{\sigma, \boldsymbol{\beta}, \nu\}$ and $\{\sigma, \nu, \boldsymbol{\beta}\}$, their reference priors are identical and of the form,

$$\pi_{R1}(\boldsymbol{\beta}, \sigma, \nu) \propto \frac{1}{\sigma} \sqrt{\delta_1(\nu)}.$$

While for the orderings of $\{\boldsymbol{\beta}, \nu, \sigma\}$, $\{\nu, \boldsymbol{\beta}, \sigma\}$ and $\{\nu, \sigma, \boldsymbol{\beta}\}$, their reference priors are all the same, given by

$$\pi_{R2}(\boldsymbol{\beta}, \sigma, \nu) \propto \frac{1}{\sigma} \sqrt{\delta_2(\nu)}.$$

In addition to the one-at-a-time group orderings, it is possible to consider other orderings for the parameters $(\boldsymbol{\beta}, \sigma, \nu)$. As Ghosh and Mukerjee (1992) stated, changing the group ordering of the parameters may yield priors with different properties. We summarize the reference priors for other orderings in the following theorem, whose proofs are straightforward and thus omitted here.

Theorem 2.1. Consider the Student-*t* linear model (2.1).

(a) The reference priors for the orderings $\{\boldsymbol{\beta}, (\sigma, \nu)\}$ and $\{(\sigma, \nu), \boldsymbol{\beta}\}$ are the same, which is just the independence Jeffreys prior $\pi_{IJ}(\boldsymbol{\beta}, \sigma, \nu)$.

(b) The reference prior for the ordering $\{\sigma, (\boldsymbol{\beta}, \nu)\}$ is given by

$$\pi_{R3}(\boldsymbol{\beta}, \sigma, \nu) \propto \frac{1}{\sigma} \left(\frac{\nu + 1}{\nu + 3} \right)^{p/2} \sqrt{\delta_1(\nu)}.$$

(c) The reference prior for the ordering $\{(\boldsymbol{\beta}, \nu), \sigma\}$ is of the form

$$\pi_{R4}(\boldsymbol{\beta}, \sigma, \nu) \propto \frac{1}{\sigma} \left(\frac{\nu + 1}{\nu + 3} \right)^{p/2} \sqrt{\delta_2(\nu)}.$$

(d) The reference prior for the ordering $\{\nu, (\boldsymbol{\beta}, \sigma)\}$ is given by

$$\pi_{R5}(\boldsymbol{\beta}, \sigma, \nu) \propto \frac{1}{\sigma^{p+1}} \sqrt{\delta_2(\nu)}.$$

(e) The reference prior for the ordering $\{(\boldsymbol{\beta}, \sigma), \nu\}$ is

$$\pi_{R6}(\boldsymbol{\beta}, \sigma, \nu) \propto \frac{1}{\sigma^{p+1}} \sqrt{\delta_1(\nu)}.$$

Clearly, all of the reference priors aforementioned can be written in a unified form

$$\pi_R(\boldsymbol{\beta}, \sigma, \nu) \propto \frac{1}{\sigma^\alpha} \pi(\nu),$$

for some constant $\alpha > 0$ and some function of $\pi(\nu)$.

3 Propriety of posterior distributions

In this section, we investigate the important problem whether the posterior distributions are proper under the reference priors obtained in Section 2. To facilitate the proofs, we consider priors for the parameters $(\boldsymbol{\beta}, \sigma^2, \nu)$ instead of $(\boldsymbol{\beta}, \sigma, \nu)$. The prior for $(\boldsymbol{\beta}, \sigma^2, \nu)$ is of the form

$$\pi(\boldsymbol{\beta}, \sigma^2, \nu) \propto \frac{1}{(\sigma^2)^a} \pi(\nu), \quad (3.1)$$

where $a > 0$ is constant, and $\pi(\nu)$ is the ‘marginal’ prior of ν .

The following theorem gives a detailed characterization for the order of the marginal posterior of ν , which plays an important role in obtaining the posterior propriety.

Lemma 3.1. *For the Student- t linear model (2.1) and the prior in (3.1), assume that $n \geq p + 1$ and $0 < a \leq 1$. Let $f(\cdot | \nu)$ denote the density function of the Gamma distribution $Ga(\nu/2, \nu/2)$. The marginal likelihood of ν is*

$$g(\nu) = n! \int_{0 < \lambda_{(1)} < \dots < \lambda_{(n)} < \infty} \left(\prod_{i=1}^{n-p} \lambda_{(i)}^{\frac{1}{2}} \right) \lambda_{(n-p)}^{\frac{-n+2a-p-2}{2}} \prod_{i=1}^n f(\lambda_{(i)} | \nu) d\lambda_{(i)},$$

(a) If $a = 1$, then

$$g(\nu) = \begin{cases} O(\nu^{n-p-1}), & \text{as } \nu \rightarrow 0, \\ O(1), & \text{as } \nu \rightarrow \infty. \end{cases}$$

(b) If $0 < a < 1$, then

$$g(\nu) = \begin{cases} O(\nu^{n+a-1}), & \text{as } \nu \rightarrow 0, \\ O(1), & \text{as } \nu \rightarrow \infty. \end{cases}$$

In order to prove Lemma 3.1, we first present a lemma, which can be found in Fernández and Steel (1999).

