



## Box–Cox elliptical distributions with application

Raúl Alejandro Morán-Vásquez<sup>1</sup> · Silvia L. P. Ferrari<sup>2</sup> 

Received: 21 May 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

### Abstract

We propose and study the class of Box–Cox elliptical distributions. It provides alternative distributions for modeling multivariate positive, marginally skewed and possibly heavy-tailed data. This new class of distributions has as a special case the class of log-elliptical distributions, and reduces to the Box–Cox symmetric class of distributions in the univariate setting. The parameters are interpretable in terms of quantiles and relative dispersions of the marginal distributions and of associations between pairs of variables. The relation between the scale parameters and quantiles makes the Box–Cox elliptical distributions attractive for regression modeling purposes. Applications to data on vitamin intake are presented and discussed.

**Keywords** Box–Cox symmetric distributions · Box–Cox transformation · Elliptical distribution · Gibbs sampling · Truncated distribution

### 1 Introduction

Multivariate positive data are frequently found in empirical studies. The statistical analysis of such data often relies on the multivariate normal distribution assumptions, ignoring characteristics of the data, namely the positive support and possible skewness and presence of outlying observations. Improvements for accomodating outliers may be achieved by replacing the multivariate normal distribution by a heavy-tailed distribution in the elliptical class of distributions, such as the multivariate  $t$  distribution (Lange et al. 1989). Further improvement may be achieved by the use of log-skew-elliptical distributions (Marchenko and Genton 2010), which are multivariate distributions with support in  $\mathbb{R}_+^p$  and accommodate heavy-tailed distributions. An alternative

---

✉ Silvia L. P. Ferrari  
[silviaferrari.usp@gmail.com](mailto:silviaferrari.usp@gmail.com)

Raúl Alejandro Morán-Vásquez  
[raul.moran@udea.edu.co](mailto:raul.moran@udea.edu.co)

<sup>1</sup> Institute of Mathematics, University of Antioquia, Medellín, Colombia

<sup>2</sup> Department of Statistics, University of São Paulo, São Paulo, Brazil

methodology for modeling multivariate positive data uses a Box–Cox transformation in each component of the vector of observations. In this approach one assumes that the vector of transformed observations follows a multivariate normal or an elliptical distribution (Quiroz et al. 1996). This assumption implies a theoretical shortcoming because the support of the transformed vector of observations is not necessarily  $\mathbb{R}^p$ . Moreover, in this approach the model parameters are interpretable only in terms of characteristics of the transformed observations (not the original variables of interest). In the univariate case, Ferrari and Fumes (2017) overcome these shortcomings by proposing the class of Box–Cox symmetric distributions. This class includes several alternative distributions, such as the Box–Cox Cole–Green (Stasinopoulos et al. 2008), Box–Cox  $t$  (Rigby and Stasinopoulos 2006), Box–Cox power exponential (Rigby and Stasinopoulos 2004; Voudouris et al. 2012) distributions, and a new distribution, the Box–Cox slash distribution, for modeling univariate positive, skewed, possibly heavy-tailed data.

In the present paper, we focus on the problem of constructing a class of multivariate distributions with support in  $\mathbb{R}_+^p$  in such a way that the marginal distributions have properties similar to those of the Box–Cox symmetric distributions, the parameters are interpretable and association among variables is controlled by association parameters. We name the proposed class of distributions the Box–Cox elliptical class of distributions. It has the log-elliptical class of distributions (Fang et al. 1990) as a special subclass and reduces to the Box–Cox symmetric class of distributions in the univariate setting. The construction of the new class is performed through an extension of the Box–Cox transformation and involves another new class of distributions defined in this paper, the class of truncated elliptical distributions. The parameters of the Box–Cox elliptical distributions are interpretable as characteristics of the original variables (not the transformed variables). Some parameters are related to quantiles of the marginal distributions, which makes the Box–Cox elliptical distributions attractive for regression modeling purposes. Several properties of the proposed distributions are derived. In particular, some properties of the log-elliptical distributions that are not available in the literature are direct consequences of properties of Box–Cox elliptical distributions stated in this paper. The flexibility of the proposed distributions for modeling multivariate positive, asymmetric data with or without the presence of outlying observations is illustrated through an analysis of real data on vitamin intake by older people.

The paper is organized as follows. In Sect. 2 we define the truncated elliptical distributions and present some properties. In Sect. 3 we define the family of the extended Box–Cox transformations, we use it to define the class of Box–Cox elliptical distributions, and we state several properties. In Sect. 4 we give interpretation for the parameters and show the relation between some parameters and quantiles of the marginal distributions. In Sect. 5 we focus on maximum likelihood estimation and present simulation studies. In Sect. 6 we present applications to real data. Finally, Sect. 7 closes the paper with concluding remarks. Technical proofs are presented in the “Appendix”.

## 2 The class of the truncated elliptical distributions

In this section, we define the class of the truncated elliptical distributions. It is needed for the definition and study of the class of the Box–Cox elliptical distributions, which

is the focus of this paper. Previous works directly related to the truncated elliptical class of distributions are e.g. Arellano-Valle et al. (2006) and Kim (2010).

We denote vectors and their components with lowercase Greek letters in bold and normal fonts, respectively. For instance, if  $\boldsymbol{\xi} \in \mathbb{R}^p$ , then  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_p)'$ . Additionally,  $\boldsymbol{\xi}_{-k} \in \mathbb{R}^{p-1}$ ,  $k = 1, \dots, p$ , is the sub-vector obtained from  $\boldsymbol{\xi}$  by excluding its  $k$ th component. Similar notations are used for random vectors, but we use capital Roman letters. Matrices are denoted by capital Greek letters in boldface and their entries in lowercase normal font Greek letters. For example, if  $\boldsymbol{\Delta}(p \times q)$  is a matrix with components in  $\mathbb{R}$ , then  $\boldsymbol{\Delta} = (\delta_{jk})_{p \times q}$ . If  $\boldsymbol{\Delta}$  is a symmetric matrix, the notation  $\boldsymbol{\Delta} > 0$  means that  $\boldsymbol{\Delta}$  is positive definite. If  $\boldsymbol{\Delta}(p \times p) > 0$ , then  $\boldsymbol{\Delta}_{-k,k} \in \mathbb{R}^{p-1}$  is the sub-vector obtained by deleting the  $k$ th component of the  $k$ th column of  $\boldsymbol{\Delta}$ ;  $\boldsymbol{\Delta}_{k,-k} = \boldsymbol{\Delta}'_{-k,k}$ ; and  $\boldsymbol{\Delta}_{-k,-k} > 0$  is the sub-matrix obtained by excluding the  $k$ th row and the  $k$ th column of  $\boldsymbol{\Delta}$ .

The elliptical distributions have been extensively studied in the statistical literature and applied in different fields; see Fang et al. (1990), Gupta et al. (2013) and references therein. From now on, whenever we say that a random vector has an elliptical distribution we assume that its probability density function (PDF) exists.

**Definition 1** The random vector  $\mathbf{X} \in \mathbb{R}^p$  has an elliptical distribution with location vector  $\boldsymbol{\mu} \in \mathbb{R}^p$  and dispersion matrix  $\boldsymbol{\Sigma}(p \times p) > 0$ , if its PDF is

$$f_{\mathbf{X}}(\mathbf{x}) = c_p \det(\boldsymbol{\Sigma})^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})), \quad \mathbf{x} \in \mathbb{R}^p. \quad (1)$$

The function  $g$ , called density generating function (DGF), is such that  $g(u) \geq 0$ , for all  $u \geq 0$ , and  $\int_0^\infty r^{p-1} g(r^2) dr < \infty$ . The normalizing constant  $c_p$  is

$$c_p = \frac{\Gamma(p/2)}{2\pi^{p/2}} \left( \int_0^\infty r^{p-1} g(r^2) dr \right)^{-1}.$$

We write  $\mathbf{X} \sim \text{E}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$ .

The univariate case of Definition 1 corresponds to a random variable  $X$  having a symmetric distribution with location parameter  $\mu \in \mathbb{R}$ , scale parameter  $\sigma > 0$  and DGF  $g$ , and we write  $X \sim \text{E}\ell_1(\mu, \sigma^2; g)$ . A detailed study about elliptical distributions can be found in Fang et al. (1990).

**Definition 2** Let  $B \subseteq \mathbb{R}^p$  be a measurable set. The random vector  $\mathbf{W} \in B$  has a truncated elliptical distribution with support  $B$  and parameters  $\boldsymbol{\mu} \in \mathbb{R}^p$  and  $\boldsymbol{\Sigma}(p \times p) > 0$ , DGF  $g$ , and we write  $\mathbf{W} \sim \text{TE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; B; g)$ , if its PDF is

$$f_{\mathbf{W}}(\mathbf{w}) = \frac{g((\mathbf{w} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{w} - \boldsymbol{\mu}))}{\int_B g((\mathbf{w} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{w} - \boldsymbol{\mu})) d\mathbf{w}}, \quad \mathbf{w} \in B, \quad (2)$$

where  $g$  is such that  $g(u) \geq 0$ , for all  $u \geq 0$ , and  $\int_0^\infty t^{p-1} g(t^2) dt < \infty$ .

If  $B = \mathbb{R}^p$  in (2), we arrive at PDF (1).

The univariate case of Definition 2 corresponds to a random variable, say  $W$ , with a truncated symmetric distribution with support  $B \subseteq \mathbb{R}$ , parameters  $\mu \in \mathbb{R}$  and  $\sigma^2 > 0$ , DGF  $g$ , and we write  $W \sim \text{TE}\ell_1(\mu, \sigma^2; B; g)$ .

Each member of the class of the truncated elliptical distributions is characterized by the DGF  $g$ . Two notable special cases are the multivariate truncated normal and truncated  $t$  distributions, which correspond to the DGF  $g(u) \propto \exp(-u/2)$  and  $g(u) \propto (1 + u/\tau)^{-(\tau+p)/2}$ , with  $\tau > 0$ , respectively. Other special cases include the following multivariate distributions: truncated power exponential ( $g(u) \propto \exp(-u^\beta/2)$ ,  $\beta > 0$ ), truncated slash ( $g(u) \propto \int_0^1 t^{p+q-1} \exp(-ut^2/2) dt$ ,  $q > 0$ ), and truncated scale mixture of normal distributions ( $g(u) \propto \int_0^\infty t^{p/2} \exp(-ut/2) dH(t)$ ,  $u \geq 0$ ,  $H$  being a cumulative distribution function (CDF) on  $(0, \infty)$ ). The DGF  $g$  may include extra parameters in PDF (2). For instance, the multivariate truncated  $t$  distribution has the degrees of freedom parameter  $\tau$ , that controls the tail behaviour. The multivariate truncated normal distribution is a limiting case of the multivariate truncated  $t$  distribution when  $\tau \rightarrow \infty$ . Some studies on multivariate truncated normal distributions are found in Birnbaum and Meyer (1953), Tallis (1961, 1963, 1965), Horrace (2005) and Manjunath and Wilhelm (2012). The multivariate truncated  $t$  distribution with rectangular support is considered in Ho et al. (2012).

Let  $W \sim \text{TE}\ell_1(\mu, \sigma^2; (a, b); g)$ . The CDF of  $W$  is given by

$$F_W(w) = \frac{F_Z\left(\frac{w-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right)}{F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right)}, \quad w \in (a, b), \quad (3)$$

where  $F_Z$  is the CDF of a random variable  $Z$  having a standard symmetric distribution,  $Z \sim \text{E}\ell_1(0, 1; g)$ . Equation (3) is also valid when  $a \rightarrow -\infty$  and/or  $b \rightarrow \infty$ . In this case, we have  $F_Z((a - \mu)/\sigma) \rightarrow 0$  and/or  $F_Z((b - \mu)/\sigma) \rightarrow 1$ .

Let  $R = I_1 \times \cdots \times I_p$  be a rectangle in  $\mathbb{R}^p$ , where  $I_1, \dots, I_p$  are intervals in  $\mathbb{R}$  (finite or infinite). With no loss of generality, assume that  $I_k = (a_k, b_k)$ ,  $k = 1, \dots, p$ .

