Bayesian Analysis (2014) **9**, Number 1, pp. 197–220

Objective Prior for the Number of Degrees of Freedom of a t Distribution

Cristiano Villa * and Stephen G. Walker †

Abstract. In this paper, we construct an objective prior for the degrees of freedom of a t distribution, when the parameter is taken to be discrete. This parameter is typically problematic to estimate and a problem in objective Bayesian inference since improper priors lead to improper posteriors, whilst proper priors may dominate the data likelihood. We find an objective criterion, based on loss functions, instead of trying to define objective probabilities directly. Truncating the prior on the degrees of freedom is necessary, as the t distribution, above a certain number of degrees of freedom, becomes the normal distribution. The defined prior is tested in simulation scenarios, including linear regression with t-distributed errors, and on real data: the daily returns of the closing Dow Jones index over a period of 98 days.

Keywords: Objective prior, t distribution, Kullback–Leibler divergence, Linear regression, Self-information loss function, Robust analysis, Financial return

1 Introduction

In disciplines such as finance and economics, extreme values tend to occur at a probability rate that is too high to be effectively modelled by distributions with appealing analytical properties, such as the normal. This is the case, for example, of financial asset returns and market index values, whose behaviour of extreme values is better represented by distributions with tails heavier than the normal distribution; in particular, see Fabozzi et al. (2010), the t distribution represents an appealing alternative. Furthermore, in Maronna (1976), Lange et al. (1989) and West (1984), it is pointed out that heavy-tailed data are more efficiently handled by regression models for which the error term is assumed to be t-distributed. In fact, it is shown that the influence of outliers is significantly reduced, leading to a more robust analysis; in particular, the smaller the number of degrees of freedom, the more robust the analysis tends to be. As such, the possibility of discerning between t distributions with different numbers of degrees of freedom, especially when the value of this parameter is small, represents an important step of the regression analysis and, in general, whenever a t model is deemed to be the most suitable in representing the observations of interest.

In this paper, we introduce an objective Bayesian prior mass function for the degrees of freedom ν of a t distribution, conditional on the mean parameter μ and variance parameter σ^2. Hence, it will be of the form $\pi(\nu|\mu, \sigma^2)$.

*School of Mathematics, Statistics and Actuarial Science, University of Kent, UK cv60@kent.ac.uk
†Department of Mathematics, and Division of Statistics and Scientific Computation, University of Texas at Austin, Texas, USA s.g.walker@math.utexas.edu

 DOI:10.1214/13-BA854

There are two fundamental aspects which have to be discussed, in our opinion, as preliminary remarks to the formal definition of the objective prior for ν. The first remark has a general characterization and it refers to a conceptual incongruence in a Bayes theorem application when the prior is defined through common objective procedures; the second remark is specific to the definition of a prior distribution for ν and it argues that this distribution should be truncated. Let us discuss the former remark first.

In a subjective Bayesian approach, $\pi(\theta)$ represents the initial degree of belief that we have about the possible values that θ can take within the parameter space. Then, by combining it with the information contained in the observed data, expressed by the likelihood function $f(x|\theta)$, the initial beliefs are updated and become the posterior probability distribution. The prior and posterior should retain the same meaning.

If the probability distribution $\pi(\theta)$ is determined through objective Bayesian methods, this distribution will often be improper. This fact raises some important concerns about defining objective probabilities *directly*. Contrary to the subjective approach, whereby the prior and posterior retain the same meaning, the same can not be said of an objective prior, for the posterior derived from it must at some point represent beliefs. We believe that the solution to this difficulty is, not to be objective in assigning a mass to every element of the parameter space, but by assigning a *worth* to every one of them. In other words, to *work* with losses instead of probabilities. Recalling that objectivity arises from the absence of knowledge, actual or alleged, about the true value of the parameter of interest, we can see the justification of the proposed approach, as we can still have an idea on the *worth* that each parameter value represents in the model. The *worth* of an element of the parameter space can be assessed by describing and evaluating what is lost if this value is removed. And by assigning the mass to each parameter value by a measure of its *worth*, we are not subject to the constraint of properness, intrinsic in a probability measure.

Let us denote by $\pi(\nu)$ the prior distribution for the discrete parameter $\nu = 1, 2, \ldots, \infty$ representing the number of degrees of freedom of a t distribution. If a prior π has been assigned then we link this to a worth of each element by means of the *self-information* loss function $-\log \pi(\nu)$. The *self-information* loss function (also known as the *log-loss* function in machine learning) measures the performance of a probability assignment with respect to an outcome. Thus, for every probability assignment $\pi = \{\pi(\nu), \nu \in N\}$ over the space $\mathcal{X}$, with $x \in \mathcal{X}$, the *self-information* loss function is defined as

$$l(\pi, \nu) = -\log \pi(\nu).$$

More details and properties of this particular loss function can be found, for example, in Merhav and Feder (1998). We can then identify an appropriate objective way to associate a loss to each ν, representing its *worth* in the model line–up, and the prior distribution $\pi(\nu)$ then follows. Furthermore, we note that in this way the Bayesian approach is conceptually consistent, as we update an initial (i.e. prior) *worth* assigned to ν, through the application of Bayes' theorem, to obtain the resulting *worth* expressed by $-\log \pi(\nu|x)$. Indeed, there is an elegant procedure akin to Bayes which works from

a loss point of view, namely that

$$-\log \pi(\nu|x) = K - \log f(x|\nu) - \log \pi(\nu)$$

which has the interpretation of

$$\text{Loss}(\nu|x, \pi) = K + \text{Loss}(\nu|x) + \text{Loss}(\nu|\pi).$$

This is a cumulative loss function for assessing the loss of ν in the presence of two pieces of mutual information x and π. Here K is a constant which does not depend on ν.

To better illustrate how an objective criterion to assign a *worth* to each element of the parameter space can be derived, the following example may be helpful. Let us assume we have a scenario where the possible models are three: f_1, f_2 and f_3, that is $t(\nu_1, \mu, \sigma^2)$, $t(\nu_2, \mu, \sigma^2)$ and $t(\nu_3, \mu, \sigma^2)$. Let us also assume that f_1 and f_2 are very similar, whilst f_3 is significantly different from the other two. For example, we can imagine that f_1 and f_2 have consecutive numbers of degrees of freedom and f_3 a much larger (or much smaller) one. We do not question the rational behind this choice of model options, we just assume that there is one. If we remove from the scenario either f_1 or f_2, as they are relatively close, there is no appreciable change in the whole structure of options, as we still have the remaining model (either f_2 or f_1) to support that specific position. On the other hand, if we remove f_3, the structure of options is considerably different from the original, as only two very similar options are left. We then see that f_3 is more *valuable* than f_1 or f_2, because, if it is removed, the scenario is significantly altered; or, alternatively, we can say that the loss in removing f_3 is higher than the loss in removing either f_1 or f_2. An important aspect is that the loss associated to each model takes into consideration the surrounding models.

The *worth* to be assigned to each model is equal to the Kullback–Leibler divergence measured from the model to its nearest neighbour. This is justified by the fact that, if the model is misspecified, the posterior distribution asymptotically accumulates at the nearest model with respect to the Kullback–Leibler divergence (Berk 1966). If we consider the family of distributions $f(\cdot|\theta)$, where $\theta \in \Theta$ is the discrete parameter characterising it, the result of Berk (1966) says that, if θ_0 is the true parameter value, and it is not considered, then the posterior distribution $\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$ will accumulate at θ', where θ' is the parameter value such that $D_{KL}(f(x|\theta_0)\|f(x|\theta'))$ attains its minimum. Thus, this divergence represents the utility (i.e. *worth*) of having θ_0 in Θ, that is $u(\theta_0)$, when it is the true parameter value. So the more isolated θ_0 is, the greater is its utility. Given that in decision theory (Berger 1985) the loss corresponds to (in general) negative the utility, we have that $-D_{KL}(f(x|\theta_0)\|f(x|\theta'))$ represents the loss in keeping θ_0.

The second remark we would like to discuss originates from the well known property of the t distribution to converge to a normal distribution when the degrees of freedom tend to infinity. That is, from a certain point in the parameter space of degrees of freedom, the distribution can be considered as normal. The key point we wish to make is that it is not fundamental where the quantification of this *turning point* is (i.e. where a t distribution turns into a normal), but the fact that there is one, and that every

t distribution with a value of ν equal or larger than this *turning point* is considered the same model, that is, a normal distribution. We take this point to be 30 based on theoretical results, see Chu (1956), and also Section 3. It follows that the set of parameter values on which the prior $\pi(\nu)$ is built becomes a finite set of models and ν translates to a label associated to each model. If we indicate the turning point as $\nu_{\max}$, the set of models is represented by $\{f_1, f_2, \ldots, f_{\nu_{\max}-1}, f_{\nu_{\max}}\}$, where the first $(\nu_{\max}-1)$ models are t distributions with degrees of freedom $\nu = 1, 2, \ldots, \nu_{\max}-1$, and $f_{\nu_{\max}} \approx N(\mu, \sigma^2)$.

