

The Use of a Log-Normal Prior for the Student t -Distribution

Se Yoon Lee

Department of Statistics, Texas A&M University, College Station, TX 77843, USA; stat_math@tamu.edu

Abstract: It is typically difficult to estimate the number of degrees of freedom due to the leptokurtic nature of the Student t -distribution. Particularly in studies with small sample sizes, special care is needed concerning prior choice in order to ensure that the analysis is not overly dominated by any prior distribution. In this article, popular priors used in the existing literature are examined by characterizing their distributional properties on an effective support where it is desirable to concentrate on most of the prior probability mass. Additionally, we suggest a log-normal prior as a viable prior option. We show that the Bayesian estimator based on a log-normal prior compares favorably to other Bayesian estimators based on the priors previously proposed via simulation studies and financial applications.

Keywords: student t -distribution; log-normal distribution; small-sample study

MSC: 62C10; 62F15; 62F10

1. Introduction

Student's t -distribution [1] occurs frequently in statistics. Its usual derivation and utility is as the sampling distribution of certain test statistics derived based on normality [2]; however, over the past decades there has been growing interest in using the t -distribution as a heavy-tailed alternative to Gaussian distribution when robustness to possible outliers is a concern [3,4]. For example, it is widely known that the fluctuations in many financial time series are not normal [5,6]. As such, t -distribution is commonly used in finance and risk management, particularly to model asset or market index returns, for which the tails of the Gaussian distribution are almost invariably found to be too thin [7–11].

We assume that random variables x_i ($i = 1, 2, \dots$) are independently and identically distributed according to the Student t -distribution $t_\nu(x)$

$$t_\nu(x) = \frac{\Gamma((\nu + 1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < x < \infty \quad (1)$$

depending on the number of degrees of freedom $\nu \in \mathbb{R}^+$. Here, the notation \mathbb{R}^+ denotes the parameter space $(0, \infty)$, $\Gamma(a)$ represents the gamma function, and the parameter ν controls the heaviness of the tails of the density, including particular cases of $\nu = 1$, where the distribution coincides with the Cauchy density, and $\nu \rightarrow \infty$, where the distribution converges to the standard normal density.

In this paper, the number of degrees of freedom ν in Student t -distributions is the parameter of main interest. If a reasonable range of the degrees of freedom ν is known, the value ν can be used as a tuning parameter in robust statistical modeling [4,12]. However, there is often very limited knowledge about the degrees of freedom ν , and it may be desirable to estimate ν based only on observed data $\mathbf{x} = (x_1, x_2, \dots, x_N)$ that are believed to be independently sampled from the t -distribution $t_\nu(x)$ (1). In particular, it is widely known that in small-sample studies the accurate estimation of the degrees of freedom ν is very difficult within both frequentist and Bayesian settings (see [3,13,14] and references therein).



Citation: Lee, S.Y. The Use of a Log-Normal Prior for the Student t -Distribution. *Axioms* **2022**, *11*, 462. <https://doi.org/10.3390/axioms11090462>

Academic Editor: Jong-Min Kim

Received: 23 July 2022

Accepted: 5 September 2022

Published: 8 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

In this paper, we re-examine a fully Bayesian way of estimating the number of degrees of freedom ν and suggest a new Bayesian estimator (i.e., posterior mean) based on a log-normal prior. To our knowledge, no previous research has studied log-normal distribution as a viable prior option in this context; hence, studying the operating characteristics of the Bayes estimator based on a log-normal prior compared to those of several existing Bayes estimators is of interest in its own right. There are ample examples of priors used in diverse applications for estimating the degrees of freedom ν [3,15–17]. Broadly, these priors fall into two classes according to whether they are (i) elicited from certain parametric distributions (e.g., exponential and gamma distributions) [15,16] or are (ii) constructed by formal rules such as the Jeffreys rule [3,18–21]. Despite certain differences in the derivation procedures between these two classes of priors, to a certain extent the motivation for using such priors is to have a robust Bayes estimator against outliers, possibly equipped with the appearance of objectivity in the statistical analysis [22,23], even when the sample size is fairly or moderately small, say, $N = 30$ or 100 .

This article is organized as follows. In Section 2, we formulate an inference problem for fully Bayesian estimation of the degrees of freedom, investigate a sufficient condition to induce a valid posterior inference, and introduce popular previously used prior distributions from the literature. Section 3 provides a state-of-the-art sampling algorithm to compute the Bayes estimator based on a log-normal prior. In Section 4, numerical experiments are conducted for the sensitivity analysis associated with the log-normal prior and for the comparison of the small-sample performance of several Bayes estimators. The performances of these Bayes estimators are further compared through a real data application in Section 5. Finally, Section 6 concludes the article.

2. Bayesian Inference

2.1. Validity of Estimation of the Degrees of Freedom

The Bayesian inference for the estimation of the number of degrees of freedom $\nu \in \mathbb{R}^+$ commences with specifying a prior density function $\pi(\nu)$ supported on the parameter space \mathbb{R}^+ , followed by evaluating the posterior density function

$$\pi(\nu|\mathbf{x}) = \frac{p(\mathbf{x}|\nu) \cdot \pi(\nu)}{m(\mathbf{x})}, \quad \nu \in \mathbb{R}^+, \tag{2}$$

where $p(\mathbf{x}|\nu) = \prod_{i=1}^N t_\nu(x_i)$ represents the likelihood based on the t distribution $t_\nu(x)$ (1). The denominator in (2), $m(\mathbf{x}) = \int p(\mathbf{x}|\nu) \cdot \pi(\nu) d\nu$ is called the marginal likelihood of observations $\mathbf{x} = (x_1, x_2, \dots, x_N)$. To have a valid posterior inference, the marginal likelihood $m(\mathbf{x})$ should be finite for all \mathbf{x} . However, as the likelihood $p(\mathbf{x}|\nu)$ converges to a positive constant, as $\nu \rightarrow +\infty$ (more precisely, it holds $\lim_{\nu \rightarrow +\infty} p(\mathbf{x}|\nu) = \phi(\mathbf{x})$ uniformly for all \mathbf{x} where the $\phi(\mathbf{x})$ represents the density of the N -dimensional multivariate standard normal distribution; see Equation (1.1) from [24]), the propriety of the posterior is highly dependent on the rate of decay of the prior $\pi(\nu)$. As such, it is nontrivial whether $m(\mathbf{x})$ is finite when the likelihood is based on the distribution of t .

In the following, we show that when a prior $\pi(\nu)$ is proper and supported on the parameter space \mathbb{R}^+ , the posterior density $\pi(\nu|\mathbf{x})$ is as well. To show this, we first prove that the two functional components of the distribution of t (1) are bounded on \mathbb{R}^+ .

Lemma 1. Consider functions $g(\nu) = (1 + x^2/\nu)^{-(\nu+1)/2}$ and $h(\nu) = \Gamma((\nu + 1)/2) / \{\sqrt{\nu\pi} \Gamma(\nu/2)\}$ defined on the domain \mathbb{R}^+ . Then, functions g and h are upper bounded on the domain \mathbb{R}^+ .

Proof. Using elementary calculus, we can show that the following three properties hold for the function $g(\nu) = (1 + x^2/\nu)^{-(\nu+1)/2}$: (i) g is continuous on \mathbb{R}^+ ; (ii) $\lim_{\nu \rightarrow 0^+} g(\nu) = 0$; and (iii) $\lim_{\nu \rightarrow +\infty} g(\nu) = e^{-x^2/2} \leq 1$. Therefore, the function $g(\nu)$ is upper bounded on the domain \mathbb{R}^+ .

