Bayesian inference for the skew-normal shape parameter: An application to change point problems

Luis M. Castro [a], Rosangela H. Loschi [b,1], and
Reinaldo B. Arellano-Valle [c]

[a] *Departamento de Estadística, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Casilla 160-C, Concepción, Chile.*

[b] *Departamento de Estatística, Universidade Federal de Minas Gerais, 31270-901 - Belo Horizonte - MG, Brazil*

[c] *Departamento de Estadítica, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile.*

Abstract

Bayesian Inference under the standard and the location-scale skew-normal distributions are presented assuming different prior specifications. Stochastic representations for the posteriors and posterior summaries are obtained in some particular cases. We apply the results to the one change point problem illustrating the results with the analysis of a simulated and some Latin American emerging market data sets.

Key words: Bayesian inference, change point analysis, Gibbs sampling, predictive and posterior distributions, truncated distribution.

---

1 Introduction

Normality can be a strong assumption for data sets arising from different areas of applications as, for instance, finance, environmental, medicine, and others. To cite an example, it is well known that, in general, the empirical distributions of stock market returns, mainly from emerging markets, are asymmetric and, in particular, they may have heavier tails than the normal distribution.

[1] Corresponding author, e-mail loschi@est.ufmg.br.

One of the greatest challenge related to data modeling is to find classes of distributions flexible enough to represent well many different data behavior such as skewness, heavy and light tails, kurtosis, etc. Recently, it is increasing the interest in classes of parametric distributions that are built by multiplying a symmetric density function (*pdf*) by a function that introduces skewness in the resulting *pdf*. This idea was firstly formalized by Azzalini (1985) who defined a class of univariate skew-normal distributions. Later, it was extended to the multivariate case by Azzalini and Dalla Valle (1996). Since then, many authors have tried to generalize these ideas by skewing arbitrary multivariate symmetric *pdf's* with very general forms of multiplicative functions see, for example, Azzalini and Capitanio (1999), Azzalini and Capitanio (2003), Branco and Dey (2001), González-Farías *et al.* (2004), Genton and Loperfido (2005), Arellano-Valle and Genton (2005) and Arellano-Valle and Azzalini (2006). An overview of these proposals can be found in the book edited by Genton (2004), in Azzalini (2005), and from a unified point of view in Arellano-Valle *et al.* (2006).

In this paper we consider that the likelihood is a member of the skew-normal class of distributions introduced by Azzalini (1985), that is, we assume that, given the location, scale and shape parameters, *i.e.,* $\xi \in \mathbb{R}$, $\omega^2 \in \mathbb{R}^+$ and $\lambda \in \mathbb{R}$, the random variable X has the skew-normal distribution denoted by $SN(\xi, \omega^2, \lambda)$ whose *pdf* is given by:

$$f(x|\xi, \omega^2, \lambda) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\lambda\left(\frac{x-\xi}{\omega}\right)\right), \quad x \in \mathbb{R}, \tag{1}$$

where ϕ and Φ denote the *pdf* and the cumulative distribution function (*cdf*) of the standard normal distribution, respectively. Note that, when $\xi = 0$ and $\omega = 1$ we obtain the standard skew-normal distribution, denoted by $SN(\lambda)$.

It is noticeable that the distribution in (1) is a particular case of the univariate skew-generalized normal (SGN) family introduced by Arellano-Valle *et al.* (2004). A random variable $X \sim SGN(\xi, \omega^2, \lambda_1, \lambda_2)$, $\xi \in \mathbb{R}$, $\omega^2 \in \mathbb{R}^+$, $\lambda_1 \in \mathbb{R}$, $\lambda_2 > 0$, if its *pdf* is

$$f(x|\xi, \omega^2, \lambda_1, \lambda_2) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\frac{\lambda_1(x-\xi)}{\sqrt{\omega^2+\lambda_2(x-\xi)^2}}\right), \quad x \in \mathbb{R}. \tag{2}$$

As a special case of such family we also have the normal one. Important properties of the SGN distributions are established by Arellano-Valle *et al.* (2004). Particularly, they establish that the SGN distribution can be represented as a shape mixture of the skew-normal distributions in (1) by taking a normal mixing distribution for the shape or skewness parameter. Shape mixtures of multivariate skew-normal distributions are also considered by Arellano-Valle *et al.* (2008) who assumed skew-normal mixing distributions for the shape parameters. From the Bayesian point of view the mixing measure is the prior distribution for λ. Consequently, Arellano-Valle *et al.* (2008) also provided important results for Bayesian inference on the shape parameter.

This paper addresses to Bayesian inference for skew-normal families. Firstly, we extend previous works by considering shape mixtures of the standard skew-normal distribution assuming as mixing measure a general prior. After that, we consider the problem of inference for the location, scale and shape parameters of the skew-normal distribution in (1). Predictive distributions and posteriors for the parameters are obtained, thus, generating classes of extended skewed distributions.

This paper also aims at providing exact posterior summaries for the shape parameter if the standard skew-normal family and particular priors for the shape parameter are under consideration. In these cases, stochastic representations for the posteriors of λ are also obtained.

It is noteworthy that the stochastic representations for the posteriors introduced here play an important role in both theoretical and practical issues. For instance, from such representations, posterior means and variances are obtained in more simple way. Also, by considering shape mixtures of skew-normal distributions, conjugate classes of distribution are obtained. However, the challenge related to their use is the computational effort that is needed in the inference process. It is noticeable that by using the stochastic representations we get a tractable way to approximate the posteriors.

In order to illustrate the results, they are applied to identify one change point in the shape parameter. We analyze two data sets: a simulated one and some Latin American emerging markets returns. It has been shown in the literature that emerging market returns experience changes in the volatility (measured through the variance) and also in their mean returns. However, it is well-known that such summaries are influenced by the asymmetry of the distribution. Thus, the identification of changes in the shape parameter is also useful to understand the phenomenon under study. Change point identification has been topic of research in many different areas. To cite a few, finance, genetics, reliability, medicine and many others. Non-Bayesian approaches for the change point problem can be found, for instance, in the papers by Dueker (1997), Hawkins (2001), Horváth and Kokoszka (2002), and Jandhyala *et al.* (2002). From the Bayesian point of view the change point problem has been considered by Barry and Hartigan (1992, 1993), Crowley (1997), Quintana and Iglesias (2003), Ruggeri and Sivaganesan (2005), Loschi and Cruz (2005) Fearnhead and Liu (2007) along with many others. A change point can be understood as the instant when a structural change is observed or simply as when an atypical return takes place. The identification of change points in financial time series analysis, for instance, become important since the rates of occurrence of atypical returns are taken into account in the evaluation of financial risks involved. In these time series, in general, a high rate of occurrence of atypical returns is common. Mainly, this phenomenon is observed in emerging markets data since these markets are more susceptible to shocks.

This paper is organized as follows. Bayesian Inference for the standard skew-normal distribution $SN(\lambda)$ and also for its location-scale extension, are presented in Sections 2 and 3, respectively. In Section 4 we consider the identification of one change point in the shape parameter. In Section 5 we apply the results to some real data sets. Section 6 closes the paper with some conclusions.

2 Bayesian inference in the standard skew-normal family

In this section, we consider shape mixtures of independent standard skew-normal distributions, say, we assume that, conditionally on λ, the random variables $X_1, \ldots, X_n \overset{iid}{\sim} SN(\lambda)$. Thus, for each observed sample $\mathbf{x} = (x_1, \ldots, x_n)^T$ of the random vector $\mathbf{X} = (X_1, \ldots, X_n)^T$, the likelihood function that is under consideration is given by:

$$L(\mathbf{x}|\lambda) = 2^n \phi_n(\mathbf{x}) \Phi_n(\lambda \mathbf{x}), \quad \lambda \in \mathbb{R}, \tag{3}$$

where $\phi_n(\mathbf{x}) = \prod_{i=1}^n \phi(x_i)$ and $\Phi_n(\lambda\mathbf{x}) = \prod_{i=1}^n \Phi(\lambda x_i)$.

To establish notation, we denote, respectively, by $\phi_n(\cdot; \boldsymbol{\mu}, \Sigma)$ $(\phi_n(\cdot))$ and $\Phi_n(\cdot;$ $\boldsymbol{\mu}, \Sigma)$ $(\Phi_n(\cdot))$ the *pdf* and *cdf* of the multivariate $N_n(\boldsymbol{\mu}, \Sigma)$ $(N_n(\mathbf{0}, I_n))$ distribution. We denote by $LTN_n(\mathbf{c}, \boldsymbol{\mu}, \Sigma)$ the distribution of a $N_n(\boldsymbol{\mu}, \Sigma)$ random vector with components truncated below $\mathbf{c} \in \mathbb{R}^n$, that is, the distribution of $\mathbf{U}^{\mathbf{c}} \overset{d}{=} \mathbf{U}|\mathbf{U} > \mathbf{c}$, where $\mathbf{U} \sim N_n(\boldsymbol{\mu}, \Sigma)$. Also, $E_\lambda(\cdot)$ denotes the expectation under the distribution of λ.

In order to obtain the posterior mean and variance of λ, for each fixed value $\mathbf{x}$ of $\mathbf{X}$, consider an unobserved random vector $\mathbf{Y}_x$ such that

$$\mathbf{Y}_x = \lambda \mathbf{x} + \boldsymbol{\epsilon}_*, \tag{4}$$

where $\boldsymbol{\epsilon}_* \sim N_n(\mathbf{0}, I_n)$ and it is independent of λ.

2.1 General prior for shape parameter

In the next proposition we consider an arbitrary mixing/prior distribution for the shape parameter λ, generalizing thus some similar results given in

Arellano-Valle *et al.* (2008). As by products, posterior inferences for λ can be obtained from these results.

Proposition 1 *Suppose that, given $\lambda \in \mathbb{R}$, the random variables $X_1, \ldots, X_n$ are iid with skew-normal distribution with parameter λ. Assume $\mathbf{Y}_x$ as defined in (4). If λ has a proper prior pdf $\pi(\cdot)$, then it follows that:*

i) the predictive pdf of $\mathbf{X}$ is given by:

$$\pi(\mathbf{x}) = 2^n \phi_n(\mathbf{x}) E_\lambda(\Phi_n(\lambda \mathbf{x})); \tag{5}$$

ii) the posterior pdf of λ is given by:

$$\pi(\lambda|\mathbf{x}) = \pi(\lambda) \frac{\Phi_n(\lambda \mathbf{x})}{E_\lambda(\Phi_n(\lambda \mathbf{x}))}; \tag{6}$$

iii) the posterior expectation of $g(\lambda)$ (if it exists) is given by:

$$E(g(\lambda)|\mathbf{x}) = E\left\{E(g(\lambda)|\mathbf{Y}_x)|\mathbf{Y}_x \geq 0\right\}, \tag{7}$$

where g is an integrable and measurable function.

