Brazilian Journal of Probability and Statistics
2015, Vol. 29, No. 4, 790–812
DOI: 10.1214/14-BJPS247

On matrix-variate Birnbaum–Saunders distributions and their estimation and application

Luis Sánchez[a], Víctor Leiva[a,b], Francisco J. Caro-Lopera[c] and Francisco José A. Cysneiros[d]

[a] *Universidad de Valparaíso*
[b] *Universidad Adolfo Ibáñez*
[c] *Universidad de Medellín*
[d] *Universidade Federal de Pernambuco*

Abstract. Diverse phenomena from the real-world can be modeled using random matrices, allowing matrix-variate distributions to be considered. The normal distribution is often employed in this modeling, but usually the mentioned random matrices do not follow such a distribution. An asymmetric non-normal model that is receiving considerable attention due to its good properties is the Birnbaum–Saunders (BS) distribution. We propose a statistical methodology based on matrix-variate BS distributions. This methodology is implemented in the statistical software R. A simulation study is conducted to evaluate its performance. Finally, an application with real-world matrix-variate data is carried out to illustrate its potentiality and suitability.

1 Introduction

Several phenomena from the real-world can be modeled by random variables that are correlated, which allows us to consider random vectors and matrices, and their corresponding multivariate and matrix-variate distributions in this modeling. Matrix-variate distributions are used in economy, physics, psychology, shape theory and in other fields; see, for example, Dryden and Mardia (1998). Multi-and-matrix-variate versions of the well-known normal (or Gaussian) distribution have been studied by a number of authors; see Johnson et al. (1994a, pp. 80–206), Kotz et al. (2000, pp. 105–333), Tulino and Verdú (2004), and Anderson et al. (2009, pp. 19–20). However, few has been made on matrix-variate non-normal distributions, despite diverse phenomena can be modeled by this type of distributions, for example, when recognition of handwriting characters is analyzed. Due to the human nature of these characters, the stochastic component is present, allowing matrix-variate models to be used.

A non-normal (asymmetric) distribution, defined on the positive real line, with two parameters (shape and scale) and skewness to the right, is the Birnbaum–Saunders (BS) model. Different aspects of the univariate BS distribution have been

Key words and phrases. Computer language, data analysis, elliptically contoured distribution, maximum likelihood estimator, Monte Carlo method, shape theory.
Received February 2014; accepted April 2014.

considered in recent decades due to its attractive properties and its relationship with the normal distribution; see Birnbaum and Saunders (1969a), Johnson et al. (1994b, pp. 651–663) and Balakrishnan et al. (2011).

A univariate generalization of the BS distribution based on elliptically contoured distributions (EC) distributions, proposed by Díaz-García and Leiva (2005), is known as the generalized BS (GBS) distribution, which includes the BS distribution as a particular case. Univariate GBS distributions have been implemented in the R software by Barros et al. (2009); see www.R-project.org. Extensions of the GBS distribution to the multivariate and matrix-variate cases have been developed by Díaz-García and Domínguez-Molina (2007), Caro-Lopera et al. (2012) and Kundu et al. (2013).

Although the BS distribution has its genesis from engineering, it has been applied to agriculture, air and water contamination, business, finance, industry, human and tree mortality, insurance, medicine, neuroscience, nutrition, pharmacology, psychology, quality control, toxicology and wind energy, among other areas; see, for example, Galea et al. (2004), Balakrishnan et al. (2009a, 2009b, 2011), Leiva et al. (2008a, 2008c, 2009, 2010, 2011a, 2011b, 2012, 2014a, 2014b, 2014d), Kotz et al. (2010), Vilca et al. (2010, 2011), Santana et al. (2011), Azevedo et al. (2012), Ferreira et al. (2012), Paula et al. (2012) and Marchant et al. (2013a, 2013b).

The objectives of this study are (i) to propose a methodology based on matrix-variate GBS distributions, (ii) to implement and evaluate the proposed methodology computationally, and (iii) to apply these results to real-world matrix-variate data. Specifically, we model data of postcodes by the landmarks of handwritten digits. These landmarks are points of correspondence on each object matching between and within populations, which allow images of handwritten digits to be displayed; see Dryden and Mardia (1998, p. 13 and pp. 318–320). The main novelty of the work proposed in this paper in relation to the existing works attributed to Caro-Lopera et al. (2012) and Kundu et al. (2013) is that we introduce a new methodology based on matrix-variate GBS distributions, including estimation, data analysis and real-world applications, whereas Caro-Lopera et al. (2012) did not developed neither estimation or data analysis, and Kundu et al. (2013) studied the multivariate case instead of the matrix-variate case, which cannot be used for analyzing the application here proposed.

The paper is organized as follows. In Section 2, we provide the tools required for developing our methodology. In Section 3, we first estimate the parameters of the matrix-variate model which allows us to propose the methodology and then we study the performance of these estimators by Monte Carlo (MC) simulations. In Section 4, we apply the proposed methodology to real-world matrix-variate data of handwritten characters, which shows its potentiality. In Section 5, we sketch some conclusions, consequences and future issues to be considered from this work.

2 The matrix-variate model

In this section, we introduce several statistical aspects related to matrix-variate GBS distributions.

2.1 A univariate GBS distribution

If $Z \sim \mathrm{N}(0, 1)$, then the random variable T given by

$$T = \beta\big[\alpha Z/2 + \sqrt{(\alpha Z/2)^2 + 1}\big]^2 \tag{2.1}$$

has a BS distribution with parameters of shape $\alpha > 0$ and scale $\beta > 0$, which is denoted by $T \sim \mathrm{BS}(\alpha, \beta)$. The random variable T has positive support and the transformation given in (2.1) is one-to-one, which allows us to establish that

$$Z = \frac{1}{\alpha}\big[\sqrt{T/\beta} - \sqrt{\beta/T}\big] \sim \mathrm{N}(0, 1).$$

Díaz-García and Leiva (2005) proposed a generalization of the transformation given in (2.1) based on the family of EC distributions. The main motivation of that generalization is to do the kurtosis of the BS distribution flexible, which improves the data modeling. A random variable Z has a standard EC distribution (symmetric in the univariate case) in $\mathbb{R}$ with kernel function g, which is denoted by $Z \sim \mathrm{S}(g)$, if its probability density function (PDF) is expressed as

$$f_Z(z) = c g\big(z^2\big), \qquad z \in \mathbb{R}, \tag{2.2}$$

where c is a normalizing constant such that $\int_{-\infty}^{\infty} g(z^2)\,\mathrm{d}z = 1/c$. From (2.1), if $Z \sim \mathrm{S}(g)$, then the random variable T has a GBS distribution with parameters of shape $\alpha > 0$, scale $\beta > 0$ and kernel g, which is denoted by $T \sim \mathrm{GBS}(\alpha, \beta; g)$. In this case, the PDF of T is given by

$$f_T(t) = \frac{c}{2\alpha\beta^{1/2}} t^{-3/2}[t + \beta] g(\kappa_t), \qquad t > 0, \tag{2.3}$$

where $\kappa_t = [\xi(t/\beta)/\alpha]^2 = [t/\beta + \beta/t - 2]/\alpha^2$, with $\xi(u) = \sqrt{u} - 1/\sqrt{u} = 2\sinh(\log(\sqrt{u}))$, and c, g are given in (2.2). If g is the normal kernel, then we have the univariate BS distribution. If g corresponds to the Student-t kernel with ν degrees of freedom, we then have the univariate BS-t distribution, which random variable is denoted by $T \sim \mathrm{BS}\text{-}t(\alpha, \beta; \nu)$; see Azevedo et al. (2012) and Paula et al. (2012).

2.2 A multivariate GBS distribution

The univariate GBS distribution can be extended to the multivariate case using EC distributions. Specifically, let $\mathbf{x} = (X_1, \ldots, X_n)^\top \in \mathbb{R}^n$ be a random vector with multivariate EC distribution, characterized by a location vector $\boldsymbol{\mu} \in \mathbb{R}^n$, a scale matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$, with $\mathrm{rank}(\boldsymbol{\Sigma}) = n$, and the corresponding kernel g, which is denoted by $\mathbf{x} \sim \mathrm{EC}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$. In this case, the PDF of $\mathbf{x}$ is defined as

$$f_{\mathbf{x}}(\mathbf{x}) = c|\boldsymbol{\Sigma}|^{-1/2} g\big([\mathbf{x}-\boldsymbol{\mu}]^\top \boldsymbol{\Sigma}^{-1}[\mathbf{x}-\boldsymbol{\mu}]\big), \qquad \mathbf{x} \in \mathbb{R}^n,$$

where c, g are as given in (2.2). Let $\mathbf{z} = (Z_1, \ldots, Z_n)^\top \sim \mathrm{EC}_n(\mathbf{0}_n, \mathbf{I}_n; g)$, with $\mathbf{0}_n$ and $\mathbf{I}_n$ being the $n \times 1$ zero vector and $n \times n$ identity matrix, $\mathbf{t} = (T_1, \ldots, T_n)^\top \in \mathbb{R}^n_+$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top \in \mathbb{R}^n_+$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^\top \in \mathbb{R}^n_+$, such that $Z_i \in \mathbb{R}$, $T_i, \alpha_i, \beta_i > 0$ satisfy the relation in (2.1), for all $i = 1, \ldots, n$. Then, the random vector $\mathbf{t}$ has a multivariate GBS distribution, denoted by $\mathbf{t} \sim \mathrm{GBS}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}; g)$, and its PDF is given by

$$f_{\mathbf{t}}(t_1, \ldots, t_n) = \frac{c}{2^n} g\left(\sum_{i=1}^n \frac{1}{\alpha_i^2}\left[\frac{t_i}{\beta_i} + \frac{\beta_i}{t_i} - 2\right]\right) \prod_{i=1}^n \frac{t_i^{-3/2}[t_i + \beta_i]}{\alpha_i\sqrt{\beta_i}},$$

with $t_i > 0$, for $i = 1, \ldots, n$.