A direct consequence of this consideration is that it reveals an important conceptual gap common to other objective approaches to derive $\pi(\nu)$. Even though it is theoretically possible to discern between two t distributions with any number of degrees of freedom, provided a sufficiently large number of observations is available, this task loses meaning when the number of degrees of freedom is large enough. It follows that, if we want to assign prior mass to models, for example, in intervals $[f_{200}, \ldots, f_{299}]$ and $[f_{300}, \ldots, f_{399}]$, this mass has to be the same for each element, as these models are in practice not distinguishable. As such, if we define a prior of ν for values that go from one to infinity, this prior has to be uniform in the interval $[\nu_{\max}, +\infty)$, and therefore improper. But, as we have discussed above, all the models in this interval are (approximatively) represented by a normal distribution and, as a result, the set of options has to be finite with the *last* element equal to a normal. Furthermore, as all the models from $f_{\nu_{\max}}$ onwards are virtually the same model (i.e. normal), if $\pi(\nu)$ is defined over the whole sample space, it means that a large amount of mass is put on the normal model. And there is no apparent justification for this approach.

Here we review some of the objective methods to assign a prior to the number of degrees of freedom of a t density that can be found in the literature. In most cases, the field of interest is when the error term of a regression model is assumed to have a t distribution.

The likelihood for ν given μ and σ^2 tends to a positive constant as $\nu \to +\infty$ (Anscombe 1967). As such, to have a proper posterior, the prior distribution has to tend to 0 as $\nu \to +\infty$. Therefore, the natural objective prior

$$\pi(\nu) \propto 1,$$

cannot be adopted as the posterior would be improper. In fact, as shown in Fernandez and Steel (1999), this behaviour of the likelihood function may lead, in general, to an improper posterior whenever the prior distribution is improper.

To overcome this issue, Jacquier et al. (2004) proposed a truncated uniform prior on the discrete integer degrees of freedom. In particular, they note that the variance of a t density exists only for values of $\nu \geq 3$. Furthermore, for values of $\nu \in [41, 50]$, the model does not have significant changes in behaviour and therefore, their discrete uniform prior is

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

However, as seen in Fonseca et al. (2008), these type of prior probabilities are inappropriate, because the estimate of the number of degrees of freedom is sensitive to the chosen truncation.

Geweke (1993) proposes a prior distribution that is exponential. In this case, the parameter ν is considered continuous and the distribution depends on a value g, which is strictly positive

$$\pi(\nu) \propto \exp\{-g\nu\} \qquad \nu > 0.$$

This prior, in our opinion, cannot be considered as strictly objective. In fact, different values of g will lead to a different distribution of the mass over small values of ν, where it is more critical to be able to estimate the number of degrees of freedom. Furthermore, as shown in Fonseca et al. (2008), the exponential prior tends to dominate the data.

In Fonseca et al. (2008), a linear regression model with p regressors and error term t-distributed is considered. The authors define two prior distributions for ν, both based on Jeffreys' prior (Jeffreys 1961): the independence Jeffreys prior

$$\pi_I(\nu) \propto \left(\frac{\nu}{\nu+3}\right)^{1/2} \left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}^{1/2} \qquad \nu > 0, \tag{1}$$

and the Jeffreys-rule prior

$$\pi_J(\nu) \propto \pi_I(\nu) \left(\frac{\nu+1}{\nu+3}\right)^{p/2} \qquad \nu > 0. \tag{2}$$

It is shown that both priors are proper, and that they lead to proper posteriors.

Prior distributions, though not objective, for the number of degrees of freedom of a t distribution, are given by Juárez and Steel (2010), where a non-hierarchical and a hierarchical prior are considered. The first is a particular gamma, with parameters 2 and 1/100, leading to the density

$$\pi_1(\nu) = \frac{\nu}{100} e^{-\nu/10}. \tag{3}$$

This prior has the property of covering a large range of relevant values of degrees of freedom and allows for all prior moments to exist. The hierarchical prior is obtained by considering an exponential distribution for the scale parameter of the gamma, with shape parameter 2. The resulting density is

$$\pi_2(\nu) = 2k \frac{\nu}{(\nu+k)^3},$$

where $k > 0$ is the hyper-parameter. The authors compared the performance of their priors with the Jeffreys' independent prior proposed by Fonseca et al. (2008), noting that there were no significant differences for values of ν below 50.

It has to be noted that in Geweke (1993), Fonseca et al. (2008) and Juárez and Steel (2010), the number of degrees of freedom is considered as continuous.

We consider the parameter space of ν to be discrete, that is restricted to positive integers. The motivation is practical. In fact, the Kullback–Leibler divergence between contiguous densities rapidly decreases to zero, making necessary large amount of information about ν (i.e. observations) in order to discern between different t distributions (Jacquier et al. 2004). We could densify the parameter space, for example $\nu = \{1, 1.5, 2, 2.5, \ldots\}$ (or even more dense), and apply our criterion to derive a prior, but the resulting increase in precision of the estimate of ν would not be of any practical use, as, for example, there is no sensible difference in having a t density with 7 degrees of freedom and one with 7.1 degrees of freedom.

The paper is organized as follows. In Section 2 we introduce the notation that will be used throughout the paper, derive the Kullback–Leibler divergence for t distributions and show the computational results related to the determination of its minimum value (for a given ν). Section 3 is dedicated to the formal definition of the objective prior for ν; here, we highlight the fact that this prior has to be truncated. In Section 4 we analyse the posterior distribution by estimating the number of degrees of freedom on simulated data. We consider the case of data t distributed, and a regression model with t-distributed errors. An analysis on actual data, in particular on daily returns of the closing Dow Jones index, is performed in Section 5, where we compare our results with the ones obtained by using other objective priors for ν found in literature. Section 6 carries the final comments and discussions.

2 Preliminaries

If random variable x has a t distribution with degrees of freedom ν, location parameter μ and scale parameter σ^2, its probability density function is represented by

$$f(x|\nu,\mu,\sigma^2) = \frac{1}{B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(\frac{1}{\nu\sigma^2}\right)^{\frac{1}{2}}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}} \qquad -\infty < x < \infty,$$

where $B(\cdot,\cdot)$ is the beta function. Both location and scale parameters are continuous, with $-\infty < \mu < \infty$ and $\sigma^2 > 0$. The density of x can equivalently be expressed in terms of the precision parameter $\lambda = 1/\sigma^2$ as follows

$$f(x|\nu,\mu,\lambda) = \frac{1}{B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(\frac{\lambda}{\nu}\right)^{\frac{1}{2}}\left(1+\frac{\lambda(x-\mu)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad -\infty < x < \infty.$$

For this section we focus on the particular case where $\mu = 0$ and $\sigma^2 = 1$; it is always possible to move from a t distribution with $\mu = 0$ and $\sigma^2 = 1$ to a t distribution with any value of the parameters (and vice versa) by simply applying the relationship $x_{\nu,\mu,\sigma^2} = \mu + \sigma x_{\nu,0,1}$. In any case, as we are interested in comparing t distributions that differ only in the number of degrees of freedom, to avoid a cumbersome notation, the t model with ν degrees of freedom and parameters μ and σ^2 is represented as f_ν in lieu of $f(x|\nu, \mu, \sigma^2)$.

In Section 1 we have introduced the objective criterion to define the prior for ν. This criterion is based on the key assumption that the posterior distribution for ν, if the true value is removed, asymptotically accumulates on the nearest model with respect to the Kullback–Leibler divergence.

Let us consider the following t distributions: f_ν and $f_{\nu'}$, with $\nu = 1, 2, \ldots$ and $\nu' = 1, 2, \ldots$. Also, we assume that $\nu \neq \nu'$ and that location and scale parameters are equal for both densities with $\mu = 0$ and $\sigma^2 = 1$. The Kullback–Leibler divergence (Kullback and Leibler (1951)) between f_ν and $f_{\nu'}$ is given by

$$
\begin{aligned}
D_{KL}(f_\nu \| f_{\nu'}) &= \int_{-\infty}^{\infty} f_\nu \log\left(\frac{f_\nu}{f_{\nu'}}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{\nu} B(1/2, \nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \log\left\{ \frac{\frac{1}{\sqrt{\nu} B(1/2, \nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{\frac{1}{\sqrt{\nu'} B(1/2, \nu'/2)} \left(1 + \frac{x^2}{\nu'}\right)^{-\frac{\nu'+1}{2}}} \right\} dx \\
&= \log\left\{ \frac{\sqrt{\nu'} B(\frac{1}{2}, \frac{\nu'}{2})}{\sqrt{\nu} B(\frac{1}{2}, \frac{\nu}{2})} \right\} - \frac{\nu+1}{2} \mathbb{E}_\nu \left[\log\left(1 + \frac{x^2}{\nu}\right) \right] + \frac{\nu'+1}{2} \mathbb{E}_\nu \left[\log\left(1 + \frac{x^2}{\nu'}\right) \right] \qquad (4)
\end{aligned}
$$

where $\mathbb{E}_\nu$ represents the expected value with respect to f_ν. To identify the nearest model, in terms of Kullback–Leibler divergence, we have numerically computed the expression in (4), for $\nu > 1$ to compare $D_{KL}(f_\nu \| f_{\nu-1})$ and $D_{KL}(f_\nu \| f_{\nu+1})$. In Table 1 we have the computational results for the first 29 values of ν. The results obtained show that $D_{KL}(f_\nu \| f_{\nu-1}) > D_{KL}(f_\nu \| f_{\nu+1})$, for any ν, and that the divergence decreases as the number of degrees of freedom tends to infinity. This latter result is intuitive and obvious, as the t distribution converges in distribution to the normal.

