Bayesian Analysis (2008) **3**, Number 3, pp. 513–540

Bayesian Inference for Shape Mixtures of Skewed Distributions, with Application to Regression Analysis

Reinaldo B. Arellano-Valle*, Luis M. Castro†, Marc G. Genton‡ and Héctor W. Gómez§

Abstract. We introduce a class of shape mixtures of skewed distributions and study some of its main properties. We discuss a Bayesian interpretation and some invariance results of the proposed class. We develop a Bayesian analysis of the skew-normal, skew-generalized-normal, skew-normal-t and skew-t-normal linear regression models under some special prior specifications for the model parameters. In particular, we show that the full posterior of the skew-normal regression model parameters is proper under an arbitrary proper prior for the shape parameter and noninformative prior for the other parameters. We implement a convenient hierarchical representation in order to obtain the corresponding posterior analysis. We illustrate our approach with an application to a real dataset on characteristics of Australian male athletes.

Keywords: Posterior analysis, regression model, shape parameter, skewness, skew-normal distribution, symmetry.

1 Introduction

The construction of flexible parametric non-Gaussian multivariate distributions has seen a growing interest in recent years because distributions of many datasets exhibit skewness as well as tails that are lighter or heavier than those of the normal distribution. Several proposals have been put forward in the literature, an overview of which can be found in the book edited by Genton (2004), in Azzalini (2005), in Arellano-Valle and Azzalini (2006), and from a unified point of view in Arellano-Valle, Branco and Genton (2006).

A fairly large class of such distributions introduced by Wang, Boyer and Genton (2004) consists of skew-symmetric (SS) distributions with probability density function (pdf) of the form

$$2f(\mathbf{z})Q(\mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^n, \tag{1}$$

*Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile, mailto:reivalle@mat.puc.cl

†Departamento de Estadística, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Concepción, Chile, mailto:luiscastroc@udec.cl

‡Department of Econometrics, University of Geneva, Geneva, Switzerland and Department of Statistics, Texas A&M University, College Station, TX, mailto:Marc.Genton@metri.unige.edu

§Departamento de Estadística, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Concepción, Chile, mailto:hgomez@udec.cl

 DOI:10.1214/08-BA320

where $f : \mathbb{R}^n \to \mathbb{R}_+$ is a continuous pdf, symmetric around zero, i.e., $f(-\mathbf{z}) = f(\mathbf{z})$ for all $\mathbf{z} \in \mathbb{R}^n$, and $Q : \mathbb{R}^n \to [0, 1]$ is a skewing function satisfying $Q(-\mathbf{z}) + Q(\mathbf{z}) = 1$ for all $\mathbf{z} \in \mathbb{R}^n$. A random vector $\mathbf{Z}$ with pdf (1) is denoted by $\mathbf{Z} \sim SS_n(f, Q)$. When f is the pdf of an elliptically contoured distribution, the family (1) defines generalized skew-elliptical distributions studied by Genton and Loperfido (2005). As noted by Ma and Genton (2004) and Azzalini and Capitanio (2003), any continuous skewing function Q can be written as $Q(\mathbf{z}) = G(w(\mathbf{z}))$, where $G : \mathbb{R} \to [0, 1]$ is the cumulative distribution function (cdf) of a continuous random variable symmetric around zero, and $w : \mathbb{R}^n \to \mathbb{R}$ is an odd continuous function, i.e., $w(-\mathbf{z}) = -w(\mathbf{z})$ for all $\mathbf{z} \in \mathbb{R}^n$. A popular choice in the literature is

$$Q(\mathbf{z}) = G(\boldsymbol{\alpha}^T\mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^n, \tag{2}$$

where the shape vector $\boldsymbol{\alpha} \in \mathbb{R}^n$ controls skewness and $\boldsymbol{\alpha} = \mathbf{0}$ corresponds to a symmetric pdf in (1). A random vector $\mathbf{Z}$ with pdf (1) and skewing function (2) is denoted by $\mathbf{Z} \sim SS_n(f, G, \boldsymbol{\alpha})$. In particular, when $f(\mathbf{z}) = \phi_n(\mathbf{z}|\mathbf{0}, \mathbf{I}_n)$, the pdf of the n-dimensional multivariate normal distribution $N_n(\mathbf{0}, \mathbf{I}_n)$ with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_n$, the identity, and $G = \Phi$, the standard normal cdf, the resulting pdf (1) is from the standard multivariate skew-normal distribution $SN_n(\mathbf{0}, \mathbf{I}_n, \boldsymbol{\alpha}) = SN_n(\boldsymbol{\alpha})$ defined by Azzalini and Dalla Valle (1996). For $n = 1$, it reduces to the univariate standard skew-normal distribution of Azzalini (1985). Note that location and scales can be introduced by means of $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}$ throughout, where $\boldsymbol{\Sigma}^{1/2}$ is the symmetric square root of $\boldsymbol{\Sigma}$, thus yielding for example $SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$ and $SS_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f, G, \boldsymbol{\alpha})$ distributions. The use of odd polynomials for the function w has been proposed by Ma and Genton (2004) and leads to flexible skew-symmetric distributions that, in addition, can exhibit multimodality.

In this paper, we consider alternative choices to linear or odd polynomials for the skewing function Q. For illustration, consider the univariate case of (1) defined by setting $n = 1$. An interesting generalization of (2) results from the choice

$$w(z) = \frac{\alpha_1 z}{\sqrt{1 + \alpha_2^2 z^2}}, \qquad z \in \mathbb{R}, \tag{3}$$

where $\alpha_1 \in \mathbb{R}$ and $\alpha_2 \geq 0$. The particular case of $f = \phi$, the standard normal pdf, and $G = \Phi$, has been studied by Arellano-Valle, Gómez and Quintana (2004), leading to a so-called skew-generalized-normal distribution, denoted by $SGN(\alpha_1, \alpha_2)$. They have shown that this distribution can be represented as a shape mixture of the skew-normal distribution, where the mixing distribution is normal. Specifically, if $Z \sim SGN(\alpha_1, \alpha_2)$, then there is a shape random variable S such that

$$[Z|S = s] \sim SN(s) \quad \text{and} \quad S \sim N(\alpha_1, \alpha_2). \tag{4}$$

In other words, this representation allows to identify the function $w(z)$ within the class defined by (1) when $f = \phi$ and $G = \Phi$. Another important example arises when we consider both scale and shape mixtures of the skew-normal distribution. For example, consider the class defined by

$$[Z|S_1 = s_1, S_2 = s_2] \sim SN(0, s_2^{-1}, s_1^{-1/2}\alpha_1), \tag{5}$$

where S_1 and S_2 are non-negative and independent random variables. Any model of the form $2f(z)G(\alpha_1 z)$, for which both f and G are scale mixtures of the normal distribution, belong to the class defined by (5). In particular, the skew-t distribution in the form introduced by Azzalini and Capitanio (2003) with pdf $2t(z;\nu)T(\sqrt{\nu+1}\,\alpha_1 z/\sqrt{\nu+z^2};\nu+1)$, where $t(z;\nu)$ and $T(z;\nu)$ are the Student-t pdf and cdf, respectively, is obtained with $S_2 \sim Gamma(\nu/2,\nu/2)$ and $S_1 \equiv 1$; see also Azzalini and Genton (2008) for additional properties. Further examples that arise from (5) are the so-called skew-normal-t and skew-t-normal distributions, with pdf's of the form $2\phi(z)T(\alpha_1 z)$ and $2t(z;\nu)\Phi(\alpha_1 z)$, and denoted by $SNT(\alpha_1,\nu)$ and $STN(\alpha_1,\nu)$, respectively. These distributions were studied by Nadarajah and Kotz (2003) and Gómez, Venegas and Bolfarine (2007), and can be obtained from (5) with $S_1 \sim Gamma(\nu/2,\nu/2)$ and $S_2 \equiv 1$, and $S_2 = S_1 \sim Gamma(\nu/2,\nu/2)$, respectively.

The representations in (4) and (5) are particularly important in a Bayesian framework, since they provide hierarchical formulations of the respective location-scale versions of (1). Moreover, they can be interpreted as Bayesian specifications of the SN model with the prior considerations for the shape/scale parameter indicated above. For instance, (4) can be applied in the Bayesian specification of the SN model for a random sample $Y_1,\ldots,Y_n$ of $Y = \mu + \sigma Z$, where $[Z|S=s] \sim SN(s)$, with a normal prior for the shape parameter S. Hence, only a prior for the location-scale parameters (μ,σ^2) needs to be elicited to complete the model specification.

This article is concerned with the subclass of skewed distributions that can be represented as shape mixtures of the family defined by (1) and (2). In particular, the representation (4) is extended to all the classes of distributions with pdf given by (1) with (3), and extensions to the multivariate setting. The rest of the article is organized as follows. For simplicity of exposition, we begin in Section 2 by considering the univariate case, where consequences from the use of symmetric location-scale mixing distributions are discussed. Next, in Section 3, the idea of shape mixtures in order to identify skewing functions for the multivariate skew-symmetric class (1) is considered. The procedure is illustrated with the multivariate SN distribution. In Sections 4 and 5, a Bayesian posterior analysis for the skew-normal, skew-generalized-normal, skew-normal-t, and skew-t-normal linear regression models is developed under some special prior specifications. In particular, we show that the full posterior of the skew-normal regression model parameters is proper under an arbitrary proper prior for the shape parameter and noninformative prior for the other parameters. An application to a dataset of Australian male athletes is presented Section 6.

2 Shape Mixtures of Univariate Skewed Distributions

2.1 Definition and properties

The shape mixtures of univariate skewed distributions form an important subclass of the skew-symmetric family of distributions defined by (1) with $n = 1$, because it allows for the specification of the skewing function Q starting from the much simpler class based on (2). This subclass is obtained as a mixture of the skewed distribution defined by (1) and (2) on the shape parameter α. Consequently, the resulting subclass contains distributions that are more flexible than the original ones.

Definition 1. The distribution of a random variable Y is a shape mixture of skew-symmetric ($SMSS$) distributions if there exists a random variable S such that the conditional distribution $[Y|S=s] \sim SS(\mu, \sigma^2, f, G, s)$ for some symmetric pdf f (which can depend on s) and cdf G (which can depend on $z = (y-\mu)/\sigma$ and/or on s).

When $f = \phi$, the distribution of the random variable Y is a shape mixture of skew-normal ($SMSN$) distributions. It follows from Definition 1 that the conditional pdf of Y given $S = s$ is of the form (1) with (2). The following result yields the unconditional distribution of Y.

Proposition 1. Let $[Y|S=s] \sim SS(\mu, \sigma^2, f, G, s)$, $s \in \mathbb{R}$, where the pdf f does not depend on s and S has cdf H. Then, $Y \sim SS(\mu, \sigma^2, f, Q)$, i.e., its pdf is of the skew-symmetric form (1), where the skewing function Q is given by

$$Q(z) = E[G(zS)] = \int_{-\infty}^{\infty} G(zs)dH(s). \tag{6}$$

Moreover, if H is absolutely continuous with density $h = H'$, then the conditional pdf of S given $Y = y$ depends on (y, μ, σ) through $z = (y-\mu)/\sigma$ only and is given by

$$f(s|z) = \frac{h(s)G(zs)}{\int_{-\infty}^{\infty} h(s)G(zs)ds}. \tag{7}$$

Proof. The proof is immediate from Definition 1 and Bayes' theorem. □

Note that any random variable $Y \sim SS(\mu, \sigma^2, f, Q)$ can be represented as $Y \stackrel{d}{=} \mu + \sigma Z$, where $Z \sim SS(0, 1, f, Q)$ and $\stackrel{d}{=}$ denotes equality in distribution, and thus its properties can be analyzed under this "standardized" version. The mixing cdf H can be chosen arbitrarily. Therefore, a subfamily of particular importance is obtained with a discrete distribution H. It yields finite shape mixtures with skewing functions of the form

$$Q(z) = \sum_{k=1}^{K} \omega_k G(\alpha_k z), \tag{8}$$

where $\omega_k \geq 0$, for all $k = 1, \ldots, K$, with $\sum_{k=1}^{K} \omega_k = 1$, and $\alpha_k \in \mathbb{R}$. Skewing functions of the form (8) can be used to obtain approximations of those of the form (6) with H continuous that cannot be computed explicitly.