Proof: Proofs of (i) and (ii) follow straightforward from some well known results of Probability Calculus and thus they are omitted. In order to prove the result in (7), firstly, notice from (4) that, for each fixed value $\mathbf{x}$, $\Phi_n(\lambda \mathbf{x}) = P(\mathbf{Y}_x \geq \mathbf{0}|\lambda) = \int_{\{\mathbf{y} \geq \mathbf{0}\}} \pi(\mathbf{w}|\lambda) d\mathbf{y}$ and $P(\mathbf{Y}_x \geq \mathbf{0}) = E_\lambda(\Phi_n(\lambda \mathbf{x}))$, where $\pi(\mathbf{y}|\lambda) = \phi_n(\mathbf{y}; \lambda \mathbf{x}, I_n)$ denotes the conditional pdf of $\mathbf{Y}_x$ given λ. Consequently, for any integrable and measurable function g, it follows from (6) that

$$\begin{aligned}
E(g(\lambda)|\mathbf{x}) &= (P(\mathbf{Y}_x \geq \mathbf{0}))^{-1} \int_{-\infty}^{\infty} g(\lambda)\pi(\lambda)\Phi_n(\lambda \mathbf{x}) d\lambda \\
&= (P(\mathbf{Y}_x \geq \mathbf{0}))^{-1} \int_{-\infty}^{\infty} g(\lambda)\pi(\lambda) \int_{\{\mathbf{y} \geq \mathbf{0}\}} \pi(\mathbf{y}|\lambda) d\mathbf{y} d\lambda \\
&= (P(\mathbf{Y}_x \geq \mathbf{0}))^{-1} \int_{\{\mathbf{y} \geq \mathbf{0}\}} \pi(\mathbf{y}) \int_{-\infty}^{\infty} g(\lambda)\pi(\lambda|\mathbf{y}) d\lambda d\mathbf{y} \\
&= \int_{\{\mathbf{y} \geq \mathbf{0}\}} \pi_0(\mathbf{y}) E(g(\lambda)|\mathbf{y}) d\mathbf{y}.
\end{aligned}$$

The proof is concluded by noticing that $\pi_0(\mathbf{y}) = \pi(\mathbf{y})(P(\mathbf{Y}_x \geq \mathbf{0}))^{-1}$ is the predictive distribution of vector $\mathbf{Y}_x$ with components truncated below $\mathbf{0}$.

Notice from Proposition 1 that the predictive distribution in (5) is a member of the fundamental skew-normal class introduced by Arellano-Valle and Genton (2005). It is also noticeable that the posterior of λ is a skewed-π type of distribution, with skewing factor $\frac{\Phi_n(\lambda \mathbf{x})}{E_\lambda(\Phi_n(\lambda \mathbf{x}))}$. In fact, the prior and the posterior of λ are in the same family of distributions, say, $\mathcal{P} = \{\pi(\lambda|\psi) = K\pi(\lambda)\Phi(\lambda\psi) :$

$\psi \in \mathcal{A}\}$ in which $K^{-1} = \int_{-\infty}^{\infty} \pi(\lambda)\Phi(\lambda\psi)d\lambda$ and $\mathcal{A}$ denotes the set of labels of the distributions. Thus, $\mathcal{P}$ and the family in (1) are conjugate (see more in Arellano-Valle *et al.* (2008)).

2.2 Skew-normal prior for shape parameter

Let us assume that the shape parameter follows a location-scale skew-normal prior distribution denoted by $\lambda \sim SN(m, v, a)$, where the location, scale and shape parameters are, respectively, $m \in \mathbb{R}$, $v \in \mathbb{R}^+$ and $a \in \mathbb{R}$ and the pdf is given by:

$$\pi(\lambda) = \frac{2}{\sqrt{v}}\phi\left(\frac{\lambda - m}{\sqrt{v}}\right)\Phi\left(\frac{a(\lambda - m)}{\sqrt{v}}\right). \tag{8}$$

Notice that this family of distributions is rich enough in forms to represent the prior opinion in several circumstances. For instance, it includes the $N(m, v)$ prior, in case we elicit $a = 0$, and also the $SN(a)$ prior whenever we declare $m = 0$ and $v = 1$.

The main contributions here are the stochastic representation for the posterior of λ, and also the posterior mean and variance of such distribution.

Along this paper, let $\mathbf{x}^0 = (\mathbf{x}^T\ 0)^T$, $\mathbf{x}^a = \left(\mathbf{x}^T\ \frac{a}{\sqrt{v}}\right)^T$ and $\mathbf{m}^* = m\left(\mathbf{0}^T\ \frac{a}{\sqrt{v}}\right)^T$.

Proposition 2 *Suppose that, given $\lambda \in \mathbb{R}$, the random variables $X_1, \ldots, X_n$ are iid with $SN(\lambda)$. If λ has the location-scale skew-normal prior pdf in (8), it follows that:*

i) the predictive pdf of $\mathbf{X}$ is

$$\pi(\mathbf{x}) = 2^n\phi_n(\mathbf{x})\Phi_{n+1}(m\mathbf{x}^0; \mathbf{0}, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT}); \tag{9}$$

ii) the posterior pdf λ is the following skew-normal pdf:

$$\pi(\lambda|\mathbf{x}) = \phi(\lambda; m, v)\frac{\Phi_{n+1}\left(\lambda\mathbf{x}^a; \mathbf{m}^*, I_{n+1}\right)}{\Phi_{n+1}(m\mathbf{x}^0; \mathbf{0}, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})}. \tag{10}$$

Proof: Note that if $\lambda \sim SN(m, v, a)$, from (4) we have that $\mathbf{Y}_\mathbf{x} \sim SN_n(m(\mathbf{x}), v(\mathbf{x}), a(\mathbf{x}))$, where $m(\mathbf{x}) = m\mathbf{x}$, $v(\mathbf{x}) = I_n + v\mathbf{x}\mathbf{x}^T$ and $a(\mathbf{x}) = av^{-1/2}[(1 + a^2v\mathbf{x}^T\mathbf{x})(1 + v\mathbf{x}^T\mathbf{x})]^{-1/2}\mathbf{x}$. Thus, the results in Proposition 2 follow straightforward from Proposition 1, since we know that if $\mathbf{Z} \sim SN_k(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$, then by applying Lemma it follows that $P(\mathbf{Z} \geq \mathbf{0}) = 2\Phi_{k+1}(\mathbf{a}; \Omega)$, where $\mathbf{a} = (\boldsymbol{\mu}^T, \boldsymbol{\alpha}^T\boldsymbol{\mu}/\sqrt{1 + \boldsymbol{\alpha}^T\Sigma\boldsymbol{\alpha}})^T$ and $\Omega = (\Omega_{ij})$, with $\Omega_{11} = 1$, $\Omega_{12} = \Omega_{21}^T = (1 + \boldsymbol{\alpha}^T\Sigma\boldsymbol{\alpha})^{-1}\Sigma$, and $\Omega_{22} = \Sigma$.

An interesting particular case is shown in the following corollary where a normal mixing measure is assumed as prior for λ.

Corollary 3 *Suppose that, given $\lambda \in \mathbb{R}$, the random variables $X_1, \ldots, X_n$ are iid with $SN(\lambda)$. If $\lambda \sim N(m, v)$ in which $m \in \mathbb{R}$ and $v \in \mathbb{R}^+$, it follows that:*

i) the predictive pdf of $\mathbf{X}$ *is given by:*

$$f(\mathbf{x}) = 2^n \phi_n(\mathbf{x}) \Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^T); \tag{11}$$

ii) the posterior pdf of λ is the following skew-normal pdf:

$$\pi(\lambda|\mathbf{x}) = \phi(\lambda; m, v) \frac{\Phi_n(\lambda\mathbf{x})}{\Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^T)}. \tag{12}$$

The proof of Corollary 3 follows straightforward from Proposition 2 by taking $a = 0$.

It is noteworthy (see Proposition 2 and Corollary 3) that the posterior of λ as well as the predictive distribution are members of the unified skew-normal (SUN) class introduced by Arellano-Valle and Azzalini (2006).

In order to obtain the posterior mean and variance of λ, the next two results will be considered. The one established in Lemma 4 gives a general result to compute the mean vector and the variance-covariance matrix of a truncated normal random vector $\mathbf{U}^\mathbf{c} \stackrel{d}{=} \mathbf{U}|\mathbf{U} \geq \mathbf{c}$, where $\mathbf{U} \sim N_n(\boldsymbol{\mu}, \Sigma)$. Its proof can be found in Appendix A.

Lemma 4 *Consider a random vector $\mathbf{U}^\mathbf{c} \stackrel{d}{=} \mathbf{U}|\mathbf{U} \geq \mathbf{c}$, where $\mathbf{U} \sim N_n(\boldsymbol{\mu}, \Sigma)$, that is, we assume that $\mathbf{U}^\mathbf{c} \sim LTN_n(\mathbf{c}, \boldsymbol{\mu}, \Sigma)$. Then, the expected vector and the variance-covariance matrix of $\mathbf{U}^\mathbf{c}$ are given, respectively, by:*

$$\begin{aligned}
E(\mathbf{U}^\mathbf{c}) &= \boldsymbol{\mu} + \Sigma \frac{\Phi_n'(\boldsymbol{\mu}; \mathbf{c}, \Sigma)}{\Phi_n(\boldsymbol{\mu}; \mathbf{c}, \Sigma)}, \\
V(\mathbf{U}^\mathbf{c}) &= \Sigma + \Sigma \left\{ \frac{\Phi_n''(\boldsymbol{\mu}; \mathbf{c}, \Sigma)}{\Phi_n(\boldsymbol{\mu}; \mathbf{c}, \Sigma)} - \left(\frac{\Phi_n'(\boldsymbol{\mu}; \mathbf{c}, \Sigma)}{\Phi_n(\boldsymbol{\mu}; \mathbf{c}, \Sigma)} \right) \left(\frac{\Phi_n'(\boldsymbol{\mu}; \mathbf{c}, \Sigma)}{\Phi_n(\boldsymbol{\mu}; \mathbf{c}, \Sigma)} \right)^T \right\} \Sigma,
\end{aligned}$$

where $\Phi_n'(\mathbf{s}; \boldsymbol{\mu}, \Sigma) = \dfrac{\partial}{\partial \mathbf{s}} \Phi_n(\mathbf{s}; \boldsymbol{\mu}, \Sigma)$ *and* $\Phi_n''(\mathbf{s}; \boldsymbol{\mu}, \Sigma) = \dfrac{\partial^2}{\partial \mathbf{s} \partial \mathbf{s}^T} \Phi_n(\mathbf{s}; \boldsymbol{\mu}, \Sigma)$.

Proposition 5 provides the stochastic representation for the posterior skewness random variable $\lambda|\mathbf{x}$, which is useful to obtain the posterior moments of λ and also in the algorithm to identify change points introduced in Section 4.2. For simplicity, throughout this paper, $\boldsymbol{V}_\mathbf{x}$ denotes the conditional random vector $\boldsymbol{V}|\mathbf{X} = \mathbf{x}$ and $\lambda_\mathbf{x}$ is the posterior skewness random variable $\lambda|\mathbf{x}$.