2.3 A matrix-variate GBS distribution

GBS distributions can be also extended to the matrix-variate case using the EC family. Specifically, let $\mathbf{X} = (X_{ij}) \in \mathbb{R}^{n\times k}$ be a random matrix with EC distribution, characterized by a location matrix $\mathbf{M} \in \mathbb{R}^{n\times k}$, scale matrices $\boldsymbol{\Omega} \in \mathbb{R}^{k\times k}$, with $\mathrm{rank}(\boldsymbol{\Omega}) = k$, and $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$, with $\mathrm{rank}(\boldsymbol{\Sigma}) = n$, and kernel g, which is denoted by $\mathbf{X} \sim \mathrm{EC}_{n\times k}(\mathbf{M}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}; g)$. In this case, the PDF of $\mathbf{X}$ is expressed as

$$f_{\mathbf{X}}(\mathbf{X}) = c|\boldsymbol{\Omega}|^{-n/2}|\boldsymbol{\Sigma}|^{-k/2} \times g\big(\mathrm{tr}\big(\boldsymbol{\Omega}^{-1}[\mathbf{X}-\mathbf{M}]^\top \boldsymbol{\Sigma}^{-1}[\mathbf{X}-\mathbf{M}]\big)\big), \qquad \mathbf{X} \in \mathbb{R}^{n\times k},$$

where c, g are given in (2.2). Now, let $\mathbf{Z} = (Z_{ij}) \sim \mathrm{EC}_{n\times k}(\mathbf{0}_{n\times k}, \mathbf{I}_k, \mathbf{I}_n; g)$, with $\mathbf{0}_{n\times k}$ being the $n \times k$ zero matrix, $\mathbf{T} = (T_{ij}) \in \mathbb{R}^{n\times k}_+$, $\mathbf{A} = (\alpha_{ij}) \in \mathbb{R}^{n\times k}_+$ and $\mathbf{B} = (\beta_{ij}) \in \mathbb{R}^{n\times k}_+$, with $T_{ij} = \beta_{ij}[\alpha_{ij} Z_{ij}/2 + \sqrt{(\alpha_{ij} Z_{ij}/2)^2 + 1}]^2$, for $\alpha_{ij}, \beta_{ij} > 0$ and $i = 1, \ldots, n$, $j = 1, \ldots, k$. Then, the random matrix $\mathbf{T}$ has a matrix-variate GBS distribution, denoted by $\mathbf{T} \sim \mathrm{GBS}_{n\times k}(\mathbf{A}, \mathbf{B}; g)$, and its PDF is given by

$$f_{\mathbf{T}}(\mathbf{t}) = \frac{c}{2^{nk}} g\left(\sum_{i=1}^n \sum_{j=1}^k \frac{1}{\alpha_{ij}^2}\left[\frac{t_{ij}}{\beta_{ij}} + \frac{\beta_{ij}}{t_{ij}} - 2\right]\right) \times \prod_{i=1}^n \prod_{j=1}^k \frac{t_{ij}^{-3/2}[t_{ij} + \beta_{ij}]}{\alpha_{ij}\sqrt{\beta_{ij}}},$$

with $t_{ij} > 0$, for $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

2.4 Generation of random matrices and moments

Below, we present a generator of random matrices from the GBS distribution. It is known that, for matrices $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{p\times n}$, $\mathbf{C} \in \mathbb{R}^{k\times k}$ and $\mathbf{D} \in \mathbb{R}^{n\times n}$, the following properties that involve the Kronecker product ($\otimes$) and the trace (tr) and vectorization (vec) operators are satisfied: (i) $\mathrm{tr}(\mathbf{U}^\top \mathbf{C}\mathbf{V}\mathbf{D}) = (\mathrm{vec}(\mathbf{U}^\top))^\top (\mathbf{C} \otimes \mathbf{D}^\top)\,\mathrm{vec}(\mathbf{V}^\top)$; (ii) $|\mathbf{C} \otimes \mathbf{D}| = |\mathbf{C}|^n |\mathbf{D}|^k$; and (iii) $(\mathbf{C} \otimes \mathbf{D})^{-1} = \mathbf{C}^{-1} \otimes \mathbf{D}^{-1}$, if $\mathbf{C}$ and $\mathbf{D}$ are invertible.

Based in Gupta and Varga (1994), we have that if $\mathbf{X} \in \mathbb{R}^{n\times k}$ is a random matrix and $\mathbf{x} = \mathrm{vec}(\mathbf{X}^\top)$, then $\mathbf{X} \sim \mathrm{EC}_{n\times k}(\mathbf{M}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}; g)$ if and only if $\mathbf{x} \sim \mathrm{EC}_{nk}(\mathrm{vec}(\mathbf{M}^\top), \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}; g)$. Hence, it is possible to generate data from a matrix-variate GBS distribution using multivariate EC random vectors. Thus, given the integer numbers n and k, $\mathbf{A} = (\alpha_{ij}) \in \mathbb{R}_+^{n\times k}$, $\mathbf{B} = (\beta_{ij}) \in \mathbb{R}_+^{n\times k}$ and g, matrix-variate data from $\mathbf{T} \sim \mathrm{GBS}_{n\times k}(\mathbf{A}, \mathbf{B}; g)$ can be generated by Algorithm 1.

To perform recognition of handwriting characters, we need the mean of $\mathbf{T} \sim \mathrm{GBS}_{n\times k}(\mathbf{A}, \mathbf{B}; g)$ given by

$$\mathrm{E}[\mathbf{T}] = 0.5\mathbf{B} \odot \left[2 + \omega_1 \mathbf{A}^{2H}\right], \tag{2.4}$$

where $\odot$ denotes the Hadamard product and $\omega_1 = \mathrm{E}[U]$, with $U \sim \mathrm{G}\chi^2(1; g)$, that is, U follows a generalized chi-squared distribution with one degree of freedom; see Fang et al. (1990) and Gupta and Varga (1993) for the central $\mathrm{G}\chi^2$ distribution, and Díaz-García et al. (2002, 2003) for the non-central case. Note that the power of a matrix in relation to the Hadamard product is simpler than that from the usual matrix product. We denote the Hadamard powers by $\mathbf{X}^{a\mathrm{H}} = (X_{ij}^a)$, for $a \in \mathbb{R}$ and so we have that: (i) $\mathbf{X}^{(1/2)\mathrm{H}} = (X_{ij}^{1/2})$ denotes the positive root of $\mathbf{X}$

Algorithm 1 Generator of matrix-variate data from the GBS distribution

1: Generate a random vector $\mathbf{z}_j = (z_{1j}, \ldots, z_{kj})^\top \in \mathbb{R}^k$ from $\mathbf{z}_j \sim \mathrm{EC}_k(\mathbf{0}_k, \mathbf{I}_k; g)$, for $j = 1, \ldots, n$;

2: Create a matrix $\mathbf{Z} = (z_{ij}) \in \mathbb{R}^{n\times k}$ with the vector $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^\top$ filling it by rows as

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1^\top \\ \vdots \\ \mathbf{z}_n^\top \end{pmatrix};$$

3: Obtain the element t_{ij} from $t_{ij} = \beta_{ij}[\alpha_{ij} z_{ij}/2 + \sqrt{(\alpha_{ij} z_{ij}/2)^2 + 1}]^2$, fixing α_{ij} and β_{ij}, for each element z_{ij} of the matrix $\mathbf{Z}$. The matrix $\mathbf{T} = (t_{ij}) \in \mathbb{R}_+^{n\times k}$ is an observation from $\mathbf{T} \sim \mathrm{GBS}_{n\times k}(\mathbf{A}, \mathbf{B}; g)$.

---

with respect to the Hadamard product, such that $\mathbf{X}^{(1/2)\mathrm{H}} \odot \mathbf{X}^{(1/2)\mathrm{H}} = \mathbf{X}$; and (ii) $\mathbf{X}^{-\mathrm{H}} = (1/X_{ij})$ denotes the inverse matrix of $\mathbf{X}$ with respect to the Hadamard product, such that $\mathbf{X} \odot \mathbf{X}^{-\mathrm{H}} = \mathbf{J}$, where $\mathbf{J}$ is an $n \times k$ matrix consisting of ones. To specific the expression given in (2.4) for the case of the $t(3), t(8), t(50)$ and $t(\infty) \equiv \mathrm{N}(0, 1)$ kernels, we have $\omega_1 = 3, 4/3, 25/24$ and 1, respectively; see Leiva et al. (2008a).