**Theorem 1** *If  $\mathbf{W} \sim \text{TE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; R; g)$ , then  $W_k | \mathbf{W}_{-k} \sim \text{TE}\ell_1(\mu_{k,-k}, \sigma_{k,-k}^2; (a_k, b_k); g_{k,-k})$ ,  $k = 1, \dots, p$ , where  $\mu_{k,-k} = \mu_k + \boldsymbol{\Sigma}_{k,-k} \boldsymbol{\Sigma}_{-k,-k}^{-1} (\mathbf{w}_{-k} - \boldsymbol{\mu}_{-k})$ ,  $\sigma_{k,-k}^2 = \sigma_{kk} - \boldsymbol{\Sigma}_{k,-k} \boldsymbol{\Sigma}_{-k,-k}^{-1} \boldsymbol{\Sigma}_{-k,k}$  and  $g_{k,-k}(u) = g(u + q(\mathbf{w}_{-k}))$ , with  $q(\mathbf{w}_{-k}) = (\mathbf{w}_{-k} - \boldsymbol{\mu}_{-k})' \boldsymbol{\Sigma}_{-k,-k}^{-1} (\mathbf{w}_{-k} - \boldsymbol{\mu}_{-k})$ .*

**Proof** See “Appendix A”. □

Theorem 1 states that if a random vector  $\mathbf{W}$  has a truncated elliptical distribution with its support being a rectangle in  $\mathbb{R}^p$ , then the conditional distribution of  $W_k$  given  $\mathbf{W}_{-k}$  is truncated symmetric with the same support of  $W_k$ . This fact is useful for obtaining the complete conditional distributions, from which random samples from (3) may be obtained using the inverse transformation method. This allows us to propose Algorithm 1 to generate random samples from the random vector  $\mathbf{W} \sim \text{TE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; R; g)$ . We construct a Markov chain by sampling from the complete conditional distributions

of  $W_k | \mathbf{W}_{-k}$ ,  $k = 1, \dots, p$ , given in Theorem 1. Let  $\mathbf{w}^{(j)}$  be a sample generated in the  $j$ th iteration,  $j = 1, \dots, n$ .

### Algorithm 1

1. Choose a starting value  $\mathbf{w}^{(0)}$  of the Markov chain.
2. Generate a random variable  $u$  from a uniform distribution  $U(0, 1)$ .
3. In each cycle  $j = 1, \dots, n$ , apply the inverse transformation method using (3) to compute

$$w_{k.-k}^{(j)} = \mu_{k.-k}^{(j)} + \sigma_{k.-k}^{(j)} F_{Z_k}^{-1} \left[ u \left\{ F_{Z_k} \left( \frac{b_k - \mu_{k.-k}^{(j)}}{\sigma_{k.-k}^{(j)}} \right) - F_{Z_k} \left( \frac{a_k - \mu_{k.-k}^{(j)}}{\sigma_{k.-k}^{(j)}} \right) \right\} + F_{Z_k} \left( \frac{a_k - \mu_{k.-k}^{(j)}}{\sigma_{k.-k}^{(j)}} \right) \right],$$

where  $Z_k \sim El_1(0, 1; g_{k.-k})$ , for  $k = 1, \dots, p$ . This is the sampled value from the conditional distribution of

$$w_k^{(j)} | w_1^{(j)}, \dots, w_{k-1}^{(j)}, w_{k+1}^{(j-1)}, \dots, w_p^{(j-1)}, \quad k = 1, \dots, p.$$

## 3 The class of the Box–Cox elliptical distributions

In this section, we define the class of the Box–Cox elliptical distributions and state several properties. First, we define the family of the extended Box–Cox transformations, which is a generalization of multivariate Box–Cox transformations given in Quiroz et al. (1996, Eq. 1.1, 1.2). Using this new family of transformations, we define the class of the Box–Cox elliptical distributions. We then present various properties of these distributions regarding a characterization through truncated elliptical distributions with rectangular support, marginal and conditional distributions, independence, and mixed moments. Some of these properties will be needed for interpreting the parameters of the Box–Cox elliptical distributions (see Sect. 4).

For each  $\xi \in \mathbb{R}^p$ , let  $\mathbf{D}_\xi$  be a diagonal matrix with diagonal elements  $\xi$ , i.e.,  $\mathbf{D}_\xi = \text{diag}\{\xi_1, \dots, \xi_p\}$ . Let  $R(\xi) = I(\xi_1) \times \dots \times I(\xi_p)$  be a rectangle in  $\mathbb{R}^p$ , where

$$I(\xi_k) = \begin{cases} (-1/\xi_k, \infty), & \xi_k > 0, \\ (-\infty, -1/\xi_k), & \xi_k < 0, \\ (-\infty, \infty), & \xi_k = 0, \end{cases} \quad (4)$$

for  $k = 1, \dots, p$ .

**Definition 3** Let  $\lambda \in \mathbb{R}^p$  and  $\mu \in \mathbb{R}_+^p$ . The extended Box–Cox transformation is defined by  $T_{\lambda, \mu} : \mathbb{R}_+^p \rightarrow R(\lambda)$  for the random vector  $\mathbf{Y} \in \mathbb{R}_+^p$  as  $T_{\lambda, \mu}(\mathbf{Y}) = \mathbf{W}$ , where  $\mathbf{W}$  is the  $p$ -dimensional vector with  $k$ th element given by

$$W_k = \begin{cases} \frac{(Y_k/\mu_k)^{\lambda_k} - 1}{\lambda_k}, & \lambda_k \neq 0, \\ \log(Y_k/\mu_k), & \lambda_k = 0, \end{cases} \quad (5)$$

for  $k = 1, \dots, p$ .

From Definition 3 we have that  $\mu_k$  is a scale parameter for  $Y_k$ , for  $k = 1, \dots, p$ . If  $\boldsymbol{\mu} = \mathbf{1} = (1, \dots, 1)'$  in (5) we obtain the multivariate Box–Cox transformation (Quiroz et al. 1996, Eq. 1.1, 1.2)). Also,  $T_{\lambda, \boldsymbol{\mu}}(\mathbf{Y}) \rightarrow T_{\mathbf{0}, \boldsymbol{\mu}}(\mathbf{Y})$  when  $\lambda \rightarrow \mathbf{0} = (0, \dots, 0)'$ . Moreover, if  $\boldsymbol{\alpha} \in \mathbb{R}_+^p$ , then  $T_{\lambda, \boldsymbol{\mu}}(\mathbf{D}_{\boldsymbol{\alpha}}\mathbf{Y}) = T_{\lambda, \mathbf{D}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\mu}}(\mathbf{Y})$ . If  $\boldsymbol{\beta} \in R(\lambda)$  and  $\boldsymbol{\gamma} = \mathbf{1} + \mathbf{D}_{\lambda}\boldsymbol{\beta}$ , then  $\mathbf{D}_{\boldsymbol{\gamma}}^{-1}(T_{\lambda, \boldsymbol{\mu}}(\mathbf{Y}) - \boldsymbol{\beta}) = T_{\lambda, \boldsymbol{\delta}}(\mathbf{Y})$ , where  $\boldsymbol{\delta} = T_{\lambda, \boldsymbol{\mu}}^{-1}(\boldsymbol{\beta})$ . These facts allow us to derive various properties of the Box–Cox elliptical distributions.

**Definition 4** We say that the random vector  $\mathbf{Y} \in \mathbb{R}_+^p$  has a Box–Cox elliptical distribution with parameters  $\boldsymbol{\mu} \in \mathbb{R}_+^p$ ,  $\lambda \in \mathbb{R}^p$ ,  $\boldsymbol{\Sigma}(p \times p) > 0$  and DGF  $g$  if  $T_{\lambda, \boldsymbol{\mu}}(\mathbf{Y}) \sim \text{TE}\ell_p(\mathbf{0}, \boldsymbol{\Sigma}; R(\lambda); g)$ , and we write  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \lambda, \boldsymbol{\Sigma}; g)$ .

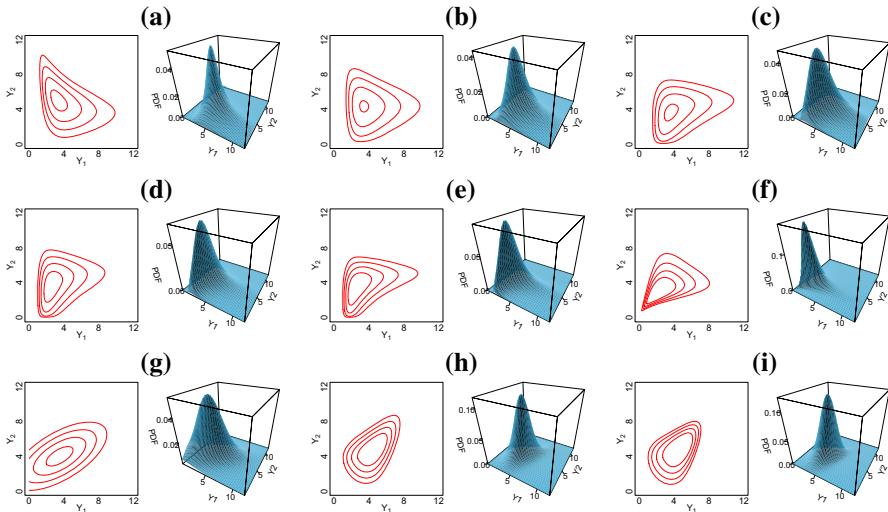
Equivalently,  $\mathbf{W} \sim \text{TE}\ell_p(\mathbf{0}, \boldsymbol{\Sigma}; R(\lambda); g)$  if  $T_{\lambda, \boldsymbol{\mu}}^{-1}(\mathbf{W}) \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \lambda, \boldsymbol{\Sigma}; g)$ . If  $\lambda = \mathbf{0}$  in Definition 4, then  $\mathbf{Y}$  follows a log-elliptical distribution with parameters  $\boldsymbol{\mu} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\Sigma}(p \times p) > 0$  and DGF  $g$  (Fang et al. 1990), and we write  $\mathbf{Y} \sim \text{LE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$ .

From Definition 4 we have that the PDF of  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \lambda, \boldsymbol{\Sigma}; g)$  is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) \prod_{k=1}^p \frac{y_k^{\lambda_k-1}}{\mu_k^{\lambda_k}}}{\int_{R(\lambda)} g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) d\mathbf{w}}, \quad \mathbf{w} = T_{\lambda, \boldsymbol{\mu}}(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}_+^p. \quad (6)$$

The case  $p = 1$  in (6) corresponds to the PDF of a positive random variable  $Y$  with a Box–Cox symmetric distribution with parameters  $\mu > 0$ ,  $\sigma > 0$ ,  $\lambda \in \mathbb{R}$  and DGF  $g$  (Ferrari and Fumes 2017), denoted by  $Y \sim \text{BCS}(\mu, \sigma, \lambda; g)$ . From Definition 4, it is clear that each member of the class of the truncated elliptical distributions has its corresponding member in the class of the Box–Cox elliptical distributions, which is identified by its DGF  $g$ . Hence, by replacing  $g(u) \propto \exp(-u/2)$ ,  $u \geq 0$ , in (6) we obtain the PDF of a random vector  $\mathbf{Y} \in \mathbb{R}_+^p$  with a multivariate Box–Cox normal distribution with parameters  $\boldsymbol{\mu} \in \mathbb{R}_+^p$ ,  $\lambda \in \mathbb{R}^p$  and  $\boldsymbol{\Sigma}(p \times p) > 0$ , denoted by  $\mathbf{Y} \sim \text{BCN}_p(\boldsymbol{\mu}, \lambda, \boldsymbol{\Sigma})$ . When  $g(u) \propto (1 + u/\tau)^{-(\tau+p)/2}$ ,  $\tau > 0$ ,  $u \geq 0$ , in (6) we have the PDF of a random vector  $\mathbf{Y} \in \mathbb{R}_+^p$  with a multivariate Box–Cox  $t$  distribution with parameters  $\boldsymbol{\mu} \in \mathbb{R}_+^p$ ,  $\lambda \in \mathbb{R}^p$ ,  $\boldsymbol{\Sigma}(p \times p) > 0$  and  $\tau > 0$  degrees of freedom, denoted by  $\mathbf{Y} \sim \text{BCt}_p(\boldsymbol{\mu}, \lambda, \boldsymbol{\Sigma}; \tau)$ . In these cases, when  $\lambda = \mathbf{0}$ , we get the PDF of  $\mathbf{Y} \in \mathbb{R}_+^p$  with multivariate log-normal and log- $t$  distributions, denoted by  $\mathbf{Y} \sim \text{LN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  and  $\mathbf{Y} \sim \text{Lt}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \tau)$ , respectively. As expected, the multivariate Box–Cox normal distribution is a limiting case of the multivariate Box–Cox  $t$  distribution as  $\tau \rightarrow \infty$ . Other members of the class of the Box–Cox elliptical distributions include the multivariate Box–Cox power exponential distribution, the multivariate Box–Cox slash distribution, and the multivariate Box–Cox scale mixture of normal distributions.

