Prediction via L-statistics in multivariate elliptical distributions

Abbreviated title: L-statistics and prediction

A. R. de Leon[1], H. Mahmoodian[2], and A. Jamalizadeh[2, *]
[1]*Department of Mathematics & Statistics, University of Calgary, Calgary, Canada*
[2]*Department of Statistics, Shahid Bahonar University, Kerman, Iran*

Abstract

We consider random vectors $\mathbf{X}_{K\times 1}$ and $\mathbf{Y}_{N\times 1}$ having a multivariate elliptical joint distribution, and derive the exact joint distribution of $\mathbf{X}$ and L-statistics from $\mathbf{Y}$, as a mixture of multivariate unified skew-elliptical distributions. This mixture representation enables us to predict $\mathbf{X}$ based on L-statistics from $\mathbf{Y}$, and vice versa, when $K = 1$ and with normal and t-distributions. Our results extend and generalize previous ones in two ways: first, we consider a general multivariate set-up for which $K \geq 1$ and $N \geq 2$, and second, we adopt the multivariate elliptical distribution to include previous multivariate normal and t-formulations as special cases. We illustrate our results using data on student test scores.

Key words: *Multivariate unified skew-elliptical distribution; order statistics; mixture distribution; linear combination; squared-error loss.*

1. Introduction

A motivation for the paper is the following problem: in the calculation of a student's overall course grade, only the 'best–n' scores among $N \geq n$ quizzes are considered along with his final test mark. It is usual practice to use the quiz scores to predict a student's final test mark. More formally, if X is the final test mark and $\mathbf{Y} = (Y_1, \cdots, Y_N)^\top$ are the quiz scores, we wish to study X in terms of some linear combination $\sum_{i=1}^{N} a_i Y_{(i)}$, where $Y_{(1)} < \cdots < Y_{(N)}$ are the ordered quiz scores. In our scenario, note that we have $a_1 = \cdots = a_n = 0$.

The above problem is akin to the following that arises in electrical engineering and discussed by Wiens et al. (2006). In processing cellular phone signals from several antennae, receivers normally select only the strongest signals to reduce signal fading. Specifically, if a receiver receives N signals, only the $n \leq N$ strongest signals will be processed, which are then combined and used to analyze and predict the transmission system's performance, as measured by $\mathbf{X} = (X_1, \cdots, X_K)^\top$, say.

Various incarnations of the above problem, simplified in some way, have been studied previously by several authors. Viana (1998) and Olkin and Viana (1995) obtained the best linear predictors in the trivariate normal distribution case with $N = 2$ and $K = 1$, where Y_1 and Y_2 are exchangeable

*Correspondence to: A. Jamalizadeh, Department of Statistics, Shahid Bahonar University, Kerman, Iran (E-mail: a.jamalizadeh@yahoo.com)

such that $(X, Y_1)^\top$ and $(X, Y_1)^\top$ share a common correlation. Loperfido (2008b) considered the same set-up and derived the exact joint distribution of X and $Y_{(2)} = max(Y_1, Y_2)$. Jamalizadeh and Balakrishnan (2009a) similarly derived the exact joint distribution of X and $\sum_{i=1}^{2} a_i Y_{(i)}$ in the case of a trivariate normal distribution for X, Y_1, Y_2, with arbitrary covariance structure. They showed that this joint distribution is a mixture of bivariate unified skew-normal distributions and obtained a predictor for X using linear combinations of order statistics from Y_1, Y_2; see also Balakrishnan et al. (2011) for the case of elliptical distributions.

Order statistics, their linear combinations (i.e., L-statistics), and their corresponding distributions have also been similarly widely studied. Early results are provided by Gupta and Pillai (1965), Basu and Ghosh (1978), Nagaraja (1982), and Balakrishnan (1993). More recent work include Genc (2006), who derived the exact distribution of L-statistics from the bivariate normal distribution; Arellano-Valle and Genton (2007,2008), who considered multivariate elliptical distributions; and Jamalizadeh and Balakrishnan (2008), who worked with bivariate skew-normal and skew-t distributions. Additional results are given by Jamalizadeh et al. (2009a,2009b), Jamalizadeh and Balakrishnan (2009b,2010), and Loperfido (2008a).

In this paper, we consider the general case of $N > 2$ and $K > 1$, and assume an elliptical joint distribution for $\mathbf{X}$ and $\mathbf{Y}$, i.e., let

$$\left(\frac{\mathbf{X}}{\mathbf{Y}}\right) = (X_1, \cdots, X_K | Y_1, \cdots, Y_N)^\top \sim \mathrm{EC}_{K+N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(K+N)}), \tag{1}$$

the $(K+N)$-dimensional elliptical distribution with location parameter $\boldsymbol{\mu}$, dispersion-shape matrix $\boldsymbol{\Sigma}$, and density generator function $h^{(K+N)}$. Note that this specification includes the multivariate normal and t-distributions, among others, and generalizes previous cases studied in the literature. With $\mathbf{Y}_{(N)} = (Y_{(1)}, \cdots, Y_{(N)})^\top$ as the vector of order statistics from $\mathbf{Y}$, we derive the exact joint distribution of $\mathbf{X}$ and $\mathbf{L}\mathbf{Y}_{(N)}$, where $\mathbf{L}$ is a $P \times N$ matrix of $rank(\mathbf{L}) = P$. We show that this joint distribution is a mixture of multivariate unified skew-elliptical distributions, and obtain in the process, in the special case when $K = 1$ and $\mathbf{L} = \mathbf{a}^\top = (a_1, \cdots, a_N)$, the best (nonlinear) predictors of X based on $\mathbf{a}^\top \mathbf{Y}_{(N)}$, and of $\mathbf{a}^\top \mathbf{Y}_{(N)}$ based on X, under square loss function in the case of normal and t-distributions. We also present a mixture representation for the joint cumulative distribution function (CDF) of X and $Y_{(r)}$, $r = 1, \cdots, N$, in terms of bivariate unified skew-elliptical distributions.

We organize the paper as follows. Section 2 presents a brief review of skew-elliptical distribution theory and presents specialized results for normal and t-distributions in the univariate and bivariate cases. The main results of the paper are then obtained in section 3. An illustration of the paper's results is then given in section 4. Section 5 then concludes the paper.

2. Skew-elliptical distributions: preliminaries

Consider the random vectors $\mathbf{X}_{K\times 1}$ and $\mathbf{Y}_{N\times 1}$ in (1) and suppose we partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as follows:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{yx}^\top \\ & \boldsymbol{\Sigma}_{yy} \end{pmatrix}, \tag{2}$$

where $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ are respective $K \times 1$ and $N \times 1$ location vectors, $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{yy}$ are respective $K \times K$ and $N \times N$ dispersion matrices, and $\boldsymbol{\Sigma}_{yx}$ is a $N \times K$ shape matrix. A random vector $\mathbf{U}_{N\times 1}$ is defined to have a multivariate unified skew-elliptical (SUE) distribution if and only if

$$\mathbf{U} \stackrel{d}{=} (\mathbf{Y} | \mathbf{X} > \mathbf{0}), \tag{3}$$

where "$\overset{d}{=}$" denotes equality in distribution, and it is understood that the inequality "$\mathbf{X} > \mathbf{0}$" must hold for each of the components of $\mathbf{X}$; we write $\mathbf{U} \sim \text{SUE}_{N,K}(\boldsymbol{\mu}_y, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yx}, h^{(K+N)})$. A closed-form expression for the corresponding probability density function (PDF) is given by Arellano-Valle and Azzalini (2006) and Arellano-Valle and Genton (2005, 2010).

Taking $h^{(K+N)}(u) = \phi^{(K+N)}(u) = (2\pi)^{-(K+N)/2}\exp(-u/2)$, $u > 0$, we obtain the multivariate unified skew-normal distribution (SUN). Given $\mathbf{U} \sim \text{SUN}_{N,K}(\boldsymbol{\mu}_y, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yx})$, its respective PDF $f_{\mathbf{U}}(\cdot)$ and moment-generating function (MGF) $M_{\mathbf{U}}(\cdot)$ are given by

$$f_{\mathbf{U}}(\mathbf{u}) = \frac{\phi_N(\mathbf{u}; \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})\Phi_K(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{yx}^\top \boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{u} - \boldsymbol{\mu}_y); \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{yx}^\top \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})}{\Phi_K(\boldsymbol{\mu}_x; \boldsymbol{\Sigma}_{xx})} \quad (4)$$

$$M_{\mathbf{U}}(\mathbf{s}) = \frac{\exp\left(\boldsymbol{\mu}_y^\top \mathbf{s} + \frac{1}{2}\mathbf{s}^\top \boldsymbol{\Sigma}_{yy}\mathbf{s}\right)\Phi_K(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{yx}^\top \mathbf{s}; \boldsymbol{\Sigma}_{xx})}{\Phi_K(\boldsymbol{\mu}_x; \boldsymbol{\Sigma}_{xx})}, \quad (5)$$

where $\phi_Q(\cdot; \boldsymbol{\Sigma})$ and $\Phi_Q(\cdot; \boldsymbol{\Sigma})$ are the PDF and CDF of the centered Q-dimensional normal distribution with dispersion matrix $\boldsymbol{\Sigma}$, respectively.