Next, we explore the three properties of the function $h(\nu)$ on the domain \mathbb{R}^+ . First, because the gamma function $\Gamma(a)$ is continuous on the domain \mathbb{R}^+ , the function

$h(\nu)$ on the domain is as well. Second, it holds that $\lim_{\nu \rightarrow 0^+} h(\nu) = 0$, because the nominator of h converges to the constant $\sqrt{\pi}$ and the denominator of h diverges to $+\infty$, as ν goes to $0+$. That is, $\lim_{\nu \rightarrow 0^+} \Gamma((\nu + 1)/2) = \sqrt{\pi}$ and $\lim_{\nu \rightarrow 0^+} \sqrt{\nu\pi} \cdot \Gamma(\nu/2) = +\infty$. The latter is true due to a property of the gamma function, $\Gamma(z) = 1/z - \gamma + (1/2)(\gamma^2 + \pi^2/6)z - (1/6)\{\gamma^3 + \gamma\pi^2/2 + 2\zeta(3)\}z^2 + O(z^3)$, where γ is the Euler–Mascheroni constant and $\zeta(z)$ is the Riemann zeta function [25]. Finally, using Stirling’s approximation [26], that is, $\Gamma(z) = \sqrt{2\pi/z} \cdot (z/e)^z \cdot (1 + O(1/z))$, with the nominator and denominator of $h(\nu)$, the following equality is obtained:

$$h(\nu) = \frac{\sqrt{\frac{4\pi}{\nu+1}} \cdot \left(\frac{\nu+1}{2e}\right)^{\frac{\nu+1}{2}} \cdot (1 + O(1/\nu))}{\sqrt{\nu\pi} \cdot \sqrt{\frac{4\pi}{\nu}} \cdot \left(\frac{\nu}{2e}\right)^{\frac{\nu}{2}} \cdot (1 + O(1/\nu))} = \frac{1}{\sqrt{2e\pi}} \cdot \sqrt{\left(1 + \frac{1}{\nu}\right)^\nu} \cdot \frac{1 + O(1/\nu)}{1 + O(1/\nu)}. \tag{3}$$

Because it generally holds that $\lim_{\nu \rightarrow +\infty} (1 + 1/\nu)^\nu = e$, we have $\lim_{\nu \rightarrow +\infty} h(\nu) = 1/\sqrt{2\pi}$. Thus, the function $h(\nu)$ is upper bounded on the domain \mathbb{R}^+ due to the three derived properties of the function. This ends the proof. \square

Lemma 1 states that the likelihood function based on Student’s t -distribution, that is, the function of ν with a fixed x , $l(\nu) = t_\nu(x)$ (1), is bounded over the parameter space \mathbb{R}^+ . Thus, we comment that the argument in [3] (“Unfortunately, the estimation of ν is not straightforward: the likelihood function tends to infinity as $\nu \rightarrow 0+$ ”) is not correct. In actuality, the likelihood function tends to zero, as $\nu \rightarrow 0+$.

We can now prove the main theorem.

Theorem 1. Suppose that $\mathbf{x} = (x_1, x_2, \dots, x_N)$ is a random sample of size N from the distribution of t (1) with degrees of freedom $\nu > 0$. Let $\pi(\nu)$ be a proper prior density supported on \mathbb{R}^+ . Then, the posterior density $\pi(\nu|\mathbf{x})$ is proper as well.

Proof. Under the formulation of Bayes’s theorem where $\pi(\nu|\mathbf{x}) = p(\mathbf{x}|\nu) \cdot \pi(\nu) / m(\mathbf{x})$, $\nu \in \mathbb{R}^+$, our eventual purpose is to prove that the marginal likelihood $m(\mathbf{x}) = \int p(\mathbf{x}|\nu) \cdot \pi(\nu) d\nu$ is finite for all values \mathbf{x} . By Lemma 1, there exists a constant C independent of ν such that

$$m(\mathbf{x}) = \int_0^\infty \prod_{i=1}^N \left[\frac{\Gamma((\nu + 1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \cdot \left(1 + \frac{x_i^2}{\nu}\right)^{-\frac{\nu+1}{2}} \right] \cdot \pi(\nu) d\nu \leq C \cdot \int_0^\infty \pi(\nu) d\nu.$$

Because the prior density $\pi(\nu)$ is proper (that is, $\int_0^\infty \pi(\nu) d\nu = 1$), the upper bound of the above inequality is finite. \square

Here, we briefly summarize theoretical results before moving to the next subsection. Generally, in most Bayesian statistical inference, a proper prior can lead to a proper posterior, particularly when Gaussian likelihood (which has very light tails) is assumed. However, this may not be obvious when dealing with a likelihood function based on a fat-tailed distribution, such as Generalized Pareto distributions [27,28], Student’s t -distribution [1], α -stable distribution [29], etc., that are frequently used in extreme value theory [30]. Theorem 1 states that a sufficient condition for a valid Bayesian inference in dealing with the t distribution is the propriety of the prior $\pi(\nu)$. On the other hand, for the case when the prior $\pi(\nu)$ is improper, several authors [3,18,19] stated that the propriety of the posterior is generally not guaranteed. Unfortunately, there is no general theorem providing simple condition under which an improper prior yields the propriety of the posterior for a particular model, and as such this must be investigated on a case-by-case basis [31].

2.2. Effective Support of the Degrees of Freedom

Given data $\mathbf{x} = (x_1, x_2, \dots, x_N) \sim t_\nu(x)$ (1) with small sample sizes, the performance of a Bayesian estimator for ν heavily relies on the suitable allocation of the prior probability mass over the support \mathbb{R}^+ . Ideally, the choice of prior is made such that most of the mass

is placed on an interval that contains a range of plausible values for ν that can generate the data \mathbf{x} before observing the data. Such an interval, denoted as ' $\mathcal{I}_e \subset \mathbb{R}^+$ ', is referred to as *effective support* of the degrees of freedom ν . The subscript ' e ' in the notation \mathcal{I}_e is noted as emphasizing 'effective'. Eventually, the prior probability on the effective support $\Pi[\nu \in \mathcal{I}_e] = \int_{\mathcal{I}_e} \pi(\nu) d\nu$ should be large enough to produce a Bayes estimator that performs well for ν .

Admittedly, a mathematical definition of the effective support \mathcal{I}_e in small-sample studies is not trivial, and may of course vary considerably for different values for observations \mathbf{x} and parameters ν (see the paper by [32] for a relevant theoretical discussion). Most works have used the interval $(0, 25) \subset \mathbb{R}^+$ as an effective support \mathcal{I}_e [3,18,19,33], although authors have used $(0, 20)$ or $(0, 30)$ as the effective support as well; throughout this paper, we use $\mathcal{I}_e = (0, 25)$. One conventional reason behind this argument is that in a small-sample case the observations $\mathbf{x} = (x_1, x_2, \dots, x_N)$ sampled from the t -distribution $t_\nu(x)$ (1) with the degrees of freedom set by either $\nu = 40$, $\nu = 50$, or any other value greater than 25 can be virtually regarded as the observations from a standard normal distribution. To our knowledge, however, no research works have statistically justified the use of $(0, 25)$, or, similarly, $(0, 20)$ or $(0, 30)$, as an effective support in the context of small-sample studies.

In this study, the Monte Carlo simulation method is used to examine the suitability of the interval $\mathcal{I}_e = (0, 25)$ for use as an effective support in small-sample cases. In general, Monte Carlo methods are widely used when the goal is to characterize certain statistical properties of a distribution under the finiteness of the sample size rather than resorting to a large-sample theory [34–37]. Experiments were designed as follows. With a choice from a list of sample size $N \in \{30, 50, 100, 200, 300, 400, 500\}$, we generated N observations $\mathbf{x} = (x_1, x_2, \dots, x_N) \sim t_{\nu_0}(x)$, with the true data generating parameter ν_0 selected from the interval $\mathcal{I}_e = (0, 25)$. Here, at each value ν_0 we simulated 100,000 replication data instances of observations \mathbf{x} . From each instance of replication data, we calculated the p -value of the Shapiro–Wilk test [38] in order to evaluate the normality of the replication data. The null and alternative hypotheses at each value $\nu_0 \in \mathcal{I}_e = (0, 25)$ are thus as follows:

- H_{0,ν_0} : a sample (x_1, x_2, \dots, x_N) simulated from $t_{\nu_0}(x)$ with the truth $\nu_0 \in \mathcal{I}_e$ deriving from a normally distributed population;
- H_{a,ν_0} : a sample (x_1, x_2, \dots, x_N) simulated from $t_{\nu_0}(x)$ with the truth $\nu_0 \in \mathcal{I}_e$ does not derive from a normally distributed population.

Note that, under the above simulation setup, the source of the non-normality from the alternative hypothesis H_{a,ν_0} is mainly due to the heavy tails of observations t . Finally, we report the median value of the p -values obtained from the replication data at each true parameter ν_0 .

Figure 1 displays the results of our experiments. Using the significance level $\alpha = 0.05$ as the criterion value (shown as the dashed horizontal line in the panel), we calculate a threshold value $\tilde{\zeta}$, dividing the interval $\mathcal{I}_e = (0, 25)$ into two sub-intervals, $\mathcal{I}_e^1 = (0, \tilde{\zeta})$ and $\mathcal{I}_e^2 = (\tilde{\zeta}, 25)$. Then, the former interval \mathcal{I}_e^1 comprises the parameters ν_0 , generating heavy-tailed t observations, while the latter interval \mathcal{I}_e^2 comprises the parameters ν_0 , generating normal observations. Obviously, the threshold value $\tilde{\zeta}$ is a function of the significance level α and the sample size N ; as such, it can be denoted as $\tilde{\zeta} = \tilde{\zeta}(\alpha, N)$, although here we use $\tilde{\zeta}$ to avoid cluttered notation. Conceptually, the threshold value $\tilde{\zeta}$ may be interpreted as the transitional point at which the tail-thickness of N observations $\mathbf{x} = (x_1, x_2, \dots, x_N) \sim t_{\nu_0}(x)$ changes from heavy tails ($\nu_0 \leq \tilde{\zeta}$) to thin tails ($\nu_0 > \tilde{\zeta}$), resulting in the threshold values $\tilde{\zeta}$ are 2.73 ($N = 30$), 3.74 ($N = 50$), 5.52 ($N = 100$), 7.83 ($N = 200$), 9.61 ($N = 300$), 11.00 ($N = 400$), and 12.21 ($N = 500$).