2.2 Symmetric location-scale mixing distributions

Another interesting subfamily of shape mixtures of skewed distributions is obtained when the skewing function (6) is computed with a symmetric location-scale mixing cdf H.

To characterize this subfamily, consider first the following general situation. Let (X_0, Z_0, S) be a random vector such that conditionally on S, the random variable $W = X_0 - SZ_0$ has a symmetric distribution, implying that for each value s of S

$$\begin{aligned} P(W \le 0|S = s) &= \int_{-\infty}^{\infty} P(X_0 \le sz|Z_0 = z, S = s) f_{Z_0|S=s}(z)dz \\ &= \int_{-\infty}^{\infty} F_{X_0|Z_0=z,S=s}(sz) f_{Z_0|S=s}(z)dz \\ &= 1/2, \end{aligned}$$

and therefore the function

$$f(z|s) = 2F_{X_0|Z_0=z,S=s}(sz) f_{Z_0|S=s}(z), \quad z \in \mathbb{R}, \tag{9}$$

defines a pdf for any value of s. Note under this general setting that the conditional cdf $F_{X_0|Z_0=z,S=s} = G_{(z,s)}$, say, and the conditional pdf $f_{Z_0|S=s} = f_{(s)}$, say, are not necessarily symmetric and can depend on (z, s) and s, respectively. Nevertheless, if they are symmetric, i.e., $f_{(s)}(-y) = f_{(s)}(y)$ and $G_{(z,s)}(-y) = 1 - G_{(z,s)}(y)$, for all y and each value of (z, s), and $G_{(-z,-s)} = G_{(-z,s)} = G_{(z,-s)} = G_{(z,s)}$ for all (z, s), then (9) becomes the pdf of an $SS(f_{(s)}, G_{(z,s)}, s)$ distribution. Note also that

$$G_{(z,s)}(sz) = F_{X_0|Z_0=z,S=s}(sz) = F_{W|Z_0=z,S=s}(0).$$

Consider now the random variable defined by $Z \stackrel{d}{=} [Z_0 \mid W \le 0]$, and note that $f(z|s)$ in (9) is simply the conditional pdf of Z given $S = s$, i.e., the pdf of $Z_s \stackrel{d}{=} [Z \mid S = s] = [Z_0 \mid W \le 0, S = s]$. Hence, the pdf of $Z \stackrel{d}{=} [Z_0 \mid W \le 0]$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f(z|s)dF_S(s) = \int_{-\infty}^{\infty} 2F_{X_0|Z_0=z,S=s}(sz) f_{Z_0|S=s}(z)dF_S(s),$$

so that $Z_s \stackrel{d}{=} [Z \mid S = s] \sim SS(f_{(s)}, G_{(z,s)}, s)$ when $[Z_0 \mid S = s] \sim f_{(s)}$ and $[X_0|Z_0 = z, S = s] \sim G_{(z,s)}$ are symmetric. If, in addition, we assume that Z_0 and S are independent, then $f_{(s)} = f_{Z_0|S=s} = f_{Z_0} = f$, say, which does not depend on s. Therefore, it follows that the pdf of Z reduces to

$$f_Z(z) = 2f(z) \int_{-\infty}^{\infty} G_{(z,s)}(sz)dF_S(s) = 2f(z)F_{W|Z_0=z}(0).$$

That is, if Z_0 and S are assumed to be independent, then the symmetry assumption on $Z_0 \sim f$ and $[X_0|Z_0 = z, S = s] \sim G_{(z,s)}$ for all (z, s), implies that $[Z_0|S = s] \sim SS(f, G_{(z,s)})$ and also that $Z \sim SS(f, Q)$, with

$$Q(z) = F_{W|Z_0=z}(0) = \int_{-\infty}^{\infty} G_{(z,s)}(sz)dH(s).$$

In addition, this symmetry assumption implies that conditionally on $S = s$, the random variable $W = X_0 - SZ_0$ is also symmetric whatever the distribution of S. Thus, we have the following results.

Proposition 2. Let (X_0, Z_0, S) be a random vector such that $Z_0 \sim f$ is independent of $S \sim H$, where f is a symmetric pdf around zero. Let $G_{(z,s)} = F_{X_0|Z_0=z,S=s}$ be the conditional cdf of X_0 given $Z_0 = z$ and $S = s$. Suppose that $G_{(z,s)}(-y) = 1 - G_{(z,s)}(y)$ and $G_{(-z,-s)}(y) = G_{(-z,s)}(y) = G_{(z,-s)}(y) = G_{(z,s)}(y)$ for all y and each value of (z, s). Let also Z be a random variable such that $Z \stackrel{d}{=} [Z_0|W \leq 0]$, where $W = X_0 - SZ_0$. Then $[Z|S = s] \sim SS(f, G_{(z,s)}, s)$ and therefore $Z \sim SS(f, Q)$, with $Q(z) = F_{W|Z_0=z}(0) = \int_{-\infty}^{\infty} G_{(z,s)}(sz)dH(s)$.

Corollary 1. Under the conditions of Proposition 2, we have the following byproducts:

(i) If X_0 is independent of Z_0, then $[Z|S = s] \sim SS(f, G_{(s)}, s)$ and so $Z \sim SS(f, Q)$, with $G_{(s)} = F_{X_0|S=s}$ and $Q(z) = F_{W|Z_0=z}(0) = \int_{-\infty}^{\infty} G_{(s)}(sz)dH(s)$.

(ii) If $X_0 \sim G$, $Z_0 \sim f$ and $S \sim H$ are independent random variables, with $f(-x) = f(x)$ and $G(-x) = 1 - G(x)$ for all x, then $[Z|S = s] \sim SS(f, G, s)$ and $Z \sim SS(f, Q)$, with $Q(z) = F_{W|Z_0=z}(0) = \int_{-\infty}^{\infty} G(sz)dH(s)$.

Under the assumptions of Proposition 2, we have that both X_0 and Z_0 are symmetric (around zero) random variables, and that Z_0 and S are independent. This implies that the random variable $W = X_0 - SZ_0$ is also symmetric whatever the distribution of S (and the same holds conditionally on S). In particular, if S is also assumed to be symmetric (around zero), then conditionally on Z_0, the random variable $W = X_0 - SZ_0$ will also be symmetric, i.e., $\int_{-\infty}^{\infty} G_{(z,s)}(sz)dH(s) = F_{W|Z_0=z}(0) = 1/2$ for all z, and so $Q(z) = 1/2$, for all z. This yields the following corollary.

Corollary 2. Let $Z \stackrel{d}{=} [Z_0|X_0 - SZ_0 \leq 0]$, where X_0, Z_0 and S are symmetric (around zero) random variables, with Z_0 and S independent. Then, Z has the same symmetric distribution as Z_0, i.e., $[Z_0|X_0 - SZ_0 \leq 0] \stackrel{d}{=} Z_0 \sim f$.

2.3 SMSS based on symmetric location-scale mixing distributions

In this section, we characterize the $SMSS$ subfamily obtained when the skewing function (6) is computed by means of a symmetric location-scale mixing cdf H, i.e., by taking

$$H(s) = H_0\left((s - \eta)/\tau\right), \quad \text{with} \quad \eta \in \mathbb{R} \quad \text{and} \quad \tau > 0,$$

where H_0 is a standardized symmetric (around zero) cdf, i.e., the cdf of $S_0 = (S - \eta)/\tau$. In such a case, the skewing function (6) can be rewritten as

$$Q(z) = \mathrm{E}[G(\tau z S_0 + z\eta)] = \int_{-\infty}^{\infty} G(z\{\tau s_0 + \eta\})dH_0(s_0). \tag{10}$$

Two important properties of this skewing function are:

(a) If $\tau = 0$ and G does not depend on s_0, then $Q(z) = G(\eta z)$ and (1) with (2) and $\alpha = \eta$ follows.

(b) If $\eta = 0$, then $Q(z) = \mathrm{E}[G(z\tau S_0)] = 1/2$, for any values of z and τ, which is a consequence of the symmetry (around zero) of G and H_0, see Corollary 2.

On the other hand, the results in Proposition 2 provide a more general and convenient expression to obtain the skewing function in (10) as indicated next. In fact, let

$$W_0 = \frac{W + \eta Z_0}{\sqrt{1+\tau^2 Z_0^2}} = \frac{X_0 - \tau S_0 Z_0}{\sqrt{1+\tau^2 Z_0^2}},$$

which is a standardized symmetric version of $W = X_0 - SZ_0$ where $S = \eta + \tau S_0$ is a symmetric location-scale random variable. Note that $W = \sqrt{1+\tau^2 Z_0^2}\, W_0 - \eta Z_0$, and so $Z \stackrel{d}{=} [Z_0 | W_0 \le \eta Z_0/\sqrt{1+\tau^2 Z_0^2}]$, whose pdf is given by (see, e.g., Arellano-Valle, del Pino and San Martín, 2002)

$$f_Z(z) = f(z) \frac{P\Big(W_0 \le \frac{\eta Z_0}{\sqrt{1+\tau^2 Z_0^2}} \Big| Z_0 = z\Big)}{P\left(W_0 \le \frac{\eta Z_0}{\sqrt{1+\tau^2 Z_0^2}}\right)},$$

where $P\big(W_0 \le \frac{\eta Z_0}{\sqrt{1+\tau^2 Z_0^2}}\big) = P(W \le 0) = 1/2$. This result is stated in the following proposition.

Proposition 3. Let (X_0, Z_0, S) be a random vector satisfying the same conditions as in Proposition 2. Suppose that $S = \eta + \tau S_0$, where $\eta \in \mathbb{R}$ and $\tau > 0$ are location and scale parameters, respectively, and S_0 is a standardized symmetric random variable. Then, the random variable $Z \stackrel{d}{=} [Z_0 | X_0 - SZ_0 \le 0]$ has an $SS(f, Q)$ distribution with $Q(z) = F_{W_0|Z_0=z}\left(\eta z/\sqrt{1+\tau^2 z^2}\right)$, i.e., with pdf given by

$$f_Z(z) = 2f(z) F_{W_0|Z_0=z}\left(\frac{\eta z}{\sqrt{1+\tau^2 z^2}}\right), \tag{11}$$

where $W_0 = \frac{X_0 - \tau S_0 Z_0}{\sqrt{1+\tau^2 Z_0^2}}$ is a standardized symmetric random variable.

Note that if W_0 is independent of Z_0, then $F_{W_0|Z_0=z} = F_{W_0} = G_0$, say, and so (11) reduces to (1) with (3).

We discuss next two interesting consequences of the assumption that the mixing cdf H is symmetric around zero, i.e., $H(-s) = 1 - H(s)$, for all s. The first one is related to the property (b) above (see also Corollary 2) and establishes that the marginal distribution of Z_0 is unaffected by the choice of a symmetric cdf for the mixing random variable S. The second consequence is that the conditional distribution of the mixing random variable S given $Z_0 = z$ belongs also to the class of the skew-symmetric distributions when H is absolutely continuous, i.e., when H has a pdf $h = H'$. Both results are very relevant from a Bayesian point of view and are summarized in the following proposition, whose proof follows directly from (11) and Bayes' theorem.

Proposition 4. Let $[Z|S = s] \sim SS(f, G, s)$, where f does not depend on s, and $S \sim H$. If H is the cdf of a symmetric distribution around zero, then the marginal distribution of Z is symmetric around zero and has pdf f. Moreover, if H has a pdf $h = H'$, then $[S|Z = z] \sim SS(h, G, z)$.

For example, consider a simple location-scale model $Y_i = \mu + \sigma Z_i$, $i = 1, \ldots, n$, with a prior $\pi(\mu, \sigma)$ for (μ, σ) and $[Z_i|S_i = \alpha_i] \overset{ind.}{\sim} SS(f, G, \alpha_i)$, $i = 1, \ldots, n$, independent of (μ, σ). We can conclude from Proposition 4 that Bayesian inference on (μ, σ) will be the same as that obtained under the symmetric location-scale model $\sigma^{-1} f((y - \mu)/\sigma)$ for the data Y_i's when a common symmetric (at zero) prior, H say, is considered for the shape parameters α_i's; see also Remark 1 in Section 4.1.