Lemma 3.2. *Let $a, b, \lambda > 0$, then*

$$\frac{\lambda^a}{a} \exp\{-b\lambda\} \leq \int_0^\lambda x^{a-1} \exp\{-bx\} dx \leq \frac{\lambda^a}{a}. \quad (3.2)$$

Proof of Lemma 3.1. Similar to Fernández and Steel (1999), the lower and upper bounds of $f_Y(y)$ are both proportional to

$$\int_{(0, \infty)} \int_{(0, \infty)^n} \left(\prod_{i \neq m_1, \dots, m_p} \lambda_i^{\frac{1}{2}} \right) \lambda_{m_{p+1}}^{\frac{-n+2a-p-2}{2}} \left(\prod_{i=1}^n f(\lambda_i | \nu) d\lambda_i \right) \pi(\nu) d\nu,$$

where

$$\prod_{i=1}^p \lambda_{m_i} \equiv \max \left\{ \prod_{i=1}^p \lambda_{l_i} : \det(x_{l_1} \dots x_{l_p}) \neq 0, l_1, \dots, l_p \in \{1, \dots, n\} \right\},$$

$$\prod_{i=1}^{p+1} \lambda_{m_i} \equiv \max \left\{ \prod_{i=1}^{p+1} \lambda_{l_i} : \det \begin{pmatrix} x_{l_1} & \cdots & x_{l_{p+1}} \\ y_{l_1} & \cdots & y_{l_{p+1}} \end{pmatrix} \neq 0, l_1, \dots, l_{p+1} \in \{1, \dots, n\} \right\},$$

and $\lambda_1, \dots, \lambda_n$ are independent and identically distributed as the Gamma distribution $Ga(\nu/2, \nu/2)$, whose density function is $f(\cdot | \nu)$.

It can be shown that, up to a set of measure zero,

$$\{\lambda_{m_1}, \dots, \lambda_{m_p}\} = \{\lambda_{(n)}, \dots, \lambda_{(n-p+1)}\}, \quad \lambda_{m_{p+1}} = \lambda_{(n-p)},$$

where $\lambda_{(1)} < \lambda_{(2)} < \dots < \lambda_{(n)}$ are the ordered statistics of $\lambda_1, \lambda_2, \dots, \lambda_n$. Denote

$$g(\nu) = \int_{(0, \infty)^n} \left(\prod_{i \neq m_1, \dots, m_p} \lambda_i^{\frac{1}{2}} \right) \lambda_{m_{p+1}}^{-\frac{n+2a-p-2}{2}} \prod_{i=1}^n f(\lambda_i | \nu) d\lambda_i, \quad (3.3)$$

then we have

$$g(\nu) = n! \int_{0 < \lambda_{(1)} < \dots < \lambda_{(n)} < \infty} \left(\prod_{i=1}^{n-p} \lambda_{(i)}^{\frac{1}{2}} \right) \lambda_{(n-p)}^{-\frac{n+2a-p-2}{2}} \prod_{i=1}^n f(\lambda_{(i)} | \nu) d\lambda_{(i)}. \quad (3.4)$$

First, we derive an upper bound for the integral in (3.4). Using the inequality of (3.2), when integrating with respect to $\lambda_{(1)}$, we have

$$\int_0^{\lambda_{(2)}} \lambda_{(1)}^{\frac{1}{2}} \lambda_{(1)}^{\frac{\nu}{2}-1} \exp \left\{ -\frac{\nu}{2} \lambda_{(1)} \right\} d\lambda_{(1)} \leq \frac{2}{\nu+1} \lambda_{(2)}^{\frac{\nu+1}{2}} \propto (\nu+1)^{-1} \lambda_{(2)}^{\frac{\nu+1}{2}}.$$

From this, when integrating with respect to $\lambda_{(2)}$, the upper bound is proportional to

$$\int_0^{\lambda_{(3)}} \lambda_{(2)}^{\frac{1}{2}} \lambda_{(2)}^{\frac{\nu}{2}-1} \exp \left\{ -\frac{\nu}{2} \lambda_{(2)} \right\} (\nu+1)^{-1} \lambda_{(2)}^{\frac{\nu+1}{2}} d\lambda_{(2)} \leq (\nu+1)^{-2} \lambda_{(3)}^{\nu+1}.$$

Inductively, after integrating with respect to $\lambda_{(3)}, \dots, \lambda_{(n-p-1)}$, the upper bound is proportional to

$$(\nu+1)^{-(n-p-1)} \lambda_{(n-p)}^{\frac{n-p-1}{2}(\nu+1)}.$$

Thus, when integrating with respect to $\lambda_{(n-p)}$, the upper bound is proportional to

$$\begin{aligned} & \int_0^{\lambda_{(n-p+1)}} \lambda_{(n-p)}^{\frac{1}{2} - \frac{n+2a-p-2}{2}} \lambda_{(n-p)}^{\frac{\nu}{2}-1} \exp \left\{ -\frac{\nu}{2} \lambda_{(n-p)} \right\} (\nu+1)^{-(n-p-1)} \lambda_{(n-p)}^{\frac{n-p-1}{2}(\nu+1)} d\lambda_{(n-p)} \\ &= \int_0^{\lambda_{(n-p+1)}} (\nu+1)^{-(n-p-1)} \lambda_{(n-p)}^{\frac{(n-p)\nu}{2} + 1 - a - 1} \exp \left\{ -\frac{\nu}{2} \lambda_{(n-p)} \right\} d\lambda_{(n-p)}. \end{aligned} \quad (3.5)$$

Note that $n \geq p+1$ and $0 < a \leq 1$, hence $\frac{(n-p)\nu}{2} + 1 - a > 0$. It follows that the upper bound for the integral in (3.5) is

$$(\nu+1)^{-(n-p-1)} \left[\frac{(n-p)\nu}{2} + 1 - a \right]^{-1} \lambda_{(n-p+1)}^{\frac{(n-p)\nu}{2} + 1 - a}.$$

Inductively, after integrating with respect to $\lambda_{(n-p+1)}, \dots, \lambda_{(n-1)}$, the upper bound is proportional to

$$(\nu + 1)^{-(n-p-1)} \lambda_{(n)}^{\frac{(n-1)\nu}{2} + 1 - a} \prod_{j=0}^{p-1} \left[\frac{(n-p+j)\nu}{2} + 1 - a \right]^{-1}.$$