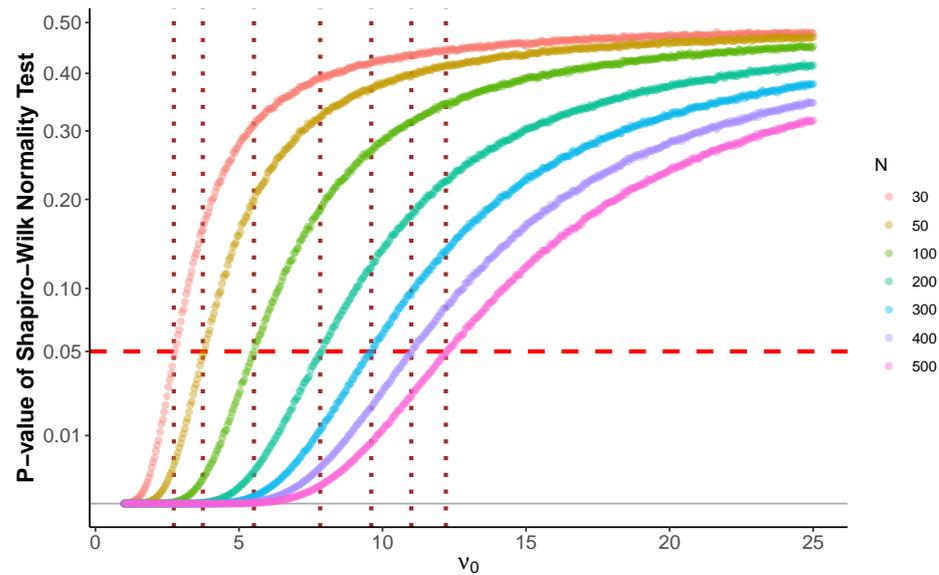


Figure 1. Median of the p -values of the Shapiro–Wilk test from 100,000 replicated observations $\mathbf{x} = (x_1, x_2, \dots, x_N) \sim t_{\nu_0}(x)$ with different sample size $N \in \{30, 50, 100, 200, 300, 400, 500\}$. The true data generating parameter ν_0 is selected from the effective support $\mathcal{I}_e = (0, 25)$.

The results of our Monte Carlo simulation experiments are summarized below:

1. As the sample size N increases from small sample sizes (i.e., $N \in \{30, 50, 100\}$) to moderate sample sizes (i.e., $N \in \{200, 300, 400, 500\}$), the threshold value $\tilde{\zeta}$ increases.
2. For the sample size $N = 30$ (and similarly for $N = 50, 100, 200, 300, 400$, and 500 , respectively), the values $\nu > 2.73$ (and similarly for $\nu > 3.74, 5.52, 7.83, 9.61, 11.00$, and 12.21 , respectively) generate the observations $\mathbf{x} = (x_1, x_2, \dots, x_N) \sim t_\nu(x)$, which are virtually distributed in a normal distribution (with the type I error 0.05).
3. The interval $\mathcal{I}_e = (0, 25)$ effectively covers a wide range of tail-thickness, from heavy-tailed to thin-tailed data, up to the maximum sample size $N = 500$ considered in the experiment. Thus, the interval $\mathcal{I}_e = (0, 25)$ can be used as an effective support for a small-sample study.

2.3. Prior Distributions for the Degrees of Freedom

Many prior distributions for the degrees of freedom ν suggested in the literature are proper distributions supported on the parameter space \mathbb{R}^+ [18,39,40]. In Theorem 1, we have shown that this is a sufficient condition for a valid posterior inference. Examples of popularly used proper priors are based on an exponential distribution [15] and a gamma distribution [16]. On the other hand, a Jeffreys prior, as suggested by [3], is improper, yet it has been shown that the posterior under the Jeffreys prior is proper. In the linear regression setup, the Jeffreys prior is called the independence Jeffreys prior [18]. Readers may refer to the papers [3,15,16] for the formulation and derivation of the priors. Considering the effective support $\mathcal{I}_e = (0, 25)$, as previously discussed, we aim to re-examine certain distributional properties of the three priors. Additionally, we study a log-normal distribution as a viable prior option. To the best of our knowledge, no previous study has reported the utility of a log-normal prior for robust Bayesian procedures. For the sake of readability, the four priors are denoted as $\pi_J(\nu)$ (4), $\pi_E(\nu)$ (5), $\pi_G(\nu)$ (6), and $\pi_L(\nu)$ (7), respectively, with the subscripts on the notations taken from the initials of the priors.

The following are the analytic expressions of the four priors:

- (a) Jeffreys prior [3]:

$$\pi_J(\nu) \propto \left(\frac{\nu}{\nu+3}\right)^{1/2} \left\{ \psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{1/2}, \quad \nu \in \mathbb{R}^+, \quad (4)$$

where $\psi(a) = d\{\log\Gamma(a)\}/da$ and $\psi'(a) = d\psi(a)/da$ are the digamma and trigamma functions, respectively. The authors of [3] developed the prior $\pi_J(\nu)$ as an objective prior on the basis of certain Jeffreys rules [23]. The Jeffreys prior may place enormously substantial mass close to the zero due to the asymptotic behavior $\pi(\nu) = O(\nu^{-1/2})$ as $\nu \rightarrow 0+$; see Corollary 1 from [3] for more details.

(b) Exponential prior [15]:

$$\pi_E(\nu) = \text{Exp}(\nu; 0.1) = \frac{1}{10}e^{-\nu/10}, \quad \nu \in \mathbb{R}^+. \tag{5}$$

The specification of the rate hyperparameter as 0.1 is recommended in [15] to avoid introducing strong prior information, for similar reasons as for using objective priors. The prior mean of ν equal to 10 and the prior variance of ν are equal to 100. Almost 92% of the prior mass is allocated to the effective support \mathcal{I}_e , $\Pi_E[\nu \in \mathcal{I}_e] = \int_0^{25} \pi_E(\nu)d\nu \approx 0.917$.

(c) Gamma prior [16]:

$$\pi_G(\nu) = \text{Ga}(\nu; 2, 0.1) = \frac{\nu}{100}e^{-\nu/10}, \quad \nu \in \mathbb{R}^+. \tag{6}$$

The authors of [16] recommend that the shape and rate hyperparameters be set to 2 and 0.1, respectively. Then, the prior mean and variance of ν are equal to 20 and 200, respectively. The gamma prior $\pi_G(\nu)$ (6) allocates nearly 70% of the prior mass to the effective support \mathcal{I}_e , $\Pi_G[\nu \in \mathcal{I}_e] = \int_0^{25} \pi_G(\nu)d\nu \approx 0.712$.

(d) Log-normal prior:

$$\pi_L(\nu) = \log \mathcal{N}(\nu; 1, 1) = \frac{1}{\nu\sqrt{2\pi}} \exp\left[-\frac{\{\log(\nu) - 1\}^2}{2}\right], \quad \nu \in \mathbb{R}^+. \tag{7}$$

We recommend setting the mean and variance hyperparameters to 1. These hyperparameters are specified on the basis of the sensitivity analysis in Section 4.1. The prior mean and variance of ν are $\exp(1 + 1/2) \approx 4.481$ and $\{\exp(1) - 1\} \cdot \exp(3) \approx 34.512$, respectively. The log-normal prior $\pi_L(\nu)$ (7) places nearly 99% of the prior mass on the effective support \mathcal{I}_e , $\Pi_L[\nu \in \mathcal{I}_e] = \int_0^{25} \pi_L(\nu)d\nu \approx 0.986$.

Table 1 summarizes the first to second moments and mass allocation of the three prior densities $\pi_E(\nu)$ (5), $\pi_G(\nu)$ (6), and $\pi_L(\nu)$ (7). As the Jeffreys prior $\pi_J(\nu)$ is improper, we do not report it. Recall that in a small-sample study the transition from fat-tailed to normal-tailed t -distributed data is typically manifested on the effective interval $\mathcal{I}_e = (0, 25)$ (refer to Figure 1). Therefore, in order to achieve robust Bayesian procedures to dynamically accommodate data with a wide range of tail thicknesses, it is desirable that most of the probability of the prior mass is reasonably placed on the effective support \mathcal{I}_e . It is notable that the log-normal and exponential priors place substantial mass (98.6% and 91.7%) on the effective support, while the gamma prior places only 71.2% of the probability mass on the effective support. In Section 4.2, we conduct simulation studies to investigate the performances of Bayes estimators based on the priors.

Table 1. Characteristics of prior densities on the interval $\mathcal{I}_e = (0, 25)$.