Proposition 5 *Let $W \in \mathbb{R}$ and $\mathbf{U^0} \in \mathbb{R}^{n+1}$ be two random variables that, conditionally in $\mathbf{X} = \mathbf{x}$, are independent. Assume that $W_{\mathbf{x}} \sim N(m, v(1+v\mathbf{x}^{aT}\mathbf{x}^a))$ and $\mathbf{U^0_x} \sim LTN_{n+1}(\mathbf{0}, m\mathbf{x}^0, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})$. Then, the posterior skewness random variable $\lambda_{\mathbf{x}}$ which distribution is given in (10) is such that*

$$\lambda_{\mathbf{x}} \stackrel{d}{=} \frac{W_{\mathbf{x}} + v\mathbf{x}^{aT}\mathbf{U^0_x}}{1 + v\mathbf{x}^{aT}\mathbf{x}^a}. \tag{13}$$

Proof: Considering the moment generating function (*mgf*) approach, we need to show that $M_{\lambda_{\mathbf{x}}}(t) = M_{W_{\mathbf{x}}}\left(\frac{t}{1+v\mathbf{x}^{aT}\mathbf{x}}\right) M_{\mathbf{U^0_x}}\left(\frac{v\mathbf{x}t}{1+v\mathbf{x}^{aT}\mathbf{x}}\right)$, for all $t \in \mathbb{R}$. The *mgf* of $\lambda_{\mathbf{x}}$ is given by:

$$M_{\lambda_{\mathbf{x}}}(t) = \exp\left\{mt + \frac{1}{2}vt^2\right\} \frac{\Phi_n(m\mathbf{x}^0 + vt\mathbf{x}^a; \mathbf{0}, I_n + v\mathbf{x}^a\mathbf{x}^{aT})}{\Phi_n(m\mathbf{x}^0; \mathbf{0}, I_n + v\mathbf{x}^a\mathbf{x}^{aT})},$$

(see Appendix B for the proof). On the other hand, from equation (26) in Appendix A, we obtain that

$$\begin{aligned} M_{\mathbf{U^0_x}}\left(\frac{v\mathbf{x}^a t}{1 + v\mathbf{x}^{aT}\mathbf{x}}\right) = \exp&\left\{tm\mathbf{x}^0 + \frac{1}{2}\frac{v^2\mathbf{x}^{aT}\mathbf{x}^a t^2}{1 + v\mathbf{x}^{aT}\mathbf{x}^a}\right\} \\ &\frac{\Phi_n(m\mathbf{x}^0 + v\mathbf{x}^a t; \mathbf{0}, I_n + v\mathbf{x}^a\mathbf{x}^{aT})}{\Phi_n(m\mathbf{x}^0; \mathbf{0}, I_n + v\mathbf{x}^a\mathbf{x}^{aT})}. \end{aligned}$$

The proof follows straightforward after some algebraic calculations.

Corollary 6 *Let $W \in \mathbb{R}$ and $\mathbf{U} \in \mathbb{R}^n$ be two random quantities that, conditionally on $\mathbf{X}$, are independent. Assume that $W_{\mathbf{x}} \sim N(m, v(1 + v\mathbf{x}^T\mathbf{x}))$ and that $\mathbf{U^0_x} \sim LTN_n(\mathbf{0}, m\mathbf{x}, I_n + v\mathbf{x}\mathbf{x}^T)$. Then, the posterior random variable λ_x which distribution is given in (12) is such that:*

$$\lambda_{\mathbf{x}} \stackrel{d}{=} \frac{W_{\mathbf{x}} + v\mathbf{x}^T\mathbf{U^0_x}}{1 + v\mathbf{x}^T\mathbf{x}}. \tag{14}$$

Proof: The proof follows straightforward from Proposition 5 whenever we assume $a = 0$.

Alternatively, the representations shown previously in Proposition 5 and Corollary 6 can be proved considering the SUN stochastic representation introduced by Arellano-Valle and Azzalini (2006). Next proposition provides an explicit expressions for the posterior mean and variance of skewness parameter λ.

Proposition 7 *Suppose that, given $\lambda \in \mathbb{R}$, $X_1, \ldots, X_n$ are iid $SN(\lambda)$ random variables. If λ has the skew-normal prior distribution given in (8), it follows that:*

i) the posterior mean of λ is given by:

$$E(\lambda|\mathbf{x}) = m + v\mathbf{x}^{aT}\boldsymbol{\xi}, \tag{15}$$

where the ith component of vector $\boldsymbol{\xi}$ is

$$\phi(mx_i^0; 0, 1 + vx_i^{a2})\frac{\Phi_n\left(m\mathbf{x}_{(i)}^0; \frac{mvx_i^0 x_i^a \mathbf{x}_{(i)}^a}{1+vx_i^{a2}}, I_n + \frac{v\mathbf{x}_{(i)}^a \mathbf{x}_{(i)}^{a\,T}}{1+vx_i^{a2}}\right)}{\Phi_{n+1}(m\mathbf{x}^0; 0, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})},$$

$\mathbf{x}_{(i)}^0$ and $\mathbf{x}_{(i)}^a$ are, respectively, the subvectors of $\mathbf{x}^0$ and $\mathbf{x}^a$ without ith component;

ii) the posterior variance of λ is given by:

$$Var(\lambda|\mathbf{x}) = v + v^2\mathbf{x}^{aT}(\Gamma - \boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{x}^a, \tag{16}$$

where the element γ_{ij} of the matrix Γ is

$$\gamma_{ij} = \begin{cases} \frac{\phi(mx_i^0;0,1+vx_i^{a2})}{\Phi_{n+1}(m\mathbf{x}^0;0,I_{n+1}+v\mathbf{x}^a\mathbf{x}^{aT})} \times \\ \qquad \left[\left(\frac{-mx_i^0}{1+vx_i^{a2}}\right)\Phi_n\left(m\mathbf{x}_{(i)}^0; \frac{mvx_i^0 x_i^a \mathbf{x}_{(i)}^a}{1+vx_i^{a2}}, I_n + \frac{v\mathbf{x}_{(i)}^a \mathbf{x}_{(i)}^{a\,T}}{1+vx_i^{a2}}\right)\right. \\ \qquad \left. + \frac{\partial}{\partial mx_i^0}\Phi_n\left(m\mathbf{x}_{(i)}^0; \frac{mvx_i^0 x_i^a \mathbf{x}_{(i)}^a}{1+vx_i^{a2}}, I_n + \frac{v\mathbf{x}_{(i)}^a \mathbf{x}_{(i)}^{a\,T}}{1+vx_i^{a2}}\right)\right], & i = j \\ \frac{\phi(mx_i^0;0,1+vx_i^{a2})}{\Phi_{n+1}(m\mathbf{x}^0;0,I_{n+1}+v\mathbf{x}^a\mathbf{x}^{aT})}\phi\left(mx_j^0; \frac{mvx_i^a x_j^a x_i^0}{1+vx_i^{a2}}, \frac{1+v(x_i^{a2}+x_j^{a2})}{1+vx_i^{a2}}\right) \\ \quad \times \Phi_{n-2}\left(m\mathbf{x}_{(ij)}^0; \frac{mv\mathbf{x}_{(ij)}^a(x_i^a x_i^0 + x_j^a x_j^0)}{1+v(x_i^{a2}+x_j^{a2})}, I_{n-1} + \frac{v\mathbf{x}_{(ij)}^a \mathbf{x}_{(ij)}^{a\,T}}{1+v(x_i^{a2}+x_j^{a2})}\right), & i \neq j, \end{cases}$$

$\mathbf{x}_{(ij)}^0$ and $\mathbf{x}_{(ij)}^a$ are, respectively, the subvectors of $\mathbf{x}^0$ and $\mathbf{x}^a$ without both the ith and jth components.

Proof: (i) From expression (13) and assuming independence, it follows that:

$$E(\lambda|\mathbf{x}) = \frac{E(W_{\mathbf{x}}) + v\mathbf{x}^{aT}E(\mathbf{U}_{\mathbf{x}}^{\mathbf{0}})}{1 + v\mathbf{x}^{aT}\mathbf{x}^a}.$$

As a consequence of Lemma 4, the expectation of $\mathbf{U}_{\mathbf{x}}^{\mathbf{0}}$ is given by:

$$E(\mathbf{U}_{\mathbf{x}}^{\mathbf{0}}) = (I_n + v\mathbf{x}^a\mathbf{x}^{aT})\frac{\Phi_n'(m\mathbf{x}^0; \mathbf{0}, I_n + v\mathbf{x}^a\mathbf{x}^{aT})}{\Phi(m\mathbf{x}^0; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^T)}.$$

The proof follows by applying Theorem 14 given in Appendix C.

(ii) Similarly, we have that the posterior variance of λ is:

$$Var(\lambda|\mathbf{x}) = \frac{Var(W_{\mathbf{x}}) + v^2\mathbf{x}^{aT}V(\mathbf{U}_{\mathbf{x}}^{\mathbf{0}})\mathbf{x}^a}{(1 + v\mathbf{x}^{aT}\mathbf{x}^a)^2}.$$

The proof is a consequence of Theorem 14 (Appendix C) and Lemma 4 which provides $Var(\mathbf{U}_{\mathbf{x}}^{\mathbf{0}})$.

Corollary 8 *Suppose that, given $\lambda \in \mathbb{R}$, the random variables $X_1, \ldots, X_n$ are iid $SN(\lambda)$. If $\lambda \sim N(m, v)$ where $m \in \mathbb{R}$ and $v \in \mathbb{R}^+$, it follows that:*

i) the posterior mean of λ is given by:

$$E(\lambda|\mathbf{x}) = m + v\mathbf{x}^T\boldsymbol{\xi}, \tag{17}$$

where the ith component of $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^T$ is:

$$\xi_i = \phi(mx_i; 0, 1 + vx_i^2)\frac{\Phi_{n-1}\left(\frac{m\mathbf{x}_{(i)}}{1+vx_i^2}; 0, I_{n-1} + \frac{v\mathbf{x}_{(i)}\mathbf{x}_{(i)}^T}{1+vx_i^2}\right)}{\Phi_n(m\mathbf{x}; 0, I_n + v\mathbf{x}\mathbf{x}^T)};$$

ii) the posterior variance of λ is given by:

$$Var(\lambda|\mathbf{x}) = v + v^2\mathbf{x}^T(\Gamma - \boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{x}, \tag{18}$$

where $\Gamma = (\gamma_{ij})$ is the $n \times n$ matrix whose entries are:

$$\gamma_{ij} = \begin{cases} \frac{\phi(mx_i;0,1+vx_i^2)}{\Phi_n(m\mathbf{x};\mathbf{0},I_n+v\mathbf{x}\mathbf{x}^T)}\left[\left(\frac{-mx_i}{1+vx_i^2}\right)\Phi_{n-1}\left(\frac{m\mathbf{x}_{(i)}}{1+vx_i^2}; \mathbf{0}, I_{n-1} + \frac{v\mathbf{x}_{(i)}\mathbf{x}_{(i)}^T}{1+vx_i^2}\right)\right. \\ \qquad \left. + \frac{\partial}{\partial mx_i}\Phi_{n-1}\left(\frac{m\mathbf{x}_{(i)}}{1+vx_i^2}; \mathbf{0}, I_{n-1} + \frac{v\mathbf{x}_{(i)}\mathbf{x}_{(i)}^T}{1+vx_i^2}\right)\right], \ i = j \\ \frac{\phi(mx_i;0,1+vx_i^2)}{\Phi_n(m\mathbf{x};\mathbf{0},I_n+v\mathbf{x}\mathbf{x}^T)}\phi\left(\frac{mx_j}{1+vx_i^2}; 0, \frac{1+v(x_i^2+x_j^2)}{1+vx_i^2}\right) \\ \qquad \times \Phi_{n-2}\left(\frac{m\mathbf{x}_{(ij)}}{1+v(x_i^2+x_j^2)}; \mathbf{0}, I_{n-2} + \frac{v\mathbf{x}_{(ij)}\mathbf{x}_{(ij)}^T}{1+v(x_i^2+x_j^2)}\right), \ i \neq j. \end{cases}$$

Remark: Notice that by applying (7), Proposition 1, and the results in Lemma 4, another interesting way to obtain the posterior mean and variance of λ is provided. From (4) it is obtained that $\mathbf{Y}_x|\lambda \sim N_n(\lambda\mathbf{x}, I_n)$. Consequently, assuming that $\lambda \sim N(m, v)$, it follows that

$$\lambda|\mathbf{y} \sim N\left(\frac{m + v\mathbf{x}^T\mathbf{y}}{1 + v\mathbf{x}^T\mathbf{x}}, \frac{v}{1 + v\mathbf{x}^T\mathbf{x}}\right).$$

Hence, from (7), the posterior mean of λ is given by:

$$E(\lambda|\mathbf{x}) = E\{E(\lambda|\mathbf{Y}_x)|\mathbf{Y}_x \geq \mathbf{0}\} = \frac{m + v\mathbf{x}^T E(\mathbf{Y}_x^{\mathbf{0}})}{1 + v\mathbf{x}^T\mathbf{x}},$$

where the truncated expectation $E(\mathbf{Y}_x^{\mathbf{0}}) = E(\mathbf{Y}_x|\mathbf{Y}_x \geq \mathbf{0})$ is provided by

Lemma 4. Similarly, we calculate $E(\lambda^2|\mathbf{x})$ and, thus, the posterior variance of λ is obtained.