2.4 Shape mixtures of univariate SN distributions

An interesting example of the previous results is the skew-generalized-normal distribution defined by (1) and (3). As indicated in (4), this model can be specified as a shape mixture of the skew-normal distribution by taking a normal mixing distribution for the shape parameter, see Arellano-Valle, Gómez and Quintana (2004). From Proposition 1, this is equivalent to considering $Y = \mu + \sigma Z$, and supposing that $[Z|S = s] \sim SS(f, G, s)$ and $S \sim H$, with $f(t) = \phi(t)$, $G(t) = \Phi(t)$ and $H(t) = \Phi(\frac{t-\eta}{\tau})$, that is, $[Z|S = s] \sim SN(s)$ and $S \sim N(\eta, \tau^2)$. Since $S_0 = \frac{S-\eta}{\tau} \sim N(0, 1)$, then by (10) it follows that (see also Ellison, 1964)

$$Q(z) = \mathrm{E}[\Phi(z\tau S_0 + z\eta)] = \Phi\left(\frac{\eta z}{\sqrt{1 + \tau^2 z^2}}\right),$$

implying the following marginal pdf for the random variable Z:

$$f_Z(z) = 2\phi(z)\Phi\left(\frac{\eta z}{\sqrt{1 + \tau^2 z^2}}\right), \quad z \in \mathbb{R}. \tag{12}$$

This is the skew-generalized-normal distribution mentioned in the introduction, and denoted by $Z \sim SGN(\eta, \tau)$. Note that $SGN(0, \tau) = N(0, 1)$, for any τ, $SGN(\eta, \infty) = N(0, 1)$, for any η, and $SGN(\eta, 0) = SN(\eta)$. A special case is obtained by letting $\tau = \sqrt{\eta^2}$, which is called curved skew-normal distribution. Further properties and applications of this model are considered in Arellano-Valle, Gómez and Quintana (2004). For instance, we note that if $Z \sim SGN(\eta, \tau)$, then $-Z \sim SGN(-\eta, \tau)$, $Z^2 \sim \chi^2_1$ and

$$[Z|S = s] \overset{d}{=} \frac{s}{\sqrt{1 + s^2}}|U| + \frac{1}{\sqrt{1 + s^2}}V, \tag{13}$$

where $S \sim N(\eta, \tau^2)$ is independent of U and V, which are independent and identically distributed (i.i.d.) $N(0, 1)$ random variables. The latter representation is useful from a computational point of view, because it implies that if $Z \sim SGN(\eta, \tau)$, then there exist random variables S and T mutually independent such that:

$$\begin{aligned}
&(i) \quad [Z|S = s, T = t] \sim N\left(\frac{st}{\sqrt{1 + s^2}}, \frac{1}{1 + s^2}\right),\\
&(ii) \quad S \sim N(\eta, \tau^2),\\
&(iii) \quad T \sim HN(0, 1),
\end{aligned}$$

where $HN(0,1)$ is the half-normal distribution. This hierarchical specification can be used to implement MCMC methods (from a Bayesian approach) or the EM algorithm (from a classical approach), in order to make inference about (η, τ) based on a random sample $Z_1, \ldots, Z_n$ from $Z \sim SGN(\eta, \tau)$.

As mentioned in the introduction, further examples belonging to the class $2f(z)G(\alpha z)$ arise by letting $S = \sqrt{W}\alpha$, for some non-negative random variable W. Moreover, the subclass of scale and shape mixtures of the skew-normal distribution can be introduced by considering (5). Hence, this is a representation of the subclass $2f(z)G(\alpha z)$ where f and G are the pdf and cdf of a distribution which is a scale mixture of the normal one.

3 Shape Mixtures of Multivariate SN Distributions

Some multivariate extensions of the skew-generalized-normal pdf in (12) are given next. In all these cases, the shape mixture idea discussed in the previous sections is adapted to independent and dependent multivariate skew-normal distributions. The resulting distributions can be interpreted as a Bayesian modeling of these independent and dependent skew-normal observations when a normal prior specification for the shape parameters is considered. Thus, as established in the following propositions, both the predictive function and the posterior pdf associated with the shape parameters define multivariate skew-generalized-normal pdf's.

For any n-dimensional vector $\mathbf{w}$, denote by $\mathbf{D}(\mathbf{w})$ the $n \times n$ diagonal matrix formed by the components $w_1, \ldots, w_n$ of $\mathbf{w}$. Then, for any two n-dimensional vectors $\mathbf{s}$ and $\mathbf{z}$, we have $\mathbf{D}(\mathbf{s})\mathbf{z} = \mathbf{D}(\mathbf{z})\mathbf{s} = (s_1 z_1, \ldots, s_n z_n)^T$. Denote by $\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and by $\Phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the pdf and the cdf of the multivariate $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, respectively. When $\boldsymbol{\mu} = \mathbf{0}$ these functions are denoted by $\phi_n(\mathbf{y}|\boldsymbol{\Sigma})$ and $\Phi_n(\mathbf{y}|\boldsymbol{\Sigma})$. The proof of the next propositions are based on the following well-known result, see, e.g., Arellano-Valle and Genton (2005). If $\mathbf{U} \sim N_k(\mathbf{c}, \mathbf{C})$ is a non-singular normal random vector, then for any vectors $\mathbf{a} \in \mathbb{R}^k$ and $n \times k$ matrix $\mathbf{B}$, we have that

$$\mathrm{E}[\Phi_n(\mathbf{B}\mathbf{U} + \mathbf{a}|\, \mathbf{d}, \mathbf{D})] = \Phi_n(\mathbf{a}|\mathbf{B}\mathbf{c} + \mathbf{d}, \mathbf{D} + \mathbf{B}\mathbf{C}\mathbf{B}^T). \tag{14}$$

Proposition 5. Suppose that $[Z_i|\mathbf{S} = \mathbf{s}] \overset{ind.}{\sim} SN(s_i)$, $i = 1, \ldots, n$, where the shape $\mathbf{S} = (S_1, \ldots, S_n)^T \sim N_n(\boldsymbol{\eta}, \boldsymbol{\Omega})$. Then, the marginal pdf of $\mathbf{Z} = (Z_1, \ldots, Z_n)^T$ is given by

$$f(\mathbf{z}|\boldsymbol{\eta}, \boldsymbol{\Omega}) = 2^n \phi_n(\mathbf{z})\Phi_n(\mathbf{D}(\boldsymbol{\eta})\mathbf{z}|\mathbf{I}_n + \mathbf{D}(\mathbf{z})\boldsymbol{\Omega}\mathbf{D}(\mathbf{z})),$$

which contains the $N_n(\mathbf{0}, \mathbf{I}_k)$ pdf for $\boldsymbol{\eta} = \mathbf{0}$ and the independent multivariate skew-normal pdf given by $2^n \phi_n(\mathbf{z})\Phi_n(\mathbf{D}(\boldsymbol{\eta})\mathbf{z})$ for $\boldsymbol{\Omega} = \mathbf{O}$, the zero matrix. Moreover, the conditional pdf of $\mathbf{S}$ given $\mathbf{Z} = \mathbf{z}$ is given by

$$f(\mathbf{s}|\mathbf{z}, \boldsymbol{\eta}, \boldsymbol{\Omega}) = \frac{\phi_n(\mathbf{s}|\boldsymbol{\eta}, \boldsymbol{\Omega})\Phi_n(\mathbf{D}(\mathbf{z})\mathbf{s})}{\Phi_n(\mathbf{D}(\boldsymbol{\eta})\mathbf{z}|\mathbf{I}_n + \mathbf{D}(\mathbf{z})\boldsymbol{\Omega}\mathbf{D}(\mathbf{z}))}.$$

Proposition 6. Suppose that $[\mathbf{Z}|\mathbf{S} = \mathbf{s}] \sim SN_n(\mathbf{s})$, i.e., its pdf is $2\phi_n(\mathbf{z})\Phi(\mathbf{s}^T\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^n$, where $\mathbf{S} = (S_1, \ldots, S_n)^T \sim N_n(\boldsymbol{\eta}, \boldsymbol{\Omega})$. Then, the marginal pdf of $\mathbf{Z} = (Z_1, \ldots, Z_n)^T$ is given by

$$f(\mathbf{z}|\boldsymbol{\eta},\boldsymbol{\Omega}) = 2\phi_n(\mathbf{z})\Phi\left(\frac{\boldsymbol{\eta}^T\mathbf{z}}{\sqrt{1+\mathbf{z}^T\boldsymbol{\Omega}\mathbf{z}}}\right),$$

which contains the $N_n(\mathbf{0},\mathbf{I}_k)$ pdf for $\boldsymbol{\eta}=\mathbf{0}$ and the dependent multivariate skew-normal pdf $2\phi_n(\mathbf{z})\Phi(\boldsymbol{\eta}^T\mathbf{z})$ for $\boldsymbol{\Omega}=\mathbf{O}$. Moreover, the conditional pdf of $\mathbf{S}$ given $\mathbf{Z}=\mathbf{z}$ is given by

$$f(\mathbf{s}|\mathbf{z},\boldsymbol{\eta},\boldsymbol{\Omega}) = \frac{\phi_n(\mathbf{s}|\boldsymbol{\eta},\boldsymbol{\Omega})\Phi(\mathbf{z}^T\mathbf{s})}{\Phi\left(\frac{\boldsymbol{\eta}^T\mathbf{z}}{\sqrt{1+\mathbf{z}^T\boldsymbol{\Omega}\mathbf{z}}}\right)}.$$

Some particular cases of the above multivariate skew-generalized-normal distributions are obtained when we assume that:

(i) $S_1,\ldots,S_n$ are i.i.d. $N(\eta,\tau^2)$ random variables, i.e., $\boldsymbol{\eta}=\eta\mathbf{1}_n$ and $\boldsymbol{\Omega}=\tau^2\mathbf{I}_n$; or

(ii) $S_i \overset{ind.}{\sim} N(\eta_i,\tau_i^2)$, $i=1,\ldots,n$, so that $\boldsymbol{\Omega}=\mathrm{diag}\{\tau_1^2,\ldots,\tau_n^2\}$; or

(iii) $S_1,\ldots,S_n$ are exchangeable normal random variables, which is equivalent to considering $\boldsymbol{\eta}=\eta\mathbf{1}_n$ and $\boldsymbol{\Omega}=\tau^2\{(1-\rho)\mathbf{I}_n+\rho\mathbf{1}_n\mathbf{1}_n^T\}$, where $\rho\in[0,1)$.

Note that (i) is a particular case of (ii), as well as of (iii).

All the above models are derived by assuming marginal skew-normal observations with different shape parameters. The situation with a common shape parameter for all the observations is considered next.

Proposition 7. Let $[Z_i|S=s] \overset{i.i.d.}{\sim} SN(s)$, $i=1,\ldots,n$, where $S\sim N(\eta,\tau^2)$. Then, the marginal pdf of $\mathbf{Z}=(Z_1,\ldots,Z_n)^T$ is given by

$$f(\mathbf{z}|\eta,\tau^2) = 2^n\phi_n(\mathbf{z})\Phi_n(\eta\mathbf{z}|\mathbf{I}_n+\tau^2\mathbf{z}\mathbf{z}^T),$$

and the conditional pdf of S given $\mathbf{Z}=\mathbf{z}$ is given by

$$f(s|\mathbf{z},\eta,\tau) = \frac{\phi(s|\eta,\tau^2)\Phi_n(\mathbf{z}s)}{\Phi_n(\eta\mathbf{z}|\mathbf{I}_n+\tau^2\mathbf{z}\mathbf{z}^T)}.$$

Proposition 8. Let $[\mathbf{Z}|S=s]\sim SN_n(s\mathbf{1}_n)$, i.e. its pdf is $2\phi_n(\mathbf{z})\Phi(ns\bar{z})$, where $\bar{z}=n^{-1}\sum_{i=1}^n z_i$ and $S\sim N(\eta,\tau^2)$. Then, the marginal pdf of $\mathbf{Z}=(Z_1,\ldots,Z_n)^T$ is given by

$$f(\mathbf{z}|\eta,\tau) = 2\phi_n(\mathbf{z})\Phi\left(\frac{n\eta\bar{z}}{\sqrt{1+n^2\tau^2\bar{z}^2}}\right),$$

and the conditional pdf of S given $\mathbf{Z}=\mathbf{z}$ is given by

$$f(s|\mathbf{z},\eta,\tau) = \frac{\phi(s|\eta,\tau^2)\Phi(n\bar{z}s)}{\Phi\left(\frac{n\eta\bar{z}}{\sqrt{1+n^2\tau^2\bar{z}^2}}\right)}.$$

An interesting fact is that the results in Propositions 7 and 8 can be obtained as particular cases of Propositions 5 and 6, respectively, when we consider the exchangeable normal distribution described in (iii) with $\rho = 1$ for the shape variables $S_1, \ldots, S_n$.