Thus, when integrating with respect to $\lambda_{(n)}$ on $(0, \infty)$ in the end, the upper bound is proportional to

$$(\nu + 1)^{-(n-p-1)} \left(\frac{\nu}{2}\right)^{-\frac{n\nu}{2} - 1 + a} \Gamma\left(\frac{n\nu}{2} + 1 - a\right) \prod_{j=0}^{p-1} \left[\frac{(n-p+j)\nu}{2} + 1 - a \right]^{-1}. \quad (3.6)$$

Therefore, the upper bound of $g(\nu)$ is proportional to (3.6) multiplied by the factor $\left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}} / [\Gamma(\frac{\nu}{2})]^n$, which is

$$g_u(\nu) \equiv (\nu + 1)^{-(n-p-1)} \left(\frac{\nu}{2}\right)^{a-1} \frac{\Gamma\left(\frac{n\nu}{2} + 1 - a\right)}{[\Gamma\left(\frac{\nu}{2}\right)]^n} \prod_{j=0}^{p-1} \left[\frac{(n-p+j)\nu}{2} + 1 - a \right]^{-1}. \quad (3.7)$$

Note that $\Gamma(a) \sim a^{-1}$, as $a \rightarrow 0$. Therefore, if $a = 1$, then

$$g_u(\nu) = O(\nu^{n-p-1}), \quad (\nu \rightarrow 0);$$

if $0 < a < 1$, then

$$g_u(\nu) = O(\nu^{n-1+a}), \quad (\nu \rightarrow 0).$$

Second, we find an lower bound for the integral in (3.4). By the inequality on the left side of (3.2), when integrating with respect to $\lambda_{(1)}$, there is

$$\begin{aligned} \int_0^{\lambda_{(2)}} \lambda_{(1)}^{\frac{1}{2}} \lambda_{(1)}^{\frac{\nu}{2} - 1} \exp\left\{-\frac{\nu}{2}\lambda_{(1)}\right\} d\lambda_{(1)} &\geq \left(\frac{\nu+1}{2}\right)^{-1} \lambda_{(2)}^{\frac{\nu+1}{2}} \exp\left\{-\frac{\nu}{2}\lambda_{(2)}\right\} \\ &\propto (\nu+1)^{-1} \lambda_{(2)}^{\frac{\nu+1}{2}} \exp\left\{-\frac{\nu}{2}\lambda_{(2)}\right\}. \end{aligned}$$

Consequently, when integrating with respect to $\lambda_{(2)}$, the lower bound is proportional to

$$\int_0^{\lambda_{(3)}} \lambda_{(2)}^{\frac{1}{2}} \lambda_{(2)}^{\frac{\nu}{2} - 1} \exp\left\{-\frac{\nu}{2}\lambda_{(2)}\right\} (\nu+1)^{-1} \lambda_{(2)}^{\frac{\nu+1}{2}} d\lambda_{(2)} \geq (\nu+1)^{-2} \lambda_{(3)}^{\nu+1} \exp\{-\nu\lambda_{(3)}\}.$$

The subsequent derivations are similar to that for the upper bound $g_u(\nu)$, then we can obtain a lower bound for $g(\nu)$, which is

$$g_l(\nu) \equiv (\nu + 1)^{-(n-p-1)} \frac{\left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}}}{\left(\frac{n\nu}{2}\right)^{\frac{n\nu}{2} + 1 - a}} \frac{\Gamma\left(\frac{n\nu}{2} + 1 - a\right)}{[\Gamma\left(\frac{\nu}{2}\right)]^n} \prod_{j=0}^{p-1} \left[\frac{(n-p+j)\nu}{2} + 1 - a \right]^{-1}.$$

If $a = 1$, then

$$g_l(\nu) = O(\nu^{n-p-1}), \quad (\nu \rightarrow 0);$$

if $0 < a < 1$, then

$$g_l(\nu) = O(\nu^{n-1+a}), \quad (\nu \rightarrow 0).$$

Till now, we have obtained the order of $g(\nu)$ as $\nu \rightarrow 0$.

Finally, we prove the order of $g(\nu)$ as $\nu \rightarrow +\infty$. According to the upper bound of (3.7), it is readily that $g(\nu)$ is finite on $(0, +\infty)$ for $0 < a \leq 1$. Moreover, it follows from (3.4) that

$$\begin{aligned} g(\nu) &\leq n! \int_{(0, +\infty)^n} \left(\prod_{i=1}^{n-p-1} \lambda_i^{\frac{1}{2}} \right) \lambda_{n-p}^{-\frac{n+2a-p-3}{2}} \prod_{i=1}^n f(\lambda_i | \nu) d\lambda_i \\ &\propto \mathbf{E} \left(\prod_{i=1}^{n-p-1} \lambda_i^{\frac{1}{2}} \right) \mathbf{E} \left(\lambda_{n-p}^{-\frac{n+2a-p-3}{2}} \right) \\ &\leq \mathbf{E} \left(\lambda_1^{\frac{1}{2}(n-p-1)} \right) \mathbf{E} \left(\lambda_{n-p}^{-\frac{n+2a-p-3}{2}} \right) \\ &\equiv g^*(\nu), \end{aligned}$$

where $\lambda_1, \dots, \lambda_n$ are independent and identically distributed as $\text{Ga}(\nu/2, \nu/2)$. When $\nu > \max\{n+2a-p-3, 0\}$, straightforward computation yields that

$$g^*(\nu) = \frac{1}{(\frac{\nu}{2})^{1-a} [\Gamma(\frac{\nu}{2})]^2} \Gamma\left(\frac{n-p-1+\nu}{2}\right) \Gamma\left(\frac{\nu-n-2a+p+3}{2}\right).$$

Using the result that $\Gamma(t-1) \sim \sqrt{2\pi t} (\frac{t}{e})^t$ as $t \rightarrow +\infty$, it can be shown that the limit $\lim_{\nu \rightarrow +\infty} g^*(\nu)$ exists and is positive. Consequently, we have

$$g(\nu) = O(1), \quad (\nu \rightarrow +\infty).$$

Thus, the proof of Lemma 3.1 is completed. \square

By Lemma 3.1, we can obtain the following result.