Prior	Mean $\mathbb{E}[\nu]$	Variance $\mathbb{V}[\nu]$	$\Pi[\nu \in (0, 10)]$	$\Pi[\nu \in (10, 25)]$	$\Pi[\nu \in \mathcal{I}_e]$
$\pi_E(\nu)$	10	100	0.632	0.285	0.917
$\pi_G(\nu)$	20	200	0.264	0.448	0.712
$\pi_L(\nu)$	4.481	34.512	0.903	0.083	0.986

NOTE: Prior mass characteristics of the Jeffreys prior $\pi_J(\nu)$ are not reported due to its impropriety.

3. Posterior Computation Using Log-Normal Priors

3.1. Elliptical Slice Sampler

In this section, we propose an efficient Markov chain Monte Carlo (MCMC) method to sample from the posterior density $\pi(v|\mathbf{x})$ (2) when provided with the log-normal prior $\pi_L(v)$ (7). Due to the non-conjugacy when sampling from the density $\pi(v|\mathbf{x})$, the first solution is to consider the Metropolis-Hastings (MH) algorithm [41,42], the performance of which can depend highly on the choice of the proposal density [43]. Instead, we can use the elliptical slice sampler (ESS) algorithm [44], which is known to be efficient when the prior distribution is a normal distribution. Conceptually, the MH and ESS algorithms are similar in that both comprise two steps, namely, a proposal step and a criterion step. A main difference between the two algorithms arises in the criterion step. If a new candidate does not pass the criterion, then the MH algorithm takes the current state as the next state, whereas the ESS re-proposes a new candidate until rejection does not take place, rendering the algorithm rejection-free. Unlike the MH algorithm, which requires the proposal variance or density, ESS is fully automated, and no tuning is required.

To adapt the ESS to simulate a Markov chain from the posterior density $\pi(v|\mathbf{x})$ (2), we first need to transform $v \in \mathbb{R}^+$ to a real-valued parameter $\eta = \log v \in \mathbb{R}$:

$$\pi(\eta|\mathbf{x}) = \pi(v|\mathbf{x})|_{v=e^\eta} \cdot |dv/d\eta| \propto \mathcal{L}(e^\eta) \cdot \log \mathcal{N}(e^\eta; 1, 1) \cdot e^\eta = \mathcal{L}(e^\eta) \cdot \mathcal{N}(\eta; 1, 1), \quad \eta \in \mathbb{R}, \tag{8}$$

where $\mathcal{L}(v) = p(\mathbf{x}|v) = \prod_{i=1}^n t_v(x_i)$. ESS can be used to sample from the transformed target density $\pi(\eta|\mathbf{x})$ (8), after which the drawn sample should be transformed back to $v = \exp \eta \in \mathbb{R}^+$. Algorithm 1 details the ESS in an algorithmic form:

Algorithm 1: ESS to sample from $\pi(v|\mathbf{x})$ (2)

Goal : Sampling from the full conditional posterior distribution

$$\pi(v|\mathbf{x}) \propto \mathcal{L}(v) \cdot \log \mathcal{N}(v; 1, 1), \quad v \in \mathbb{R}^+,$$

where $\mathcal{L}(v) = p(\mathbf{x}|v) = \prod_{i=1}^n t_v(x_i)$.

Input : Current state $v^{(s)}$.

Output : A new state $v^{(s+1)}$.

- a. Change of variable ($\eta = \log v$): $\eta^{(s)} = \log v^{(s)}$.
- b. Choose an ellipse: $\rho \sim \mathcal{N}(1, 1)$.
- c. Define a criterion function: $\alpha(\eta, \eta^{(s)}) = \min\{\mathcal{L}(e^\eta) / \mathcal{L}(e^{\eta^{(s)}}), 1\} : \mathbb{R} \rightarrow [0, 1]$.
- d. Choose a threshold and fix: $u \sim \text{Unif}[0, 1]$.
- e. Draw an initial proposal η^* :

$$\phi \sim \text{Unif}(-\pi, \pi]; \eta^* = (\eta^{(s)} - 1) \cos \phi + (\rho - 1) \sin \phi + 1.$$

- f. **if** ($u < \alpha(\eta^*, \eta^{(s)})$) **{** $\eta^{(s+1)} = \eta^*$ **}** **else** **{**
 Define a bracket : $(\phi_{\min}, \phi_{\max}) = (-\pi, \pi)$.
while ($u \geq \alpha(\eta^*, \eta^{(s)})$) **{**
 Shrink the bracket and try a new point :
if ($\phi > 0$) $\phi_{\max} = \phi$ **else** $\phi_{\min} = \phi$
 $\phi \sim \text{Unif}(\phi_{\min}, \phi_{\max})$; $\eta^* = (\eta^{(s)} - 1) \cos \phi + (\rho - 1) \sin \phi + 1$
}
 $\eta^{(s+1)} = \eta^*$
}
 - g. Change of variable ($v = \exp \eta$): $v^{(s+1)} = \exp \eta^{(s+1)}$.
-

In Algorithm 1, the logarithm of the ratio part in the criterion function $\alpha(\eta, \eta^{(s)})$ can be detailed as follows:

$$\log \left(\frac{\mathcal{L}(e^\eta)}{\mathcal{L}(e^{\eta^{(s)}})} \right) = N \cdot \left[\log \left(\frac{\Gamma((e^\eta + 1)/2)}{\sqrt{e^\eta \pi} \cdot \Gamma(e^\eta/2)} \right) - \log \left(\frac{\Gamma((e^{\eta^{(s)}} + 1)/2)}{\sqrt{e^{\eta^{(s)}} \pi} \cdot \Gamma(e^{\eta^{(s)}}/2)} \right) \right] - \left[\frac{e^\eta + 1}{2} \sum_{i=1}^N \log \left(1 + \frac{x_i^2}{e^\eta} \right) - \frac{e^{\eta^{(s)}} + 1}{2} \sum_{i=1}^N \log \left(1 + \frac{x_i^2}{e^{\eta^{(s)}}} \right) \right].$$

To calculate the above quantity using R statistical software, the built-in function *lgamma* is recommended in order to produce stable calculation of the logarithm of the gamma function $\Gamma(a)$ evaluated at $(e^\eta + 1)/2$, $e^\eta/2$, $(e^{\eta^{(s)}} + 1)/2$, and $e^{\eta^{(s)}}/2$.

3.2. Bayesestdft R Package

We developed an R package called *bayesestdft* to provide Bayesian tools to estimate the degrees of freedoms by sampling from the posterior distribution $\pi(v|\mathbf{x}) \propto p(\mathbf{x}|v) \cdot \pi(v)$ (2) with the likelihood $p(\mathbf{x}|v) = \prod_{i=1}^n t_v(x_i)$, the log-normal prior $\pi(v) = \log \mathcal{N}(v; \mu, \sigma^2)$, and hyperparameters with mean μ and variance σ^2 . Note that with the specification $\mu = \sigma^2 = 1$ we have $\pi_L(v)$ (7). The function *BayesLNP*(*y, ini.nu, S, mu, rho.sq*) implements the ESS (Algorithm 1) with the following inputs:

- *y*: *N*-dimensional vector of continuous observations supported on \mathbb{R} , $\mathbf{x} = (x_1, \dots, x_N)$.
- *ini.nu*: the initial posterior sample value, $v^{(1)}$ (Default = 1).
- *S*: the number of posterior samples, *S* (Default = 1000).
- *mu*: mean of the log-normal prior density, μ (Default = 1).
- *sigma.sq*: variance of the log-normal prior density, σ^2 (Default = 1).

The output of the function *BayesLNP* is the *S*-dimensional vector of posterior samples, that is, $\{v^{(s)}\}_{s=1}^S$, and is drawn from the posterior density $\pi(v|\mathbf{x})$.

In order to demonstrate the estimation performance of the Bayes estimator based on the log-normal prior $\pi_L(v)$ (7), we conducted the following simulation experiments. We generated $N = 100$ observations $\mathbf{x} = (x_1, \dots, x_N) \sim t_{v_0}(x)$ with the truth value v_0 specified by $v_0 = 0.1, 1$, and 5 , respectively, and then estimated the parameter v using the function *BayesLNP* for each of the three scenarios. To that end, we used the following command:

```
R> library(devtools)
R> devtools::install_github("yain22/bayesestdft")
R> library(bayesestdft)
R> x1 = rt(n=100,df=0.1) ; x2 = rt(n=100,df=1) ; x3 = rt(n=100,df=5)
R> nu.1 = BayesLNP(x1) ; nu.2 = BayesLNP(x2) ; nu.3 = BayesLNP(x3)
```

The outputs *nu.1*, *nu.2*, and *nu.3* are $S = 1000$ number of posterior samples from each of the scenarios. Figure 2 displays the trace plots after burning the first hundred posterior samples. The results show that ESS (Algorithm 1) possesses a good mixing property and a reasonably high accuracy. More thorough simulation studies are described in Section 4.

