South African Statistical Journal
Vol. 56, No. 2, pp. 91–120
https://doi.org/10.37920/sasj.2022.56.2.2

A simulation study to compare reference and other priors in the case of a standard univariate Student t-distribution

A. J. van der Merwe[1], *M. J. von Maltitz*[1], *J. H. Meyer*[2] *and P. C. N. Groenewald*[1]

[1]Department of Mathematical Statistics and Actuarial Science, University of the Free State, Bloemfontein, South Africa
[2]Department of Mathematics and Applied Mathematics, University of the Free State, Bloemfontein, South Africa

In this paper, reference and probability-matching priors are derived for the univariate Student t-distribution. These priors generally lead to procedures with properties frequentists can relate to while still retaining Bayes validity. The priors are tested by performing simulation studies. The focus is on the relative mean squared error from the posterior median and on the frequentist coverage of the 95% credibility intervals for a sample of size $n = 30$. Average interval lengths of the credibility intervals as well as the modes of the interval lengths based on 2 000 simulations are also considered. The performance of the priors is also tested on real data, namely daily logarithmic returns of IBM stocks.

Keywords: Coverage percentages, Credibility intervals, Log-returns, Mean squared error, Probability-matching prior, Reference priors, t-distribution.

1. Introduction

In most applied as well as theoretical research works, the residual terms in linear models are assumed to be normally distributed. However, such assumptions may not be appropriate in many practical situations (see, for example, Gnanadesikan and Kettenring, 2005; Zellner, 1976). Many economic and business data, for example stock return data, exhibit heavy (or fat) tail distributions and cannot be effectively modelled by the normal distribution. The usual approach is to test for outliers and to delete them from the data set. However, it is not always easy to identify outliers. Alternatively then, the Student t-distribution can be used instead of the normal distribution to model data with heavy (or fat) tails. The use of the t-distribution reduces the influence of outliers and thus makes the statistical analysis more robust (Fonseca et al., 2008). The robustness of the t-distribution and its suitability to model outliers have been thoroughly discussed in the literature and it has been applied in disciplines such as stock return data (Blattberg and Gonedes, 1974; Zellner, 1976), medicine (Liu, 1997), global navigation satellite systems (Vaneck et al., 1996), finance and biology (Fernández and Steel, 1998) and portfolio optimisation (Kotz and Nadarajah, 2004). According to Fonseca et al. (2008) the degree of robustness of the analysis is directly related to the number of degrees of freedom, ν, with smaller

Corresponding author: M. J. von Maltitz (vmaltitzmj@ufs.ac.za)
MSC2020 subject classifications: 62C10, 62C15, 65C60.

values of ν implying a more robust analysis (Villa and Walker, 2014). Since the degree of robustness of the analysis is related to the number of degrees of freedom, the emphasis of our research will be mainly on the estimation of ν. It should also be noted that this research will not complicate the analysis with any estimation of linear model slope parameters. The estimation of such parameters is a trivial extension to the methodology presented, and details of such estimation can be found in von Maltitz (2015).[1]

Unfortunately, the estimation of the number of degrees of freedom of the t-distribution is not easy. The reason for this is the illogical behaviour of the likelihood function for ν for given location and scale parameters. More precisely, The likelihood function of ν does not always tend to zero as ν goes to infinity, but it rather tends to a positive constant. To overcome the fact that the likelihood function does not vanish in the tail, a prior distribution that tends to zero as ν tends to infinity should be used to form a proper posterior distribution. The uniform prior will result in an improper posterior distribution for ν and can therefore not be used. It is for this reason that non-informative priors are derived in this paper. For further discussion on proper and improper priors for ν (see, for example, Fonseca et al., 2008, 2014; Vallejos and Steel, 2013; Villa and Walker, 2014).

The manuscript is organised as follows. In Section 2, reference and probability-matching priors are given for the parameters ν, μ and σ^2 of the univariate t-distribution. The proofs of these priors are given in Appendix A, and in Appendix B it is shown that the priors tend to zero as ν tends to infinity, and that the reference priors result in proper posterior distributions. In Section 3, simulation studies are performed for the standard t-distribution ($\mu = 0$ and $\sigma^2 = 1$) based on the non-informative priors defined in Section 2 and on priors previously proposed. We assess the simulations using the relative square-rooted mean squared error ($\sqrt{MSE(\nu)}/\nu$) from the posterior median and the frequentist coverage of the 95% credibility intervals for a sample of size $n = 30$. Average interval lengths based on 2 000 simulations are also considered. An application is given in Section 4.

2. Reference and probability-matching priors

Reference and probability-matching priors generally lead to procedures with properties frequentists can relate to while still retaining Bayesian validity. The derivation of the reference priors of Berger and Bernardo (1992) depends on the ordering of the parameters and how the parameter vector is divided into sub-vectors. The reference prior maximises the difference in information about the parameters provided by the prior and the posterior (Pearn and Wu, 2005), i.e., the reference prior provides as little information as possible about the parameters of interest.

The probability-matching prior (another non-informative prior) on the other hand provides accurate frequentist intervals and is also used for comparisons in Bayesian analysis. Datta and Ghosh (1995) provided a method for finding probability-matching priors by deriving a differential equation that a prior must satisfy if the posterior probability of a one-sided credibility interval for a parametric function and its frequentist probability agree up to $O(n^{-1})$, where n is the sample size. The following theorems can now be stated.

---

[1] In short, the mean parameter μ is replaced by a conditional mean $X\boldsymbol{\beta}$, and the degrees of freedom available for estimation is reduced by the increased dimensions of $\boldsymbol{\beta}$. The Gibbs sampling process for $\boldsymbol{\beta}$, given in von Maltitz (2015), is not substantially more complicated than for μ alone, but the added parameterisation is not necessary in research that is solely focused on the robustness of the error distribution.

Theorem 2.1. *The reference prior for the orderings* $\{\nu, \mu, \sigma^2\}$, $\{\mu, \nu, \sigma^2\}$ *and* $\{\nu, \sigma^2, \mu\}$ *is given by*

$$p_2\left(\nu, \mu, \sigma^2\right) \propto \sigma^{-2}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}$$

and the reference prior for the orderings $\{\mu, \sigma^2, \nu\}$, $\{\sigma^2, \mu, \nu\}$ *and* $\{\sigma^2, \nu, \mu\}$ *is given by*

$$p_1\left(\mu, \sigma^2, \nu\right) \propto \sigma^{-2}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right]^{\frac{1}{2}},$$

where $\Gamma(\cdot), \Psi(\cdot)$ *and* $\Psi'(\cdot)$ *are the gamma, digamma and trigamma functions, i.e.,* $\Psi(a) = \frac{d}{da}\log\Gamma(a)$ *and* $\Psi'(a) = \frac{d}{da}\Psi(a)$*, and the ordering of, for example,* $\{\nu, \mu, \sigma^2\}$*, means that* ν *is considered to be the most important parameter,* μ *the second most important, and* σ^2 *the least important parameter.*

Proof. See proof in Appendix A.

Theorem 2.2. *The prior* $p_2\left(\nu, \mu, \sigma^2\right)$ *is also a probability-matching prior for* ν*.*

Proof. See proof in Appendix A.

Theorem 2.3. *The reference priors tend to zero as* ν *tends to infinity.*

Proof. See proof in Appendix B.

Theorem 2.4. *In the case of the standard univariate* t*-distribution the reference priors result in proper posterior distributions for* ν*.*

Proof. See proof in Appendix B.

It should be noted that Wang and Yang (2016) also derived the two one-at-a-time reference priors $p_1(\nu, \mu, \sigma^2)$ and $p_2(\nu, \mu, \sigma^2)$, and proved that $p_2(\nu, \mu, \sigma^2)$ is a first-order probability-matching prior. The proofs of reference and probability-matching priors (Theorems 2.1 and 2.2) were originally derived in von Maltitz (2015), an unpublished PhD thesis. The proof of Theorem 2.2 is also given in an unpublished technical report by van der Merwe et al. (2021).

3. Simulation study

3.1 Priors compared

The Student t linear model for independent data x_i, $i = 1, 2, \ldots, n$, and $n > 2$, is given as follows:

$$f\left(x_i|\mu, \sigma^2, \nu\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\nu^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)\sigma}\left[\nu + \left(\frac{x_i - \mu}{\sigma}\right)^2\right]^{-\frac{1}{2}(\nu+1)}.$$

We consider six priors for this model. Let us introduce these as $p_i(\nu, \mu, \sigma^2)$, $i = 1, 2, \ldots, 6$. As mentioned in Theorem 2.1,

$$p_1\left(\nu, \mu, \sigma^2\right) \propto \sigma^{-2}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right]^{\frac{1}{2}}$$

is a reference prior with respect to the orderings $\{\mu, \sigma^2, \nu\}$, $\{\sigma^2, \mu, \nu\}$ and $\{\sigma^2, \nu, \mu\}$. From the Fisher information matrix given in (1) in Appendix A, it can be seen that it is also a Jeffreys prior for ν if μ and σ^2 are considered to be known. Liu (1997) proposed the prior

$$\pi(\nu) \propto \left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+d}{2}\right) - \frac{2d(\nu+d+4)}{\nu(\nu+d)(\nu+d+2)}\right]^{\frac{1}{2}}$$

for the multivariate t-distribution, where d is the dimension of the multivariate distribution. The prior $\pi(\nu)$ is obtained by applying Jeffreys' rule (Box and Tiao, 2011). Villa and Rubio (2018) included it in their simulation study on objective priors for the number of degrees of freedom of a multivariate t-distribution. If $d = 1$, $\pi(\nu)$ simplifies to $p_1(\nu)$.

The prior

$$p_2\left(\nu, \mu, \sigma^2\right) \propto \sigma^{-2}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}},$$

on the other hand, is a probability-matching prior for ν as well as a reference prior for the parameter orderings $\{\nu, \mu, \sigma^2\}$, $\{\mu, \nu, \sigma^2\}$ and $\{\nu, \sigma^2, \mu\}$ (see Theorems 2.1 and 2.2).

The prior

$$p_3\left(\nu, \mu, \sigma^2\right) \propto \sigma^{-2}\left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}$$

is the independence Jeffreys prior and

$$p_4\left(\nu, \mu, \sigma^2\right) \propto \sigma^{-3}\left(\frac{\nu+1}{\nu+3}\right)^{\frac{1}{2}}\left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}$$

is the Jeffreys-rule prior. Both of these priors were derived by Fonseca et al. (2008). The Jeffreys-rule prior is proportional to the square root of the determinant of the Fisher information matrix, while the independence Jeffreys prior is obtained by assuming that the priors for μ and $\left(\sigma^2, \nu\right)$ are independent, i.e., $p_3\left(\nu, \mu, \sigma^2\right) = p_3(\mu)\, p_3\left(\nu, \sigma^2\right)$. From the Fisher information matrix defined in (1) it therefore follows that

$$p_3(\mu) \propto \sqrt{\det\left[I(\theta)\right]_{22}} \propto 1$$

and

$$p_3\left(\nu, \sigma^2\right) \propto \sqrt{\left[I(\theta)\right]_{11}\left[I(\theta)\right]_{33} - \left[I(\theta)\right]_{13}^2}.$$

The exponential prior $p_5\left(\nu, \mu, \sigma^2\right) \propto \sigma^{-2} e^{-\xi\nu}$, where $\xi = 0.1$, was derived by Geweke (1993), but according to Fonseca et al. (2008) and Villa and Walker (2014) this prior is too informative and is found to dominate the data.

