# Bayesian Sensitivity Analysis and Model Comparison for Skew Elliptical Models

I. Vidal [a,*,1], P. Iglesias [b,2], M. D. Branco [c] and
R. B. Arellano-Valle [b,3]

[a]*Instituto de Matemática y Física, Universidad de Talca, Casilla 721, Chile*
[b]*Pontificia Universidad Católica de Chile*
[c]*Universidade de São Paulo, Brazil*

**Abstract**

In this work we approach the problem of model comparison between skew families. For the univariate skew model, we measure the sensitivity of the skewness parameter using the $L_1$-distance between symmetric and asymmetric models and we obtain explicit expressions for some of these models. The main result is that the $L_1$-distance between a representable elliptical distribution and a representable skew elliptical distribution remains invariant and it equals to the $L_1$-distance between the normal and skew normal densities. We also use the Bayes factor to test asymmetry and present some simulation results for the skew-normal and skew-t distributions obtaining expected results for adequate prior distribution. An application in stock markets is also considered.

*Key words:* Skew distribution, skew-normal distribution, representable skew elliptical distribution, Bayes factor, $L_1$-distance.
*AMS subject classification number:* 62F15 and 62F03.
*Short running title:* Bayesian Sensitivity Analysis and Model Comparison.

# 1 Introduction

In a substantial number of applications symmetric (elliptical) models, in particular the normal model, have been found to be restrictive and more realistic models are needed. The main focus in the present work is when the random quantities exhibits skewness. Recently, new alternative models have been developed with the goal of preserving the good properties of the elliptical models and also to be more flexible to model more realistically a data set. These more general models allow us to control the skewness and the kurtosis of the distribution and it includes the normal case as a special one. However, these more flexible models increase the mathematical complexity. Computational techniques can solve partially the problem, even though, some mathematical calculation needs to be done in order to obtain accurate results.

The idea proposed by Azzalini (1985) in the context of the normal distribution, introduce skewness in a symmetric distribution in the following way. If $f$ and $g$ are symmetric probability density function (pdf) around zero and $G$ is a continuous cumulative distribution function (cdf) associated with $g$, then

$$\frac{2}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) G\left(\lambda\frac{x-\mu}{\sigma}\right) \tag{1}$$

is a skew pdf for any $\lambda \in \mathbb{R}$. Where $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter and $\lambda$ is a skewness parameter. When $\lambda = 0$ we obtain the symmetric location-scale pdf, $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$. Different choices of the $f$ and $G$ functions give us important special cases. For example, if $f = \phi$ and $G = \Phi$, the pdf and cdf of normal distribution respectively, we have the skew-normal distribution which is denoted by $SN(\lambda, \mu, \sigma)$.

The elliptical model given by Kelker (1970) is another well known generalization of the normal model. This model has been studied, for example, by Cambanis et al. (1981), Fang et al. (1990) and Arellano-Valle (1994). The elliptical model includes a vast variety of important distributions (the Student-t distribution, double exponential, Pearson type II) and also has good properties, for example they are closed under marginalization and conditioning. The symmetry of the normal model is preserved, but different kurtosis coefficient are allowed.

Extension of the normal model connecting the two ideas, skewness and heavy tails, have been studied by Branco and Dey (2001) and Genton and Loperfido (2001). An interesting special case is the skew-t distribution, which could be represented by a mixture of skew normal models.

In this paper we approach the problem of model comparison within skew-elliptical families. In Section 2, we measure the sensitivity of the skewness

parameter using the $L_1$-distance between the symmetric and asymmetric models. Computation of the Bayes factor to examine asymmetry is presented in Section 3. Also, in Section 4 we present simulation results for the skew-normal and skew-t distributions. An application in Chilean's stock markets is also considered.

## 2   Sensitivity Analysis for the Skewness Parameter

In this section, we analyze the sensitivity of the skewness parameter on the model given by (1), by measuring this sensitivity using the $L_1$ distance between the following models

$$
\begin{aligned}
M_0 &: \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) \\
M_1 &: \frac{2}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) G\left(\lambda \frac{x-\mu}{\sigma}\right),
\end{aligned}
\tag{2}
$$

where the factor $2G\left(\lambda \frac{x-\mu}{\sigma}\right)$ can be interpreted as the asymmetry degree.

Interesting questions are, for example: 1.How much different are $M_0$ and $M_1$? 2.Is it possible to obtain an expression as function of $\lambda$?

In Figure 1, we plot the skew normal pdf for three different values of $\lambda$.
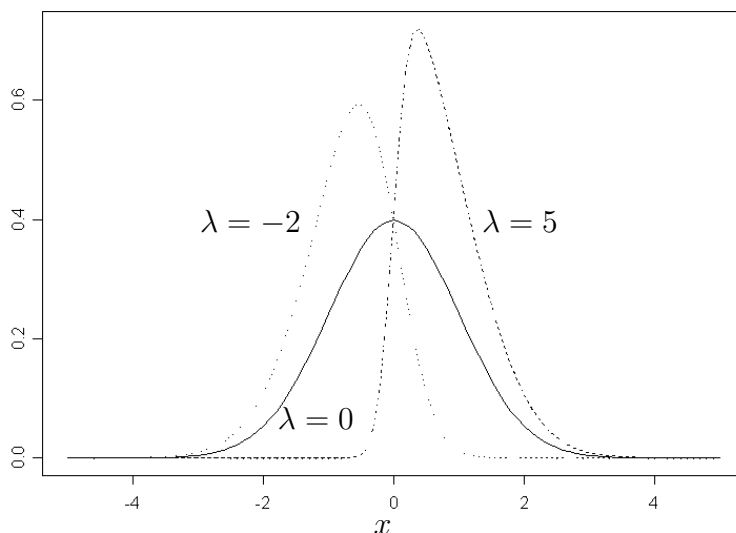


Fig. 1. Skew-normal densities for $\lambda = -2$, $\lambda = 0$ and $\lambda = 5$.

3

Although, there are many ways to measure the distance between two pdf's, we use here the $L_1$ distance because it has an easy and nice interpretation (see, for example, Peng and Dey (1995), Weiss (1996) and Arellano-Valle et al. (2000)). The $L_1$ distance between two densities $f_1$ and $f_2$ is given by

$$L_1\left(f_1, f_2\right) = \frac{1}{2} \int \left|f_1\left(x\right) - f_2\left(x\right)\right| dx = \sup_{A \in \mathcal{B}} \left|\mathbb{P}\left(A \,|\, f_1\right) - \mathbb{P}\left(A \,|\, f_2\right)\right|,$$

where $\mathcal{B}$ are the Borel's sets and $\mathbb{P}\left(\cdot \,|\, f\right)$ denotes the probability measure defined by the density $f$. Thus, $L_1\left(f_1, f_2\right)$ is an upper bound on the differences $\left|\mathbb{P}\left(A \,|\, f_1\right) - \mathbb{P}\left(A \,|\, f_2\right)\right|$ for any set $A$. Also, the $L_1$ distance is bounded and takes values in $[0, 1]$, where $L_1\left(f_1, f_2\right) = 0$ implies that $f_1\left(x\right) = f_2\left(x\right)$ for all $x$ values, and $L_1\left(f_1, f_2\right) = 1$ indicates that the supports of the two densities are disjoint, indicating maximal discrepancy. Generally, it is difficult to obtain explicit expressions for the $L_1$ distance, even in simple cases. However, in our case, the following proposition provides an useful expression to compute and understand this distance.

**Proposition 1** *For any $\mu$ and $\sigma$ fixed, the $L_1$ distance between $M_0$ and $M_1$, specified in (2), is*

$$L_1(M_0, M_1) = \mathbb{E}_{f^*}\left[G\left(\left|\lambda\right| Z\right)\right] - \frac{1}{2}, \tag{3}$$

*where $f^*\left(z\right) = 2f\left(z\right) I_{[0,+\infty)}\left(z\right)$ (f left truncated of zero).*

**PROOF.** From (2) and by letting $z = \frac{x-\mu}{\sigma}$, we have

$$L_1(M_0, M_1) = \int_{\mathbb{R}} \left|\frac{1}{2} - G\left(\lambda z\right)\right| f\left(z\right) dz$$
$$= \int_{-\infty}^{0} \left|\frac{1}{2} - G\left(\lambda z\right)\right| f\left(z\right) dz + \int_{0}^{\infty} \left|\frac{1}{2} - G\left(\lambda z\right)\right| f\left(z\right) dz.$$

Now, by the symmetry of $g$, we notice that if $z > 0$ and $\lambda > 0$ or if $z < 0$ and $\lambda < 0$, then

$$\frac{1}{2} - G\left(\lambda z\right) = -\int_{0}^{\lambda z} g\left(u\right) du \quad \text{and} \quad \left|\frac{1}{2} - G\left(\lambda z\right)\right| = \int_{0}^{\lambda z} g\left(u\right) du.$$

On the other hand, if $z < 0$ and $\lambda > 0$ or if $z > 0$ and $\lambda < 0$, then

$$\frac{1}{2} - G\left(\lambda z\right) = \int_{\lambda z}^{0} g\left(u\right) du = \int_{0}^{-\lambda z} g\left(u\right) du > 0.$$

Thus, if $\lambda > 0$, then

4

$$L_1(M_0, M_1) = \int_{-\infty}^0 \int_0^{-\lambda z} g(u)\, du\, f(z)\, dz + \int_0^\infty \int_0^{\lambda z} g(u)\, du\, f(z)\, dz$$

$$= 2 \int_0^\infty \int_0^{\lambda z} g(u)\, du\, f(z)\, dz = \mathbb{E}_{f^*}[G(\lambda Z)] - \frac{1}{2}.$$

Similarly, if $\lambda < 0$, then

$$L_1(M_0, M_1) = 2 \int_0^\infty \int_0^{-\lambda z} g(u)\, du\, f(z)\, dz = \mathbb{E}_{f^*}[G(|\lambda| Z)] - \frac{1}{2}.$$

□

Note that $L_1(M_0, M_1) = L_1(\lambda)$ does not depend on $\mu$ and $\sigma$.

**Corollary 2** *Given $L_1$ as in (3), then*

$$\sup_\lambda L_1(\lambda) = \frac{1}{2}.$$

**PROOF.** Since $G$ is a cdf, follows that $G(|\lambda| z) \to 1$ when $|\lambda| \to \infty$. On the other hand, $\mathbb{E}_{f^*}[G(|\lambda| Z)]$ exist because $0 < G(x) < 1$ for all $x \in \mathbb{R}$. Therefore

$$\lim_{|\lambda| \to \infty} \mathbb{E}_{f^*}[G(|\lambda| Z)] = \mathbb{E}_{f^*}\left[\lim_{|\lambda| \to \infty} G(|\lambda| Z)\right] = 1.$$

□

The result in Corollary 2 implies that the biggest difference between the probabilities from $M_0$ and $M_1$, assigned to any set $A$, is at most $\frac{1}{2}$. The next examples show us some special cases where $L_1$ distance can be obtained in a closed form. In these examples we assume, without loss of generality, $\mu = 0$ and $\sigma = 1$.