3 Bayesian inference in the location-scale skew-normal family

In this section, we consider shape mixtures of independent location-scale skew-normal distributions, say, we assume that, conditionally on ξ, ω^2 and λ, the random variables $X_1, \ldots, X_n \stackrel{iid}{\sim} SN(\xi, \omega^2, \lambda)$. Thus, for each observed sample $\mathbf{x} = (x_1, \ldots, x_n)^T$ of the random vector $\mathbf{X} = (X_1, \ldots, X_n)^T$, the likelihood function that is under consideration is given by:

$$L(\mathbf{x}|\xi, \omega^2, \lambda) = \left(\frac{2}{\omega}\right)^n \phi_n\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega}\right) \Phi_n\left(\lambda\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega}\right)\right). \tag{19}$$

We provide the posterior kernels for the location, scale and shape parameters assuming that, a prior, ξ and ω^2 are independent from λ. We also consider a improper prior for (ξ, ω^2) and different priors for λ.

Along this paper, denote the Euclidian norm of vector $\mathbf{a}$ by $||\mathbf{a}||$ and let $t_n(\cdot; \mathbf{a}, \mathbf{B}, \mathbf{c})$ and $T_n(\cdot; \mathbf{a}, \mathbf{B}, \mathbf{c})$ be the *pdf* and *cdf* of an n-variate student-t distribution with location vector $\mathbf{a} \in \mathbb{R}^n$, scale matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ and degree of freedom $c \in \mathbb{R}^+$.

3.1 General prior for shape parameter

The analysis in this section take into consideration an arbitrary proper prior $\pi(\lambda)$ for λ.

Proposition 9 *Suppose that, given $\xi \in \mathbb{R}$, $\omega^2 \in \mathbb{R}^+$ and $\lambda \in \mathbb{R}$, the random variables $X_1, \ldots, X_n$ are iid with $SN(\xi, \omega^2, \lambda)$. If ξ and ω^2 has the joint improper prior $\pi(\xi, \omega^2) = \omega^{-2}$ and λ has a proper prior $\pi(\lambda)$, then*

i) the joint posterior pdf of (ξ, ω^2) is

$$\pi(\xi, \omega^2|\mathbf{x}) \propto \frac{1}{\omega^{n+2}} \phi_n\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega}\right) E_\lambda\left[\Phi_n\left(\lambda\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega}\right)\right)\right];$$

ii) the posterior pdf of λ is

$$\pi(\lambda|\mathbf{x}) \propto \pi(\lambda) T_n\left(\lambda(\mathbf{x} - \bar{x}\mathbf{1}_n); \mathbf{0}, \left(\frac{||\mathbf{x}||^2 + n\bar{x}^2}{n-1}\right)[I_n + n^{-1}\lambda^2 \mathbf{1}_n \mathbf{1}_n^T], n-1\right).$$

Proof: (i) The proof follows straightforward from (19) and by considering some well-known results of probability calculus from which it can be noticed that the joint distribution of $(\mathbf{x}, \xi, \omega^2, \lambda)$ is given by:

$$\pi(\xi, \omega^2, \lambda, \mathbf{x}) = \frac{2^n}{\omega^{n+2}} \phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega} \right) \Phi_n \left(\lambda \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega} \right) \right) \pi(\lambda). \tag{20}$$

(ii) It follows from (20), Lemma 12 and some results of probability calculus that the posterior *pdf* of λ is given by:

$\pi(\lambda|\mathbf{x}) \propto \pi(\lambda) \int_{\omega^2 \in \mathbb{R}^+} \int_{\xi \in \mathbb{R}} \frac{1}{\omega^{n+2}} \exp\left\{ \frac{||\mathbf{x} - \xi \mathbf{1}_n||^2}{2\omega^2} \right\} \Phi_n \left(\lambda \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega} \right) \right) d\xi d\omega^2$

$$\begin{aligned}
&= \pi(\lambda) \int_{\omega^2 \in \mathbb{R}^+} \frac{1}{\omega^{n+1}} \exp\left\{ \frac{||\mathbf{x}||^2 - n\bar{x}}{2\omega^2} \right\} \int_{\xi \in \mathbb{R}} \frac{\sqrt{n}}{\omega} \phi \left(\frac{\xi - \bar{x}}{\omega/\sqrt{n}} \right) \Phi_n \left(\lambda \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega} \right) \right) d\xi d\omega^2 \\
&= \pi(\lambda) \int_{\omega^2 \in \mathbb{R}^+} \frac{1}{\omega^{n+1}} \exp\left\{ \frac{||\mathbf{x}||^2 - n\bar{x}}{2\omega^2} \right\} \Phi_n \left(\lambda(\mathbf{x} - \bar{x}\mathbf{1}_n); 0, \omega^2 [I_n + n^{-1}\lambda^2 \mathbf{1}_n \mathbf{1}_n^T] \right) d\omega^2 \\
&= \pi(\lambda) \int_{\mathbf{t} \leq \lambda(\mathbf{x} - \bar{x}\mathbf{1}_n)} \int_{\omega^2 \in \mathbb{R}^+} \frac{1}{\omega^{n+1}} \exp\left\{ \frac{||\mathbf{x}||^2 - n\bar{x}}{2\omega^2} \right\} \phi_n \left(\mathbf{t}; 0, \omega^2 [I_n + n^{-1}\lambda^2 \mathbf{1}_n \mathbf{1}_n^T] \right) d\mathbf{t} d\omega^2 \\
&= \pi(\lambda) \int_{\mathbf{t} \leq \lambda(\mathbf{x} - \bar{x}\mathbf{1}_n)} t_n \left(\mathbf{t}; 0, \left(\frac{||\mathbf{x}||^2 + n\bar{x}^2}{n-1} \right) [I_n + n^{-1}\lambda^2 \mathbf{1}_n \mathbf{1}_n^T], n-1 \right) d\mathbf{t},
\end{aligned}$$

which concludes the proof.

3.2 Skew-normal prior for shape parameter

Let us assume particular priors for the shape parameter. We start eliciting a location-scale skew-normal prior for λ, say, we assume that $\lambda \sim SN(m, v, a)$ which density is given in (8).

Proposition 10 *Suppose that, given $\xi \in \mathbb{R}$, $\omega^2 \in \mathbb{R}^+$ and $\lambda \in \mathbb{R}$, the random variables $X_1, \ldots, X_n$ are iid with $SN(\xi, \omega^2, \lambda)$. If ξ and ω^2 has the improper prior $\pi(\xi, \omega^2) = \omega^{-2}$ and $\lambda \sim SN(m, v, a)$, where $m \in \mathbb{R}$, $v \in \mathbb{R}^+$ and $a \in \mathbb{R}$, then*

i) the joint posterior pdf of (ξ, ω^2) is such that

$$\pi(\xi, \omega^2|\mathbf{x}) \propto \frac{1}{\omega^{n+2}} \phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega} \right) \Phi_{n+1} \left(m\mathbf{x}^a_{\xi,\omega}; \mathbf{m}^*, I_{n+1} + v\mathbf{x}^a_{\xi,\omega} \mathbf{x}^{a\,T}_{\xi,\omega} \right)$$

where $\mathbf{x}^a_{\xi,\omega} = \left(\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega} \right)^T \; \frac{a}{\sqrt{v}} \right)^T$;

ii) the posterior pdf of λ *is*

$$\pi(\lambda|\mathbf{x}) \propto \phi\left(\frac{\lambda - m}{v}\right) \Phi_n\left(\frac{a(\lambda - m)}{v}\right)$$
$$T_n\left(\lambda(\mathbf{x} - \bar{x}\mathbf{1}_n); \mathbf{0}, \left(\frac{||\mathbf{x}||^2 + n\bar{x}}{n-1}\right)[I_n + n^{-1}\lambda^2\mathbf{1}_n\mathbf{1}_n^T], n-1\right).$$

Proof: Proofs of (i) and (ii) follow straightforward from Proposition 9.

Corollary 11 *Suppose that, given* $\xi \in \mathbb{R}$, $\omega^2 \in \mathbb{R}^+$ *and* $\lambda \in \mathbb{R}$, *the random variables* $X_1, \ldots, X_n$ *are iid with* $SN(\xi, \omega^2, \lambda)$. *If* ξ *and* ω^2 *has the improper prior* $\pi(\xi, \omega^2) = \omega^{-2}$ *and* $\lambda \sim N(m, v)$, *then*

i) the joint posterior pdf of (ξ, ω^2) *is given by:*

$$\pi(\xi, \omega^2|\mathbf{x}) \propto \frac{1}{\omega^{n+2}}\phi_n\left(\frac{\mathbf{x} - \xi\mathbf{1}_n}{\omega}\right)$$
$$\Phi_{n+1}\left(\frac{m(\mathbf{x} - \xi\mathbf{1}_n)}{\omega}; I_n + v\left(\frac{\mathbf{x} - \xi\mathbf{1}_n}{\omega}\right)\left(\frac{\mathbf{x} - \xi\mathbf{1}_n}{\omega}\right)^T\right);$$

ii) the posterior pdf of λ *is*

$$\pi(\lambda|\mathbf{x}) \propto \phi\left(\frac{\lambda - m}{v}\right)$$
$$T_n\left(\lambda(\mathbf{x} - \bar{x}\mathbf{1}_n); \mathbf{0}, \left(\frac{||\mathbf{x}||^2 + n\bar{x}}{n-1}\right)[I_n + n^{-1}\lambda^2\mathbf{1}_n\mathbf{1}_n^T], n-1\right).$$

Proof: The proofs of (i) and (ii) are obtained from Proposition 10 by setting $a = 0$.

4 Change point in the shape parameter

The results introduced previously are considered here to identify change points in the skewness or shape parameter whenever data are sequentially observed. The interest here is focused on the one change point problem.

4.1 Model specification

Denote by K the random instant when a change takes place and by λ_1 and λ_2, the shape parameters before and after the change point, respectively. We assume that, given ξ, ω^2, λ_1, λ_2 and $K = k$, the random vectors $\mathbf{X}_{(k)} =$

$(X_1, \ldots, X_k)$ and $\mathbf{X}_{(n-k)} = (X_{k+1}, \ldots, X_n)$ are independent and that, for $k = 1, \ldots, n$, it follows that:

$$\begin{aligned}\mathbf{X}_{(k)}|\xi, \omega^2, \lambda_1, \lambda_2, k &\overset{ind.}{\sim} SN(\xi, \omega^2, \lambda_1);\\ \mathbf{X}_{(n-k)}|\xi, \omega^2, \lambda_1, \lambda_2, k &\overset{ind.}{\sim} SN(\xi, \omega^2, \lambda_2).\end{aligned}$$

Consequently, the likelihood that is under consideration is:

$$L(\mathbf{x}|\xi, \omega^2, \lambda_1, \lambda_2, K = k) = \left(\frac{2}{\omega}\right)^n \phi_n(\mathbf{z}_{(n)})\Phi_{(k)}(\lambda_1\mathbf{z}_{(k)})\Phi_{(n-k)}(\lambda_2\mathbf{z}_{(n-k)}), \quad (21)$$

where $\mathbf{z}_{(l)} = (\mathbf{x}_{(l)} - \xi\mathbf{1}_l)/\omega$.