2.5 Relation between matrix-variate and univariate GBS distributions

Matrix-variate and univariate EC distributions are related as follows. Let $\mathbf{X} \sim \mathrm{EC}_{n\times k}(\mathbf{M}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}; g)$, with $\mathbf{X} = (X_{ij})$, $\mathbf{M} = (m_{ij})$, $\boldsymbol{\Omega} = (\omega_{ij})$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$. Then, $X_{ij} \sim \mathrm{EC}_1(m_{ij}, \omega_{ij}\sigma_{ij}; g)$, for $i = 1, \ldots, n$ and $j = 1, \ldots, k$; see Gupta and Varga (1994). Thus, if $\mathbf{Z} = (Z_{ij}) \sim \mathrm{EC}_{n\times k}(\mathbf{0}_{n\times k}, \mathbf{I}_k, \mathbf{I}_n; g)$, $Z_{ij} \sim \mathrm{EC}_1(0, 1; g) \equiv \mathrm{S}(g)$. Let $\mathbf{T} \sim \mathrm{GBS}_{n\times k}(\mathbf{A}, \mathbf{B}; g)$, with $\mathbf{T} = (T_{ij})$, $\mathbf{A} = (\alpha_{ij})$ and $\mathbf{B} = (\beta_{ij})$. Because $\mathbf{T} = \psi(\mathbf{Z})$, where $\psi(\cdot)$ is an one-to-one transformation from $\mathbb{R}_+^{n\times k}$ to $\mathbb{R}^{n\times k}$ and $\mathbf{Z} \sim \mathrm{EC}_{n\times k}(\mathbf{0}_{n\times k}, \mathbf{I}_k, \mathbf{I}_n; g)$, or more precisely, T_{ij} and Z_{ij} satisfy the relation given in (2.1), for α_{ij} and β_{ij}, where $Z_{ij} \sim \mathrm{S}(g)$, we can conclude that $T_{ij} \sim \mathrm{GBS}(\alpha_{ij}, \beta_{ij}; g)$, for $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

3 Estimation

In this section, we estimate the matrix-variate GBS parameters with the (ML) maximum likelihood method and evaluate their performance by simulation.

3.1 ML method

Based on the results given in Section 2.5, the ML estimates of the matrix-variate GBS parameters can be obtained by means of the ML estimates of the univariate GBS parameters as follows. Let $\mathbf{T}^{(1)}, \ldots, \mathbf{T}^{(m)}$ be a random sample of size m from $\mathbf{T} \sim \mathrm{GBS}_{n\times k}(\mathbf{A}, \mathbf{B}; g)$, where $\mathbf{T}^{(l)} = (T_{ij}^{(l)})$, for $l = 1, \ldots, m$, $\mathbf{A} = (\alpha_{ij})$ and $\mathbf{B} = (\beta_{ij})$. Then, the ML estimators of the matrix-variate GBS parameters can be obtained as $\widehat{\mathbf{A}} = (\widehat{\alpha}_{ij})$ and $\widehat{\mathbf{B}} = (\widehat{\beta}_{ij})$, where $\widehat{\alpha}_{ij}$ and $\widehat{\beta}_{ij}$ are the ML estimators of the parameters of the $\mathrm{GBS}(\alpha_{ij}, \beta_{ij}; g)$ distribution for the random sample $T_{ij}^{(1)}, \ldots, T_{ij}^{(m)}$, with $i = 1, \ldots, n$ and $j = 1, \ldots, k$. Note that the problem of estimation in a matrix-variate GBS distribution of dimension $n \times k$ can be solved by nk estimation problems in univariate GBS distributions due to the result provided in Section 2.5. Thus, to estimate the matrix-variate GBS parameters, we need simply a method to estimate the univariate GBS parameters. Several efforts have been carried out to estimate the univariate GBS parameters, but some numerical problems remain, which we propose to solve in the third point considered next.

First, it is well known the ML estimates of the GBS distribution do not have explicit form. For the BS case, Birnbaum and Saunders (1969b) proposed a method to find the ML estimates of α and β, $\widehat{\alpha}$ and $\widehat{\beta}$ say, which provides an explicit form for $\widehat{\alpha}$, but the ML estimate of β must be obtained numerically. Birnbaum and Saunders' method is summarized in Algorithm 2. Birnbaum and Saunders (1969b) proved that, under certain conditions, the sequence $\{\widehat{\beta}_{r+1}, r = 0, 1, \ldots\}$ given in Step 2 of Algorithm 2 converges to the ML estimator of β. Algorithm 2 is implemented in the R computer language; see Barros et al. (2009).

Second, Balakrishnan et al. (2009b) used the EM approach to estimate the parameters of BS distributions based in scale mixtures of normal models, between which the BS-t distribution is a particular case. They proved that, using this approach for the BS-t distribution with ν degrees of freedom, a similar method to that proposed by Birnbaum and Saunders (1969b) and detailed in Algorithm 2 is obtained, but now $h(y)$ given in its Step 2 is replaced by

$$h_u(y) = y^2 - y[2r_u\bar{u} + K(y)] + r_u[s_u + \bar{u}K(y)],$$

where $s_u = [1/n]\sum_{i=1}^n \widehat{u}_i t_i$, $r_u = [(1/n)\sum_{i=1}^n \widehat{u}_i t_i^{-1}]^{-1}$ and $\bar{u} = [1/n]\sum_{i=1}^n \widehat{u}_i$, with $\widehat{u}_i = [\nu + 1]/[\nu + \widehat{\kappa}_{t_i}]$ and κ_{t_i} as given in (2.3). Balakrishnan et al. (2009b) proposed as starting values to find the ML estimates of α and β their corresponding values from the BS distribution. Balakrishnan et al.'s method can be seen as a generalization of Birnbaum and Saunders' method, because if $\widehat{u}_i = 1$, for $i = 1, \ldots, n$ (degenerate case), Algorithm 2 is obtained.

Algorithm 2 ML estimation of the univariate BS parameters

1: Estimate the parameter α of the BS distribution with a sample of observations $t_1, \ldots, t_n$ by $\widehat{\alpha} = [s/\widehat{\beta} + \widehat{\beta}/r - 2]^{1/2}$, where s and r are arithmetic and harmonic means of $t_1, \ldots, t_n$ given by $s = [1/n]\sum_{i=1}^n t_i$ and $r = [(1/n)\sum_{i=1}^n t_i^{-1}]^{-1}$, and $\widehat{\beta}$ is the ML estimate of β;

2: Consider the starting value $\widehat{\beta}_0 = [sr]^{1/2}$ and calculate the ML estimate of β by

$$\widehat{\beta}_{r+1} = \widehat{\beta}_r + \frac{h(\widehat{\beta}_r)}{h'(\widehat{\beta}_r)}, \qquad r = 0, 1, \ldots,$$

where $h(y) = y^2 - y[2r + K(y)] + r[s + K(y)]$, with $K(y) = [(1/n)\sum_{i=1}^n (y + t_i)^{-1}]^{-1}$, and $h'(y) = [y - r][1 - K'(y)] + y - r - K(y)]$, with $K'(y) = K^2(y)[1/n]\sum_{i=1}^n (y + t_i)^{-2}$;

3: Repeat Step 2 of Algorithm 2 until to reach convergence such as that inherited from the R function `uniroot()`, in which a solution for $h(y) = 0$ converges when $h(y_{r+1}) = 0$, or when $|y_{r+1} - y_r| < 10^{-5}$, where y_{r+1} is the current value of y and y_r its previous value.

Third, Step 2 of Algorithm 2 corresponds to Newton–Raphson's iterations to solve $h(y) = 0$, which does not always converge. Thus, to overcome this convergence problem, we propose to use a search procedure of zeros. Brent (1973) developed this type of procedures combining the bisection, inverse quadratic interpolation and secant techniques, which guarantees the convergence to zero of the function, it does not require a starting value and is implemented in the `R` language by the function `uniroot()`. We use Brent's procedure to obtain the ML estimates of the parameters of the BS distribution and the criterion of convergence of the function `uniroot()`. We estimate the parameters of the BS-t distribution with Balakrishnan et al.'s method, using as starting values those obtained from the BS case and a similar criterion of convergence, selecting the parameter ν of the BS-t distribution with a non-failing optimum criterion such as described in Barros et al. (2009).