Juárez and Steel (2010) considered a non-hierarchical and a hierarchical prior. The first is a gamma prior with parameters 1 and $1/100$. The hierarchical prior is obtained by considering an exponential distribution for the scale parameter of the gamma prior with slope parameter a. In other words,

$$p_6(\nu) = \int_0^\infty p(\nu|a)\, p(a|\tilde{d})\, da,$$

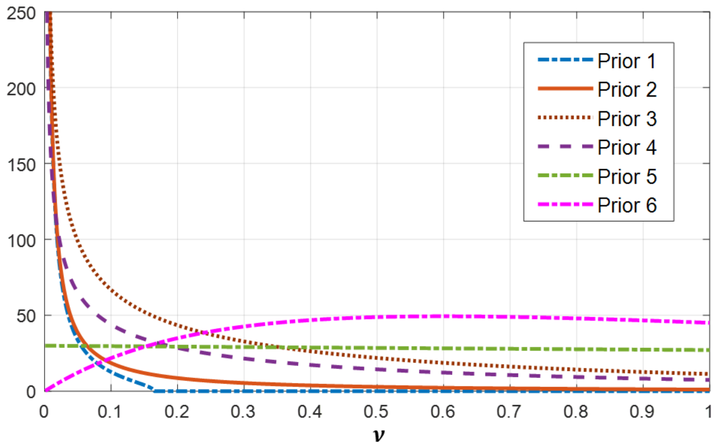

Figure 1. Comparison of the six prior distributions in terms of ν.

where $p(\nu|a) = a^2\nu e^{-a\nu}$ and $p(a|\tilde{d}) = \tilde{d}e^{-a\tilde{d}}$. The resulting prior is therefore $p_6(\nu,\mu,\sigma^2) \propto \sigma^{-2}2\nu\tilde{d}/(\nu+\tilde{d})^3$ for $\nu > 0$ and $\tilde{d} > 0$. The parameter a controls the mode $(\tilde{d}/2)$ and the median $(1+\sqrt{2})\tilde{d}$. Rubio and Steel (2015) mentioned that if $\tilde{d} = 1.2$ then it will be a continuous alternative to the discrete objective prior proposed by Villa and Walker (2014), so $p_6(\nu,\mu,\sigma^2) \propto \sigma^{-2}2\nu\tilde{d}/(\nu+\tilde{d})^3$, where $\tilde{d} = 1.2$.

Thus, the six priors that are used in the simulation study for comparison are:

1. $p_1\left(\nu,\mu,\sigma^2\right) \propto \sigma^{-2}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right]^{\frac{1}{2}}$,

2. $p_2\left(\nu,\mu,\sigma^2\right) \propto \sigma^{-2}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}$,

3. $p_3\left(\nu,\mu,\sigma^2\right) \propto \sigma^{-2}\left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}$,

4. $p_4\left(\nu,\mu,\sigma^2\right) \propto \sigma^{-3}\left(\frac{\nu+1}{\nu+3}\right)^{\frac{1}{2}}\left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}$,

5. $p_5\left(\nu,\mu,\sigma^2\right) \propto \sigma^{-2}e^{-\xi\nu}$, where $\xi = 0.1$,

6. $p_6\left(\nu,\mu,\sigma^2\right) \propto \sigma^{-2}\frac{2\nu\tilde{d}}{(\nu+\tilde{d})^3}$, where $\tilde{d} = 1.2$.

For the standard univariate t-distribution ($\mu = 0, \sigma^2 = 1$), the priors will be denoted by $p_i(\nu)$, $i = 1,2,\ldots,6$. In Figure 1 a visualisation of the priors $p_i(\nu), i = 1,\ldots,6$, is given.

It is observed from Figure 1 that $p_5(\nu)$ and $p_6(\nu)$ differ from the other priors. Even though the shapes of the densities of priors $p_1(\nu), p_2(\nu), p_3(\nu)$ and $p_4(\nu)$ are quite similar, the behaviours of the resulting posterior distributions can be quite different, especially if the sample sizes are small.

The priors $p_i(\nu,\mu,\sigma^2)$, $i = 1,2,\ldots,6$, can be written as $p(\nu,\mu,\sigma^2) \propto \sigma^{-2c}p(\nu)$, where $p(\nu)$ is the component of the prior that depends on ν. It is proved by Vallejos and Steel (2013) that the posterior distribution of (ν,μ,σ^2) is not proper if $c \geq 1$ and the range of ν is $(0,\infty)$. In particular, the

Jeffreys-rule prior for which $a = \frac{3}{2}$ does not lead to a proper posterior distribution. Our simulation study is however based on the standard univariate t-distribution ($\mu = 0, \sigma^2 = 1$), and as shown by Fonseca et al. (2008), in this case

$$p_4\left(\nu, \mu, \sigma^2\right) \propto \sigma^{-3}\left(\frac{\nu+1}{\nu+3}\right)^{\frac{1}{2}}\left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2\left(\nu+3\right)}{\nu\left(\nu+1\right)^2}\right]^{\frac{1}{2}}$$

is a proper density function of ν. The Jeffreys-rule prior is therefore included in our simulation study. The fact that the reference priors result in proper posterior distributions in the case of the standard univariate t-distribution and the fact that $p_i(\nu, \mu, \sigma^2) = \sigma^{-2a} p_i(\nu), i = 1, 2$, where $a = 1$, should have been an indication that $p_1(\nu, \mu, \sigma^2)$ and $p_2(\nu, \mu, \sigma^2)$ will lead to proper posterior distributions. It is therefore not surprising that He et al. (2021) proved that for the Student t linear model, for $i = 1, 2, \ldots, n$ and $n > 2$, the posterior distributions of (ν, μ, σ^2) are proper under the reference priors $p_1(\nu, \mu, \sigma^2)$ and $p_2(\nu, \mu, \sigma^2)$.

3.2 Frequentist properties

In this subsection we summarise the frequentist properties of the priors for ν in the case of the univariate standard t-distribution. The focus is on the relative square rooted mean squared error from the median of the posterior distribution of ν. The index is denoted by $\sqrt{MSE\left(\nu\right)}/\nu$, where $MSE = E\left(\nu - m\right)^2$, and m is the median of the posterior distribution. The frequentist coverage percentages of the 95% credibility intervals for a sample of size $n = 30$ as well as the interval lengths and modes of the interval lengths based on 2 000 simulations are also considered.

To describe the simulation procedure more clearly the following should also be mentioned. The simulation study was done for 25 different values of ν, i.e., $\nu = 1, 2, 3, \ldots, 25$. We start by drawing a sample of $n = 30$ observations $(x_1, x_2, \ldots, x_n)$ from a standard t-distribution ($\mu = 0$ and $\sigma^2 = 1$) with ν degrees of freedom. By using the observations in the sample, the likelihood function is calculated as

$$L\left(\nu | x_1, \ldots, x_n\right) = \frac{\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^n \nu^{\frac{n\nu}{2}}}{\left[\Gamma\left(\frac{\nu}{2}\right)\right]^n \left[\Gamma\left(\frac{1}{2}\right)\right]^n} \prod_{i=1}^{n}\left(\nu + x_i^2\right)^{-\frac{1}{2}\left(\nu+1\right)}.$$

The posterior distributions $p_i\left(\nu | x_1, \ldots, x_n\right), i = 1, 2, \ldots, 6$, are obtained by multiplying the likelihood function with each of the six prior distributions. For each of the six posterior distributions, the parameter estimates (mean, median, mode, variance, 95% credibility intervals and interval lengths) are calculated for that simulation run.

For a specific ν, this process is replicated 2 000 times. More specifically, this means that 2 000 samples of size $n = 30$ each are drawn from a standard t-distribution with ν degrees of freedom to obtain 2 000 posterior distribution sets $p_i\left(\nu | x_1, \ldots, x_n\right)$, $i = 1, 2, \ldots, 6$. The parameter estimates for each of the 12 000 (= 2 000 × 6) posterior distributions are therefore calculated. This is done for each of the 25 values of ν. In total, 300 000 (= 2 000 × 6 × 25) posterior distributions are used in the simulation study.

Our focus is on the special case $\mu = 0$ and $\sigma^2 = 1$ to estimate the degrees of freedom of the Student t-distribution. Other statisticians, for example Fonseca et al. (2008), Villa and Walker (2014) and Villa and Rubio (2018), also used the standard t-distribution in their simulation studies. This might

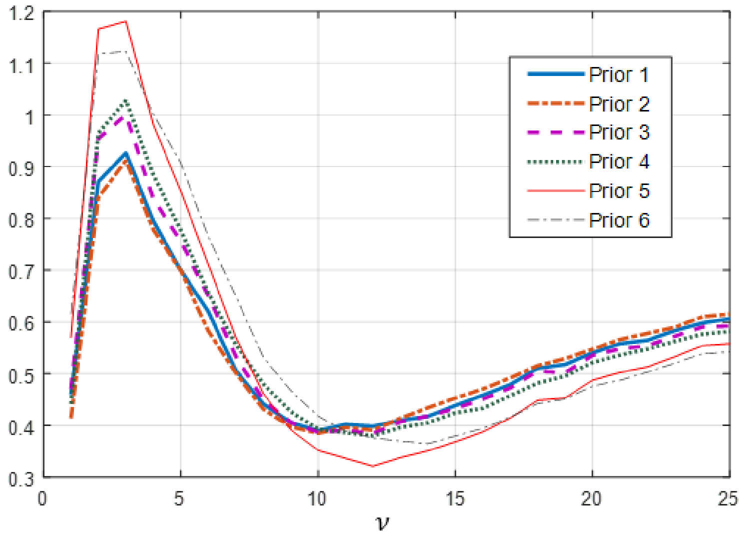

Figure 2. Relative root mean squared errors for the posterior median of ν.

Table 1. Averages of the relative root mean squared errors for ν.

ν	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**	**Prior 5**	**Prior 6**
Mean (1 to 10)	0.6115	0.5942	0.6445	0.6608	0.7238	0.7596
Mean (11 to 25)	0.4988	0.5067	0.4903	0.4790	0.4378	0.4434

be a limitation of our research, but, from the relationship $x(\nu, \mu, \sigma^2) = \mu + \sigma(\nu, 0, 1)$ the simulation results will at least not be affected by changes in the location parameter μ. At this stage it is not clear to us how much changes in the scale parameter σ will affect the simulation results, although Villa and Walker (2014) suggest that changes in the scale parameter will not significantly affect the posterior distributions.

From Table 7 in Appendix D one can argue that reference priors $p_1(\nu)$ and $p_2(\nu)$ are on average the two best priors. The prior $p_2(\nu)$, which is also a probability-matching prior, is somewhat better than $p_1(\nu)$. The prior $p_6(\nu)$ is performing worst in this study. These results are more succinctly illustrated in Figure 2 and Table 1.

It should be reiterated that researchers are usually interested in t-distributions with a small number of degrees of freedom, and from Table 1 it can be seen that $p_2(\nu)$ is particularly good if ν is small (1 to 10); For large values of ν (11 to 25) we have that $p_5(\nu)$ and $p_6(\nu)$ seem to be the best priors. However, it is clear from Figure 2 that for ν between two and six the $\sqrt{MSE(\nu)}/\nu$ values for priors $p_5(\nu)$ and $p_6(\nu)$ are larger than those of the objective priors.