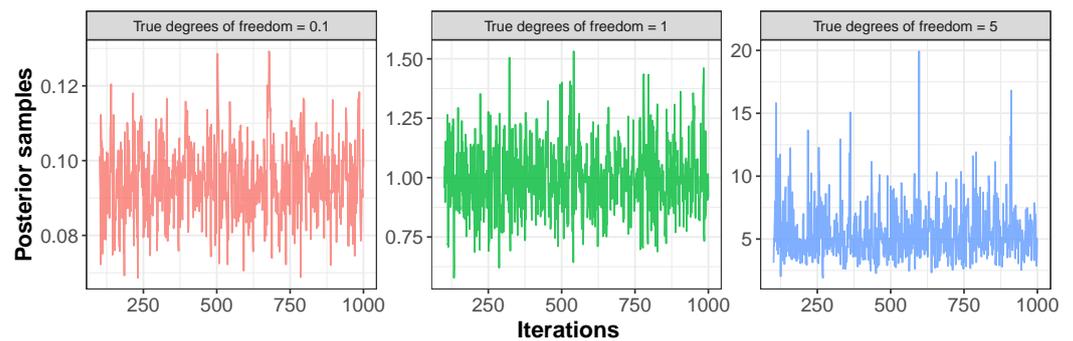


Figure 2. Trace plots for the three simulation experiments. Training data generated from Student t -distribution $t_{\nu_0}(x)$ with $\nu_0 = 0.1$ (left), $\nu_0 = 1$ (middle), and $\nu_0 = 5$ (right).

The package *bayesestdft* includes a function *BayesJeffreys* to implement an MCMC algorithm to sample from the posterior distribution based on the Jeffreys prior $\pi_J(\nu)$ (4). The sampling engines are a random walk Metropolis algorithm [45] and a Metropolis-adjusted Langevin algorithm (MALA) [46,47]. The relevant gradient calculation for implementing MALA is performed using the *numDeriv* R package; users can see the manual in *help(BayesJeffreys)* for more detail.

4. Numerical Studies

4.1. Sensitivity Analysis for a Log-Normal Prior

This subsection presents frequentist properties of Bayes estimators of the parameter ν based on a log-normal prior $\pi(\nu; \mu, \sigma^2) = \log \mathcal{N}(\nu; \mu, \sigma^2)$ with four different choices of the hyperparameters $(\mu, \sigma^2) \in \{(0, 1), (0, 2), (1, 1), (1, 2)\}$. The purpose of this analysis is to measure the impact of the four log-normal priors on the posterior inference about ν and select the most promising hyperparameters out of the four choices, which are then coherently used in the subsequent analyses. As for performance metrics, we report the frequentist mean squared error (MSE) and the frequentist coverage of 95% credible intervals. These are widely used in assessing the accuracy of robust Bayesian procedures [3,19]. For the MSE, we report the median value of the MSEs based on 1000 replications.

The detailed simulation procedures are explained here. We considered drawing independent and identically distributed N samples $\mathbf{x} = (x_1, x_2, \dots, x_N)$ from the student t -distribution $t_{\nu_0}(x)$ (1) with the true data generating parameter ν_0 specified from the effective support $\mathcal{I}_e = (0, 25)$. After specifying a prior $\pi(\nu)$ from the four prior options above (which only differ in the hyperparameters), we obtain the posterior mean $\hat{\nu} = \mathbb{E}[\nu|\mathbf{x}]$ and the 95% credible interval $(L_\alpha(\mathbf{x}), U_\alpha(\mathbf{x}))$ ($\alpha = 0.95$) based on posterior samples $\{\nu^{(s)}\}_{s=1}^S \sim \pi(\nu|\mathbf{x})$. For the purpose of the stabilization to the stationary distribution, we draw 1,000 samples from the posterior $\pi(\nu|\mathbf{x})$, followed by 500 burn-in and 10 thinning. As a result, for a single replicated data $\mathbf{x} \sim t_{\nu_0}(x)$ at each evaluation point $\nu_0 \in \mathcal{I}_e = (0, 25)$, we can calculate the square root of the relative MSE $\sqrt{MSE(\nu_0)}/\nu_0 = |\hat{\nu} - \nu_0|/\nu_0$ and a coverage indicator $\delta(\nu_0) = I[\nu_0 \in (L_\alpha(\mathbf{x}), U_\alpha(\mathbf{x}))]$, where $\delta(\nu_0) = 1$ if $\nu_0 \in (L_\alpha(\mathbf{x}), U_\alpha(\mathbf{x}))$ and 0 otherwise. In particular, the frequentist coverage of 95% credible interval (which is mathematically defined as $Pr_{\nu_0}[\nu_0 \in (L_\alpha(\mathbf{x}), U_\alpha(\mathbf{x}))]$) can be approximated by taking the mean of the resulting values $\delta(\nu_0)$ across the replications. A smaller value of the relative MSE indicates more accurate estimation. For the frequentist coverage of the 95% credible interval, a value closer to 0.95 indicates better coverage performance.

The results of the simulation are shown in Figures 3 and 4. As expected, for all prior choices the relative MSEs are smaller for $N = 100$ than for $N = 30$, and the coverage properties are closer to 95% for $N = 100$ than for $N = 30$ across the values $\nu_0 \in \mathcal{I}_e = (0, 25)$. For both $N = 30$ and $N = 100$ the performance of the posterior mean based on the standard log-normal prior $\pi(\nu) = \log \mathcal{N}(\nu; 0, 1)$ is not good if ν_0 is greater than 10. This is because the standard log-normal prior places too much mass on the interval $(0, 10)$, and hence the estimation performance deteriorates as ν_0 becomes larger. Although the prior

$\pi(v) = \log \mathcal{N}(v; 1, 2)$ produces the best frequentist coverage with a 95% credible interval over the wide range of the effective support \mathcal{I}_e , its MSE is higher than those of other priors for $v_0 \in (3, 15)$ ($N = 30$) and $v_0 \in (5, 17)$ ($N = 100$). The performance based on the priors $\pi(v) = \log \mathcal{N}(v; 1, 1)$ and $\pi(v) = \log \mathcal{N}(v; 0, 2)$ in terms of MSE is quite similar for $N = 30$, although the former performs better than the latter for $N = 100$ on the values $v_0 \in (0, 17)$. Based on these results, we opted to use $\pi_L(v) = \log \mathcal{N}(v; 1, 1)$ (7) as the default choice of the log-normal prior, as it produces reasonably stable estimation on the effective support $\mathcal{I}_e = (0, 25)$ compared to the other choices.

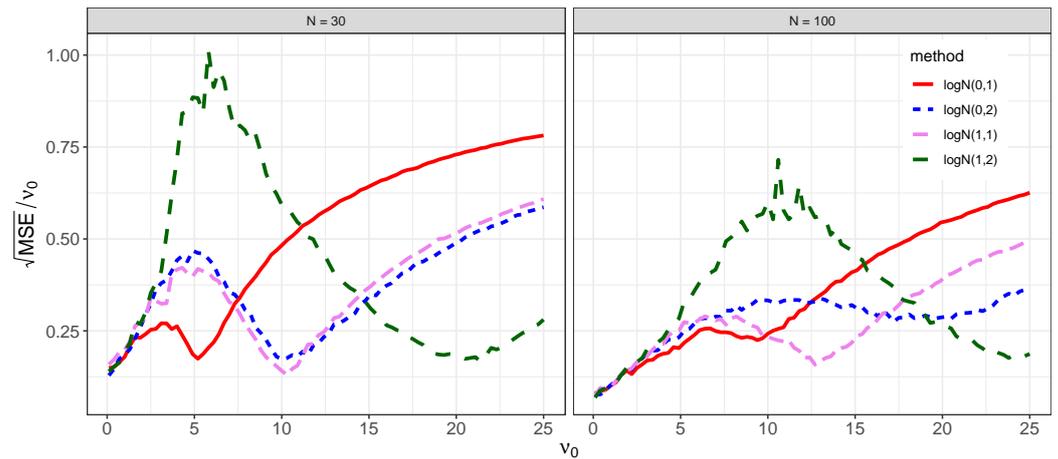


Figure 3. Frequentist properties of the posterior mean $\hat{v} = \mathbb{E}[v|x]$ based on the four log-normal priors $\pi(v) = \log \mathcal{N}(v; \mu, \sigma^2)$ with $(\mu, \sigma^2) \in \{(0, 1), (0, 2), (1, 1), (1, 2)\}$ for sample sizes $N = 30$ and $N = 100$. The y-axis value is the square root of the relative MSE and the x-axis value is the true data-generating parameter on the effective support $\mathcal{I}_e = (0, 25)$.