Theorem 3.1. *For the Student- t linear model (2.1), assume that $n > p+1$, then the posterior distributions of $(\boldsymbol{\beta}, \sigma, \nu)$, under the reference priors π_{R1} , π_{R2} , π_{R3} , and π_{R4} , are all proper.*

Proof. It suffices to show the propriety of the posterior distribution under the prior π_{R4} in that the proofs for the other cases are similar. The prior π_{R4} for $(\boldsymbol{\beta}, \sigma, \nu)$ is equivalent to the following prior for $(\boldsymbol{\beta}, \sigma^2, \nu)$

$$\pi(\boldsymbol{\beta}, \sigma^2, \nu) \propto \frac{1}{\sigma^2} \pi(\nu), \quad (3.8)$$

where

$$\pi(\nu) = \left(\frac{\nu+1}{\nu+3} \right)^{p/2} \{\delta_2(\nu)\}^{1/2}.$$

Under the prior (3.8), it follows from Lemma 3.1 that $g(\nu) = O(\nu^{n-p-1})$, as $\nu \rightarrow 0$ and $g(\nu) = O(1)$ as $\nu \rightarrow +\infty$. On the other hand, it can be shown that $\pi(\nu) = O(\nu^{-1})$,

as $\nu \rightarrow 0$ and $\pi(\nu) = O(\nu^{-2})$ as $\nu \rightarrow +\infty$. Therefore, when $n > p + 1$, we have $\int_0^{+\infty} g(\nu)\pi(\nu)d\nu < +\infty$. This completes the proof of Theorem 3.1. \square

Remark 3.1. For the special case that $n = p + 1$, the proof of Theorem 3.1 shows that π_{R1} , π_{R2} , π_{R3} , and π_{R4} yield improper posteriors.

In order to study the propriety of the posterior under the priors π_{R5} and π_{R6} , we give the following theorem, whose proof is analogous to Theorem 1 in Vallejos and Steel (2013).

Theorem 3.2. For the Student- t linear model (2.1) and the prior in (3.1), assume that $n \geq p + 1$ and $a > 1$, then a necessary condition for posterior propriety is $\pi(\nu) = 0$ a.e. on $(0, (2a - 2)/(n - p)]$.

According to Theorem 3.2, we have the following result.

Theorem 3.3. For the Student- t linear model (2.1), assume that $n \geq p + 1$, then under the reference priors π_{R5} and π_{R6} , the posterior distributions are both improper.

4 Frequentist properties

We now investigate the frequentist performance of Bayesian estimators of ν based on those reference priors in Section 3, as well as the discrete prior (say, π_{VW}) proposed in Villa and Walker (2014). Two simulation studies are carried out here.

4.1 Independent and identically distributed Student- t sample

In the first simulation study, as in Villa and Walker (2014), we draw an independent and identically distributed sample from a standard Student- t distribution with location parameter $\mu = 0$ and scale parameter $\sigma = 1$. Under this case, the available reference priors are only π_{R1} , π_{R2} and π_{IJ} . Two different sample sizes are considered, which are $n = 30$ and $n = 100$. The simulations are performed for $\nu = 1, 2, \dots, 25$, where the number of replications is taken as 5000. The frequentist mean squared error and the frequentist coverage probability of 95% credible interval are included to compare the estimators of ν based on different priors.

The square root of the relative mean squared error of ν , $\sqrt{\text{MSE}(\nu)}/\nu$, for the posterior medians from the four priors when $n = 30$ and $n = 100$ are shown in Figures 1 (a) and 1 (b), respectively. For the case of $n = 30$, when the true value of ν is not greater than 13, the performance of π_{R1} , π_{R2} and π_{IJ} outperforms that of π_{VW} in an obvious way. In addition, the prior π_{R2} behaves slightly better than the others at this time. While as the value of ν gets large, the relative mean squared errors based on the prior π_{VW} become smallest. For the case of $n = 100$, the situation is analogous.

Figure 2 shows the frequentist coverage probabilities of 95% credible intervals for ν . In terms of coverage, the priors π_{R1} , π_{R2} and π_{IJ} perform relatively robust, the coverage probabilities are close to the nominal level 0.95 whether $n = 30$ or $n = 100$. While for

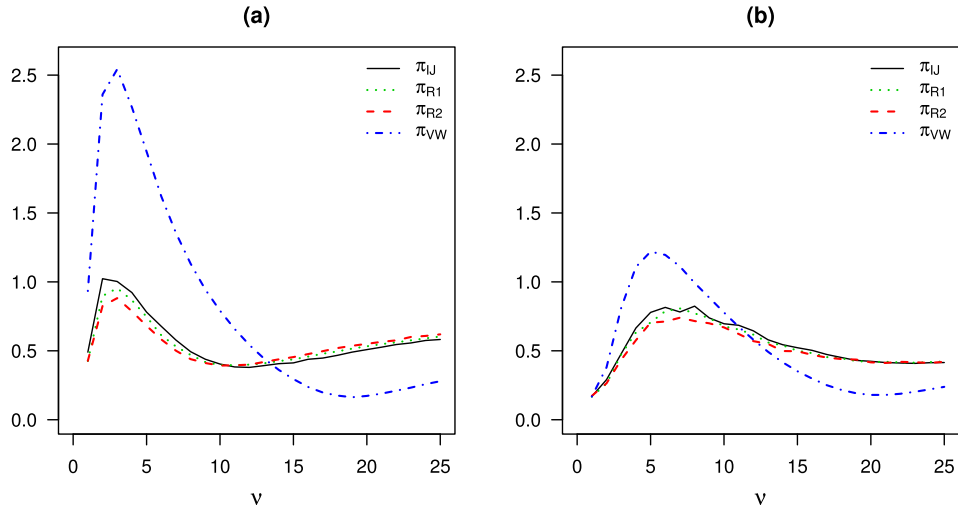


Figure 1: Square root of relative mean squared error of estimators of ν based on the independent Jeffreys prior π_{IJ} (solid), the reference prior π_{R1} (dotted), the reference prior π_{R2} (dashed), and Villa and Walker's prior π_{VW} (dash-dotted). Panel (a) is for $n = 30$, panel (b) is for $n = 100$.