In Table 8 in Appendix D, the frequentist coverage percentages of the 95% credibility intervals are given for a sample size of $n = 30$ and 2 000 simulations. These values are illustrated in Figure 3.

According to Fonseca et al. (2008, Figure 2, p. 329), the coverage percentages of the 95% credibility intervals for ν in the case of the Jeffreys-rule prior, are poor for $n = 30$. They also mentione that for the Geweke prior $p_5(\nu)$, the frequentist coverage is much smaller than the nominal

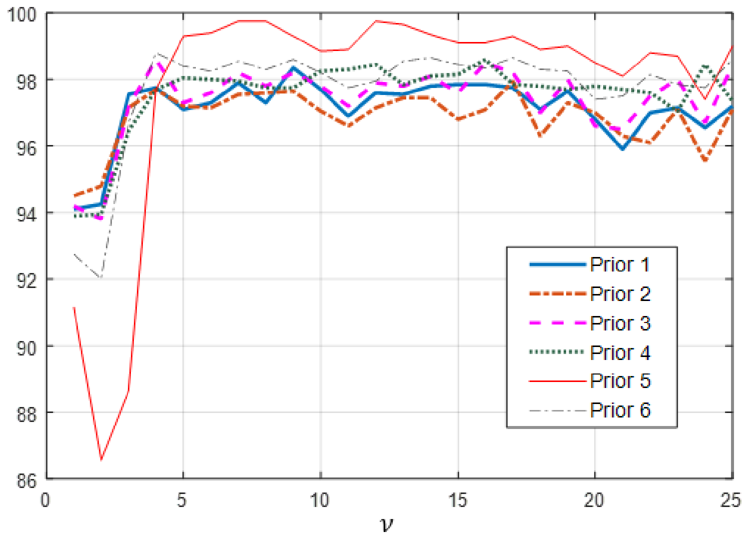

Figure 3. Frequentist coverage percentages of the 95% credibility intervals for ν.

level for small ν and is undesirably close to 1 for $\nu > 6$. The results of Fonseca et al. (2008) differ somewhat from our results given in Table 8 and Figure 3. It can be seen that for $\nu \leq 3$ the frequentist coverage percentage of the Geweke prior is smaller than the nominal level and for $\nu \geq 4$ it is on average 98.98%. The coverage percentages of the Jeffreys-rule prior $p_4(\nu)$, however, do not differ much from those of the other objective priors ($p_1(\nu)$, $p_2(\nu)$ and $p_3(\nu)$). In the case of the coverage percentages, the reference (or probability-matching) prior $p_2(\nu)$ seems to be the best because it has on average a 96.86% coverage.

From Table 9 in Appendix D, illustrated in Figure 4, it can be observed that $p_2(\nu)$ has the shortest average interval lengths of all the objective priors. The prior that gives the shortest interval lengths is however $p_5(\nu)$, the Geweke prior, with interval lengths on average two and a half to three times shorter than those of the objective priors and with a coverage percentage of more than 95%. The worst performing prior seems to be $p_6(\nu)$.

Although the interval lengths of the objective priors for most of the 2 000 simulations are quite small, a few extremely large lengths can have a big influence on the average interval length. A large interval length will occur if the observations in the sample are of such a nature that it is not clear if the data were drawn from a normal or t-distribution. It is for this reason that the modes of the interval lengths are given in Table 2 and Figure 5.

As before, the reference priors $p_1(\nu)$ and $p_2(\nu)$ seem to be the two best priors because the modes of their interval lengths are in general the smallest. The prior $p_2(\nu)$ seems to be somewhat better than $p_1(\nu)$. The modes of the interval lengths of the priors $p_3(\nu)$ and $p_4(\nu)$ (the independence Jeffreys and the Jeffreys-rule priors) change dramatically for $\nu \geq 10$. From Table 2 it is clear that $p_5(\nu)$, the Geweke prior, is the worst prior for $5 \leq \nu \leq 8$. It does well for $1 \leq \nu \leq 4$ and seems to do better than most of the priors for $\nu > 10$. The prior $p_6(\nu)$ again seems to perform worst in this study.

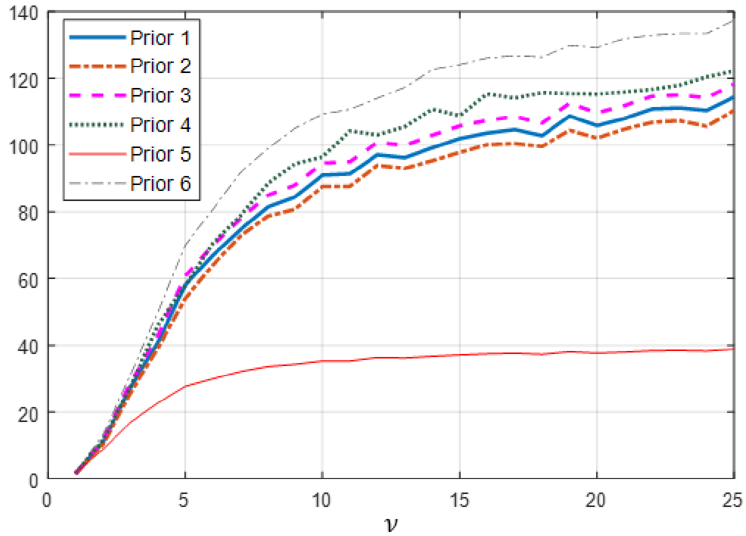

Figure 4. Average interval lengths of the 95% credibility intervals for ν.

Table 2. Mode of interval lengths of the 95% credibility intervals for ν.

ν	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**	**Prior 5**	**Prior 6**
1	1.3918	0.9694	0.8824	1.0916	1.1767	1.1350
2	2.2132	2.1039	2.2849	2.2145	2.5818	2.2474
3	5.1277	5.1071	2.5923	5.3916	4.1166	5.4524
4	5.5141	5.8186	5.6744	6.1332	7.0209	6.2217
5	5.8345	6.4857	6.3385	6.8614	40.3448	6.9601
6	8.4593	6.7914	7.0016	7.1734	39.8707	7.2716
7	7.1246	6.5109	9.9536	6.8896	42.4445	9.7214
8	6.9421	7.2842	7.8857	10.4286	40.2139	10.5493
9	9.5604	7.1040	10.2161	7.5794	41.2099	162.6129
10	10.6371	9.9450	151.4783	156.0163	40.6707	178.7038

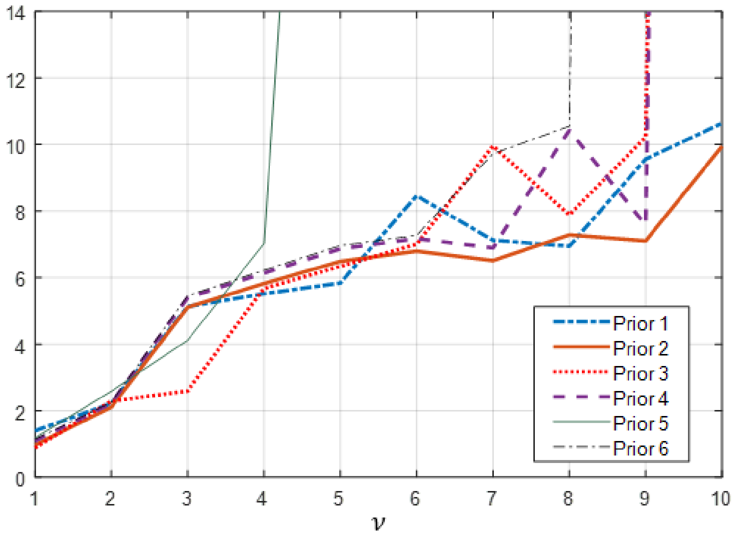

Figure 5. Mode of interval lengths of the 95% credibility intervals for ν.

4. Application

To compare the six priors on real data, the Student t-distribution will be applied to model daily log returns of IBM data. The data set contains 2528 observations for the period from the 3rd of January 1969 to the 31st of December 1998. The data are available from the 'Ecdat' R package (R Core Team, 2013; Croissant and Graves, 2020). In Figure 6, a histogram of the data is shown. It is clear that the data exhibit a heavy (or fat) tail distribution.

By using the prior

$$p_2\left(\nu, \mu, \sigma^2\right) \propto \sigma^{-2}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}$$

and Gibbs sampling, the posterior distributions of the parameters μ, σ^2 and ν are obtained and illustrated in Figures 7, 8 and 9. The resulting posterior estimates of μ, σ^2 and ν for the six priors are summarised in Tables 3 to 5. The conditional posterior distributions that were used in the Gibbs sampling procedure are given in Appendix C.

From Tables 3 to 5 it is clear that for this example similar results are obtained with the different priors due to the large sample size. To obtain a better idea of the differences between the six priors, a random sample of $n = 100$ observations of the daily log returns of the IBM data are analysed. The resulting posterior estimates of ν for the six priors are summarised in Table 6.

It is interesting to note that the 95% credibility intervals of ν for the full data set are on average 4.7 times shorter than those for the sample of 100 observations. This is close to what one would expect given that confidence/credibility intervals are generally proportional to the square root of the sample size. In this case $\sqrt{2528}/\sqrt{100} \approx 5$, which lines up with the observed value.[2] Also, the posterior distributions in the case of the full data set are more symmetrical.

[2] We would like to acknowledge the input of a referee for this comment.

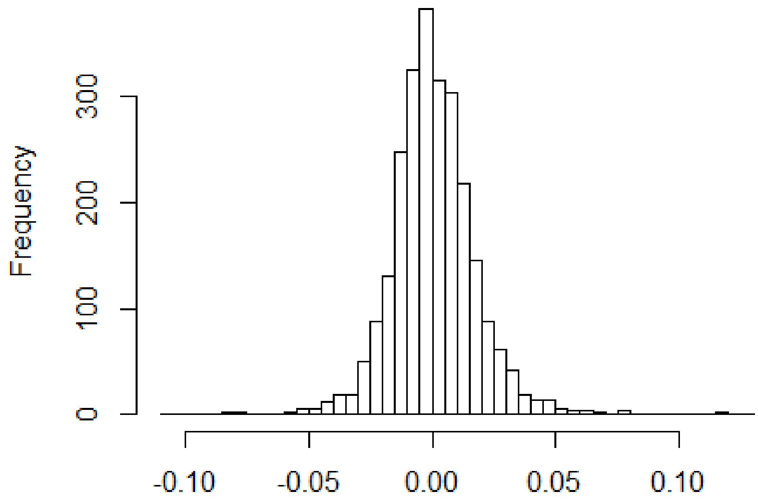

Figure 6. IBM daily log-returns, 1969-1998.

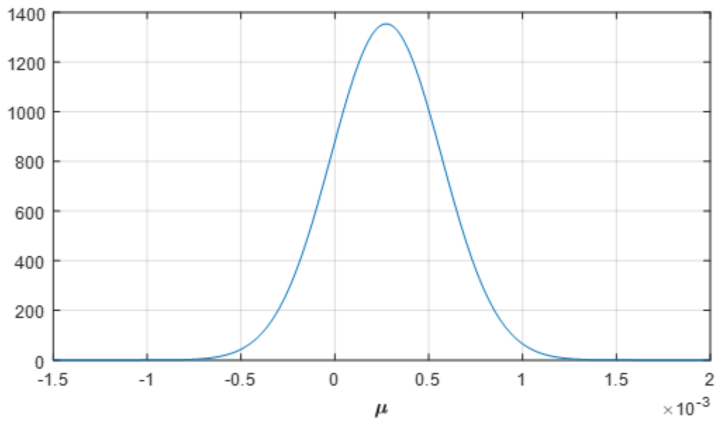

Figure 7. Posterior distribution of μ using Prior $p_2(\nu,\mu,\sigma^2)$. Mean = 2.76×10^{-4}, Median = 2.71×10^{-4}, Mode = 2.67×10^{-4}, Var = 8.80×10^{-8}; 95% Equal-tail Interval = $(-3.11\times10^{-4}; 8.53\times10^{-4})$.