Figure 1 shows plots of the PDF of  $\mathbf{Y} \sim \text{BCt}_2(\boldsymbol{\mu}, \lambda, \boldsymbol{\Sigma}; \tau)$  for different parameter values. The legend indicates the values of all the parameters considered in the first plot



**Fig. 1** Contour plots at levels 0.04, 0.02, 0.01, 0.005 and joint PDF of  $\mathbf{Y} \sim \text{BCt}_2(\boldsymbol{\mu}, \boldsymbol{\lambda}; \boldsymbol{\Sigma}; \tau)$ , where **a**  $\mu_1 = 5, \mu_2 = 4, \lambda_1 = -1, \lambda_2 = 1.5, \sigma_{11} = 0.5, \sigma_{22} = 0.3, \sigma_{12} = -0.2, \tau = 3$ , **b**  $\sigma_{12} = 0$ , **c**  $\sigma_{12} = 0.2$ , **d**  $\mu_1 = 3.5, \mathbf{e} \sigma_{22} = 0.2, \mathbf{f} \lambda_2 = -1.5, \mathbf{g} \lambda_1 = \lambda_2 = 1, \mathbf{h} \lambda_2 = 2, \mathbf{i} \tau = 10$

and the value of the parameter that is changed from a plot to the next (in alphabetical order). Note that the parameter  $\sigma_{12}$  impacts the association between the marginal distributions of  $Y_1$  and  $Y_2$  (Fig. 1a–c), ranging from negative association ( $\sigma_{12} < 0$ ) to positive association ( $\sigma_{12} > 0$ ). The parameter  $\mu_1$  affects the scale of the marginal distribution of  $Y_1$  (Fig. 1c, d). The parameter  $\sigma_{22}$  influences the dispersion of the marginal distribution of  $Y_2$  (Fig. 1d, e). The parameters  $\lambda_1$  and  $\lambda_2$  control the skewness of the respective marginal distribution of  $Y_1$  and  $Y_2$  (Fig. 1e–h). In Fig. 1g, for which  $\lambda_1 = \lambda_2 = 1$ , it is clear that the contour lines are (truncated) ellipses; this fact is stated in item 3 of Theorem 3. Additionally, as the degrees of freedom parameter grows, the contour lines corresponding to the bivariate Box–Cox  $t$  distributions tend to the contour lines of bivariate Box–Cox normal distributions. Moreover, the tails of the Box–Cox  $t$  distributions seem to be heavier for smaller values of  $\tau$  (Fig. 1h, i).

Definition 4 characterizes the Box–Cox elliptical distributions from truncated elliptical distributions with support in  $R(\lambda)$  and parameter  $\boldsymbol{\mu} = \mathbf{0}$ . In Theorem 2, we present a characterization of Box–Cox elliptical distributions from truncated elliptical distributions with support in  $R(\lambda)$  and parameter  $\boldsymbol{\mu} = \boldsymbol{\xi}$ .

**Theorem 2** *Let  $\boldsymbol{\mu} \in \mathbb{R}_+^p, \boldsymbol{\lambda} \in \mathbb{R}^p, \boldsymbol{\xi} \in R(\lambda), \boldsymbol{\alpha} = \mathbf{1} + \mathbf{D}_\lambda \boldsymbol{\xi}$  and  $\boldsymbol{\Sigma}(p \times p) > 0$ . Then,  $T_{\lambda, \boldsymbol{\mu}}(\mathbf{Y}) \sim TE\ell_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}; R(\lambda); g)$  if and only if  $\mathbf{Y} \sim BCE\ell_p(\boldsymbol{\delta}, \boldsymbol{\lambda}, \mathbf{D}_\alpha^{-1} \boldsymbol{\Sigma} \mathbf{D}_\alpha^{-1}; g)$ , where  $\boldsymbol{\delta} = T_{\lambda, \boldsymbol{\mu}}^{-1}(\boldsymbol{\xi})$ .*

**Proof** See “Appendix B”. □

In Theorem 3 we state various distributional results concerning the Box–Cox elliptical distributions. Items 1 and 2 consider some transformations of Box–Cox elliptical random vectors, and item 3 states that the class of the truncated elliptical distributions

with support in  $\mathbb{R}_+^p$  and parameter  $\boldsymbol{\mu} \in \mathbb{R}_+^p$  is obtained from the class of the Box–Cox elliptical distributions.

**Theorem 3** Let  $\boldsymbol{\mu} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\lambda} \in \mathbb{R}^p$ ,  $\boldsymbol{\Sigma}(p \times p) > 0$  and  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ .

1. If  $\boldsymbol{\alpha} \in \mathbb{R}_+^p$ , then  $\mathbf{D}_{\boldsymbol{\alpha}}\mathbf{Y} \sim \text{BCE}\ell_p(\mathbf{D}_{\boldsymbol{\alpha}}\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ .
2. If  $\boldsymbol{\beta} \in \mathbb{R}^p \setminus \mathbf{0}$  and  $\mathbf{U} \in \mathbb{R}_+^p$  is the random vector with components  $U_k = (Y_k/\mu_k)^{\beta_k}$ ,  $k = 1, \dots, p$ , then  $\mathbf{U} \sim \text{BCE}\ell_p(\mathbf{1}, \mathbf{D}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\lambda}, \mathbf{D}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{\beta}}; g)$ .
3. If  $\boldsymbol{\lambda} = \mathbf{1}$ , then  $\mathbf{Y} \sim \text{TE}\ell_p(\boldsymbol{\mu}, \mathbf{D}_{\boldsymbol{\mu}}\boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{\mu}}; \mathbb{R}_+^p; g)$ .

**Proof** See “Appendix C”. □

In order to state results on marginal and conditional distributions let us consider the partitions of  $\mathbf{Y} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\mu} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\lambda} \in \mathbb{R}^p$  and  $\boldsymbol{\Sigma}(p \times p) > 0$  as

$$\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2)', \quad \boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2)', \quad \boldsymbol{\lambda} = (\boldsymbol{\lambda}'_1, \boldsymbol{\lambda}'_2)', \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \quad (7)$$

with  $\mathbf{Y}_1 \in \mathbb{R}_+^r$ ,  $\mathbf{Y}_2 \in \mathbb{R}_+^{p-r}$ ,  $\boldsymbol{\mu}_1 \in \mathbb{R}_+^r$ ,  $\boldsymbol{\mu}_2 \in \mathbb{R}_+^{p-r}$ ,  $\boldsymbol{\lambda}_1 \in \mathbb{R}^r$ ,  $\boldsymbol{\lambda}_2 \in \mathbb{R}^{p-r}$ ,  $\boldsymbol{\Sigma}_{11}(r \times r) > 0$ ,  $\boldsymbol{\Sigma}_{22}((p-r) \times (p-r)) > 0$  and  $\boldsymbol{\Sigma}_{12}(r \times (p-r))$  such that  $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}'_{12}$ . The rectangle  $R(\boldsymbol{\lambda})$  can be written as  $R(\boldsymbol{\lambda}) = R(\boldsymbol{\lambda}_1) \times R(\boldsymbol{\lambda}_2)$ , where  $R(\boldsymbol{\lambda}_1) = I(\lambda_1) \times \dots \times I(\lambda_r) \in \mathbb{R}^r$  and  $R(\boldsymbol{\lambda}_2) = I(\lambda_{r+1}) \times \dots \times I(\lambda_p) \in \mathbb{R}^{p-r}$ .

Let  $\mathbf{Y} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\mu} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\lambda} \in \mathbb{R}^p$ ,  $\boldsymbol{\Sigma}(p \times p) > 0$  partitioned as in (7) and such that  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ . The marginal PDF of  $\mathbf{Y}_1$  is given by

$$f_{Y_1}(\mathbf{y}_1) = \frac{\left\{ \int_{R(\boldsymbol{\lambda}_2)} g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) d\mathbf{w}_2 \right\} \prod_{k=1}^r \frac{y_k^{\lambda_k-1}}{\mu_k^{\lambda_k}}}{\int_{R(\boldsymbol{\lambda})} g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) d\mathbf{w}}, \quad \mathbf{y}_1 \in \mathbb{R}_+^r, \quad (8)$$

where  $\mathbf{w} = (\mathbf{w}'_1, \mathbf{w}'_2)'$ , with  $\mathbf{w}_1 = T_{\lambda_1, \mu_1}(\mathbf{y}_1)$  and  $\mathbf{w}_2 = T_{\lambda_2, \mu_2}(\mathbf{y}_2)$ . Clearly, the marginal PDF (8) is not necessarily of the form (6). This form is possible when  $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ , i.e., when the matrix  $\boldsymbol{\Sigma}(p \times p) > 0$  is block-diagonal. In Theorem 4 this fact is stated.

**Theorem 4** Let  $\mathbf{Y} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\mu} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\lambda} \in \mathbb{R}^p$ ,  $\boldsymbol{\Sigma}(p \times p) > 0$  partitioned as in (7) and such that  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ . If  $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ , then  $\mathbf{Y}_1 \sim \text{BCE}\ell_r(\boldsymbol{\mu}_1, \boldsymbol{\lambda}_1, \boldsymbol{\Sigma}_{11}; g_1)$ , where

$$g_1(u) = \int_{T(R(\boldsymbol{\lambda}_2))} g(u + \mathbf{s}' \mathbf{s}) ds, \quad u \geq 0,$$

with  $T : \mathbb{R}^{p-r} \rightarrow \mathbb{R}^{p-r}$  being the transformation  $T(\mathbf{x}) = \boldsymbol{\Sigma}_{22}^{-1/2} \mathbf{x}$ .

**Proof** See “Appendix D”. □

When the matrix  $\boldsymbol{\Sigma}(p \times p) > 0$  is a diagonal matrix, all the marginal distributions are Box–Cox symmetric distributions. This fact is stated in Corollary 1.

**Corollary 1** Let  $\mu \in \mathbb{R}_+^p$ ,  $\lambda \in \mathbb{R}^p$ ,  $\Sigma = \text{diag}\{\sigma_{11}, \dots, \sigma_{pp}\} > 0$  and  $\mathbf{Y} \sim \text{BCE}\ell_p(\mu, \lambda, \Sigma; g)$ . Then,  $Y_k \sim \text{BCS}(\mu_k, \sqrt{\sigma_{kk}}, \lambda_k; g_k)$ ,  $k = 1, \dots, p$ , where

$$g_k(u) = \int_{R(\Sigma_{-k,-k}^{1/2} \lambda_{-k})} g(u + s's) \, ds, \quad u \geq 0.$$

**Proof** Simply let  $\mathbf{Y}_1 = Y_k$ ,  $\mathbf{Y}_2 = \mathbf{Y}_{-k}$ ,  $\mu_1 = \mu_k$ ,  $\mu_2 = \mu_{-k}$ ,  $\lambda_1 = \lambda_k$ ,  $\lambda_2 = \lambda_{-k}$ ,  $\Sigma_{11} = \sigma_{kk}$  and  $\Sigma_{22} = \Sigma_{-k,-k}$ ,  $k = 1, \dots, p$ , in Theorem 4.  $\square$

In Theorem 4 we stated that if  $\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2)' \sim \text{BCE}\ell_p(\mu, \lambda, \Sigma; g)$ , then the sub-vector  $\mathbf{Y}_1$  has a Box–Cox elliptical distribution if  $\Sigma_{12} = \mathbf{0}$ . Note that  $\mathbf{Y}_1$  has a distribution in the Box–Cox elliptical class but not necessarily with the same parent distribution as  $\mathbf{Y}$  (e.g. normal,  $t$ , power exponential). The condition in Theorem 4, although sufficient, is not necessary for the subclass of the log-elliptical distributions. Indeed, if  $\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2)' \sim \text{LE}\ell_p(\mu, \Sigma; g)$ , then the sub-vector  $\mathbf{Y}_1$  has a log-elliptical distribution for any  $\Sigma(p \times p) > 0$  (Fang et al. 1990, Sec. 2.8). Moreover, the distribution of  $\mathbf{Y}_1$  is log-elliptical with the same parent distribution as  $\mathbf{Y}$  if the DGF  $g$  is that of multivariate scale mixture of normal distributions, as we establish in Theorem 5.

**Theorem 5** Let  $\mathbf{Y} \in \mathbb{R}_+^p$ ,  $\mu \in \mathbb{R}_+^p$ ,  $\Sigma(p \times p) > 0$  partitioned as in (7) and such that  $\mathbf{Y} \sim \text{LE}\ell_p(\mu, \Sigma; g)$ , with  $g(u) \propto \int_0^\infty t^{p/2} \exp(-ut/2) \, dH(t)$ ,  $u \geq 0$ ,  $H$  being a CDF on  $(0, \infty)$ . Then,  $\mathbf{Y}_1 \sim \text{LE}\ell_r(\mu_1, \Sigma_{11}; g)$ .

**Proof** See “Appendix E”.  $\square$

The following log-elliptical distributions have DGF as multivariate scale mixture of normal distributions and therefore satisfy the conditions in Theorem 5: multivariate log-normal distribution ( $H$  is the CDF of a degenerate distribution at  $t = 1$ ), the multivariate log- $t$  distribution ( $H$  is the CDF of a gamma distribution with shape parameter  $\tau/2$  and scale parameter  $2/\tau$ ,  $\tau > 0$ ), the multivariate log-slash distribution ( $H$  is the CDF of  $T = U^{2/q}$ ,  $q > 0$ , with  $U \sim \text{U}(0, 1)$ ), and the multivariate log-power exponential distribution for  $0 < \beta \leq 1$  ( $H$  is the CDF with PDF  $h(t) = \frac{1}{2}t^{-3/2}h_\beta(t^{-1/2})$ ,  $0 < \beta < 1$ , with  $h_\beta$  given in Gómez-Sánchez-Manzano et al. (2008, Eq. 3). If  $\beta = 1$ ,  $H$  is as in the multivariate log-normal distribution case). However, Theorem 5 does not apply to the multivariate log-power exponential distribution for  $\beta > 1$ .