In Section 1, we have anticipated that the prior we propose is truncated, and that this is done to avoid assigning more mass than appropriate to the normal model. As such, the Kullback–Leibler divergence at the points of the parameter space near to and at the truncation have to be discussed separately. First, we note that the minimum Kullback–Leibler divergence at the truncation point is given by

ν	$D_{KL}(f_\nu \| f_{\nu-1})$	$D_{KL}(f_\nu \| f_{\nu+1})$	ν	$D_{KL}(f_\nu \| f_{\nu-1})$	$D_{KL}(f_\nu \| f_{\nu+1})$
2	0.0621	0.0192	17	1.8943×10^{-05}	1.5465×10^{-05}
3	0.0136	0.0059	18	1.5155×10^{-05}	1.2496×10^{-05}
4	0.0047	0.0024	19	1.2268×10^{-05}	1.0207×10^{-05}
5	0.0020	0.0012	20	1.0037×10^{-05}	8.4196×10^{-06}
6	0.0010	6.3640×10^{-04}	21	8.2910×10^{-06}	7.0071×10^{-06}
7	5.7683×10^{-04}	3.7607×10^{-04}	22	6.9087×10^{-06}	5.8791×10^{-06}
8	3.4728×10^{-04}	2.3658×10^{-04}	23	5.8028×10^{-06}	4.9693×10^{-06}
9	2.2150×10^{-04}	1.5632×10^{-04}	24	4.9096×10^{-06}	4.2291×10^{-06}
10	1.4789×10^{-04}	1.0746×10^{-04}	25	4.1819×10^{-06}	3.6217×10^{-06}
11	1.0249×10^{-04}	7.6319×10^{-05}	26	3.5841×10^{-06}	3.1197×10^{-06}
12	7.3261×10^{-05}	5.5705×10^{-05}	27	3.0894×10^{-06}	2.7018×10^{-06}
13	5.3751×10^{-05}	4.1614×10^{-05}	28	2.6773×10^{-06}	2.3516×10^{-06}
14	4.0326×10^{-05}	3.1717×10^{-05}	29	2.3316×10^{-06}	2.0564×10^{-06}
15	3.0844×10^{-05}	2.4599×10^{-05}	30	2.0399×10^{-06}	1.8061×10^{-06}
16	2.3993×10^{-05}	1.9373×10^{-05}			

Table 1: Comparison of the Kullback–Leibler divergence $D_{KL}(f_\nu \| f_{\nu-1})$ and $D_{KL}(f_\nu \| f_{\nu+1})$ for $\nu = 2, \ldots, 30$. The distance from ν to $\nu + 1$ is smaller than the distance from ν to $\nu - 1$, for any value of ν.

$$
\begin{aligned}
D_{KL}(N_{0,1} \| f_\nu) &= \int_{-\infty}^{\infty} N_{0,1} \log\left(\frac{N_{0,1}}{f_\nu}\right) dx \\
&= \log\left\{\frac{\sqrt{\nu} B(\frac{1}{2}, \frac{\nu}{2})}{\sqrt{2\pi}}\right\} - \frac{1}{2}\mathbb{E}_N\left(x^2\right) + \frac{\nu+1}{2}\mathbb{E}_N\left\{\log\left(1 + \frac{x^2}{\nu}\right)\right\}, \quad (5)
\end{aligned}
$$

where $N_{0,1}$ is the standard normal, and $\mathbb{E}_N$ represents the expected value with respect to $N_{0,1}$. If we indicate by $f_{\nu_{\max}}$ the normal model at the truncation point, the nearest distribution to $f_{\nu_{\max-1}}$ is $f_{\nu_{\max-2}}$, as the numerical computation in Table 2 shows. The results can be summarised as follows. If the set of densities is given by $\{f_1, f_2, \ldots, f_{\nu_{\max-1}}, f_{\nu_{\max}}\}$, with $f_{\nu_{\max}} \approx N(0,1)$, the minimum divergence for $\nu = 1, \ldots, \nu_{\max-2}$ is $D_{KL}(f_\nu \| f_{\nu+1})$; for $f_{\nu_{\max-1}}$ and $f_{\nu_{\max}}$ it is $D_{KL}(f_\nu \| f_{\nu-1})$.

3 The Objective prior

To define the prior mass function for the degrees of freedom ν of a t distribution, we need to make the following considerations. We assume that the location parameter μ and the scale parameter σ^2 (or, equivalently, the precision λ) are known. Let us consider a random variable x with a t distribution with parameters ν, μ and σ^2. Therefore, for $\nu \to +\infty$ we have $x \xrightarrow{d} N(\mu, \sigma^2)$. It is common practice to assume normality for $\nu \geq 30$. Chu (1956) shows that the proportional error in using the distribution function

ν	$D_{KL}(f_{\nu_{\max}-1} \| f_{\nu_{\max}-2})$	$D_{KL}(f_{\nu_{\max}-1} \| f_{\nu_{\max}})$
30	2.0399×10^{-06}	0.0021
60	1.3121×10^{-07}	0.0005×10^{-04}
90	2.6168×10^{-08}	0.0002×10^{-04}
120	8.3194×10^{-09}	0.0001×10^{-04}
150	3.4174×10^{-09}	7.9029×10^{-05}
180	1.6513×10^{-09}	5.4735×10^{-05}

Table 2: Comparison of the Kullback–Leibler divergence from $f_{\nu_{\max}-1}$ to $f_{\nu_{\max}-2}$ and from $f_{\nu_{\max}-1}$ to $f_{\nu_{\max}}$, with $f_{\nu_{\max}} \approx N(0,1)$. It can be noted that the last t distribution is closer to the t distribution on its left than to the standard normal.

of a standard normal, $\Phi(x)$, as an approximation to the distribution function of x, $F(x)$, is smaller than $1/\nu$ for every $\nu \geq 8$, where the proportional error is defined as $E = |(F(x)/\Phi(x)) - 1|$. In fact, the approximation of a t distribution to a normal density is always to a certain level of precision and, apart from computational limitations, it is always possible to find a sample size large enough to be able to discriminate the two distributions for a given precision level. In any case, the prior mass function for the parameter ν is defined over a set of models composed by t distributions with increasing number of degrees of freedom and, as a final model, a normal distribution. This normal distribution can be seen as the model that incorporates all the remaining t distributions for which we assess that the value of ν is too high to make them distinguishable from a normal. Therefore, the prior $\pi(\nu)$ is a function that associates a mass to each model in the finite set $\{f_1, f_2, \ldots, f_{\nu_{\max}-1}, f_{\nu_{\max}}\}$, where f_ν (for $\nu = 1, \ldots, \nu_{\max}-1$) is a t distribution with ν degrees of freedom, and $f_{\nu_{\max}}$ is the normal distribution $N(\mu, \sigma^2)$.

For the remainder of this section, we focus on the special case where $\mu = 0$ and $\sigma^2 = 1$, as this simplifies the notation and does not result in any loss of generality. We have introduced in Section 1 the fact that, if the true model is removed from the set of all possible models, then the posterior distribution will tend to accumulate on the nearest model in terms of the Kullback–Leibler divergence, see also Dmochowski (1996). Then, minus the divergence represents the loss we would incur if the removed model is the true one, that is

$$l(\nu) = \begin{cases} -D_{KL}(f_\nu \| f_{\nu-1}) & \text{if } \nu \geq \nu_{\max} - 1 \\ -D_{KL}(f_\nu \| f_{\nu+1}) & \text{if } \nu < \nu_{\max} - 1, \end{cases}$$

and the derivation of the prior probability from this loss is given by the *self-information* loss function

$$-\log \pi(\nu) = \begin{cases} -D_{KL}(f_\nu \| f_{\nu-1}) & \text{if } \nu \geq \nu_{\max} - 1 \\ -D_{KL}(f_\nu \| f_{\nu+1}) & \text{if } \nu < \nu_{\max} - 1. \end{cases}$$

The prior mass to be put on each model in the set of options is given by

$$\pi(\nu) \propto \begin{cases} \exp\left\{D_{KL}(f_\nu \| f_{\nu-1})\right\} & \text{if } \nu \geq \nu_{\max} - 1 \\ \exp\left\{D_{KL}(f_\nu \| f_{\nu+1})\right\} & \text{if } \nu < \nu_{\max} - 1. \end{cases} \tag{6}$$