In the case of the t_ν-kernel distribution with generator

$$h^{(K+N)}(u) = t_\nu^{(K+N)}(u) = \frac{\Gamma\left(\frac{\nu+K+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{K+N}{2}}}\left(1 + \frac{u}{\nu}\right)^{-(\nu+K+N)/2}, \quad (6)$$

for $u \geq 0$, $\nu > 0$, where $\Gamma(\cdot)$ is the gamma function, we generate the multivariate unified skew-t (SUT) distribution, with PDF

$$\begin{aligned} f_{\mathbf{U}}(\mathbf{u}) &= T_K\left(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{yx}^\top \boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{u} - \boldsymbol{\mu}_y); \frac{\nu + (\mathbf{u} - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{u} - \boldsymbol{\mu}_y)}{\nu + N}(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{yx}^\top \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}), \nu + N\right) \\ &\quad \times \frac{t_N(\mathbf{u}; \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}, \nu)}{T_K(\boldsymbol{\mu}_x; \boldsymbol{\Sigma}_{xx}, \nu)}, \end{aligned} \quad (7)$$

where $t_Q(\cdot; \boldsymbol{\Sigma}, \nu)$ and $T_Q(\cdot; \boldsymbol{\Sigma}, \nu)$ are the respective PDF and CDF of the centered Q-dimensional t-distribution with scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν; we write $\mathbf{U} \sim \text{SUT}_{N,K}(\boldsymbol{\mu}_y, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yx}, \nu)$.

These distributions were developed recently in Jamalizadeh and Balakrishnan (2011), who obtained marginal and conditional distributions, linear transformations, moments, etc. In what follows, we study univariate and bivariate SUN and SUT distributions and derive moment expressions, among others, for later use in section 3.

2.1. Univariate case

We now consider the univariate class of SUE distributions that arises from (1) with $K \geq 1$ and $N = 1$. In this case, (2) becomes

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \mu_y \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\sigma}_{xy} \\ & \sigma_{yy} \end{pmatrix}. \quad (8)$$

For $U \sim \text{SUN}_{1,K}(\mu_y, \boldsymbol{\mu}_x, \sigma_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\sigma}_{xy})$, we get

$$E(U) = \mu_y + \frac{1}{\Phi_K(\boldsymbol{\mu}_x; \boldsymbol{\Sigma}_{xx})}\sum_{i=1}^{K}\frac{\sigma_{xy,i}}{\sqrt{\sigma_{xx,ii}}}\phi\left(\frac{\mu_{x,i}}{\sqrt{\sigma_{xx,ii}}}\right)\Phi_{K-1}\left(\boldsymbol{\mu}_{x,-i} - \frac{\mu_{x,i}}{\sigma_{xx,ii}}\boldsymbol{\sigma}_{xx,-ii}; \boldsymbol{\Sigma}_{xx,-i|i}\right), \quad (9)$$

where, for some i,

$$\boldsymbol{\sigma}_{xy} = \begin{pmatrix} \sigma_{xy,i} \\ \boldsymbol{\sigma}_{xy,-i} \end{pmatrix}, \quad \boldsymbol{\mu}_x = \begin{pmatrix} \mu_{x,i} \\ \boldsymbol{\mu}_{x,-i} \end{pmatrix}, \quad \boldsymbol{\Sigma}_{xx} = \begin{pmatrix} \sigma_{xx,ii} & \boldsymbol{\sigma}_{xx,-ii}^\top \\ & \boldsymbol{\Sigma}_{xx,-i-i} \end{pmatrix},$$

with $\boldsymbol{\Sigma}_{xx,-i|i} = \boldsymbol{\Sigma}_{xx,-i-i} - \boldsymbol{\sigma}_{xx,-ii}\boldsymbol{\sigma}_{xx,-ii}^\top/\sigma_{xx,ii}$, and with $\phi(\cdot)$ the standard normal PDF. Expression (9) is easily obtained by differentiating (5) and using the following:

$$\begin{aligned} \frac{\partial}{\partial s}\Phi_K(\boldsymbol{\mu}_x + s\boldsymbol{\sigma}_{xy}; \boldsymbol{\Sigma}_{xx}) &= \sum_{i=1}^K \frac{\sigma_{xy,i}}{\sqrt{\sigma_{xx,ii}}}\phi\left(\frac{\mu_{x,i} + s\sigma_{xy,i}}{\sqrt{\sigma_{xx,ii}}}\right)\Phi_{K-1}\left(\left\{\boldsymbol{\sigma}_{xy,-i} - \frac{\sigma_{xy,i}}{\sigma_{xx,ii}}\boldsymbol{\sigma}_{xx,-ii}\right\}s\right. \\ &\quad \left. + \boldsymbol{\mu}_{x,-i} - \frac{\mu_{x,i}}{\sigma_{xx,ii}}\boldsymbol{\sigma}_{xx,-ii}; \boldsymbol{\Sigma}_{xx,-i|i}\right). \end{aligned}$$

Next, if $U \sim \mathrm{SUT}_{1,K}(\mu_y, \boldsymbol{\mu}_x, \sigma_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\sigma}_{xy}, \nu)$, we similarly obtain

$$\begin{aligned} E(U) &= \mu_y + \frac{\nu^{\nu/2}\Gamma\left(\frac{\nu-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)T_K(\boldsymbol{\mu}_x; \boldsymbol{\Sigma}_{xx}, \nu)}\sum_{i=1}^K \frac{\sigma_{xy,i}}{\sqrt{\sigma_{xx,ii}}}\left(\nu + \frac{\mu_{x,i}^2}{\sigma_{xx,ii}}\right)^{-(\nu-1)/2} \\ &\quad \times T_{K-1}\left(\frac{\sqrt{\nu-1}}{\sqrt{\nu + \frac{\mu_{x,i}^2}{\sigma_{xx,ii}}}}\left(\boldsymbol{\mu}_{x,-i} - \frac{\mu_{x,i}}{\sigma_{xx,ii}}\boldsymbol{\sigma}_{xx,-ii}\right); \boldsymbol{\Sigma}_{xx,-i|i}, \nu - 1\right). \end{aligned} \tag{10}$$

This follows from (9) and from the result that for a χ_ν^2 random variable νV with ν degrees of freedom,

$$E\left\{V^{-1/2}\phi(aV^{1/2})\Phi_Q(V^{1/2}\mathbf{b}; \boldsymbol{\Sigma})\right\} = \frac{\nu^{\nu/2}\Gamma\left(\frac{\nu-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}(\nu + a^2)^{-(\nu-1)2}T_Q\left(\frac{\sqrt{\nu-1}}{\sqrt{\nu+a^2}}\mathbf{b}; \boldsymbol{\Sigma}, \nu - 1\right),$$

for any real number a, any $Q \times 1$ real vector $\mathbf{b}$, and any positive definite $Q \times Q$ matrix $\boldsymbol{\Sigma}$.

2.2. Bivariate case

Next, we consider the case $N = 2$ and $K \geq 1$. Let

$$\boldsymbol{\Sigma}_{yy} = \begin{pmatrix} \sigma_{yy,11} & \sigma_{yy,12} \\ & \sigma_{yy,22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{yx} = \begin{pmatrix} \boldsymbol{\sigma}_{yx1}^\top \\ \boldsymbol{\sigma}_{yx2}^\top \end{pmatrix}. \tag{11}$$

For $\mathbf{U} = (U_1, U_2)^\top \sim \mathrm{SUN}_{2,K}(\boldsymbol{\mu}_y, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yx})$, it follows from Arellano-Valle and Genton (2010) that $U_1 \sim \mathrm{SUN}_{1,K}(\mu_{y,1}, \boldsymbol{\mu}_x, \sigma_{yy,11}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\sigma}_{yx1})$. In addition, we get

$$U_2|U_1 = u_1 \sim \mathrm{SUN}_{1,K}(\mu_y^{2\cdot1}(u_1), \boldsymbol{\mu}_x^{2\cdot1}(u_1), \sigma_{yy}^{22\cdot1}, \boldsymbol{\Sigma}_{xx}^{2\cdot1}, \boldsymbol{\sigma}_{yx}^{2\cdot1}), \tag{12}$$

where $\mu_y^{2\cdot1}(u_1) = \mu_{y,2} + \sigma_{yy,12}(u_1 - \mu_{y,1})/\sigma_{yy,11}$, $\boldsymbol{\mu}_x^{2\cdot1}(u_1) = \boldsymbol{\mu}_x + \boldsymbol{\sigma}_{yx1}(u_1 - \mu_{y,1})/\sigma_{yy,11}$, $\sigma_{yy}^{22\cdot1} = \sigma_{yy,22} - \sigma_{yy,12}^2/\sigma_{yy,11}$, $\boldsymbol{\Sigma}_{xx}^{2\cdot1} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\sigma}_{yx1}\boldsymbol{\sigma}_{yx1}^\top/\sigma_{yy,11}$, and $\boldsymbol{\sigma}_{yx}^{2\cdot1} = \boldsymbol{\sigma}_{yx2} - \sigma_{yy,12}\boldsymbol{\sigma}_{yx1}/\sigma_{yy,11}$. It can also be shown that

$$E(U_2|U_1 = u_1) = \mu_y^{2\cdot1}(u_1) + \frac{1}{\Phi_K(\boldsymbol{\mu}_x^{2\cdot1}(u_1); \boldsymbol{\Sigma}_{xx}^{2\cdot1})}\sum_{i=1}^K \frac{\sigma_{yx,i}^{2\cdot1}}{\sqrt{\sigma_{xx,ii}^{2\cdot1}}}\phi\left(\frac{\mu_{x,i}^{2\cdot1}(u_1)}{\sqrt{\sigma_{xx,ii}^{2\cdot1}}}\right)$$