*2.1   Some Examples*

**Example 3 (Uniform)** *Let $f(x) = \frac{1}{2}I_{[-1,1]}(x)$ and $G(x) = \frac{x+1}{2}I_{[-1,1]}(x) + I_{(1,+\infty)}(x)$, the pdf and the cdf of the uniform distribution $U_{[-1,1]}$, respectively. Then, $f^*(x) = I_{[0,1]}(x)$, and for any $\lambda > 0$,*

$$G(\lambda x) = \frac{\lambda x + 1}{2}I_{[-1,1]}(\lambda x) + I_{(1,+\infty)}(\lambda x)$$

$$= \frac{\lambda x + 1}{2}I_{[-\frac{1}{\lambda},\frac{1}{\lambda}]}(x) + I_{(\frac{1}{\lambda},+\infty)}(x).$$

*Therefore, after some computations we obtain*

$$L_1(\lambda) = \begin{cases} \frac{1}{2} - \frac{1}{4|\lambda|} & \text{if} \ \ |\lambda| > 1 \\ \frac{|\lambda|}{4} & \text{if} \ \ |\lambda| \leq 1. \end{cases}$$

*The solid line in Figure 2 shows this function.*

**Example 4 (Double Exponential)** *Let* $f(x) = \frac{1}{2}e^{-|x|}I_{\mathbb{R}}(x)$ *and* $G(x) = \frac{e^x}{2}I_{(-\infty,0)}(x) + \left(1 - \frac{e^{-x}}{2}\right)I_{[0,+\infty)}(x)$, *the pdf and the cdf of the double exponential distribution, respectively. Then,* $f^*(x) = e^{-x}I_{[0,+\infty)}(x)$ *and, for each* $\lambda > 0$ *and* $x > 0$ *it follows that*

$$G(\lambda x) = 1 - \frac{1}{2}e^{-\lambda x}.$$

*Then,*

$$L_1(\lambda) = \int_0^\infty \left(1 - \frac{1}{2}e^{-|\lambda|x}\right)e^{-x}dx - \frac{1}{2} = \frac{|\lambda|}{2(|\lambda|+1)}.$$

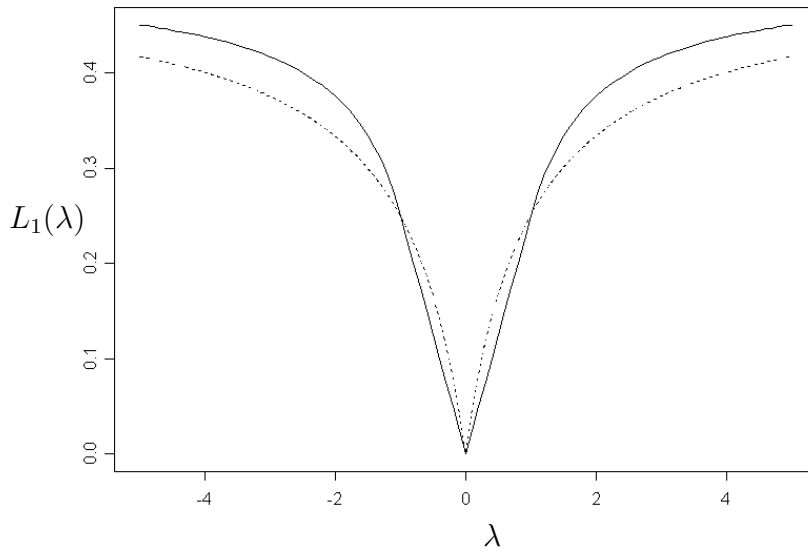*The dashed line in the Figure 2 shows the function given above.*



Fig. 2. $L_1$-distance: solid line for the uniform and skew-uniform densities, and dashed line for double exponential and skew-double-exponential densities.

Hereafter we will denote by $\phi_k(\mathbf{x} - \boldsymbol{\mu} \,|\, \boldsymbol{\Sigma})$ and $\boldsymbol{\Phi}_k(\mathbf{x} - \boldsymbol{\mu} \,|\, \boldsymbol{\Sigma})$ the pdf and the cdf of the $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, respectively. Moreover, when $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_k$ we will denote these functions as $\phi_k(\mathbf{x})$ and $\boldsymbol{\Phi}_k(\mathbf{x})$, respectively. Also, $\mathbb{1}_n$ will denote a vector with $n$ ones.

**Example 5 (Normal)** *The $L_1$ distance between the normal and skew-normal distributions can be also computed. In fact, using (3) we have after some algebraic manipulations that*

$$\mathbb{E}_{\phi^*}[\phi(|\lambda|X)] = 2\mathbf{\Phi}_2\left(\mathbf{0}\,|\,\mathbf{\Omega}\right), \quad \text{where} \quad \mathbf{\Omega} = \begin{pmatrix} 1 & |\lambda| \\ |\lambda| & 1 + \lambda^2 \end{pmatrix}$$

*and $\phi^*\left(x\right) = 2\phi\left(x\right)I_{[0,+\infty)}\left(x\right).$ Thus, using the well known result*

$$\mathbf{\Phi}_2\left(\mathbf{0}\,|\,\mathbf{\Omega}\right) = \frac{1}{4} + \frac{1}{2\pi}\arcsin\left(\frac{|\lambda|}{\sqrt{1+\lambda^2}}\right)$$

*(see Fang et al. (1990)), we have that*

$$L_1(\lambda) = \frac{1}{\pi}\arcsin\left(\frac{|\lambda|}{\sqrt{1+\lambda^2}}\right).$$

*Also, since $\arcsin\theta + \arccos\theta = \frac{\pi}{2}$, this $L_1$ distance can be written as*

$$L_1(\lambda) = \frac{1}{2} - \frac{1}{\pi}\arccos\left(\frac{|\lambda|}{\sqrt{1+\lambda^2}}\right).$$

Extensions of the skew normal model are the skew elliptical models. An interesting subclass of this family is given by representable skew elliptical distributions, which are defined below, where the notation $x \perp\!\!\!\perp y$ means that $x$ and $y$ are independent.

**Definition 6** *The random variable $X|\lambda, \mu, \sigma$ has a representable skew elliptical distribution if its pdf can be written as*

$$f_{X|\lambda,\mu,\sigma}\left(x\right) = \int_0^\infty \frac{2}{\sigma\sqrt{\omega}}\phi\left(\frac{x-\mu}{\sigma\sqrt{\omega}}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma\sqrt{\omega}}\right)dH\left(\omega\right),$$

*where $H$ is the cdf of a non-negative random variable $\omega$ such that $\omega\perp\!\!\!\perp(\lambda,\mu,\sigma)$.*

An equivalent definition is the following: $X|\lambda, \mu, \sigma$ is a representable skew elliptical random variable if and only if there exists a non-negative random variable random variable $\omega \sim H$, with $\omega\perp\!\!\!\perp(\lambda,\mu,\sigma)$, such that $X|\lambda,\mu,\sigma,\omega \sim SN\left(\lambda,\mu,\sigma\sqrt{\omega}\right)$. Properties and examples of this distribution can be found in Branco and Dey (2001). Now, using representable skew elliptical distributions, (2) can be rewritten as

$$M_0 : \int_0^\infty \frac{1}{\sigma\sqrt{\omega}} \phi\left(\frac{x-\mu}{\sigma\sqrt{\omega}}\right) dH(\omega) \tag{4}$$

$$M_1 : \int_0^\infty \frac{2}{\sigma\sqrt{\omega}} \phi\left(\frac{x-\mu}{\sigma\sqrt{\omega}}\right) \Phi\left(\lambda\frac{x-\mu}{\sigma\sqrt{\omega}}\right) dH(\omega).$$

In the next proposition we prove that the $L_1$ distance between the models $M_0$ and $M_1$ considered in (4) is equal to the $L_1$ distance between the normal and skew-normal distributions (see Example 5).

**Proposition 7** *The $L_1$ distance between the models $M_0$ and $M_1$ specified in (4) is*

$$L_1(\lambda) = \frac{1}{\pi} \arcsin\left(\frac{|\lambda|}{\sqrt{1+\lambda^2}}\right).$$

**PROOF.** By letting $z = \frac{x-\mu}{\sigma}$, and making some calculations,

$$L_1(M_0, M_1) = h_1(\lambda) + h_2(\lambda),$$

where

$$h_1(\lambda) = \int_{-\infty}^0 \left| \int_0^\infty \left[\frac{1}{2} - \Phi\left(\lambda\frac{z}{\sqrt{\omega}}\right)\right] \frac{1}{\sqrt{\omega}} \phi\left(\frac{z}{\sqrt{\omega}}\right) dH(\omega) \right| dz$$

and

$$h_2(\lambda) = \int_0^\infty \left| \int_0^\infty \left[\frac{1}{2} - \Phi\left(\lambda\frac{z}{\sqrt{\omega}}\right)\right] \frac{1}{\sqrt{\omega}} \phi\left(\frac{z}{\sqrt{\omega}}\right) dH(\omega) \right| dz.$$

Now, If $\lambda > 0$ and $z < 0$ we have

$$\frac{1}{2} - \Phi\left(\lambda\frac{z}{\sqrt{\omega}}\right) = \Phi\left(-\lambda\frac{z}{\sqrt{\omega}}\right) - \frac{1}{2} > 0.$$

Considering the following change of variable $z = -\sqrt{\omega}y$, we have

$$h_1(\lambda) = \int_0^\infty \left[\int_0^\infty \Phi(\lambda y)\phi(y)\,dy - \frac{1}{4}\right] dH(\omega).$$

For $z > 0$, $\frac{1}{2} - \Phi\left(\lambda\frac{z}{\sqrt{\omega}}\right) < 0$ and considering the change of variable $z = \sqrt{\omega}y$, we have

$$h_2(\lambda) = \int_0^\infty \left[\int_0^\infty \Phi(\lambda y)\phi(y)\,dy - \frac{1}{4}\right] dH(\omega).$$

Therefore, for $\lambda > 0$, we have

$$L_1(M_0, M_1) = \mathbb{E}_{\phi^*}[\Phi(\lambda Y)] - \frac{1}{2},$$

where $\phi^*(x) = 2\phi(x) I_{[0,+\infty)}(x)$.