The prior for values of $\nu < \nu_{\max} - 1$ is obtained by replacing equation (4) in the first of (6)

$$\begin{aligned} \pi(\nu) \propto & \frac{\sqrt{\nu+1}B\left(\frac{1}{2}, \frac{\nu+1}{2}\right)}{\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \exp\left\{-\frac{\nu+1}{2}\mathbb{E}_\nu\left[\log\left(1+\frac{x^2}{\nu}\right)\right]\right. \\ & \left. +\frac{\nu+2}{2}\mathbb{E}_\nu\left[\log\left(1+\frac{x^2}{\nu+1}\right)\right]\right\}. \end{aligned} \tag{7}$$

The prior mass for $\nu_{\max} - 1$ is given by replacing (4) in the second of (6), for which we set $\nu' = \nu_{\max} - 2$:

$$\begin{aligned} \pi(\nu_{-1}) \propto & \frac{\sqrt{\nu_{-1}-1}B\left(\frac{1}{2}, \frac{\nu_{-1}-1}{2}\right)}{\sqrt{\nu_{-1}}B\left(\frac{1}{2}, \frac{\nu_{-1}}{2}\right)} \exp\left\{-\frac{\nu_{-1}+1}{2}\mathbb{E}_{\nu_{-1}}\left[\log\left(1+\frac{x^2}{\nu_{-1}}\right)\right]\right. \\ & \left. +\frac{\nu_{-1}}{2}\mathbb{E}_{\nu_{-1}}\left[\log\left(1+\frac{x^2}{\nu_{-1}-1}\right)\right]\right\}. \end{aligned} \tag{8}$$

Note that, for simplicity in the notation, in equation (8) we have replaced $\nu_{\max} - 1$ by ν_{-1}. Finally, the prior for $\nu_{\max}$ is obtained by replacing (5), for which $\nu = \nu_{\max} - 1$, in the second equation of (6), obtaining

$$\pi(\nu_{\max}) \propto \frac{\sqrt{\nu_{-1}}B\left(\frac{1}{2}, \frac{\nu_{-1}}{2}\right)}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\mathbb{E}_N\left(x^2\right) + \frac{\nu_{-1}}{2}\mathbb{E}_N\left[\log\left(1+\frac{x^2}{\nu_{-1}}\right)\right]\right\}. \tag{9}$$

To have a picture of the prior on ν, we have plotted its behaviour for three distinctive values of $\nu_{\max}$; in particular, in Figure 1 we have explored the cases where the prior has been truncated at $\nu = 30, 60$ and 90. The prior puts the highest value of mass on the first model, the t distribution with one degree of freedom, and gradually decreases toward one as ν increases. This is a direct consequence of the fact that the models become more and more similar to each other, resulting in a Kullback–Leibler divergence converging to 0. The priors look uniform for $\nu > 5$; however this is a perception caused by the fact that the scale is distorted by the larger values of the prior for the small values. While the prior does look uniform, it is not and the subtle differences are sufficient for the prior not to behave as a uniform prior. And something close to uniform for high degrees of freedom is coherent. For if mass $\pi(\nu)$ has been put on ν then one would expect the

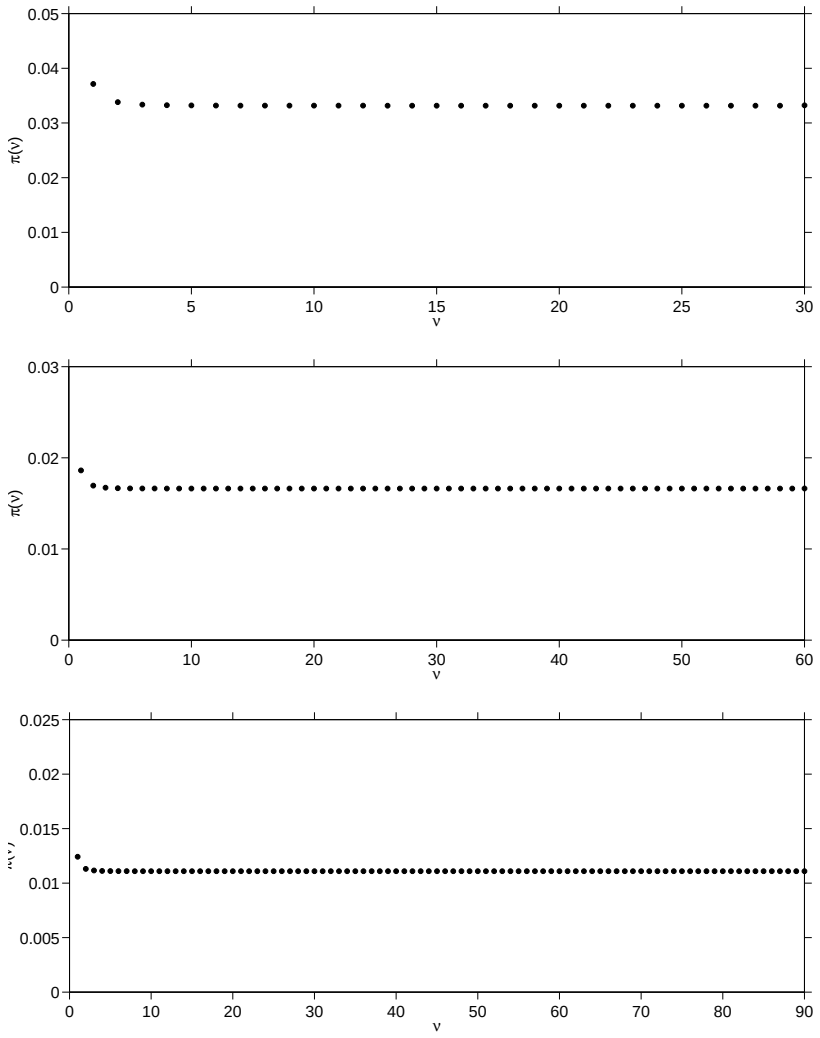

Figure 1: Normalised prior distributions for ν truncated at $\nu_{\max} = 30$, $\nu_{\max} = 60$ and $\nu_{\max} = 90$.

mass on $\pi(\nu+1)$ to be very similar simply because the f_ν and $f_{\nu+1}$ are almost the same density.

The prior distribution has also been analysed for t distributions with different values of μ and σ^2. We have observed that the prior is not affected by changes in the location parameter μ. Although the scale parameter σ^2 has some effect on the prior, that is a larger mass is assigned to values of $\nu \leq 5$ for increasing values of σ^2, there is no change in the tail of the distribution. However, the posterior is not significantly affected by

this, given that the main effect of the prior on the posterior is in the tails, where the priors are remarkably similar (Berger et al. 2012).

4 Posterior analysis

4.1 Sampling algorithm

By combining the likelihood function for parameter ν (given μ and σ^2) for a t distribution, that is

$$L(\nu|\mu,\sigma^2,x) = \prod_{i=1}^{n}\left\{\frac{1}{B\left(1/2,\nu/2\right)}\left(\frac{1}{\nu\sigma^2}\right)^{1/2}\left(1+\frac{(x_i-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}\right\},$$

with the appropriate prior for ν in (7), (8) or (9), in which we have included parameters μ and σ^2, we obtain, respectively, the following three posterior distributions

$$\begin{aligned}\pi(\nu|\mu,\sigma^2,x) \propto &\prod_{i=1}^{n}\left\{\frac{1}{B\left(1/2,\nu/2\right)}\left(\frac{1}{\nu\sigma^2}\right)^{1/2}\left(1+\frac{(x_i-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}\right\}\times\\ &\frac{\sqrt{\sigma^2(\nu+1)}B\left(\frac{1}{2},\frac{\nu+1}{2}\right)}{\sqrt{\sigma^2\nu}B\left(\frac{1}{2},\frac{\nu}{2}\right)}\exp\left\{-\frac{\nu+1}{2}\mathbb{E}_\nu\left[\log\left(1+\frac{(x-\mu)^2}{\sigma^2\nu}\right)\right]\right.\\ &\left.+\frac{\nu+2}{2}\mathbb{E}_\nu\left[\log\left(1+\frac{(x-\mu)^2}{\sigma^2(\nu+1)}\right)\right]\right\},\end{aligned}$$

for values of $\nu = 1,\ldots,\nu_{\max}-2$;

$$\begin{aligned}\pi(\nu_{-1}|\mu,\sigma^2,x) \propto &\prod_{i=1}^{n}\left\{\frac{1}{B\left(1/2,\nu/2\right)}\left(\frac{1}{\nu\sigma^2}\right)^{1/2}\left(1+\frac{(x_i-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}\right\}\times\\ &\frac{\sqrt{\sigma^2(\nu_{-1}-1)}B\left(\frac{1}{2},\frac{\nu_{-1}-1}{2}\right)}{\sqrt{\sigma^2\nu_{-1}}B\left(\frac{1}{2},\frac{\nu_{-1}}{2}\right)}\exp\left\{-\frac{\nu_{-1}+1}{2}\times\right.\\ &\left.\mathbb{E}_{\nu_{-1}}\left[\log\left(1+\frac{(x-\mu)^2}{\sigma^2\nu_{-1}}\right)\right]+\frac{\nu_{-1}}{2}\mathbb{E}_{\nu_{-1}}\left[\log\left(1+\frac{(x-\mu)^2}{\sigma^2(\nu_{-1}-1)}\right)\right]\right\},\end{aligned}$$