$$\times\Phi_{K-1}\left(\boldsymbol{\mu}_{x,-1}^{2\cdot1}(u_1)-\frac{\mu_{x,i}^{2\cdot1}(u_1)}{\sigma_{xx,ii}^{2\cdot1}}\boldsymbol{\sigma}_{xx,-ii}^{2\cdot1};\boldsymbol{\Sigma}_{xx,-i|i}^{2\cdot1}\right), \tag{13}$$

where $\sigma_{xx,ii}^{2\cdot1}=\sigma_{xx,ii}-\sigma_{yx1,i}/\sigma_{yy,11}$, $\boldsymbol{\mu}_{x,-i}^{2\cdot1}(u_1)=\boldsymbol{\mu}_{x,-i}+\boldsymbol{\sigma}_{yx1,-i}(u_1-\mu_{y,1})/\sigma_{yy,11}$, $\boldsymbol{\sigma}_{xx,-ii}^{2\cdot1}=\boldsymbol{\sigma}_{xx,-ii}-\sigma_{yx1,i}\boldsymbol{\sigma}_{yx1,-i}/\sigma_{yy,11}$,

$$\boldsymbol{\Sigma}_{xx,-i|i}^{2\cdot1}=\boldsymbol{\Sigma}_{xx,-i-i}-\frac{\boldsymbol{\sigma}_{yx1,-i}\boldsymbol{\sigma}_{yx1,-i}^{\top}}{\sigma_{yx,11}}-\frac{\left(\boldsymbol{\sigma}_{xx,-ii}-\frac{\sigma_{yx1,i}\boldsymbol{\sigma}_{yx1,-i}}{\sigma_{yy,11}}\right)\left(\boldsymbol{\sigma}_{xx,-ii}-\frac{\sigma_{yx1,i}\boldsymbol{\sigma}_{yx1,-i}}{\sigma_{yy,11}}\right)^{\top}}{\sigma_{xx,ii}-\frac{\sigma_{yx1,i}}{\sigma_{yy,11}}},$$

and where $\boldsymbol{\sigma}_{yx1}$, $\boldsymbol{\sigma}_{yx2}$, $\boldsymbol{\mu}_x$, and $\boldsymbol{\Sigma}_{xx}$ are similarly partitioned as in section 2.1. Analogous results may be similarly obtained for $\mathbf{U}=(U_1,U_2)^{\top}\sim\mathrm{SUT}_{2,K}(\boldsymbol{\mu}_y,\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_{yy},\boldsymbol{\Sigma}_{xx},\boldsymbol{\Sigma}_{yx},\nu)$:

$$U_1\sim\mathrm{SUT}_{1,K}(\mu_{y,1},\boldsymbol{\mu}_x,\sigma_{yy,11},\boldsymbol{\Sigma}_{xx},\boldsymbol{\sigma}_{yx1},\nu), \tag{14}$$

$$U_2|U_1=u_1\sim\mathrm{SUT}_{1,K}\left(\begin{array}{c}\mu_y^{2\cdot1}(u_1),\boldsymbol{\mu}_x^{2\cdot1}(u_1),q_1(u_1,\nu)\sigma_{yy}^{22\cdot1},\\ q_1(u_1,\nu)\boldsymbol{\Sigma}_{xx}^{2\cdot1},q_1(u_1,\nu)\boldsymbol{\sigma}_{yx}^{2\cdot1},\nu+1\end{array}\right), \tag{15}$$

$$\begin{aligned}E(U_2|U_1=u_1)&=\mu_y^{2\cdot1}(u_1)+\frac{(\nu+1)^{(\nu+1)/2}q_1^{1/2}(u_1,\nu)\Gamma\left(\frac{\nu}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu+1}{2}\right)T_K(\boldsymbol{\mu}_x^{2\cdot1}(u_1);q_1(u_1,\nu)\boldsymbol{\Sigma}_x^{2\cdot1},\nu+1)}\sum_{i=1}^{K}\frac{\sigma_{yx,i}^{2\cdot1}}{\sqrt{\sigma_{xx,ii}^{2\cdot1}}}\\&\times T_{K-1}\left(\frac{\sqrt{\nu}\left(\boldsymbol{\sigma}_{x,-i}^{2\cdot1}(u_1)-\frac{\sigma_{xi}^{2\cdot1}(u_1)}{\sigma_{xx,ii}^{2\cdot1}}\boldsymbol{\sigma}_{xx,-ii}^{2\cdot1}\right)}{\sqrt{\nu+1+\frac{\{\mu_{x,i}^{2\cdot1}(u_1)\}^2}{q_1(u_1,\nu)\sigma_{xx,ii}^{2\cdot1}}}};q_1(u_1,\nu)\boldsymbol{\Sigma}_{xx,-i|i}^{2\cdot1},\nu\right)\\&\times\left(\nu+1+\frac{\{\mu_{x,i}^{2\cdot1}(u_1)\}^2}{q_1(u_1,\nu)\sigma_{xx,ii}^{2\cdot1}}\right)^{-\nu/2},\end{aligned} \tag{16}$$

where $\sigma_{yx,i}^{2\cdot1}$, $\sigma_{xx,ii}^{2\cdot1}$, $\mu_{x,i}^{2\cdot1}(u_1)$, $\boldsymbol{\mu}_{x,-1}^{2\cdot1}(u_1)$, $\sigma_{yx,-ii}^{2\cdot1}$, and $\boldsymbol{\Sigma}_{xx,-i|i}^{2\cdot1}$ are as defined previously, and

$$q_1(u_1,\nu)=\frac{\nu+\frac{(u_1-\mu_{x,1})^2}{\sigma_{yy,11}}}{\nu+1}. \tag{17}$$

Note that the above show that the class of SUN and SUT distributions are conveniently closed under marginalization and conditionalization.

3. Main results

Assume (1) holds, where $\boldsymbol{\Sigma}$ is positive definite. In this section, we show that $\mathbf{X}$ and $\mathbf{LY}_{(N)}$ are jointly distributed according to a mixture of SUE distributions, where $\mathbf{Y}_{(N)}=(Y_{(1)},\cdots,Y_{(N)})^{\top}$ is the vector of order statistics from $\mathbf{Y}$, and $\mathbf{L}$ is a $P\times N$ matrix of $rank(\mathbf{L})=P$. To this end, let $\mathbf{Y}_{(N)}\in\mathcal{P}(\mathbf{Y})$, where $\mathcal{P}(\mathbf{Y})=\{\mathbf{Y}_i:\mathbf{Y}_i=\mathbf{P}_i\mathbf{Y},i=1,\cdots,N!\}$ is the collection of vectors $\mathbf{Y}_i$ corresponding to the $N!$ different permutations of the components of $\mathbf{Y}$, with $\mathbf{P}_i$ a $N\times N$ permutation matrix such that $\mathbf{P}_i\neq\mathbf{P}_{i'}$, for all $i\neq i'$. Further, let $\mathbf{D}$ be an $(N-1)\times N$ difference matrix such that $\mathbf{DY}=(Y_2-Y_1,Y_3-Y_2,\cdots,Y_N-Y_{N-1})$, i.e., row i of $\mathbf{D}$ is given by $\mathbf{e}_{i+1}^{\top}-\mathbf{e}_i^{\top}$, $i=1,\cdots,N$, where $\mathbf{e}_1,\cdots,\mathbf{e}_N$ are N-dimensional unit basis vectors. We give below our main result on the joint, marginal and conditional distributions of $\mathbf{X}$ and $\mathbf{LY}_{(N)}$.

Proposition 1 *Assume (1) holds. Then the following are true:*

(i) *the joint CDF* $F_{\mathbf{X},\mathbf{LY}_{(N)}}(\cdot)$ *and joint PDF* $f_{\mathbf{X},\mathbf{LY}_{(N)}}(\cdot)$ *of* $\mathbf{X}$ *and* $\mathbf{LY}_{(N)}$ *is then given by*

$$F_{\mathbf{X},\mathbf{LY}_{(N)}}(\mathbf{x},\mathbf{w}) = \sum_{i=1}^{N!} \pi_i F_{N-1,K+P}^{h^{(K+N+P-1)}}(\mathbf{x},\mathbf{w};\boldsymbol{\Theta}_i), \tag{18}$$

$$f_{\mathbf{X},\mathbf{LY}_{(N)}}(\mathbf{x},\mathbf{w}) = \sum_{i=1}^{N!} \pi_i f_{N-1,K+P}^{h^{(K+N+P-1)}}(\mathbf{x},\mathbf{w};\boldsymbol{\Theta}_i), \tag{19}$$

where $F_{N-1,K+P}^{h^{(K+N+P-1)}}(\cdot;\boldsymbol{\Theta}_i)$ *and* $f_{N-1,K+P}^{h^{(K+N+P-1)}}(\cdot;\boldsymbol{\Theta}_i)$ *are the CDF and PDF of* $\mathrm{SUE}_{N-1,K+P}(\boldsymbol{\Theta}_i, h^{(K+N+P-1)})$, *and*

$$\pi_i = G_{N-1}^{h^{(N-1)}}(\boldsymbol{\eta}_i;\boldsymbol{\Gamma}_i), \tag{20}$$

with $G_{N-1}^{h^{(N-1)}}(\cdot;\boldsymbol{\Gamma}_i)$ *the CDF of* $\mathrm{EC}_{N-1}(\mathbf{0},\boldsymbol{\Gamma}_i,h^{(N-1)})$, $\boldsymbol{\Theta}_i = \{\boldsymbol{\xi}_i,\boldsymbol{\eta}_i,\boldsymbol{\Omega}_i,\boldsymbol{\Gamma}_i,\boldsymbol{\Lambda}_i\}$,

$$\boldsymbol{\xi}_i = \begin{pmatrix} \boldsymbol{\mu}_x \\ \mathbf{L}\boldsymbol{\mu}_{y,i} \end{pmatrix}, \quad \boldsymbol{\eta}_i = \mathbf{D}\boldsymbol{\mu}_{y,i}, \quad \boldsymbol{\Omega}_i = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{yx,i}^\top \mathbf{L}^\top \\ & \mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top \end{pmatrix},$$
$$\boldsymbol{\Gamma}_i = \mathbf{D}\boldsymbol{\Sigma}_{yy,i}\mathbf{D}^\top, \quad \boldsymbol{\Lambda}_i = \begin{pmatrix} \boldsymbol{\Sigma}_{yx,i}^\top \mathbf{D}^\top \\ \mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top \end{pmatrix},$$