3.2 Simulation

To evaluate the performance of the estimation method described in Section 3.1, we use MC simulations and the generator of univariate GBS random numbers proposed by Leiva et al. (2008b). The simulations consider the setting: (S1) sample size $m \in \{10, 25, 100\}$, covering small, moderate and large sizes; (S2) shape parameter $\alpha \in \{0.2, 0.5, 1.0\}$, considering low, moderate and high asymmetry, respectively, fixing the scale parameter at $\beta = 1.0$, without loss of generality; and (S3) kernel $g \in \{t(3), t(8), t(50)\}$, corresponding to high, moderate and low kurtosis, respectively, in relation to the normal case $t(\infty) \equiv \mathrm{N}(0, 1)$. The quality of the method is studied by the empirical relative bias (RS) and root of the mean square error ($\sqrt{\mathrm{MSE}}$) of the ML estimators. The sample is generated from the GBS model with a specific kernel (normal or $t(\nu)$), called "true kernel", and the estimation of parameters is computed from samples obtained using the same or another kernel, called "assumed kernel". The empirical RS and $\sqrt{\mathrm{MSE}}$ are averages of 5000 MC replications for each combination of m, α, g (settings S1–S3). The results of the simulations are presented in Tables 1 and 2 for $\widehat{\alpha}$ and $\widehat{\beta}$, respectively. Other results (omitted here) based on the direct maximization of the log-likelihood function of α and β (by the L-BFGS-B procedure) showed, in some cases, convergence problems. Furthermore, the method proposed in this paper based on Algorithm 2 and Brent's procedure has a processing time smaller than that based on the L-BFGS-B procedure. From Tables 1 and 2, note that when the true and assumed models are the same, the performance of the method is evaluated using the empirical RS of the estimators of α and β. We get the results expected when the empirical RS is analyzed. For instance, it decreases when m increases, increases when the asymmetry increases and decreases when the kurtosis increases. A misspecification of the GBS model (i.e., when true and assumed models are different) introduces empirical RS and $\sqrt{\mathrm{MSE}}$ greater in the estimation of α than in β. The sensitivity of the estimation method is studied by the empirical $\sqrt{\mathrm{MSE}}$ of the estimators of α

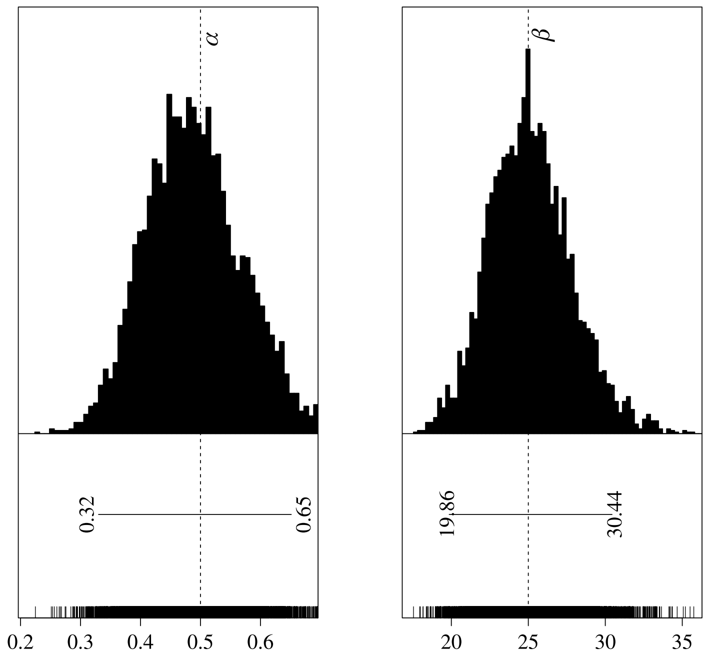

Figure 1 *Empirical distributions of the ML estimators of* α *and* β *from a BS-t*$(0.5, 25; 8)$ *distribution.*

and β. In general, when the assumed model becomes away from the true model, the empirical $\sqrt{\text{MSE}}$ increases, such as expected, but it decreases when m increases, which also occurs when the asymmetry decreases. Furthermore, in the case of the estimation of α, the empirical $\sqrt{\text{MSE}}$ decreases when the kurtosis increases. Figure 1 shows histograms of the empirical distributions of the ML estimators of α and β for a particular setting. Note that the shape of these histograms is close to the normal distribution. The lines constructed below of the histograms represent the asymptotic confidence intervals of level 95% given by $\widehat{\theta} \pm 1.96\widehat{\text{SE}}(\widehat{\theta})$, where $\theta = \alpha$ or β, and $\widehat{\theta}$ and $\widehat{\text{SE}}(\widehat{\theta})$ are the estimated parameter and estimated standard error (SE) of the estimator, respectively. These results show an empirical behavior for the distribution of the ML estimators of α and β that is expected for the corresponding asymptotic distributions.

4 Application

In this section, we use matrix-variate GBS distributions to model real-world data of handwritten digits and compare them to some symmetric distributions. We choose the best distribution and estimate the mean shape of digit 3.

Table 1 *Empirical RS and* $\sqrt{\text{MSE}}$ *of the estimator of* α *for the indicated values and distributions* ($\beta = 1.0$)

			True distribution							
			BS-t(2)		BS-t(8)		BS-t(50)		BS	
α	m	Assumed distribution	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$
0.2	10	BS-t(2)	−0.0254	0.1316	−0.2691	0.0446	−0.3191	0.0392	−0.3276	0.0383
		BS-t(8)	0.5647	0.2646	−0.0554	0.0530	−0.1422	0.0429	−0.1563	0.0413
		BS-t(50)	0.9221	0.3327	0.0358	0.0601	−0.0726	0.0457	−0.0897	0.0436
		BS	0.9761	0.3412	0.0539	0.0617	−0.0587	0.0464	−0.0764	0.0442
	25	BS-t(2)	0.0123	0.2017	−0.2313	0.0281	−0.2799	0.0250	−0.2885	0.0246
		BS-t(8)	0.5160	0.2476	−0.0190	0.0332	−0.1046	0.0272	−0.1189	0.0263
		BS-t(50)	1.1313	0.3517	0.0893	0.0386	−0.0257	0.0289	−0.0440	0.0278
		BS	1.2787	0.3744	0.1150	0.0406	−0.0080	0.0295	−0.0274	0.0282
	100	BS-t(2)	−0.0008	0.0225	−0.2170	0.0140	−0.2641	0.0124	−0.2724	0.0122
		BS-t(8)	0.4798	0.0391	−0.0043	0.0165	−0.0894	0.0135	−0.1036	0.0130
		BS-t(50)	1.1158	0.1791	0.1135	0.0195	−0.0060	0.0145	−0.0247	0.0138
		BS	1.6546	0.3259	0.1459	0.0212	0.0139	0.0148	−0.0063	0.0140
0.5	10	BS-t(2)	−0.0082	0.3643	−0.2704	0.1115	−0.3202	0.0980	−0.3287	0.0957
		BS-t(8)	0.5895	0.6055	−0.0570	0.1323	−0.1433	0.1072	−0.1573	0.1033
		BS-t(50)	0.8309	0.6762	0.0314	0.1484	−0.0747	0.1139	−0.0916	0.1089
		BS	0.8703	0.6875	0.0485	0.1518	−0.0612	0.1156	−0.0785	0.1104
	25	BS-t(2)	0.0100	0.3931	−0.2319	0.0703	−0.2804	0.0625	−0.2890	0.0614
		BS-t(8)	0.5488	0.5562	−0.0197	0.0828	−0.1050	0.0679	−0.1193	0.0659
		BS-t(50)	1.0830	0.7379	0.0874	0.0961	−0.0266	0.0723	−0.0448	0.0694
		BS	1.1876	0.7759	0.1123	0.1005	−0.0091	0.0737	−0.0283	0.0705
	100	BS-t(2)	−0.0009	0.0562	−0.2171	0.0350	−0.2642	0.0311	−0.2725	0.0304
		BS-t(8)	0.4847	0.1902	−0.0044	0.0412	−0.0895	0.0337	−0.1037	0.0326
		BS-t(50)	1.1845	0.5709	0.1131	0.0488	−0.0062	0.0362	−0.0249	0.0345
		BS	1.5832	0.7445	0.1451	0.0528	0.0136	0.0370	−0.0065	0.0351

Table 1 (*Continued*)

			True distribution							
			BS-t(2)		BS-t(8)		BS-t(50)		BS	
α	m	Assumed distribution	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$
1.0	10	BS-t(2)	0.0170	0.7314	−0.2730	0.2241	−0.3227	0.1971	−0.3311	0.1924
		BS-t(8)	0.5684	1.0649	−0.0599	0.2663	−0.1457	0.2151	−0.1596	0.2075
		BS-t(50)	0.7488	1.1769	0.0220	0.2930	−0.0799	0.2278	−0.0962	0.2181
		BS	0.7807	1.1981	0.0376	0.2985	−0.0671	0.2309	−0.0839	0.2208
	25	BS-t(2)	0.0108	0.6485	−0.2332	0.1408	−0.2816	0.1252	−0.2901	0.1231
		BS-t(8)	0.5707	0.9780	−0.0210	0.1657	−0.1061	0.1360	−0.1202	0.1320
		BS-t(50)	1.0324	1.3098	0.0835	0.1909	−0.0286	0.1446	−0.0465	0.1389
		BS	1.1173	1.3858	0.1070	0.1986	−0.0115	0.1473	−0.0304	0.1410
	100	BS-t(2)	−0.0011	0.1125	−0.2174	0.0701	−0.2645	0.0622	−0.2728	0.0609
		BS-t(8)	0.4885	0.3765	−0.0047	0.0824	−0.0898	0.0675	−0.1039	0.0652
		BS-t(50)	1.2133	1.0781	0.1121	0.0974	−0.0067	0.0724	−0.0254	0.0690
		BS	1.5297	1.3597	0.1435	0.1049	0.0130	0.0740	−0.0070	0.0702

Table 2 *Empirical RS and* $\sqrt{\text{MSE}}$ *of the estimator of* β *for the indicated values and distributions* ($\beta = 1.0$)