for $\nu = \nu_{\max}-1$; and

$$\pi(\nu_{\max}|\mu,\sigma^2,x) \propto \prod_{i=1}^{n}\left\{\frac{1}{B\left(1/2,\nu/2\right)}\left(\frac{1}{\nu\sigma^2}\right)^{1/2}\left(1+\frac{(x_i-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}\right\}\times \frac{\sqrt{\sigma^2\nu_{-1}}B\left(\frac{1}{2},\frac{\nu_{-1}}{2}\right)}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\mathbb{E}_N\left[(x-\mu)^2/\sigma^2\right] +\frac{\nu_{-1}}{2}\mathbb{E}_N\left[\log\left(1+\frac{(x-\mu)^2}{\sigma^2\nu_{-1}}\right)\right]\right\}$$

for $\nu = \nu_{\max}$. It has to be noted that the posterior distribution is proper as it is finite. Furthermore, the actual posterior for the general case, that is when $\mu \neq 0$ and $\sigma^2 \neq 1$, needs to take into consideration the priors for these parameters. We have chosen proper priors for both the location and the scale parameters so that the posterior distributions are proper as well. In particular, $\pi(\mu)$ is normally distributed and $\pi(\sigma^2)$ has an inverse gamma distribution, both with large variance. However, we have also run the simulations with the well known objective priors, that is $\pi(\mu) \propto 1$ and $\pi(\sigma^2) \propto 1/\sigma^2$, and no significant differences were seen.

The above expressions are not analytically tractable. Thus, to study the posterior distribution of the number of degrees of freedom ν, it is necessary to use Monte Carlo methods.

4.2 Independent and identically distributed sample

For the first simulation study, we have considered drawing an independent and identically distributed sample from a t density with known location and scale parameter, that is $\mu = 0$ and $\sigma^2 = 1$. To be able to compare the results with the objective priors proposed by Fonseca et al. (2008), we have obtained the frequentist mean squared error from the median of the posterior distribution for ν, and the frequentist coverage of the 95% credible intervals. The simulation has been performed for $\nu = 1, \ldots, 20$. We have considered both a relatively small sample size, $n = 30$, and a relatively large sample size $n = 100$. In both cases, our prior has been truncated at $\nu_{\max} = 31$, meaning that we consider $f_{31} \approx N(0, 1)$.

The results of the simulation for $n = 100$ are shown in Figure 2. Although the values of the number of degrees of freedom in the simulations are discrete (i.e. integers from 1 to 20), the plots in the figure represent continuous lines, as this allows for a clearer basis for the comparison. The plot on the left shows the relative mean squared error from the median for ν. This index is given by $\sqrt{MSE(\nu)}/\nu$, with $MSE(\nu) = \mathbb{E}\{(\nu - m)^2\}$, where m represents the median of the posterior $\pi(\nu|x)$, $\nu = 1, \ldots, 20$. We have compared our prior with the prior proposed by Fonseca et al. (2008), that is independence Jeffreys' prior (1) and Jeffreys-rule prior (2). From the simulation results, the performance of the posterior median is good for all the priors if we exclude a slightly smaller relative mean squared error of Jeffreys-rule prior and a frequentist coverage below the 95% threshold for the independence Jeffreys' prior in the initial region of the parameter space.

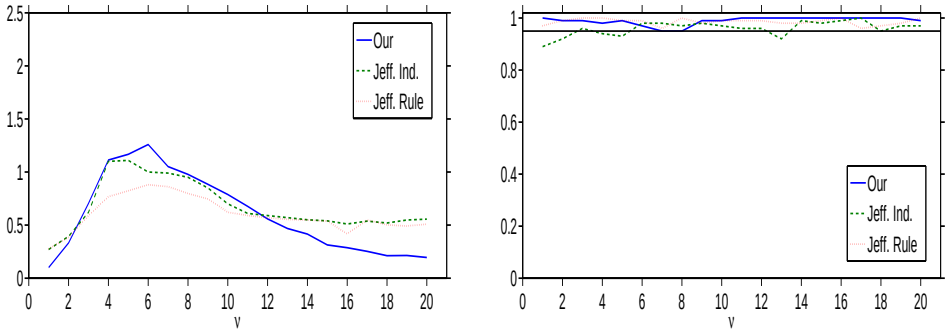

Figure 2: Frequentist properties of our prior (continuous), independence Jeffreys' prior (dashed) and Jeffreys-rule prior (dotted), for $n = 100$. The left figure shows the square root of the relative mean squared error from the median of the posterior for ν. The right figure shows the frequentist coverage of the 95% credible intervals for ν.

Figure 3 shows the results for $n = 30$. As for the previous case, the plots report results by means of continuous lines to ease the analysis and the comparisons. Given that the sample size is relatively small, the mean squared error tends to be larger, in particular for values of ν smaller than 10. The frequentist performance for Jeffreys-rule prior is poor for any value simulated of ν. Our prior and independence Jeffreys' prior have similar performance, with a better relative mean squared error for the latter. To

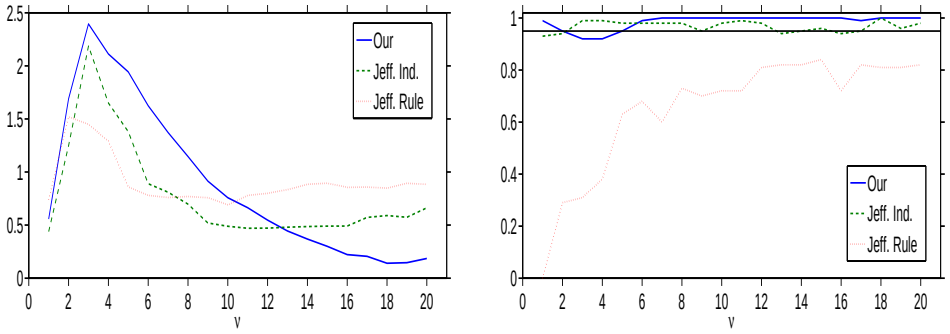

Figure 3: Frequentist properties of our prior (continuous), independence Jeffreys' prior (dashed) and Jeffreys-rule prior (dotted), for $n = 30$. The left figure shows the square root of the relative mean squared error from the median of the posterior for ν. The right figure shows the frequentist coverage of the 95% credible intervals for ν.

have a feeling for the posterior for ν, in Figure 4 we have plotted the results for one simulated sample. In particular, from a t distribution with $\nu = 3$, $\mu = 0$ and $\sigma^2 = 1$: $x \sim t(3, 0, 1)$. The figure includes progressive median, the posterior samples and the histogram of the posterior distribution. In addition, we have reported the statistics of the posterior distribution in Table 3. In the last simulations study for an i.i.d. sample,

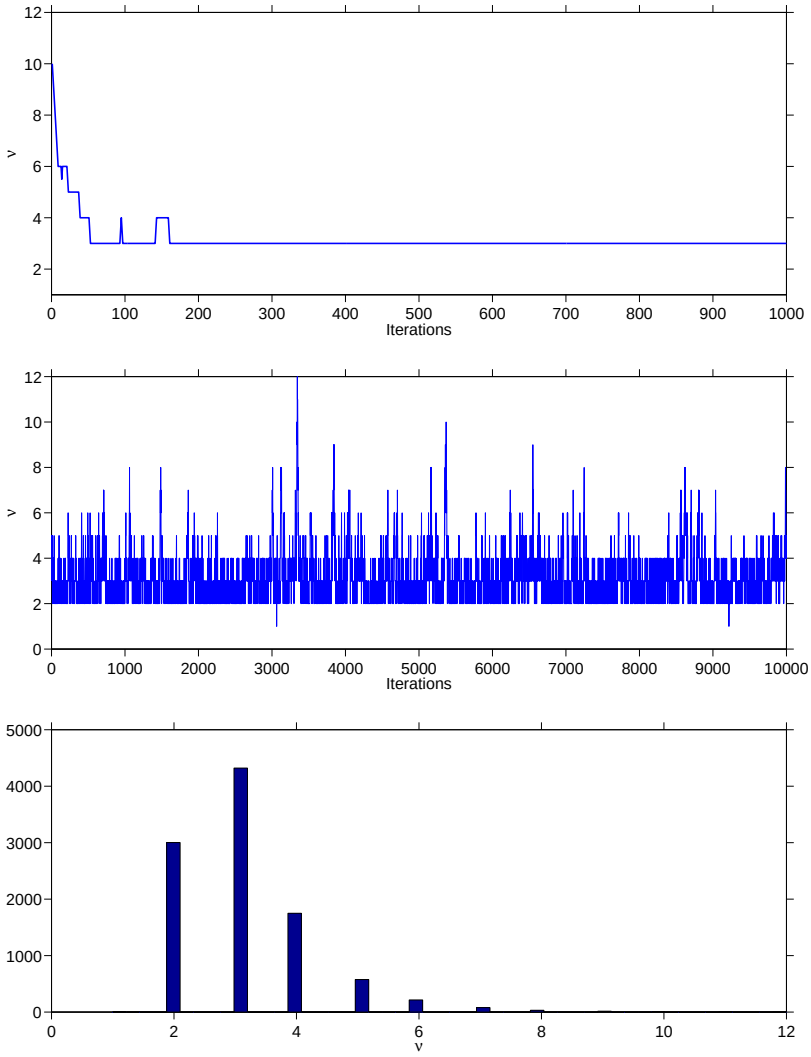

Figure 4: Progressive median (top), posterior sample (middle) and posterior histogram (bottom) of the parameter ν for an independent sample of size $n = 100$ drawn from a t distribution with $\nu = 3$, $\mu = 0$ and $\sigma^2 = 1$.