$\boldsymbol{\mu}_{y,i} = \mathbf{P}_i\boldsymbol{\mu}_y$, $\boldsymbol{\Sigma}_{yy,i} = \mathbf{P}_i\boldsymbol{\Sigma}_{yy}\mathbf{P}_i^\top$, *and* $\boldsymbol{\Sigma}_{yx,i} = \mathbf{P}_i\boldsymbol{\Sigma}_{yx}$;

(ii) *the marginal CDF of* $\mathbf{LY}_{(N)}$ *is given by*

$$F_{\mathbf{LY}_{(N)}}(\mathbf{w}) = \sum_{i=1}^{N!} \pi_i F_{N-1,P}^{h^{(N+P-1)}}(\mathbf{w};\boldsymbol{\Theta}_i^1), \tag{21}$$

where $\boldsymbol{\Theta}_i^1 = \{\mathbf{L}\boldsymbol{\mu}_{y,i}, \mathbf{D}\boldsymbol{\mu}_{y,i}, \mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top, \mathbf{D}\boldsymbol{\Sigma}_{yy,i}\mathbf{D}^\top, \mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{D}^\top\}$;

(iii) *the conditional CDF of* $\mathbf{X}$ *given* $\mathbf{LY}_{(N)} = \mathbf{w}$ *is given by*

$$F_{\mathbf{X}|\mathbf{LY}_{(N)}}(\mathbf{x}|\mathbf{w}) = \sum_{i=1}^{N!} \pi_i F_{N-1,K}^{h_{q_{2i}(\mathbf{w})}^{(K+N-1)}}(\mathbf{x};\boldsymbol{\Theta}_i^{1\cdot 2}), \tag{22}$$

where $F_{N-1,K}^{h_{q_{2i}(\mathbf{w})}^{(K+N-1)}}(\cdot;\boldsymbol{\Theta}_i^{1\cdot 2})$ *is the CDF of* $\mathrm{SUE}_{N-1,K}(\boldsymbol{\Theta}_i^{1\cdot 2}, h_{q_{2i}(\mathbf{w})}^{(K+N-1)})$ *with conditional density generator function* $h_{q_{21}(\mathbf{w})}^{(K+N-1)}$, *and* $\boldsymbol{\Theta}_i^{1\cdot 2} = \{\boldsymbol{\xi}_i^{1\cdot 2}(\mathbf{w}), \boldsymbol{\eta}_i^{1\cdot 2}(\mathbf{w}), \boldsymbol{\Omega}_i^{11\cdot 2}, \boldsymbol{\Gamma}_i^{1\cdot 2}, \boldsymbol{\Lambda}_i^{1\cdot 2}\}$, *with* $\boldsymbol{\xi}_i^{1\cdot 2}(\mathbf{w}) = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{yx,i}^\top\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top)^{-1}(\mathbf{w}-\mathbf{L}\boldsymbol{\mu}_{y,i})$, $\boldsymbol{\eta}_i^{1\cdot 2}(\mathbf{w}) = \mathbf{D}\boldsymbol{\mu}_{y,i} + \mathbf{D}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top)^{-1}(\mathbf{w}-\mathbf{L}\boldsymbol{\mu}_{y,i})$, $\boldsymbol{\Omega}_i^{11\cdot 2} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{yx,i}^\top\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top)^{-1}\mathbf{L}\boldsymbol{\Sigma}_{yx,i}$, $\boldsymbol{\Gamma}_i^{1\cdot 2} = \boldsymbol{\Gamma}_i - \mathbf{D}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top)^{-1}\mathbf{L}\boldsymbol{\Sigma}_{yy,i}\ \mathbf{D}^\top$, $\boldsymbol{\Lambda}_i^{1\cdot 2} = \boldsymbol{\Sigma}_{yx,i}^\top\mathbf{D}^\top - \boldsymbol{\Sigma}_{yx,i}^\top\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top)^{-1}\mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{D}^\top$, *and* $q_{2i}(\mathbf{w}) = (\mathbf{w}-\mathbf{L}\boldsymbol{\mu}_{y,i})^\top(\mathbf{L}\boldsymbol{\Sigma}_{yy,i}\ \mathbf{L}^\top)^{-1}(\mathbf{w}-\mathbf{L}\boldsymbol{\mu}_{y,i})$;

(iv) *the conditional CDF of* $\mathbf{LY}_{(N)}$ *given* $\mathbf{X}=\mathbf{x}$ *is given by*

$$F_{\mathbf{LY}_{(N)}|\mathbf{X}}(\mathbf{w}|\mathbf{x}) = \sum_{i=1}^{N!} \pi_i F_{N-1,P}^{h_{q_1(\mathbf{x})}^{(N+P-1)}}(\mathbf{w};\boldsymbol{\Theta}_i^{2\cdot 1}), \qquad (23)$$

where $F_{N-1,P}^{h_{q_1(\mathbf{x})}^{(N+P-1)}}(\cdot;\boldsymbol{\Theta}_i^{2\cdot 1})$ *is the CDF of* $\mathrm{SUE}_{N-1,P}(\boldsymbol{\Theta}_i^{2\cdot 1}, h_{q_1(\mathbf{x})}^{(N+P-1)})$ *with conditional density generator function* $h_{q_1(\mathbf{x})}^{(N+P-1)}$, *and* $\boldsymbol{\Theta}_i^{2\cdot 1} = \{\boldsymbol{\xi}_i^{2\cdot 1}(\mathbf{x}), \boldsymbol{\eta}_i^{2\cdot 1}(\mathbf{x}), \boldsymbol{\Omega}_i^{22\cdot 1}, \boldsymbol{\Gamma}_i^{2\cdot 1}, \boldsymbol{\Lambda}_i^{2\cdot 1}\}$, *with* $\boldsymbol{\xi}_i^{2\cdot 1}(\mathbf{x}) = \boldsymbol{\mu}_{y,i} + \mathbf{L}\boldsymbol{\Sigma}_{yx,i}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)$, $\boldsymbol{\eta}_i^{2\cdot 1}(\mathbf{x}) = \mathbf{D}\boldsymbol{\mu}_{y,i} + \mathbf{D}\boldsymbol{\Sigma}_{yx,i}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)$, $\boldsymbol{\Omega}_i^{22\cdot 1} = \mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top - \mathbf{L}\boldsymbol{\Sigma}_{yx,i}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{yx,i}^\top\mathbf{L}^\top$, $\boldsymbol{\Gamma}_i^{2\cdot 1} = \boldsymbol{\Gamma}_i - \mathbf{D}\boldsymbol{\Sigma}_{yx,i}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{yx,i}^\top\mathbf{D}^\top$, $\boldsymbol{\Lambda}_i^{2\cdot 1} = \mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{D}^\top - \mathbf{L}\boldsymbol{\Sigma}_{yx,i}^\top\boldsymbol{\Sigma}_{xx}^{-1}\,\boldsymbol{\Sigma}_{yx,i}^\top\mathbf{D}^\top$, *and* $q_1(\mathbf{x}) = (\mathbf{x}-\boldsymbol{\mu}_x)^\top\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)$.

Proof: For (i), note first that

$$F_{\mathbf{X},\mathbf{LY}_{(N)}}(\mathbf{x},\mathbf{w}) = \sum_{i=1}^{N!} P(\mathbf{DY}_i \geq \mathbf{0})P(\mathbf{X}\leq\mathbf{x}, \mathbf{LY}_i \leq \mathbf{w}|\mathbf{DY}_i \geq \mathbf{0}), \qquad (24)$$

where the inequalities hold componentwise. Next, note that

$$\begin{pmatrix} \mathbf{DY}_i \\ \mathbf{X} \\ \mathbf{LY}_i \end{pmatrix} \sim \mathrm{EC}_{K+N+P-1}\left(\begin{pmatrix} \boldsymbol{\eta}_i \\ \boldsymbol{\xi}_i \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Gamma}_i & \boldsymbol{\Lambda}_i^\top \\ & \boldsymbol{\Omega}_i \end{pmatrix}, h^{(K+N+P-1)}\right),$$

for $i=1,\cdots,N!$. For the ith term of (24), we have $P(\mathbf{DY}_i\geq\mathbf{0}) = \pi_i$ and by (3), we get

$$P(\mathbf{X}\leq\mathbf{x}, \mathbf{LY}_i\leq\mathbf{w}|\mathbf{DY}_i\geq\mathbf{0}) = F_{N-1,K+P}^{h^{(K+N+P-1)}}(\mathbf{x},\mathbf{w};\boldsymbol{\Theta}_i),$$

which proves (i). Parts $(ii)-(iv)$ follow in a straightforward manner from the mixture representation (18) and results in Jamalizadeh and Balakrishnan (2011) on marginal and conditional distributions of SUE distributions. ■