			True distribution							
			BS-t(2)		BS-t(8)		BS-t(50)		BS	
α	m	Assumed distribution	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$
0.2	10	BS-t(2)	0.0132	0.6450	0.0026	0.0751	0.0022	0.0724	0.0022	0.0720
		BS-t(8)	0.0249	0.7793	0.0022	0.0706	0.0018	0.0654	0.0019	0.0646
		BS-t(50)	0.0403	0.8386	0.0021	0.0723	0.0018	0.0646	0.0018	0.0635
		BS	0.0432	0.8492	0.0022	0.0731	0.0018	0.0646	0.0018	0.0635
	25	BS-t(2)	0.0215	1.4371	0.0009	0.0472	0.0009	0.0458	0.0010	0.0455
		BS-t(8)	0.0322	1.5952	0.0010	0.0444	0.0008	0.0413	0.0009	0.0408
		BS-t(50)	0.0532	1.5799	0.0011	0.0457	0.0008	0.0407	0.0008	0.0399
		BS	0.0586	1.5607	0.0011	0.0465	0.0008	0.0408	0.0008	0.0399
	100	BS-t(2)	−0.0006	0.0261	−0.0004	0.0235	−0.0004	0.0229	−0.0004	0.0228
		BS-t(8)	−0.0002	0.0297	−0.0004	0.0221	−0.0004	0.0207	−0.0004	0.0204
		BS-t(50)	0.0083	0.3012	−0.0004	0.0227	−0.0003	0.0204	−0.0003	0.0200
		BS	0.0327	0.5134	−0.0004	0.0232	−0.0003	0.0204	−0.0003	0.0200
0.5	10	BS-t(2)	0.0694	2.1200	0.0167	0.1905	0.0150	0.1832	0.0149	0.1821
		BS-t(8)	0.1556	2.3517	0.0139	0.1766	0.0120	0.1625	0.0119	0.1605
		BS-t(50)	0.2018	2.4160	0.0144	0.1821	0.0115	0.1600	0.0115	0.1572
		BS	0.2098	2.4263	0.0147	0.1844	0.0115	0.1602	0.0115	0.1571
	25	BS-t(2)	0.0899	4.5456	0.0063	0.1184	0.0061	0.1147	0.0062	0.1140
		BS-t(8)	0.1565	4.2842	0.0060	0.1103	0.0051	0.1019	0.0051	0.1004
		BS-t(50)	0.2162	3.3221	0.0065	0.1146	0.0049	0.1001	0.0048	0.0980
		BS	0.2302	3.1044	0.0068	0.1172	0.0049	0.1002	0.0048	0.0979
	100	BS-t(2)	−0.0001	0.0655	0.0001	0.0586	−0.00004	0.0570	−0.00002	0.0567
		BS-t(8)	0.0016	0.0822	−0.0001	0.0544	−0.0001	0.0506	−0.0001	0.0499
		BS-t(50)	0.1121	1.7014	0.0000	0.0564	−0.0001	0.0497	−0.0001	0.0487
		BS	0.1867	1.8125	0.0001	0.0584	−0.0001	0.0498	−0.0001	0.0486

Table 2 (*Continued*)

			True distribution							
		Assumed distribution	BS-t (2)		BS-t (8)		BS-t (50)		BS	
α	m		RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$	RS	$\sqrt{\text{MSE}}$
1.0	10	BS-t (2)	0.2416	4.6552	0.0626	0.3900	0.0577	0.3739	0.0572	0.3718
		BS-t (8)	0.4556	4.7577	0.0501	0.3481	0.0422	0.3138	0.0417	0.3097
		BS-t (50)	0.5134	4.6820	0.0534	0.3604	0.0407	0.3080	0.0400	0.3017
		BS	0.5224	4.6610	0.0546	0.3645	0.0408	0.3082	0.0400	0.3015
	25	BS-t (2)	0.2313	9.0645	0.0248	0.2347	0.0234	0.2268	0.0235	0.2254
		BS-t (8)	0.3729	7.4173	0.0211	0.2112	0.0174	0.1918	0.0170	0.1887
		BS-t (50)	0.4742	4.6997	0.0233	0.2227	0.0166	0.1873	0.0159	0.1826
		BS	0.4865	4.3616	0.0245	0.2285	0.0166	0.1875	0.0159	0.1824
	100	BS-t (2)	0.0042	0.1303	0.0031	0.1136	0.0028	0.1105	0.0028	0.1097
		BS-t (8)	0.0337	1.3959	0.0022	0.1019	0.0018	0.0936	0.0018	0.0922
		BS-t (50)	0.3020	2.9321	0.0030	0.1080	0.0019	0.0914	0.0017	0.0891
		BS	0.3837	2.8438	0.0037	0.1136	0.0019	0.0916	0.0018	0.0889

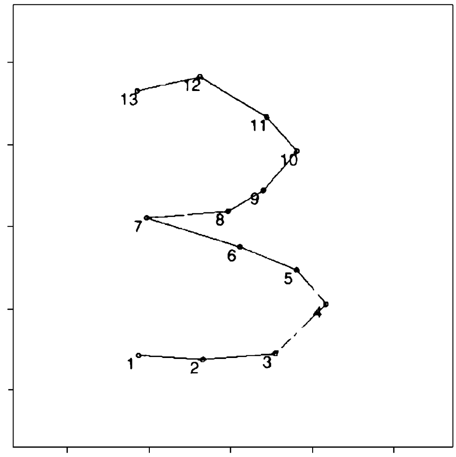

Figure 2 *Landmarks for handwritten British postcodes of digit* 3.

4.1 The data set

Dryden and Mardia (1998, pp. 318–320) presented landmark data corresponding to handwritten British postcodes of digit 3. A landmark is a point of correspondence on each object that matches between and within populations; see Dryden and Mardia (1998, p. 13). Figure 2 shows 13 landmarks of an image of handwritten digit 3. Landmark 1 is at the extreme bottom left; landmark 4 is at the maximum curvature of the bottom arc; landmark 7 is at the extreme of the central protrusion; landmark 10 is at the maximum curvature of the top arc; and landmark 13 is the extreme top left point. The other landmarks are pseudo-landmarks, localized at approximately equal intervals between the previous landmarks.

The data set contain $m = 30$ handwritten records of digit 3, with $n = 13$ landmarks and $k = 2$ dimensions, because the handwritten digit 3 is considered in a Cartesian system. The data set is presented in the Appendix, where each row corresponds to an observation of the handwritten digit 3, and the coordinates are of the type $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Here, we estimate the mean shape of the handwritten digit 3. Thus, first, we suppose that the data follow a specific matrix-variate asymmetric or symmetric distribution within a set of possible options, which are the BS and BS-t asymmetric models and N(0, 1) and Student-t symmetrical models. Then, we choose the distribution that fits the data better using model selection criteria. Thus, with the best model, we estimate the mean digit 3 shape.

4.2 Estimation

First, we estimate the parameters of the EC distributions, that is, of the normal and Student-$t(\nu)$ models, with $\nu \in \{3, 8, 50\}$, considering high ($\nu = 3$), moderate ($\nu = 8$) and low ($\nu = 50$) kurtosis levels. The scale matrices are chosen as $\mathbf{\Omega} =$

$\sigma^2\mathbf{I}_2$, where σ^2 corresponds to a dispersion parameter, and $\boldsymbol{\Sigma} = \mathbf{I}_{13}$. Then, the parameters to estimate are σ^2 and $\mathbf{M}$ (the mean of the distribution).

Because of the equivalence between the matrix-variate and multivariate EC models given in Section 2.5, we develop the estimation procedure using equivalent multivariate EC models. Thus, the models under which we do the estimation are $\mathrm{EC}_{26}(\boldsymbol{\mu}, \sigma^2\mathbf{I}_{26}; g)$, where $\boldsymbol{\mu} = \mathrm{vec}(\mathbf{M}^\top)$. By using the R software, the ML estimates of the corresponding parameters are displayed in Table 3.

Now, using the estimation method proposed in Section 3.1, we fit the matrix-variate GBS distribution using N(0,1), $t(3)$, $t(8)$ and $t(50)$ kernels. As mentioned, the values $\nu \in \{3, 8, 50\}$ are chosen to vary the kurtosis level of the GBS model from high ($\nu = 3$) to low ($\nu = 50$) kurtosis levels. Once again using the R software, the ML estimates of the parameter matrices $\mathbf{A}$ and $\mathbf{B}$ of the matrix-variate GBS distribution for the indicated kernel (g) are displayed in Table 3.

4.3 Selection of the best model

To select the best model (from the set of considered models) for digit 3 data, we use selection criteria based on loss of information, such as Akaike (AIC) and Schwarz's Bayesian (BIC). These criteria allow us to compare models for the same data set, which are given by

$$\mathrm{AIC} = 2p - 2\ell(\widehat{\boldsymbol{\theta}})$$

and

$$\mathrm{BIC} = p\log(m) - 2\ell(\widehat{\boldsymbol{\theta}}),$$

where $\widehat{\boldsymbol{\theta}}$ is the estimated parameter, $\ell(\widehat{\boldsymbol{\theta}})$ is the log-likelihood function evaluated at $\widehat{\boldsymbol{\theta}}$, m is the sample size and p is the number of parameters of the model. A smaller value for AIC or BIC is an indication of a better model. In Table 5, we present AIC and BIC values for the different matrix-variate models presented in Section 4.2, from where, according to both criteria, is concluded that the BS-$t(3)_{13\times 2}$ distribution is the best model.