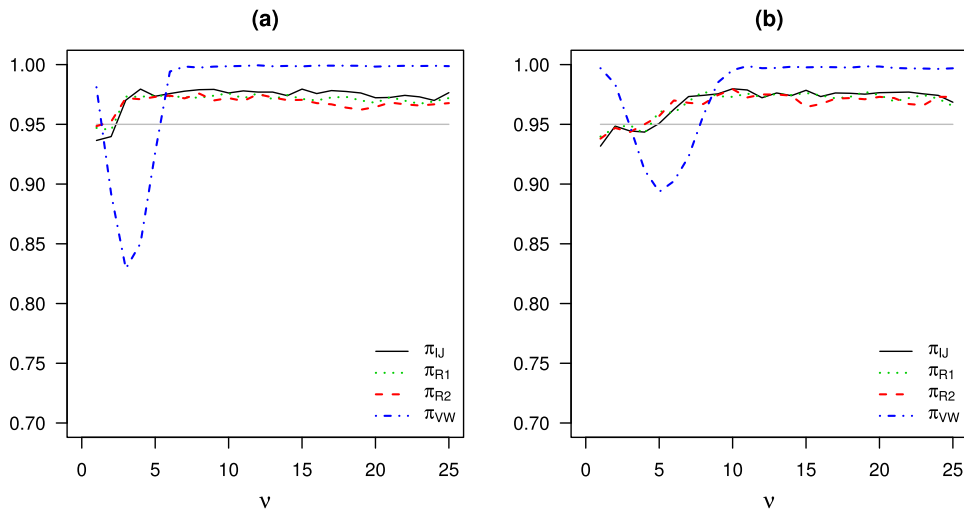


Figure 2: The frequentist coverage of the 95% credible intervals for ν based on the independent Jeffreys prior π_{IJ} (solid), the reference prior π_{R1} (dotted), the reference prior π_{R2} (dashed) and the Villa and Walker's prior π_{VW} (dash-dotted). Panel (a) is for $n = 30$, panel (b) is for $n = 100$.

the prior π_{VW} , there is a drop for small values of ν . In addition, as ν gets larger, the coverage probabilities under π_{VW} are undesirably approaching 1.

4.2 Student- t regression model

In the second simulation study, following Fonseca et al. (2008), we consider samples from a Student- t regression model in (2.1). The configuration of simulation is set as follows:

$$p = 5, \quad \mathbf{x}_i = (1, x_{i1}, x_{i2}, x_{i3}, x_{i4})', \quad i = 1, \dots, n, \quad \boldsymbol{\beta} = (2, 1, 0.3, 0.9, 1)', \quad \sigma^2 = 1.5.$$

Also, it is assumed that x_{ij} 's are independent and identically distributed as standard normal.

We employ the Metropolis–Hasting (M–H) within Gibbs sampling to generate samples of posterior distributions. Note that t distribution can be considered as a scale mixture of normals, the full conditional distributions for Gibbs sampling are given by

$$\boldsymbol{\beta} \mid \boldsymbol{\Lambda}, \sigma^2, \nu, \mathbf{y} \sim N(\hat{\boldsymbol{\beta}}, \sigma^2(\mathbf{X}'\boldsymbol{\Lambda}\mathbf{X})^{-1}), \quad (4.1)$$

$$\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\Lambda}, \mathbf{y} \sim IG\left(\frac{n}{2}, \frac{s}{2}\right), \quad (4.2)$$

$$\lambda_i \mid \boldsymbol{\beta}, \sigma^2, \nu, \mathbf{y} \sim Ga\left(\frac{\nu+1}{2}, \frac{d_i+\nu}{2}\right), \quad i = 1, \dots, n, \quad (4.3)$$

$$\pi(\nu \mid \lambda_1, \dots, \lambda_n) \propto \frac{\left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}}}{[\Gamma(\frac{\nu}{2})]^n} \left(\prod_{i=1}^n \lambda_i^{\frac{\nu}{2}-1}\right) \exp\left(-\frac{\nu}{2} \sum_{i=1}^n \lambda_i\right) \pi(\nu), \quad (4.4)$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Lambda}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Lambda}\mathbf{y}$, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, $s = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Lambda}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, $d_i = (y_i - \mathbf{x}_i'\boldsymbol{\beta})^2/\sigma^2$, and $IG(a, b)$ stands for the inverse Gamma distribution with the shape parameter a and the scale parameter b .

While for the conditional distribution (4.4), the M–H algorithm is used to generate random numbers. Following the idea of Kang et al. (2018), the proposal distribution is taken as the truncated normal distribution $N_{(0 < \nu < 300)}(\mu_\nu, \tau_\nu^2)$, where $\mu_\nu = x - q'(x)/q''(x)$, $\tau_\nu^2 = -1/q''(x)$, and $q(v) = \log(\pi(\nu \mid \lambda_1, \dots, \lambda_n))$. Specifically, the sampling algorithm is as follows:

- Generate a candidate parameter value ν_{cand} from $N_{(0 < \nu < 300)}(\mu_\nu, \tau_\nu^2)$, where $\mu_\nu = v_j - q'(v_j)/q''(v_j)$, $\tau_\nu^2 = -1/q''(v_j)$;
- Generate u from the uniform distribution on $(0, 1)$, and let $v_{j+1} = \nu_{\text{cand}} \cdot \mathbf{1}(\alpha > u) + v_j \cdot \mathbf{1}(\alpha \leq u)$, where

$$\alpha = \min\left(1, \exp\left\{\frac{q(\nu_{\text{cand}}) \log(p(v_j))}{q(v_j) \log(p(\nu_{\text{cand}}))}\right\}\right),$$

and $p(\cdot)$ is the density function of $N_{(0 < \nu < 300)}(\mu_\nu, \tau_\nu^2)$. That is to say, the candidate value ν_{cand} is accepted with probability α .