In Theorem 6 we state that, if  $\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2)'$  has a Box–Cox elliptical distribution, then the conditional distribution of  $\mathbf{Y}_1|\mathbf{Y}_2$  is Box–Cox elliptical.

**Theorem 6** Let  $\mathbf{Y} \in \mathbb{R}_+^p$ ,  $\mu \in \mathbb{R}_+^p$ ,  $\lambda \in \mathbb{R}^p$ ,  $\Sigma(p \times p) > 0$  partitioned as in (7) and such that  $\mathbf{Y} \sim \text{BCE}\ell_p(\mu, \lambda, \Sigma; g)$ . Let  $\mu_1(\mathbf{w}_2) = \Sigma_{12}\Sigma_{22}^{-1}\mathbf{w}_2 \in R(\lambda_1)$  and  $\alpha(\mathbf{w}_2) = \mathbf{1} + \mathbf{D}_{\lambda_1}\mu_1(\mathbf{w}_2)$ , with  $\mathbf{w}_2 = T_{\lambda_2, \mu_2}(\mathbf{y}_2)$ , then  $\mathbf{Y}_1|\mathbf{Y}_2 = \mathbf{y}_2 \sim \text{BCE}\ell_r(\delta_1, \lambda_1, \mathbf{D}_{\alpha(\mathbf{w}_2)}^{-1}\Sigma_{11,2}\mathbf{D}_{\alpha(\mathbf{w}_2)}^{-1}; g_{q(\mathbf{w}_2)})$ , where  $\delta_1 = T_{\lambda_1, \mu_1}^{-1}(\mu_1(\mathbf{w}_2))$ ,  $\Sigma_{11,2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$  and  $g_{q(\mathbf{w}_2)}(u) = g(u + q(\mathbf{w}_2))$ ,  $u \geq 0$ , with  $q(\mathbf{w}_2) = \mathbf{w}_2'\Sigma_{22}^{-1}\mathbf{w}_2$ .

**Proof** See “Appendix F”.  $\square$

If  $\boldsymbol{\Sigma}_{12} = \mathbf{0}$  in Theorem 6, then  $\mathbf{Y}_1|\mathbf{Y}_2 = \mathbf{y}_2 \sim \text{BCE}\ell_r(\boldsymbol{\mu}_1, \boldsymbol{\lambda}_1, \boldsymbol{\Sigma}_{11}; g_{\mathbf{q}(\mathbf{w}_2)})$ . By comparing this conditional distribution with the marginal distribution of  $\mathbf{Y}_1$  given in Theorem 4, we have that, if  $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ ,  $\mathbf{Y}_1|\mathbf{Y}_2$  and  $\mathbf{Y}_1$  have the same distribution if the DGFs  $g_{\mathbf{q}(\mathbf{w}_2)}$  and  $g_1$  coincide. This characterizes the independence of the sub-vectors  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$ , as we state in Theorem 7.

**Theorem 7** *Let  $\mathbf{Y} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\mu} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\lambda} \in \mathbb{R}^p$ ,  $\boldsymbol{\Sigma}(p \times p) > 0$  partitioned as in (7) and such that  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ . Then,  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  are independent if and only if  $\mathbf{Y} \sim \text{BCN}_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma})$  and  $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ .*

**Proof** See “Appendix G”.  $\square$

In Theorem 8 we give an expression for mixed moments of Box–Cox elliptical random vectors.

**Theorem 8** *Let  $\mathbf{h} \in \mathbb{R}^p$ ,  $\boldsymbol{\mu} \in \mathbb{R}_+^p$ ,  $\boldsymbol{\lambda} \in \mathbb{R}^p$ ,  $\boldsymbol{\Sigma}(p \times p) > 0$ ,  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$  and  $\mathbf{U} \sim \text{BCE}\ell_p(\mathbf{1}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ . If  $E(\prod_{k=1}^p U_k^{h_k}) < \infty$ , then*

$$E\left(\prod_{k=1}^p Y_k^{h_k}\right) = \left(\prod_{k=1}^p \mu_k^{h_k}\right) E\left(\prod_{k=1}^p U_k^{h_k}\right).$$

**Proof** See “Appendix H”.  $\square$

The computation of mixed moments of a random vector  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$  from mixed moments of  $\mathbf{U} \sim \text{BCE}\ell_p(\mathbf{1}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$  as indicated in Theorem 8 is possible using Monte Carlo integration. By using Algorithm 1, one may generate a random sample of size  $n$  of the random vector  $\mathbf{W} = T_{\boldsymbol{\lambda}, \mathbf{1}}(\mathbf{U}) \sim \text{TE}\ell_p(\mathbf{0}, \boldsymbol{\Sigma}; R(\boldsymbol{\lambda}); g)$ , say  $\mathbf{w}_1, \dots, \mathbf{w}_n$ , where  $\mathbf{w}_i = (w_{i1}, \dots, w_{ip})'$ ,  $i = 1, \dots, n$ . If  $n$  is large,

$$E\left(\prod_{k=1}^p Y_k^{h_k}\right) \approx \frac{1}{n} \sum_{i=1}^n \prod_{k=1}^p (\mu_k u_k(w_{ik}))^{h_k},$$

with  $u_k(w_{ik}) = T_{\lambda_k, 1}^{-1}(w_{ik})$ ,  $i = 1, \dots, n$ ;  $k = 1, \dots, p$ .

Let  $\boldsymbol{\lambda} = \mathbf{0}$  (i.e.,  $\mathbf{Y} \sim \text{LE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$ ) in Theorem 8. We have that

$$E\left(\prod_{k=1}^p Y_k^{h_k}\right) = \left(\prod_{k=1}^p \mu_k^{h_k}\right) M_X(\mathbf{h}),$$

whenever  $M_X$ , the moment generating function of  $X \sim \text{E}\ell_p(\mathbf{0}, \boldsymbol{\Sigma}; g)$ , exists (Fang et al. 1990, Sec. 2.8). Another consequence of Theorem 8 is that the covariance matrices of  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$  and  $\mathbf{U} \sim \text{BCE}\ell_p(\mathbf{1}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ , denoted by  $\boldsymbol{\Sigma}_Y$  and  $\boldsymbol{\Sigma}_U$ , respectively, are such that  $\boldsymbol{\Sigma}_Y = D_{\boldsymbol{\mu}} \boldsymbol{\Sigma}_U D_{\boldsymbol{\mu}}$ . Moreover, the correlation matrices of  $\mathbf{Y}$  and  $\mathbf{U}$  are equal.

## 4 Parameter interpretation

From Definition 4 we have that the distribution of a random vector  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$  is characterized by a random vector  $\mathbf{W} = T_{\boldsymbol{\lambda}, \boldsymbol{\mu}}(\mathbf{Y}) \sim \text{TE}\ell_p(\mathbf{0}, \boldsymbol{\Sigma}; R(\boldsymbol{\lambda}); g)$ . In such a characterization, the parameter vectors  $\boldsymbol{\mu} \in \mathbb{R}_+^p$  and  $\boldsymbol{\lambda} \in \mathbb{R}^p$  are introduced through an extended Box–Cox transformation (Definition 3), in such a way that  $\mu_k$  and  $\lambda_k$ ,  $k = 1, \dots, p$ , are parameters involved in the transformation of  $Y_k$  only; hence these parameters are characteristics of the distribution of  $Y_k$ . Also, the marginal distributions of the components of  $\mathbf{W}$  are associated through  $\boldsymbol{\Sigma}(p \times p) > 0$ , which implies that the marginal distributions of the components of  $\mathbf{Y}$  are associated through this matrix as well. Hence,  $\mu_k$  and  $\lambda_k$ ,  $k = 1, \dots, p$ , are, respectively the scale parameter and skewness parameter (power transformation for marginal symmetry) of the distribution of  $Y_k$ ;  $\sigma_{jk}$ ,  $j \neq k$ , is the association parameter between  $Y_j$  and  $Y_k$ .

The parameters  $\mu_k$  and  $\sigma_{kk}$ ,  $k = 1, \dots, p$ , are related with quantiles of  $Y_k$ . In order to establish these relations, let the marginal PDF of  $Y_k$  be written as

$$f_{Y_k}(y_k) = \frac{g_{\boldsymbol{\gamma}_k}(s_k) \frac{y_k^{\lambda_k-1}}{\sqrt{\sigma_{kk}} \mu_k^{\lambda_k}}}{\int_{I(\lambda_k \sqrt{\sigma_{kk}})} g_{\boldsymbol{\gamma}_k}(s_k) ds_k}, \quad s_k = \sigma_{kk}^{-1/2} T_{\lambda_k, \mu_k}(y_k), \quad y_k > 0, \quad (9)$$

with  $I(\lambda_k \sqrt{\sigma_{kk}})$  defined in (4) and

$$g_{\boldsymbol{\gamma}_k}(u_k) = \int_{R(\boldsymbol{\Delta}_{-k, -k} \boldsymbol{\lambda}_{-k})} g((1 + \boldsymbol{\gamma}_k \boldsymbol{\gamma}'_k) u_k^2 - 2 \boldsymbol{\gamma}_k \boldsymbol{\Omega}_k u_k \mathbf{w} + \mathbf{w}' \boldsymbol{\Omega}'_k \boldsymbol{\Omega}_k \mathbf{w}) d\mathbf{w}, \quad (10)$$

where  $u_k \in I(\lambda_k \sqrt{\sigma_{kk}})$ ,  $\boldsymbol{\Delta} = \text{diag}\{\sqrt{\sigma_{11}}, \dots, \sqrt{\sigma_{pp}}\}$ ,  $\boldsymbol{\Omega}_k = (\boldsymbol{\Sigma}_{-k, -k} - \sigma_{kk}^{-1} \boldsymbol{\Sigma}_{-k, k} \boldsymbol{\Sigma}_{k, -k})^{-1/2} \boldsymbol{\Delta}_{-k, -k}$  and  $\boldsymbol{\gamma}_k = \sigma_{kk}^{-1/2} \boldsymbol{\Sigma}_{k, -k} \boldsymbol{\Omega}_k \boldsymbol{\Delta}_{-k, -k}^{-1}$ .

PDF (9) can be built from a random variable  $U_k$  defined in  $\mathbb{R}$  with CDF

$$F_{U_k}(u_k) = c_k \int_{-\infty}^{u_k} g_{\boldsymbol{\gamma}_k}(t) dt, \quad (11)$$

where  $c_k^{-1} = \int_{-\infty}^{\infty} g_{\boldsymbol{\gamma}_k}(t) dt$  (see details in “Appendix I”). An interesting case occurs when the integral that involves  $g_{\boldsymbol{\gamma}_k}$  has integration region  $R(\boldsymbol{\Delta}_{-k, -k} \boldsymbol{\lambda}_{-k}) = \mathbb{R}^{p-1}$ , i.e., when  $\boldsymbol{\Delta}_{-k, -k} \boldsymbol{\lambda}_{-k} = \mathbf{0}$ . In this case,  $U_k \sim \text{El}_1(0, 1; \tilde{g})$ , with

$$\tilde{g}(u) = \int_{\mathbb{R}^{p-1}} g(u + \mathbf{w}' \mathbf{w}) d\mathbf{w}, \quad u \geq 0. \quad (12)$$

In Theorem 9 we show that all the quantiles of the univariate marginal distributions of Box–Cox elliptical random vectors are proportional to the respective component of  $\boldsymbol{\mu}$ .

**Theorem 9** Let  $\mu \in \mathbb{R}_+^p$ ,  $\lambda \in \mathbb{R}^p$ ,  $\Sigma(p \times p) > 0$  and  $\mathbf{Y} \sim \text{BCE}\ell_p(\mu, \lambda, \Sigma; g)$ . The  $\alpha$ -quantile  $y_{k,\alpha}$  of  $Y_k$ ,  $\alpha \in (0, 1)$ ,  $k = 1, \dots, p$ , satisfies

$$y_{k,\alpha} = \begin{cases} \mu_k (1 + \lambda_k \sqrt{\sigma_{kk}} s_{k,\alpha})^{1/\lambda_k}, & \lambda_k \neq 0, \\ \mu_k \exp(\sqrt{\sigma_{kk}} s_{k,\alpha}), & \lambda_k = 0, \end{cases} \quad (13)$$

with

$$s_{k,\alpha} = \begin{cases} F_{U_k}^{-1}(\alpha + (1 - \alpha)F_{U_k}(-1/\lambda_k \sqrt{\sigma_{kk}})), & \lambda_k > 0, \\ F_{U_k}^{-1}((1 + \alpha)F_{U_k}(-1/\lambda_k \sqrt{\sigma_{kk}}) - 1), & \lambda_k < 0, \\ F_{U_k}^{-1}(\alpha), & \lambda_k = 0, \end{cases} \quad (14)$$

where  $F_{U_k}$  is the CDF given in (11).