In order to evaluate the magnitude of the differences between two values of the BIC, the Bayes factor (BF) can be used. The BF allows us to compare M_1 (model considered as correct) to M_2 (model to be contrasted with M_1), which is given by

$$B_{12} = \mathbb{P}(D|\mathrm{M}_1)/\mathbb{P}(D|\mathrm{M}_2), \tag{4.1}$$

where D is the data set assumed to be generated from one of two hypothetical models (M_1 and M_2). Based on (4.1), we can use the approximation

$$2\log(B_{12}) \approx 2\big[\ell(\widehat{\boldsymbol{\theta}}_1) - \ell(\widehat{\boldsymbol{\theta}}_2)\big] - [p_1 - p_2]\log(m), \tag{4.2}$$

where $\ell(\widehat{\boldsymbol{\theta}}_k)$ is the log-likelihood function for the parameter $\boldsymbol{\theta}_k$ under the model M_k evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_k$ and p_k is the dimension of $\boldsymbol{\theta}_k$, for $k = 1, 2$. Note that

Table 3 *ML estimates of the indicated parameter and model for digit 3 data*

N(0, 1)		$t(3)$		$t(8)$		$t(50)$	
$\widehat{\mathbf{M}}_1$		$\widehat{\mathbf{M}}_2$		$\widehat{\mathbf{M}}_3$		$\widehat{\mathbf{M}}_4$	
13.3667	38.4333	14.0576	38.8459	13.9340	38.8275	13.6284	38.6976
19.3667	40.3333	19.8927	40.2032	19.8352	40.3077	19.6380	40.4264
27.4000	39.4667	27.2996	39.0374	27.3968	39.1596	27.5206	39.3787
31.6000	34.4000	30.8704	34.0054	31.0851	34.0686	31.4969	34.2284
29.0333	29.1667	28.2512	28.7918	28.4093	28.8254	28.7728	28.9533
23.9000	26.4333	22.9471	26.2481	23.1013	26.2279	23.5003	26.2710
18.0000	25.5000	17.3756	25.9259	17.4012	25.7506	17.5804	25.4958
22.5000	23.9333	22.0327	24.4572	22.0840	24.3266	22.2579	24.0694
25.6667	21.5333	25.5007	21.8989	25.5180	21.8218	25.5839	21.6507
27.1667	18.3333	27.4817	18.3404	27.4373	18.3328	27.3152	18.3221
24.2000	15.4333	24.5527	15.2791	24.5181	15.3117	24.3837	15.3874
17.6667	14.5000	17.8412	14.5514	17.8616	14.5111	17.8136	14.4816
11.6333	15.2000	12.1077	15.6311	12.0995	15.4718	11.9306	15.2417
$\widehat{\sigma_1^2} = 19.7130$		$\widehat{\sigma_2^2} = 12.0416$		$\widehat{\sigma_3^2} = 13.0665$		$\widehat{\sigma_4^2} = 16.3321$	
BS		BS-$t(3)$		BS-$t(8)$		BS-$t(50)$	
$\widehat{\mathbf{A}}_1$		$\widehat{\mathbf{A}}_2$		$\widehat{\mathbf{A}}_3$		$\widehat{\mathbf{A}}_4$	
0.4365	0.0952	0.3469	0.0611	0.3906	0.0738	0.4275	0.0899
0.2439	0.0837	0.2032	0.0636	0.2272	0.0723	0.2411	0.0812
0.1748	0.0873	0.1411	0.0761	0.1579	0.0825	0.1715	0.0864
0.1347	0.1016	0.1169	0.0874	0.1272	0.0953	0.1334	0.1004
0.1519	0.1413	0.1339	0.1147	0.1442	0.1302	0.1506	0.1394
0.2038	0.1583	0.1468	0.1193	0.1746	0.1419	0.1979	0.1556
0.3977	0.1510	0.2608	0.1203	0.3225	0.1374	0.3822	0.1486
0.2377	0.1557	0.1776	0.1186	0.2088	0.1397	0.2324	0.1530
0.1847	0.1884	0.1489	0.1437	0.1670	0.1669	0.1813	0.1844
0.1864	0.2170	0.1541	0.1738	0.1714	0.1942	0.1836	0.2124
0.2229	0.2760	0.1729	0.2065	0.1975	0.2431	0.2180	0.2702
0.3097	0.3212	0.2270	0.2376	0.2662	0.2796	0.3010	0.3136
0.7378	0.3617	0.4089	0.2570	0.5250	0.3068	0.6936	0.3508
BS		BS-$t(3)$		BS-$t(8)$		BS-$t(50)$	
$\widehat{\mathbf{B}}_1$		$\widehat{\mathbf{B}}_2$		$\widehat{\mathbf{B}}_3$		$\widehat{\mathbf{B}}_4$	
12.1970	38.2601	13.0256	38.7383	12.6100	38.5673	12.2790	38.3511
18.8071	40.1926	19.4314	40.5242	19.0513	40.3918	18.8464	40.2420
26.9874	39.3170	27.3225	39.6436	27.1965	39.4449	27.0338	39.3382
31.3159	34.2235	31.1008	34.4611	31.2372	34.3058	31.3037	34.2357
28.7021	28.8785	28.4134	29.0507	28.5938	28.9357	28.6848	28.8861
23.4136	26.1060	23.1543	26.7615	23.3910	26.3956	23.4309	26.1558
16.6738	25.2126	17.3786	25.7152	17.1638	25.4417	16.7972	25.2529
21.8816	23.6468	22.0528	24.1690	22.0225	23.8762	21.9151	23.6854
25.2363	21.1577	25.4177	21.5491	25.3520	21.3491	25.2618	21.1947
26.7029	17.9115	26.9817	18.0472	26.7753	18.0002	26.7094	17.9350
23.6129	14.8665	24.0705	15.2160	23.8649	15.0828	23.6660	14.9091
16.8562	13.7868	17.4440	14.3519	17.2214	14.1087	16.9387	13.8518
9.0338	14.2620	11.2807	15.3127	10.5798	14.8229	9.3736	14.3765

Table 4 *Interpretation of* $2\log(B_{12})$ *associated with the BF*

$2\log(B_{12})$	Evidence in favor of M_1
<0	Negative (M_2 is accepted)
[0, 2)	Weak
[2, 6)	Positive
[6, 10)	Strong
≥10	Very strong

Table 5 *Values of AIC, BIC and* $2\log(B_{12})$ *between* M_1 (*BS-*$t(3)$) *and* M_2 *for the indicated model with digit 3 data*

Matrix-variate model (M_2)	AIC	BIC	$2\log(B_{12})$	Evidence in favor of M_1
$t(3)$	4474.200	4512.032	1811.809	Very strong
$t(8)$	4475.490	4513.322	1813.099	Very strong
$t(50)$	4531.207	4569.039	1868.816	Very strong
Normal	4592.943	4630.775	1930.552	Very strong
BS-$t(3)$	2627.361	2700.223	–	–
BS-$t(8)$	3414.185	3487.048	786.825	Very strong
BS-$t(50)$	4281.617	4354.479	1654.256	Very strong
BS	4643.942	4716.804	2016.581	Very strong

the approximation given in (4.2) is computed subtracting the BIC value from the model M_2, given by $\text{BIC}_2 = p_2 \log(m) - 2\ell(\widehat{\boldsymbol{\theta}}_2)$, to the BIC value of the model M_1, given by $\text{BIC}_1 = p_1 \log(m) - 2\ell(\widehat{\boldsymbol{\theta}}_1)$.

In general, the BF is informative because it presents ranges in which the level of superiority of a model with respect to another can be quantified. An interpretation of the BF is displayed in Table 4; see Vilca et al. (2011).

Table 5 presents the values of the BF which are useful for comparing the model BS-$t(3)_{13\times 2}$ (model M_1, used as reference) to the rest of models (each one considered as model M_2). Note that, in all the cases, there is a very strong evidence in favor of the model M_1 instead of any of the others. Interesting, the GBS model, that is asymmetrical, fits best than the EC models, which, as it is known, are symmetrical. Therefore, by these three criteria (AIC, BIC and BF), the model BS-$t(3)_{13\times 2}$ is preferred, so that we consider it as the best one for modeling the matrix-variate data of digit 3.

4.4 Statistical analysis

By using the invariance property of the ML estimators, the estimated mean of the handwritten digit 3 can be obtained by replacing the corresponding estimates in the mean shape. Figure 3 (first panel left) shows the digit 3 estimate under the normal, $t(3), t(8)$ and $t(50)$ models. From this figure, it is clear that negligible