Similarly, for $\lambda < 0$, we can show that

$$L_1(M_0, M_1) = 2 \int_0^\infty \left[ \int_0^\infty \Phi(-\lambda y) \phi(y) \, dy - \frac{1}{4} \right] dH(\omega) = \mathbb{E}_{\phi^*} \left[ \Phi(|\lambda| Y) \right] - \frac{1}{2}.$$

The proof follows from Example 5. $\quad\square$

Although we obtained closed expressions for the $L_1$ distance in a variety of interesting cases, in other cases numerical methods are necessary to calculate this distance. The following one is an example of this situation.

**Example 8 (Student-t)** *Let* $f(x) = t_1(x - \mu \,|\sigma, \nu)$, *the pdf of the univariate Student-t distribution with location parameter* $\mu$, *scale parameter* $\sigma$ *and* $\nu$ *degree of freedom, and* $G(x) = T_1(x - \mu \,|\sigma, \nu)$, *the corresponding cdf. Assuming that* $\mu = 0$ *and* $\sigma = 1$, *then* $f^*(x) = 2t_1(x \,|\nu) I_{[0,+\infty)}(x)$. *Using Equation (3), we have*

$$L_1(\lambda) = \mathbb{E}_{f^*} \left[ T_1(|\lambda| X| \nu) \right] - \frac{1}{2} = 2 \int_0^\infty T_1(|\lambda| x| \nu) \, t_1(x |\nu) \, dx - \frac{1}{2}.$$

*Note that in this case the skew-t pdf* $\frac{2}{\sigma} t_1(x - \mu \,|\sigma, \nu) T_1(x - \mu \,|\sigma, \nu)$ *is not representable according to Definition 6.*

Figure 3 presents curves of $L_1(\lambda)$ for two groups of models. The solid line shows the distance between $N(0,1)$ and $SN(\lambda)$ densities. The dashed line shows the $L_1$ distance between the Cauchy distribution and $2t_1(x |\nu) T_1(\lambda x |\nu)$ with $\nu = 1$. We calculated $L_1(\lambda)$, given in Example 8, using the S-PLUS integration function, which implements adaptive 15-point Gauss-Kronrod quadrature based on the Fortran function *dqage* and *dqagie* from QUADPACK (Piessens et al. (1983)) in NETLIB (Dongarra and Grosse (1987)).

In the next section we study the influence of $\lambda$ on the posterior distribution for the location-scale parameters.

## 2.2 $L_1$-distance for Posterior Distribution of $(\mu, \sigma)$ under Skew-normal Model.

An important way to measure the sensitivity of the $\lambda$ parameter is to consider its effect on the posterior distribution of $(\mu, \sigma)$. With this objective in mind, we give the following results to compute the posterior distributions by adopting a normal-inverse-gamma prior distribution for $(\mu, \sigma^2)$. Let denote by $\pi(\mu, \sigma | \lambda = 0, \mathbf{x})$ the associated posterior distribution to the corresponding symmetric model with $\lambda = 0$.
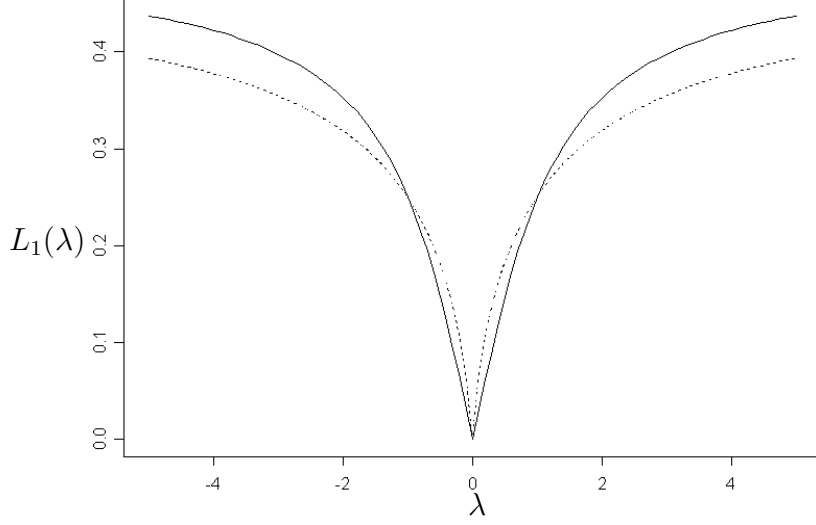
Fig. 3. $L_1$-distance: solid line for the normal and skew-normal densities, and dashed line for Cauchy and skew-Cauchy densities.

**Proposition 9** *If $X \,|\, \lambda, \mu, \sigma \sim SN(\lambda, \mu, \sigma)$ then, under the prior assumptions $\lambda \perp\!\!\!\perp (\mu, \sigma)$, $\mu \,|\, \sigma \sim N\left(m, \frac{\sigma^2}{v}\right)$ and $\sigma^{-2} \sim Ga\,(a, b)$, it follows that*

$$\pi\left(\mu, \sigma \,|\, \lambda, \mathbf{x}\right) = k\left(\mu, \sigma, \lambda, \mathbf{x}\right) \times \pi\left(\mu, \sigma \,|\, \lambda = 0, \mathbf{x}\right),$$

*where*

$$k\left(\mu, \sigma, \lambda, \mathbf{x}\right) = \frac{\Phi_n\left(\lambda \frac{\mathbf{x} - \mu \mathbb{1}_n}{\sigma}\right)}{\mathbf{T}_n\left(\lambda \sqrt{n + 2a} \frac{\mathbf{x} - \hat{\mu} \mathbb{1}_n}{r} \,\middle|\, \boldsymbol{\Sigma}, n + 2a\right)},$$

*with $\hat{\mu} = \frac{n\bar{x} + mv}{n + v}$, $r^2 = ns^2 + \frac{nv}{n+v}(m - \bar{x})^2 + 2b$, $s^2 = n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ and $\mathbf{T}_n(\cdot \,|\, \boldsymbol{\Sigma}, \nu)$ is the cdf of the multivariate $t_n\,(\mathbf{0}, \boldsymbol{\Sigma}, \nu)$ distribution with $\boldsymbol{\Sigma} = \mathbf{I}_n + \frac{\lambda^2}{v+n} \mathbb{1}_n \mathbb{1}_n^t$ and $\nu = n + 2a$.*

**PROOF.** See Appendix A.  □

In the following the notations $\boldsymbol{\mu}$ and $\hat{\boldsymbol{\mu}}$ will be used to indicate $\mu \mathbb{1}_n$ and $\hat{\mu} \mathbb{1}_n$, respectively.

**Corollary 10** *Under conditions of Proposition 9 we have*

$$\pi\left(\mu \,|\, \lambda, \mathbf{x}\right) = k_1\left(\mu, \lambda, \mathbf{x}\right) \times \pi\left(\mu \,|\, \lambda = 0, \mathbf{x}\right)$$

$$\pi\left(\sigma \,|\, \lambda, \mathbf{x}\right) = k_2\left(\sigma, \lambda, \mathbf{x}\right) \times \pi\left(\sigma \,|\, \lambda = 0, \mathbf{x}\right),$$

10

*where*

$$k_1\left(\mu,\lambda,\mathbf{x}\right) = \frac{\mathbf{T}_n\left(\frac{\lambda\sqrt{n+2a+1}}{\sqrt{r^2+(n+v)(\mu-\hat{\mu})^2}}\left(\mathbf{x}-\boldsymbol{\mu}\right)\middle|\,\mathbf{I}_n, n+2a+1\right)}{\mathbf{T}_n\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\boldsymbol{\mu}}}{r}\middle|\,\boldsymbol{\Sigma}, n+2a\right)}$$

*and*

$$k_2\left(\sigma,\lambda,\mathbf{x}\right) = \frac{\boldsymbol{\Phi}_n\left(\lambda\frac{\mathbf{x}-\hat{\boldsymbol{\mu}}}{\sigma}\middle|\,\boldsymbol{\Sigma}\right)}{\mathbf{T}_n\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\boldsymbol{\mu}}}{r}\middle|\,\boldsymbol{\Sigma}, n+2a\right)},$$

*with $\hat{\mu}$ and $r^2$ defined as in Proposition 9.*

The terms $k\left(\mu,\sigma,\lambda,\mathbf{x}\right)$, $k_1\left(\mu,\lambda,\mathbf{x}\right)$ and $k_2\left(\sigma,\lambda,\mathbf{x}\right)$ are the perturbation functions, introduced by Kass et al. (1989) and Weiss (1996). These terms can be interpreted as sensitivity factors because the sensitivity of the skewness parameter depends only on them.

Note that these factors tend to one when $n$ increases. In other words, for large values of the sample size $n$, the two posterior distributions are similar. Also, the previous proposition allows us to calculate the two conditional posterior distributions, $\pi\left(\mu\middle|\sigma,\lambda,\mathbf{x}\right)$ and $\pi\left(\sigma\middle|\mu,\lambda,\mathbf{x}\right)$ which are necessary, for example, in Gibbs sampler-type algorithms.

Proposition 9 enables us to calculate the $L_1$ distance between the posterior distributions $\pi\left(\mu,\sigma\middle|\lambda=0,\mathbf{x}\right)$ and $\pi\left(\mu,\sigma\middle|\lambda,\mathbf{x}\right)$, which is given by

$$\begin{aligned}
L_1(\lambda) &= \frac{1}{2}\int_{-\infty}^{\infty}\int_0^{\infty}\left|\pi\left(\mu,\sigma\middle|\lambda=0,\mathbf{x}\right) - \pi\left(\mu,\sigma\middle|\lambda,\mathbf{x}\right)\right|d\mu d\sigma \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\int_0^{\infty}\left|1 - k\left(\mu,\sigma,\lambda,\mathbf{x}\right)\right|\pi\left(\mu,\sigma\middle|\lambda=0,\mathbf{x}\right)d\mu d\sigma \\
&= \frac{1}{2}\mathbb{E}\left[\left|1 - k\left(M,S,\lambda,\mathbf{x}\right)\right|\right],
\end{aligned}$$

where the expectation is taken under $(M, S^2)$, which has the normal-inverse-gamma posterior distribution.