In the simulation, the size for Gibbs sampling is 5000 after 1000 burn-in samples. The square root of the relative mean squared errors of ν for the posterior medians based on the priors π_{IJ} , π_{R1} , π_{R2} , π_{R3} and π_{VW} are shown in Figure 3. The simulation results

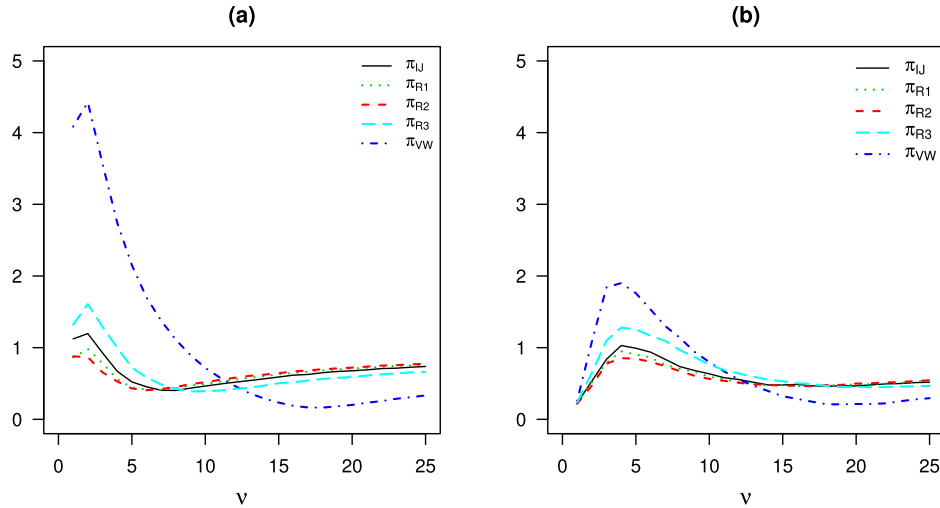


Figure 3: Square root of relative mean squared error of estimators of ν under Student- t regression based on the independent Jeffreys prior π_{IJ} (solid), the reference priors π_{R1} (dotted), π_{R2} (dashed), π_{R3} (long-dashed), and the Villa and Walker's prior π_{VW} (dash-dotted). Panel (a) is for $n = 30$, panel (b) is for $n = 100$.

for the reference prior π_{R4} are similar to π_{R3} , which are not reported here. For $n = 30$ or $n = 100$, generally speaking, the performance of π_{IJ} , π_{R1} , π_{R2} and π_{R3} is better than π_{VW} when the true values of ν are not greater than 11. Furthermore, the prior π_{R2} behaves best under this case. While for large values of ν , the behavior of π_{VW} shows its superiority.

Figures 4 (a) and (b) show the frequentist coverage probabilities of 95% credible intervals for ν based on the five priors for $n = 30$ and $n = 100$, respectively. It can be seen that the performance of the priors π_{IJ} , π_{R1} , π_{R2} and π_{R3} is similar to each other, and the corresponding coverage probabilities are relatively close to the nominal level. While the behavior of π_{VW} in this circumstance is analogous to that in the first simulation study. That is, there is a drop for small values of ν , and the coverage probabilities are undesirably close to 1 for large values of ν .

Based on the two simulation studies, in the light of the relative mean squared error, it can be seen that the prior π_{R2} behaves best for small values of ν , while the prior π_{VW} outperforms the others for large values of ν . Therefore, it is suggested to use the reference priors if there is evidence that the model is a Student- t .

5 Real data analysis

To show how the objective priors discussed above work in practice, we consider two data sets from financial markets. The first data set relates to the U.S. Treasury bond prices.

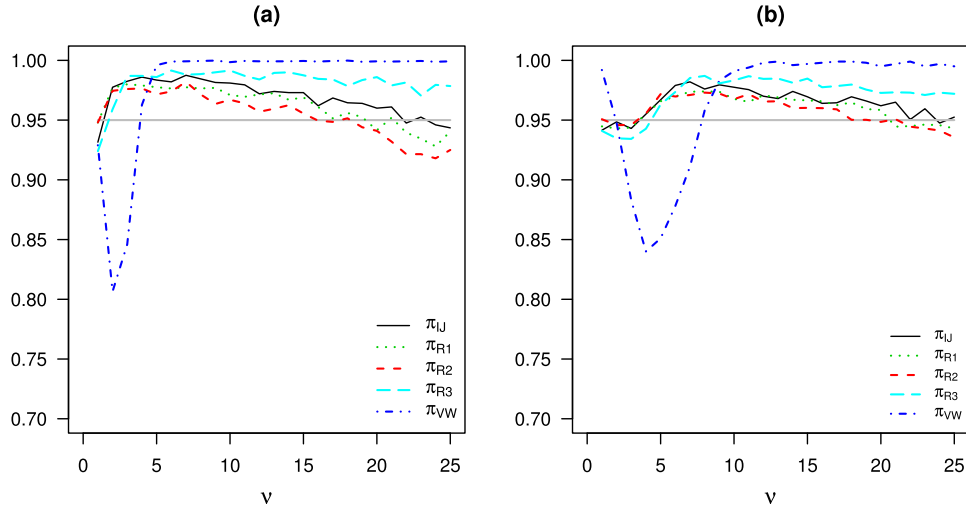


Figure 4: The frequentist coverage of the 95% credible intervals for ν under Student- t regression based on the independent Jeffreys prior π_{IJ} (solid), the reference priors π_{R1} (dotted), π_{R2} (dashed), π_{R3} (long-dashed), and the Villa and Walker's prior π_{VW} (dash-dotted). Panel (a) is for $n = 30$, panel (b) is for $n = 100$.