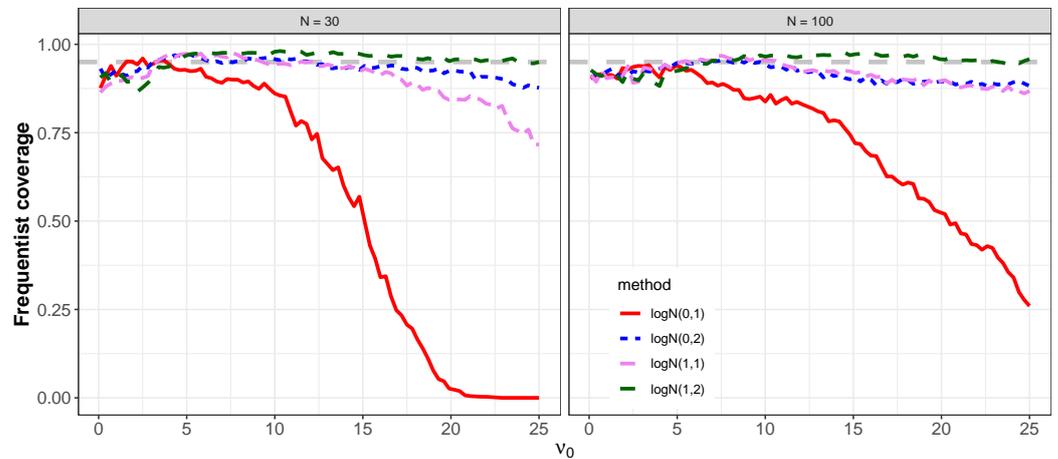


Figure 4. Frequentist coverage of 95% credible intervals for the degrees of freedom based on the four log-normal priors $\pi(v) = \log \mathcal{N}(v; \mu, \sigma^2)$ with $(\mu, \sigma^2) \in \{(0, 1), (0, 2), (1, 1), (1, 2)\}$ for sample sizes $N = 30$ and $N = 100$. The horizontal dashed line represents 0.95 target coverage.

The sensitivity analysis described in this section considered four choices of hyperparameters. While it would obviously be more desirable to select the hyperparameters from a larger number of options, the results show that our choice of $\pi(v) = \log \mathcal{N}(v; 1, 1)$ produces reasonably accurate estimation compared to other priors in both numerical and real data studies.

4.2. Numerical Comparison Using the Jeffreys, Exponential, Gamma, and Log-Normal Priors

This subsection compares frequentist properties of posterior inference based on the four priors $\pi_J(\nu)$ (4), $\pi_E(\nu)$ (5), $\pi_G(\nu)$ (6), and $\pi_L(\nu)$ (7). The simulation designs are explained in Section 4.1. Additionally, we consider the maximum likelihood estimator (MLE) of ν to examine the performance of the frequentist approach compared to Bayesian approaches in estimating the parameter ν . For the MLE, in order to assess the coverage of 95% confidence interval, we used a 95% bootstrap confidence interval based on the MLE of ν [48], as the exact 95% confidence interval was not available. For implementation, the functions *BayesJeffreys* and *BayesLNP* were used to obtain Bayes estimators for ν based on the priors $\pi_J(\nu)$ (4) and $\pi_L(\nu)$ (7), for which the sampling engines were ESS and MALA, respectively. To obtain Bayes estimators based on the priors $\pi_E(\nu)$ (5) and $\pi_G(\nu)$ (6), we used *Stan* [49], which uses a Hamiltonian Monte Carlo algorithm [50]. Finally, the MLE of ν was computed by using the function *fitdistr(densfun = "t")* within *library(MASS)*.

The results of the simulation are shown in Figures 5 and 6. As expected, when using Bayesian and frequentist methods, the relative MSEs are smaller for $N = 100$ than for $N = 30$ across the values $\nu_0 \in \mathcal{I}_e$. For the Bayesian methods, the coverage properties tend to improve (that is, are closer to 0.95) for $N = 100$ than for $N = 30$. In contrast, when using frequentist methods the coverage property becomes more conservative (that is, less than 0.95) for $N = 100$ than for $N = 30$ across the values $\nu_0 \in \mathcal{I}_e$. The Bayes estimator based on the Jeffreys prior $\pi_J(\nu)$ (4) results in relatively lower MSE then the MLE over the values $\nu_0 \in \mathcal{I}_e$, in agreement with [3]. This is to some extent expected, as MLE can generally suffer in small-sample studies [51]. It is important to note that no Bayes estimator dominates other estimators over the entire interval \mathcal{I}_e . In other words, each Bayes estimator has its own region where the estimator is non-inferior to others.

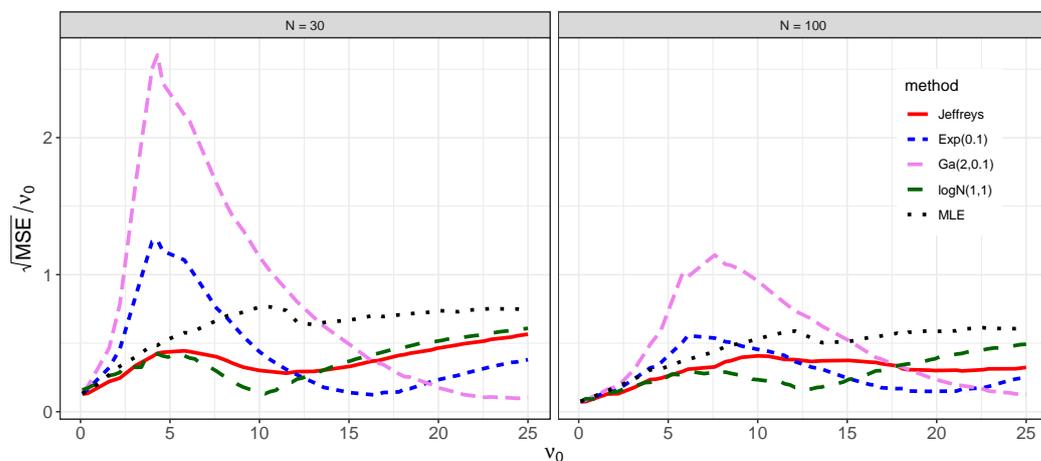


Figure 5. Frequentist properties of the posterior mean based on the four priors ($\pi_J(\nu)$ (4), $\pi_E(\nu)$ (5), $\pi_G(\nu)$ (6), and $\pi_L(\nu)$ (7)) and MLE of ν , for sample sizes $N = 30$ and $N = 100$.

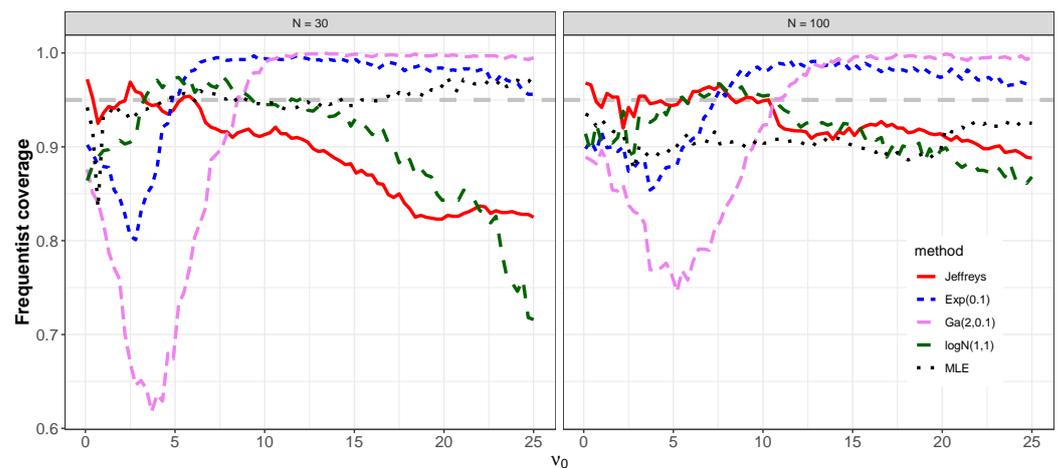


Figure 6. Frequentist coverage of 95% credible intervals for the degrees of freedom based on the four priors ($\pi_J(\nu)$ (4), $\pi_E(\nu)$ (5), $\pi_G(\nu)$ (6), and $\pi_L(\nu)$ (7)) for sample sizes $N = 30$ and $N = 100$. The 95% bootstrap confidence interval based on MLE of ν is used to describe the 95% coverage for the frequentist analysis.