In all of the above cases, multivariate location-scale skew-generalized-normal distributions for $\mathbf{Z}$ can be obtained through an affine linear transformation of the form $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}$, for a given location vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and non-singular scale matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$. In Proposition 5, however, another possibility is to incorporate the scale matrix $\boldsymbol{\Sigma}$ directly in the conditional distribution of $\mathbf{Z}$ given $\mathbf{S} = \mathbf{s}$. Thus, many other multivariate families of skewed distributions can be obtained in the same way.

Finally, as was mentioned at the beginning of this section, Bayesian inference on the shape (vector of) parameter(s) $\mathbf{S}$ can be obtained from the above results. In fact, if we consider that $\mathbf{Z} = (Z_1, \ldots, Z_n)^T$ is an observed vector of data from an independent or a dependent skew-normal distribution with shape parameter $\mathbf{S}$, then the resulting marginal and conditional pdf's of $\mathbf{Z}$ and $\mathbf{S}$ given $\mathbf{Z} = \mathbf{z}$ correspond to the predictive and posterior distributions of $\mathbf{S}$, respectively, when a normal prior for the shape parameter is considered. Both of these distributions belong to the so-called unified skew-normal (SUN) distributions discussed by Arellano-Valle and Azzalini (2006). Exploring these aspects in connection with the conjugacy theory can be an interesting topic of investigation under the more general situations where an independent or a dependent location-scale skew-normal model is considered.

4 Bayesian Inference for SN Regression Models

Inference on linear regression models and related problems have been approached from a Bayesian point of view under the assumption that the error terms are symmetrically distributed. Most of the research has been developed under a multivariate spherical normal distribution for the error vector. However, there are some extensions of the normal linear model to the spherical linear model (see, for example, Arellano-Valle, del Pino and Iglesias, 2006, and references therein), where the attention is focused on the robustness of the inferential normal theory. Moreover, Sahu, Dey and Branco (2003) implemented a posterior regression analysis by considering skewed distributions for the error terms. More recently, other authors have also been working on this subject in an even more general context than a simple linear regression model; see, e.g., Ma, Genton and Davidian (2004), Arellano-Valle, Bolfarine and Lachos (2007) and Ghosh, Branco and Chakraborty (2007). In this section, we consider a Bayesian analysis of the linear regression model for observations $(Y_i, \mathbf{x}_i)$, $\mathbf{x}_i \in \mathbb{R}^k$, $i = 1, \ldots, n$, when the error terms are i.i.d. with a skew-normal distribution.

Consider the linear regression model

$$Y_i = \mathbf{x}_i^T\boldsymbol{\beta} + \sigma\varepsilon_i, \quad i = 1, \ldots, n,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. $SN(\alpha)$ random errors, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T$, σ^2 and α are unknown parameters. That is, we consider the skew-normal linear regression model

$$[Y_i|\boldsymbol{\beta}, \sigma^2, \alpha] \overset{ind.}{\sim} SN(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2, \alpha), \quad i = 1, \ldots, n, \tag{15}$$

whose likelihood function is

$$f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \alpha) = \frac{2^n}{\sigma^n}\, \phi_n\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right) \Phi_n\left(\alpha\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right)\right), \tag{16}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^T$ and $\mathbf{X}$ is the $n \times k$ matrix with rows $\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T$.

In order to obtain posterior inferences on functions of $(\boldsymbol{\beta}, \sigma^2, \alpha)$, we assume that

$$\alpha \perp\!\!\!\perp (\boldsymbol{\beta}, \sigma^2), \tag{17}$$

where the symbol $\perp\!\!\!\perp$ is used to indicate independence, and we consider the following scenarios for the specifications of the prior distributions $\pi(\alpha)$ and $\pi(\boldsymbol{\beta}, \sigma^2)$:

$$(i) \quad \pi(\boldsymbol{\beta}, \sigma^2) = \text{arbitrary}, \quad \alpha \sim N(a, b^2); \tag{18}$$

$$(ii) \quad \pi(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}, \quad \alpha \sim N(a, b^2); \tag{19}$$

$$(iii) \quad \pi(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}, \quad \pi(\alpha) = \text{arbitrary}. \tag{20}$$

Even when a Gibbs sampling scheme based on the corresponding conditional distributions can be implemented (see Section 4.2) in order to obtain the required posterior analysis, it is possible to use the results obtained in the previous sections to derive some partial analytically tractable expressions under the above prior specifications.

4.1 The marginal likelihood function of $(\boldsymbol{\beta}, \sigma^2)$ when $\alpha \sim N(a, b^2)$

The main objective of this section is to obtain the marginal likelihood function of $(\boldsymbol{\beta}, \sigma^2)$ under the prior specifications given by (17) and (18). As a byproduct of this result, we show that under the particular prior specifications in (19) we obtain a proper predictive function, which guarantees that the posterior distribution of $(\boldsymbol{\beta}, \sigma^2, \alpha)$ is also proper.

Proposition 9. Consider the skew-normal linear regression model in (15). Then, under the prior specifications in (17) and (18), the marginal likelihood function of $(\boldsymbol{\beta}, \sigma^2)$ is

$$f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) = \frac{2^n}{\sigma^n}\, \phi_n\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right) \Phi_n\left(a\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right) \middle|\, \mathbf{I}_n + b^2\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right)\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right)^T\right).$$

Proof. Let $Z = \frac{\alpha - a}{b}$ and $\mathbf{t} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/\sigma$. Since $\alpha \sim N(a, b^2)$, we have from (16) that $f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) = \frac{2^n}{\sigma^n}\, \phi_n(\mathbf{t})\mathrm{E}[\Phi_n(\mathbf{b}Z + \mathbf{a})]$, where $\mathbf{a} = a\mathbf{t}$, $\mathbf{b} = b\mathbf{t}$ and $Z \sim N(0, 1)$. Thus, the proof follows from (14). □

Remark 1. An alternative marginal likelihood is obtained when in (15) the common shape parameter α is replaced by α_i and $\alpha_i \overset{ind.}{\sim} N(a_i, b_i^2)$, $i = 1, \ldots, n$, in (18) with

$(\alpha_1, \ldots, \alpha_n)$ independent of $(\boldsymbol{\beta}, \sigma^2)$. In this case, Proposition 5 yields $[Y_i|\boldsymbol{\beta}, \sigma^2] \overset{ind.}{\sim} SGN(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2, a_i, b_i^2)$, $i = 1, \ldots, n$, whose joint pdf is

$$f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) = \frac{2^n}{\sigma^n}\phi_n\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right)\prod_{i=1}^n \Phi\left(\frac{a_i\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1 + b_i^2\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)^2}}\right).$$

Thus, the special prior specification $\alpha_i \overset{ind.}{\sim} N(0, b_i^2)$, $i = 1, \ldots, n$, is equivalent to considering the standard normal linear regression model for the conditional distribution of $Y_1, \ldots, Y_n$ given $(\boldsymbol{\beta}, \sigma^2)$. An analogous result is obtained when $b_i^2 \to \infty$, $i = 1, \ldots, n$, which can be interpreted as a diffuse joint prior distribution for $\alpha_1, \ldots, \alpha_n$.

A consequence of Proposition 9 is given in the following corollary. It establishes the propriety of predictive functions, and thus of the posterior distributions of $\boldsymbol{\beta}$, σ^2 and α, when, in addition to the normal prior distribution for α, we consider also the usual noninformative prior distribution for $(\boldsymbol{\beta}, \sigma^2)$. We generalize this result in the next subsection for an arbitrary proper prior distribution for the shape parameter α.

Corollary 3. If in Proposition 9 we consider the prior specifications in (19), then the posterior distribution of $(\boldsymbol{\beta}, \sigma^2, \alpha)$ is proper.

Proof. By Proposition 9, we have under (19) that the predictive function is

$$\begin{aligned} f(\mathbf{y}) &= \int_{\mathbb{R}^k}\int_0^\infty \frac{2^n}{\sigma^{n+2}}\,\phi_n\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right) \\ &\quad \times\Phi_n\left(a\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right)\Big|\,\mathbf{I}_n + b^2\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right)\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right)^T\right)d\sigma^2 d\boldsymbol{\beta} \\ &\leq 2^n\int_{\mathbb{R}^k}\int_0^\infty \frac{1}{\sigma^{n+2}}\,\phi_n\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\sigma}\right)d\sigma^2 d\boldsymbol{\beta} \\ &< \infty, \end{aligned}$$

where we used the fact that this last integral corresponds to the predictive function under the standard symmetric normal linear regression model $[Y_1, \ldots, Y_n|\boldsymbol{\beta}, \sigma^2] \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$ and the usual noninformative prior distribution $\pi(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}$, which is well-known to be proper. □

4.2 A Gibbs sampling scheme for the SN regression model

In this section, we give the conditional distributions needed to implement a Gibbs sampling procedure in order to obtain the required posterior analysis of the SN linear regression model (15) when the prior specifications in (17) and (19) are considered. For this objective, we note first that an appropriate use of the stochastic representation in

(13), conditionally on $S = \alpha$, yields the following equivalent specification of (15):

$$(i) \quad [Y_i|\boldsymbol{\beta}, \sigma^2, \alpha, \tau_i] \overset{ind.}{\sim} N\left(\frac{\alpha\tau_i}{\sqrt{1+\alpha^2}} + \mathbf{x}_i^T\boldsymbol{\beta}, \frac{\sigma^2}{1+\alpha^2}\right),\ i = 1, \ldots, n,$$

$$(ii) \quad \tau_i \overset{i.i.d.}{\sim} HN(0,1) \text{ and } \tau_i \perp\!\!\!\perp (\boldsymbol{\beta}, \sigma^2, \alpha),\ i = 1, \ldots, n.$$

Then it is straightforward to obtain the required conditional distributions to implement a Gibbs sampling scheme. In fact, considering the transformations $\omega_i = \psi\tau_i$, $i = 1, \ldots, n$, where ψ is a new scale parameter defined by

$$\psi = \frac{\sigma}{\sqrt{1+\alpha^2}}, \tag{21}$$

the above model representation can be rewritten as

$$[Y_i|\boldsymbol{\beta}, \psi^2, \alpha, \omega_i] \overset{ind.}{\sim} N(\alpha\omega_i + \mathbf{x}_i^T\boldsymbol{\beta}, \psi^2),\ i = 1, \ldots, n, \tag{22}$$

$$[\omega_i|\psi^2] \overset{i.i.d.}{\sim} HN(0, \psi^2) \text{ and } \omega_i \perp\!\!\!\perp (\boldsymbol{\beta}, \alpha),\ i = 1, \ldots, n. \tag{23}$$

Moreover, it is easy to show from (17), (19) and (21) that the prior specification associated with the parameters $\boldsymbol{\beta}$, ψ^2 and α is such that

$$\pi(\boldsymbol{\beta}, \psi^2|\alpha) \propto \frac{1}{\psi^2} \quad \text{and} \quad \alpha \sim N(a, b^2). \tag{24}$$

Therefore, in terms of the new parameterization $(\beta_1, \ldots, \beta_k, \psi^2, \alpha, \omega_1, \ldots, \omega_n)$, the following conditional posterior distributions are necessary to implement a Gibbs sampling procedure:

$$[\beta_i|\boldsymbol{\beta}_{-i}, \psi^2, \alpha, \boldsymbol{\omega}, \mathbf{y}]; \quad [\psi^2|\boldsymbol{\beta}, \alpha, \boldsymbol{\omega}, \mathbf{y}]; \quad [\alpha|\boldsymbol{\beta}, \psi^2, \boldsymbol{\omega}, \mathbf{y}]; \quad [\omega_i|\boldsymbol{\beta}, \psi^2, \alpha, \omega_{-i}, \mathbf{y}];$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T$ and for any vector $\mathbf{u} = (u_1, \ldots, u_p)^T$, the vector $\mathbf{u}_{-i}$ is defined by $(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_p)^T$.

Since under (24) the propriety of the marginal posterior distributions is guaranteed by Corollary 3, the conditional posterior distributions necessary to implement a Gibbs sampling procedure with the objective of obtaining the required posterior analysis are established in the next proposition. The proof of this proposition follows from standard algebraic manipulations, and therefore is omitted.