The second data set is a sample of the daily closing prices of the Brazilian IBOVESPA, which is an index of about 50 stocks traded on the São Paulo Stock, Mercantile & Futures Exchange.

5.1 Application to the US treasury bond prices

The first data set is from Siegel (1977), which has been previously analyzed by Sheather (2009) and Yang and Yuan (2017). As in those previous analyses, we take the bid price as the dependent variable and the coupon rate as the regressor variable. It is shown in Yang and Yuan (2017) that the normal model does not fit the data well and some heavy-tailed distributions may be appropriate. Yang and Yuan (2017) then analysed the data by using scale mixtures of normal regression models including the Student- t regression model. However, it should be pointed out that the degrees of freedom in their analysis are assumed to be known as 1.

Now, if we set $\nu = 1$ in the Student- t regression model (2.1), then it can be seen that the independent Jeffreys prior and the corresponding reference priors in π_{R1} , π_{R2} , π_{R3} and π_{R4} for (β, σ) are the same, which is

$$\pi^*(\beta, \sigma) \propto \frac{1}{\sigma}. \quad (5.1)$$

Using the prior (5.1), we can obtain that a Bayesian estimator (i.e., the posterior median) of σ^2 is 0.2650.

Prior	Parameters	Median	95% Credible Interval
π_{R2}	ν	0.7940	[0.4501, 1.3842]
	σ^2	0.2049	[0.0455, 0.6643]
	β_1	58.0622	[55.3476, 60.8905]
	β_2	4.7778	[4.4833, 5.0510]
π_{IJ}	ν	0.8254	[0.4686, 1.4828]
	σ^2	0.2124	[0.0530, 0.6906]
	β_1	58.1029	[55.3294, 60.9247]
	β_2	4.7738	[4.4800, 5.0552]
π_{VW}	ν	1	[1, 1]
	σ^2	0.2785	[0.1003, 0.7583]
	β_1	58.1969	[55.2476, 61.1049]
	β_2	4.7621	[4.4628, 5.0609]

Table 1: Posterior summaries under the priors π_{R2} , π_{IJ} and π_{VW} for the real data in Section 5.1.

Next, we analyze the data by using the model (2.1) with unknown degrees of freedom. Table 1 displays posterior summaries based on the reference priors π_{R2} and π_{IJ} as well as the prior π_{VW} . It can be seen that the priors π_{R2} and π_{IJ} lead to similar results for the four parameters, and the posterior medians of ν , 0.7940 and 0.8254, are close to 1 assumed in Yang and Yuan (2017). In addition, the Bayesian estimators of σ^2 under π_{R2} and π_{IJ} are 0.2049 and 0.2124, respectively. As can be expected, the estimators of σ^2 are smaller than that under the case where the degrees of freedom are specified as 1. While for the discrete prior π_{VW} , the Bayesian estimator of ν is 1, and the corresponding 95% credible interval is only a single point 1. The estimator of σ^2 under π_{VW} is 0.2785, which is the biggest among the three priors.

5.2 Application to the Brazilian IBOVESPA

To further compare the performance of the objective priors on real data when the “true” degrees of freedom may be slightly larger, in this subsection, we consider the daily closing prices of the IBOVESPA in Brazilian stock market. Specifically, the data set contains 100 observations from April 20 to September 13, 2001, which is part of a wider sample used in Abanto-Vallea et al. (2012). Throughout, we work with the daily returns as a percentage, $y_d = 100 \times (\log(P_d) - \log(P_{d-1}))$, where P_d is the closing price on day d . For the transformed data, a direct calculation yields that the kurtosis is much larger than 3. Thus, we use the following t -regression model to fit the data:

$$y_d = \mu + \varepsilon_d, \quad d = 1, \dots, 100,$$

where ε_d 's are independent and identically distributed according to the t distribution with location zero, scale parameter σ and degrees of freedom ν .

The results based on the priors π_{R2} and π_{IJ} , as well as the prior π_{VW} , are listed in Table 2. It can be seen from Table 2 that the Bayesian estimators of the parameters μ , ν and σ^2 are close to each other for the priors π_{R2} and π_{IJ} , which are somewhat

different from that based on the prior π_{VW} . Further, the credible intervals based on π_{R2} are relatively shorter compared with the other priors, especially for π_{VW} .

Prior	Parameters	Median	95% Credible Interval
π_{R2}	ν	4.1861	[2.1445, 12.2255]
	σ^2	1.8362	[1.1501, 2.8657]
	μ	-0.1638	[-0.4797, 0.1509]
π_{IJ}	ν	4.2678	[2.1185, 13.4195]
	σ^2	1.8368	[1.1180, 2.9335]
	μ	-0.1648	[-0.4937, 0.1538]
π_{VW}	ν	7	[3, 36]
	σ^2	2.1149	[1.3336, 3.2636]
	μ	-0.1922	[-0.5229, 0.1286]

Table 2: Posterior summaries under the priors π_{R2} , π_{IJ} and π_{VW} for the real data in Section 5.2.

6 Concluding remarks

Following the seminal work of Fonseca et al. (2008), in this paper, we have derived all the reference priors for the Student- t linear regression model with unknown degrees of freedom. And the posterior propriety under a larger class of priors is then validated. The frequentist property of Bayesian estimators based on the reference priors is investigated by simulation study. The findings of this paper indicate that it is feasible to use the reference analysis for estimating the unknown degrees of freedom ν from both theoretical and practical viewpoints.

To conclude, it should be pointed out that the posterior propriety in this paper is based on the condition that $n \geq p + 1$. Nowadays, there is a necessity to be able to deal with large amount of data (in terms of covariates). For the high-dimensional cases, there is some work along this topic in the literature. For example, Clarke and Ghosal (2010) considered the posterior properties based on reference priors for exponential families with increasing dimension. The problem for high-dimensional Student- t regression deserves a deeper study, which will be left for our future endeavor.