Proposition 1 is a generalization of Jamalizadeh and Balakrishnan (2009b) and Balakrishnan et al. (2011), and gives the relevant joint, marginal, and conditional distributions as mixtures of SUE distributions. It can also be specialized to the case of normal and t-distributions; hence, it can be viewed as extensions of Viana (1998) and Olkin and Viana (1995). In either case, we need only replace $F_{N-1,K+P}^{h^{(K+N+P-1)}}(\cdot;\boldsymbol{\Theta}_i)$, the CDF of a SUE distribution with parameter $\boldsymbol{\Theta}_i$ and density generator function $h^{(K+N+P-1)}$ with the CDF $F_{N-1,K+P}^{\phi^{(K+N+P-1)}}(\cdot;\boldsymbol{\Theta}_i)$ of a SUN distribution (with density generator function $\phi^{(K+N+P-1)}$) in the normal case, or with the CDF $F_{N-1,K+P}^{t_\nu^{(K+N+P-1)}}(\cdot;\boldsymbol{\Theta}_i)$ of a SUT distribution (with density generator function $t_\nu^{(K+N+P-1)}$) in the case of t. It is also important to mention that the density generator function should be replaced by the characteristic generator function when $\boldsymbol{\Sigma}$ is singular.

3.1. Special case: results for X and $\mathbf{a}^\top\mathbf{Y}_{(N)}$

In this section, we consider the special case $K=P=1$ with $\mathbf{L}=\mathbf{a}^\top = (a_1,\cdots,a_N)$. Provided $\sum_{i=1}^N a_i \neq 0$, the joint distribution of X and $\mathbf{a}^\top\mathbf{Y}_{(N)}$ is given by (18) in Proposition 1. If

$\sum_{i=1}^{N} a_i = 0$, then the bivariate SUE distributions in (18) are singular, in which case the density generator function $h^{(N+1)}$ is replaced by the characteristic generator function $\varphi^{(N+1)}$. Corresponding marginal and conditional distributions likewise follow from Proposition 1. For example, consider the exchangeable case

$$\begin{pmatrix} X \\ \mathbf{Y} \end{pmatrix} \sim \mathrm{EC}_{N+1}\left(\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \mu \mathbf{1}_N \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \delta\tau\sigma\mathbf{1}_N^\top \\ & \tau^2\{(1-\rho)\mathbf{I}_N + \rho\mathbf{1}_N\mathbf{1}_N^\top\} \end{pmatrix}, h^{(N+1)}\right), \qquad (25)$$

where $\tau > 0$, $|\rho| < 1$, $\sqrt{N}|\delta| \le \sqrt{1+\rho(N-1)}$, $\mathbf{1}_N = (1, \cdots, 1)^\top$ is the $N \times 1$ summing vector, and $\mathbf{I}_N = diag(1, \cdots, 1)$ is the $N \times N$ identity matrix. This equicorrelation structure for $\mathbf{Y}$ is found most frequently in familial studies in genetics, for example, and in animal teratology, where data arise in clusters from litters. Then, it follows from Proposition 1 that $X|\mathbf{a}^\top\mathbf{Y}_{(N)} = w \sim \mathrm{SUE}_{N-1,1}(\boldsymbol{\Theta}^{1\cdot 2}, h_{q_2(w)}^{(N)})$, where $\boldsymbol{\Theta}^{1\cdot 2} = \{\xi^{1\cdot 2}(w), \boldsymbol{\eta}^{1\cdot 2}(w), \omega^{11\cdot 2}, \boldsymbol{\Gamma}^{1\cdot 2}, \boldsymbol{\lambda}^{1\cdot 2}\}$, with

$$\begin{aligned}
\xi^{1\cdot 2}(w) &= \mu_x + \frac{\delta\sigma\sum_{i=1}^{N} a_i}{\tau\left\{(1-\rho)\sum_{i=1}^{N} a_i^2 + \rho\left(\sum_{i=1}^{N} a_i\right)^2\right\}}\left(w - \mu\sum_{i=1}^{N} a_i\right), \\
\boldsymbol{\eta}^{1\cdot 2}(w) &= \frac{(1-\rho)\mathbf{D}\mathbf{a}}{(1-\rho)\sum_{i=1}^{N} a_i^2 + \rho\left(\sum_{i=1}^{N} a_i\right)^2}\left(w - \mu\sum_{i=1}^{N} a_i\right), \\
\omega^{11\cdot 2} &= \sigma^2 - \frac{\delta^2\sigma^2\left(\sum_{i=1}^{N} a_i\right)^2}{(1-\rho)\sum_{i=1}^{N} a_i^2 + \rho\left(\sum_{i=1}^{N} a_i\right)^2}, \\
\boldsymbol{\Gamma}^{1\cdot 2} &= \tau^2(1-\rho)\mathbf{D}\mathbf{D}^\top - \frac{\tau^2(1-\rho)^2\mathbf{D}\mathbf{a}\mathbf{a}^\top\mathbf{D}^\top}{(1-\rho)\sum_{i=1}^{N} a_i^2 + \rho\left(\sum_{i=1}^{N} a_i\right)^2}, \\
\boldsymbol{\lambda}^{1\cdot 2} &= -\frac{\delta\tau\sigma(1-\rho)\mathbf{D}\mathbf{a}\sum_{i=1}^{N} a_i}{(1-\rho)\sum_{i=1}^{N} a_i^2 + \rho\left(\sum_{i=1}^{N} a_i\right)^2}, \\
q_2(w) &= \frac{\left(w - \mu\sum_{i=1}^{N} a_i\right)^2}{\tau\left\{(1-\rho)\sum_{i=1}^{N} a_i^2 + \rho\left(\sum_{i=1}^{N} a_i\right)^2\right\}}.
\end{aligned}$$

These results may be specialized as well to the normal and t-cases as generalizations of previous results by Jamalizadeh and Balakrishnan (2009b), Balakrishnan et al. (2011), Viana (1998), and Olkin and Viana (1995), among others.

We now give the best (non-linear) predictors $E(X|\mathbf{a}^\top\mathbf{Y}_{(N)} = w)$ and $E(\mathbf{a}^\top\mathbf{Y}_{(N)}|X = x)$ of X and $\mathbf{a}^\top\mathbf{Y}_{(N)}$, respectively, in the normal and t-cases under squared error loss. Let

$$\mathbf{D} = \begin{pmatrix} \mathbf{d}_j^\top \\ \mathbf{D}_{-j} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_{j+1}^\top - \mathbf{e}_j^\top \\ \mathbf{D}_{-j} \end{pmatrix}, \qquad (26)$$

where $\mathbf{D}_{-j}$ is the matrix obtained from $\mathbf{D}$ by deleting row j, where $\mathbf{e}_1, \cdots, \mathbf{e}_N$ are N-dimensional unit basis vectors.

Proposition 2 *Suppose X and $\mathbf{Y}$ have a $(N+1)$-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then, under squared error loss, the best (non-linear) predictor of X based on $\mathbf{a}^\top \mathbf{Y}_{(N)} = w$ is*

$$
\begin{aligned}
E(X|\mathbf{a}^\top \mathbf{Y}_{(N)} = w) &= \sum_{i=1}^{N!} \pi_i \xi_i^{1\cdot 2}(w) + \sum_{i=1}^{N!} \frac{\pi_i}{\Phi_{N-1}(\boldsymbol{\eta}_i^{1\cdot 2}(w); \boldsymbol{\Gamma}_i^{1\cdot 2})} \sum_{j=1}^{N-1} \frac{\lambda_{i,j}^{1\cdot 2}}{\sqrt{\gamma_{i,jj}^{1\cdot 2}}} \\
&\quad \times \phi\left(\frac{\eta_{i,j}^{1\cdot 2}(w)}{\sqrt{\gamma_{i,jj}^{1\cdot 2}}}\right) \Phi_{N-2}\left(\boldsymbol{\eta}_{i,-j}^{1\cdot 2}(w) - \frac{\eta_{i,j}^{1\cdot 2}(w)}{\gamma_{i,jj}^{1\cdot 2}} \boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}; \boldsymbol{\Gamma}_{i,-j|j}^{1\cdot 2}\right); \qquad (27)
\end{aligned}
$$

if, on the other hand, X and $\mathbf{Y}$ have a $(N+1)$-dimensional t_ν-distribution with mean $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}$, then the best (non-linear) predictor of X based on $\mathbf{a}^\top \mathbf{Y}_{(N)} = w$ becomes

$$
\begin{aligned}
E(X|\mathbf{a}^\top \mathbf{Y}_{(N)} = w) &= \sum_{i=1}^{N!} \pi_i \xi_i^{1\cdot 2}(w) + \frac{(\nu+1)^{(\nu+1)/2}\Gamma\left(\frac{\nu}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu+1}{2}\right)} \sum_{i=1}^{N!} \frac{\pi_i q_{2i}^{1/2}(w,\nu)}{T_{N-1}\left(\boldsymbol{\eta}_i^{1\cdot 2}(w); \frac{q_{2i}(w,\nu)\boldsymbol{\Gamma}_i^{1\cdot 2}}{\nu+1}, \right)} \\
&\quad \times \sum_{j=1}^{N-1} \frac{\lambda_{i,j}^{1\cdot 2}}{\sqrt{\gamma_{i,jj}^{1\cdot 2}}} T_{N-2}\left(\begin{array}{c} \frac{\sqrt{\nu}}{\sqrt{\nu+1+\frac{\{\eta_{i,j}^{1\cdot 2}(w)\}^2}{q_{2i}(w,\nu)\gamma_{i,jj}^{1\cdot 2}}}} \left(\boldsymbol{\eta}_{i,-j}^{1\cdot 2}(w) - \frac{\eta_{i,j}^{1\cdot 2}(w)}{\gamma_{i,jj}^{1\cdot 2}} \boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}\right); \\ q_{2i}(w,\nu)\boldsymbol{\Gamma}_{i,-j|j}^{1\cdot 2}, \nu \end{array}\right) \\
&\quad \times \left(\nu + 1 + \frac{\{\eta_{i,j}^{1\cdot 2}(w)\}^2}{q_{2i}(w,\nu)\gamma_{i,jj}^{1\cdot 2}}\right)^{-\nu/2}. \qquad (28)
\end{aligned}
$$