In order to construct the prior distribution of $(\xi, \omega^2, \lambda_1, \lambda_2, K)$, let us assume that K is independent of ξ, ω^2, λ_1 and λ_2. Also, given $K = k$, assume that the vector of parameters (ξ, ω^2) is independent of the shape parameters λ_1 and λ_2 and that λ_1 is independent of λ_2. Consider that, a prior, $\pi(\xi, \omega^2) = w^{-2}$ and denote by $\pi(k)$, $\pi(\lambda_1)$ and $\pi(\lambda_2)$, the prior specifications for K, λ_1 and λ_2, respectively.

As a consequence of such specifications, it follows that the predictive distribution for $\mathbf{x}$ is given by:

$$\pi(\mathbf{x}) = 2^n \sum_{k=1}^{n} \pi(k) \int_{-\infty}^{\infty} \int_{0}^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) H_{\lambda_1,\lambda_2}(\mathbf{z}) d\omega^2 d\xi,$$

where $H_{\lambda_1,\lambda_2}(\mathbf{z}) = E_{\lambda_1}\left(\Phi_k(\lambda_1\mathbf{z}_{(k)})\right) E_{\lambda_2}\left(\Phi_{n-k}(\lambda_2\mathbf{z}_{(n-k)})\right)$ and the expectation $E_{\lambda_i}(\cdot)$ is computed under the prior of λ_i, $i = 1, 2$. A straightforward application of Bayes's theorem provides that the posteriors of K, λ_1 and (ξ, ω^2) are given, respectively, by:

$$\pi(K = k|\mathbf{x}) = \frac{\pi(k) \int_{-\infty}^{\infty} \int_0^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) H_{\lambda_1,\lambda_2}(\mathbf{z}) d\omega^2 d\xi}{\sum_{k=1}^{n} \pi(k) \int_{-\infty}^{\infty} \int_0^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) H_{\lambda_1,\lambda_2}(\mathbf{z}) d\omega^2 d\xi}, \quad (22)$$

$$\pi(\lambda_1|\mathbf{x}) = \frac{\pi(\lambda_1) \sum_{k=1}^{n} \pi(k) \int_{-\infty}^{\infty} \int_0^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) G_{\lambda_2}(\mathbf{z}, \lambda_1) d\omega^2 d\xi}{\sum_{k=1}^{n} \pi(k) \int_{-\infty}^{\infty} \int_0^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) H_{\lambda_1,\lambda_2}(\mathbf{z}) d\omega^2 d\xi}, \quad (23)$$

$$\pi(\xi, \omega^2|\mathbf{x}) = \frac{\omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) \sum_{k=1}^{n} \left[\pi(k) H_{\lambda_1,\lambda_2}(\mathbf{z})\right]}{\sum_{k=1}^{n} \pi(k) \int_{-\infty}^{\infty} \int_0^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) H_{\lambda_1,\lambda_2}(\mathbf{z}) d\omega^2 d\xi}, \quad (24)$$

where $G_{\lambda_2}(\mathbf{z}, \lambda_1) = \Phi_k(\lambda_1\mathbf{z}_{(k)}) E_{\lambda_2}\left(\Phi_{n-k}(\lambda_2\mathbf{z}_{(n-k)})\right)$. Moreover, let $\mathbf{Y}_{\mathbf{z}_{(k)}} =$

$\lambda_1 \mathbf{z}_{(k)} + \epsilon_*$ and $\epsilon_* \sim N_k(\mathbf{0}, I_k)$. Then, from Proposition 1 it follows that:

$$E_{\lambda_1}\left(\lambda_1 \frac{\Phi_k(\lambda_1 \mathbf{z}_{(k)})}{E_{\lambda_1}\left(\Phi_k(\lambda_1 \mathbf{z}_{(k)})\right)}\right) = E\left\{E(\lambda_1 | \mathbf{Y}_{\mathbf{z}_{(k)}}) | \mathbf{Y}_{\mathbf{z}_{(k)}} \geq 0\right\} = E(\lambda_1 | \mathbf{z}_{(k)}),$$

Consequently, the posterior expectation of shape parameter λ_1 is given by:

$$E(\lambda_1|\mathbf{x}) = \frac{\sum_{k=1}^{n} \pi(k) \int_{-\infty}^{\infty} \int_0^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_n) E(\lambda_1 | \mathbf{z}_{(k)}) H_{\lambda_1,\lambda_2}(\mathbf{z}) d\omega^2 d\xi}{\sum_{k=1}^{n} \int_{-\infty}^{\infty} \int_0^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) \pi(k) H_{\lambda_1,\lambda_2}(\mathbf{z}) d\omega^2 d\xi}. \quad (25)$$

The posterior distribution of λ_2 and its expectation can be obtained similarly.

Let us assume, as a particular case, that $\lambda_i \sim \mathcal{N}(m_i, v_i)$, for $i = 1, 2$, that will be considered in the applications in next section. In this case, the posteriors and the posterior expectation of λ_1 are obtained from (22) to (25) by noticing that:

$$\begin{aligned}
E_{\lambda_1}\left[\Phi_k(\lambda_1 \mathbf{z}_{(k)})\right] &= \Phi_k(m_1 \mathbf{z}_{(k)}; \mathbf{0}, I_k + v \mathbf{z}_{(k)} \mathbf{z}_{(k)}^T), \\
E_{\lambda_2}\left[\Phi_{n-k}(\lambda_2 \mathbf{z}_{(n-k)})\right] &= \Phi_{n-k}(m_2 \mathbf{z}_{(n-k)}; \mathbf{0}, I_{n-k} + v \mathbf{z}_{(n-k)} \mathbf{z}_{(n-k)}^T), \\
E(\lambda_1 | \mathbf{z}_k) &= \frac{m_1 + v_1 \mathbf{z}_{(k)}^T E(\mathbf{Y}_x^{\mathbf{0}})}{1 + v \mathbf{z}_{(k)}^T \mathbf{z}_{(k)}},
\end{aligned}$$

where $E(\mathbf{Y}_{z_k}^{\mathbf{0}}) = E(\mathbf{Y}_{z_k} | \mathbf{Y}_{z_k} \geq \mathbf{0})$ is provided in Lemma 4.

4.2 Gibbs sampling scheme

Notice from (22), (23) and (24) that is computationally taxing to obtain the posteriors. In this paper, we consider a Metropolis-within-Gibbs algorithm to approximate such distributions. Notice that for general priors for the shape parameters the full conditionals reduce analytically to the following distributions:

$$\begin{aligned}
&\pi(K = k | \xi, \omega^2, \lambda_1, \lambda_2, \mathbf{x}) \propto \pi(k) \Phi_k(\lambda_1 \mathbf{z}_{(k)}) \Phi_{n-k}(\lambda_2 \mathbf{z}_{(n-k)}) \\
&\pi(\lambda_1 | \xi, \omega^2, \lambda_2, K = k, \mathbf{x}) \propto \pi(\lambda_1) \Phi_k(\lambda_1 \mathbf{z}_{(k)}), \\
&\pi(\lambda_2 | \xi, \omega^2, \lambda_1, K = k, \mathbf{x}) \propto \pi(\lambda_2) \Phi_{n-k}(\lambda_2 \mathbf{z}_{(n-k)}), \\
&\pi(\xi | \omega^2, \lambda_1, \lambda_2, K = k, \mathbf{x}) \propto \phi_n(\mathbf{z}_{(n)}) \Phi_k(\lambda_1 \mathbf{z}_{(k)}) \Phi_{n-k}(\lambda_2 \mathbf{z}_{(n-k)}), \\
&\pi(\omega^2 | \xi, \lambda_1, \lambda_2, K = k, \mathbf{x}) \propto \omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) \Phi_k(\lambda_1 \mathbf{z}_{(k)}) \Phi_{n-k}(\lambda_2 \mathbf{z}_{(n-k)}).
\end{aligned}$$

Particularly, eliciting skew-normal priors for the shape parameters, that is, considering $\lambda_i \sim SN(m_i, v_i, a_i)$, $i = 1, 2$, the full conditionals become:

$$
\begin{aligned}
&\pi(\lambda_1|\xi, \omega^2, \lambda_2, K = k, \mathbf{x}) \propto \phi(\lambda_1, m_1, v_1)\Phi_{k+1}(\lambda_1 \mathbf{z}^*_{(k)}), \\
&\pi(\lambda_2|\xi, \omega^2, \lambda_1, K = k, \mathbf{x}) \propto \phi(\lambda_2, m_2, v_2)\Phi_{n-k+1}(\lambda_2 \mathbf{z}^*_{(n-k)}), \\
&\pi(\xi|\omega^2, \lambda_1, \lambda_2, K = k, \mathbf{x}) \propto \phi\left(\frac{\xi - \bar{x}}{\omega/\sqrt{n}}\right)\Phi_n(\xi\lambda^* + \mathbf{x}^*), \\
&\pi(\omega^2|\xi, \lambda_1, \lambda_2, K = k, \mathbf{x}) \propto IG\left(\frac{n}{2}, \frac{||\mathbf{x} - \xi\mathbf{1}_n||^2}{2}\right)\Phi_k(\lambda_1 \mathbf{z}_{(k)})\Phi_{n-k}(\lambda_2 \mathbf{z}_{(n-k)}),
\end{aligned}
$$

where $IG(a/2, b/2)$ denotes an Inverted-gamma density with parameters $a > 0$ and $b > 0$, $\mathbf{z}^*_{(k)} = (z_1, \ldots, z_k, \frac{a_1}{\sqrt{v_1}})$, $\mathbf{z}^*_{(n-k)} = (z_{k+1}, \ldots, z_n, \frac{a_2}{\sqrt{v_2}})$, $\lambda^* = \left(-\frac{\lambda_1 \mathbf{1}_k^T}{\omega} \; -\frac{\lambda_2 \mathbf{1}_{n-k}^T}{\omega}\right)^T$ and $\mathbf{x}^* = \left(-\frac{\lambda_1}{\omega}\mathbf{x}_{(k)}^T \; -\frac{\lambda_2}{\omega}\mathbf{x}_{(n-k)}^T\right)^T$. The full conditionals for normal priors are obtained assuming $a_1 = a_2 = 0$.

It can be noticed that, in these particular cases, the full conditionals of λ_1, λ_2 and ξ are skew-normal distributions. Even thought, it is not easy to sample directly from such distributions. We consider the stochastic representations provided in previous sections to performing efficiently such a task. Samples from the full conditional of ω^2 are obtained using a Metropolis-Hastings step considering a Gamma as the proposal distribution. Finally, the approach suggested by Carlin *et al.* (1992) is used to sample from the full conditional of K .

5 Applications

The ultimate goal in this section is to identify one change point in the return of the stock prices indexes of the four most important Latin American markets. These time series was previously analyzed by Loschi *et al.* (2005) that considered the product partition model (Barry and Hartigan , 1992) to identify multiple change points in the mean returns and volatilities of such markets. Loschi *et al.* (2005) assume that the return series were normally distributed and concluded that all these markets experienced changes in both, the mean return and the volatility, in similar periods.