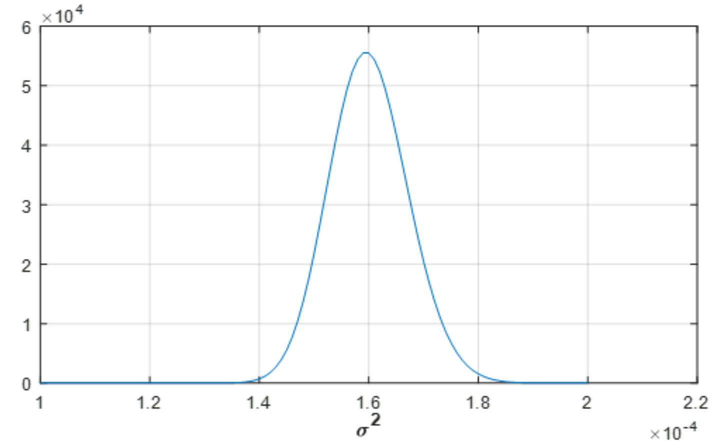

Figure 8. Posterior distribution of σ^2 using Prior $p_2(\nu, \mu, \sigma^2)$. Mean = 1.60×10^{-4}, Median = 1.59×10^{-4}, Mode = 1.59×10^{-4}, Var = 5.28×10^{-11}; 95% Equal-tail Interval = $(1.46\times10^{-4}; 1.75\times10^{-4})$.

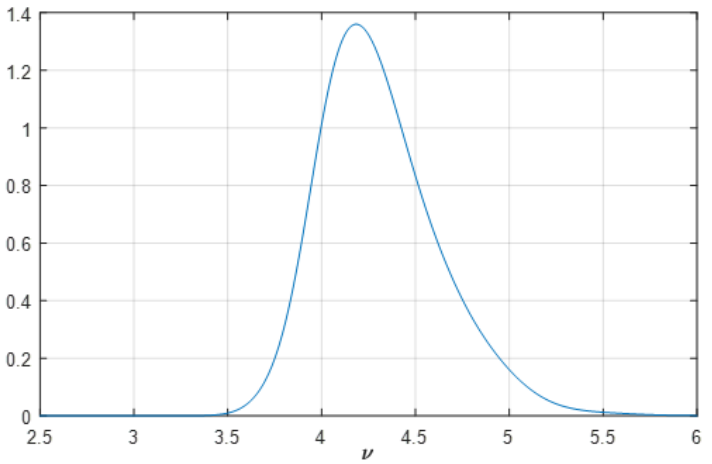

Figure 9. Posterior distribution of ν using Prior $p_2(\nu, \mu, \sigma^2)$. Mean = 4.29, Median = 4.25, Mode = 4.19, Var = 0.10; 95% Equal-tail Interval = $(3.75; 4.99)$.

Table 3. Posterior estimates for μ obtained with six different priors for ν.

Prior	Mean $\times 10^4$	Median $\times 10^4$	95% Credibility Interval $\times 10^4$
1	2.7244	2.6725	(-3.114; 8.474)
2	2.7593	2.7058	(-3.107; 8.527)
3	2.7595	2.7061	(-3.065; 8.500)
4	2.7234	2.6700	(-3.124; 8.489)
5	2.7174	2.6660	(-3.153; 8.495)
6	2.7457	2.6970	(-3.124; 8.508)

Table 4. Posterior estimates for σ^2 obtained with six different priors for ν.

Prior	Mean $\times 10^4$	Median $\times 10^4$	95% Credibility Interval $\times 10^4$
1	1.6039	1.5962	(1.463; 1.749)
2	1.6021	1.5947	(1.462; 1.747)
3	1.6019	1.5955	(1.462; 1.745)
4	1.6013	1.5939	(1.463; 1.743)
5	1.6049	1.5972	(1.463; 1.750)
6	1.6036	1.5964	(1.464; 1.745)

Table 5. Posterior estimates for ν obtained with six different priors for ν.

Prior	Mean	Median	95% Credibility Interval
1	4.3105	4.2713	(3.7697; 5.0315)
2	4.2869	4.2474	(3.7541; 4.9855)
3	4.2908	4.2568	(3.7631; 4.9795)
4	4.2795	4.2450	(3.7525; 4.9652)
5	4.3126	4.2738	(3.7646; 5.0179)
6	4.2984	4.2650	(3.7678; 4.9703)

Table 6. Posterior estimates for ν obtained with six different priors for ν — random sample of 100 observations.

Prior	Mean	Median	95% Credibility Interval
1	3.6454	3.33	(1.851; 7.044)
2	3.6118	3.28	(1.797; 7.136)
3	3.7471	3.41	(1.830; 7.217)
4	3.7894	4.43	(1.860; 7.919)
5	4.3722	3.95	(2.064; 9.112)
6	3.9485	3.59	(1.940; 7.561)

5. Discussion

The Student t-distribution is of great importance for many economic and business data because it reduces the influence of outliers in model estimation and thus makes statistical analysis more robust.

Unfortunately, the estimation of ν, the number of degrees of freedom of the t-distribution, is not easy. The reason for this is the illogical behaviour of the likelihood function for ν. To overcome the fact that the likelihood function does not vanish in the tail, a prior distribution that tends to zero as ν tends to infinity should be applied. It is for this reason that two non-informative priors have been derived for the parameters of the Student t-distribution, $p_1(\nu, \mu, \sigma^2)$ and $p_2(\nu, \mu, \sigma^2)$. Both of these priors are reference priors while $p_2(\nu, \mu, \sigma^2)$ is also a probability-matching prior.

Our simulation studies illustrate the good frequentist properties of the posterior distributions associated with these priors, focusing on the relative square-rooted mean squared error from the posterior median and the 95% credibility intervals for a sample of size $n = 30$ based on 2 000 simulations. We have compared the frequentist properties of the two reference priors to four other priors (the Jeffreys-rule prior, the independence Jeffreys prior, the exponential prior and a hierarchical prior). Overall, the two reference priors seem to give better results, especially if $1 \leq \nu \leq 10$. For $n = 30$ our simulation results are for all practical purposes the same as those obtained by He et al. (2021). These authors also came to the conclusion that the reference priors are the best. However, they did not do any simulation studies on the "Average Interval Lengths of the 95% Credibility Intervals for ν" or on the "Modes of the Interval Lengths".

In Section 4 the six priors are compared on a real data set of the daily log-returns of IBM stock. A sample of $n = 100$ observations is also analysed. The results for the sample show that the posterior estimates of ν for the reference priors, probability matching prior, the Jeffreys-rule prior, and the independence Jeffreys prior are for all practical purposes the same, but differ somewhat from those of the exponential and hierarchical priors.

Acknowledgement. The authors would like to thank the editor and reviewers for their valuable comments, which helped to improve the quality of this paper.

A. Reference and probability-matching priors

This appendix provides derivations for the reference and probability-matching priors for the univariate Student t-distribution. As in the case of the Jeffreys priors, the derivations of these priors are based on the Fisher information matrix. Differentiation of the log likelihood functions twice with respect to the unknown parameters and taking expected values give the Fisher information matrix

$$\{I(\theta)\}_{ij} = E_{X|\theta}\left[-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\{L(\boldsymbol{\theta};\mathbf{x})\}\right],$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]' = [\nu, \mu, \sigma]'$ and

$$L(\theta, \mathbf{x}) = \prod_{i=1}^{n} f\left(x_i|\mu, \sigma^2, \nu\right) = \frac{\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^n \nu^{n\nu/2}}{\left[\Gamma\left(\frac{\nu}{2}\right)\right]^n\left[\Gamma\left(\frac{1}{2}\right)\right]^n\sigma^n}\prod_{i=1}^{n}\left[\nu+\left(\frac{x_i-\mu}{\sigma}\right)^2\right]^{-\frac{1}{2}(\nu+1)}.$$

Proof of Theorem 2.1. First we prove that

$$p_2\left(\nu, \mu, \sigma^2\right) \propto \sigma^{-2}\left[\Psi'\left(\frac{\nu}{2}\right)-\Psi'\left(\frac{\nu+1}{2}\right)-\frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}.$$

The Fisher information matrix for the ordering $\{\nu, \mu, \sigma\}$ is given in Fonseca et al. (2008) as

$$I(\nu, \mu, \sigma) = \begin{bmatrix} \frac{n}{4}\left[\Psi'\left(\frac{\nu}{2}\right)-\Psi'\left(\frac{\nu+1}{2}\right)-\frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right] & 0 & \frac{-2n}{\sigma(\nu+1)(\nu+3)} \\ 0 & \frac{n(\nu+1)}{\sigma^2(\nu+3)} & 0 \\ \frac{-2n}{\sigma(\nu+1)(\nu+3)} & 0 & \frac{2n\nu}{\sigma^2(\nu+3)} \end{bmatrix} \tag{1}$$

$$= \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix}.$$

To calculate the reference prior for the ordering $\{\nu, \mu, \sigma\}$, we must first calculate

$$h_1 = F_{11} - \begin{bmatrix} F_{12} & F_{13} \end{bmatrix}\begin{bmatrix} F_{22} & F_{23} \\ F_{32} & F_{33} \end{bmatrix}^{-1}\begin{bmatrix} F_{21} \\ F_{31} \end{bmatrix}$$

from the information matrix in (1). Therefore,

$$\begin{aligned} h_1 &= \frac{n}{4}\left[\Psi'\left(\frac{\nu}{2}\right)-\Psi'\left(\frac{\nu+1}{2}\right)-\frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right] \\ &\quad - \begin{bmatrix} 0 & \frac{-2n}{\sigma(\nu+1)(\nu+3)} \end{bmatrix}\begin{bmatrix} \frac{\sigma^2(\nu+3)}{n(\nu+1)} & 0 \\ 0 & \frac{\sigma^2(\nu+3)}{2n\nu} \end{bmatrix}\begin{bmatrix} 0 \\ \frac{-2n}{\sigma(\nu+1)(\nu+3)} \end{bmatrix} \\ &= \frac{n}{4}\left[\Psi'\left(\frac{\nu}{2}\right)-\Psi'\left(\frac{\nu+1}{2}\right)-\frac{2(\nu+3)}{\nu(\nu+1)^2}\right] \end{aligned}$$

and

$$h_1^{\frac{1}{2}} \propto \left[\Psi'\left(\frac{\nu}{2}\right)-\Psi'\left(\frac{\nu+1}{2}\right)-\frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}}.$$

Now $p\,(\nu) \propto h_1^{1/2}$.

Consider

$$\begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} - (F_{33})^{-1} \begin{bmatrix} F_{13} \\ F_{23} \end{bmatrix} \begin{bmatrix} F_{31} & F_{32} \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}.$$

Since $h_2 = H_{22} = F_{22} - F_{23}/F_{32}F_{33}$ does not contain μ, $h_2^{1/2} \propto c$ and $p(\mu|\nu) \propto c$. Further, $h_3 = F_{33} = 2n\mu/\sigma^2(\nu+3)$ and $p(\sigma|\nu,\mu) \propto h_3^{1/2} = \sigma^{-1}$.