**Proof** See “Appendix J”.  $\square$

In Theorem 9 we stated that, if  $\mathbf{Y} \sim \text{BCE}\ell_p(\mu, \lambda, \Sigma; g)$ , all the quantiles of  $Y_k$ ,  $k = 1, \dots, p$ , particularly the median, are proportional to  $\mu_k$ . This feature of the class of Box–Cox elliptical distributions makes it attractive for regression modeling purposes. For instance, assume that, for fixed  $k$ ,  $\log \mu_k = \sum_{j=1}^q x_j \beta_j$ , where the betas are unknown regression parameters and the  $x$ s are fixed covariates. In this case,  $\exp(\beta_j)$  is the multiplicative effect of a one unit increase in  $x_j$  on the quantiles of  $Y_k$ .

In Corollary 2 we establish conditions under which the quantiles of  $Y_k$  can be calculated from quantiles of standard symmetric distributions.

**Corollary 2** Let  $\mu \in \mathbb{R}_+^p$ ,  $\lambda \in \mathbb{R}^p$ ,  $\Sigma(p \times p) > 0$  and  $\mathbf{Y} \sim \text{BCE}\ell_p(\mu, \lambda, \Sigma; g)$ . If  $\lambda = \mathbf{0}$  (i.e.  $\mathbf{Y} \sim \text{LE}\ell_p(\mu, \Sigma; g)$ ) or  $\lambda_j \sqrt{\sigma_{jj}} \rightarrow 0$ ,  $j = 1, \dots, p$ , then the  $\alpha$ -quantile  $y_{k,\alpha}$  of  $Y_k$ ,  $\alpha \in (0, 1)$ ,  $k = 1, \dots, p$ , is given by  $y_{k,\alpha} = \mu_k \exp(\sqrt{\sigma_{kk}} q_\alpha)$ , where  $q_\alpha$  is the  $\alpha$ -quantile of a standard symmetric distribution with DGF given by (12).

**Proof** Let  $\lambda = \mathbf{0}$  or  $\lambda_j \sqrt{\sigma_{jj}} \rightarrow 0$ ,  $j = 1, \dots, p$  in Theorem 9. From (13) and (14) it follows that  $y_{k,\alpha} = \mu_k \exp(\sqrt{\sigma_{kk}} q_\alpha)$ , where  $q_\alpha = F_U^{-1}(\alpha)$ , with  $U \sim \text{El}_1(0, 1; \tilde{g})$ , with  $\tilde{g}$  being a DGF given by (12). This fact follows because  $R(\Delta_{-k,-k} \lambda_{-k}) = \mathbb{R}^{p-1}$  when  $\lambda = \mathbf{0}$ , or  $R(\Delta_{-k,-k} \lambda_{-k}) \rightarrow \mathbb{R}^{p-1}$  when  $\lambda_j \sqrt{\sigma_{jj}} \rightarrow 0$ ,  $j = 1, \dots, p$ .  $\square$

Let  $\mathbf{Y} \sim \text{BCE}\ell_p(\mu, \lambda, \Sigma; g)$ . A coefficient of variation based on quantiles for  $Y_k$ ,  $k = 1, \dots, p$ , is defined as (Rigby and Stasinopoulos 2006)

$$\text{CV}_{Y_k} = \frac{3}{4} \frac{(y_{k,3/4} - y_{k,1/4})}{y_{k,1/2}}.$$

Corollary 2 allows interpretation of the parameters  $\mu_k$  and  $\sigma_{kk}$  from their relations with quantiles of  $Y_k$ ,  $k = 1, \dots, p$ . In fact, if  $\lambda = \mathbf{0}$  (i.e.  $\mathbf{Y} \sim \text{LE}\ell(\mu, \Sigma; g)$ ), then  $\mu_k = y_{k,1/2}$  and  $\text{CV}_{Y_k} = 1.5 \sinh(\sqrt{\sigma_{kk}} q_{3/4})$ , where  $q_{3/4}$  is the third quartile of a standard symmetric distribution with DGF given in (12). Also, if  $\lambda \approx \mathbf{0}$  or  $\lambda_j \sqrt{\sigma_{jj}} \approx 0$ ,  $j = 1, \dots, p$ , then  $\mu_k \approx y_{k,1/2}$  and  $\text{CV}_{Y_k} \approx 1.5 \sinh(\sqrt{\sigma_{kk}} q_{3/4})$ . Hence, in these cases,  $\mu_k$  is equal or approximately equal to the median of  $Y_k$ . Moreover,  $\text{CV}_{Y_k}$  depends on  $\sigma_{kk}$  through the hyperbolic sine function, which is a monotonically increasing function. Therefore,  $\sigma_{kk}$  can be seen as a relative dispersion parameter of the distribution of  $Y_k$ .

## 5 Parameter estimation

Let  $y_1, \dots, y_n$  be the observed values of a random sample  $Y_1, \dots, Y_n$  of  $Y \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ , with  $Y_i = (Y_{i1}, \dots, Y_{ip})'$ ,  $i = 1, \dots, n$ . Let  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_q)'$  be the vector of extra parameters induced by the DGF  $g$ . The maximum likelihood estimators of  $\boldsymbol{\mu}$ ,  $\boldsymbol{\lambda}$ ,  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\eta}$ , denoted by  $\hat{\boldsymbol{\mu}}$ ,  $\hat{\boldsymbol{\lambda}}$ ,  $\hat{\boldsymbol{\Sigma}}$  and  $\hat{\boldsymbol{\eta}}$ , respectively, will be such that maximize the log-likelihood function  $\ell = \sum_{i=1}^n \ell_i$ , with

$$\begin{aligned} \ell_i = & -\log \left\{ \int_{R(\boldsymbol{\lambda})} g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) d\mathbf{w} \right\} + \log\{g(\mathbf{w}'_i \boldsymbol{\Sigma}^{-1} \mathbf{w}_i)\} \\ & + \sum_{k=1}^p (\lambda_k - 1) \log y_{ik} - \sum_{k=1}^p \lambda_k \log \mu_k, \end{aligned} \quad (15)$$

where  $\mathbf{w}_i = T_{\boldsymbol{\lambda}, \boldsymbol{\mu}}(y_i)$ . There is no closed form for the maximum likelihood estimators  $\hat{\boldsymbol{\mu}}$ ,  $\hat{\boldsymbol{\lambda}}$ ,  $\hat{\boldsymbol{\Sigma}}$  and  $\hat{\boldsymbol{\eta}}$ , but they can be computed using numerical optimization algorithms implemented in computer packages. The number of parameters to be estimated is  $p(p+5)/2 + q$ .

Let  $\boldsymbol{\mu}^{(0)}, \boldsymbol{\lambda}^{(0)}, \boldsymbol{\Sigma}^{(0)}$  and  $\boldsymbol{\eta}^{(0)}$  be the initial values for the estimation of  $\boldsymbol{\mu}$ ,  $\boldsymbol{\lambda}$ ,  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\eta}$ , respectively. For the choice of  $\mu_k^{(0)}, \lambda_k^{(0)}$  and  $\sigma_{kk}^{(0)}$ ,  $k = 1, \dots, p$ , we suggest the estimates obtained by fitting a Box–Cox symmetric distribution to the  $k$ th component of  $\mathbf{Y}$ , i.e. the estimated parameters of  $Y_k \sim \text{BCS}(\mu_k, \sqrt{\sigma_{kk}}, \lambda_k; g)$ . As initial values for  $\sigma_{jk}$ , we suggest  $\sigma_{jk}^{(0)} = 0$ ,  $j \neq k$ . Initial values for the extra parameters (if any),  $\eta_j^{(0)}$ ,  $j = 1, \dots, q$ , will depend on the family of distributions considered. For instance, for the multivariate Box–Cox  $t$  distribution we propose as initial value for the degrees of freedom parameter,  $\tau^{(0)}$ , the corresponding estimate obtained by fitting a multivariate  $t$  distribution to the vector  $\mathbf{X} = T_{\boldsymbol{\lambda}^{(0)}, \boldsymbol{\mu}^{(0)}}(\mathbf{Y})$ .

The main difficulty in implementing an optimization scheme is due to the need of an efficient computation of the integral  $\int_{R(\boldsymbol{\lambda})} g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) d\mathbf{w}$ , that appears in (15). This integral depends on the complexity and structure of the DGF  $g$  and is computed over  $R(\boldsymbol{\lambda})$ . Hence, the vector of the extra parameters  $\boldsymbol{\eta}$ , the matrix  $\boldsymbol{\Sigma}$  and the vector  $\boldsymbol{\lambda}$  are involved in the estimation procedure through this integral. Genz and Bretz (2009) propose algorithms to efficiently compute this type of integral over rectangles when  $g$  is the DGF of the multivariate normal and  $t$  families. In the class of the log-elliptical distributions ( $\boldsymbol{\lambda} = \mathbf{0}$ ) the integral disappears making the estimation process much easier. In this case, the logarithm of the likelihood function is given by  $\ell = \sum_{i=1}^n \ell_i$ , where  $\ell_i$ ,  $i = 1, \dots, n$ , is

$$\ell_i = -\frac{1}{2} \log(\det(\boldsymbol{\Sigma})) + \log\{g(\mathbf{w}'_i \boldsymbol{\Sigma}^{-1} \mathbf{w}_i)\} - \sum_{k=1}^p \log y_{ik},$$

with  $\mathbf{w}_i = T_{\mathbf{0}, \boldsymbol{\mu}}(y_i)$ . Here, the unknown quantities to be estimated are  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\eta}$ , i.e.  $p(p+3)/2 + q$  parameters.

To evaluate the proposed estimation procedure we conducted simulations with bivariate log-normal, log- $t$ , Box–Cox normal and Box–Cox  $t$  distributions, different

sample sizes, namely  $n = 125, 250, 500$ , and  $N = 5000$  Monte Carlo replicates. The random samples of  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$  were generated using Algorithm 2.

### Algorithm 2

1. Generate a random sample of size  $n$ , say  $\mathbf{w}_1, \dots, \mathbf{w}_n$ , of  $\mathbf{W} \sim \text{TEL}_p(\mathbf{0}, \boldsymbol{\Sigma}; R(\boldsymbol{\lambda}); g)$  using Algorithm 1.
2. Compute  $\mathbf{y}_1 = T_{\boldsymbol{\lambda}, \boldsymbol{\mu}}^{-1}(\mathbf{w}_1), \dots, \mathbf{y}_n = T_{\boldsymbol{\lambda}, \boldsymbol{\mu}}^{-1}(\mathbf{w}_n)$ . From Definition 4,  $\mathbf{y}_1, \dots, \mathbf{y}_n$  is a random sample of  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ .

In each simulation experiment we used the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) optimization algorithm to maximize the log-likelihood function with the initial values proposed above. The integral in (15) was efficiently evaluated using algorithms proposed by Genz and Bretz (2009). All the computations were conducted in the R software (R Core Team 2016).

Let  $\hat{\theta}_1, \dots, \hat{\theta}_N$  be the ordered estimated values of a scalar parameter, say  $\theta$ , in  $N$  Monte Carlo simulated samples. Let  $M(\hat{\theta})$  be the median of  $\{\hat{\theta}_1, \dots, \hat{\theta}_N\}$ . The median bias, denoted by  $\text{MB}(\hat{\theta})$ , is given by  $\text{MB}(\hat{\theta}) = M(\hat{\theta}) - \theta$ . The median absolute deviation, denoted by  $\text{MAD}(\hat{\theta})$ , is defined as the median of  $\{|\hat{\theta}_1 - M(\hat{\theta})|, \dots, |\hat{\theta}_N - M(\hat{\theta})|\}$ . Also, let  $\text{IQR}(\hat{\theta})$  be the interquartile range of  $\{\hat{\theta}_1, \dots, \hat{\theta}_N\}$ . These summaries of the estimates were computed for each simulation experiment and reported in Table 1. The figures in this table suggest a suitable behavior of the estimation procedure, because the median biases are close to zero and the median absolute deviations and interquartile ranges get smaller as  $n$  grows.