Normality can be such a strong assumption for emerging markets stock returns since the empirical distributions of them often exhibits skewness and tails that are lighter or heavier than normal distribution. Moreover, it is well-known that the mean and the variance can be influenced by the asymmetry of the distribution. Consequently, changes in the skewness parameter can lead to changes in such parameters.

In the following, we assume a location-scale skew-normal likelihood to describe the data behavior. Thus, the results introduced in the previous sections are used. We focus our attention in the identification of one change point in the skewness parameter of such distribution extending previous works.

In order to evaluate the efficiency of our approach, we start by canalizing a generated data set.

In both cases, we assume the uniform prior for K and the joint improper prior for (ξ, ω^2) considered in Section 3.

For the MCMC scheme, we generated samples of size $20,000$ and, after convergence has been reached, we discarded the initial $5,000$ iterations as the burn-in. To avoid correlation, a lag of 10 was considered.

5.1 Analysis for simulated data set

The data set we analyze here consists of a sequence of size $n = 120$ of skew-normal observations, which experiences a change in the skewness parameter at position 40 (see Figure 1). It is generated from the following distributions:

$$\begin{aligned} X_1, \ldots, X_{40} \mid \xi, \omega^2, \lambda_1, \lambda_2, K &\overset{iid}{\sim} SN(5, 2.5^2, -2), \\ X_{41}, \ldots, X_{120} \mid \xi, \omega^2, \lambda_1, \lambda_2, K &\overset{iid}{\sim} SN(5, 2.5^2, 2). \end{aligned}$$

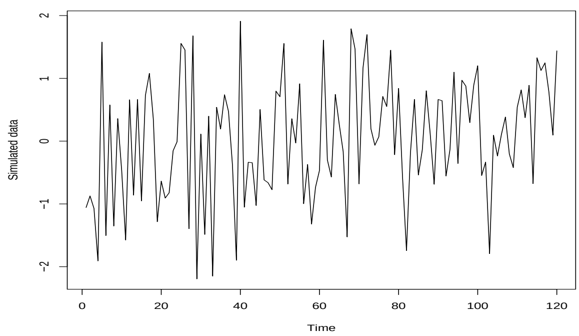

Fig. 1. Simulated skew-normal data

As prior specifications, we also assume that λ_1 and λ_2 are identically distributed with normal distribution with mean equal to 0 and variance equal to 10. The posterior results for both models are summarized in Figure 2 and Table 1

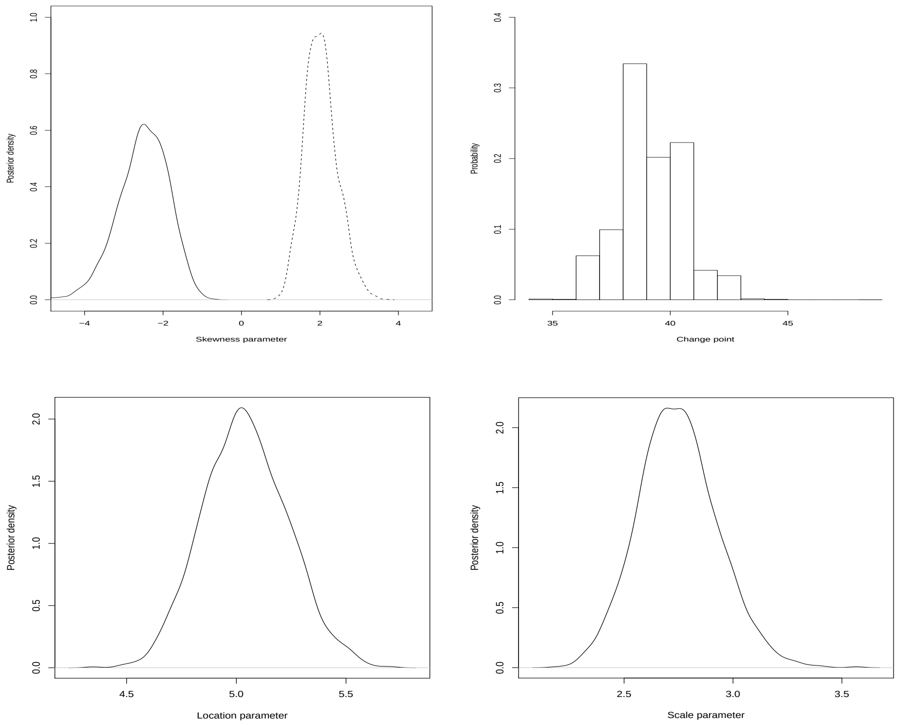

Fig. 2. Posteriors for λ_1 (solid line), λ_2 (dotted line), K, ξ and ω

Table 1
Posterior summaries.

Param.	Mean	St.Dev.	$P_{2.5}$	$P_{97.5}$
ξ	5.04	0.19	4.67	5.43
ω	2.75	0.18	2.41	3.12
λ_1	-2.51	0.64	-3.89	-1.38
λ_2	2.03	0.42	1.26	2.93
K	39.68	1.43	37.0	43.0

It can be noticed from Figure 2 that the posteriors for λ_1 and λ_2 are asymmetric. If compared with the posterior of λ_1, the posterior of λ_2 tends to put most of its mass in high values. The posterior of K points out that the probability of instant 40 being a change point is close to 20.0% but, most probably, the change took place early. It can also be observed that the posterior mean of K (39.68) is very close to the true change point. For the other parameters, we notice that, although the posterior means are close to the true parameters, they are overestimated by them. A possible explanation for this issue is the asymmetry of the posteriors. As expected, we also notice that the uncertainty about the parameters become small in the posterior evaluation. It is noticeable that all the posteriors present smaller variances than we observed a prior.

5.2 *Analysis for Latin American emerging indexes*

It is well known that emerging markets are more susceptible to the political scenario than developed markets thus their indexes tend to present more atypical observations as well as structural changes.

In this section, we consider the return series of the main stock indexes of four Latin American markets, say, the MERVAL (*Índice de Mercado de Valores de Buenos Aires*) of Argentina, the IBOVESPA (*Índice da Bolsa de Valores do Estado de São Paulo*) of Brazil, the IPSA (*Índice de Precios Selectivos de Acciones*) of Chile and the IPyC (*Índice de Precios y Cotizaciones*) of Mexico. The stock returns are recorded weekly from October 31, 1995 to October 31, 2000.

The data sets consist of the return series that are defined using the transformation $R_t = (P_t - P_{t-1})/P_{t-1}$ where P_t is the last price observed in the tth week. We assume that these series have a skew-normal distributions with unknown location, scale and skewness parameters.

We also assume, as prior specification, that λ_1 and λ_2 are identically distributed such that $\lambda_i \sim N(m, 10)$. The parameter m is different for each return series and it is equal to the cumulant estimator of the shape parameter of the return series (see Castro *et al.* (2007)), say, $m = -1.51$, -2.71, -1.05 and -1.41 for MERVAL, IBOVESPA, IPSA and IPyC, respectively.

According to some expert's opinion, changes in the behavior of stock returns are mainly consequence of crises or events that occur in other countries. Great financial crises involving emerging markets occurred in January, 1995 (Mexico's Crisis), August, 1997 (Asia's Crisis) and July, 1998 (Russia's Crisis). Also, in January, 1999, the Minas Gerais (Brazil) State Governor stopped paying Minas Gerais's debt with other countries and Argentina's crisis started around December, 1999. These important events are country specific. However, they can spread out across countries with similar economy eventually producing changes in their behavior (Lopes and Migon , 2002). The posterior means of K (see Table 2) point out that the changes in the skewness parameter of MERVAL, IBOVESPA, IPSA and IPyC occured, respectively, in the 2nd week, January, 1997; the 4th week, July, 1998; the 4th week, July 1997 and the 2nd week, December, 1997. Notice that all indexes, except MERVAl, have experienced a change in periods close to some international crisis. However, it is noticeable from Figure 3 that the probablitity of each week being a change point is smaller than 10.0% $(2,5\%)$, for IBOVESPA and IPSA (for MERVAL and IPyC), which means that the evidence in favour of a change point is not strong.

It is noteworthy that IBOVESPA presents the highest mean return (0.05) as

Table 2
Posterior summaries for each index.

	Parameter				
	ξ	ω	λ_1	λ_2	K
MERVAL					
Mean	0.01	0.06	−0.36	−0.41	106.76
St.Dev.	0.03	0.01	1.65	0.85	47.44
$P_{2.5}$	−0.03	0.05	−2.71	−1.62	4
$P_{97.5}$	0.05	0.07	2.30	0.89	228
IBOVESPA					
Mean	0.05	0.07	−3.17	−0.95	128.16
St.Dev.	0.01	0.01	0.51	0.21	12.75
$P_{2.5}$	0.04	0.06	−4.19	−1.41	80
$P_{97.5}$	0.06	0.08	−2.24	−0.61	137
IPSA					
Mean	−0.02	0.04	2.74	0.73	89.81
St.Dev.	0.01	0.01	0.54	0.17	7.19
$P_{2.5}$	−0.03	0.03	1.72	0.41	74
$P_{97.5}$	−0.01	0.05	3.81	1.08	103
IPyC					
Mean	0.01	0.06	0.17	−0.19	102.07
St.Dev.	0.04	0.01	1.66	0.96	57.44
$P_{2.5}$	−0.04	0.05	−2.11	−1.74	5
$P_{97.5}$	0.05	0.08	2.75	1.11	237

well as the highest volatility (0.07) and that IPSA has the lowest mean return (-0.02) and volatility (0.04). Also notice that the volatility and mean return for MERVAL and IPyC are equal and that IPSA is the only index that presents a negative mean return.

Taking into consideration the asymmetry before and after the change point, we perceive that they are both negative for MERVAL and IBOVESPA and, for IPSA, they are both positive. Moreover, for IBOVESPA and IPSA, we notice that the asymmetry of the returns is small after the change point and

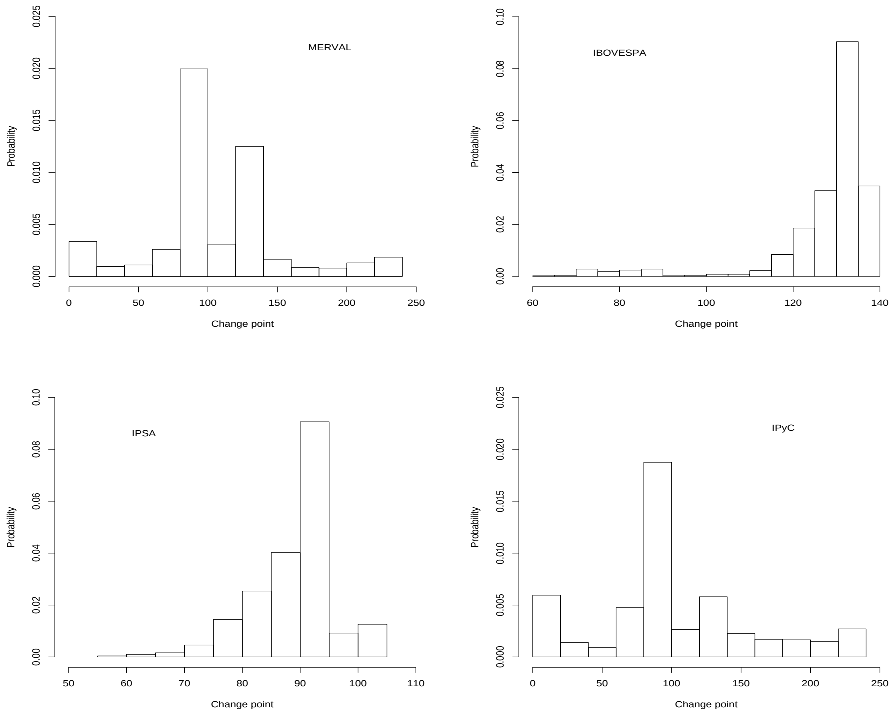

Fig. 3. Posteriors for the change point

the diference between the posterior means for the skewness parameter before and after the change is high for both indexes. The opposite is observed for MERVAL and IPyC. In fact, for MERVAL and IPyC, the posterior means for the skewness parameter before and after the change point are close, that is, there is not a strong evidence of skewness in such indexes.