Now, using the results in Corollary 10, we have that the $L_1$ distance between $\pi\left(\mu\middle|\lambda=0,\mathbf{x}\right)$ and $\pi\left(\mu\middle|\lambda,\mathbf{x}\right)$ is

$$L_1(\lambda) = \frac{1}{2}\mathbb{E}\left[\left|1 - k_1\left(M,\lambda,\mathbf{x}\right)\right|\right],$$

11

where the expectation is taken under $M \sim t_1\left(\hat{\mu}, \frac{r^2}{(n+v)(n+2a)}, n+2a\right)$. Similarly, the $L_1$ distance between $\pi\left(\sigma|\,\lambda=0,\mathbf{x}\right)$ and $\pi\left(\sigma|\,\lambda,\mathbf{x}\right)$ is

$$L_1(\lambda) = \frac{1}{2}\mathbb{E}\left[\left|1 - k_2\left(S^{-1},\lambda,\mathbf{x}\right)\right|\right],$$

where the expected value is taken assuming $S^2 \sim Ga\left(\frac{n}{2}+a, \frac{r^2}{2}\right)$.

The results of Proposition 9 and Corollary 10 depend on the sensitivity factors $k\left(\mu,\sigma,\lambda,\mathbf{x}\right)$, $k_1\left(\mu,\lambda,\mathbf{x}\right)$ and $k_2\left(\sigma,\lambda,\mathbf{x}\right)$. Therefore, these sensitivity factors can be used to study the influence of the skewness parameter $\lambda$ over the joint and marginal posterior distributions of $\mu$ and $\sigma$. For example, Figures 4 and 5 show $\log\left[\bar{k}_2\left(\sigma,\lambda,\mathbf{x}\right)\right]$, where $\bar{k}_2\left(\sigma,\lambda,\mathbf{x}\right)$ is the average of $k_2\left(\sigma,\lambda,\mathbf{x}_1\right),\ldots,$ $k_2\left(\sigma,\lambda,\mathbf{x}_{50}\right)$, and $\mathbf{x}_1,\ldots,\mathbf{x}_{50}$ are data sets with sample size 100 drawn from the $N\left(0,1\right)$ distribution. The $k_2\left(\sigma,\lambda,\mathbf{x}_j\right)$, $j=1,\ldots,50$, were computed for $\lambda$ from 0 to 5 and $\sigma$ from 0 to 10, both with step 0.25. Figure 4 shows the great influence of the $\lambda$ parameter.
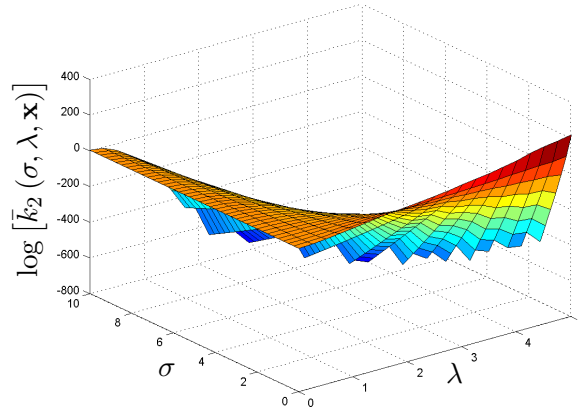


Fig. 4. Surface of $\log\left[\bar{k}_2\left(\sigma,\lambda,\mathbf{x}\right)\right]$ for generated data.

Note also that there are values of $\lambda$ and $\sigma$ for which the posterior distribution of $\sigma$ equals to $\pi\left(\sigma|\,\lambda=0,\mathbf{x}\right)$. They arise for the $k_2\left(\sigma,\lambda,\mathbf{x}\right)=1$ contour line, or equivalently, $\log\left[k_2\left(\sigma,\lambda,\mathbf{x}\right)\right]=0$. Figure 5 shows several contour lines from $\log\left[\bar{k}_2\left(\sigma,\lambda,\mathbf{x}\right)\right]$.

## 3 Bayes Factor

In this section we consider the Bayes factor to perform model comparison. Information about the Bayes factor can be found in Kass and Raftery (1995), Lavine and Schervish (1999) and Berger and Pericchi (2001)). Liseo and Loperfido (2002) used the Bayes factor by adopting reference priors to compare
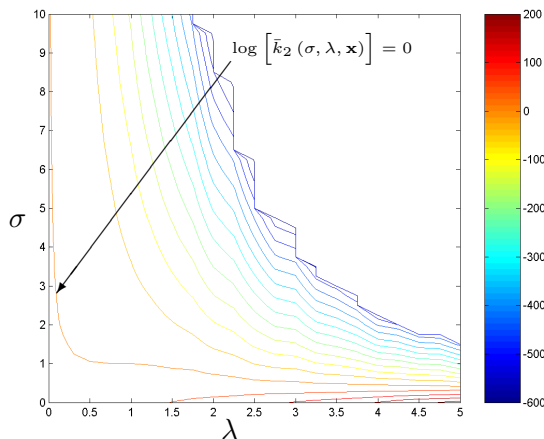
Fig. 5. Contour lines of $\log\left[\bar{k}_2\left(\sigma,\lambda,\mathbf{x}\right)\right]$ for generated data.

normal versus skew normal models. Here we study the Bayes factor using an specific informative prior for the skewness parameter. In Section 3.1 we study the Bayes factor under representable skew elliptical distribution.

We assume that the data set $\mathbf{x}=(x_1,\ldots,x_n)$ comes from $i.i.d.$ random samples given by the models $M_0$ or $M_1$ in (2). Then, the Bayes factor to compare these models, in favor of the $M_0$ (the symmetric model), is given by

$$BF=\frac{\int \sigma^{-n}\left[\prod_{i=1}^{n}f\left(\frac{x_i-\mu}{\sigma}\right)\right]\pi\left(\mu,\sigma\right)d\mu d\sigma}{2^n\int \sigma^{-n}\left[\prod_{i=1}^{n}f\left(\frac{x_i-\mu}{\sigma}\right)G\left(\lambda\frac{x_i-\mu}{\sigma}\right)\right]\pi\left(\mu,\sigma,\lambda\right)d\mu d\sigma d\lambda}, \qquad (5)$$

where $\pi$ is the prior distribution adopted for the respective parameter. As we may note from expression (5), a closed form of Bayes factor is not possible to obtain in general. Also its numeric calculation is complex to implement. However, when $\mu$ and $\sigma$ are known, the Bayes factor have the following simple expression

$$BF_{\mu,\sigma}=\frac{1}{2^n\int\left[\prod_{i=1}^{n}G\left(\lambda\frac{x_i-\mu}{\sigma}\right)\right]\pi\left(\lambda\right)d\lambda}. \qquad (6)$$

Note that the integral associated to (6) can be computed as

$$\int\left[\prod_{i=1}^{n}G\left(\lambda\frac{x_i-\mu}{\sigma}\right)\right]\pi\left(\lambda\right)d\lambda=\mathbb{P}\left(\mathbf{Z}-\Lambda\frac{\mathbf{x}-\boldsymbol{\mu}}{\sigma}\leq\mathbf{0}\right), \qquad (7)$$

where $\mathbf{Z}=(Z_1,\ldots,Z_n)\perp\!\!\!\perp\Lambda$, with $Z_1,\ldots,Z_n\overset{iid}{\sim}G$ and $\Lambda$ has prior pdf $\pi\left(\lambda\right)$. Thus, for the particular case where $G=\Phi$ and $\pi\left(\lambda\right)=\phi\left(\lambda-m\,|v^2\right)$, we have that

$$\mathbf{Z}-\Lambda\frac{\mathbf{x}-\boldsymbol{\mu}}{\sigma}\sim N_n\left[m\frac{\mathbf{x}-\boldsymbol{\mu}}{\sigma},\mathbf{I}_n+(v/\sigma)^2\left(\mathbf{x}-\boldsymbol{\mu}\right)\left(\mathbf{x}-\boldsymbol{\mu}\right)^t\right],$$

13

so that (6) reduces to

$$BF_{\mu,\sigma} = \left[ 2^n \mathbf{\Phi}_n \left( -m \frac{\mathbf{x} - \boldsymbol{\mu}}{\sigma} \,\bigg|\, \mathbf{I}_n + (v/\sigma)^2 (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^t \right) \right]^{-1}.$$

From a practical point of view it is natural to consider the sign of the $\lambda$ parameter known, i.e., we know the direction of the skewness. In this case, the comparison to be considered is an unilateral test for the parameter $\lambda$. For example, if $G = \Phi$ and $\lambda > 0$ with $\lambda^2 \sim Ga(a, b)$, we have, from (7) that

$$\begin{aligned}
\mathbb{P} \left( \mathbf{Z} - \Lambda \frac{\mathbf{x} - \boldsymbol{\mu}}{\sigma} \leq \mathbf{0} \right) &= \int_0^\infty \mathbb{P} \left( \mathbf{Z} \leq \Lambda \frac{\mathbf{x} - \boldsymbol{\mu}}{\sigma} \,\bigg|\, \Lambda = \lambda \right) \pi(\lambda)\, d\lambda \\
&= \int_0^\infty \mathbf{\Phi}_n \left( \lambda \frac{\mathbf{x} - \boldsymbol{\mu}}{\sigma} \right) g\left( \lambda^2 \,|\, a, b \right) d\lambda^2 \\
&= \int_0^\infty \mathbf{\Phi}_n \left( \frac{\mathbf{x} - \boldsymbol{\mu}}{\sigma} \,\bigg|\, \lambda^{-2} \mathbf{I}_n \right) g\left( \lambda^2 \,|\, a, b \right) d\lambda^2 \\
&= \mathbf{T}_n \left( \frac{\mathbf{x} - \boldsymbol{\mu}}{\sigma} \sqrt{\frac{a}{b}} \,\bigg|\, \mathbf{I}_n, 2a \right),
\end{aligned}$$

where $g\left( \cdot \,|\, a, b \right)$ denotes the pdf of the $Ga(a, b)$ distribution. Thus, (6) can be written as

$$BF_{\mu,\sigma} = \frac{1}{2^n \mathbf{T}_n \left( \frac{\mathbf{x} - \boldsymbol{\mu}}{\sigma} \sqrt{\frac{a}{b}} \,\big|\, \mathbf{I}_n, 2a \right)}. \tag{8}$$

A similar expression is obtained when $\lambda < 0$ and $\lambda^2 \sim Ga(a, b)$.

In order to obtain numerical computation of (8) we can use the fact that the Student-t distribution can be specified as a mixture of the normal distribution, so that we can rewrite (8) as

$$\left\{ \int_0^\infty \prod_{i=1}^n \left[ 2\Phi \left( \frac{x_i - \mu}{\sigma} \sqrt{\frac{\omega a}{b}} \right) \right] g\left( \omega \,|\, a, a \right) d\omega \right\}^{-1}. \tag{9}$$

From expressions (6) and (8) we can note that the symmetric pdf $\sigma^{-1} f\left( \frac{x_i - \mu}{\sigma} \right)$ is not necessary.