All relevant quantities in (27)–(28) are as defined in Proposition 1 and in section 2.2; for example,

$$
\begin{aligned}
\boldsymbol{\Gamma}_{i,-j|j}^{1\cdot 2} &= \mathbf{D}_{-j}\boldsymbol{\Sigma}_{yy,i}\mathbf{D}_{-j}^\top - \frac{\mathbf{D}_{-j}\boldsymbol{\Sigma}_{yy,i}\mathbf{a}\mathbf{a}^\top\boldsymbol{\Sigma}_{yy,i}\mathbf{D}_{-j}^\top}{\mathbf{a}^\top\boldsymbol{\Sigma}_{yy,i}\mathbf{a}} - \frac{\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}\left(\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}\right)^\top}{\gamma_{i,jj}^{1\cdot 2}} \\
&= \boldsymbol{\Gamma}_{i,-j-j}^{1\cdot 2} - \frac{\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}\left(\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}\right)^\top}{\gamma_{i,jj}^{1\cdot 2}}, \qquad (29) \\
\boldsymbol{\eta}_i^{1\cdot 2}(\mathbf{w}) &= \begin{pmatrix} \eta_{i,j}^{1\cdot 2}(w) \\ \boldsymbol{\eta}_{i,-j}^{1\cdot 2}(w) \end{pmatrix} = \mathbf{D}\boldsymbol{\mu}_{y,i} + \mathbf{D}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top)^{-1}(\mathbf{w} - \mathbf{L}\boldsymbol{\mu}_{y,i}), \qquad (30)
\end{aligned}
$$

with $\mathbf{D}$ given by (26). The best (non-linear) predictor $E(\mathbf{a}^\top \mathbf{Y}_{(N)}|X = x)$ of $\mathbf{a}^\top \mathbf{Y}_{(N)}$ based on $X = x$ in both the normal and t-cases is obtained by replacing "w" with "x", changing the superscript "$1\cdot 2$" to "$2\cdot 1$", and replacing "$q_{2i}(w,\nu)$" in (28) with "$q_1(x,\nu)$".

Proof: Expressions (27) and (28) are straightforward from (13) and (16). ∎

Proposition 2 extends results in Loperfido (2008b), Jamalizadeh and Balakrishnan (2009a), and Balakrishnan et al. (2011), to the case $N \geq 2$.

3.2. Special case: results for X and $Y_{(r)}$

The joint CDF of X and $Y_{(r)}$ can be obtained from Proposition 1 or from section 3.1, as a mixture containing $N!$ terms, by taking $\mathbf{a} = \mathbf{e}_r$, $r = 1, \cdots, N$. In this section, we present an alternative approach for deriving this joint CDF with only $N\binom{N-1}{r-1} < N!$ terms. To do this, let $\mathbf{S}_{j_1\cdots j_m} = diag(s_1, \cdots, s_m)$ be such that

$$s_i = \begin{cases} 1 & \text{for } i = j_1, \cdots, j_m \\ -1 & \text{otherwise} \end{cases},$$

for integer $m \in [0, N-1]$ and ordered integers $1 \le j_1 < \cdots < j_m \le N$. In particular, $\mathbf{S}_{j_1\cdots j_{N-1}} = \mathbf{I}_{N-1}$ and $\mathbf{S}_{j_0} = -\mathbf{I}_{N-1}$. Further, let $\mathbf{X}_{j_1\cdots j_m} = (X_{j_1}, \cdots, X_{j_m})^\top$, and for $i = 1, \cdots, N$, let the vector $\mathbf{X}_{-i-j_1-\cdots-j_m}$ $(j_k \ne i, k = 1, \cdots, m)$ be obtained from $\mathbf{X}$ by deleting $X_i, X_{j_1}, \cdots, X_{j_m}$. Consider also the partitions

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_{-i} \\ Y_i \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{y,-i} \\ \mu_{y,i} \end{pmatrix}, \ \boldsymbol{\Sigma}_{yy} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy,-i-i} & \boldsymbol{\sigma}_{yy,-ii} \\ & \sigma_{yy,ii} \end{pmatrix}, \tag{31}$$

so that

$$\begin{pmatrix} X \\ \mathbf{Y}_{-i} \\ Y_i \end{pmatrix} \sim \mathrm{EC}_{m+N}\left(\begin{pmatrix} \mu_x \\ \boldsymbol{\mu}_{y,-i} \\ \mu_{y,i} \end{pmatrix}, \begin{pmatrix} \sigma_{xx} & \boldsymbol{\sigma}_{yx,-i}^\top & \sigma_{yx,i} \\ & \boldsymbol{\Sigma}_{yy,-i-i} & \boldsymbol{\sigma}_{yy,-ii} \\ & & \sigma_{yy,i} \end{pmatrix}, h^{(N+1)}\right). \tag{32}$$

The following proposition gives the joint CDF of X and $Y_{(r)}$ as a mixture of SUE distributions with only $N\binom{N-1}{r-1} < N!$ terms. The result is technically elegant and computationally simpler than what we get from Proposition 1.

Proposition 3 *For $r = 1, \cdots, N$, and $j_k \ne i$, the joint CDF of X and $Y_{(r)}$ is given by*

$$F_{X,Y_{(r)}}(x, y) = \sum_{i=1}^{N} \sum_{\substack{j_1<\cdots<j_{r-1} \\ 1\le j_k\le N, j_k\ne i}} \pi_{i,j_1\cdots j_{r-1}} F_{N-1,2}^{h^{(N+1)}}(x, y; \boldsymbol{\Theta}_{i,j_1\cdots j_{r-1}}), \tag{33}$$

where $\boldsymbol{\Theta}_{i,j_1\cdots j_{r-1}} = \{\boldsymbol{\xi}_{i,j_1\cdots j_{r-1}}, \boldsymbol{\eta}_{i,j_1\cdots j_{r-1}}, \boldsymbol{\Omega}_{i,j_1\cdots j_{r-1}}, \boldsymbol{\Gamma}_{i,j_1\cdots j_{r-1}}, \boldsymbol{\Lambda}_{i,j_1\cdots j_{r-1}}\}$, $F_{N-1,2}^{h^{(N+1)}}(\cdot; \boldsymbol{\Theta}_{i,j_1\cdots j_{r-1}})$ is the CDF of $\mathrm{SUE}_{N-1,2}(\boldsymbol{\Theta}_{i,j_1\cdots j_{r-1}}, h^{(N+1)})$, and

$$\pi_{i,j_1\cdots j_{r-1}} = G_{N-1}^{h^{(N-1)}}(\mathbf{S}_{j_1\cdots j_{r-1}}(\mu_{y,i}\mathbf{1}_{N-1} - \boldsymbol{\mu}_{y,-i}); \boldsymbol{\Gamma}_{i,j_1\cdots j_{r-1}}), \tag{34}$$

with $G_{N-1}^{h^{(N-1)}}(\cdot; \boldsymbol{\Gamma}_{i,j_1\cdots j_{r-1}})$ the CDF of $\mathrm{EC}_{N-1}(\mathbf{0}, \boldsymbol{\Gamma}_{i,j_1\cdots j_{r-1}}, h^{(N+1)})$, $\boldsymbol{\xi}_{i,j_1\cdots j_{r-1}} = (\mu_x, \mu_{y,i})^\top$, $\boldsymbol{\eta}_{i,j_1\cdots j_{r-1}} = \mathbf{S}_{j_1\cdots j_{r-1}}(\mu_{y,i}\mathbf{1}_{N-1} - \boldsymbol{\mu}_{y,-i})$,

$$\begin{aligned}
\boldsymbol{\Omega}_{i,j_1\cdots j_{r-1}} &= \begin{pmatrix} \sigma_{xx} & \sigma_{yx,i} \\ & \sigma_{yy,ii} \end{pmatrix}, \\
\boldsymbol{\Gamma}_{i,j_1\cdots j_{r-1}} &= \mathbf{S}_{j_1\cdots j_{r-1}}(\sigma_{yy,ii}\mathbf{1}_{N-1}\mathbf{1}_{N-1}^\top + \boldsymbol{\Sigma}_{yy,-i-i} - \mathbf{1}_{N-1}\boldsymbol{\sigma}_{yy,-ii}^\top - \boldsymbol{\sigma}_{yy,-ii}\mathbf{1}_{N-1}^\top)\mathbf{S}_{j_1\cdots j_{r-1}}, \\
\boldsymbol{\Lambda}_{i,j_1\cdots j_{r-1}} &= \begin{pmatrix} (\sigma_{yx,i}\mathbf{1}_{N-1}^\top - \boldsymbol{\sigma}_{yx,-i})^\top\mathbf{S}_{j_1\cdots j_{r-1}} \\ (\sigma_{yy,ii}\mathbf{1}_{N-1}^\top - \boldsymbol{\sigma}_{yy,-ii})^\top\mathbf{S}_{j_1\cdots j_{r-1}} \end{pmatrix} = \begin{pmatrix} (\sigma_{yx,i}\mathbf{1}_{N-1}^\top - \boldsymbol{\sigma}_{yx,-i})^\top\mathbf{S}_{j_1\cdots j_{r-1}} \\ \boldsymbol{\lambda}_{i,j_1\cdots j_{r-1}}^\top \end{pmatrix}.
\end{aligned}$$