Proposition 10. Consider the conditional representation (22)-(23) of the SN linear regression model (15), with the prior specifications in (24). Then,

$$\begin{aligned}
[\boldsymbol{\beta}|\psi^2, \alpha, \boldsymbol{\omega}, \mathbf{y}] &\sim N_k\left(\widehat{\boldsymbol{\beta}}(\mathbf{y}) - \alpha\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega}), \psi^2(\mathbf{X}^T\mathbf{X})^{-1}\right), \\
[\boldsymbol{\omega}|\boldsymbol{\beta}, \psi^2, \alpha, \mathbf{y}] &\sim TN_n\left(\mathbf{0}; \frac{\alpha}{1+\alpha^2}\boldsymbol{\epsilon}, \frac{\psi^2}{1+\alpha^2}\mathbf{I}_n\right), \\
[\psi^2|\boldsymbol{\beta}, \alpha, \boldsymbol{\omega}, \mathbf{y}] &\sim IG\left(n, \frac{\|\boldsymbol{\epsilon} - \alpha\boldsymbol{\omega}\|^2 + \|\boldsymbol{\omega}\|^2}{2}\right), \\
[\alpha|\boldsymbol{\beta}, \psi^2, \boldsymbol{\omega}, \mathbf{y}] &\sim N\left(\frac{b^2\boldsymbol{\omega}^T\boldsymbol{\epsilon} + a\psi^2}{b^2\|\boldsymbol{\omega}\|^2 + \psi^2}, \frac{b^2\psi^2}{b^2\|\boldsymbol{\omega}\|^2 + \psi^2}\right),
\end{aligned}$$

where $\widehat{\boldsymbol{\beta}}(\mathbf{z}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{z}$, $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, $IG(\alpha, \gamma)$ is the Inverse Gamma distribution and $TN_n(\mathbf{c}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution truncated below at $\mathbf{c}$.

An application based on the results of Proposition 10 is given in Section 6.

4.3 Posterior analysis under an arbitrary proper prior for α

In this section, we consider a different approach for studying the posterior distributions of $\boldsymbol{\beta}$, σ^2 and α, based on the prior specifications given by (17) and (20), i.e., by considering that these parameters are independent with a noninformative prior for $(\boldsymbol{\beta}, \sigma^2)$ and an arbitrary proper prior for the shape parameter α. We note first that for $\alpha = 0$, the skew-normal likelihood in (16) reduces to the standard symmetric normal likelihood function

$$f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \alpha = 0) = \frac{1}{\sigma^n}\phi_n\left(\frac{\mathbf{y} - \mathbf{X}\beta}{\sigma}\right). \tag{25}$$

Consequently, under the noninformative prior distribution $\pi(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}$, we obtain the following well-known posterior distributions for $\boldsymbol{\beta}$ and σ^2 :

$$[\boldsymbol{\beta}|\mathbf{y}, \alpha = 0] \sim t_k(\widehat{\boldsymbol{\beta}}, S^2(\mathbf{X}^T\mathbf{X})^{-1}, n-k), \quad [\sigma^2|\mathbf{y}, \alpha = 0] \sim IG\left(\frac{n-k}{2}, \frac{(n-k)S^2}{2}\right),$$

where $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $IG(a, b)$ denote the p-variate Student-t and Inverse Gamma distributions, respectively, and

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \quad \text{and} \quad S^2 = \frac{\|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\|^2}{n-k},$$

are the ordinary least squares estimators of $\boldsymbol{\beta}$ and σ^2, respectively. In the sequel, we denote by $\pi(\boldsymbol{\beta}|\mathbf{y}, \alpha = 0)$ and $\pi(\sigma^2|\mathbf{y}, \alpha = 0)$ the corresponding pdf of the conditional posterior distributions above and by $T_p(\mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = T_p(\mathbf{z} - \boldsymbol{\mu}|\boldsymbol{\Sigma}, \nu)$ the cdf of the multivariate Student-t distribution $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. With these ingredients we obtain the following results.

Proposition 11. Consider the skew-normal likelihood (16) and the prior specifications (20). Then, the full posterior of $(\boldsymbol{\beta}, \sigma^2, \alpha)$ is proper.

Proof. Since $\pi(\boldsymbol{\beta}, \sigma^2, \alpha|\mathbf{y}) \propto \sigma^{-(n+2)}\phi_n(\mathbf{z})\Phi_n(\alpha\mathbf{z})\pi(\alpha)$, where $\mathbf{z} = \sigma^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, it is straightforward to see that the marginal posterior of $(\boldsymbol{\beta}, \sigma^2)$ is given by

$$\pi(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) \propto \pi(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \alpha = 0)\mathrm{E}_\alpha\{\Phi_n(\alpha\mathbf{z})\}.$$

Here $\mathrm{E}_\alpha\{\Phi_n(\alpha\mathbf{z})\} = \int_{-\infty}^{\infty}\Phi_n(\alpha\mathbf{z})\pi(\alpha)d\alpha$, and

$$\pi(\boldsymbol{\beta}, \sigma|\mathbf{y}, \alpha = 0) \propto \sigma^{-(n+2)}\phi_n(\mathbf{e}/\sigma)\phi_n(\boldsymbol{\beta}|\widehat{\boldsymbol{\beta}}, \sigma^2(\mathbf{X}\mathbf{X}^T)^{-1}),$$

where $\mathbf{e} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$, is the posterior of $(\boldsymbol{\beta}, \sigma^2)$ under the noninformative prior $\pi(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}$, which is known to be proper. Therefore, the proof follows by noting that $0 \leq \mathrm{E}_\alpha\{\Phi_n(\alpha\mathbf{z})\} \leq 1$ for all $\mathbf{z}$ since $\pi(\alpha)$ is proper. □

Proposition 12. Under the skew-normal likelihood (16) and the prior specifications in (20), it follows that

$$\pi(\boldsymbol{\beta}|\mathbf{y},\alpha) \propto \pi(\boldsymbol{\beta}|\mathbf{y},\alpha=0)\ T_n(\alpha\boldsymbol{\epsilon}\mid n^{-1}\|\boldsymbol{\epsilon}\|^2\mathbf{I}_n, n), \tag{26}$$

$$\pi(\sigma^2|\mathbf{y},\alpha) \propto \pi(\sigma^2|\mathbf{y},\alpha=0)\ \Phi_n(\alpha\mathbf{e}\mid \sigma^2(\mathbf{I}_n+\alpha^2\mathbf{P})), \tag{27}$$

$$\pi(\alpha|\mathbf{y}) \propto \pi(\alpha)\ T_n(\alpha\mathbf{e}\mid S^2(\mathbf{I}_n+\alpha^2\mathbf{P}), n-k), \tag{28}$$

where $\boldsymbol{\epsilon}=\mathbf{y}-\mathbf{X}\boldsymbol{\beta}$, $\mathbf{e}=\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}$ and $\mathbf{P}=\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

Proof. Let $\boldsymbol{\epsilon}=\mathbf{y}-\mathbf{X}\boldsymbol{\beta}$. From (20) and (25), we have that

$$\begin{aligned}\pi(\boldsymbol{\beta},\alpha|\mathbf{y}) &\propto \pi(\alpha)\int_0^\infty \frac{1}{\sigma^{n+2}}\phi_n\left(\frac{\boldsymbol{\epsilon}}{\sigma}\right)\Phi_n\left(\frac{\alpha\boldsymbol{\epsilon}}{\sigma}\right)d\sigma^2\\ &= \pi(\alpha)\int_{\mathbf{u}<\alpha\boldsymbol{\epsilon}}\int_0^\infty \frac{e^{-\frac{\|\boldsymbol{\epsilon}\|^2+\|\mathbf{u}\|^2}{2\sigma^2}}}{\sigma^{2n+2}}d\sigma^2 d\mathbf{u}.\end{aligned}$$

With the change of variable $v=\frac{\|\boldsymbol{\epsilon}\|^2+\|\mathbf{u}\|^2}{2\sigma^2}$ in the first of the integrals above, we have that

$$\pi(\boldsymbol{\beta},\alpha|\mathbf{y}) \propto \pi(\alpha)\int_{\mathbf{u}<\alpha\boldsymbol{\epsilon}}\{\|\boldsymbol{\epsilon}\|^2+\|\mathbf{u}\|^2\}^{-n}d\mathbf{u}. \tag{29}$$

Then, it follows that $\pi(\boldsymbol{\beta}|\mathbf{y},\alpha)\propto\int_{\mathbf{u}<\alpha\boldsymbol{\epsilon}}\{\|\boldsymbol{\epsilon}\|^2+\|\mathbf{u}\|^2\}^{-n}d\mathbf{u}$, which yields (26) after considering the following well-known relation:

$$\|\boldsymbol{\epsilon}\|^2=\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2=\|\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\|^2+\|\mathbf{X}\widehat{\boldsymbol{\beta}}-\mathbf{X}\boldsymbol{\beta}\|^2. \tag{30}$$

Now, letting $\mathbf{t}=\mathbf{u}-\alpha\mathbf{e}$, where $\mathbf{e}=\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}$, we have from (29) that

$$\pi(\alpha|\mathbf{y}) \propto \pi(\alpha)\int_{\mathbf{t}<\alpha\mathbf{e}}\int_{\mathbb{R}^k}\{\|\boldsymbol{\epsilon}\|^2+\|\mathbf{t}-\alpha\mathbf{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\|^2\}^{-n}d\boldsymbol{\beta}d\mathbf{t}.$$

Thus, noting that (30) implies

$$\|\boldsymbol{\epsilon}\|^2+\|\mathbf{t}-\alpha\mathbf{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\|^2=(n-k)S^2+\mathbf{t}^T(\mathbf{I}_n+\alpha^2\mathbf{P})^{-1}\mathbf{t}+(1+\alpha^2)(\boldsymbol{\beta}-\mathbf{b})^T(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta}-\mathbf{b}), \tag{31}$$

where $\mathbf{b}=\widehat{\boldsymbol{\beta}}+\frac{\alpha}{1+\alpha^2}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{t}$, and using the properties of the Student-t distribution to solve the first of the integrals above, we have that

$$\pi(\alpha|\mathbf{y}) \propto (1+\alpha^2)^{-\frac{k}{2}}\pi(\alpha)\int_{\mathbf{t}<\alpha\mathbf{e}}\{(n-k)S^2+\mathbf{t}^T(\mathbf{I}_n+\alpha^2\mathbf{P})^{-1}\mathbf{t}\}^{-\frac{n+n-k}{2}}d\mathbf{t},$$

from where (28) follows. Finally, taking again $\mathbf{t}=\mathbf{u}-\alpha\mathbf{e}$, we have from (20), (25) and (31) that

$$\pi(\sigma^2|\mathbf{y},\alpha) \propto \frac{e^{-\frac{(n-k)S^2}{2\sigma^2}}}{\sigma^{2n+2}}\int_{\mathbf{t}<\alpha\mathbf{e}}e^{-\frac{\mathbf{t}^T(\mathbf{I}_n+\alpha^2\mathbf{P})^{-1}\mathbf{t}}{2\sigma^2}}\int_{\mathbb{R}^k}e^{-\frac{(1+\alpha^2)(\boldsymbol{\beta}-\mathbf{b})^T(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta}-\mathbf{b})}{2\sigma^2}}d\boldsymbol{\beta}d\mathbf{t},$$

from where we obtain (27). □

According to Proposition 11, from (26) to (28) we can implement an MCMC scheme based on the substitution procedure for any given proper prior distribution $\pi(\alpha)$.

5 Bayesian Inference for SMSN Regression Models

In this section, we present a Bayesian specification of three models that have been considered in the literature, by considering the shape mixture representation discussed in Section 2. These models are the skew-generalized-normal (SGN; see also Section 2.4), the skew-normal-t (SNT) and the skew-t-normal (STN) presented in the Introduction. As in the previous section, we focus our study on the linear regression model

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma\varepsilon_i, \qquad \varepsilon_i \overset{ind}{\sim} SN(0, 1/v_i, s_i), \tag{32}$$

$i = 1, \ldots, n$, with different prior specifications for the parameters v_i and s_i, $i = 1, \ldots, n$. The results will be applied in the next section to the Australian athletes dataset.