### 3.1  Bayes Factor for Representable Skew Elliptical Distributions

In the later section we discussed the difficulties to obtain a general form for the Bayes factor. Therefore, it is important to consider restrictions in the functions $f$ and $G$ to make the calculations of (5) simpler and also to keep the class of the asymmetric distributions general. The class of representable skew elliptical distributions fulfills these requirements.

14

**Proposition 11** *Let* $\mathbf{x} = (x_1, \ldots, x_n)$ *be a random sample from a representable skew elliptical distribution. If a priori* $\mu|\sigma \sim N\left(m, \frac{\sigma^2}{v}\right)$ *and* $\sigma^{-2} \sim Ga(a,b)$*, then the Bayes factor (5) is given by*

$$BF(\mathbf{x}) = \frac{\int \cdots \int r^{-n-2a} \left[(\eta + v) \prod_{i=1}^{n} \omega_i\right]^{-\frac{1}{2}} dH(\omega_1) \cdots dH(\omega_n)}{2^{\frac{n}{2}} \int \cdots \int r^{-n-2a} \left[(\eta + v) \prod_{i=1}^{n} \omega_i\right]^{-\frac{1}{2}} g(\boldsymbol{\omega}) dH(\omega_1) \cdots dH(\omega_n)}.$$

*where*

$$g(\boldsymbol{\omega}) = \int \mathbf{T}_n \left(\sqrt{n + 2a} \frac{\mathbf{x} - \hat{\mu} \mathbb{1}_n}{r} \middle| \mathbf{0}, \boldsymbol{\Sigma}, n + 2a\right) \pi(\lambda) d\lambda,$$

$\boldsymbol{\Sigma} = \frac{1}{\lambda^2} \mathbf{D}(\boldsymbol{\omega}) + \frac{1}{v+\eta} \mathbb{1}_n \mathbb{1}_n^t$, $\mathbf{D}(\boldsymbol{\omega}) = diag(\omega_1, \ldots, \omega_n)$, $\eta = \sum_{i=1}^{n} \omega_i^{-1}$, $\hat{\mu} = \frac{\eta}{\eta + v} \left(\sum_{i=1}^{n} \nu_i x_i + vm\right)$, $r^2 = \eta S_\omega^2 + \frac{\eta v}{\eta + v} \left(m - \sum_{i=1}^{n} \nu_i x_i\right)^2 + 2b$, $S_\omega^2 = \sum_{i=1}^{n} \nu_i \left(x_i - \sum_{j=1}^{n} \nu_j x_j\right)^2$ *and* $\nu_i = \frac{\omega_i}{\eta}$ *for each* $i = 1, \ldots, n$.

**PROOF.** See Appendix B. $\square$

Note that $g$ is a mixture of Student-t distributions. In the special case where $\omega_1 = \omega_2 = \cdots = \omega_n = 1$, the result of Proposition 11 agrees with Liseo and Loperfido (2002) result.

## 4 Simulation Results

In this section we perform a simulation study to describe the behavior of the Bayes factor given by (6). In Subsection 4.1, we calculate the integral given in (9) using the MATLAB integration function (*quad*) based on the recursive adaptive Simpson quadrature method.

### 4.1 Normal Versus Skew-Normal Models

For each value of $\lambda$, $\lambda = 0.0, 0.1, \ldots, 0.5$, we generated 1000 independent data sets, $y_1, \ldots, y_n$, from the pdf $2\phi(y) \Phi(\lambda y)$ with $n = 10, 50, 100$. Then, for each data set we calculated (8) by considering $\mu = 0$, $\sigma = 1$ and $\lambda > 0$ with $\lambda^2 \sim Ga(1, b)$, where $b = 0.1, 1, 5$. Figure 6 shows the curves corresponding to the three prior distributions for $\lambda$ considered in this simulation study.

When $\lambda^2 \sim Ga(a, b)$, the prior pdf for $\lambda$ is given by

$$f(\lambda | a, b) = \frac{2b^a}{\Gamma(a)} \left(\lambda^2\right)^{a - \frac{1}{2}} \exp\left(-b\lambda^2\right),$$
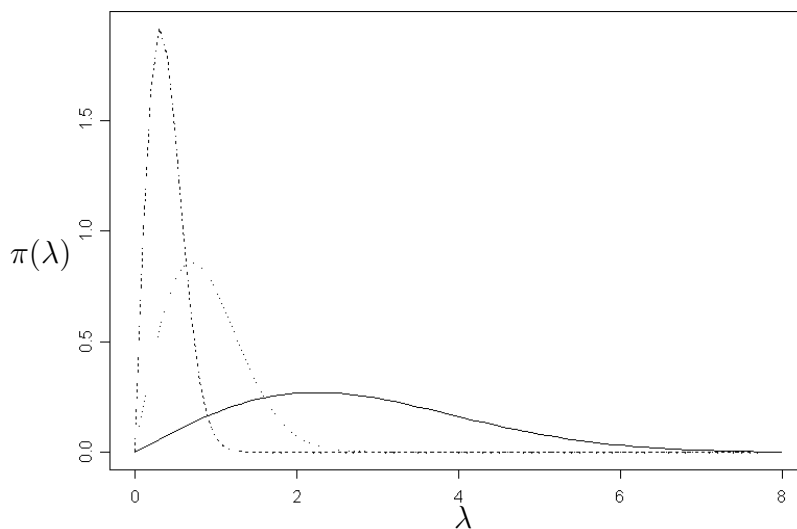
Fig. 6. Prior distributions for $\lambda$: solid line for $b = 0.1$, dotted line for $b = 1$ and dashed for $b = 5$.

with mean and variance given, respectively, by

$$\mathbb{E}(\lambda) = \operatorname{sgn}(\lambda)\frac{\Gamma\left(a + \frac{1}{2}\right)}{\sqrt{b}\,\Gamma(a)} \quad \text{and} \quad \mathbb{V}(\lambda) = \frac{a}{b} - \frac{\Gamma^2\left(a + \frac{1}{2}\right)}{b\,\Gamma^2(a)},$$

where $\operatorname{sgn}(\lambda)$ denote the sign assigned to the parameter $\lambda$. We note that the prior variance of $\lambda$ is always smaller than $\frac{a}{b}$. Thus, if we want to have a big prior variance, then we have to consider $a$ much bigger than $b$. Therefore, in this case the prior mean will be big also. The values for $\mathbb{E}(\lambda)$ and $\mathbb{V}(\lambda)$ considered in this simulation study are presented in the Table 1.

Table 1
Mean and variance for prior distributions in simulation study.

| $b$ | $\mathbb{E}(\lambda)$ | $\mathbb{V}(\lambda)$ | $\frac{a}{b}$ |
|---|---|---|---|
| 0.1 | 2.8 | 2.15 | 10 |
| 1 | 0.89 | 0.21 | 1 |
| 5 | 0.4 | 0.04 | 0.2 |

We computed the Bayes factor for each sample size $n$ and values of $\lambda$ specified above. Here, we considered, as strong evidence in favor of the asymmetric model if $\hat{BF} < 0.5$, and in favor of the symmetric model if $\hat{BF} > 2$. Table 2 displays, in Column 4 and 5, the percentage of the samples that presented evidence in favor the asymmetric and symmetric models, respectively; from Column 6 to 8, the $25^{th}$ percentile, median and $75^{th}$ percentile of the Bayes factor values, respectively.

16

Table 2
Simulation results for Normal versus Skew-normal comparison.

| $b$ | $n$ | $\lambda$ | $\hat{BF} < 0.5$ | $\hat{BF} > 2$ | $25^{th}$ p. | Median | $75^{th}$ p. |
|---|---|---|---|---|---|---|---|
| | | 0 | 4.4 | 89.4 | 8.4193 | 28.4309 | 66.0642 |
| | 10 | 0.1 | 5.1 | 86.9 | 6.9357 | 21.2874 | 52.7914 |
| | | 0.3 | 14.0 | 72.3 | 1.6913 | 7.3060 | 24.1999 |
| | | 0.5 | 24.3 | 58.9 | 0.5541 | 3.5752 | 12.0782 |
| | | 0 | 0.5 | 98.4 | 60.5811 | 158.1004 | 319.9009 |
| 0.1 | 50 | 0.1 | 1.7 | 95.2 | 21.4082 | 67.7103 | 169.2791 |
| | | 0.3 | 12.4 | 72.4 | 1.6805 | 8.4769 | 38.5235 |
| | | 0.5 | 47.1 | 33.3 | 0.0636 | 0.6080 | 3.7100 |
| | | 0 | 0.2 | 99.3 | 121.7138 | 310.8615 | 643.1082 |
| | 100 | 0.1 | 2.3 | 95.2 | 27.9092 | 103.7148 | 261.5884 |
| | | 0.3 | 25.8 | 56.9 | 0.4623 | 3.3920 | 18.7598 |
| | | 0.5 | 81.9 | 9.1 | 0.0016 | 0.0240 | 0.2455 |
| | | 0 | 10.4 | 68.5 | 1.4078 | 3.7729 | 7.8112 |
| | 10 | 0.1 | 12.4 | 63.3 | 1.2134 | 2.9230 | 6.3771 |
| | | 0.3 | 28.1 | 37.5 | 0.4165 | 1.2357 | 3.2828 |
| | | 0.5 | 42.4 | 23.8 | 0.2186 | 0.7602 | 1.8649 |
| | | 0 | 2.1 | 91.5 | 6.9877 | 17.8707 | 36.4650 |
| 1 | 50 | 0.1 | 8.7 | 76.6 | 2.2403 | 7.7218 | 19.0447 |
| | | 0.3 | 38.5 | 36.8 | 0.2347 | 1.0252 | 3.6563 |
| | | 0.5 | 73.6 | 8.6 | 0.0130 | 0.1075 | 0.5436 |
| | | 0 | 1.1 | 96.5 | 13.2021 | 34.0111 | 76.7385 |
| | 100 | 0.1 | 7.2 | 81.6 | 3.1714 | 9.8879 | 25.5317 |
| | | 0.3 | 53.5 | 27.1 | 0.0539 | 0.3897 | 2.3266 |
| | | 0.5 | 93.8 | 2 | 0.0002 | 0.0035 | 0.0342 |
| | | 0 | 11.8 | 38.0 | 0.8493 | 1.5772 | 2.5897 |
| | 10 | 0.1 | 13.8 | 30.0 | 0.7558 | 1.3369 | 2.2663 |
| | | 0.3 | 33.2 | 12.5 | 0.4136 | 0.7660 | 1.4244 |
| | | 0.5 | 42.9 | 5.1 | 0.2952 | 0.5855 | 1.0312 |
| | | 0 | 7.2 | 73.0 | 1.8322 | 4.1703 | 7.7046 |
| 5 | 50 | 0.1 | 16.7 | 50.3 | 0.7710 | 2.0095 | 4.3708 |
| | | 0.3 | 56.5 | 14.9 | 0.1107 | 0.3736 | 1.2628 |
| | | 0.5 | 87.7 | 1.5 | 0.0110 | 0.0523 | 0.2118 |
| | | 0 | 4.3 | 83.0 | 3.0349 | 7.2055 | 14.2496 |
| | 100 | 0.1 | 17.7 | 57.4 | 0.8089 | 2.6087 | 6.1743 |
| | | 0.3 | 72.4 | 7.0 | 0.0232 | 0.1234 | 0.5619 |
| | | 0.5 | 98.4 | 0.0 | 0.0002 | 0.0019 | 0.0136 |

Note that for each sample size, the $\hat{BF}$ decreases when the value of $\lambda$ increases and for $\lambda = 0$ tends to be quite big. This desired behavior is appreciated better for higher sample sizes. In general, the calculated Bayes factors show correct evidence when $\lambda = 0$ and when $\lambda \geq 0.4$, and while the sample size increases, this evidence improves. For $\lambda \geq 1$, the Bayes factor values are almost zero. As we have expected, we can see the strong dependence of the values of Bayes factor from prior specification for small sample sizes. Similar results are

obtained for negatives values of $\lambda$.