we have analysed the behaviour of our prior distribution, truncated at $\nu_{\max} = 31$, with data simulated from a t distribution with 50 degrees of freedom. It can then be treated as if the data was originated by a standard normal model. In this circumstance, it is more meaningful to analyse one sample only. In fact, the frequentist coverage of the 95% credible interval is zero, as the true value ($\nu = 50$, in this case) is never included in the interval. In Figure 5, we have plotted the posterior distribution for ν. The

Parameter	Mean	Median	C.I. (95%)
ν	3.12	3	(2, 6)
μ	0.08	0.08	(-0.17, 0.33)
σ^2	1.09	1.07	(0.70, 1.59)

Table 3: Posterior mean, median and 95% credible interval for the simulated data from a t distribution with $\nu = 3$, $\mu = 0$ and $\sigma^2 = 1$, using our prior.

posterior distribution tends to accumulate towards the truncation point, suggesting a normal model or a t density with a relatively high number of degrees of freedom, which is approximatively equivalent.

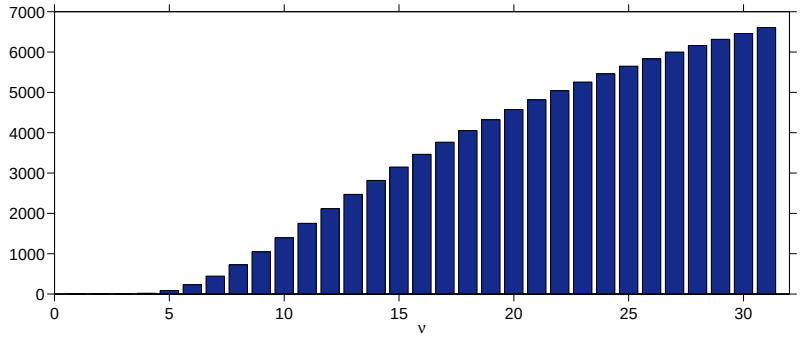

Figure 5: Posterior distribution of the parameter ν with data simulated from a t density with 50 degrees of freedom.

4.3 Regression model

The second simulation study we carried out is on a regression model where the errors are t distributed. This is quite typical when financial quantities are involved. In fact, the distribution of these quantities tends to have heavier tails than a normal density; therefore, departure from normality for the errors has to be expected, and the likely presence of outliers needs to be considered in order to have robust estimates. In particular, it has been shown that the t distribution is, quite often, more appropriate than the normal distribution to model the error terms of a linear regression model. For our simulation study, we have considered a model with four covariates

$$y_i \mid x_i \sim t(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_4 x_{4i}, \sigma^2 | \nu) \qquad i = 1, \ldots, n,$$

where β_0, β_1, β_2, β_3 and β_4 are the regression parameters, σ^2 the regression variance and ν the number of degrees of freedom of the t-distributed errors. For the purpose of this simulation, we have set $\beta_0 = 2$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = 0.9$, $\beta_4 = 1$, $\sigma^2 = 1.5$ and

$\nu = 1, \ldots, 20$. Furthermore, to have a direct comparison with the results in Fonseca et al. (2008), we have assumed the covariates are independent. Similarly as in Section 4.2, the analysis has been carried out for a relatively large sample size ($n = 100$) and a relatively small sample size ($n = 30$). The results of the simulations are shown in Figure 6 where, as done in Section 4.2, although the parameter is discrete, we have used continuous lines to improve the readability of the graphs. The figure shows the

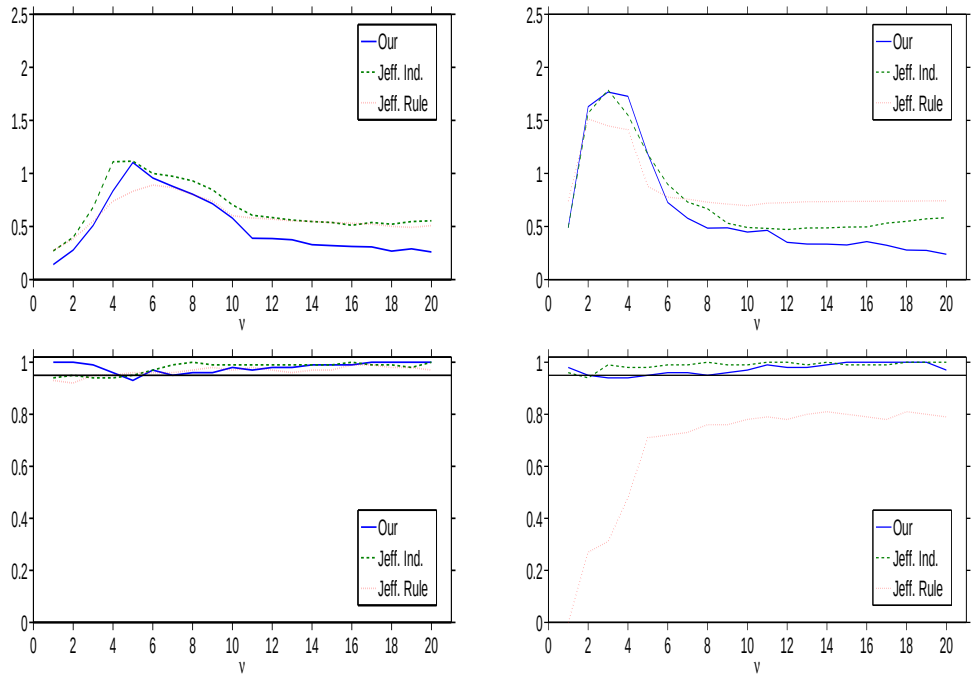

Figure 6: Frequentist result of the simulation of the regression model with $n = 100$ (left side) and $n = 30$ (right side). For each sample size, we have the mean squared error from the median (top graph) and the frequentist coverage of the 95% credible interval (bottom graph). Each plot has our prior (continuous), independent Jeffreys' prior (dashed) and Jeffreys-rule prior (dotted).

frequentist performances of our prior together with the ones of the independent Jeffreys' prior and Jeffreys-rule prior. We note that the mean squared error, in general, tends to be relatively large for values of ν between 2 and 8; as expected, the value of the index is larger for the case $n = 30$ than for the case $n = 100$. In terms of coverage, our prior performs well in both cases; the independence Jeffreys' prior performs well too, whilst the Jeffreys-rule prior has a drop in the performance when $n = 30$.

We have selected a particular sample from a regression model with one covariate to illustrate the behaviour of the posterior for ν. This model has, in particular, $\beta_0 = \beta_1 = 10$, $\sigma^2 = 4$ and $\nu = 5$. Figure 7 shows the chains of the simulation for each parameter, alongside the histogram of the posterior. The sample has size $n = 100$ and has been