5.1 Skew-generalized-normal distributions

We consider the SGN linear regression model

$$[Y_i|\boldsymbol{\beta}, \sigma^2, \alpha_1, \alpha_2] \overset{ind.}{\sim} SGN(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2, \alpha_1, \alpha_2), \quad i = 1, \ldots, n, \tag{33}$$

whose likelihood function is

$$f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \alpha_1, \alpha_2) = (2/\sigma)^n \;\; \phi_n(\mathbf{z})\, \Phi_n\left(\alpha_1\mathbf{z}; \mathbf{I}_n + \alpha_2\mathbf{D}^2(\mathbf{z})\right),$$

where $\mathbf{z} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/\sigma$ and, as was defined earlier, $\mathbf{D}(\mathbf{z}) = \mathrm{diag}(z_1, \ldots, z_n)$. From the specification (32) and the results in Section 2.4 (see, e.g., (13)), we can write (33) as

$$[Y_i|\boldsymbol{\beta}, \sigma^2, \lambda_i] \quad \overset{ind.}{\sim} SN(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2, \lambda_i), \quad i = 1, \ldots, n, \tag{34}$$

$$[\lambda_i|\alpha_1, \alpha_2] \quad \overset{i.i.d.}{\sim} N(\alpha_1, \alpha_2), \quad i = 1, \ldots, n, \tag{35}$$

where $v_i = 1$ and $s_i = \lambda_i$, $i = 1, \ldots, n$. Therefore, letting $\psi_i^2 = \sigma^2/(1+\lambda_i^2)$, $i = 1, \ldots, n$, a convenient hierarchical representation of (33) is

$$[Y_i|\boldsymbol{\beta}, \sigma^2, \xi_i, \lambda_i] \quad \overset{ind.}{\sim} N(\mathbf{x}_i^T\boldsymbol{\beta} + \lambda_i\xi_i, \psi_i^2), \quad i = 1, \ldots, n, \tag{36}$$

$$[\xi_i \perp\!\!\!\perp \lambda_i|\psi_i^2], \qquad [\xi_i|\psi_i^2] \overset{ind.}{\sim} HN(0, \psi_i^2), \quad i = 1, \ldots, n, \tag{37}$$

$$[\psi_i^2|\lambda_i] \qquad \overset{ind.}{\sim} G_i, \quad [\lambda_i|\alpha_1, \alpha_2] \overset{i.i.d.}{\sim} N(\alpha_1, \alpha_2), \quad i = 1, \ldots, n, \tag{38}$$

where the G_i's are conditional distributions on the λ_i, which are determined by the prior distribution of σ^2.

We consider also the following prior specifications:

$$\pi(\boldsymbol{\beta}, \sigma^2, \alpha_1, \alpha_2) \propto \frac{1}{\sigma^2}\pi(\alpha_1|\alpha_2)\pi(\alpha_2), \tag{39}$$

with $\alpha_1 \sim N(a_1, \alpha_2)$ and $\alpha_2 \sim IG(a_2/2, b_2/2)$. Note that for $a_1 = 0$ it follows from (34)-(35) and by considering the results in Section 3 (see Proposition 5) that $[Y_i|\boldsymbol{\beta}, \sigma^2] \overset{ind.}{\sim} N(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2)$, $i = 1, \ldots, n$. Hence, in that situation the posterior inferences on $(\boldsymbol{\beta}, \sigma^2)$ must be based on the standard normal regression model. Moreover, since we are considering the improper prior $\pi(\sigma^2) \propto 1/\sigma^2$ for σ^2, we have then in (38) that the distributions G_i's are such that

$$\pi(\psi_i^2|\lambda_i) \propto \frac{1}{\psi_i^2}, \quad i = 1, \ldots, n. \tag{40}$$

The conditional posterior distributions necessary to implement a Gibbs sampling procedure with the aim of obtaining the required posterior analysis are established in the next proposition, whose proof follows from standard algebraic manipulations, and therefore is omitted.

Proposition 13. Consider the conditional representation (36)-(38) of the *SGN* linear regression model (33), with the prior specifications in (39)-(40). Then,

$$\begin{aligned}
[\boldsymbol{\beta}|\boldsymbol{\psi}, \alpha_1, \alpha_2, \boldsymbol{\xi}, \boldsymbol{\lambda}, \mathbf{y}] &\sim N_k\left(\widehat{\boldsymbol{\beta}}(\mathbf{y}) - \widehat{\boldsymbol{\beta}}(\mathbf{D}(\boldsymbol{\lambda})\boldsymbol{\xi}), (\mathbf{X}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\right),\\
[\boldsymbol{\xi}|\boldsymbol{\beta}, \boldsymbol{\psi}, \alpha_1, \alpha_2, \boldsymbol{\lambda}, \mathbf{y}] &\sim TN_n\left(\mathbf{0}; [\mathbf{I}_n + \mathbf{D}(\boldsymbol{\lambda})\mathbf{D}(\boldsymbol{\lambda})]^{-1}\mathbf{D}(\boldsymbol{\lambda})\boldsymbol{\epsilon}, [\mathbf{I}_n + \mathbf{D}(\boldsymbol{\lambda})\mathbf{D}(\boldsymbol{\lambda})]^{-1}\mathbf{D}(\boldsymbol{\psi})\right),\\
[\boldsymbol{\psi}|\boldsymbol{\beta}, \alpha_1, \alpha_2, \boldsymbol{\xi}, \boldsymbol{\lambda}, \mathbf{y}] &\sim \prod_{i=1}^{n} IG\left(1, \frac{(\epsilon_i - \lambda_i\xi_i)^2 + \xi_i^2}{2}\right),\\
[\boldsymbol{\lambda}|\boldsymbol{\beta}, \boldsymbol{\psi}, \alpha_1, \alpha_2, \boldsymbol{\xi}, \mathbf{y}] &\sim N_n\left(\mathbf{D}(\boldsymbol{\tau})[\mathbf{D}(\boldsymbol{\xi})\mathbf{D}(\boldsymbol{\psi})^{-1}\boldsymbol{\epsilon} + \alpha_1\alpha_2^{-1}\mathbf{I}_n\mathbf{1}_n], \mathbf{D}(\boldsymbol{\tau})\right),\\
[\alpha_1|\boldsymbol{\beta}, \boldsymbol{\psi}, \alpha_2, \boldsymbol{\xi}, \boldsymbol{\lambda}, \mathbf{y}] &\sim N\left(\frac{\boldsymbol{\lambda}^T\mathbf{1}_n + a_1}{n+1}, \frac{\alpha_2}{n+1}\right),\\
[\alpha_2|\boldsymbol{\beta}, \boldsymbol{\psi}, \alpha_1, \boldsymbol{\xi}, \boldsymbol{\lambda}, \mathbf{y}] &\sim IG\left(\frac{a_2 + n + 1}{2}, \frac{\|\boldsymbol{\lambda} - \alpha_1\mathbf{1}_n\|^2 + (\alpha_1 - a_1)^2 + b_2}{2}\right),
\end{aligned}$$

where $\widehat{\boldsymbol{\beta}}(\mathbf{z}) = (\mathbf{X}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\mathbf{z}$, $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{\psi} = (\psi_1^2, \ldots, \psi_n^2)^T$, $\boldsymbol{\tau} = \left(\frac{\alpha_2\psi_1^2}{\alpha_2\xi_1^2 + \psi_1^2}, \ldots, \frac{\alpha_2\psi_n^2}{\alpha_2\xi_n^2 + \psi_n^2}\right)^T$, $IG(\alpha, \gamma)$ is the Inverse Gamma distribution and $TN_n(\mathbf{c}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution truncated below at $\mathbf{c}$.

5.2 Skew-normal-t distributions

We start with the following result.

Proposition 14. If $[Z|S = s] \sim SN(\alpha_1\sqrt{s})$, where $S \sim Gamma(\nu/2, \nu/2)$, then marginally, the pdf of Z is given by

$$f_Z(z) = 2\phi(z)T(\alpha_1 z; \nu)$$

where $T(\cdot; \nu)$ is the cdf of the standard Student-t distribution with ν degrees of freedom.

Proof. The marginal pdf of Z is given by

$$\begin{aligned} f_Z(z) &= 2\phi(z)\int_0^\infty \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} s^{\nu/2-1} e^{-\nu s/2}\Phi(\alpha_1\sqrt{s}z)ds \\ &= 2\phi(z)\int_0^\infty \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} s^{\nu/2-1} e^{-\nu s/2}\int_{t\le z}\alpha_1\sqrt{s}\phi(\alpha_1\sqrt{s}t)dtds. \end{aligned}$$

Letting $w = t - z$, we have

$$f_Z(z) = 2\phi(z)\int_0^\infty \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} s^{\nu/2-1} e^{-\nu s/2}\int_{w\le 0}\alpha_1\sqrt{s}\phi(\alpha_1\sqrt{s}(w+z))dwds. \quad (41)$$

Applying Fubini's theorem and after some algebraic manipulations, we obtain from (41) that

$$f_Z(z) = 2\phi(z)\int_{w\le 0}\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\frac{\alpha_1}{\sqrt{\pi\nu}}\left[1+\frac{(\alpha_1(w+x))^2}{\nu}\right]^{-(\nu+1)/2} dw,$$

thus concluding the proof. □

The skew-normal-t (SNT) linear regression model is defined as

$$[Y_i|\boldsymbol{\beta},\sigma^2,\alpha_1,\nu] \overset{ind.}{\sim} SNT(\mathbf{x}_i^T\boldsymbol{\beta},\sigma^2,\alpha_1,\nu), \quad i=1,\dots,n, \quad (42)$$

whose likelihood function is

$$f(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\alpha_1,\nu) = (2/\sigma)^n\ \phi_n(\mathbf{z})\prod_{i=1}^n T(\alpha_1 z_i;\nu),$$

where $\mathbf{z} = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/\sigma$. Using Proposition 14 and specifications given in (32), we can rewrite (42) as

$$\begin{aligned} [Y_i|\boldsymbol{\beta},\sigma^2,\lambda_i,\alpha_1] &\overset{ind.}{\sim} SN(\mathbf{x}_i^T\boldsymbol{\beta},\sigma^2,\alpha_1\sqrt{\lambda_i}), \quad i=1,\dots,n, \\ [\lambda_i|\nu] &\overset{i.i.d.}{\sim} Gamma(\nu/2,\nu/2), \quad i=1,\dots,n, \end{aligned}$$

where $v_i = 1$ and $s_i = \alpha_1\sqrt{\lambda_i}$, $i=1,\dots,n$. Thus, letting $\psi_i^2 = \sigma^2/(1+\alpha_1^2\lambda_i)$, $i = 1,\dots,n$, we propose the following hierarchical representation of (42):

$$[Y_i|\boldsymbol{\beta},\sigma^2,\alpha_1,\xi_i,\lambda_i] \overset{ind.}{\sim} N(\mathbf{x}_i^T\boldsymbol{\beta}+\alpha_1\sqrt{\lambda_i}\xi_i,\psi_i^2), \quad i=1,\dots,n, \quad (43)$$

$$[\xi_i \perp\!\!\!\perp (\lambda_i,\alpha_1)|\psi_i^2], \quad [\xi_i|\psi_i^2] \overset{ind.}{\sim} HN(0,\psi_i^2), \quad i=1,\dots,n, \quad (44)$$

$$[\psi_i^2|\alpha_1,\lambda_i] \overset{ind.}{\sim} G_i, \quad [\lambda_i|\nu] \overset{i.i.d.}{\sim} Gamma(\nu/2,\nu/2), \quad i=1,\dots,n, \quad (45)$$

where the G_i's are conditional distributions on (α_1,λ_i) and are determined by the prior distribution of σ^2. Considering the same scheme for prior specifications as in the SGN case with

$$\alpha_1 \sim N(a_1,b_1^2), \quad \nu \sim Exponential(c_1/2), \quad (46)$$

we obtain the following full conditional distributions.