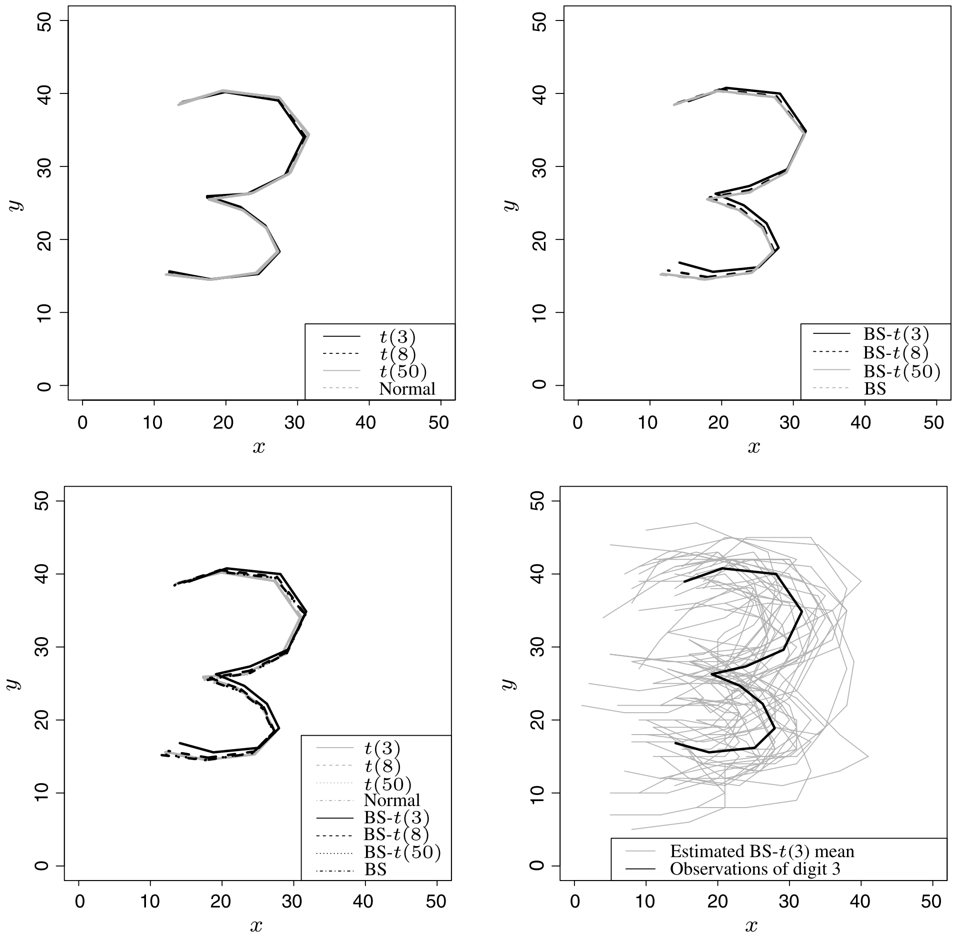

Figure 3 *Plots of the estimated mean shape located at coordinates* (x, y) *for the indicated matrix–variate model and superposition of it on digit* 3 *data.*

differences among the estimates based on symmetric models of the digit 3 shape are observed. However, the graphical comparison provided in Figure 3 (first panel right) for the estimate of the mean digit 3 shape based on the BS, BS-$t(3)$, BS-$t(8)$, and BS-$t(50)$ models, establishes that the corresponding estimate for the best selected model, that is, the BS-$t(3)$ model, is different from the others. Thus, collecting the eight models in Figure 3 (second panel left), a remarkable difference for the estimate of the mean digit 3 shape from the BS-$t(3)$ model is observed. Focusing on the best model, the estimated mean digit 3 shape under the BS-$t(3)_{13\times 2}$ model is obtained by replacing the ML estimates $\widehat{\mathbf{A}}_3$ and $\widehat{\mathbf{B}}_3$ in the expression given by (2.4), reaching the estimated mean with values given in (4.3).

$$\widehat{\mu} = \begin{pmatrix} 15.3762 & 38.9555 \\ 20.6353 & 40.7703 \\ 28.1387 & 39.9878 \\ 31.7388 & 34.8560 \\ 29.1770 & 29.6243 \\ 23.9033 & 27.3330 \\ 19.1515 & 26.2733 \\ 23.0961 & 24.6791 \\ 26.2634 & 22.2163 \\ 27.9433 & 18.8654 \\ 25.1499 & 16.1890 \\ 18.7918 & 15.5673 \\ 14.1099 & 16.8298 \end{pmatrix}. \tag{4.3}$$

Figure 3 (second panel right) displays a graphical plot that superimposes the estimate of the mean digit 3 shape (bold curve) on the observations of the sample (gray curves). From this figure, we detect a clear tendency to enlarge more the upper curve of digit 3 that of the below part and, visibly, a suitable estimate for the mean digit 3 shape is obtained.

5 Conclusions and future work

We proposed a methodology by using matrix-variate Birnbaum–Saunders distributions, which was based on an estimation method for the parameters of matrix-variate Birnbaum–Saunders and Birnbaum–Saunders–Student-t distributions. We evaluated the quality of these estimators by a Monte Carlo study, which showed their good performance. We applied the proposed methodology to real-world matrix-variate data of handwritten characters, comparing some matrix-variate symmetric and asymmetric models, which illustrated its potentiality. Interesting consequences of this work can be implemented in future studies, some of them can avoid complex open problems in shape theory under matrix-variate generalized Birnbaum–Saunders distributions. For instance, it seems that the methodology proposed for landmark data behaves similarly to certain classic invariant distributions based on elliptically contoured models. The proof of this heuristic equivalence for invariant statistics is a shortcut for some of the open problems proposed in Caro-Lopera et al. (2012), which are unsolved until now, because they require a special algebra and group and integration theories involving Hadamard products. Once this is solved, a comparison between two generalized Birnbaum–Saunders populations of landmark data can be a feasible task.

Appendix: Data of handwritten digit 3

This appendix provides the data used in the application, where each of $m = 30$ rows is an observation with $n = 13$ landmarks (x_i, y_i), for $i = 1, \ldots, 13$, in $k = 2$ dimensions. For example, the first coordinate of the first landmark is (9, 27) and so far by rows for the other coordinates.

9	27	12	31	17	36	26	39	34	37	36	33	38	27	35	19	30	15	21	14	21	8	16	6	8	5
17	40	21	38	26	36	27	32	25	28	22	27	19	29	24	25	26	20	28	16	26	13	18	14	15	17
19	38	24	38	29	33	30	29	27	24	21	25	17	26	27	24	30	22	31	19	31	16	27	15	24	15
9	40	15	43	24	41	29	36	24	30	20	26	12	22	20	22	24	20	21	16	18	14	13	12	9	10
14	41	21	42	29	42	35	37	32	33	26	30	16	26	25	26	29	24	33	20	30	16	23	11	16	12
24	39	28	40	35	38	38	35	34	30	29	27	22	24	27	24	29	22	31	19	28	15	20	11	13	12
9	39	15	39	21	40	25	36	23	31	21	27	19	25	21	25	23	24	25	22	22	19	15	17	8	17
8	38	14	41	25	43	29	38	25	33	18	29	8	28	12	27	16	25	18	23	13	21	7	21	1	22
4	34	12	39	22	42	31	36	27	30	23	28	11	25	20	25	22	24	22	22	19	19	13	18	8	18
21	36	25	37	31	36	33	32	32	28	29	25	27	22	29	21	31	20	31	18	28	16	24	16	20	16
14	40	20	39	25	37	27	31	26	28	20	29	16	31	21	28	25	23	28	16	25	13	17	15	13	18
12	40	20	42	30	42	36	33	31	24	23	22	16	23	25	22	31	18	33	13	31	9	24	8	17	8
9	35	17	36	26	34	30	31	26	27	20	25	13	27	19	25	23	21	26	15	22	12	12	12	7	13
17	38	24	39	30	37	34	34	31	28	22	25	16	28	21	26	27	24	30	20	26	15	18	14	10	17
21	35	27	36	36	35	39	28	38	22	34	18	28	19	31	18	33	17	31	15	26	15	20	17	14	20
16	40	20	43	25	39	27	31	24	24	19	21	17	23	19	22	21	21	23	21	22	18	19	16	15	16
15	41	21	45	34	44	40	39	36	35	26	30	16	29	24	25	28	20	31	16	28	14	21	14	12	12
11	42	22	42	32	39	35	34	32	29	25	26	20	27	25	26	31	23	35	19	31	14	21	12	16	15
5	44	15	43	24	41	29	36	22	28	13	28	5	29	14	28	24	26	29	22	26	19	17	17	10	20
14	37	19	39	25	38	28	32	25	26	20	22	14	23	17	23	21	20	23	17	21	15	16	15	11	15
16	35	22	38	30	36	32	29	29	23	23	20	17	20	20	19	24	17	26	14	21	11	16	12	12	15
14	38	17	40	25	42	28	38	27	32	24	28	20	25	23	25	26	24	28	21	24	18	18	17	10	18
7	40	13	43	22	45	31	42	27	38	21	34	13	32	18	31	24	30	27	27	23	23	15	22	6	22
14	35	21	36	26	34	31	30	28	26	25	22	21	18	21	17	22	16	23	15	20	12	13	10	5	10
10	46	17	47	27	43	29	36	26	30	22	29	16	28	20	27	21	25	23	21	21	19	15	20	9	20
18	39	24	42	33	41	38	35	37	30	32	28	28	27	33	22	37	18	41	15	37	13	29	11	21	12
18	38	22	42	30	42	34	36	33	32	29	30	22	28	25	26	28	24	28	20	27	19	22	18	18	18
9	41	17	43	30	40	34	31	30	23	23	19	11	19	15	17	18	13	21	10	17	8	12	7	5	7
8	36	12	42	20	43	25	38	24	35	23	33	21	32	20	31	20	30	20	27	16	25	9	24	2	25
19	41	24	45	33	45	38	38	36	31	28	27	21	23	24	22	26	20	28	17	26	14	20	13	14	11

Acknowledgments

The authors wish to thank the Editor-in-Chief, the Associate Editor and two anonymous referees for their constructive comments on an earlier version of the manuscript, which resulted in this improved version. L. Sánchez gratefully acknowledges financial support from a CONICYT scholarship of the Chilean government for his studies of master in statistics in the University of Valparaíso, Chile. This work was partially supported by project grant 158 from the University of Medellín, Colombia (F. J. Caro-Lopera) and the University of Valparaíso, Chile (V. Leiva). Research work of V. Leiva was partially supported by grant FONDECYT 1120879 from the Chilean government. Research work of F. J. A. Cysneiros and