From this it follows that

$$p_2\,(\nu,\mu,\sigma) = p\,(\nu)\,p\,(\mu|\nu)\,p\,(\sigma|\nu,\mu) \propto \sigma^{-1}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2\,(\nu+3)}{\nu\,(\nu+1)^2}\right]^{\frac{1}{2}}.$$

Similarly,

$$p_2\left(\nu,\mu,\sigma^2\right) \propto \sigma^{-2}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2\,(\nu+3)}{\nu\,(\nu+1)^2}\right]^{\frac{1}{2}}$$

because it is known that if $p\,(\sigma) \propto \sigma^{-1}$, then $p(\sigma^2) \propto \sigma^{-2}$, and for $p(\sigma) \propto \sigma^{-2}$ it follows that $p(\sigma^2) \propto \sigma^{-3}$.

By using the Fisher information matrices $I(\nu,\sigma,\mu)$ and $I(\mu,\nu,\sigma)$ it can be shown that $p_2(\nu,\mu,\sigma)$ is also a reference prior for the orderings $\{\nu,\sigma,\mu\}$ and $\{\mu,\nu,\sigma\}$.

Now let us prove that

$$p_1\,(\sigma,\mu,\nu) \propto \sigma^{-1}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2\,(\nu+5)}{\nu\,(\nu+1)\,(\nu+3)}\right]^{\frac{1}{2}},$$

which is equivalent to proving that

$$p_1\left(\sigma^2,\mu,\nu\right) \propto \sigma^{-2}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2\,(\nu+5)}{\nu\,(\nu+1)\,(\nu+3)}\right]^{\frac{1}{2}}.$$

The Fisher information matrix for the ordering $\{\sigma,\nu,\mu\}$ is given by

$$I\,(\sigma,\nu,\mu) = \begin{bmatrix} \frac{2n\nu}{\sigma^2(\nu+3)} & \frac{-2n}{\sigma(\nu+1)(\nu+3)} & 0 \\ \frac{-2n}{\sigma(\nu+1)(\nu+3)} & \frac{n}{4}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right] & 0 \\ 0 & 0 & \frac{n(\nu+1)}{\sigma^2(\nu+3)} \end{bmatrix}$$
$$= \begin{bmatrix} \tilde{F}_{11} & \tilde{F}_{12} & \tilde{F}_{13} \\ \tilde{F}_{21} & \tilde{F}_{22} & \tilde{F}_{23} \\ \tilde{F}_{31} & \tilde{F}_{32} & \tilde{F}_{33} \end{bmatrix}.$$

Now,

$$\tilde{h}_1 = \tilde{F}_{11} - \begin{bmatrix} \tilde{F}_{12} & \tilde{F}_{13} \end{bmatrix} \begin{bmatrix} \tilde{F}_{22} & \tilde{F}_{23} \\ \tilde{F}_{32} & \tilde{F}_{33} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{F}_{21} \\ \tilde{F}_{31} \end{bmatrix},$$

whence

$$\tilde{h}_1 = \frac{2n\nu}{\sigma^2(\nu+3)} - \begin{bmatrix} \frac{-2n}{\sigma(\nu+1)(\nu+3)} & 0 \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & \frac{\sigma^2(\nu+3)}{n(\nu+1)} \end{bmatrix} \begin{bmatrix} 0 \\ \frac{-2n}{\sigma(\nu+1)(\nu+3)} \end{bmatrix},$$

where $A = 4n^{-1}[\psi'(\nu/2) - \psi'((\nu+1)/2) - 2(\nu+5)/\nu(\nu+1)(\nu+3)]^{-1}$. Therefore,

$$\tilde{h}_1 = \frac{2n\nu}{\sigma^2(\nu+3)} - \frac{4n^2}{\sigma^2(\nu+1)^2(\nu+3)}\left(\frac{4}{n}\right)\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right]^{-1}.$$

This means that $\tilde{h}_1 \propto \sigma^{-2}$, so that

$$\tilde{h}_1^{\frac{1}{2}} \propto \sigma^{-1} = p(\sigma).$$

Consider

$$\begin{bmatrix} \tilde{F}_{11} & \tilde{F}_{12} \\ \tilde{F}_{21} & \tilde{F}_{22} \end{bmatrix} - \left(\tilde{F}_{33}\right)^{-1} \begin{bmatrix} \tilde{F}_{13} \\ \tilde{F}_{23} \end{bmatrix} \begin{bmatrix} \tilde{F}_{31} & \tilde{F}_{32} \end{bmatrix} = \begin{bmatrix} \tilde{H}_{11} & \tilde{H}_{12} \\ \tilde{H}_{21} & \tilde{H}_{22} \end{bmatrix}.$$

Therefore,

$$\tilde{H} = \begin{bmatrix} \frac{2n\nu}{\sigma^2(\nu+3)} & \frac{-2n}{\sigma(\nu+1)(\nu+3)} \\ \frac{-2n}{\sigma(\nu+1)(\nu+3)} & \frac{n}{4}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right] \end{bmatrix} - \frac{\sigma^2(\nu+3)}{n(\nu+1)} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \end{bmatrix},$$

and hence

$$\tilde{h}_2^{\frac{1}{2}} = \tilde{H}_{22}^{\frac{1}{2}} \propto \left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right]^{\frac{1}{2}}.$$

We also have that $\tilde{h}_3^{\frac{1}{2}} = \left(\tilde{F}_{33}\right)^{\frac{1}{2}} \propto c$, so that

$$p_1(\sigma,\nu,\mu) \propto \sigma^{-1}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right]^{\frac{1}{2}}.$$

By using the Fisher information matrices $I(\mu,\sigma,\nu)$ and $I(\sigma,\mu,\nu)$ it can be shown that $p_1(\sigma,\mu,\nu)$ is also a reference prior for the orderings $\{\mu,\sigma,\nu\}$ and $\{\sigma,\mu,\nu\}$.

■

Proof of Theorem 2.2. To derive the probability-matching prior $P_M(\nu,\mu,\sigma)$, we need the inverse of the Fisher information matrix,

$$I^{-1}(\mu,\sigma,\nu) = \begin{bmatrix} \frac{\sigma^2(\nu+3)}{n(\nu+1)} & 0 & 0 \\ 0 & \frac{n}{4D}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right] & \frac{2n}{D\sigma(\nu+1)(\nu+3)} \\ 0 & \frac{2n}{D\sigma(\nu+1)(\nu+3)} & \frac{2n\nu}{D\sigma^2(\nu+3)} \end{bmatrix},$$

where

$$D = \frac{n^2\nu}{2\sigma^2(\nu+3)}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right] - \frac{4n^2}{\sigma^2(\nu+1)^2(\nu+3)^2}$$
$$= \frac{n^2\nu}{2\sigma^2(\nu+3)}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right].$$

Let $t(\boldsymbol{\theta}) = \nu$, where $t(\boldsymbol{\theta})$ is the parameter of interest. From this it follows that $\frac{\partial t(\boldsymbol{\theta})}{\partial \nu} = 1$, $\frac{\partial t(\boldsymbol{\theta})}{\partial \mu} = 0$, $\frac{\partial t(\boldsymbol{\theta})}{\partial \sigma} = 0$, and

$$\nabla_t'(\boldsymbol{\theta}) = \left[\begin{array}{ccc} \frac{\partial t(\boldsymbol{\theta})}{\partial \mu} & \frac{\partial t(\boldsymbol{\theta})}{\partial \sigma} & \frac{\partial t(\boldsymbol{\theta})}{\partial \nu} \end{array}\right] = \left[\begin{array}{ccc} 0 & 0 & 1 \end{array}\right].$$

Therefore,

$$\nabla_t'(\boldsymbol{\theta})\, I^{-1}(\boldsymbol{\theta}) = \left[\begin{array}{ccc} 0 & \frac{2n}{D\sigma(\nu+1)(\nu+3)} & \frac{2n\nu}{D\sigma^2(\nu+3)} \end{array}\right],$$

which means that

$$\left[\nabla_t'(\boldsymbol{\theta})\, I^{-1}(\boldsymbol{\theta})\, \nabla_t(\boldsymbol{\theta})\right]^{\frac{1}{2}} = \left(\frac{2n\nu}{D\sigma^2(\nu+3)}\right)^{\frac{1}{2}}.$$

Hence,

$$\zeta'(\boldsymbol{\theta}) = \frac{\nabla_t'(\boldsymbol{\theta})\, I^{-1}(\boldsymbol{\theta})}{\left[\nabla_t'(\boldsymbol{\theta})\, I^{-1}(\boldsymbol{\theta})\, \nabla_t(\boldsymbol{\theta})\right]^{\frac{1}{2}}}$$
$$= \left[\begin{array}{ccc} \zeta_1(\boldsymbol{\theta}) & \zeta_2(\boldsymbol{\theta}) & \zeta_3(\boldsymbol{\theta}) \end{array}\right]$$
$$= \left[\begin{array}{ccc} 0 & \frac{(2n)^{\frac{1}{2}}}{D^{\frac{1}{2}}\nu^{\frac{1}{2}}(\nu+1)(\nu+3)^{\frac{1}{2}}} & \frac{(2n\nu)^{\frac{1}{2}}}{D^{\frac{1}{2}}\sigma(\nu+3)^{\frac{1}{2}}} \end{array}\right].$$

This indicates that the probability-matching prior is

$$p_M(\boldsymbol{\theta}) = p_M(\nu,\mu,\sigma) \propto D^{\frac{1}{2}}\frac{(\nu+3)^{\frac{1}{2}}}{\nu^{\frac{1}{2}}}$$
$$\propto \sigma^{-1}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right]^{\frac{1}{2}},$$

because the differential equation $\frac{\partial}{\partial\mu}[\zeta_1(\boldsymbol{\theta})p(\boldsymbol{\theta})] + \frac{\partial}{\partial\sigma}[\zeta_2(\boldsymbol{\theta})p(\boldsymbol{\theta})] + \frac{\partial}{\partial\nu}[\zeta_3(\boldsymbol{\theta})p(\boldsymbol{\theta})] = 0$. The probability-matching prior is therefore the same as the reference priors for the orderings $\{\nu,\mu,\sigma\}$, $\{\mu,\nu,\sigma\}$, and $\{\nu,\sigma,\mu\}$. ■

B. Reference prior behaviour

Proof of Theorem 2.3. The proof is the same as that of Corollary 1 in Fonseca et al. (2008). Consider

$$[p_2(\nu)]^2 = \left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right].$$

By using Stirling's asymptotic formula $\Psi'(a) \approx a^{-1} + (2a^2)^{-1}$, for large a it follows that

$$\Psi'\left(\frac{\nu}{2}\right) \approx \left(\frac{\nu}{2}\right)^{-1} + \left[2\left(\frac{\nu}{2}\right)^2\right]^{-1} = \frac{2}{\nu} + \frac{2}{\nu^2}$$

and

$$\Psi'\left(\frac{\nu+1}{2}\right) \approx \frac{2}{\nu+1} + \frac{2}{(\nu+1)^2}.$$

Therefore,

$$\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) \approx \frac{2\nu^2+6\nu+2}{\nu^2(\nu+1)^2}$$

and

$$[p_2(\nu)]^2 = \frac{2\nu^2+6\nu+2}{\nu^2(\nu+1)^2} - \frac{2(\nu+3)}{\nu(\nu+1)^2} = \frac{2}{\nu^2(\nu+1)^2} = O\left(\nu^{-4}\right).$$

Therefore, $p_2(\nu) = O(\nu^{-2})$ as $\nu \to \infty$.