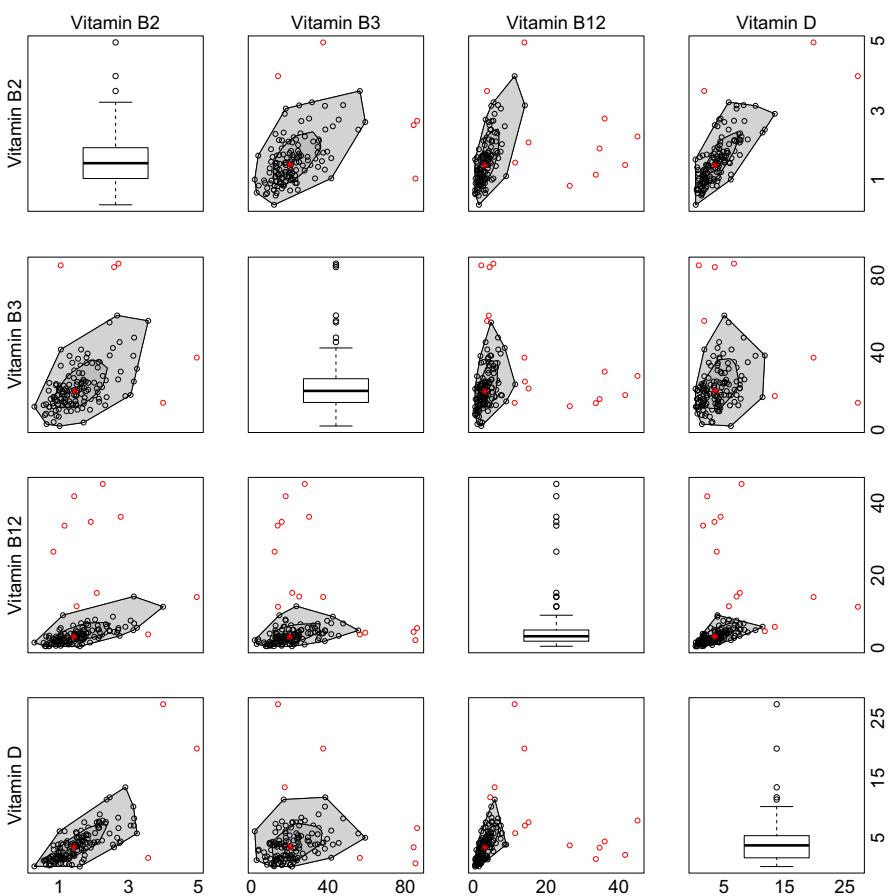
## 6 Application

The dataset refers to observations of vitamins B2 (in mg), B3 (in mg), B12 (in mcg) and D (in mcg) intakes based on the first 24-h dietary recall interview for  $n = 136$  older men. The bagplots (Rousseeuw et al. 1999) shown in Fig. 2 indicate that the vitamin intakes are positively correlated, their bivariate distributions are skewed, and that outliers are present.

For each pair of variables, we fitted bivariate log-normal, log- $t$ , Box–Cox normal and Box–Cox  $t$  distributions, and the respective marginal independent distributions; we denote these distributions by  $\text{LN}_2$ ,  $\text{Lt}_2$ ,  $\text{BCN}_2$ ,  $\text{BCt}_2$ ,  $\text{Ind-LN}_1$ ,  $\text{Ind-Lt}_1$ ,  $\text{Ind-BCN}_1$  and  $\text{Ind-BCt}_1$ , respectively. Table 2 shows the Akaike information criterion (AIC) for each fit. The smallest AIC is marked in bold and indicates the distribution that best fits the data. The figures in this table indicate that the bivariate distributions provide better fit when compared with the respective marginal independent distributions. This is not surprising since there is evidence of association among the variables. Additionally, Table 2 indicates that the bivariate Box–Cox  $t$  distribution gives the best fit for the pairs of variables: vitamins B2–D, B3–D and B12–D. Also, the bivariate log- $t$  distribution provides the best fit for the pairs: vitamins B2–B3, B2–B12 and B3–B12. Hence, the bivariate distributions based on the  $t$  distribution provide better fit than those based on the normal distribution. This fact is due to the presence of extreme outliers (Fig. 2).

**Table 1** Median bias (MB), median absolute deviation (MAD) and interquartile range (IQR) of the parameter estimators

n	Bivariate log-normal						Bivariate Box–Cox <i>t</i>						Bivariate Box–Cox normal																				
	$\mu_1$			$\mu_2$			$\sigma_{11}$			$\sigma_{12}$			$\sigma_{22}$			$\lambda_1$			$\lambda_2$			$\sigma_{11}$			$\sigma_{12}$			$\sigma_{22}$					
	8	8	8	0.8	0.8	0.5	1	1	1	0.2	0.02	−0.03	0.08	−0.01	−0.04	−0.01	0.06	0.06	0.05	0.1	0.1	0.4	0.3	0.4	0.3	0.1	0.1	0.3	0.3	0.3			
125	MB	−0.02	0.00	−0.02	0.02	0.02	−0.03	0.08	−0.01	−0.04	−0.01	−0.04	−0.01	−0.04	−0.01	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05			
	MAD	0.59	0.66	0.07	0.07	0.09	0.09	0.72	0.42	0.25	0.13	0.31	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16			
	IQR	1.19	1.33	0.15	0.14	0.18	0.18	2.96	0.84	0.51	0.27	2.80	0.61	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36			
250	MB	−0.01	0.00	−0.01	0.01	−0.01	0.01	0.05	0.00	−0.02	−0.01	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03		
	MAD	0.42	0.47	0.05	0.05	0.07	0.07	0.50	0.29	0.19	0.10	0.24	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11		
	IQR	0.85	0.93	0.10	0.10	0.13	0.13	1.46	0.58	0.38	0.20	0.96	0.28	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24			
500	MB	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.01	−0.02	−0.01	0.05	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02		
	MAD	0.30	0.33	0.04	0.04	0.05	0.05	0.40	0.20	0.13	0.07	0.20	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07		
	IQR	0.60	0.66	0.08	0.07	0.09	0.09	0.94	0.40	0.27	0.15	0.56	0.17	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16		
n	Bivariate log- <i>t</i>						Bivariate Box–Cox <i>t</i>						Bivariate Box–Cox normal																				
	$\mu_1$	$\mu_2$	$\sigma_{11}$	$\sigma_{12}$	$\sigma_{22}$	$\tau$	$\mu_1$	$\mu_2$	$\sigma_{11}$	$\sigma_{12}$	$\sigma_{22}$	$\lambda_1$	$\lambda_2$	$\sigma_{11}$	$\sigma_{12}$	$\sigma_{22}$	$\mu_1$	$\mu_2$	$\sigma_{11}$	$\sigma_{12}$	$\sigma_{22}$	$\mu_1$	$\mu_2$	$\sigma_{11}$	$\sigma_{12}$	$\sigma_{22}$	$\mu_1$	$\mu_2$	$\sigma_{11}$	$\sigma_{12}$	$\sigma_{22}$		
	7	10	1.2	0.6	1.4	5	5	4	5	5	5	−1	−1	0.6	0.6	0.8	125	125	125	125	125	125	125	125	125	125	125	125	125	125	125	125	125
125	MB	0.01	0.02	0.00	−0.01	−0.01	0.25	−0.17	−0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	MAD	0.97	1.45	0.22	0.17	0.26	1.76	0.91	0.58	0.08	0.09	0.09	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04		
	IQR	1.31	1.98	0.30	0.22	0.35	2.63	1.80	1.15	0.15	0.17	0.17	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08		
250	MB	−0.01	−0.01	0.00	0.00	0.00	0.00	0.15	−0.07	−0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	MAD	0.69	1.05	0.16	0.12	0.19	1.18	0.63	0.41	0.05	0.06	0.06	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03		
	IQR	0.93	1.43	0.22	0.16	0.25	1.66	1.26	0.82	0.10	0.12	0.12	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05		
500	MB	0.00	0.00	0.00	0.00	0.00	0.00	0.07	−0.02	−0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	MAD	0.46	0.72	0.12	0.08	0.13	0.82	0.45	0.29	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	
	IQR	0.62	0.97	0.16	0.11	0.18	1.13	0.90	0.58	0.07	0.09	0.09	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	



**Fig. 2** Bagplot matrix; nutritional data

Table 3 gives the estimates (and standard errors) of the parameters of the best fitting model as indicated in Table 2. The standard errors were obtained asymptotically through the observed information matrix, which is calculated numerically based on the finite difference method (Nocedal and Wright 2006, Sect. 8.1) implemented within the `mle` function of R (R Core Team 2016) for maximum likelihood estimation. It is noteworthy that the estimated degrees of freedom parameter varies from 4 to 8, indicating that heavier-than-normal distributions are better suited for fitting the data.

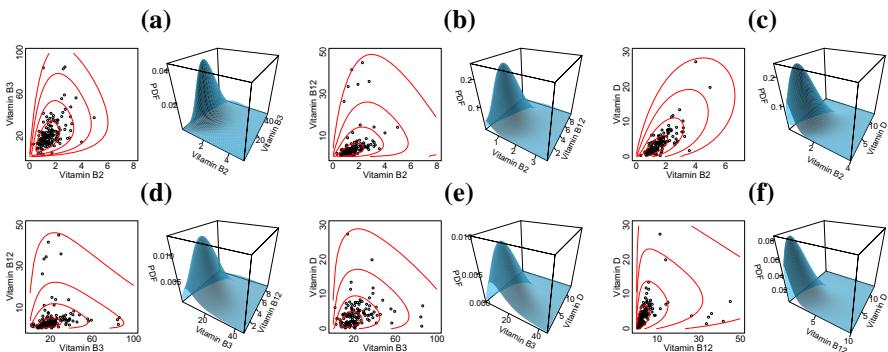
For the bivariate log-*t* distribution fitted to the pair of vitamins B2–B3 the estimates of  $\mu_1$  and  $\mu_2$  are  $\hat{\mu}_1 = 1.45$  and  $\hat{\mu}_2 = 19.91$  and correspond to estimates of the median intake of vitamins B2 and B3 in the population. These estimates are close to the corresponding sample medians (1.49 and 19.99, respectively). The estimates of the relative dispersion parameters are  $\hat{\sigma}_{11} = 0.16$  and  $\hat{\sigma}_{22} = 0.23$ ; hence the relative dispersion of vitamin B2 is estimated to be smaller than that of vitamin B3. For the intake of vitamins B12–D the best fit is achieved by the bivariate Box–Cox *t* distribution. Note that the estimated parameters satisfy  $\hat{\lambda}_1 \sqrt{\hat{\sigma}_{11}} = -0.13$  and  $\hat{\lambda}_2 \sqrt{\hat{\sigma}_{22}} = 0.10$ , that are

**Table 2** AIC for the fitting distributions; nutritional data

Variables	Ind-LN <sub>1</sub>	LN <sub>2</sub>	Ind-L <sub>1</sub>	L <sub>f2</sub>	Ind-BCN <sub>1</sub>	BCN <sub>2</sub>	Ind-BC <sub>1</sub>	BC <sub>2</sub>
B2-B3	1329.87	1291.88	1320.10	<b>1279.44</b>	1328.90	1289.57	1323.15	1281.66
B2-B12	946.51	900.99	941.11	<b>883.24</b>	945.13	895.20	944.11	885.68
B2-D	916.83	814.27	919.47	802.72	914.99	803.79	918.08	<b>796.84</b>
B3-B12	1726.19	1710.38	1707.90	<b>1689.32</b>	1720.67	1702.24	1710.27	1690.79
B3-D	1696.51	1689.65	1686.26	1679.20	1690.52	1683.20	1684.75	<b>1676.79</b>
B12-D	1313.15	1250.09	1307.27	1226.64	1306.76	1233.10	1305.23	<b>1223.88</b>

**Table 3** Estimates of the parameters (and standard errors) of the best fitting distribution: nutritional data

Variables	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}_{11}$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\hat{\tau}$
B2-B3 (L <sub>T2</sub> )	1.45 (0.05)	19.91 (0.90)	—	—	0.16 (0.02)	0.10 (0.02)	0.23 (0.04)	6.22 (2.18)
B2-B12 (L <sub>T2</sub> )	1.46 (0.06)	3.10 (0.20)	—	—	0.15 (0.02)	0.16 (0.03)	0.43 (0.08)	4.57 (1.36)
B2-D (BC <sub>T2</sub> )	1.45 (0.06)	3.42 (0.23)	0.19 (0.17)	0.31 (0.11)	0.16 (0.03)	0.22 (0.04)	0.48 (0.08)	7.96 (3.57)
B3-B12 (L <sub>T2</sub> )	20.10 (0.91)	3.13 (0.20)	—	—	0.20 (0.04)	0.13 (0.03)	0.42 (0.08)	3.96 (1.12)
B3-D (BC <sub>T2</sub> )	19.86 (0.94)	3.30 (0.23)	0.15 (0.14)	0.24 (0.12)	0.24 (0.04)	0.12 (0.03)	0.51 (0.08)	7.42 (3.03)
B12-D (BC <sub>T2</sub> )	3.10 (0.20)	3.42 (0.24)	-0.19 (0.10)	0.15 (0.11)	0.45 (0.08)	0.31 (0.06)	0.47 (0.08)	5.50 (1.99)



**Fig. 3** Scatter plots overlaid with contour plots and joint PDF of the best fitting distributions; nutritional data

close to zero. Hence,  $\hat{\mu}_1 = 3.10$  and  $\hat{\mu}_2 = 3.42$  are expected to be close to the sample median of vitamins B12 and D intakes respectively, and this is in fact the case (the sample medians are 3.24 and 3.80, respectively). Since  $\hat{\sigma}_{11} = 0.45$  and  $\hat{\sigma}_{11} = 0.47$ , we have that the relative dispersions of vitamins B12 and D intakes are similar.