The followings key points summarize the results of the simulations:

- (i) Bayes estimators based on the log-normal prior $\pi_L(\nu)$ (7) and the Jeffreys prior $\pi_J(\nu)$ (4) have almost equal performance, and outperform other estimators for $\nu_0 \in (0, 5)$ ($N = 30$) and $\nu_0 \in (0, 6)$ ($N = 100$).
- (ii) The Bayes estimator based on the log-normal prior $\pi_L(\nu)$ (7) outperforms other estimators for $\nu_0 \in (5, 12)$ ($N = 30$) and $\nu_0 \in (6, 15)$ ($N = 100$).
- (iii) The Bayes estimator based on the exponential prior $\pi_E(\nu)$ (5) outperforms other estimators for $\nu_0 \in (12, 19)$ ($N = 30$) and $\nu_0 \in (15, 22.5)$ ($N = 100$);
- (iv) The Bayes estimator based on the gamma prior $\pi_G(\nu)$ (6) outperforms other estimators for $\nu_0 \in (19, 25)$ ($N = 30$) and $\nu_0 \in (22.5, 25)$ ($N = 100$).

5. Real Data Analysis

To further assess the performance of the Bayes estimators based on the priors $\pi_J(\nu)$ (4), $\pi_E(\nu)$ (5), $\pi_G(\nu)$ (6), and $\pi_L(\nu)$ (7), we analyzed a sample of the daily index values from four countries: the United States (S&P500), Japan (NIKKEI225), Germany (DAX Index), and South Korea (KOSPI). In particular, we considered the data from 2 June 2009 to 30 October 2009, amounting to around 100 observations. The actual analysis was performed on the daily log-rate returns multiplied by 100, that is, $x_i = \log(X_{i+1}/X_i) \times 100$, where X_i is the market index on the i -th trading day. The transformed data for the period of interest are plotted in Figure 7. It can be seen that the series are stationary and that their variances can be reasonably considered as constant over the relevant period. The dataset used here can be loaded using `data(index_return)` in the R package `bayesestdft`.

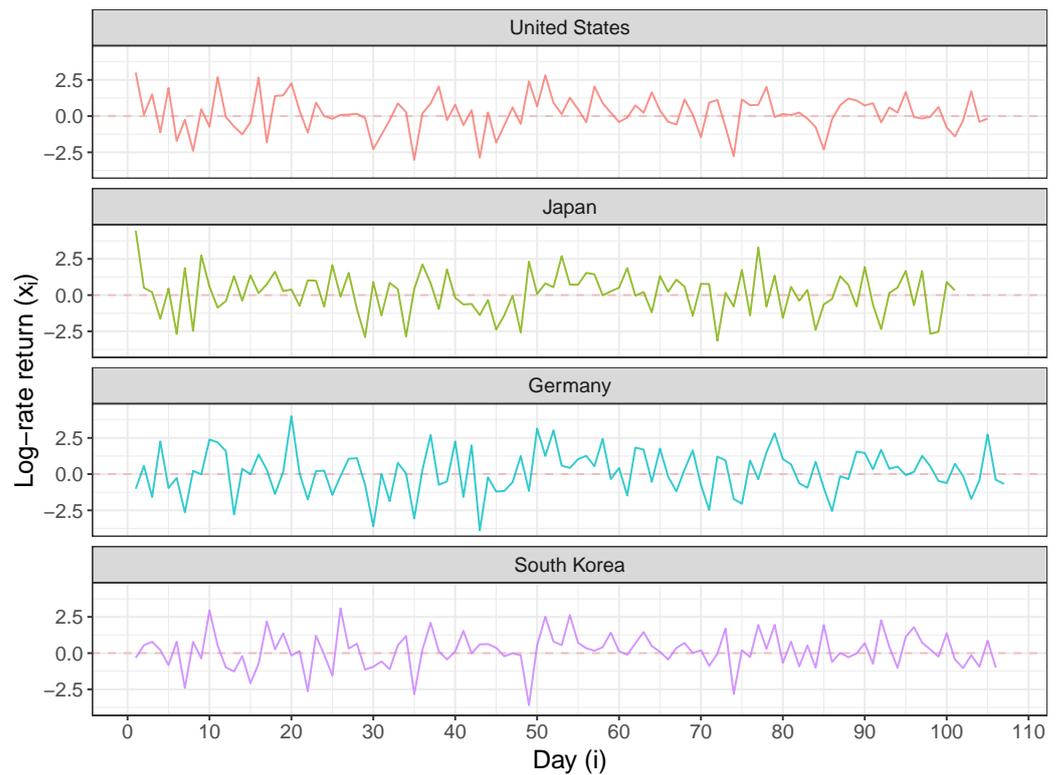


Figure 7. Daily returns of the index values from four countries for the period spanning 6 May 2009 to 30 September 2009.

Table 2 lists the basic descriptive statistics of the index return series from the four countries. Note that the kurtosis is larger than 3 for every country, and even though the distribution of the returns does not have tails much heavier than a normal distribution it seems to be appropriate to consider a t model. Several researchers have found that Student’s t -distribution is the best marginal distribution for index returns [11,52]. Specifically, the model is $x_i \sim t_\nu(x)$, ($i = 1, \dots, N$), with sample sizes $N = 105$ (United States), 101 (Japan), 107 (Germany), and 106 (South Korea) and the goal of estimating the degrees of the freedom $\nu > 0$.

Table 2. Descriptive statistics of the daily index returns.

Statistics	United States	Japan	Germany	South Korea
Mean	0.1753	0.1199	0.1368	0.1695
Variance	1.5040	2.1411	2.2459	1.4347
Skewness	−0.2052	−0.1317	−0.1334	−0.2722
Kurtosis	3.2536	3.0309	3.0086	3.8631

The results of the posterior inference as summarized by posterior mean and 95% credible interval on the parameter ν are reported in Table 3. Additionally, in order to compare the model fit, we report the deviance information criterion (DIC) based on the posterior samples; see Equation (10) in [53] for the analytic formula of the DIC. A smaller value of DIC indicates better modeling fitting. The best inference result is in bold in the table from each country. It can be seen that the Bayes estimator based on the log-normal prior $\pi_L(\nu)$ (7) performs the best for the series from the United States, Japan, and South Korea, while the Jeffreys prior $\pi_J(\nu)$ (4) produces the best model fitting for the series from Germany.

Table 3. Estimation results based on the four priors.

Prior	United States	Japan	Germany	South Korea
$\pi_J(v)$	5.63 (2.67, 11.59) 344.15	3.66 (2.08, 6.63) 373.56	3.31 (1.89, 5.73) 401.24	5.56 (2.64, 11.28) 338.20
$\pi_E(v)$	7.82 (3.10, 17.91) 344.05	4.37 (2.54, 7.28) 374.01	3.91 (2.20, 6.88) 402.66	7.08 (2.92, 14.18) 338.06
$\pi_G(v)$	10.52 (3.82, 25.91) 346.47	4.94 (2.54, 10.12) 377.59	4.26 (2.39, 7.62) 403.62	9.86 (3.82, 25.88) 340.65
$\pi_L(v)$	6.24 (2.78, 14.63) 343.55	3.93 (2.21, 7.85) 373.55	3.56 (1.95, 6.09) 402.1	5.74 (2.88, 11.31) 337.71

NOTE: Contents of table are posterior mean of v , 95% credible interval of v , and DIC. The best outcome in terms of DIC for each country is in bold.

6. Concluding Remarks

In this paper, we studied three popular existing priors, namely, the Jeffreys [3], exponential [15], and gamma prior [16] distributions, and suggested a log-normal distribution as an alternative to produce accurate estimation for the number of degrees of freedom for Student's t -distribution in a small-sample study. The Jeffreys prior has no hyperparameter. Estimation results based on the exponential and gamma priors can be sensitive to the choice of their hyperparameters, and we therefore set these values as the original authors suggested [15,16]. The use of a log-normal prior represents a new trial; hence, we performed a sensitivity analysis to select reasonable hyperparameters. The posterior computation algorithm used to calculate the Bayes estimator based on the log-normal prior possesses good operating characteristics in terms of balancing sampling and estimation accuracy without requiring expert-tuning. We were able to fairly compare the small-sample performance of the four priors through simulation studies for both an effective support and a real data application. The results show that the performance of the Bayes estimator based on the log-normal prior is reasonably good compared to the others. This elucidates the usefulness of the log-normal prior for more complex model settings such as linear regression, nonparametric regression, time-series analysis, and machine learning models, when t errors would be more desirable than using Gaussian errors to carry out robust Bayesian analyses [54].

Funding: This research received no external funding.

Data Availability Statement: Code is publicly available in developed R package *bayesestdft*. Visit Github repository (<https://github.com/yain22/bayesestdft/tree/master/data>, accessed on 22 July 2022) for the datasets analysed during the current study.

Conflicts of Interest: The author declares no conflict of interest.