In a similar way it can be proved that $[p_1(\nu)]^2 = 2(5\nu+3)/\nu^2(\nu+1)^2(\nu+3)$, which means that $p_1(\nu) = O(\nu^{-2})$ as $\nu \to \infty$. ∎

Proof of Theorem 2.4. The proof will be given for $p_1(\nu)$. The proof for $p_2(\nu)$ follows in a similar way. The posterior for ν is as follows:

$$p_1(\nu|data) = \tilde{k}\left[\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right]^{\frac{1}{2}} \times \frac{\Gamma\left(\frac{\nu+1}{2}\right)^n \nu^{n\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)^n \Gamma\left(\frac{1}{2}\right)^n}\left[\prod_{i=1}^n\left(\nu + x_i^2\right)\right]^{-\frac{1}{2}(\nu+1)},$$

where $\tilde{k}$ is the normalising constant. We then have that

$$p_1(\nu|data) \le \tilde{k}\left[\Psi'\left(\frac{\nu}{2}\right)\right]^{\frac{1}{2}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)^n \nu^{n\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)^n\left(\sqrt{\pi}\right)^n}\left[\prod_{i=1}^n\left(\nu + x_i^2\right)\right]^{-\frac{1}{2}(\nu+1)}.$$

Since $(\nu^\nu)^{n/2} \to 1^{n/2} = 1$ if $\nu \to 0^+$, it follows that, if $\nu \to 0^+$, then

$$\left[\prod_{i=1}^n\left(\nu + x_i^2\right)\right]^{-\frac{1}{2}(\nu+1)} \to \left[\prod_{i=1}^n x_i^2\right]^{-\frac{1}{2}}.$$

It is therefore only necessary to consider

$$\lim_{\nu\to 0^+} \frac{[\Psi'(\nu)]^{\frac{1}{2}}\,\Gamma\left(\nu+\frac{1}{2}\right)^n}{\Gamma(\nu)^n}.$$

Since $\Psi(\nu) = \frac{d}{d\nu}[\ln\Gamma(\nu)] = \Gamma'(\nu)/\Gamma(\nu)$, it follows that

$$\Psi'(\nu) = \frac{\Gamma''(\nu)\Gamma(\nu) - [\Gamma'(\nu)]^2}{[\Gamma(\nu)]^2}.$$

Therefore,

$$\lim_{\nu\to 0^+} \frac{[\Psi'(\nu)]^{\frac{1}{2}}\left[\Gamma\left(\nu+\frac{1}{2}\right)\right]^n}{[\Gamma(\nu)^n]} = \lim_{\nu\to 0^+}\left\{\frac{\Gamma''(\nu)\Gamma(\nu) - [\Gamma'(\nu)]^2}{[\Gamma(\nu)]^2}\cdot\frac{\left[\Gamma\left(\nu+\frac{1}{2}\right)\right]^{2n}}{[\Gamma(\nu)]^{2n}}\right\}^{\frac{1}{2}}. \tag{2}$$

The following formulae are valid:

$$-\frac{\Gamma'(\nu)}{\Gamma(\nu)} = \frac{1}{\nu} + \gamma + \sum_{n=1}^{\infty}\left(\frac{1}{n+\nu} - \frac{1}{n}\right), \quad \nu > 0, \tag{3}$$

where $\gamma = 0.5772$ is Euler's constant. It can also be shown that if $\nu > 0$, we have that

$$\sum_{n=1}^{\infty}\left(\frac{1}{n+\nu} - \frac{1}{n}\right) = -\Psi(1+\nu) - \gamma$$

and

$$\frac{1}{\Gamma(\nu)} = \nu \exp\left[\gamma\nu - \sum_{k=2}^{\infty}\frac{(-1)^k\zeta(k)\nu^k}{k}\right], \tag{4}$$

where $\zeta(k)$ is Riemann's zeta function. Therefore (3) $\times$ (4) gives

$$-\frac{\Gamma'(\nu)}{[\Gamma(\nu)]^2} = [1 + \nu\gamma - \nu\Psi(1+\nu) - \nu\gamma] \times \exp\left[\gamma\nu - \sum_{k=2}^{\infty}\frac{(-1)^k\zeta(k)\nu^k}{k}\right].$$

Since $\Psi(1) = -\gamma$, it follows that, as $\nu \to 0^+$,

$$-\frac{\Gamma'(\nu)}{[\Gamma(\nu)]^2} \to 1.$$

Therefore,

$$\frac{\Gamma'(\nu)}{[\Gamma(\nu)]^2} \to -1. \tag{5}$$

From (3) it follows that

$$-\Gamma'(\nu) = \Gamma(\nu)\left[\frac{1}{\nu} + \gamma - \Psi(1+\nu) - \gamma\right] = \Gamma(\nu)\left[\frac{1}{\nu} - \Psi(1+\nu)\right], \quad \nu > 0.$$

Therefore,

$$-\Gamma''(\nu) = \Gamma'(\nu)\left[\frac{1}{\nu} - \Psi(1+\nu)\right] + \Gamma(\nu)\left[-\frac{1}{\nu^2} - \Psi'(1+\nu)\right]$$

and

$$\frac{-\Gamma''(\nu)}{[\Gamma(\nu)]^3} = \frac{\Gamma'(\nu)}{[\Gamma(\nu)]^2} \cdot \frac{1}{\Gamma(\nu)} \left[\frac{1}{\nu} - \Psi(1+\nu) \right] + \frac{1}{[\Gamma(\nu)]^2} \left[-\frac{1}{\nu^2} - \Psi'(1+\nu) \right].$$

Remember that $\Psi'(1) = \frac{1}{6}\pi^2$. By making use of (5) and the fact that $\nu\Gamma(\nu) \to 1$ as $\nu \to 0^+$, it follows that

$$\frac{-\Gamma''(\nu)}{[\Gamma(\nu)]^3} \to (-1)(1-0) + (-1-0) = -2.$$

Therefore, as $\nu \to 0^+$,

$$\frac{\Gamma''(\nu)}{[\Gamma(\nu)]^3} \to 2. \tag{6}$$

Substitute (6) into (2) and assume that $n \geq 2$. Then,

$$\lim_{\nu \to 0^+} \left\{ \left[\frac{\Gamma''(\nu)}{\Gamma(\nu)[\Gamma(\nu)]^2} - \frac{\Gamma'(\nu)}{[\Gamma(\nu)]^2} \cdot \frac{\Gamma'(\nu)}{[\Gamma(\nu)]^2} \right] \cdot \frac{\left[\Gamma\left(\nu + \frac{1}{2}\right)\right]^{2n}}{[\Gamma(\nu)]^{2n-2}} \right\}^{\frac{1}{2}} = \left\{ \left[2 - (-1)^2 \right] \cdot 0 \right\}^{\frac{1}{2}} = 0,$$

which follows from the fact that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $2n - 2 > 0$, therefore

$$\frac{\left[\Gamma\left(\nu + \frac{1}{2}\right)\right]^{2n}}{[\Gamma(\nu)]^{2n-2}} \to \frac{\left(\sqrt{\pi}\right)^{2n}}{\infty} = 0,$$

if $n \geq 2$. This means that $p_1(\nu|data) \to 0$ if $\nu \to 0^+$. A similar proof can be made for $p_2(\nu|data)$. ■

C. Gibbs sampling

If $x_i|\mu, \sigma^2, \lambda_i \sim N(\mu, \sigma^2/\lambda_i), i = 1, 2, \ldots, n$, and $\nu\lambda_i \sim \chi^2_\nu$, then $x_i|\mu, \sigma^2, \nu \sim t_\nu(\mu, \sigma^2)$. If the prior $p_l(\mu, \sigma^2, \nu) \propto \sigma^{-2} p_l(\nu), l = 1, 2, 3, 5, 6$, is used, then the following conditional posterior distributions can be derived:

$$\mu|\sigma^2, H, \mathbf{x} \sim N\left[(\mathbf{1}'H\mathbf{1})^{-1} \mathbf{1}'H\mathbf{x}, \sigma^2 (\mathbf{1}'H\mathbf{1})^{-1} \right], \tag{7}$$

where $\mathbf{1} = [1, 1, \ldots, 1]'$, $\mathbf{x} = [x_1, x_2, \ldots, x_n]'$, and $H = \text{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n]'$;

$$\sigma^2|\mu, H, \mathbf{x} \sim \frac{(\mathbf{x} - \mu\mathbf{1})' H (\mathbf{x} - \mu\mathbf{1})}{\chi^2_n}; \tag{8}$$

$$\lambda_i|\mu, \sigma^2, \nu, x_i \sim \frac{\chi^2_{\nu+1}}{\nu + \left(\frac{x_i - \mu}{\sigma}\right)^2}, i = 1, 2, \ldots, n; \tag{9}$$

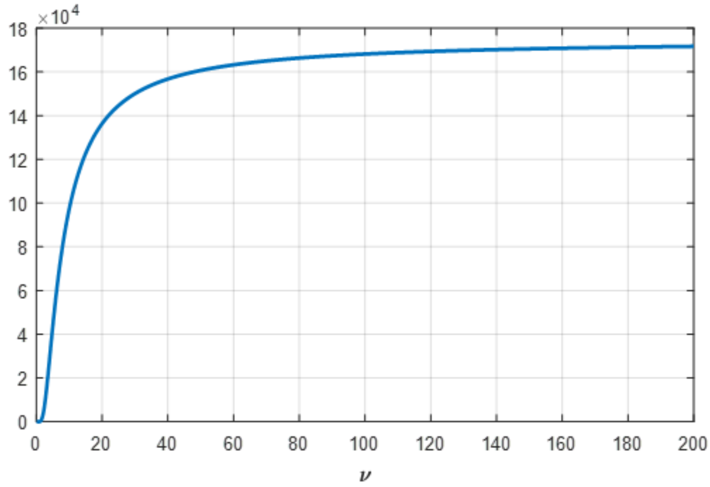

Figure 10. Likelihood function for ν, first sample.

and

$$p\left(\nu|\lambda_1,\lambda_2,\ldots,\lambda_n\right) \propto \frac{\nu^{n\nu/2}}{2^{n\nu/2}\left[\Gamma\left(\frac{\nu}{2}\right)\right]^n}\left[\prod_{i=1}^{n}\lambda_i^{\frac{1}{2}(\nu-1)}\right]e^{-\frac{\nu}{2}\sum_{i=1}^{n}\lambda_i}p_j\left(\nu\right). \tag{10}$$

By using (7), (8), (9) and (10) and then Gibbs sampling, the unconditional posterior distributions of μ, σ^2 and ν can be obtained.

In the case of $p_4(\mu,\sigma^2,\nu) \propto \sigma^{-3}p_4(\nu)$, the degrees of freedom of the chi-square distribution in (8) changes to $n+1$.

D. Complete simulation results

For transparency, the simulation data used for Figures 2 to 4 are given in Tables 7 to 9.

E. Behaviour of the likelihood

During the course of the compilation of this paper, a reviewer had some questions concerning the ill behaviour of a likelihood, as mentioned in Section 1. A simulation study is presented here to explain the matter. In particular, we will discuss the behaviour of the likelihood function and the posterior distributions for two prior distributions, namely $p_2\left(\nu\right)$ and $p_5\left(\nu\right)$, the prior distributions that seemed to perform best in Section 3 for intermediate and extreme degrees of freedom, respectively.