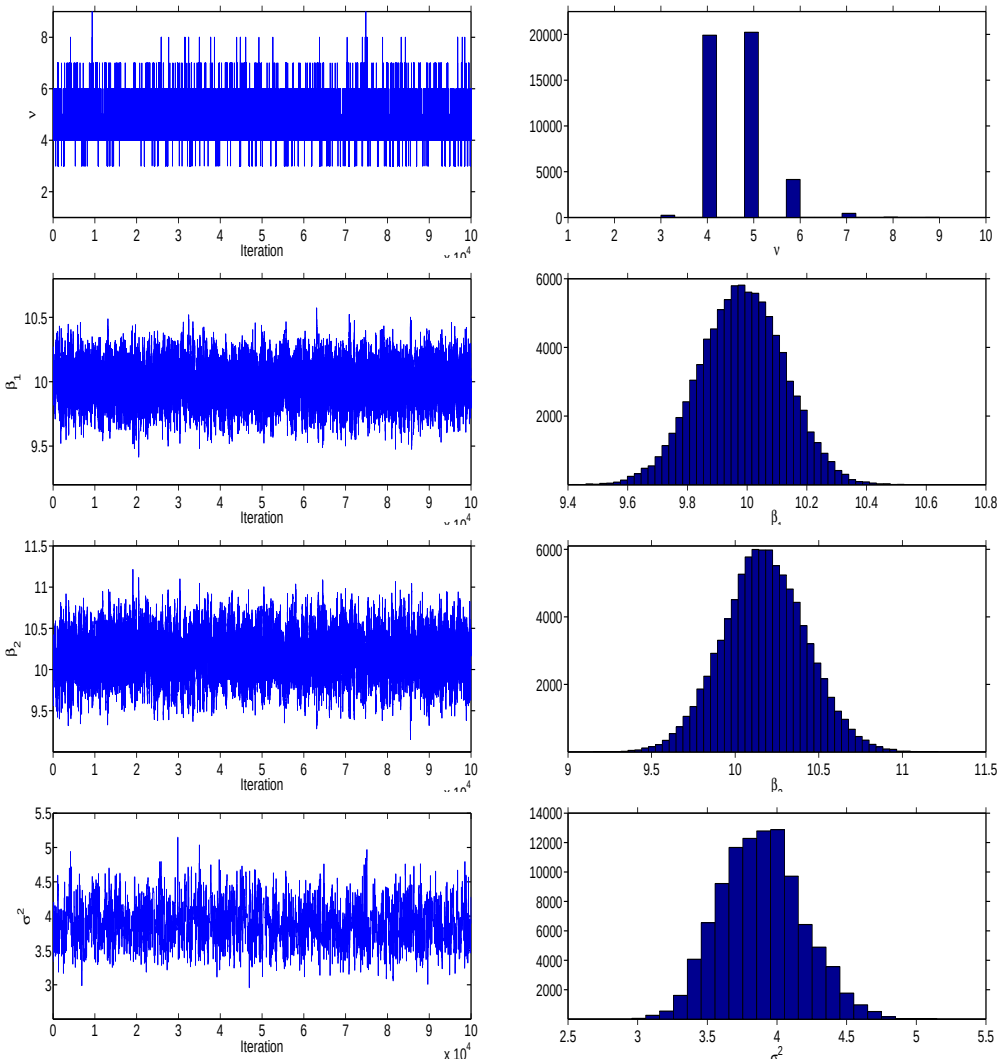

Figure 7: Sample (left) and histogram (right) of the posterior distributions for regression parameters ν (top), β_0 (middle-top), β_1 (middle-bottom) and σ^2 (bottom). The parameters of the regression model from which the data were sampled were $\nu = 5$, $\beta_0 = 10$, $\beta_1 = 10$ and $\sigma^2 = 4$.

drawn from a regression model with a t distributed error term with $\nu = 5$. Table 4 reports the summary statistics of the four posteriors.

Parameter	Mean	Median	C.I. (95%)
ν	5	4.67	(4, 6)
β_0	9.99	9.99	(9.70, 10.26)
β_1	10.17	10.17	(9.68, 10.67)
σ^2	3.89	3.87	(3.36, 4.50)

Table 4: Posterior median and 95% credible interval for the regression simulation. The parameters were set to $\nu = 5$, $\beta_0 = 10$, $\beta_1 = 10$ and $\sigma^2 = 4$.

5 Application

To illustrate the proposed prior on real data, we analyse a sample of the daily closing values of the Dow Jones Industrial Average index of the U.S. stock market. In particular, we consider the data from 11 November 2008 to 4 May 2009, that is 98 observations. This data sample is part of a wide sample analysed in Lin et al. (2012), which ranged from 22 October 2008 to 22 October 2009. Given that the objective of Lin et al. (2012) was to estimate variance change-points in the series, we have focussed our analysis on a subset with estimated constant variance. The actual analysis has been performed on the daily returns, multiplied by 100. That is, $X_d = \{(Y_{d+1} - Y_d)/Y_d\}\,100$, where Y_d is the market index at day d. The transformed data, for the period of interest, is plotted in Figure 8. It can be noted that the series is stationary, and that its variance can be reasonably considered as constant (for the period). In Table 5 we have reported some

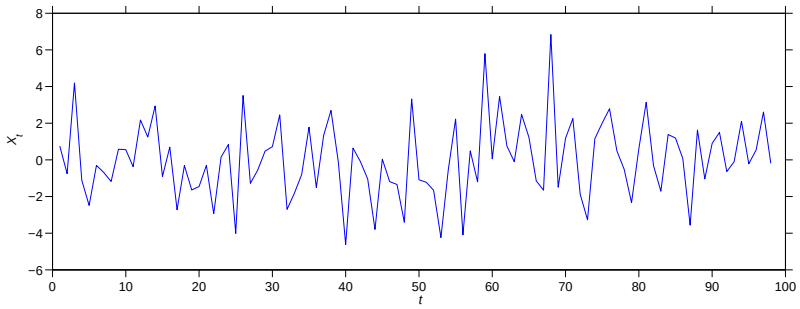

Figure 8: Daily returns (multiplied by 100) of the closing Dow Jones index from 11 November 2008 to 4 May 2009.

basic descriptive statistics of the series. The kurtosis is larger than 3 and even though the distribution of the returns does not have tails much heavier than a normal, it seems to be appropriate to consider a t model. Specifically, the model is

$$X_d = \mu + \varepsilon_d \qquad d = 1, \ldots, 98,$$

where $\varepsilon_d \sim t(0, \sigma^2, \nu)$. The results of the simulation are compared, when appropriate, with the ones in (Lin et al. 2012).

Mean	0.0035
Variance	4.4813
Skewness	0.3216
Kurtosis	3.5626

Table 5: Descriptive statistics of the daily Dow Jones index returns from 11 November 2008 to 4 May 2009.

Parameter	Mean	Median	C.I. (95%)
ν	9.96	8	(2, 26)
μ	-0.05	-0.05	(-0.45, 0.36)
σ^2	3.07	3.21	(0.03, 5.61)

Table 6: Mean, median and 95% credible interval for the number of degrees of freedom, location and scale parameters for the daily returns of the Dow Jones index, from 11 November 2008 to 4 May 2009.

We have obtained the posterior distributions for the three parameters by Markov chain Monte Carlo simulation methods. In Figure 9 we have plotted the posterior sample, the progressive median and the posterior histogram of the number of degrees of freedom ν only. As the posterior distribution of ν is skewed, the median represents the appropriate estimate of the true value of the parameter. The posterior statistics of the parameters are reported in Table 6. The results from (Lin et al. 2012) are, $\nu = 8.4873$, $\mu = -0.0406$ and $\sigma^2 = 3.3749$. It has to be noted that the estimate of the degrees of freedom and the mean μ are relative to a larger data set, in particular, for the first 133 observations. However, the authors conclude that the number of degrees of freedom for the whole data set is homogeneous and in the range 6.68–8.49. The median of the posterior distribution, representing our estimate of the parameter value, is 8 degrees of freedom. We can then conclude that our estimate of ν is in agreement with Lin et al. (2012).

We have analysed the data by adopting a prior different from ours. In addition to the independence Jeffreys' prior and the Jeffreys-rule prior proposed by Fonseca et al. (2008), we have considered the non-hierarchical prior proposed by Juárez and Steel (2010) in (3). The resulting posterior statistics are summarised in Table 7. We see that both the independence Jeffreys' and the Jeffreys-rule prior give estimation results that do not differ from ours, considering that our prior assumes ν discrete whilst both Jeffreys' do not. However, the credible interval of the Jeffreys-rule prior is larger than the one obtained with our prior and independence Jeffreys'. For the Dow Jones index data analysed here, the posterior median of ν obtained by applying the gamma prior proposed by Juárez and Steel (2010) is in contrast with our results.

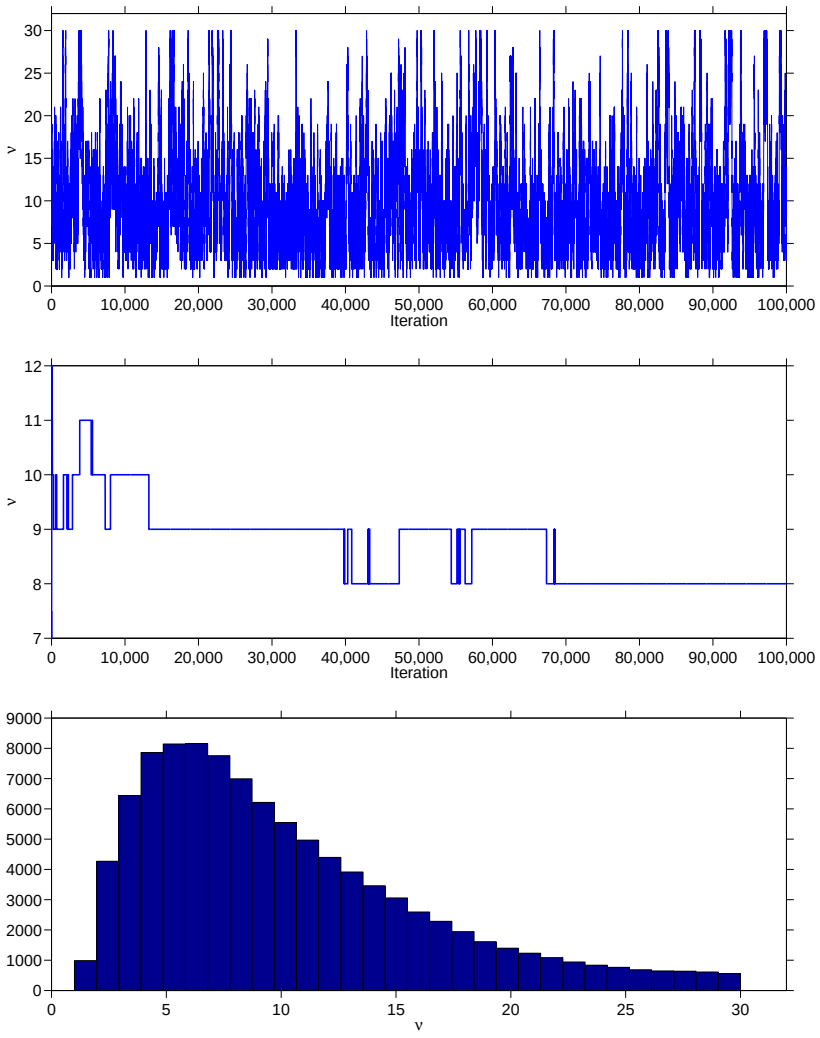

Figure 9: Posterior samples (top), progressive median (middle) and posterior histogram (bottom) for the parameter ν.