Figure 3 shows contour plots of the fitted distributions superimposed to the scatter plots of the data, and the corresponding PDFs. The plots suggest a reasonable fit for all the pairs of variables.

## 7 Final remarks

In this paper we presented a new class of multivariate distributions, the class of Box–Cox elliptical distributions, that is suitable for modeling multivariate positive, marginally asymmetric, possibly heavy-tailed data. The construction of the Box–Cox distributions uses an extended multivariate Box–Cox transformation and the class of truncated elliptical distributions, both defined in this paper. We show that the class of Box–Cox elliptical distributions has as special cases the classes of the log-elliptical and Box–Cox symmetric distributions. The Box–Cox elliptical distributions allow easy parameter interpretation, a desirable feature for modeling purposes.

Starting from a study of the class of truncated elliptical distributions, we defined and studied the Box–Cox elliptical distributions. Specifically, we stated useful properties and discussed maximum likelihood estimation issues, generation of random samples, interpretation of parameters, and applications.

There are some open problems that will be addressed in future papers. The efficiency of the implementation of maximum likelihood estimation depends on the efficient computation of the integral involved in (15). The methods proposed by Genz and Bretz (2009) to efficiently compute the integral when  $g$  is the DGF of the multivariate normal and  $t$  distributions allowed us to implement maximum likelihood estimation for the parameters of the multivariate Box–Cox normal and Box–Cox  $t$  distributions. Computational aspects of these distributions depend on computations related to the multivariate truncated normal and truncated  $t$  distributions, for which recent works have been developed (e.g. Ho et al. 2012; Kan and Robotti 2017). The efficient com-

putation of the integral for other DGFs will provide the implementation of maximum likelihood estimation for other distributions in the Box–Cox elliptical class, such as the multivariate Box–Cox power exponential and Box–Cox slash distributions. The relation of the scale parameters to quantiles of the marginal distributions permits the construction of Box–Cox elliptical regression models that are able to model the relationship between covariates and quantiles of the response variables. The study of regression models based on Box–Cox elliptical distributions, including measurement error models, spatial models, mixed models, etc., is perfectly feasible and of interest but it is beyond the scope of the present paper.

**Acknowledgements** We thank José Eduardo Corrente for providing the data used in this study. Funding was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq (Grant No. 304388-2014-9) and Fundação de Amparo à Pesquisa do Estado de São Paulo—FAPESP (Grant No. 2012/21788-2). The first author received Ph.D. scholarships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—CAPES—and CNPq. We are thankful to the Editor, Associate Editor and the anonymous reviewers for their helpful comments and suggestions.

## Appendix A: Proof of the Theorem 1

The conditional PDF of  $W_k | \mathbf{W}_{-k}$ ,  $k = 1, \dots, p$ , is given by

$$f_{W_k | \mathbf{W}_{-k}}(w_k) = \frac{g((\mathbf{w} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu}))}{\int_{a_k}^{b_k} g((\mathbf{w} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu})) d\mathbf{w}_k}, \quad w_k \in (a_k, b_k).$$

From the identity  $(\mathbf{w} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu}) = [(w_k - \mu_{k,-k})/\sigma_{k,-k}]^2 + q(\mathbf{w}_{-k})$ , we get the result.

## Appendix B: Proof of the Theorem 2

If  $\mathbf{W} = T_{\lambda, \boldsymbol{\mu}}(\mathbf{Y}) \sim \text{TEL}_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}; R(\lambda); g)$ , then its PDF is given by

$$f_{\mathbf{W}}(\mathbf{w}) = \frac{g((\mathbf{w} - \boldsymbol{\xi})' \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\xi}))}{\int_{R(\lambda)} g((\mathbf{w} - \boldsymbol{\xi})' \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\xi})) d\mathbf{w}}, \quad \mathbf{w} \in R(\lambda). \quad (16)$$

Let  $V : R(\lambda) \rightarrow R(\lambda)$  be the transformation defined as  $V(\mathbf{w}) = \mathbf{D}_{\boldsymbol{\alpha}}^{-1}(\mathbf{w} - \boldsymbol{\xi})$ , and let  $\mathbf{U} = V(\mathbf{W})$ , with Jacobian  $J(\mathbf{w} \rightarrow \mathbf{u}) = \prod_{k=1}^p (1 + \lambda_k \xi_k)$ . The PDF of  $\mathbf{U}$  is

$$f_{\mathbf{U}}(\mathbf{u}) = \frac{g(\mathbf{u}' (\mathbf{D}_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\alpha}}^{-1})^{-1} \mathbf{u})}{\int_{R(\lambda)} g(\mathbf{u}' (\mathbf{D}_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\alpha}}^{-1})^{-1} \mathbf{u}) d\mathbf{u}}, \quad \mathbf{u} \in R(\lambda).$$

Hence,  $\mathbf{U} \sim \text{TEL}_p(\mathbf{0}, \mathbf{D}_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\alpha}}^{-1}; R(\lambda); g)$ . Because  $\mathbf{U} = V(T_{\lambda, \boldsymbol{\mu}}(\mathbf{Y})) = T_{\lambda, \boldsymbol{\mu}}^{-1}(\mathbf{Y})$ , where  $\boldsymbol{\delta} = T_{\lambda, \boldsymbol{\mu}}^{-1}(\boldsymbol{\xi})$ , then from Definition 4 we have  $\mathbf{Y} = T_{\lambda, \boldsymbol{\delta}}^{-1}(\mathbf{U}) \sim \text{BCE}\ell_p(\boldsymbol{\delta}, \lambda, \mathbf{D}_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\alpha}}^{-1}; g)$ .

On the other hand, if  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\delta}, \boldsymbol{\lambda}, \mathbf{D}_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\alpha}}^{-1}; g)$ , then its PDF is

$$f_Y(\mathbf{y}) = \frac{g(\mathbf{w}'(\mathbf{D}_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\alpha}}^{-1})^{-1} \mathbf{w}) \prod_{k=1}^p \frac{y_k^{\lambda_k-1}}{\delta_k^{\lambda_k}}}{\int_{R(\boldsymbol{\lambda})} g(\mathbf{w}'(\mathbf{D}_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\alpha}}^{-1})^{-1} \mathbf{w}) d\mathbf{w}}, \quad \mathbf{w} = T_{\boldsymbol{\lambda}, \boldsymbol{\delta}}(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}_+^p. \quad (17)$$

Now, from the transformation  $\mathbf{W} = T_{\boldsymbol{\lambda}, \boldsymbol{\mu}}(\mathbf{Y})$ , with Jacobian  $J(\mathbf{y} \rightarrow \mathbf{w}) = \prod_{k=1}^p \mu_k (1 + \lambda_k w_k)^{1/\lambda_k - 1}$ , in the PDF (17) we arrive at PDF (16).

## Appendix C: Proof of Theorem 3

1. From  $\mathbf{T} = \mathbf{D}_{\boldsymbol{\alpha}} \mathbf{Y}$ , with Jacobian  $J(\mathbf{y} \rightarrow \mathbf{t}) = \prod_{k=1}^p \alpha_k^{-1}$ , in (6), we get the PDF of  $\mathbf{T}$  as

$$f_T(\mathbf{t}) = \frac{g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) \prod_{k=1}^p \frac{t_k^{\lambda_k-1}}{(\alpha_k \mu_k)^{\lambda_k}}}{\int_{R(\boldsymbol{\lambda})} g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) d\mathbf{w}}, \quad \mathbf{t} \in \mathbb{R}_+^p,$$

where  $\mathbf{w} = T_{\boldsymbol{\lambda}, \boldsymbol{\mu}}(\mathbf{D}_{\boldsymbol{\alpha}}^{-1} \mathbf{t}) = T_{\boldsymbol{\lambda}, \mathbf{D}_{\boldsymbol{\alpha}} \boldsymbol{\mu}}(\mathbf{t})$ . Hence,  $\mathbf{T} = \mathbf{D}_{\boldsymbol{\alpha}} \mathbf{Y} \sim \text{BCE}\ell_p(\mathbf{D}_{\boldsymbol{\alpha}} \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ .

2. Note that the PDF of  $\mathbf{Y}$ , given in (6), can be expressed as

$$f_Y(\mathbf{y}) = \frac{g(\mathbf{v}'(\mathbf{D}_{\boldsymbol{\beta}} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\beta}})^{-1} \mathbf{v}) \prod_{k=1}^p \frac{|y_k|^{\lambda_k-1}}{\mu_k^{\lambda_k}}}{\int_{R(\mathbf{D}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\lambda})} g(\mathbf{v}'(\mathbf{D}_{\boldsymbol{\beta}} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\beta}})^{-1} \mathbf{v}) d\mathbf{v}}, \quad \mathbf{y} \in \mathbb{R}_+^p,$$

where  $\mathbf{v} = \mathbf{D}_{\boldsymbol{\beta}} T_{\boldsymbol{\lambda}, \boldsymbol{\mu}}(\mathbf{y})$  has its  $k$ th component given by

$$v_k = \begin{cases} \frac{[(y_k/\mu_k)^{\beta_k}]^{\lambda_k/\beta_k} - 1}{\lambda_k/\beta_k}, & \lambda_k \neq 0, \\ \log(y_k/\mu_k)^{\beta_k}, & \lambda_k = 0, \end{cases}$$

for  $k = 1, \dots, p$ . From  $U_k = (y_k/\mu_k)^{\beta_k}$ ,  $k = 1, \dots, p$ , with Jacobian  $J(\mathbf{y} \rightarrow \mathbf{u}) = \prod_{k=1}^p \mu_k \beta_k^{-1} u_k^{1/\beta_k - 1}$ , we arrive at the desired result.

3. Plugging  $\boldsymbol{\lambda} = \mathbf{1}$  in (6) we have that the PDF of  $\mathbf{Y}$  is

$$f_Y(\mathbf{y}) = \frac{g((\mathbf{y} - \boldsymbol{\mu})'(\mathbf{D}_{\boldsymbol{\mu}} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\mu}})^{-1}(\mathbf{y} - \boldsymbol{\mu})) \prod_{k=1}^p \frac{1}{\mu_k}}{\int_{R(\mathbf{1})} g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) d\mathbf{w}}, \quad \mathbf{y} \in \mathbb{R}_+^p.$$

From the change of variables  $\mathbf{w} = \mathbf{D}_{\boldsymbol{\mu}}^{-1}(\mathbf{y} - \boldsymbol{\mu})$  we arrive at the desired result.

## Appendix D: Proof of Theorem 4

Plugging  $\Sigma_{12} = \mathbf{0}$  in (8), and then making the change of variables  $s = T(\mathbf{w}_2) = \Sigma_{22}^{-1/2} \mathbf{w}_2$ , the marginal PDF of  $\mathbf{Y}_1$  is

$$f_{\mathbf{Y}_1}(\mathbf{y}_1) = \frac{\left\{ \int_{R(\lambda_2)} g(\mathbf{w}'_1 \Sigma_{11}^{-1} \mathbf{w}_1 + \mathbf{w}'_2 \Sigma_{22}^{-1} \mathbf{w}_2) d\mathbf{w}_2 \right\} \prod_{k=1}^r \frac{y_k^{\lambda_k-1}}{\mu_k^{\lambda_k}}}{\int_{R(\lambda_1)} \left\{ \int_{R(\lambda_2)} g(\mathbf{w}'_1 \Sigma_{11}^{-1} \mathbf{w}_1 + \mathbf{w}'_2 \Sigma_{22}^{-1} \mathbf{w}_2) d\mathbf{w}_2 \right\} d\mathbf{w}_1} \\ = \frac{g_1(\mathbf{w}'_1 \Sigma_{11}^{-1} \mathbf{w}_1) \prod_{k=1}^r \frac{y_k^{\lambda_k-1}}{\mu_k^{\lambda_k}}}{\int_{R(\lambda_1)} g_1(\mathbf{w}'_1 \Sigma_{11}^{-1} \mathbf{w}_1) d\mathbf{w}_1}, \quad \mathbf{w}_1 = T_{\lambda_1, \mu_1}(\mathbf{y}_1), \quad \mathbf{y}_1 \in \mathbb{R}_+^r.$$

Note that  $g_1(u) = \int_{T(R(\lambda_2))} g(u+s's) ds \leq \int_{\mathbb{R}^{p-r}} g(u+s's) ds = h_1(u)$ ,  $u \geq 0$ , where  $h_1$  is such that  $\int_0^\infty t^{r-1} h_1(t^2) dt < \infty$  (Fang et al. 1990, Sec. 2.2). This completes the proof.

## Appendix E: Proof of Theorem 5

Because  $\mathbf{Y} \sim \text{LE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$ , with  $g(u) \propto \int_0^\infty t^{p/2} \exp(-ut/2) dH(t)$ ,  $u \geq 0$ , then  $\mathbf{X} = T_{\mathbf{0}, \boldsymbol{\mu}}(\mathbf{Y}) \sim \text{E}\ell_p(\mathbf{0}, \boldsymbol{\Sigma}; g)$ . Thus,  $\mathbf{X}_1 = T_{\mathbf{0}, \boldsymbol{\mu}_1}(\mathbf{Y}_1) \sim \text{E}\ell_p(\mathbf{0}, \boldsymbol{\Sigma}_{11}; g)$  (Kano 1994). Hence,  $\mathbf{Y}_1 \sim \text{LE}\ell_r(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}; g)$ .