The data in Table 10 represent a sample of $n = 30$ observations drawn from a standard univariate Student t-distribution with $\nu = 3$ degrees of freedom.

In Figure 10 the strange behaviour of the likelihood function for ν is illustrated. All the observations except one (-2.5400) are concentrated around zero and this might be the reason for this behaviour. In other words the data in Table 10 are of such a nature that it is impossible for the likelihood function to distinguish between a normal distribution and a (heavy-tailed) t-distribution.

According to Fonseca et al. (2008), maximum likelihood estimation for the Student t-distribution

Table 7. Relative root mean squared errors ($\sqrt{MSE\ (\nu)}/\nu$) for six priors for ν.

ν	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**	**Prior 5**	**Prior 6**
1	0.4593	0.4138	0.4707	0.4416	0.5703	0.6153
2	0.8714	0.8408	0.9546	0.9650	1.1661	1.1180
3	0.9264	0.9120	1.0001	1.0280	1.1807	1.1226
4	0.7946	0.7779	0.8377	0.8840	0.9811	1.0013
5	0.6995	0.6991	0.7560	0.7768	0.8525	0.9068
6	0.6202	0.5825	0.6501	0.6563	0.7110	0.7649
7	0.5062	0.5014	0.5335	0.5565	0.5705	0.6498
8	0.4416	0.4316	0.4495	0.4805	0.4620	0.5320
9	0.4055	0.3975	0.4058	0.4268	0.3916	0.4674
10	0.3899	0.3850	0.3871	0.3928	0.3519	0.4180
11	0.4022	0.3961	0.3905	0.3860	0.3367	0.3893
12	0.3982	0.3909	0.3841	0.3802	0.3214	0.3760
13	0.4098	0.4138	0.4069	0.3961	0.3378	0.3698
14	0.4174	0.4341	0.4181	0.4053	0.3510	0.3648
15	0.4385	0.4528	0.4327	0.4242	0.3685	0.3797
16	0.4583	0.4703	0.4508	0.4332	0.3873	0.3942
17	0.4789	0.4918	0.4724	0.4573	0.4143	0.4160
18	0.5107	0.5155	0.5050	0.4819	0.4486	0.4425
19	0.5178	0.5291	0.5021	0.4953	0.4534	0.4510
20	0.5403	0.5473	0.5353	0.5217	0.4875	0.4761
21	0.5581	0.5662	0.5479	0.5359	0.5023	0.4871
22	0.5646	0.5778	0.5541	0.5474	0.5133	0.5032
23	0.5834	0.5896	0.5725	0.5626	0.5327	0.5204
24	0.5978	0.6101	0.5902	0.5764	0.5541	0.5386
25	0.6064	0.6153	0.5921	0.5816	0.5582	0.5418
Mean	**0.5439**	**0.5417**	**0.5520**	**0.5517**	**0.5522**	**0.5699**

Table 8. Coverage percentages of the 95% credibility intervals for ν.

ν	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**	**Prior 5**	**Prior 6**
1	94.10	94.50	94.20	93.90	91.15	92.75
2	94.25	94.80	93.80	93.95	86.55	92.00
3	97.55	97.15	97.20	96.45	88.65	96.70
4	97.75	97.70	98.60	97.70	97.70	98.80
5	97.10	97.20	97.30	98.05	99.30	98.40
6	97.30	97.15	97.60	98.00	99.40	98.25
7	97.90	97.55	98.20	97.95	99.75	98.55
8	97.30	97.60	97.80	97.75	99.75	98.30
9	98.35	97.65	98.20	97.75	99.30	98.60
10	97.70	97.05	97.80	98.25	98.85	98.20
11	96.90	96.60	97.20	98.30	98.90	97.75
12	97.60	97.15	97.90	98.45	99.75	97.95
13	97.55	97.45	97.80	97.85	99.65	98.55
14	97.80	97.45	98.10	98.10	99.35	98.65
15	97.85	96.80	97.60	98.15	99.10	98.45
16	97.85	97.10	98.50	98.60	99.10	98.35
17	97.75	97.95	98.20	97.85	99.30	98.65
18	97.10	96.30	97.00	97.80	98.90	98.30
19	97.65	97.30	98.00	97.70	99.00	98.25
20	96.80	97.00	96.60	97.80	98.50	97.40
21	95.90	96.30	96.50	97.70	98.10	97.50
22	97.00	96.10	97.50	97.60	98.80	98.15
23	97.15	97.10	98.00	97.05	98.70	97.85
24	96.55	95.55	96.70	98.45	97.40	97.75
25	97.20	97.10	98.40	97.35	99.00	98.60
Mean	**97.118**	**96.864**	**97.388**	**97.540**	**97.758**	**97.708**

Table 9. Average interval lengths of the 95% credibility intervals for ν.

ν	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**	**Prior 5**	**Prior 6**
1	1.4863	1.4705	1.5211	1.6613	1.6616	1.7346
2	10.9029	10.2188	11.4415	11.4660	8.7877	13.1349
3	27.2776	25.6413	28.6591	27.4616	16.8940	31.0931
4	40.8181	39.1123	42.8603	46.0271	22.6912	49.8889
5	58.0892	54.0122	60.7860	57.8947	27.6600	70.0064
6	66.8981	64.0113	69.8454	70.4983	29.9378	80.6659
7	74.5588	72.6509	77.8293	78.6948	32.0676	91.6501
8	81.3848	78.6516	84.8817	88.3929	33.6076	98.8977
9	84.4323	80.7490	87.9760	94.3561	34.2755	105.0209
10	90.9408	87.5548	94.5730	96.4120	35.2627	109.2556
11	91.3351	87.6219	94.9579	104.1987	35.2547	110.6545
12	97.0578	93.8699	100.7701	103.0313	36.2704	114.0066
13	96.1922	93.0008	99.8689	105.4707	36.1493	117.2605
14	99.1662	95.2863	102.9146	110.7050	36.6309	122.5355
15	101.8962	97.7551	105.7096	108.7843	37.0405	124.0829
16	103.5499	100.0941	107.4031	115.3113	37.3834	126.0403
17	104.6025	100.4079	108.4889	114.0209	37.5741	126.7225
18	102.7811	99.5750	106.6120	115.6712	37.2385	126.3335
19	108.6713	104.4110	112.4989	115.4082	38.0129	129.7908
20	105.8102	102.0438	109.6180	115.2661	37.5957	129.2041
21	107.9256	104.7400	111.7811	115.8514	37.8954	131.9269
22	110.7806	106.7750	114.5896	116.6591	38.3138	132.9114
23	111.1055	107.3003	114.9879	117.8916	38.4163	133.3977
24	110.3250	105.5900	114.1435	120.3860	38.2195	133.3991
25	114.3742	110.4022	118.2419	122.1614	38.8475	137.3235
Mean	**84.0945**	**80.9178**	**87.3184**	**90.9473**	**32.1475**	**101.8775**

Table 10. First sample of $n = 30$ observations from a t-distribution with $\mu = 0$, $\sigma^2 = 1$, and $\nu = 3$.

0.1189	0.1150	0.8309	1.0141	−0.9620	0.4662
−2.5400	0.3361	0.8219	0.3253	−0.6136	0.6204
−0.5659	0.3238	−0.1484	−0.8110	0.2741	−0.4700
−0.6983	−1.0854	0.2471	1.1470	−0.3032	0.1980
0.4189	0.3930	−0.2992	1.0051	−0.2657	0.4598

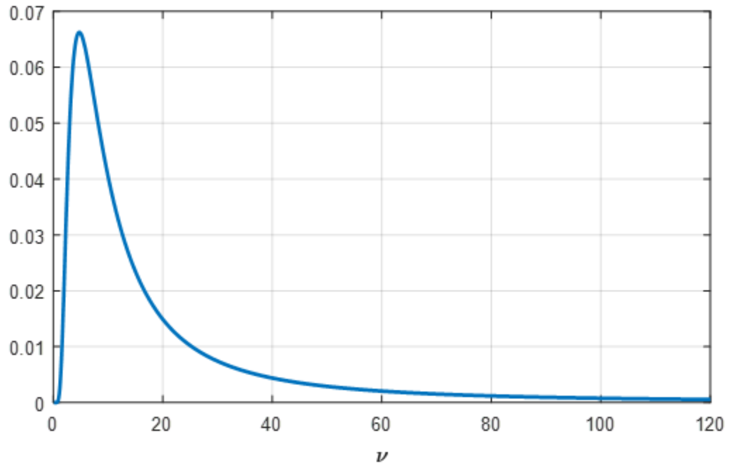

Figure 11. Posterior of ν with prior $p_2(\nu)$, first sample. Mean = 22.80; median = 11.57; mode = 4.75; Var = 905.61; 95% equal-tail interval = $(2.48, 121.76)$, length = 119.28; 95% HPD interval = $(1.72, 92.40)$, length = 90.68.

is very problematic because the likelihood function is ill-behaved for ν close to zero and may be ill-behaved when $\nu \to \infty$.

Our experience is that for a sample size of $n = 30$ and degrees of freedom $\nu = 3$, the poor behaviour of the likelihood function occurs in more than 50% of the simulation runs.

Consider the following two priors: (i) $p_2(\nu) \propto [\Psi'(\frac{1}{2}\nu) - \Psi'(\frac{1}{2}(\nu+1)) - 2(\nu+3)/\nu(\nu+1)^2]^{1/2}$, and (ii) $p_5(\nu) \propto e^{-0.1\nu}$. As mentioned before, $p_2(\nu)$ is a reference prior and $p_5(\nu)$ is the exponential prior.

In Figure 11 the posterior distribution $p_2\,(\nu|x_1,\ldots,x_{30})$ for the data in Table 10 is illustrated and in Figure 14 the posterior distribution $p_5\,(\nu|x_1,\ldots,x_{30})$ for the exponential prior is shown.

From Figure 11 it can be seen that the mode of 4.75 corresponds reasonably well with the true parameter value of $\nu = 3$. Also, the 95% credibility intervals contain the true parameter value of $\nu = 3$. In the case of the exponential prior (Figure 12) the mode of 8.72 is two and a half times as large as the true parameter and is therefore not a good estimate thereof. The true parameter value is also not contained in the 95% equal-tailed interval. The behaviour of the likelihood function is also the reason for the large interval lengths of the 95% credibility intervals.

In Table 11, a second sample of $n = 30$ observations drawn from a standard univariate Student t-distribution with $\nu = 3$ degrees of freedom is illustrated and in Figure 13 the likelihood function is given.

Figure 13 is an example of a likelihood function that behaves well. The largest values (3.8614, 7.3637) and smallest values (-4.6084, -4.6674) are an indication that the observations are sampled from a heavy-tailed distribution. The likelihood on its own, however, cannot be used as a posterior distribution because it is improper.