6 Summary and conclusions

In this paper we dealt with the problem of Bayesian inference under standard and location-scale skew-normal family. We considered a joint improper prior for the location and scale parameter and, for the shape parameter, a general, a skew-normal and a normal priors were assumed. Under the standard skew-normal family, we obtained exact posteriors for the shape parameter, and assuming particular priors, stochastic representations for them were obtained. In these cases, by considering such representations, we were able to find the posterior means and variances for the shape parameter. We applied such results to identify one change point in the skewness of emerging market return series.

We noticed that those stochastic representations also played a fundamental role in practical issues. Since posteriors of the shape parameter are skewed distributions, the challenge related to its use in practical situations is the computational effort that is needed in the inference process, particularly, in change point identification. The stochastic representations provided a tractable way to approximate the posteriors in that case. In the application, we obtained a weak evidence in favour of a change in the skewness parameter of all Latin American indexes. We perceived that IBOVESPA and IPSA has similar behaviour regarding to the asymmetry before and after the change point being the skewness parameter negative for IBOVESPA and positive for IPSA. We also perceive that MERVAL and IPyC present a quite similar behaviour being the returns of such indexes almost symmetric.

Interesting topic for future research in the area is to apply the results in multiple change point identification.

Acknowledgments

The authors would like to express their gratitude to Pilar Iglesias (PUC - Chile) for her comments and suggestions in earlier versions of this paper. L. M. Castro thanks to the Comisión Nacional de Ciencia y Tecnología - CONICYT for partially supporting his Ph.D. studies at the Pontificia Universidad Católica de Chile. The research of R. B. Arellano-Valle was supported in part by FONDECYT (Chile), grants 1085241, 1040865, 7060133 and 7070137. The work of R.H. Loschi was partially supported by CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil, grants 3004505/2006-4, 472066/2004-8 and 472877/2006-2.

References

Arellano-Valle, R.B., del Pino, G., San Martín, E. (2002) Definition and probabilistic properties of skew-distributions. *Statist. Probab. Lett.*, **58**, 111–121.

Arellano-Valle, R.B., Gómez, H., Quintana, F. (2004) A new class of skew-normal distributions. *Communs. Statist. Theory Meth.*, **33**, 1465–1480.

Arellano-Valle, R.B., Genton, M.G. (2005) On fundamental skew distributions. *J. Multivariate Anal.*, **96**, 93–116.

Arellano-Valle, R.B., Azzalini, A. (2006) On the unification of families of skew-normal distributions. *Scand. J. Statist.*, **33**, 561–574.

Arellano-Valle, R.B., Branco, M.D., Genton, M.G. (2006) A unified view on skewed distributions arising from selections. *Canadian J. Statist.*, **34**, 581–601.

Arellano-Valle, R. B, Genton, M. G. and Loschi, R. H. (2008) Shape mixtures of multivariate skew-normal distributions. *J. Multivariate Anal.*, doi: 10.1016/j.jmva.2008.03.009.

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171–178.

Azzalini, A. (2005) The skew-normal distribution and related multivariate families. With discussion by Marc G. Genton and a rejoinder by the author. *Scand. J. Statist.*, **32**, 159–200.

Azzalini, A., Dalla Valle, A. (1996) The multivariate skew-normal distribution. *Biometrika*, **83**, 715–726.

Azzalini, A., Capitanio, A. (1999) Distributions generated by perturbation of symmetry with emphasis on multivariate skew t distribution. *J. R. Statist. Soc. B*, **65**, 367–389.

Azzalini, A., Capitanio, A. (1999) Statistical applications of the multivariate skew-normal distribution. *J. R. Statist. Soc. B*, **61**, 579–602.

Barry, D., Hartigan, J.A. (1992) Product partition models for change point problems. *Ann. of Statist.*, **20**(1), 260–279.

Barry, D., Hartigan, J.A. (1993) A Bayesian analysis for change point problem. *J. Am. Statist. Assoc.*, **88**(421), 309–319.

Branco, M.D., Dey, D.K. (2001) A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.*, **79**, 93–113.

Carlin, B.P., Gelfand, A.E., Smith, A.F.M. (1992) Hierarchical Bayesian analysis of changepoint problems. *Appl. Statist.*, **41**(2), 389–405.

Crowley, E.M. (1997) Product partition models for normal means. *J. Am. Statist. Assoc.*, **92**(437), 192–198.

Castro, M., San Martín, E., Arellano-Valle, R. (2007) On the statistical meaning of parameters of the skew-normal experiments. *Technical Report.* Pontificia Universidad Catolica de Chile.

Domínguez-Molina, A., Gonzalez-Farías, G., Gupta,A. (2003) The multivariate closed-skew normal distribution. *Technical Report N 03-12*, Department of Mathematics and Statistics, Bowling Green State University.

Domínguez-Molina, A., Gonzalez-Farías, G., Ramos-Quiroga,R. (2004) Skew-normality in stochastic frontier analysis. In *Skew-elliptical distributions and their applications: A journey beyond normality*, M. G. Genton (ed.), Chapman & Hall / CRC, Boca Raton, FL, 416 pp.

Dueker, M.J. (1997) Markov switching in GARCH processes and mean-reverting stock-market volatility. *J. Bus. and Econ. Statist.*, **15**(1), 26–34.

Fearnhead, P. and Liu, Z. (2007) On-line inference for multiple changepoint problems, *J. R. Statist. Soc. B*, **69**, 589–605.

Genton, M.G. (2004) *Skew-elliptical distributions and their applications: A journey beyond normality.* Edited Volume, Chapman & Hall / CRC, Boca Raton, FL, 416 pp.

Genton, M.G., Loperfido, N. (2005) Generalized skew-elliptical distributions and their quadratic forms. *Ann. Inst. Statist. Math.*, **57**, 389–401.

Gonzalez-Farías, G., Domínguez-Molina, A., Gupta, A. K. (2004) The closed skew-normal distribution. In *Skew-Elliptical Distributions and Their Appli-*

cations: A Journey Beyond Normality, Genton, M. G. (Ed.), Chapman & Hall / CRC, Boca Raton, FL, pp. 25−42.

Hawkins, D.M. (2001) Fitting multiple change-point models to data. *Comp. Statist. Data Anal.*, **37**(3), 323–341.

Horváth, L., Kokoszka P. (2002) Change-point detection with non-parametric regression. *Statistics*, **36**(1), 9–31.

Jandhyala, V.K., Fotopoulos S.B., Hawkins DM. (2002) Detection and estimation of abrupt changes in the variability of a process. *Comp. Statist. Data Anal.*, **40**(1), 1–19.

Lopes, H. F. and H. S. Migon. (2002) Comovements and Contagion in Emergin Markets: Stock Indexes Volatilities *Case Studies on Bayesian Statistics VI*, Eds. Gatsonis, C. et al, Lecture Notes in Statistics Series 167.

Loschi, R.H., and Cruz, F.R.B. (2005) Extension to the product partition model: Computing the probability of a change, *Comp. Statist. Data Anal.*, **48**(2) 255–268.

Loschi, R.H., Cruz, F.R.B., Arellano-Valle, R.B., 2005. Multiple change point analysis for the regular exponential family using the product partition model. *Journal of Data Science*, **3**(3), 305–330.

Quintana, F., Iglesias, P. (2003) Nonparametric Bayesian clustering and product partition models, *J. R. Statist. Soc. B* **65**, 557–574.

Ruggeri, F. and Sivaganesan, S. (2005) On modeling change points in non-homogeneous Poisson processes, *Statist. Infer. Stoch. Proc.*, **8**, 311–329.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc. Ser. B*, **64**, 583–640.

Tallis, G.M. (1961) The moment generating function of the truncated muiltinormal distribution. *J. R. Statist. Soc. B*, **23**(1), 223-229.

Appendix A

In this appendix we prove Lemma 4. Assume that $\mathbf{U^c} \sim LTN_n(\mathbf{c}, \boldsymbol{\mu}, \mathbf{\Sigma})$. From some results presented in Domínguez-Molina *et al.* (2003, 2004) (see also Tallis (1961)) it follows that the mgf of $\mathbf{U^c}$ is given by:

$$M_{\mathbf{U^c}}(\mathbf{t}) = e^{\mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}} \frac{\Phi_n(\boldsymbol{\mu} + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)}{\Phi_n(\boldsymbol{\mu}; \mathbf{c}, \Sigma)}. \tag{26}$$

Let us consider the cumulant function $K_{\mathbf{U^c}}(\mathbf{t}) = \log M_{\mathbf{U^c}}(\mathbf{t})$ and let $K'_{\mathbf{U^c}}(\mathbf{t}) = \dfrac{\partial}{\partial \mathbf{t}} K_{\mathbf{U^c}}(\mathbf{t})$ and $K''_{\mathbf{U^c}}(\mathbf{t}) = \dfrac{\partial}{\partial \mathbf{t}^T} K'_{\mathbf{U^c}}(\mathbf{t})$. From (26), it follows that

$$K'_{\mathbf{U^c}}(\mathbf{t}) = \boldsymbol{\mu} + \Sigma \mathbf{t} + \mathbf{\Sigma} \frac{\Phi'_n(\boldsymbol{\mu} + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)}{\Phi_n(\boldsymbol{\mu} + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)} \tag{27}$$

and

$$\begin{aligned}
K''_{\mathbf{U^c}}(\mathbf{t}) &= \Sigma + \Sigma \frac{\partial}{\partial \mathbf{t}^T} \frac{\Phi'_n(\boldsymbol{\mu} + \Sigma\mathbf{t}; \mathbf{c}, \Sigma)}{\Phi_n(\boldsymbol{\mu} + \Sigma\mathbf{t}; \mathbf{c}, \Sigma)} \\
&= \Sigma + \Sigma \left\{ \frac{\Phi''_n(\boldsymbol{\mu} + \Sigma\mathbf{t}; \mathbf{c}, \Sigma)}{\Phi_n(\boldsymbol{\mu} + \Sigma\mathbf{t}; \mathbf{c}, \Sigma)} \Sigma \right. \\
&\quad \left. - \frac{\Phi'_n(\boldsymbol{\mu} + \Sigma\mathbf{t}; \mathbf{c}, \Sigma)\Phi'_n(\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{t}; \mathbf{c}, \Sigma)^T}{[\Phi_n(\boldsymbol{\mu} + \Sigma\mathbf{t}; \mathbf{c}, \Sigma)]^2} \Sigma \right\}. \qquad (28)
\end{aligned}$$

Since $E(\mathbf{U^c}) = K'_{\mathbf{U^c}}(\mathbf{0})$ and $Var(\mathbf{U^c}) = K''_{\mathbf{U^c}}(\mathbf{0})$, the proof follows by letting $\mathbf{t} = \mathbf{0}$ in expressions (27) and (28).