## 4.2 Student-t Versus Skew-Student-t Models

For each $\lambda = 0.0, 0.1, \ldots, 0.5$, we generated 1000 independent data sets, $y_1, \ldots, y_n$, with $n = 10, 50, 100$, from the pdf $2t_1 \left( y \left| \nu \right. \right) T_1 \left( \lambda y \left| \nu \right. \right)$. We considered $\nu = 1, 10, 20$ and calculated (6) by considering $\lambda > 0$ with $\lambda^2 \sim Ga\left( 1, 1 \right)$. Thus, Equation (6) becomes in

$$BF_{\mu,\sigma} = \frac{1}{2^{n+1} \int \left[ \prod_{i=1}^n T_1 \left( \lambda y_i \left| \nu \right. \right) \right] \lambda \exp\left( -\lambda^2 \right) d\lambda}.$$

Table 3 exhibits the results of the performed simulations for each different values of $\nu$. Similar results to the normal case were obtained, highlighting those better results obtained for small values of $\nu$, as we have expected from Figure 3. We note that, in this case as well as in the normal case, the BF has a nice behavior for small values of $\lambda$.

## 5   An Application

In this section we use the result (8) in a real data set. The data set becomes from the Chilean's stock markets and it consists in the monthly rentability of five Chilean companies measured between March, 1990 and April, 1999. The sample size is $n = 110$ for each company. Table 4 presents some descriptive statistics, including the skewness and the kurtosis. These descriptive statistics were calculated using S-Plus software.

The Bayes factor under prior distribution $\lambda^2 \sim Ga\left( 1, 1 \right)$ is presented for each company in Table 5. We consider the $\lambda$ sign known, $\lambda > 0$. It could be justified because in this period the Chilean companies presented an affluent economy.

From Table 5, we can see the only data set in favor of the symmetry assumption is the data from Iansa company. Also, the positive asymmetry is more evident in the company Chilquinta following by company Cervezas.

## 6   Final Remarks

In this work we evaluate the sensitivity of the skewness parameter using the $L_1$-distance between the symmetric and asymmetric models. We also compute

Table 3
Simulation results for Student versus Skew-student comparison.

| $\nu$ | $n$ | $\lambda$ | $\hat{BF} < 0.5$ | $\hat{BF} > 2$ | $25^{th}$ p. | Median | $75^{th}$ p. |
|---|---|---|---|---|---|---|---|
| | | 0 | 11.7 | 71.7 | 1.5312 | 6.0873 | 24.8469 |
| | 10 | 0.1 | 26.0 | 48.0 | 0.4909 | 1.7623 | 5.1188 |
| | | 0.3 | 42.7 | 23.2 | 0.2176 | 0.6198 | 1.7899 |
| | | 0.5 | 60.8 | 14.5 | 0.1032 | 0.3198 | 0.9775 |
| | | 0 | 1.4 | 95.6 | 29.7884 | 133.4624 | 594.1999 |
| 1 | 50 | 0.1 | 29.4 | 49.9 | 0.3438 | 1.9686 | 8.9924 |
| | | 0.3 | 79.7 | 8.4 | 0.0049 | 0.0427 | 0.3417 |
| | | 0.5 | 93.4 | 2.4 | 0.0004 | 0.0041 | 0.0370 |
| | | 0 | 0.3 | 98.7 | 148.4 | 608.8 | 2737.9 |
| | 100 | 0.1 | 41.9 | 38.2 | 0.1 | 0.8 | 6.2 |
| | | 0.3 | 94.9 | 2.8 | 0.0 | 0.0 | 0.0 |
| | | 0.5 | 99.6 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | 0 | 10.3 | 69.3 | 1.5578 | 4.3809 | 9.2094 |
| | 10 | 0.1 | 14.7 | 56.4 | 0.9371 | 2.5278 | 6.4705 |
| | | 0.3 | 29.2 | 38.7 | 0.4043 | 1.2866 | 3.3779 |
| | | 0.5 | 46.7 | 21.1 | 0.1758 | 0.5682 | 1.6956 |
| | | 0 | 2.6 | 91.2 | 7.0075 | 19.2956 | 43.3332 |
| 10 | 50 | 0.1 | 25.5 | 45.6 | 0.4844 | 1.6349 | 4.6176 |
| | | 0.3 | 41.5 | 33.0 | 0.1552 | 0.7937 | 3.2732 |
| | | 0.5 | 78.3 | 6.5 | 0.0078 | 0.0593 | 0.3789 |
| | | 0 | 2.1 | 94.7 | 15.2573 | 38.3028 | 87.9117 |
| | 100 | 0.1 | 7.5 | 81.7 | 3.4485 | 11.4648 | 32.0435 |
| | | 0.3 | 58.3 | 19.0 | 0.0294 | 0.2585 | 1.3151 |
| | | 0.5 | 95.6 | 1.1 | 0.0002 | 0.0022 | 0.0241 |
| | | 0 | 9.2 | 67.2 | 1.3782 | 3.6613 | 8.6506 |
| | 10 | 0.1 | 15.0 | 58.0 | 0.8423 | 2.6160 | 6.5973 |
| | | 0.3 | 27.4 | 36.1 | 0.4471 | 1.2501 | 3.1800 |
| | | 0.5 | 45.2 | 23.2 | 0.1925 | 0.6397 | 1.8215 |
| | | 0 | 2.3 | 92.1 | 6.7759 | 17.6408 | 37.4567 |
| 20 | 50 | 0.1 | 8.9 | 77.3 | 2.2818 | 7.9469 | 19.0925 |
| | | 0.3 | 39.3 | 36.8 | 0.1860 | 0.9700 | 3.3768 |
| | | 0.5 | 79.2 | 7.6 | 0.0079 | 0.0624 | 0.3695 |
| | | 0 | 0.8 | 95.3 | 14.8925 | 35.9162 | 76.7650 |
| | 100 | 0.1 | 6.9 | 79.6 | 2.6278 | 10.5830 | 28.1940 |
| | | 0.3 | 53.9 | 24.1 | 0.0466 | 0.3766 | 1.9179 |
| | | 0.5 | 95.0 | 1.3 | 0.0002 | 0.0027 | 0.0332 |

the Bayes factor to test skewness and present simulation results for the skew-normal and the skew-$t$ distributions obtaining expected results.

In Section 2, we present two expressions for the $L_1$ distance between posterior distributions of $\mu$ and $\sigma$. In order to determine the sensitivity of the $\lambda$ parameter concerning these posterior distributions, the numeric calculation would be most useful. Also, we showed that the $L_1$ distance between repre-

Table 4
Some descriptive statistics for the monthly rentability of five Chilean companies.

| Company | Mean | s.d. | Skewness | Kurtosis |
|---|---|---|---|---|
| Cementos | 0.01347727 | 0.13283164 | 1.33268186 | 3.31334209 |
| Cervezas | 0.02022545 | 0.11423326 | 0.51903086 | 0.66034736 |
| Chilquinta | 0.02778182 | 0.13638980 | 0.68802875 | 2.21338495 |
| Copec | 0.01201909 | 0.10339547 | 0.71021655 | 0.98810570 |
| Iansa | 0.0006200 | 0.1131153 | 0.3621623 | 0.2541416 |

Table 5
Bayes factors by company.

| | $\lambda^2 \sim Ga\left(1,1\right)$ |
|---|---|
| Company | $\hat{BF}_{\lambda>0}$ |
| Cementos | 0.6297 |
| Cervezas | 0.3033 |
| Chilquinta | 0.1936 |
| Copec | 0.5601 |
| Iansa | 1.3841 |

sentable skew elliptical distribution and representable elliptical distribution is equal to the $L_1$ distance between normal and skew normal distribution. On the other hand, in Section 3, different prior conditions were assumed obtaining different Bayes Factor expressions. Similar results could be obtained for other prior conditions, for instance, we could assume that $|\lambda| \sim Ga\left(a,b\right)$ or $\lambda \sim$ Half-Normal.

In other directions we could study the computation of default Bayes factors for testing skewness, for example by considering noninformative prior distribution $\pi\left(\lambda\right)$ in (6). A simple computation shows that if we consider an improper prior distribution $\pi\left(\lambda\right) \propto 1$, then $\int_{-\infty}^{\infty}\left[\prod_{i=1}^{n}G\left(\lambda y_i\right)\right]^b d\lambda$ diverge with $0 < b \leq 1$. Therefore, the Intrinsic Bayes factor and Fractional Bayes factor can not be calculated.