In Figures 14 and 15 the posterior distributions $p_2(\nu|x_1,\ldots,x_{30})$ and $p_5(\nu|x_1,\ldots,x_{30})$ are displayed for the second sample. It is clear from these figures that, although the 95% credibility intervals

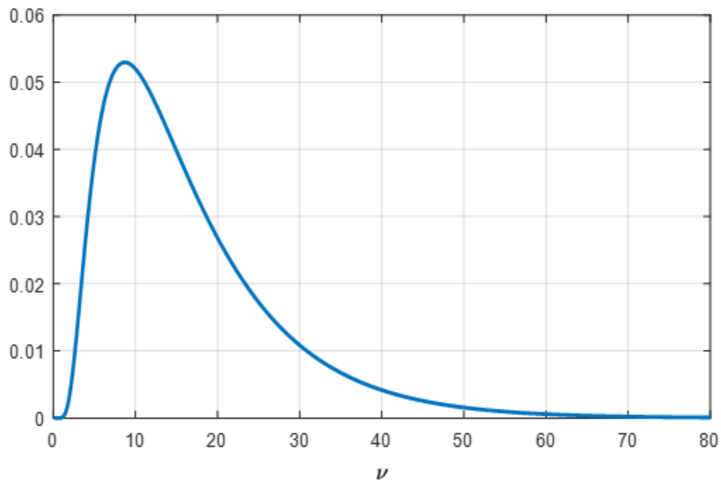

Figure 12. Posterior of ν with prior $p_5(\nu)$, first sample. Mean = 16.73; median = 13.88; mode = 8.72; Var = 122.45; 95% equal-tail interval = $(3.73, 45.39)$, length = 41.66; 95% HPD interval = $(2.55, 39.40)$, length = 36.85.

Table 11. Second sample of $n = 30$ observations from a t-distribution with $\mu = 0$, $\sigma^2 = 1$, and $\nu = 3$.

3.8614	−1.1644	1.5364	−1.9872	0.9092	0.2249
1.1924	0.5385	0.8573	1.1897	1.1392	−0.1857
0.8187	7.3637	−1.9517	0.6999	0.4705	−0.6518
1.1640	0.7606	−0.7805	0.9881	−4.6084	−0.8601
−0.6030	−1.1308	−4.6674	−1.5709	−1.3119	−0.8824

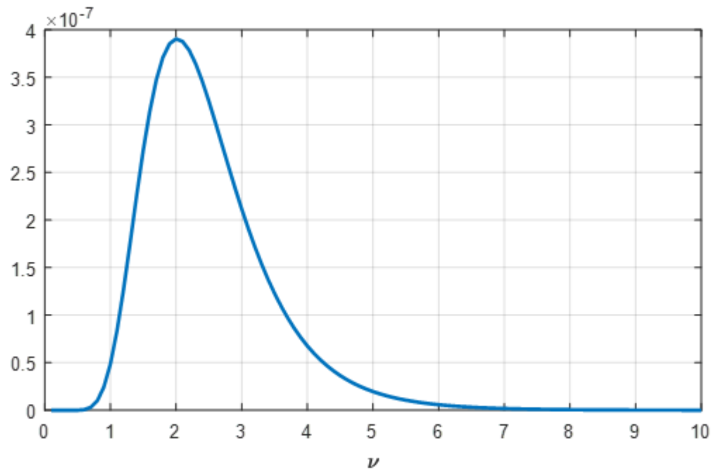

Figure 13. Likelihood function for ν, second sample.

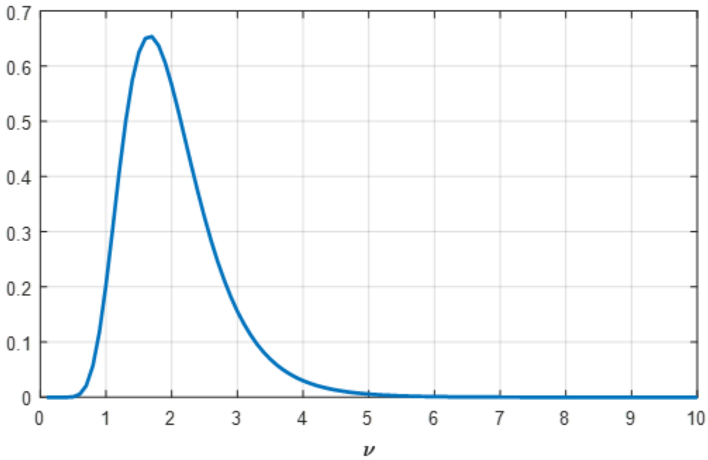

Figure 14. Posterior of ν with prior $p_2(\nu)$, second sample. Mean = 2.02; median = 1.84; mode = 1.66; Var = 0.55; 95% equal-tail interval = $(0.92, 3.77)$, length = 2.85; 95% HPD interval = $(0.86, 3.56)$, length = 2.71.

are now much shorter, they do cover the true parameter value.

As mentioned before, the focus of our simulation study in Section 3.2 is on the relative square rooted mean squared error from the median of the posterior distribution of ν and on the frequentist coverage percentages of the 95% equal-tail credibility intervals for samples of size $n = 30$ based on 2 000 simulations. The interval lengths are also considered. The relative square rooted mean squared error is defined by $\sqrt{MSE(\nu)}/\nu$, where $MSE = E(\nu - m)^2$ and m is the median of the posterior distribution. The results in Table 7 are based on 2 000 simulations, but now consider only the two data sets given in Tables 10 and 11, for illustrative purposes. For the posterior distribution $p_2(\nu|x_1, \ldots, x_{30})$, $E(\nu - m)^2 = \frac{1}{2}[(3 - 11.57)^2 + (3 - 1.84)^2] \approx 37.4$, and the relative square rooted mean squared error is $\sqrt{MSE(\nu)}/\nu = \frac{1}{3}(37.4)^{1/2} \approx 2.04$. For the posterior distribution $p_5(\nu|x_1, \ldots, x_{30})$, $E(\nu - m)^2 = \frac{1}{2}[(3 - 13.88)^2 + (3 - 2.20)^2] \approx 59.51$, and the relative square rooted mean squared error is $\sqrt{MSE(\nu)}/\nu = \frac{1}{3}(59.51)^{1/2} \approx 2.57$.

Now let us turn to the results in Table 8, based on 2 000 simulations. For illustrative purposes again, for the two data sets in this appendix, the coverage percentages of the 95% equal-tail credibility intervals for ν are 100% for $p_2(\nu|x_1, \ldots, x_{30})$, and 50% for $p_5(\nu|x_1, \ldots, x_{30})$.

Now consider the results in Table 9 for the full 2 000 simulations. Again, for illustration, looking at only the two samples in this appendix, the average interval lengths of the 95% equal-tail credibility intervals are $\frac{1}{2}(119.28 + 2.85) \approx 61.07$ for $p_2(\nu|x_1, \ldots, x_{30})$, and $\frac{1}{2}(41.66 + 3.49) \approx 22.58$ for $p_5(\nu|x_1, \ldots, x_{30})$.

Finally, the modes in Table 2 are obtained from the histograms of the 2 000 interval lengths for each ν. It makes no sense to look at the modes for only two simulations, so we will not illustrate that here.

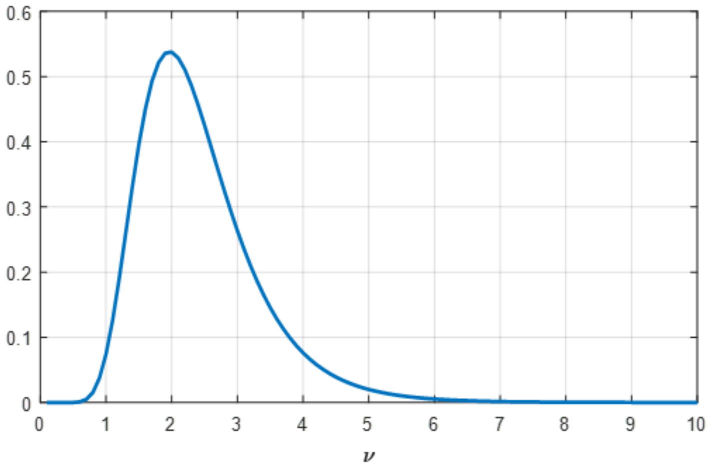

Figure 15. Posterior of ν with prior $p_5(\nu)$, second sample. Mean = 2.41; median = 2.20; mode = 1.96; Var = 0.83; 95% equal-tail interval = $(1.09, 4.59)$, length = 3.49; 95% HPD interval = $(1.01, 4.33)$, length = 3.32.

References

Berger, J. and Bernardo, J. (1992). On the development of reference priors. *Bayesian Statistics*, **4**, 35–60.

Blattberg, R. C. and Gonedes, N. J. (1974). A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business*, **47**, 244–280.

Box, G. E. P. and Tiao, G. C. (2011). *Bayesian Inference in Statistical Analysis*. Wiley, New York.

Croissant, Y. and Graves, S. (2020). *Ecdat: Data Sets for Econometrics*. R package version 0.3-9. URL: *https://CRAN.R-project.org/package=Ecdat*

Datta, G. and Ghosh, J. (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika*, **82**, 37–45.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, **93**, 359–371.

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.

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**, 252–252.

Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, **8**, S19–S40.

Gnanadesikan, R. and Kettenring, J. R. (2005). Bayesian computation for logistic regression. *Computational Statistics & Data Analysis*, **48**, 857–868.

He, D., Sun, D., and He, L. (2021). Objective Bayesian analysis for the Student-t linear regression. *Bayesian Analysis*, **16**, 129 – 145.

JUÁREZ, M. A. AND STEEL, M. F. J. (2010). Non-Gaussian dynamic Bayesian modelling for panel data. *Journal of Applied Econometrics*, **25**, 1128–1154.

KOTZ, S. AND NADARAJAH, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge University Press, Cambridge.

LIU, C. (1997). ML estimation of the multivariate t distribution and the EM algorithm. *Journal of Multivariate Analysis*, **63**, 296–312.

PEARN, W. AND WU, C. (2005). A Bayesian approach for assessing process precision based on multiple samples. *European Journal of Operational Research*, **165**, 685–695.

R CORE TEAM (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org/*

RUBIO, F. J. AND STEEL, M. F. J. (2015). Bayesian modelling of skewness and kurtosis with two-piece scale and shape distributions. *Electron. J. Statist.*, **9**, 1884–1912.

VALLEJOS, C. A. AND STEEL, M. F. J. (2013). On posterior propriety for the Student-t linear regression model under Jeffreys priors.
URL: *https://arxiv.org/abs/1311.1454v2*

VAN DER MERWE, A. J., VON MALTITZ, M. J., AND MEYER, J. H. (2021). Reference and probability-matching priors for the parameters of a univariate Student t-distribution.
URL: *http://arxiv.org/abs/2104.07386v1*

VANECK, T. W., RODRIGUEZ-ORTIZ, C. D., SCHMIDT, M. C., AND MANLEY, J. E. (1996). Automated bathymetry using an autonomous surface craft. *NAVIGATION*, **43**, 407–419.

VILLA, C. AND RUBIO, F. J. (2018). Objective priors for the number of degrees of freedom of a multivariate t distribution and the t-copula. *Computational Statistics & Data Analysis*, **124**, 197–219.

VILLA, C. AND WALKER, S. G. (2014). Objective prior for the number of degrees of freedom of a t distribution. *Bayesian Analysis*, **9**, 197–220.

VON MALTITZ, M. J. (2015). *Extending the Reach of Sequential Regression Multiple Imputation*. Ph.D. thesis, University of the Free State, Bloemfontein, South Africa.

WANG, M. AND YANG, M. (2016). Posterior property of Student-t linear regression model using objective priors. *Statistics & Probability Letters*, **113**, 23–29.

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.

Manuscript received 2021-04-28, revised 2022-02-14, accepted 2022-06-16.