V. Leiva was partially supported by Capes, CNPq and FACEPE from the Brazilian government.

References

Anderson, G. W., Guionnet, A. and Zeitouni, O. (2009). *An Introduction to Random Matrices*. Cambridge, UK: Cambridge Univ. Press. MR2760897

Azevedo, C., Leiva, V., Athayde, E. and Balakrishnan, N. (2012). Shape and change point analyses of the Birnbaum–Saunders-*t* hazard rate and associated estimation. *Computational Statistics and Data Analysis* **56**, 3887–3897. MR2957839

Balakrishnan, N., Leiva, V., Sanhueza, A. and Cabrera, E. (2009a). Mixture inverse Gaussian distribution and its transformations, moments and applications. *Statistics* **43**, 91–104. MR2499364

Balakrishnan, N., Leiva, V., Sanhueza, A. and Vilca, F. (2009b). Estimation in the Birnbaum–Saunders distribution based on scale-mixture of normals and the EM-algorithm. *Statistics and Operations Research Transactions* **33**, 171–192. MR2643505

Balakrishnan, N., Gupta, R., Kundu, D. Leiva, V. and Sanhueza, A. (2011). On some mixture models based on the Birnbaum–Saunders distribution and associated inference. *Journal of Statistical Planning and Inference* **141**, 2175–2190. MR2775197

Barros, M., Paula, G. A. and Leiva, V. (2009). An R implementation for generalized Birnbaum–Saunders distributions. *Computational Statistics and Data Analysis* **53**, 1511–1528. MR2657109

Birnbaum, Z. W. and Saunders, S. C. (1969a). A new family of life distributions. *Journal of Applied Probability* **6**, 319–327. MR0253493

Birnbaum, Z. W. and Saunders, S. C. (1969b). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability* **6**, 328–347. MR0251807

Brent, R. (1973). *Algorithms for Minimization without Derivatives*. Englewood Cliffs, NJ: Prentice-Hall. MR0339493

Caro-Lopera, F. J., Leiva, V. and Balakrishnan, N. (2012). Connection between the Hadamard and matrix products with an application to a matrix-variate Birnbaum–Saunders distribution. *Journal of Multivariate Analysis* **104**, 126–139. MR2832191

Díaz-García, J. A. and Domínguez-Molina, J. R. (2007). A new family of life distributions for dependent data: Estimation. *Computational Statistics and Data Analysis* **51**, 5927–5939. MR2407689

Díaz-García, J. A., Leiva, V. and Galea, M. (2002). Singular elliptical distribution: Density and applications. *Communications in Statistics: Theory and Methods* **31**, 665–681. MR1905138

Díaz-García, J. A., Galea, M. and Leiva, V. (2003). Influence diagnostics for elliptical multivariate linear regression models. *Communications in Statistics: Theory and Methods* **32**, 625–641. MR1965909

Díaz-García, J. A. and Leiva, V. (2005). A new family of life distributions based on elliptically contoured distributions. *Journal of Statistical Planning and Inference* **128**, 445–457. MR2102769

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Chichester: Wiley. MR1646114

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. London: Chapman. MR1071174

Ferreira, M., Gomes, M. I. and Leiva, V. (2012). On an extreme value version of the Birnbaum–Saunders distribution. *Revstat Statistical Journal* **10**, 181–210. MR2960731

Galea, M., Leiva, V. and Paula, G. (2004). Influence diagnostics in log-Birnbaum–Saunders regression models. *Journal of Applied Statistics* **31**, 1049–1064. MR2100116

Gupta, A. K. and Varga, T. (1993). *Elliptically Contoured Models in Statistics*. Boston: Kluwer. MR1256485

Gupta, A. K. and Varga, T. (1994). A new class of matrix variate elliptically contoured distributions. *Journal of the Italian Statistical Society* **3**, 255–270. MR1279976

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994a). *Continuous Univariate Distributions*. **1**. New York: Wiley. MR1299979

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994b). *Continuous Univariate Distributions*. **2**. New York: Wiley. MR1299979

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions*. **1**. New York: Wiley. MR1788152

Kotz, S., Leiva, V. and Sanhueza, A. (2010). Two new mixture models related to the inverse Gaussian distribution. *Methodology and Computing in Applied Probability* **12**, 199–212. MR2580102

Kundu, D., Balakrishnan, N. and Jamalizadeh, A. (2013). Generalized multivariate Birnbaum–Saunders distributions and related inferential issues. *Journal of Multivariate Analysis* **116**, 230–244. MR3049902

Leiva, V., Barros, M., Paula, G. A. and Sanhueza, A. (2008a). Generalized Birnbaum–Saunders distributions applied to air pollutant concentration. *Environmetrics* **19**, 235–249. MR2420468

Leiva, V., Sanhueza, A., Sen, P. K. and Paula, G. A. (2008b). Random number generators for the generalized Birnbaum–Saunders distribution. *Journal of Statistical Computation and Simulation* **78**, 1105–1118. MR2474416

Leiva, V., Sanhueza, A., Silva, A. and Galea, M. (2008c). A new three-parameter extension of the inverse Gaussian distribution. *Statistics and Probability Letters* **78**, 1266–1273. MR2444316

Leiva, V., Sanhueza, A. and Angulo, J. M. (2009). A length-biased version of the Birnbaum–Saunders distribution with application in water quality. *Stochastic Environmental Research and Risk Assessment* **23**, 299–307. MR2476075

Leiva, V., Sanhueza, A., Kotz, S. and Araneda, N. (2010). A unified mixture model based on the inverse Gaussian distribution. *Pakistan Journal of Statistics* **26**, 445–460. MR2756967

Leiva, V., Soto, G., Cabrera, E. and Cabrera, G. (2011a). New control charts based on the Birnbaum–Saunders distribution and their implementation. *Revista Colombiana de Estadística* **34**, 147–176. MR3075136

Leiva, V., Athayde, E., Azevedo, C. and Marchant, C. (2011b). Modeling wind energy flux by a Birnbaum–Saunders distribution with unknown shift parameter. *Journal of Applied Statistics* **38**, 2819–2838. MR2859837

Leiva, V., Ponce, M. G., Marchant, C. and Bustos, O. (2012). Fatigue statistical distributions useful for modeling diameter and mortality of trees. *Revista Colombiana de Estadística* **35**, 349–367. MR3075150

Leiva, V., Marchant, C., Saulo, H., Aslam, M. and Rojas, F. (2014a). Capability indices for Birnbaum–Saunders processes applied to electronic and food industries. *Journal of Applied Statistics* **41**, 1881–1902. MR3292616

Leiva, V., Rojas, E., Galea, M. and Sanhueza, A. (2014b). Diagnostics in Birnbaum–Saunders accelerated life models with an application to fatigue data. *Applied Stochastic Models in Business and Industry* **30**, 115–131. MR3191346

Leiva, V., Santos-Neto, M., Cysneiros, F. J. A. and Barros, M. (2014d). Birnbaum–Saunders statistical modelling: A new approach. *Statistical Modelling* **14**, 21–48. MR3179546

Marchant, C., Bertin, K., Leiva, V. and Saulo, H. (2013a). Generalized Birnbaum–Saunders kernel density estimators and an analysis of financial data. *Computational Statistics and Data Analysis* **63**, 1–15. MR3040246

Marchant, C., Leiva, V., Cavieres, M. F. and Sanhueza, A. (2013b). Air contaminant statistical distributions with application to PM10 in Santiago, Chile. *Reviews of Environmental Contamination and Toxicology* **223**, 1–31.

Paula, G. A., Leiva, V., Barros, M. and Liu, S. (2012). Robust statistical modeling using the Birnbaum–Saunders-t distribution applied to insurance. *Applied Stochastic Models in Business and Industry* **28**, 16–34. MR2898899

Santana, L., Vilca, F. and Leiva, V. (2011). Influence analysis in skew-Birnbaum–Saunders regression models and applications. *Journal of Applied Statistics* **38**, 1633–1649. MR2819378

Tulino, A. M. and Verdú, S. (2004). *Random Matrix Theory and Wireless Communications*. Hanover, MA: Now Publishers Inc.

Vilca, F., Sanhueza, A., Leiva, V. and Christakos, G. (2010). An extended Birnbaum–Saunders model and its application in the study of environmental quality in Santiago, Chile. *Stochastic Environmental Research and Risk Assessment* **24**, 771–782.

Vilca, F., Santana, L., Leiva, V. and Balakrishnan, N. (2011). Estimation of extreme percentiles in Birnbaum–Saunders distributions. *Computational Statistics and Data Analysis* **55**, 1665–1678. MR2748670

L. Sánchez
Instituto de Estadística
Universidad de Valparaíso
Chile
E-mail: ldaniel9.24@gmail.com

V. Leiva
Facultad de Ingeniería y Ciencias
Universidad Adolfo Ibáñez
Chile
E-mail: victorleivasanchez@gmail.com
URL: www.victorleiva.cl

F. J. Caro-Lopera
Departamento de Ciencias Básicas
Universidad de Medellín
Colombia
E-mail: fjcaro@udem.edu.co

F. J. A. Cysneiros
Departamento de Estatística
Universidade Federal de Pernambuco
Brazil
E-mail: cysneiros@de.ufpe.br