## A    Proof of Proposition 9

Since

$$\pi\left(\mu,\sigma\,|\,\lambda,\mathbf{x}\right) \propto \frac{1}{\sigma^n}\prod_{i=1}^{n}\phi\left(\frac{x_i-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x_i-\mu}{\sigma}\right)\pi\left(\mu\,|\,\sigma\right)\pi\left(\sigma\right)$$

$$\propto \frac{1}{\sigma^{n+2(a+1)}}\phi_n\left(\lambda\frac{\mathbf{x}-\mu\mathbb{1}_n}{\sigma}\right)\boldsymbol{\Phi}_n\left(\lambda\frac{\mathbf{x}-\mu\mathbb{1}_n}{\sigma}\right)$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left[v\left(\mu-m\right)^2+2b\right]\right\},$$

we have, after some simple algebraic manipulations, that

$$\pi\left(\mu,\sigma\,|\,\lambda,\mathbf{x}\right) \propto \frac{1}{\sigma^{n+2(a+1)}} \exp\left\{-\frac{r^2}{2\sigma^2}\right\} \phi\left(\sqrt{n+v}\,\frac{\mu-\hat{\mu}}{\sigma}\right) \boldsymbol{\Phi}_n\left(\lambda\frac{\mathbf{x}-\mu\mathbb{1}_n}{\sigma}\right),$$

where $\hat{\mu} = \frac{n\bar{x}+mv}{n+v}$, $r^2 = ns^2 + \frac{nv}{n+v}\left(m-\bar{x}\right)^2 + 2b$ and $s^2 = n^{-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$. The normalizing constant is

$$m\left(\mathbf{x}\right) = \int_0^\infty \frac{1}{\sigma^{n+2a+1}} \exp\left\{-\frac{r^2}{2\sigma^2}\right\} h\left(\sigma\right) d\sigma,$$

where

$$h\left(\sigma\right) = \int_{-\infty}^{\infty} \frac{1}{\sigma} \phi\left(\sqrt{n+v}\,\frac{\mu-\hat{\mu}}{\sigma}\right) \boldsymbol{\Phi}_n\left(\lambda\frac{\mathbf{x}-\mu\mathbb{1}_n}{\sigma}\right) d\mu.$$

The change of variable $y = \sqrt{n+v}\,\frac{\mu-\hat{\mu}}{\sigma}$ implies

$$
\begin{aligned}
h\left(\sigma\right) &= \frac{1}{\sqrt{n+v}} \int_{-\infty}^{\infty} \phi(y)\boldsymbol{\Phi}_n\left(\lambda\left(\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{\sigma} - \frac{y\mathbb{1}_n}{\sqrt{n+v}}\right)\right) dy \\
&= \frac{1}{\sqrt{n+v}} \int_{-\infty}^{\infty} \phi(y)\mathbb{P}\left(\mathbf{U} \le \lambda\left(\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{\sigma} - \frac{Y\mathbb{1}_n}{\sqrt{n+v}}\right)\,\bigg|\,Y = y\right) dy \\
&= \frac{1}{\sqrt{n+v}} \mathbb{P}\left(\mathbf{U} + \lambda\frac{Y\mathbb{1}_n}{\sqrt{n+v}} \le \lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{\sigma}\right)
\end{aligned}
$$

where $\mathbf{U} \sim N_n(\mathbf{0},\mathbf{I}_n)$ and $Y \sim N(0,1)$ are independent random quantities. Since $\mathbf{U} + \lambda\frac{Y\mathbb{1}_n}{\sqrt{n+v}} \sim N_n\left(\mathbf{0},\boldsymbol{\Sigma}\right)$, where $\boldsymbol{\Sigma} = \mathbf{I}_n + \frac{\lambda^2}{n+v}\mathbb{1}_n\mathbb{1}_n^t$, then

$$h\left(\sigma\right) = \frac{1}{\sqrt{n+v}} \boldsymbol{\Phi}_n\left(\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{\sigma}\,\bigg|\,\boldsymbol{\Sigma}\right),$$

and so

$$m\left(\mathbf{x}\right) = \frac{1}{\sqrt{v+n}} \int_0^\infty \frac{1}{\sigma^{n+2a+1}} \exp\left\{-\frac{r^2}{2\sigma^2}\right\} \boldsymbol{\Phi}_n\left(\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{\sigma}\,\bigg|\,\boldsymbol{\Sigma}\right) d\sigma.$$

Making now the change of variable $s = \frac{r^2}{\sigma^2}$, we obtain

21

$$m\left(\mathbf{x}\right) = \frac{1}{2r^{n+2a}\sqrt{v+n}} \int_0^\infty s^{\frac{n+2a}{2}-1} \exp\left\{-\frac{s}{2}\right\} \mathbf{\Phi}_n\left(\lambda\sqrt{s}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r} \middle| \mathbf{\Sigma}\right) ds$$

$$= \frac{2^{\frac{n+2a}{2}-1}\Gamma\left(\frac{n+2a}{2}\right)}{r^{n+2a}\sqrt{v+n}} \int_0^\infty g\left(s\middle|\frac{n+2a}{2},\frac{1}{2}\right) \mathbb{P}\left(\frac{1}{\sqrt{S}}\mathbf{U} \le \lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r} \middle| S = s\right) ds$$

$$= \frac{2^{\frac{n+2a}{2}-1}\Gamma\left(\frac{n+2a}{2}\right)}{r^{n+2a}\sqrt{v+n}} \mathbb{P}\left(\sqrt{\frac{n+2a}{S}}\mathbf{U} \le \lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)$$

$$= \frac{2^{\frac{n+2a}{2}-1}\Gamma\left(\frac{n+2a}{2}\right)}{r^{n+2a}\sqrt{v+n}} \mathbf{T}_n\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r} \middle| \mathbf{\Sigma}, n+2a\right)$$

where $g\left(s\middle|\frac{n+2a}{2},\frac{1}{2}\right)$ denotes the pdf of $S \sim Ga\left(\frac{n+2a}{2},\frac{1}{2}\right)$ and is independent of $\mathbf{U} \sim N_n\left(\mathbf{0},\mathbf{\Sigma}\right)$, which implies that $\sqrt{\frac{n+2a}{S}}\mathbf{U} \sim \mathbf{T}_n\left(\mathbf{0},\mathbf{\Sigma}, n+2a\right)$. Therefore,

$$\pi\left(\mu,\sigma\middle|\lambda,\mathbf{x}\right) = \frac{r^{n+2a}\sqrt{v+n}}{2^{\frac{n+2a}{2}-1}\Gamma\left(\frac{n+2a}{2}\right)} \frac{1}{\sigma^{n+2(a+1)}} \exp\left\{-\frac{r^2}{2\sigma^2}\right\} \phi\left(\sqrt{n+v}\frac{\mu-\hat{\mu}}{\sigma}\right)$$

$$\times \frac{\mathbf{\Phi}_n\left(\lambda\frac{\mathbf{x}-\mu\mathbb{1}_n}{\sigma}\right)}{\mathbf{T}_n\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r} \middle| \mathbf{\Sigma}, n+2a\right)}.$$

But, for $\lambda = 0$ (the normal model), it is well know that

$$\pi\left(\mu,\sigma\middle|\lambda = 0,\mathbf{x}\right) = \frac{r^{n+2a}\sqrt{v+n}}{2^{\frac{n+2a}{2}-1}\Gamma\left(\frac{n+2a}{2}\right)} \frac{1}{\sigma^{n+2(a+1)}} \exp\left\{-\frac{r^2}{2\sigma^2}\right\} \phi\left(\sqrt{n+v}\frac{\mu-\hat{\mu}}{\sigma}\right),$$

which concludes the proof.

## B    Proof of Proposition 11

After some algebraic manipulations, the likelihood of $M_1$ model is

$$f\left(\mathbf{x}\middle|\lambda,\mu,\sigma,\omega_1,\ldots,\omega_n\right) = \prod_{i=1}^n \frac{2}{\sigma\sqrt{\omega_i}}\phi\left(\frac{x_i-\mu}{\sigma\sqrt{\omega_i}}\right) \Phi\left(\lambda\frac{x_i-\mu}{\sigma\sqrt{\omega_i}}\right)$$

$$= \left(\frac{2}{\pi\sigma^2}\right)^{\frac{n}{2}} |\mathbf{D}\left(\boldsymbol{\omega}\right)|^{-\frac{1}{2}} \exp\left(-\frac{\eta}{2\sigma^2}S_\omega^2\right)$$

$$\exp\left[-\frac{\eta}{2\sigma^2}\left(\mu-\sum_{i=1}^n \nu_i x_i\right)^2\right] \mathbf{\Phi}_n\left(\frac{\lambda}{\sigma}\left[\mathbf{D}\left(\boldsymbol{\omega}\right)\right]^{-\frac{1}{2}}\left(\mathbf{x}-\mu\mathbb{1}_n\right)\right),$$

where $\mathbf{D}\left(\boldsymbol{\omega}\right) = \text{diag}\left(\omega_1,\ldots,\omega_n\right), \eta = \sum_{i=1}^n \omega_i^{-1}, S_\omega^2 = \sum_{i=1}^n \nu_i x_i^2 - \left(\sum_{i=1}^n \nu_i x_i\right)^2 = \sum_{i=1}^n \nu_i\left(x_i - \sum_{j=1}^n \nu_j x_j\right)^2$ and $\nu_i = \frac{\omega_i}{\eta}$ for each $i = 1,\ldots,n$.

Now, considering $\mu\,|\sigma \sim N\left(m, \frac{\sigma^2}{v}\right)$, we obtain

$$
\begin{aligned}
f\left(\mathbf{x}\,|\,\lambda, \sigma, \boldsymbol{\omega}\right) &= \int_{-\infty}^{\infty} f\left(\mathbf{x}\,|\,\lambda, \mu, \sigma, \boldsymbol{\omega}\right) \phi\left(\mu - m\,\middle|\,\frac{\sigma^2}{v}\right) d\mu \\
&= \sqrt{\frac{2^n v}{(\pi \sigma^2)^n}}\,\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}} h\left(\lambda, \sigma, \boldsymbol{\omega}\right) \\
&\quad \exp\left\{-\frac{\eta}{2\sigma^2}\left[S_\omega^2 + \frac{v}{\eta+v}\left(m - \sum_{i=1}^{n} \nu_i x_i\right)^2\right]\right\}
\end{aligned}
\tag{B.1}
$$

where

$$
\hat{\mu} = \frac{\eta \sum_{i=1}^{n} \nu_i x_i + vm}{\eta + v}
$$

and

$$
h\left(\lambda, \sigma, \boldsymbol{\omega}\right) = \int_{-\infty}^{\infty} \frac{1}{\sigma}\phi\left(\sqrt{\eta + v}\,\frac{(\mu - \hat{\mu})}{\sigma}\right) \boldsymbol{\Phi}_n\left(\frac{\lambda}{\sigma}\left[\mathbf{D}\left(\boldsymbol{\omega}\right)\right]^{-\frac{1}{2}}\left(\mathbf{x} - \mu \mathbb{1}_n\right)\right) d\mu.
$$