Proposition 15. Consider the conditional representation (43)-(45) of the SNT linear regression model (42), with the prior specifications in (46). Then,

$$
\begin{aligned}
[\boldsymbol{\beta}|\boldsymbol{\psi},\alpha_1,\nu,\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim N_k\left(\widehat{\boldsymbol{\beta}}(\mathbf{y})-\widehat{\boldsymbol{\beta}}(\alpha_1\boldsymbol{\kappa}),(\mathbf{X}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\right),\\
[\boldsymbol{\xi}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_1,\nu,\boldsymbol{\lambda},\mathbf{y}] &\sim TN_n\left(\mathbf{0};\alpha_1[\mathbf{I}_n+\alpha_1^2\mathbf{D}(\boldsymbol{\lambda})]^{-1}\mathbf{D}(\boldsymbol{\lambda})^{1/2}\boldsymbol{\epsilon},[\mathbf{I}_n+\alpha_1^2\mathbf{D}(\boldsymbol{\lambda})]^{-1}\mathbf{D}(\boldsymbol{\psi})\right),\\
[\boldsymbol{\psi}|\boldsymbol{\beta},\alpha_1,\nu,\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim \prod_{i=1}^n IG\left(1,\frac{(\epsilon_i-\alpha_1\kappa_i)^2+\xi_i^2}{2}\right),\\
[\boldsymbol{\lambda}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_1,\nu,\boldsymbol{\xi},\mathbf{y}] &\sim \prod_{i=1}^n Gamma\left(\frac{\nu}{2},\frac{\alpha_1^2\xi_i^2+\psi_i^2\nu}{2\psi_i^2}\right)\exp\left\{\frac{\alpha_1\kappa_i\epsilon_i}{\psi_i^2}\right\},\\
[\alpha_1|\boldsymbol{\beta},\boldsymbol{\psi},\nu,\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim N\left(\frac{a_1+b_1^2\boldsymbol{\kappa}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\boldsymbol{\epsilon}}{1+b_1^2\boldsymbol{\kappa}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\boldsymbol{\kappa}},\frac{b_1^2}{1+b_1^2\boldsymbol{\kappa}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\boldsymbol{\kappa}}\right),\\
[\nu|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_1,\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim \exp\left\{-\nu\left(\frac{\boldsymbol{\lambda}^T\mathbf{1}_n+c_1}{2}\right)\right\}\prod_{i=1}^n\lambda_i^{\nu/2-1},
\end{aligned}
$$

where $\widehat{\boldsymbol{\beta}}(\mathbf{z})=(\mathbf{X}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\mathbf{z}$, $\boldsymbol{\epsilon}=\mathbf{y}-\mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{\psi}=(\psi_1^2,\ldots,\psi_n^2)^T$, $\boldsymbol{\kappa}=(\lambda_1^{1/2}\xi_1,\ldots,\lambda_n^{1/2}\xi_n)^T$ and $TN_n(\mathbf{c};\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the $N_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution truncated below at $\mathbf{c}$.

5.3 Skew-t-normal distributions

We start with the following result.

Proposition 16. If $[Z|S=s]\sim SN(0,s^{-1},\alpha_1 s^{-1/2})$, where $S\sim Gamma(\nu/2,\nu/2)$, then marginally, the pdf of Z is given by

$$f_Z(z)=2t(z;\nu)\Phi(\alpha_1 z),$$

where $t(\cdot;\nu)$ denotes the pdf of the standard Student-t distribution with ν degrees of freedom.

Proof. The proof is straightforward using the well-known result about the predictive distribution of the normal-gamma Bayesian model (see Bernardo and Smith, 2000), i.e.

$$f_Z(z) = 2\Phi(\alpha_1 z)\int_0^\infty N(z;0,s^{-1})Ga(s;\nu/2,\nu/2)ds.$$

□

The skew-t-normal (STN) linear regression model is defined as

$$[Y_i|\boldsymbol{\beta},\sigma^2,\alpha_1,\nu]\overset{ind.}{\sim} STN(\mathbf{x}_i^T\boldsymbol{\beta},\sigma^2,\alpha_1,\nu),\quad i=1,\ldots,n, \tag{47}$$

whose likelihood function is

$$f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \alpha_1, \nu) = (2/\sigma)^n \left\{ \prod_{i=1}^{n} t\,(z_i; \nu) \right\} \Phi_n(\alpha_1 \mathbf{z}),$$

and $\mathbf{z} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/\sigma$. Using Proposition 16 and the model specifications given in (32), (47) can also be rewritten as

$$\begin{aligned} [Y_i|\boldsymbol{\beta}, \sigma^2, \lambda_i, \alpha_1] &\overset{ind.}{\sim} SN(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2\lambda_i^{-1}, \alpha_1\lambda_i^{-1/2}), \quad i = 1, \ldots, n, \\ [\lambda_i|\nu] &\overset{i.i.d.}{\sim} Gamma(\nu/2, \nu/2), \quad i = 1, \ldots, n, \end{aligned}$$

where $v_i = \lambda_i$ and $s_i = \alpha_1\lambda_i^{-1/2}$, $i = 1, \ldots, n$. Letting $\psi_i^2 = \sigma^2/(\lambda_i + \alpha_1^2)$, $i = 1, \ldots, n$, we propose the following hierarchical representation of (47):

$$[Y_i|\boldsymbol{\beta}, \sigma^2, \alpha_1, \xi_i, \lambda_i] \overset{ind.}{\sim} N(\mathbf{x}_i^T\boldsymbol{\beta} + \alpha_1\xi_i, \psi_i^2), \quad i = 1, \ldots, n, \tag{48}$$

$$[\xi_i \perp\!\!\!\perp (\lambda_i, \alpha_1)|\psi_i^2], \quad [\xi_i|\psi_i^2] \overset{ind.}{\sim} HN(0, \psi_i^2), \quad i = 1, \ldots, n, \tag{49}$$

$$[\psi_i^2|\alpha_1, \lambda_i] \overset{ind.}{\sim} G_i, \quad [\lambda_i|\nu] \overset{i.i.d.}{\sim} Gamma(\nu/2, \nu/2), \quad i = 1, \ldots, n. \tag{50}$$

Again, the G_i's are conditional distributions on (α_1, λ_i) and are determined by the prior distribution of σ^2. The prior scheme for this model is the same as for the SNT regression model. Note that this regression model considers both a mixture of the shape parameter and a mixture of the scale parameter for the skew-normal model.

Proposition 17. Consider the conditional representation (48)-(50) of the STN linear regression model (47), with the prior specifications in (46). Then,

$$\begin{aligned} [\boldsymbol{\beta}|\boldsymbol{\psi}, \alpha_1, \nu, \boldsymbol{\xi}, \boldsymbol{\lambda}, \mathbf{y}] &\sim N_k\left(\widehat{\boldsymbol{\beta}}(\mathbf{y}) - \alpha_1\widehat{\boldsymbol{\beta}}(\boldsymbol{\xi}), (\mathbf{X}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\right), \\ [\boldsymbol{\xi}|\boldsymbol{\beta}, \boldsymbol{\psi}, \alpha_1, \nu, \boldsymbol{\lambda}, \mathbf{y}] &\sim TN_n\left(\mathbf{0}; \frac{\alpha_1}{1+\alpha_1^2}\boldsymbol{\epsilon}, \frac{1}{1+\alpha_1^2}\mathbf{D}(\boldsymbol{\psi})\right), \\ [\boldsymbol{\psi}|\boldsymbol{\beta}, \alpha_1, \nu, \boldsymbol{\xi}, \boldsymbol{\lambda}, \mathbf{y}] &\sim \prod_{i=1}^{n} IG\left(1, \frac{(\epsilon_i - \alpha_1\xi_i)^2 + \xi_i^2}{2}\right), \\ [\boldsymbol{\lambda}|\boldsymbol{\beta}, \boldsymbol{\psi}, \alpha_1, \nu, \boldsymbol{\xi}, \mathbf{y}] &\sim \prod_{i=1}^{n} Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \\ [\alpha_1|\boldsymbol{\beta}, \boldsymbol{\psi}, \nu, \boldsymbol{\xi}, \boldsymbol{\lambda}, \mathbf{y}] &\sim N\left(\frac{a_1 + b_1^2\boldsymbol{\xi}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\boldsymbol{\epsilon}}{1 + b_1^2\boldsymbol{\xi}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\boldsymbol{\xi}}, \frac{b_1^2}{1 + b_1^2\boldsymbol{\xi}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\boldsymbol{\xi}}\right), \\ [\nu|\boldsymbol{\beta}, \boldsymbol{\psi}, \alpha_1, \boldsymbol{\xi}, \boldsymbol{\lambda}, \mathbf{y}] &\sim \exp\left\{-\nu\left(\frac{\boldsymbol{\lambda}^T\mathbf{1}_n + c_1}{2}\right)\right\}\prod_{i=1}^{n}\lambda_i^{\nu/2-1}, \end{aligned}$$

where $\widehat{\boldsymbol{\beta}}(\mathbf{z}) = (\mathbf{X}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{D}(\boldsymbol{\psi})^{-1}\mathbf{z}$, $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{\psi} = (\psi_1^2, \ldots, \psi_n^2)^T$ and $TN_n(\mathbf{c}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution truncated below at $\mathbf{c}$.

Remark 2. As a summary, note that, if we combine scale and shape mixtures, then all the above regression models, including the SN one, can be represented jointly by assuming that $[\varepsilon_i|v_i, s_i] \overset{i.i.d.}{\sim} SN(0, 1/v_i, s_i)$, $i = 1, \dots, n$, and considering that:

1. For the SN: $v_i = 1$ and $s_i = \alpha$ for all i;
2. For the SGN: $v_i = 1$ and $s_i = \lambda_i$ with $\lambda_i \overset{i.i.d.}{\sim} N(\alpha_1, \alpha_2)$ for all i;
3. For the SNT: $v_i = 1$ and $s_i = \alpha_1\sqrt{\lambda_i}$ with $\lambda_i \overset{i.i.d.}{\sim} Gamma(\nu/2, \nu/2)$ for all i;
4. For the STN: $v_i = \lambda_i$ and $s_i = \alpha_1\lambda_i^{-1/2}$, and $\lambda_i \overset{i.i.d.}{\sim} Gamma(\nu/2, \nu/2)$ for all i.

6 An Application to Australian Athletes Data

In order to illustrate our results, particularly the Bayesian specification of skew-normal shape mixtures models given in Section 5, we consider a dataset from Cook and Weisberg (1994) on characteristics of Australian athletes available from the Australian Institute of Sport (AIS). Specifically, we consider the variables lean body mass (Lbm), height (Ht) and weight (Wt) associated with $n = 102$ Australian male athletes. Table 1 presents a summary of the basic descriptive statistics for these variables.

Table 1: Descriptive statistics of the AIS dataset: sample mean $\overline{z}$, sample standard deviation s, and sample skewness and kurtosis coefficients $\sqrt{b_1}$ and b_2, respectively.

Variable	$\overline{z}$	s	$\sqrt{b_1}$	b_2
Lbm	74.66	9.89	0.28	0.71
Ht	185.50	7.90	0.07	0.06
Wt	82.52	12.40	0.40	0.49

Enhanced athletic performance is known to be linked to increase in lean body mass, the difference between total body weight and body fat. In order to study the relationship between the lean body mass of the AIS male athletes and their height and weight, we consider a linear regression model through the origin given by

$$Lbm_i = \beta_1 Ht_i + \beta_2 Wt_i + \sigma\varepsilon_i, \quad i = 1, \dots, 102,$$

assuming $[\varepsilon_i|v_i, s_i]$ are independent with $SN(0, 1/v_i, s_i)$ with prior specification for v_i and s_i, $i = 1, \dots, n$ such that we obtain the skew-normal shape mixtures regression models studied in Section 5, say, $SGN(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2, \alpha_1, \alpha_2)$, $SNT(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2, \alpha_1, \nu)$ and $STN(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2, \alpha_1, \nu)$. To elicitate the prior distributions for these models, we start with the Bayesian approach to fit a skew-normal (SN) model. We consider an improper prior for $(\boldsymbol{\beta}, \sigma^2)$ and a normal prior centered at 0 with variance 10 for the shape parameter. Such a prior specification reflects our belief in favor of normality of the data. Using Proposition 10, we implement a Gibbs sampling algorithm with the software R. For this algorithm, we run parallel chains of $50,000$ iterations and discard the first $25,000$ as the burn-in period with lags of 10 iterations to avoid autocorrelation. The posterior summaries of this model are reported in Table 2.

Table 2: Bayesian estimates for the AIS dataset under a skew-normal linear regression model.