References

- Student. The probable error of a mean. *Biometrika* **1908**, 1–25.
- Casella, G.; Berger, R.L. *Statistical Inference*; Cengage Learning: Boston, MA, USA, 2021.
- Fonseca, T.C.; Ferreira, M.A.; Migon, H.S. Objective Bayesian analysis for the Student- t regression model. *Biometrika* **2008**, *95*, 325–333. [\[CrossRef\]](#)
- Lange, K.L.; Little, R.J.; Taylor, J.M. Robust statistical modeling using the t distribution. *J. Am. Stat. Assoc.* **1989**, *84*, 881–896. [\[CrossRef\]](#)
- Teichmoeller, J. A note on the distribution of stock price changes. *J. Am. Stat. Assoc.* **1971**, *66*, 282–284. [\[CrossRef\]](#)
- Nolan, J.P. Financial modeling with heavy-tailed stable distributions. *Wiley Interdiscip. Rev. Comput. Stat.* **2014**, *6*, 45–55. [\[CrossRef\]](#)
- Fernández, C.; Steel, M.F. Multivariate Student- t regression models: Pitfalls and inference. *Biometrika* **1999**, *86*, 153–167. [\[CrossRef\]](#)
- Zhu, D.; Galbraith, J.W. A generalized asymmetric Student- t distribution with application to financial econometrics. *J. Econom.* **2010**, *157*, 297–305. [\[CrossRef\]](#)
- Vrontos, I.D.; Dellaportas, P.; Politis, D.N. Full Bayesian inference for GARCH and EGARCH models. *J. Bus. Econ. Stat.* **2000**, *18*, 187–198.

10. Bollerslev, T. A conditionally heteroskedastic time series model for speculative prices and rates of return. *Rev. Econ. Stat.* **1987**, *542*–547. [[CrossRef](#)]
11. Hurst, S.R.; Platen, E. The marginal distributions of returns and volatility. *Lect. Notes Monogr. Ser.* **1997**, *31*, 301–314.
12. West, M. Outlier models and prior distributions in Bayesian linear regression. *J. R. Stat. Soc. Ser.* **1984**, *46*, 431–439. [[CrossRef](#)]
13. Liu, C.; Rubin, D.B. ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Stat. Sin.* **1995**, *5*, 19–39.
14. Villa, C.; Rubio, F.J. Objective priors for the number of degrees of freedom of a multivariate t distribution and the t-copula. *Comput. Stat. Data Anal.* **2018**, *124*, 197–219. [[CrossRef](#)]
15. Fernández, C.; Steel, M.F. On Bayesian modeling of fat tails and skewness. *J. Am. Stat. Assoc.* **1998**, *93*, 359–371.
16. Juárez, M.A.; Steel, M.F. Model-based clustering of non-Gaussian panel data based on skew-t distributions. *J. Bus. Econ. Stat.* **2010**, *28*, 52–66. [[CrossRef](#)]
17. Jacquier, E.; Polson, N.G.; Rossi, P.E. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *J. Econom.* **2004**, *122*, 185–212. [[CrossRef](#)]
18. He, D.; Sun, D.; He, L. Objective Bayesian analysis for the Student-t linear regression. *Bayesian Anal.* **2021**, *16*, 129–145. [[CrossRef](#)]
19. Villa, C.; Walker, S.G. Objective prior for the number of degrees of freedom of at distribution. *Bayesian Anal.* **2014**, *9*, 197–220. [[CrossRef](#)]
20. Jeffreys, H. *Theory of Probability*; Oxford University Press: New York, NY, USA, 1961; p. 472.
21. Kass, R.E.; Wasserman, L. The selection of prior distributions by formal rules. *J. Am. Stat. Assoc.* **1996**, *91*, 1343–1370. [[CrossRef](#)]
22. Berger, J. The case for objective Bayesian analysis. *Bayesian Anal.* **2006**, *1*, 385–402. [[CrossRef](#)]
23. Consonni, G.; Fouskakis, D.; Liseo, B.; Ntzoufras, I. Prior distributions for objective Bayesian analysis. *Bayesian Anal.* **2018**, *13*, 627–679. [[CrossRef](#)]
24. Finner, H.; Dickhaus, T.; Roters, M. Asymptotic tail properties of Student’s t-distribution. *Commun. Stat. Methods* **2008**, *37*, 175–179. [[CrossRef](#)]
25. Abramowitz, M.; Stegun, I.A.; Romer, R.H. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; USA Department of Commerce: Washington, DC, USA, 1988.
26. Jameson, G.J. A simple proof of Stirling’s formula for the gamma function. *Math. Gaz.* **2015**, *99*, 68–74. [[CrossRef](#)]
27. Lee, S.; Kim, J.H. Exponentiated generalized Pareto distribution: Properties and applications towards extreme value theory. *Commun. Stat. Theory Methods* **2019**, *48*, 2014–2038. [[CrossRef](#)]
28. Castillo, E.; Hadi, A.S. Fitting the generalized Pareto distribution to data. *J. Am. Stat. Assoc.* **1997**, *92*, 1609–1620. [[CrossRef](#)]
29. DuMouchel, W.H. On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *Ann. Stat.* **1973**, *1*, 948–957. [[CrossRef](#)]
30. De Haan, L.; Ferreira, A.; Ferreira, A. *Extreme Value Theory: An Introduction*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 21.
31. Northrop, P.J.; Attalides, N. Posterior propriety in Bayesian extreme value analyses using reference priors. *Stat. Sin.* **2016**, *26*, 721–743. [[CrossRef](#)]
32. Chu, J. Errors in Normal Approximations to the t, tau, and Similar Types of Distribution. *Ann. Math. Stat.* **1956**, *27*, 780–789. [[CrossRef](#)]
33. Wang, M.; Yang, M. Posterior property of Student-t linear regression model using objective priors. *Stat. Probab. Lett.* **2016**, *113*, 23–29. [[CrossRef](#)]
34. Metropolis, N.; Ulam, S. The monte carlo method. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341. [[CrossRef](#)] [[PubMed](#)]
35. Thomas, D.R.; Rao, J. Small-sample comparisons of level and power for simple goodness-of-fit statistics under cluster sampling. *J. Am. Stat. Assoc.* **1987**, *82*, 630–636. [[CrossRef](#)]
36. Razali, N.M.; Wah, Y.B. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *J. Stat. Model. Anal.* **2011**, *2*, 21–33.
37. Massey, F.J., Jr. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68–78. [[CrossRef](#)]
38. Shaphiro, S.; Wilk, M. An analysis of variance test for normality. *Biometrika* **1965**, *52*, 591–611. [[CrossRef](#)]
39. Ho, K.W. The use of Jeffreys priors for the Student-t distribution. *J. Stat. Comput. Simul.* **2012**, *82*, 1015–1021. [[CrossRef](#)]
40. Geweke, J. Bayesian treatment of the independent Student-t linear model. *J. Appl. Econom.* **1993**, *8*, S19–S40. [[CrossRef](#)]
41. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [[CrossRef](#)]
42. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [[CrossRef](#)]
43. Chib, S.; Greenberg, E. Understanding the metropolis-hastings algorithm. *Am. Stat.* **1995**, *49*, 327–335.
44. Murray, I.; Prescott Adams, R.; MacKay, D.J. Elliptical slice sampling. *arXiv* **2010**, arXiv:1001.0175.
45. Gustafson, P. A guided walk Metropolis algorithm. *Stat. Comput.* **1998**, *8*, 357–364. [[CrossRef](#)]
46. Ma, Y.A.; Chen, Y.; Jin, C.; Flammarion, N.; Jordan, M.I. Sampling can be faster than optimization. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 20881–20885. [[CrossRef](#)]
47. Dwivedi, R.; Chen, Y.; Wainwright, M.J.; Yu, B. Log-concave sampling: Metropolis-Hastings algorithms are fast! In Proceedings of the Conference on Learning Theory (PMLR), Stockholm, Sweden, 6–9 July 2018; pp. 793–797.

48. Efron, B. Bootstrap methods: Another look at the jackknife. In *Breakthroughs in Statistics*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 569–593.
49. Stan Development Team. *RStan: The R Interface to Stan*, R Package version; Stan Development Team: 2016; Volume 2, p. 522. Available online: <https://mc-stan.org> (accessed on 22 July 2022).
50. Hoffman, M.D.; Gelman, A. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **2014**, *15*, 1593–1623.
51. Van de Schoot, R.; Kaplan, D.; Denissen, J.; Asendorpf, J.B.; Neyer, F.J.; Van Aken, M.A. A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Dev.* **2014**, *85*, 842–860. [[CrossRef](#)]
52. Praetz, P.D. The distribution of share price changes. *J. Bus.* **1972**, *45*, 49–55. [[CrossRef](#)]
53. Gelman, A.; Hwang, J.; Vehtari, A. Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **2014**, *24*, 997–1016. [[CrossRef](#)]
54. Berger, J.O.; Moreno, E.; Pericchi, L.R.; Bayarri, M.J.; Bernardo, J.M.; Cano, J.A.; De la Horra, J.; Martín, J.; Ríos-Insúa, D.; Betrò, B.; et al. An overview of robust Bayesian analysis. *Test* **1994**, *3*, 5–124. [[CrossRef](#)]