Appendix B

In this appendix, we obtain the mgf for the posterior skewness parameter $\lambda_{\mathbf{x}} = \lambda|\mathbf{x}$. We consider the following well known Lemma which can also be used to show most of the result considered in this paper.

Lemma 12 *If* $\mathbf{U} \sim N_n(\mathbf{c}, \mathbf{C})$ *is a non-singular, normal random vector, then for any fixed* $m \times 1$ *vector* $\mathbf{a}$ *and* $m \times n$ *matrix* $\mathbf{A}$, *we have*

$$E[\Phi_m(\mathbf{AU} + \mathbf{a}; \mathbf{b}, B)] = \Phi_m(\mathbf{Ac} + \mathbf{a}; \mathbf{b}, B + \mathbf{ACA}^T).$$

The mgf of $\lambda_{\mathbf{x}}$ assuming a skew-normal prior, $\lambda \sim \mathcal{SN}(m, v, a)$, is given by.

$$\begin{aligned}
M_{\lambda_{\mathbf{x}}}(t) &= \int_{-\infty}^{\infty} e^{\lambda t} \phi(\lambda; m, v) \frac{\Phi_{n+1}(\lambda\mathbf{x}^a; \mathbf{m}^*, I_{n+1})}{\Phi_{n+1}(m\mathbf{x}^0; \mathbf{0}, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})} d\lambda \\
&= e^{tm + \frac{t^2 v}{2}} \int_{-\infty}^{\infty} \phi(\lambda; m + tv, v) \frac{\Phi_{n+1}(\lambda\mathbf{x}^a; \mathbf{m}^*, I_{n+1})}{\Phi_{n+1}(m\mathbf{x}^0; \mathbf{0}, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})} d\lambda \\
&= e^{tm + \frac{t^2 v}{2}} \frac{E[\Phi_{n+1}(\lambda\mathbf{x}^a; \mathbf{m}^*, I_{n+1})]}{\Phi_{n+1}(m\mathbf{x}^0; \mathbf{0}, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})}.
\end{aligned}$$

Applying again Lemma 12 it follows that:

$$M_{\lambda_{\mathbf{x}}}(t) = e^{tm + \frac{t^2 v}{2}} \frac{\Phi_{n+1}(m\mathbf{x}^0 + tv\mathbf{x}^a; \mathbf{0}, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})}{\Phi_{n+1}(m\mathbf{x}^0; \mathbf{0}, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})}.$$

Appendix C

In this appendix, we present a theorem that is an important tool to calculate the mean vector and the variance-covariance matrix of truncated multivariate normal distributions. The following preliminary lemma is considered in its proof.

Denote by $\mathbf{s}_{(i)}$ and $\mathbf{s}_{(ij)}$ the sub vectors of $\mathbf{s} = (s_1, \ldots, s_n)^T$ without the components s_i and (s_i, s_j), respectively. Let f and F be the pdf and the cdf of a random vector $\mathbf{U} \in \mathbb{R}^n$. For simplicity, denote by $f(s_i)$ $(F(s_i))$, the i-th marginal pdf (cdf) of f (F) and by $f(s_j|s_i)$ $(F(s_j|s_i))$, the j-th conditional pdf (cdf) of f (F), given the i-th component $\mathbf{U} \in \mathbb{R}^n$. The functions $f'(s_i)$ and $F'(\cdot|s_i)$ are the partial derivatives of f or F, respectively, with respect to s_i.

Lemma 13 *Let $\mathbf{U} \in \mathbb{R}^n$ be an absolutely continuous random vector and denote, respectively, by f and F its pdf and cdf. Suppose that f and F are differentiable functions. Then, for all $\mathbf{s} = (s_1, \ldots, s_n)^T \in \mathbb{R}^n$, it follows that:*

i) $\dfrac{\partial}{\partial s_i} F(\mathbf{s}) = f(s_i) F(\mathbf{s}_{(i)}|s_i)$;

ii) $\dfrac{\partial^2}{\partial s_i^2} F(\mathbf{s}) = f'(s_i) F(\mathbf{s}_{(i)}|s_i) + f(s_i) F'(\mathbf{s}_{(i)}|s_i)$; *and*

iii) $\dfrac{\partial^2}{\partial s_i \partial s_j} F(\mathbf{s}) = f(s_i) f(s_j|s_i) F(\mathbf{s}_{(j)}|s_i, s_j)$.

Proof: (i) Since $F(\mathbf{s}) = P(U_i \leq s_i|\mathbf{U}_{(i)} \leq \mathbf{s}_{(i)}) P(\mathbf{U}_{(i)} \leq \mathbf{s}_{(i)})$, we have

$$\begin{aligned}\frac{\partial}{\partial s_i} F(\mathbf{s}) &= \frac{\partial}{\partial s_i} P(U_i \leq s_i|\mathbf{U}_{(i)} \leq \mathbf{s}_{(i)}) P(\mathbf{U}_{(i)} \leq \mathbf{s}_{(i)}) \\ &= f(s_i|\mathbf{U}_{(i)} \leq \mathbf{s}_{(i)}) P(\mathbf{U}_{(i)} \leq \mathbf{s}_{(i)}). \end{aligned} \tag{29}$$

Moreover, for a random vector $(\mathbf{V}, \mathbf{W}) \in \mathbb{R}^{p+q}$, such that $\mathbf{V}$ has pdf $f_{\mathbf{V}}$, it is well known that:

$$f(\boldsymbol{v}|\mathbf{W} \in A) = f_{\mathbf{V}}(\mathbf{v}) \frac{P(\mathbf{W} \in A|\mathbf{V} = \mathbf{v})}{P(\mathbf{W} \in A)}, \tag{30}$$

for any subset A of $\mathbb{R}^q$ (see Arellano-Valle *et al.* (2002)). Thus, it follows from (29) and (30) that

$$\frac{\partial}{\partial s_i} F(\mathbf{s}) = f(s_i) F(\mathbf{s}_{(i)}|s_i),$$

(ii) It follows straightforward from (i) that

$$
\begin{aligned}
\frac{\partial^2}{\partial s_i^2}F(\mathbf{s}) &= \left\{\frac{\partial}{\partial s_i}f(s_i)\right\}F(\mathbf{s}_{(i)}|s_i) + f(s_i)\frac{\partial}{\partial s_i}F(\mathbf{s}_{(i)}|s_i)\\
&= f'(s_i)F(\mathbf{s}_{(i)}|s_i) + f(s_i)F'(\mathbf{s}_{(i)}|s_i).
\end{aligned}
$$

(iii) Similarly we have in (i), $F(\mathbf{s}_{(i)}|s_i) = P(U_j \leq s_j|U_i = s_i, \mathbf{U}_{(j)} \leq \mathbf{s}_{(j)})P(\mathbf{U}_{(j)} \leq \mathbf{s}_{(j)}|U_i = s_i)$. Thus it follows from (i) and (30) that:

$$
\begin{aligned}
\frac{\partial^2}{\partial s_i \partial s_j}F(\mathbf{s}) &= \frac{\partial}{\partial s_j}f(s_i)F(\mathbf{s}_{(i)}|s_i)\\
&= f(s_i)\frac{\partial}{\partial s_j}P(U_j \leq s_j|U_i = s_i, \mathbf{U}_{(j)} \leq \mathbf{s}_{(j)})P(\mathbf{U}_{(j)} \leq \mathbf{s}_{(j)}|U_i = s_i)\\
&= f(s_i)f(s_j|U_i = s_i, \mathbf{U}_{(j)} \leq \mathbf{s}_{(j)})P(\mathbf{U}_{(j)} \leq \mathbf{s}_{(j)}|U_i = s_i,)\\
&= f(s_i)f(s_j|s_i)F(\mathbf{s}_{(ij)}|s_i, s_j).
\end{aligned}
$$

Some useful results related to the normal distribution are pointed out in next theorem (see also Domínguez-Molina *et al.* (2003, 2004)).

Theorem 14 *Assume that* $\mathbf{U} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$. *Let* $\mathbf{U}_{(i)}$ *and* $\mathbf{U}_{(ij)}$ *be the sub vectors of* $\mathbf{U}$ *without the components* U_i *and* (U_i, U_j), *respectively. Consider* $\boldsymbol{\mu}_{(i).i} = E(\mathbf{U}_{(i)}|U_i = s_i)$, $\Sigma_{(i).i} = Var(\mathbf{U}_{(i)}|U_i = s_i)$, $\boldsymbol{\mu}_{(ij).ij} = E(\mathbf{U}_{(ij)}|U_i = s_i, U_j = s_j)$ *and* $\Sigma_{(ij).ij} = Var(\mathbf{U}_{(ij)}|U_i = s_i, U_j = s_j)$. *Then, it follows that:*

i) $\displaystyle \frac{\partial}{\partial s_i}\Phi_n(\mathbf{s}; \boldsymbol{\mu}, \Sigma) = \phi(s_i; \mu_i, \sigma_{ii})\Phi_{n-1}(\mathbf{s}_{(i)}; \boldsymbol{\mu}_{(i).i}, \Sigma_{(i).i})$;

ii) $$\frac{\partial^2}{\partial s_i^2}\Phi_n(\mathbf{s}; \boldsymbol{\mu}, \Sigma) = \phi(s_i; \mu_i, \sigma_{ii})\Bigg[\left(\frac{\mu_i - s_i}{\sigma_{ii}}\right)\Phi_{n-1}(\mathbf{s}_{(i)}; \boldsymbol{\mu}_{(i).i}, \Sigma_{(i).i}) + \frac{\partial}{\partial s_i}\Phi_{n-1}(\mathbf{s}_{(i)}; \boldsymbol{\mu}_{(i).i}, \Sigma_{(i).i})\Bigg];$$

iii) $$\frac{\partial^2}{\partial s_i \partial s_j}\Phi_n(\mathbf{s}; \boldsymbol{\mu}, \Sigma) = \phi(s_i; \mu_i, \sigma_{ii})\phi(s_j; \mu_{j.i}, \sigma_{j.i})\Phi_{n-2}(\mathbf{s}_{(ij)}; \boldsymbol{\mu}_{(ij).ij}, \Sigma_{(ij).ij})$$

where for any subset I *of* $\{1, \ldots, n\}$, *the conditional mean vector and variance-covariance matrix of* $(\mathbf{U}_{(I)}|\mathbf{U}_I = \mathbf{s}_I)$ *are given, respectively, by:* $\boldsymbol{\mu}_{(I).I} = \boldsymbol{\mu}_{(I)} + \Sigma_{(I)I}\Sigma_{II}^{-1}(\mathbf{s}_I - \boldsymbol{\mu}_I)$ *and* $\Sigma_{(I).I} = \Sigma_{(I)(I)} - \Sigma_{(I)I}\Sigma_{II}^{-1}\Sigma_{I(I)}$.

Proof: The proof follows from Lemma 13 and from the well know marginal-conditional decomposition of multivariate normal. In fact, if

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{pmatrix} \sim \mathcal{N}_{l+(n-l)} \left(\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right),$$

then $\phi_n(\mathbf{u}; \boldsymbol{\mu}, \Sigma) = \phi_l(\mathbf{u}_1; \boldsymbol{\mu}_1, \Sigma_{11})\phi_{n-l}(\mathbf{u}_2; \boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{2.1})$, where $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{u}_1 - \boldsymbol{\mu}_1)$ and $\Sigma_{2.1} = \boldsymbol{\Sigma}_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ are the mean vector and variance-covariance matrix of $(\mathbf{U}_2|\mathbf{U}_1 = \mathbf{u}_1)$, respectively.