## Appendix F: Proof of Theorem 6

The conditional PDF of  $\mathbf{Y}_1|\mathbf{Y}_2$  is given by

$$f_{\mathbf{Y}_1|\mathbf{Y}_2}(\mathbf{y}_1) = \frac{g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) \prod_{k=1}^r \frac{y_k^{\lambda_k-1}}{\mu_k^{\lambda_k}}}{\int_{R(\lambda_1)} g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) d\mathbf{w}_1}, \quad \mathbf{w}_1 = T_{\lambda_1, \mu_1}(\mathbf{y}_1), \quad \mathbf{y}_1 \in \mathbb{R}_+^r. \quad (18)$$

Because  $\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w} = \mathbf{u}'_1 (\mathbf{D}_{\boldsymbol{\alpha}(\mathbf{w}_2)}^{-1} \boldsymbol{\Sigma}_{11 \cdot 2} \mathbf{D}_{\boldsymbol{\alpha}(\mathbf{w}_2)}^{-1})^{-1} \mathbf{u}_1 + q(\mathbf{w}_2)$ , where  $\mathbf{u}_1 = \mathbf{D}_{\boldsymbol{\alpha}(\mathbf{w}_2)}^{-1} (\mathbf{w}_1 - \boldsymbol{\mu}_1(\mathbf{w}_2)) = T_{\lambda_1, \delta_1}(\mathbf{y}_1)$ , (18) can be expressed as

$$f_{\mathbf{Y}_1|\mathbf{Y}_2}(\mathbf{y}_1) = \frac{g_q(\mathbf{w}_2) (\mathbf{u}'_1 (\mathbf{D}_{\boldsymbol{\alpha}(\mathbf{w}_2)}^{-1} \boldsymbol{\Sigma}_{11 \cdot 2} \mathbf{D}_{\boldsymbol{\alpha}(\mathbf{w}_2)}^{-1})^{-1} \mathbf{u}_1) \prod_{k=1}^r \frac{y_k^{\lambda_k-1}}{\mu_k^{\lambda_k} (1 + \lambda_k \mu_{1k}(\mathbf{w}_2))}}{\int_{R(\lambda_1)} g_q(\mathbf{w}_2) (\mathbf{u}'_1 (\mathbf{D}_{\boldsymbol{\alpha}(\mathbf{w}_2)}^{-1} \boldsymbol{\Sigma}_{11 \cdot 2} \mathbf{D}_{\boldsymbol{\alpha}(\mathbf{w}_2)}^{-1})^{-1} \mathbf{u}_1) d\mathbf{u}_1}, \quad \mathbf{y}_1 \in \mathbb{R}_+^p,$$

where  $\mathbf{u}_1 = T_{\lambda_1, \delta_1}(\mathbf{y}_1)$ . Since  $\prod_{k=1}^r \mu_k^{\lambda_k} (1 + \lambda_k \mu_{1k}(\mathbf{w}_2)) = \prod_{k=1}^r \delta_k^{\lambda_k}$  the proof is complete.

## Appendix G: Proof of Theorem 7

$Y_1$  and  $Y_2$  are independent if, and only if, the PDF of  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$  given in (6) is such that  $f_{\mathbf{Y}}(\mathbf{y}) = f_{Y_1}(y_1)f_{Y_2}(y_2)$ . This condition is satisfied if, and only if,  $\boldsymbol{\Sigma}_{12} = \mathbf{0}$  and the DGF  $g$  satisfies the functional equation  $g(u+v) = g(u)g(v)$ , with  $u \geq 0$  and  $v \geq 0$ , for which  $g(u) = \exp(-ku)$ , for some  $k \geq 0$ , is a solution (Gupta et al. 2013, Sec. 1.3). From  $\int_0^\infty t^{p-1} \exp(-kt^2) dt = 2^{p/2-1} \Gamma(p/2)$ , we find that  $k = 1/2$ . Hence,  $Y_1$  and  $Y_2$  are independent if, and only if,  $\boldsymbol{\Sigma}_{12} = \mathbf{0}$  and  $\mathbf{Y} \sim \text{BCN}_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma})$ .

## Appendix H: Proof of Theorem 8

From (6) we have

$$\mathbb{E}\left(\prod_{k=1}^p Y_k^{h_k}\right) = \frac{\int_{\mathbb{R}_+^p} g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) \prod_{k=1}^p \frac{y_k^{\lambda_k+h_k-1}}{\mu_k^{\lambda_k}} d\mathbf{y}}{\int_{R(\boldsymbol{\lambda})} g(\mathbf{w}' \boldsymbol{\Sigma}^{-1} \mathbf{w}) d\mathbf{w}},$$

where  $\mathbf{w} = T_{\boldsymbol{\lambda}, \boldsymbol{\mu}}(\mathbf{y})$ . By making the change of variables  $\mathbf{u} = \mathbf{D}_{\boldsymbol{\mu}}^{-1} \mathbf{y}$  we arrive at the desired result.

## Appendix I: Marginal PDF of $Y_k$

The function  $g_{\boldsymbol{\gamma}_k}$  given in (10) can be defined in  $\mathbb{R}$ . Hence, we can define a random variable  $U_k \in \mathbb{R}$  from the PDF

$$f_{U_k}(u_k) = c_k g_{\boldsymbol{\gamma}_k}(u_k), \quad u_k \in \mathbb{R},$$

where  $c_k^{-1} = \int_{-\infty}^\infty g_{\boldsymbol{\gamma}_k}(t) dt$ . The CDF of  $U_k$  is given by (11). We now define  $S_k \in I(\lambda_k \sqrt{\sigma_{kk}})$  as a random variable  $U_k$  truncated on  $I(\lambda_k \sqrt{\sigma_{kk}})$ . The PDF of  $S_k$  is given by

$$f_{S_k}(s_k) = \frac{g_{\boldsymbol{\gamma}_k}(s_k)}{\int_{I(\lambda_k \sqrt{\sigma_{kk}})} g_{\boldsymbol{\gamma}_k}(s_k) ds_k}, \quad s_k \in I(\lambda_k \sqrt{\sigma_{kk}}). \quad (19)$$

From the transformation  $S_k = \sigma_{kk}^{-1/2} T_{\lambda_k, \mu_k}(Y_k)$ , with Jacobian  $J(s_k \rightarrow y_k) = \sigma_{kk}^{-1/2} \mu_k^{-\lambda_k} y_k^{\lambda_k-1}$ , we arrive at the PDF of  $Y_k$  given in (9).

## Appendix J: Proof of Theorem 9

Because  $\mathbf{Y} \sim \text{BCE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}; g)$ , the PDF of  $Y_k, k = 1, \dots, p$ , is given by (9), where  $S_k = \sigma_{kk}^{-1/2} T_{\lambda_k, \mu_k}(Y_k)$  has PDF given in (19). The CDF of  $S_k$  is

$$F_{S_k}(s_k) = \begin{cases} \frac{F_{U_k}(s_k) - F_{U_k}(-1/\lambda_k \sqrt{\sigma_{kk}})}{1 - F_{U_k}(-1/\lambda_k \sqrt{\sigma_{kk}})}, & \lambda_k > 0, \\ \frac{1 - F_{U_k}(-1/\lambda_k \sqrt{\sigma_{kk}}) + F_{U_k}(s_k)}{F_{U_k}(-1/\lambda_k \sqrt{\sigma_{kk}})}, & \lambda_k < 0, \\ F_{U_k}(s_k), & \lambda_k = 0, \end{cases} \quad (20)$$

from which we have that the  $\alpha$ -quantile  $y_{k,\alpha}$  of  $Y_k, \alpha \in (0, 1)$ , is such that  $P(Y_k \leq y_{k,\alpha}) = \alpha$ , or equivalently  $P[S_k \leq \sigma_{kk}^{-1/2} T_{\lambda_k, \mu_k}(y_{k,\alpha})] = \alpha$ . Hence,

$$y_{k,\alpha} = \begin{cases} \mu_k (1 + \lambda_k \sqrt{\sigma_{kk}} s_{k,\alpha})^{1/\lambda_k}, & \lambda_k \neq 0, \\ \mu_k \exp(\sqrt{\sigma_{kk}} s_{k,\alpha}), & \lambda_k = 0, \end{cases}$$

where  $s_{k,\alpha}$  is such that  $F_{S_k}(s_{k,\alpha}) = \alpha$ , with  $F_{S_k}$  given in (20). Therefore,  $s_{k,\alpha}$  is given by

$$s_{k,\alpha} = \begin{cases} F_{U_k}^{-1}(\alpha + (1 - \alpha)F_{U_k}(-1/\lambda_k \sqrt{\sigma_{kk}})), & \lambda_k > 0, \\ F_{U_k}^{-1}((1 + \alpha)F_{U_k}(-1/\lambda_k \sqrt{\sigma_{kk}}) - 1), & \lambda_k < 0, \\ F_{U_k}^{-1}(\alpha), & \lambda_k = 0, \end{cases}$$

where  $F_{U_k}$  is the CDF given in (11).

## References

Arellano-Valle RB, Branco MD, Genton MG (2006) A unified view on skewed distributions arising from selections. *Can J Stat* 34(4):581–601

Birnbaum Z, Meyer PL (1953) On the effect of truncation in some or all coordinates of a multi-normal population. *J Indian Soc Agric Stat* 5:17–28

Fang KT, Kotz S, Ng KW (1990) Symmetric multivariate and related distributions. Chapman and Hall, London

Ferrari SLP, Fumes G (2017) Box–Cox symmetric distributions and applications to nutritional data. *Adv Stat Anal* 101:321–344

Genz A, Bretz F (2009) Computation of multivariate normal and  $t$  probabilities. Springer, Heidelberg

Gómez-Sánchez-Manzano E, Gómez-Villegas MA, Marín JM (2008) Multivariate exponential power distributions as mixtures of normal distributions with Bayesian applications. *Commun Stat Theory Methods* 37(6):972–985

Gupta AK, Varga T, Bodnar T (2013) Elliptically contoured models in statistics and portfolio theory. Springer, New York

Ho HJ, Lin TI, Chen HY, Wang WL (2012) Some results on the truncated multivariate  $t$  distribution. *J Stat Plan Inference* 142:25–40

Horrace WC (2005) Some results on the multivariate truncated normal distribution. *J Multivar Anal* 94:209–221

Kan R, Robotti C (2017) On moments of folded and truncated multivariate normal distributions. *J Comput Graph Stat* 26(4):930–934

Kano Y (1994) Consistency property of elliptical probability density functions. *J Multivar Anal* 51:139–147

Kim H-J (2010) A class of weighted multivariate elliptical models useful for robust analysis of nonnormal and bimodal data. *J Korean Stat Soc* 39:83–92

Lange KL, Little RJA, Taylor JMG (1989) Robust statistical modeling using the  $t$  distribution. *J Am Stat Assoc* 84(408):881–896

Manjunath BG, Wilhelm S (2012). Moments calculation for the double truncated multivariate normal density. Working paper [arXiv:1206.5387v1](https://arxiv.org/abs/1206.5387v1)

Marchenko YV, Genton MG (2010) Multivariate log-skew-elliptical distributions with applications to precipitation data. *Environmetrics* 21(3–4):318–340

Nocedal J, Wright SJ (2006) Numerical optimization. Springer, New York

Quiroz AJ, Nakamura M, Pérez FJ (1996) Estimation of a multivariate Box–Cox transformation to elliptical symmetry via the empirical characteristic function. *Ann Inst Stat Math* 48(4):687–709

R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna <https://www.R-project.org/>

Rigby RA, Stasinopoulos DM (2004) Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. *Stat Med* 23(19):3053–3076

Rigby RA, Stasinopoulos DM (2006) Using the Box–Cox  $t$  distribution in GAMLSS to model skewness and kurtosis. *Stat Model* 6(3):209–229

Rousseeuw PJ, Ruts I, Tukey JW (1999) The Bagplot: a bivariate boxplot. *Am Stat* 53(4):382–387

Stasinopoulos DM, Rigby RA, Akantziliotou C (2008). Instructions on how to use the GAMLSS package in R. London. <http://www.gamlss.org>

Tallis GM (1961) The moment generating function of the truncated multi-normal distribution. *J R Stat Soc* 23:223–229

Tallis GM (1963) Elliptical and radial truncation in normal populations. *Ann Math Stat* 34(3):940–944

Tallis GM (1965) Plane truncation in normal populations. *J R Stat Soc* 27(2):301–307

Voudouris V, Gilchrist R, Rigby RA, Sedgwick J, Stasinopoulos DM (2012) Modelling skewness and kurtosis with the BCPE density in GAMLSS. *J Appl Stat* 39(6):1279–1293