References

- Abanto-Vallea, C. A., Migona, H. S., and Lachosb, V. H. (2012). “Stochastic volatility in mean models with heavy-tailed distributions.” *Brazilian Journal of Probability and Statistics*, 26: 402–422. MR2949086. doi: <https://doi.org/10.1214/11-BJPS169.142>
- Berger, J. O. and Bernardo, J. M. (1992). “Ordered group reference priors with application to the multinomial problem.” *Biometrika*, 79: 25–37. MR1158515. doi: <https://doi.org/10.1093/biomet/79.1.25>. 130

- Berger, J. O., Bernardo, J. M., and Sun, D. C. (2009). “The formal definition of reference priors.” *The Annals of Statistics*, 37: 905–938. MR2502655. doi: <https://doi.org/10.1214/07-AOS587>. 130
- Bernardo, J. M. (1979). “Reference posterior distributions for Bayesian inference (C/R p128-147).” *Journal of the Royal Statistical Society, Series B*, 41: 113–128. MR0547240. 130
- Clarke, B. and Ghosal, S. (2010). “Reference priors for exponential families with increasing dimension.” *Electronic Journal of Statistics*, 4: 737–780. MR2678969. doi: <https://doi.org/10.1214/10-EJS569>. 143
- Fernández, C. and Steel, M. J. F. (1999). “Multivariate Student- t regression models: pitfalls and inference.” *Biometrika*, 86: 153–167. MR1688079. doi: <https://doi.org/10.1093/biomet/86.1.153>. 133
- 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. doi: <https://doi.org/10.1093/biomet/asn001>. 130, 131, 132, 139, 143
- Fonseca, T. C. O., Ferreira, M. A. R., and Migon, H. S. (2014). “Amendments and corrections: Objective Bayesian analysis for the Student- t regression model.” *Biometrika*, 101: 252. 132
- Geweke, J. (1993). “Bayesian treatment of the independent Student- t linear model.” *Journal of Applied Econometrics*, 8: S19–S40. 130
- Ghosh, J. K. and Mukerjee, R. (1992). “Non-informative priors (with discussion).” In *Bayesian Statistics 4*, 177–194. MR1380277. 132
- Jacquier, E., Polson, N. G., and Rossi, P. E. (2004). “Bayesian analysis of stochastic volatility models with fat-tails and correlated errors.” *Journal of Econometrics*, 122: 185–212. MR2083256. doi: <https://doi.org/10.1016/j.jeconom.2003.09.001>. 129
- Kang, S. M., Liu, G. Y., Qi, H., and Wang, M. (2018). “Bayesian variance change-point detection in linear models with symmetric heavy-tailed errors.” *Computational Economics*, 52: 459–477. 139
- Lange, K. L., Little, R. J., and Taylor, J. M. G. (1989). “Robust statistical modeling using the t distribution.” *Journal of the American Statistical Association*, 84: 881–896. MR1134486. 129
- Liu, C. H. and Rubin, D. B. (1995). “ML estimation of the multivariate t distribution with unknown degrees of freedom using EM and its extensions, ECM and ECME.” *Statistica Sinica*, 5: 19–39. MR1329287. 129
- Sheather, S. (2009). *A modern approach to regression with R*. New York: Springer. MR2730111. doi: <https://doi.org/10.1007/978-0-387-09608-7>. 141
- Siegel, A. (1977). *Practical business statistics*. Boston: Irwin McGraw-Hill. 141
- Singh, R. S. (1988). “Estimation of error variance in linear regression models with

- errors having multivariate Student- t distribution with unknown degrees of freedom.” *Economics Letters*, 27: 47–53. MR0962533. doi: [https://doi.org/10.1016/0165-1765\(88\)90218-2](https://doi.org/10.1016/0165-1765(88)90218-2). 129
- Vallejos, C. A. and Steel, M. F. J. (2013). “On posterior propriety for the Student- t linear regression model under Jeffreys priors.” arXiv:1311.1454v2. MR2949504. doi: <https://doi.org/10.1080/00949655.2011.563239>. 130, 132, 137
- Villa, C. and Walker, S. G. (2014). “Objective prior for the number of degree of freedom of a t distribution.” *Bayesian Analysis*, 9: 197–220. MR3188305. doi: <https://doi.org/10.1214/13-BA854>. 130, 137
- Wang, M. and Yang, M. G. (2016). “Posterior propriety of Student- t linear regression model using objective priors.” *Statistics and Probability Letters*, 113: 23–39. MR3480390. doi: <https://doi.org/10.1016/j.spl.2016.02.003>. 130, 132
- Wang, S. G. and Ip, W. C. (2003). “Inconsistency of estimate of the degree of freedom of multivariate Student- t disturbances in linear regression models.” *Economics Letters*, 80: 383–389. MR1999676. doi: [https://doi.org/10.1016/S0165-1765\(03\)00117-4](https://doi.org/10.1016/S0165-1765(03)00117-4). 129
- West, M. (1984). “Outlier models and prior distributions in Bayesian linear regression.” *Journal of the Royal Statistical Society, Series B*, 46: 431–439. MR0790630. 129
- Yang, F. and Yuan, H. (2017). “A non-iterative Bayesian sampling algorithm for linear regression models with scale mixtures of normal distributions.” *Computational Economics*, 49: 579–597. MR3547960. doi: <https://doi.org/10.1080/00949655.2016.1232675>. 141, 142
- Zellner, A. (1976). “Bayesian and non-Bayesian analysis of the regression model with multivariate Student t error terms.” *Journal of the American Statistical Association*, 71: 400–405. MR0405699. 129

Acknowledgments

The authors are very grateful to the Editor-in-Chief, the Editor, the Associate Editor and the two anonymous Reviewers for their valuable comments and suggestions on earlier versions of this paper.