The marginal CDF *of* $Y_{(r)}$ *is readily obtained from (33) as*

$$F_{Y_{(r)}}(y) = \sum_{i=1}^{N} \sum_{\substack{j_1<\cdots<j_{r-1} \\ 1\le j_k\le N, j_k\neq i}} \pi_{i,j_1\cdots j_{r-1}} F_{N-1,1}^{h^{(N+1)}}(y;\boldsymbol{\Theta}_{i,j_1\cdots j_{r-1}}^{(r)}), \tag{35}$$

where $\boldsymbol{\Theta}_{i,j_1\cdots j_{r-1}}^{(r)} = \{\mu_{y,i}, \boldsymbol{\eta}_{i,j_1\cdots,j_{r-1}}, \sigma_{yy,ii}, \boldsymbol{\Gamma}_{i,j_1,\cdots,j_{r-1}}, \boldsymbol{\lambda}_{i,j_1,\cdots,j_{r-1}}\}$.

Proof: We have

$$F_{X,Y_{(r)}}(x,y) = \sum_{i=1}^{N} P(X\le x, Y_i\le y, Y_i = Y_{(r)}). \tag{36}$$

The ith term of the RHS of (34) is

$$\begin{aligned} P(X\le x, Y_i\le y, Y_i=Y_{(r)}) &= \sum_{\substack{j_1<\cdots<j_{r-1} \\ 1\le j_k\le N, j_k\neq i}} P\begin{pmatrix} X<x, Y_i\le y, \\ max(\mathbf{Y}_{j_1\cdots j_{r-1}}) < Y_i < min(\mathbf{Y}_{-i-j_1-\cdots-j_{r-1}}) \end{pmatrix} \\ &= \sum_{\substack{j_1<\cdots<j_{r-1} \\ 1\le j_k\le N, j_k\neq i}} P(X<x, Y_i\le y|\mathbf{S}_{j_1\cdots j_{r-1}}\{\mathbf{1}_{N-1}Y_i - \mathbf{Y}_{-i}\} > \mathbf{0}) \\ &\quad \times P(\mathbf{S}_{j_1\cdots j_{r-1}}\{\mathbf{1}_{N-1}Y_i - \mathbf{Y}_{-i}\} > \mathbf{0}). \end{aligned}$$

Now, we have for $i = 1,\cdots,N$,

$$\begin{pmatrix} \mathbf{S}_{j_1\cdots j_{r-1}}\{\mathbf{1}_{N-1}Y_i - \mathbf{Y}_{-i}\} \\ (X,Y_i)^\top \end{pmatrix} \sim \mathrm{EC}_{N+1}\left(\begin{pmatrix} \boldsymbol{\eta}_{i,j_1\cdots j_{r-1}} \\ \boldsymbol{\xi}_{i,j_1\cdots j_{r-1}} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Gamma}_{i,j_1\cdots j_{r-1}} & \boldsymbol{\Lambda}_{i,j_1\cdots j_{r-1}}^\top \\ & \boldsymbol{\Omega}_{i,j_1\cdots j_{r-1}} \end{pmatrix}, h^{(N+1)}\right),$$

so that

$$P(X<x, Y_i\le y|\mathbf{S}_{j_1\cdots j_{r-1}}\{\mathbf{1}_{N-1}Y_i - \mathbf{Y}_{-i}\} > \mathbf{0}) = F_{N-1,2}^{h^{(N+1)}}(x,y;\boldsymbol{\Theta}_{i,j_1\cdots j_{r-1}}),$$

and since $P(\mathbf{S}_{j_1\cdots j_{r-1}}\{\mathbf{1}_{N-1}Y_i - \mathbf{Y}_{-i}\} > \mathbf{0}) = \pi_{i,j_1\cdots j_{r-1}}$, the proof is complete. ■

Proposition 3 extends Loperfido's (2008b) result to the general contralateral data set-up and for an arbitrary multivariate skew-elliptical distribution with an arbitrary correlation structure. Best nonlinear predictors of X and of $Y_{(r)}$ based on X and on $Y_{(r)}$, respectively, are also conveniently obtained from Proposition 3. To do this for the case $r = N$, consider the following partitions for $j \neq i$:

$$\begin{aligned} \boldsymbol{\mu}_{y,-i} &= \begin{pmatrix} \mu_{y,j} \\ \boldsymbol{\mu}_{y,-i-j} \end{pmatrix}, \ \boldsymbol{\Sigma}_{yy,-i-i} = \begin{pmatrix} \sigma_{yy,jj} & \boldsymbol{\sigma}_{yy;-i-j,j}^\top \\ & \boldsymbol{\Sigma}_{yy;-i-j,-i-j} \end{pmatrix}, \\ \boldsymbol{\sigma}_{yy,-ii} &= \begin{pmatrix} \sigma_{yy,ji} \\ \boldsymbol{\sigma}_{yy;-i-j,i} \end{pmatrix}, \ \boldsymbol{\sigma}_{yx,-i} = \begin{pmatrix} \sigma_{yx,j} \\ \boldsymbol{\sigma}_{yx,-i-j} \end{pmatrix}. \end{aligned}$$

Assuming joint normality for X and $\mathbf{Y}$, we get

$$
\begin{aligned}
E(X|Y_{(N)}=y) &= \sum_{i=1}^{N}\pi_i\left\{\xi_i^{1\cdot 2}(y)+\frac{1}{\Phi_{N-1}\left(\boldsymbol{\eta}_i^{1\cdot 2}(y);\boldsymbol{\Gamma}_i^{1\cdot 2}\right)}\sum_{j\neq i}\frac{\lambda_{i,j}^{1\cdot 2}\phi\left(\frac{\eta_{i,j}^{1\cdot 2}(y)}{\sqrt{\gamma_{i,jj}^{1\cdot 2}}}\right)}{\sqrt{\gamma_{i,jj}^{1\cdot 2}}}\right.\\
&\qquad \left.\Phi_{N-2}\left(\boldsymbol{\eta}_{i,-j}^{1\cdot 2}(y)-\frac{\eta_{i,j}^{1\cdot 2}(y)}{\gamma_{i,jj}^{1\cdot 2}}\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2};\boldsymbol{\Gamma}_{i,-j|j}^{1\cdot 2}\right)\right\}, \qquad (37)
\end{aligned}
$$

where $\pi_i = G_{N-1}^{h^{(N-1)}}(\boldsymbol{\eta}_i;\boldsymbol{\Gamma}_i)$, $\xi_i^{1\cdot 2}(y)=\mu_x+\sigma_{yx,i}(y-\mu_{y,i})/\sigma_{yy,ii}$, $\lambda_{i,j}^{1\cdot 2}=\sigma_{yx,i}-\sigma_{yx,j}-\sigma_{yx,i}(\sigma_{yy,ii}-\sigma_{yy,ji})/\sigma_{yy,ii}$, $\gamma_{i,jj}^{1\cdot 2}=\sigma_{yy,ii}+\sigma_{yy,jj}-2\sigma_{yy,ji}-(\sigma_{yy,ii}-\sigma_{yy,ji})^2/\sigma_{yy,ii}$, $\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}=\sigma_{yy,ii}\mathbf{1}_{N-2}^\top+\boldsymbol{\sigma}_{yy;-i-j,j}^\top-\boldsymbol{\sigma}_{yy;-i-j,i}^\top-\sigma_{yy,ji}\mathbf{1}_{N-2}^\top-(\sigma_{yy,ii}-\sigma_{yy,ji})(\sigma_{yy,ii}\mathbf{1}_{N-2}-\boldsymbol{\sigma}_{yy;-i-j,i})^\top/\sigma_{yy,ii}$,

$$
\begin{aligned}
\boldsymbol{\eta}_i^{1\cdot 2}(y) &= \begin{pmatrix}\eta_{i,j}^{1\cdot 2}(y)\\ \boldsymbol{\eta}_{i,-j}^{1\cdot 2}(y)\end{pmatrix} = \mu_{y,i}\mathbf{1}_{N-1}-\boldsymbol{\mu}_{y,-i}+\frac{\sigma_{yy,ii}\mathbf{1}_{N-1}-\boldsymbol{\sigma}_{yy,-ii}}{\sigma_{yy,ii}}(y-\mu_{y,i}),\\
\boldsymbol{\Gamma}_i^{1\cdot 2} &= \boldsymbol{\Gamma}_i-\frac{(\sigma_{yy,ii}\mathbf{1}_{N-1}-\boldsymbol{\sigma}_{yy,-ii})(\sigma_{yy,ii}\mathbf{1}_{N-1}-\boldsymbol{\sigma}_{yy,-ii})^\top}{\sigma_{yy,ii}},\\
\boldsymbol{\Gamma}_{i,-j-j}^{1\cdot 2} &= \sigma_{yy,ii}\mathbf{1}_{N-2}\mathbf{1}_{N-2}^\top+\boldsymbol{\Sigma}_{yy;-i-j,-i-j}-\mathbf{1}_{N-2}\boldsymbol{\sigma}_{yy;-i-j,i}-\boldsymbol{\sigma}_{yy;-i-j,i}\mathbf{1}_{N-2}^\top\\
&\quad -\frac{(\sigma_{yy,ii}\mathbf{1}_{N-2}-\boldsymbol{\sigma}_{yy;-i-j,i})(\sigma_{yy,ii}\mathbf{1}_{N-2}-\boldsymbol{\sigma}_{yy;-i-j,i})^\top}{\sigma_{yy,ii}},
\end{aligned}
$$