6 Discussion

The adoption of t distributed models is an important area of application in finance. This can either be the application of t-distributed random variable to model a certain quantity, such as financial returns, or the assumption that the errors of a linear regression model should have heavier tails than the ones of the more commonly adopted normal distribution. While objective priors for continuous parameters, such as the mean or the variance, can be obtained with several popular approaches, the estimation of the

Prior	Median	C.I. (95%)
$\pi_I(\nu)$	7.30	(3.80, 25.44)
$\pi_J(\nu)$	8.63	(3.46, 31.98)
$\pi_1(\nu)$	15.32	(4.90, 28.89)

Table 7: Posterior statistics obtained by using the independence Jeffreys' prior ($\pi_I(\nu)$), the Jeffreys-rule prior ($\pi_J(\nu)$) and the non-hierarchical gamma prior proposed by Juárez and Steel (2010) ($\pi_1(\nu)$).

number of degrees of freedom of a t distribution is not so straightforward.

The contribution of this paper is threefold. It introduces a new approach to define objective priors based, not on probabilities, but on loss functions, via the *worth* of a particular parameter value being included in the model. The second contribution is that this approach can be consistently applied to any discrete parameter space and does not require any parameter manipulation to be effective. In particular, we have applied it (Villa and Walker 2013) to the discrete scenarios discussed in Berger et al. (2012). The last important result is that an objective prior on the number of degrees of freedom of a t distribution has to be truncated. This is a consequence of the fact that the t distribution converges, in distribution, to the normal distribution. Therefore, for a sufficiently large number of degrees of freedom, the model can be considered as normal and it represents the last element in the set of the option models. We have performed simulations for different truncation points of the prior distribution, namely for $\nu_{\max} = 60$ and $\nu_{\max} = 90$. The mean squared error increases when the truncation point increases. This is due to the fact that the uncertainty about the parameter value becomes larger and larger. However, estimates are not affected for $\nu = 1, \ldots, 20$, in terms of frequentist coverage of credible intervals. We would also add that taking the truncation point up to 60 implies an interest in discriminating between a t_{45} and a t_{50}, for example. This is not practical or desirable.

As mentioned above, our objective approach can be applied to any discrete parameter space. Therefore, it seems appropriate to briefly discuss the main differences between our approach and the one proposed in Berger et al. (2012). We have seen that our approach depends only on the choice of the model. Once this has been selected, the objective prior on the discrete parameter space is obtained by minimising the Kullback–Leibler divergence from the model defined by each element in the parameter space. In Villa and Walker (2013) we have demonstrated our approach on the five models considered in Berger et al. (2012): a population size model, the univariate and the multivariate hypergeometric models, the binomial-beta model and the binomial model. The essence of the Berger et al. (2012) approach is to embed the discrete model into a continuous one such that the structure is preserved. Then, reference analysis is applied to the continuous model. The authors identify four different embedding methods, as it had not been possible to identify one method that can be applied to all the five considered models. This obviously differs from our approach. Furthermore, Berger et al. (2012) may obtain more than one prior per model, and further analysis is required to choose

the most convenient prior distribution. We believe that the innovation of our approach is mainly in this aspect, as the methods to choose between priors add subjectivity to the whole process.

The efficiency of the designed prior for the number of degrees of freedom of a t distribution has been demonstrated through two simulations. The first one is based on data simulated from a t density with given parameter values, and the second from data simulated from a given regression model. We have also performed an analysis on real data: the daily returns of the closing Dow Jones index over a period of 98 days.

It is worth mentioning that we are currently working on applying the proposed objective approach to continuous parameter spaces. The criterion in assigning a mass to a parameter value on the basis of its *worth*, represented by the Kullback–Leibler divergence to the nearest model, is still valid, as the following result in Blyth (1994) supports:

$$\lim_{\delta \to 0} \frac{1}{\delta^2} D_{KL}(f(x|\theta) \| f(x|\theta + \delta)) = \sum_{i,j} I_{i,j}(\theta),$$

where $I_{i,j}(\theta)$ is the (i,j)-th element of the Fisher information matrix. In Brown and Walker (2012) it has been shown that the approach, in the continuous case, may yield Jeffreys' prior.

References

Anscombe, F. J. (1967). "Topics in the investigation of linear relations fitted by the method of least squares." *Journal of the Royal Statistical Society, Series B*, 29(1): 1–52. 200

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag. 199

Berger, J. O., Bernardo, J. M., and Sun, D. (2012). "Objective priors for discrete parameter spaces." *Journal of the American Statistical Association*, 107(498): 636–648. 208, 218

Berk, R. H. (1966). "Limiting behaviour of posterior distributions when the model is incorrect." *Annals of Mathematical Statistics*, 37: 51–58. 199

Blyth, S. (1994). "Local divergence and association." *Biometrika*, 81(3): 579–584. 219

Brown, P. J. and Walker, S. G. (2012). "Bayesian priors from loss matching." *International Statistical Review*, 80(1): 60–82. 219

Chu, J. T. (1956). "Errors in normal approximations to the y, τ, and similar types of distribution." *Annals of Mathematical Statistics*, 27: 780–789. 200, 204

Dmochowski, J. (1996). "Intrinsic priors via Kullback–Leibler geometry." In *Bayesian Statistics 5*, 543–549. London: Oxford University Press. 205

Fabozzi, F. J., Focardi, S. M., Höchstötter, M., and Rachev, S. T. (2010). *Probability and Statistics for Finance*. Hoboken, New Jersey: John Wiley & Sons, Inc. 197

Fernandez, C. and Steel, M. F. (1999). "Multivariate Student-t regression models: pitfalls and inference." *Biometrika*, 86(1): 153–167. 200

Fonseca, T. C. O., Ferreira, M. A. R., and Migon, H. S. (2008). "Objective Bayesian analysis for the Student-t regression model." *Biometrika*, 95(2): 325–333. 201, 202, 209, 213, 216

Geweke, J. (1993). "Bayesian treatment of the independent Student-t linear model." *Journal of Applied Econometrics*, 8: S19–S40. 201, 202

Jacquier, E., Polson, N. G., and Rossi, P. E. (2004). "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors." *Journal of Econometrics*, 122: 185–212. 200, 202

Jeffreys, H. (1961). *Theory of Probability*. New York: Oxford University Press, 3rd edition. 201

Juárez, M. A. and Steel, M. F. J. (2010). "Model-based clustering of non-Gaussian panel data based on skew-t distributions." *Journal of Business and Economic Statistics*, 28(1): 52–66. 201, 202, 216, 218

Kullback, S. and Leibler, R. A. (1951). "On information and sufficiency." *Annals of Mathematical Statistics*, 22: 79–86. 203

Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989). "Robust statistical modelling using the t distribution." *Journal of the American Statistical Association*, 84(408): 881–896. 197

Lin, J. G., Chen, J., and Lin, Y. (2012). "Bayesian analysis of Student t linear regression with unknown change-point and application to stock data analysis." *Computational Economics*, 40: 203–217. 215, 216

Maronna, R. A. (1976). "Robust m-estimators of multivariate location and scatter." *Annals of Statistics*, 4: 51–67. 197

Merhav, N. and Feder, M. (1998). "Universal prediction." *IEEE Transactions on Information Theory*, 44: 2124–2147. 198

Villa, C. and Walker, S. G. (2013). "An objective approach to prior mass functions for discrete parameter spaces." Under Revision by *Journal of the American Statistical Association*. 218

West, M. (1984). "Outlier models and prior distributions in Bayesian linear regression." *Journal of the Royal Statistical Society, Series B*, 46: 431–439. 197

Acknowledgments

The authors are grateful for the comments of two referees, an Associate Editor and the Editor for constructive comments on earlier versions of the paper.