Making the change of variable $y = \sqrt{\eta + v}\,\frac{(\mu - \hat{\mu})}{\sigma}$, we obtain

$$
\begin{aligned}
h\left(\lambda, \sigma, \boldsymbol{\omega}\right) &= \frac{1}{\sqrt{\eta + v}} \int_{-\infty}^{\infty} \boldsymbol{\Phi}_n\left(\lambda\left[\mathbf{D}\left(\boldsymbol{\omega}\right)\right]^{-\frac{1}{2}}\left(\frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{\sigma} - \frac{y\mathbb{1}_n}{\sqrt{\eta+v}}\right)\right) \phi\left(y\right) dy \\
&= \frac{1}{\sqrt{\eta + v}} \int_{-\infty}^{\infty} \mathbb{P}\left(\mathbf{U} \leq \lambda\left[\mathbf{D}\left(\boldsymbol{\omega}\right)\right]^{-\frac{1}{2}}\left(\frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{\sigma} - \frac{Y\mathbb{1}_n}{\sqrt{\eta+v}}\right)\,\middle|\,Y = y\right) \phi\left(y\right) dy \\
&= \frac{1}{\sqrt{\eta + v}} \mathbb{P}\left(\frac{1}{\lambda}\left[\mathbf{D}\left(\boldsymbol{\omega}\right)\right]^{\frac{1}{2}}\mathbf{U} + \frac{Y\mathbb{1}_n}{\sqrt{\eta+v}} \leq \frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{\sigma}\right),
\end{aligned}
$$

where $\mathbf{U} \sim N_n\left(\mathbf{0}, \mathbf{I}_n\right)$ and $Y \sim N\left(0, 1\right)$ are independent random quantities. Since $\frac{1}{\lambda}\left[\mathbf{D}\left(\boldsymbol{\omega}\right)\right]^{\frac{1}{2}}\mathbf{U} + \frac{Y\mathbb{1}_n}{\sqrt{\eta+v}} \sim N_n\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$, where $\boldsymbol{\Sigma} = \lambda^{-2}\mathbf{D}\left(\boldsymbol{\omega}\right) + \frac{1}{v+\eta}\mathbb{1}_n \mathbb{1}_n^t$, then

$$
h\left(\lambda, \sigma, \boldsymbol{\omega}\right) = \frac{1}{\sqrt{\eta + v}} \boldsymbol{\Phi}_n\left(\frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{\sigma}\,\middle|\,\mathbf{0}, \boldsymbol{\Sigma}\right).
$$

Replacing $h\left(\lambda, \sigma, \boldsymbol{\omega}\right)$ in (B.1) we obtain

$$
\begin{aligned}
f\left(\mathbf{x}\,|\,\lambda, \sigma, \boldsymbol{\omega}\right) &= \sqrt{\frac{2^n v\,\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-1}}{(\eta + v)\,(\pi \sigma^2)^n}}\,\boldsymbol{\Phi}_n\left(\frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{\sigma}\,\middle|\,\mathbf{0}, \boldsymbol{\Sigma}\right) \\
&\quad \exp\left\{-\frac{\eta}{2\sigma^2}\left[S_\omega^2 + \frac{v}{\eta+v}\left(m - \sum_{i=1}^{n} \nu_i x_i\right)^2\right]\right\}.
\end{aligned}
$$

Then, since $\sigma^{-2} \sim Ga\left(a, b\right)$, one obtains

$$f\left(\mathbf{x}\mid \lambda, \boldsymbol{\omega}\right) = \frac{2b^a}{\Gamma\left(a\right)} \sqrt{\frac{2^n v \left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-1}}{\pi^n \left(v + \eta\right)}}$$

$$\int_0^\infty \sigma^{-n-1-2a} \exp\left(-\frac{r^2}{2\sigma^2}\right) \boldsymbol{\Phi}_n\left(\left.\frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{\sigma}\right| \mathbf{0}, \boldsymbol{\Sigma}\right) d\sigma,$$

where $r^2 = \eta S_\omega^2 + \frac{\eta v}{\eta + v}\left(m - \sum_{i=1}^n \nu_i x_i\right)^2 + 2b$.

A change of variable $s = r^2/\sigma^2$, allows us to write

$$f\left(\mathbf{x}\mid \lambda, \boldsymbol{\omega}\right) = \frac{b^a}{r^{n+2a}\Gamma\left(a\right)} \sqrt{\frac{2^n v \left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-1}}{\pi^n \left(v + \eta\right)}}$$

$$\int_0^\infty s^{\frac{n}{2}+a-1} \exp\left(-\frac{s}{2}\right) \boldsymbol{\Phi}_n\left(\left.\frac{\sqrt{s}\left(\mathbf{x} - \hat{\mu}\mathbb{1}_n\right)}{r}\right| \mathbf{0}, \boldsymbol{\Sigma}\right) ds$$

$$= \frac{2^{n+a}b^a\Gamma\left(\frac{n}{2}+a\right)}{r^{n+2a}\Gamma\left(a\right)} \sqrt{\frac{v \left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-1}}{\pi^n \left(v + \eta\right)}}$$

$$\int_0^\infty \mathbb{P}\left(\left.\frac{1}{\sqrt{S}}\mathbf{U} \leq \left.\frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{r}\right| S = s\right.\right) g\left(s \left| \frac{n}{2} + a, \frac{1}{2}\right.\right) ds$$

$$= \frac{2^{n+a}b^a\Gamma\left(\frac{n}{2}+a\right)}{r^{n+2a}\Gamma\left(a\right)} \sqrt{\frac{v \left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-1}}{\pi^n \left(v + \eta\right)}}$$

$$\mathbb{P}\left(\left.\sqrt{\frac{n+2a}{S}}\mathbf{U} \leq \sqrt{n+2a}\frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{r}\right| S = s\right)$$

$$= \frac{2^{n+a}b^a\Gamma\left(\frac{n}{2}+a\right)}{r^{n+2a}\Gamma\left(a\right)} \sqrt{\frac{v \left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-1}}{\pi^n \left(v + \eta\right)}}$$

$$\mathbf{T}_n\left(\left.\sqrt{n+2a}\frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{r}\right| \mathbf{0}, \boldsymbol{\Sigma}, n + 2a\right),$$

where $g\left(s \left| \frac{n}{2} + a, \frac{1}{2}\right.\right)$ denotes the pdf of $S \sim Ga\left(\frac{n}{2} + a, \frac{1}{2}\right)$ and is independent of $\mathbf{U} \sim N_n\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$, which implies that $\sqrt{\frac{n+2a}{S}}\mathbf{U} \sim \mathbf{T}_n\left(\mathbf{0}, \boldsymbol{\Sigma}, n + 2a\right)$.

In a similar way, for the $M_0$ model we obtain

$$f\left(\mathbf{x}\mid \boldsymbol{\omega}\right) = \frac{2^{\frac{n}{2}+a}b^a\Gamma\left(\frac{n}{2}+a\right)}{r^{n+2a}\Gamma\left(a\right)} \sqrt{\frac{v \left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-1}}{\pi^n \left(v + \eta\right)}}.$$

Therefore, the Bayes factor is

$$BF\left(\mathbf{x}\right) = \frac{\int \cdots \int f\left(\mathbf{x} | \boldsymbol{\omega}\right) dH\left(\omega_1\right) \cdots dH\left(\omega_n\right)}{\int \left[\int \cdots \int f\left(\mathbf{x} | \lambda, \boldsymbol{\omega}\right) dH\left(\omega_1\right) \cdots dH\left(\omega_n\right)\right] \pi\left(\lambda\right) d\lambda}$$

$$= \frac{\int \cdots \int r^{-n-2a} \left[(\eta + v) \prod_{i=1}^{n} \omega_i\right]^{-\frac{1}{2}} dH\left(\omega_1\right) \cdots dH\left(\omega_n\right)}{2^{\frac{n}{2}} \int \cdots \int r^{-n-2a} \left[(\eta + v) \prod_{i=1}^{n} \omega_i\right]^{-\frac{1}{2}} g\left(\boldsymbol{\omega}\right) dH\left(\omega_1\right) \cdots dH\left(\omega_n\right)},$$

where

$$g\left(\boldsymbol{\omega}\right) = \int \mathbf{T}_n \left( \sqrt{n + 2a} \frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{r} \middle| \mathbf{0}, \boldsymbol{\Sigma}, n + 2a \right) \pi\left(\lambda\right) d\lambda.$$

# References

Arellano-Valle, R. B., 1994. Distribuicões elípticas: Propiedades e aplicacões a modelos de regressão. Ph.D. thesis, IME-USP, São Paulo, Brasil, in Portugues.

Arellano-Valle, R. B., Galea-Rojas, M., Iglesias, P., 2000. Bayesian sensitivity analysis in elliptical linear regression models. Journal of Statistical Planning and Inference 86, 175–199.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12, 171–178.

Berger, J. O., Pericchi, L. R., 2001. Objective Bayesian methods for model selection: Introduction and comparison. In: Lahiri, P. (Ed.), Model Selection. Vol. 38. Institute of Mathematical Statistics, Lecture Notes Monograph Series, with discussion.

Branco, M. D., Dey, D. K., 2001. A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis 79, 99–113.

Cambanis, S., Huang, S., Simons, G., 1981. On the theory of elliptical contoured distributions. Journal of Multivariate Analysis 11, 368–385.

Dongarra, J. J., Grosse, E., 1987. Distribution of mathematical software via electronic mail. Communications of the ACM 30, 403–407.

Fang, K. T., Kotz, S., Ng, K. W., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall, London/New York.

Genton, M. G., Loperfido, N. M. R., 2001. Generalized skew-elliptical distributions and their quadratic forms. Tech. rep., North Carolina State University.

Kass, R. E., Raftery, A. E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.

Kass, R. E., Tierney, L., Kadane, J. B., 1989. Approximate methods for assessing influence and sensitivity in Bayesian analysis. Biometrika 76, 663–674.

Kelker, D., 1970. Distribution theory of spherical distributions and a location-scale parameter generalization. Sankhyã A 32, 419–430.

Lavine, M., Schervish, M., 1999. Bayes factors: what they are and what they are not. The American Statistician 53, 119–122.

Liseo, B., Loperfido, N., 2002. Default Bayesian analysis of the skew-normal distribution. Tech. rep., Università di Roma "La Sapienza".

Peng, F., Dey, D. K., 1995. Bayesian analysis of outlier problems using divergences measures. The Canadian Journal of Statistics 23, 199–213.

Piessens, R., deDoncker Kapenga, E., Uberhuber, C., Kahaner, D., 1983. QUADPACK: a Subroutine Package for Automatic Integration. Springer, Berlin.

Weiss, R. E., 1996. An approach to Bayesian sensitivity analysis. Journal of the Royal Statistical Society B 58, 739–750.