	SN			
Parameter	mean	sd	2.5%	97.5%
β_1	0.06	0.01	0.04	0.08
β_2	0.81	0.03	0.76	0.86
σ	3.49	0.31	2.90	4.16
α	-3.97	1.13	-6.51	-2.15

In Figure 1 we provide a normal QQ-plot for the standardized residuals under a normal model. This plot shows strong evidence against normality. According to the Bayesian estimates reported in Table 2, the AIS dataset shows evidence of skewness, because the parameter α has a posterior mean equal to -3.97 and the credibility interval is $[-6.51, -2.15]$. Using this information, we fit the above shape mixtures of skew-normal linear regression models following a Bayesian point of view. For each model, we place essentially improper priors on the regression parameters and on the scale parameter σ. Based on the posterior summaries of the shape parameter of the skew-normal model, we adopt a prior distribution centered at -4 for α_1 with variance equal to 1. For the parameter α_2 in the SGN model, we adopt an Inverse Gamma $IG(3,2)$ distribution. This prior elicitation means that we fix both the prior mean and prior variance of α_2 equal to 1. In the SNT and STN models, we adopt an *Exponential*(0.10) distribution truncated below 2 for the parameter ν.

We implement the Bayesian approach using the full conditional distributions given in Propositions 13, 15 and 17, by means of a Metropolis-Hastings algorithm within Gibbs sampling. Furthermore, the convenient hierarchical representation of the SGN, SNT and STN models allows to use WinBUGS as an alternative for implementing the Bayesian approach. In both computational schemes, similar results were obtained.

For the full conditionals of the parameters λ and ν provided in Propositions 15 and 17, we consider a log-normal candidate centered at the logarithm of the previous sample and with variance allowing a rejection rate of 40% for the Metropolis step. As before, we run parallel chains with $50,000$ iterations, discarding the first $25,000$ as the burn-in and considering a lag of 10 iterations to avoid autocorrelation. The results are summarized in Table 3. In order to compare the models, we compute the deviance information criterion (DIC), see, e.g., Spiegelhalter, Best, Carlin and van der Linde (2002). In addition, we also adopt a cross validation criterion to compare the models, computing pseudo Bayes factors (PBF) (see, e.g., Geisser and Eddy, 1979; Gelfand and Dey, 1994) and using the log-marginal pseudo likelihood (LPML) (see, e.g., Ghosh and Gönen, 2008; Branscum and Hanson, 2008) through the conditional predictive ordinates (Chen,

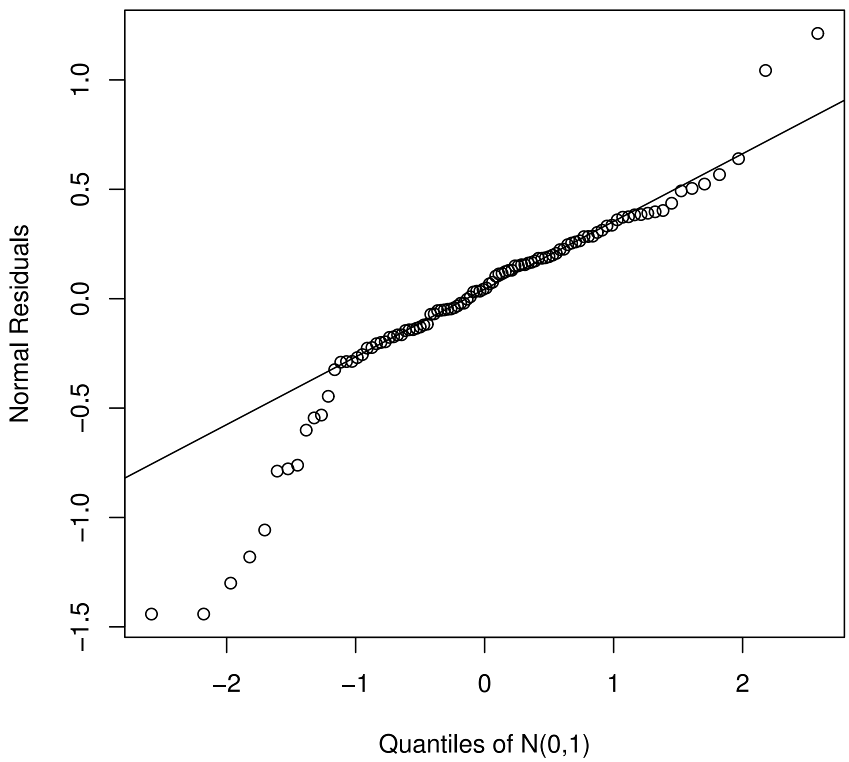

Figure 1: Normal QQ-plot for standardized residuals.

Shao and Ibrahim, 2000, pp. 307–315).

According to the DIC values reported in Table 3, we conclude that the SGN model is better than the other shape mixtures models for this dataset. Also, using the information provided by LPML, we compare the SGN model against SNT and STN computing PBF. The respective values of LPML are given in Table 3 and lead to a $2\log PBF$ of 59.64 (SGN model against SNT model) and 50.76 (SGN model against STN model) which are interpreted as strong evidence in favor of the SGN model. With respect to the SNT and STN models, note that, while the posterior estimates of the parameters α_1 and ν are quite similar in both models, the effect of these parameters on the skewness and kurtosis coefficients is very different; for more details see Nadarajah and Kotz (2003) and Gómez, Venegas and Bolfarine (2007).

Table 3: Bayesian estimates for the AIS dataset under various linear regression models: skew-generalized-normal (SGN), skew-normal-t (SNT), and skew-t-normal (STN).

SGN	β_1	β_2	σ	α_1	α_2	ν
Mean	0.05	0.81	1.14	−1.01	0.99	–
SD	0.01	0.03	0.16	0.41	0.32	–
$P_{2.5}$	0.03	0.76	0.86	−1.88	0.50	–
$P_{97.5}$	0.07	0.86	1.51	−0.26	1.75	–
DIC	−52.88					
$LPML$	−319.76					
SNT	β_1	β_2	σ	α_1	α_2	ν
Mean	0.05	0.81	0.86	−0.61	–	12.37
SD	0.01	0.02	0.13	0.33	–	10.32
$P_{2.5}$	0.03	0.76	0.67	−1.33	–	2.30
$P_{97.5}$	0.07	0.85	1.16	−0.04	–	38.9
DIC	145.16					
$LPML$	−349.58					
STN	β_1	β_2	σ	α_1	α_2	ν
Mean	0.05	0.81	0.72	−0.59	–	11.87
SD	0.01	0.02	0.07	0.31	–	9.90
$P_{2.5}$	0.03	0.76	0.60	−1.29	–	2.26
$P_{97.5}$	0.07	0.85	0.87	−0.06	–	38.29
DIC	177.79					
$LPML$	−345.14					

Acknowledgments

The authors thank a referee and the Associate Editor for constructive comments. They would like to express their gratitude to Pilar Iglesias (PUC - Chile) for her comments and suggestions on earlier versions of this paper. The research of R. B. Arellano-Valle was supported in part by FONDECYT (Chile), 1040865/7060133, 1085241 and 1030588. L. M. Castro thanks the Comisión Nacional de Ciencia y Tecnología - CONICYT for partially supporting his Ph.D. studies at the Pontificia Universidad Católica de Chile and the International Society for Bayesian Analysis for awarding him the Pilar Iglesias Travel Award to attend the ISBA World Meeting 2008. The work of M. G. Genton was partially supported by NSF grant DMS-0504896 and by a grant from the Swiss National Science Foundation. The research of H. W. Gómez was supported by FONDECYT (Chile) 1060727.

References

[1] Arellano-Valle, R. B. and Azzalini, A. (2006). "On the unification of families of skew-normal distributions." *Scandinavian Journal of Statistics*, 33: 561–574.

[2] Arellano-Valle, R. B., Bolfarine, H., and Lachos, V. H. (2007). "Bayesian inference for skew-normal linear mixed models." *Journal of Applied Statistics*, 34: 663–682.

[3] Arellano-Valle, R. B., Branco, M. D., and Genton, M. G. (2006). "A unified view on skewed distributions arising from selections." *The Canadian Journal of Statistics*, 34: 581–601.

[4] Arellano-Valle, R. B., del Pino, G., and Iglesias, P. (2006). "Bayesian inference in spherical linear models: robustness and conjugate analysis." *Journal of Multivariate Analysis*, 97: 179–197.

[5] Arellano-Valle, R. B., del Pino, G., and San Martín, E. (2002). "Definition and probabilistic properties of skew-distributions." *Statistics and Probability Letters*, 58: 111–121.

[6] Arellano-Valle, R. B. and Genton, M. G. (2005). "On fundamental skewed distributions." *Journal of Multivariate Analysis*, 96: 93–116.

[7] Arellano-Valle, R. B., Gómez, H. W., and Quintana, F. A. (2004). "A new class of skew-normal distributions." *Communications in Statistics, Theory and Methods*, 33: 1465–1480.

[8] Azzalini, A. (1985). "A class of distributions which includes the normal ones." *Scandinavian Journal of Statistics*, 12: 171–178.

[9] — (2005). "The skew-normal distribution and related multivariate families. With discussion by Marc G. Genton and a rejoinder by the author." *Scandinavian Journal of Statistics*, 32: 159–200.

[10] Azzalini, A. and Capitanio, A. (2003). "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution." *Journal of the Royal Statistical Society Series B*, 65: 367–389.

[11] Azzalini, A. and Dalla Valle, A. (1996). "The multivariate skew-normal distribution." *Biometrika*, 83: 715–726.

[12] Azzalini, A. and Genton, M. G. (2008). "Robust likelihood methods based on the skew-t and related distributions." *International Statistical Review*, 76: 106–129.

[13] Bernardo, J. and Smith, A. (2000). *Bayesian Theory*. New York: Wiley, second edition.

[14] Branscum, A. and Hanson, T. (2008). "Bayesian nonparametric meta-analysis using Polya tree mixture models." *Biometrics*, in press.

[15] Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.

[16] Cook, R. D. and Weisberg, S. (1994). *An Introduction to Regression Graphics*. New York: Wiley.

[17] Ellison, B. E. (1964). "Two theorems for inferences about the normal distribution with applications in acceptance sampling." *Journal of the American Statistical Association*, 59: 89–95.

[18] Geisser, S. and Eddy, W. (1979). "A predictive approach to model selection." *Journal of the American Statistical Association*, 74: 153–160.

[19] Gelfand, A. E. and Dey, D. K. (1994). "Bayesian model choice: asymptotics and exact calculations." *Journal of the Royal Statistical Society Series B*, 56: 501–514.

[20] Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, Florida: Chapman & Hall / CRC. Edited Volume.

[21] Genton, M. G. and Loperfido, N. (2005). "Generalized skew-elliptical distributions and their quadratic forms." *Annals of the Institute of Statistical Mathematics*, 57: 389–401.

[22] Ghosh, P., Branco, M. D., and Chakraborty, H. (2007). "Bivariate random effect model using skew normal distribution with application to HIV-RNA." *Statistics in Medicine*, 26: 1255–1267.

[23] Ghosh, P. and Gönen, M. (2008). "Bayesian modeling of multivariate average bioequivalence." *Statistics in Medicine*, 27: 2402–2419.

[24] Gómez, H. W., Venegas, O., and Bolfarine, H. (2007). "Skew-symmetric distributions generated by the distribution function of the normal distribution." *Environmetrics*, 18: 395–407.

[25] Ma, Y. and Genton, M. G. (2004). "A flexible class of skew-symmetric distributions." *Scandinavian Journal of Statistics*, 31: 459–468.

[26] Ma, Y., Genton, M. G., and Davidian, M. (2004). "Linear mixed effects models with flexible generalized skew-elliptical random effects." In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, 339–358. Boca Raton, Florida: Genton, M. G., Ed., Chapman & Hall / CRC.

[27] Nadarajah, S. and Kotz, S. (2003). "Skewed distributions generated by the normal kernel." *Statistics and Probability Letters*, 65: 269–277.

[28] Sahu, S. K., Dey, D. K., and Branco, M. D. (2003). "A new class of multivariate skew distributions with application to Bayesian regression models." *The Canadian Journal of Statistics*, 31: 129–150.

[29] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). "Bayesian measures of model complexity and fit (with discussion)." *Journal of the Royal Statistical Society Series B*, 64: 583–640.

[30] Wang, J., Boyer, J., and Genton, M. G. (2004). "A skew-symmetric representation of multivariate distributions." *Statistica Sinica*, 14: 1259–1270.