with $\boldsymbol{\Gamma}_{i,-j|j}^{1\cdot 2}$ given in (29). Similarly, the predictor $E(Y_{(N)}|X=x)$ can be analogously obtained from (37) by replacing "y" with "x" and changing the superscript "$1\cdot 2$" to "$2\cdot 1$", where $\xi_i^{2\cdot 1}(x)=\mu_{y,i}+\sigma_{yx,i}(x-\mu_x)/\sigma_{xx}$, $\lambda_{i,j}^{2\cdot 1}=\sigma_{yy,ii}-\sigma_{yy,ji}-\sigma_{yx,i}(\sigma_{yx,i}-\sigma_{yx,j})/\sigma_{xx}$, $\gamma_{i,jj}^{2\cdot 1}=\sigma_{yy,ii}+\sigma_{yy,jj}-2\sigma_{yy,ji}-(\sigma_{yx,i}-\sigma_{yx,j})^2/\sigma_{xx}$, $\boldsymbol{\gamma}_{i,-jj}^{2\cdot 1}=\sigma_{yy,ii}\mathbf{1}_{N-2}^\top+\boldsymbol{\sigma}_{yy;-i-j,j}^\top-\boldsymbol{\sigma}_{yy;-i-j,i}^\top-\sigma_{yy,ji}\mathbf{1}_{N-2}^\top-(\sigma_{yx,i}-\sigma_{yx,j})(\sigma_{yx,i}\mathbf{1}_{N-2}-\boldsymbol{\sigma}_{yx,-i-j})^\top/\sigma_{xx}$,

$$
\begin{aligned}
\boldsymbol{\eta}_i^{2\cdot 1}(x) &= \begin{pmatrix}\eta_{i,j}^{2\cdot 1}(x)\\ \boldsymbol{\eta}_{i,-j}^{2\cdot 1}(x)\end{pmatrix} = \mu_{y,i}\mathbf{1}_{N-1}-\boldsymbol{\mu}_{y,-i}+\frac{\sigma_{yx,i}\mathbf{1}_{N-1}-\boldsymbol{\sigma}_{yx,-i}}{\sigma_{xx}}(x-\mu_x),\\
\boldsymbol{\Gamma}_i^{2\cdot 1} &= \boldsymbol{\Gamma}_i-\frac{(\sigma_{yx,i}\mathbf{1}_{N-1}-\boldsymbol{\sigma}_{yx,-i})(\sigma_{yx,i}\mathbf{1}_{N-1}-\boldsymbol{\sigma}_{yx,-i})^\top}{\sigma_{xx}},\\
\boldsymbol{\Gamma}_{i,-j-j}^{2\cdot 1} &= \sigma_{yy,ii}\mathbf{1}_{N-2}\mathbf{1}_{N-2}^\top+\boldsymbol{\Sigma}_{yy;-i-j,-i-j}-\mathbf{1}_{N-2}\boldsymbol{\sigma}_{yy;-i-j,i}-\boldsymbol{\sigma}_{yy;-i-j,i}\mathbf{1}_{N-2}^\top\\
&\quad -\frac{(\sigma_{yx,i}\mathbf{1}_{N-2}-\boldsymbol{\sigma}_{yx,-i-j})(\sigma_{yx,i}\mathbf{1}_{N-2}-\boldsymbol{\sigma}_{yx,-i-j})^\top}{\sigma_{xx}}.
\end{aligned}
$$

Note that these predictors are computationally simpler than the corresponding ones given in Proposition 2.

In the exchangeable case, it is easy to see that X and $Y_{(r)}$ are jointly $\mathrm{SUE}_{N-1,2}(\boldsymbol{\Theta},h^{(N+1)})$, where

$$
\boldsymbol{\Theta} = \left\{\begin{pmatrix}\mu_x\\ \mu\end{pmatrix},\mathbf{0},\begin{pmatrix}\sigma^2 & \delta\tau\sigma\\ & \tau^2\end{pmatrix},\tau^2(1-\rho)(\mathbf{I}_{N-1}+\rho\mathbf{1}_{N-1}\mathbf{1}_{N-1}^\top),\begin{pmatrix}\mathbf{0}^\top\\ \tau^2(1-\rho)\mathbf{1}_{N-1}^\top\end{pmatrix}\right\}.
$$

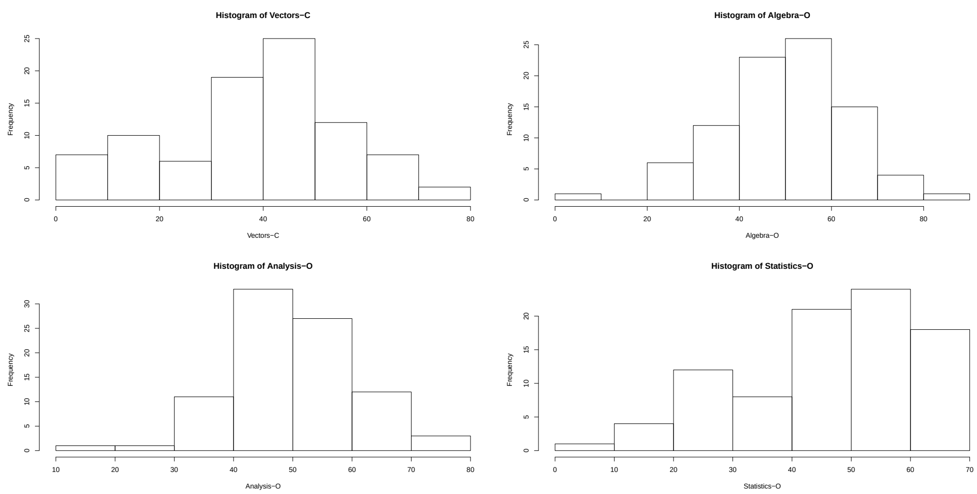

Figure 1: Histograms of examination scores in (a) "Vectors", (b) "Algebra", (c) "Analysis", and (d) "Statistics".

That is, X and $Y_{(r)}$ have an exact bivariate SUE joint distribution under exchangeability. In this case with $r = N$, the best predictors above reduce to

$$\begin{aligned} E(X|Y_{(N)} = y) &= \mu_x + \delta\sigma \left(\left(\frac{y-\mu}{\tau} \right) - \frac{(N-1)\sqrt{\frac{1-\rho}{1+\rho}}\,\phi\left(\sqrt{\frac{1-\rho}{1+\rho}}\tau(y-\mu) \right)}{\Phi_{N-1}\left(\begin{array}{c} (1-\rho)\tau^2(y-\mu)\mathbf{1}_{N-2}; \\ (1-\rho)\tau^2\left\{ \mathbf{I}_{N-2} + \rho\mathbf{1}_{N-2}\mathbf{1}_{N-2}^\top \right\} \end{array} \right)} \right) \\ &\quad \times\Phi_{N-2}\left(\begin{array}{c} \frac{(1-\rho)^2\tau^2(y-\mu)}{1+\rho}\mathbf{1}_{N-2}; \\ \rho(1-\rho)\tau^2\left\{ \frac{1}{\rho}\mathbf{I}_{N-2} + \frac{1+2\rho}{1+\rho}\mathbf{1}_{N-2}\mathbf{1}_{N-2}^\top \right\} \end{array} \right), \qquad (38) \\ E(Y_{(N)}|X = x) &= \mu + \delta\tau\left(\frac{x-\mu_x}{\sigma} \right) + \frac{(N-1)\tau\sqrt{1-\rho}}{2\sqrt{\pi}}. \qquad (39) \end{aligned}$$

Note that (38) and (39) reproduce earlier results for $N = 2$ given, for example, by Viana (1998) and Olkin and Viana (1995). Proposition 3 may also be used to obtain the best predictors of X and of $Y_{(N)}$ based on $Y_{(N)}$ and on X, respectively, in the case of a joint t-distribution for X and $\mathbf{Y}$.

4. Illustration: test scores data

We now use the examination scores (in %) data in Mardia et al. (1979, p. 3) to illustrate the results in previous sections. The data were collected from 87 students who wrote examinations on 5 subjects, namely, "Statistics", "Vectors", "Algebra", "Analysis", and "Mechanics". We consider only the first four subjects in what follows, and treat "Statistics" score as the variable of primary interest (X) and scores in "Vectors" (Y_1), "Algebra" (Y_2), and "Analysis" (Y_3), as auxiliary variables; thus, we have $N = 3$ and $K = 1$. Our interest lies in predicting X using an L-statistic $\mathbf{a}^\top\mathbf{Y}_{(3)}$, where $\mathbf{a} = (a_1, a_2, a_3)^\top$ is a real vector and $\mathbf{Y}_{(3)} = (Y_{(1)}, Y_{(2)}, Y_{(3)})^\top$ is the vector of order statistics.

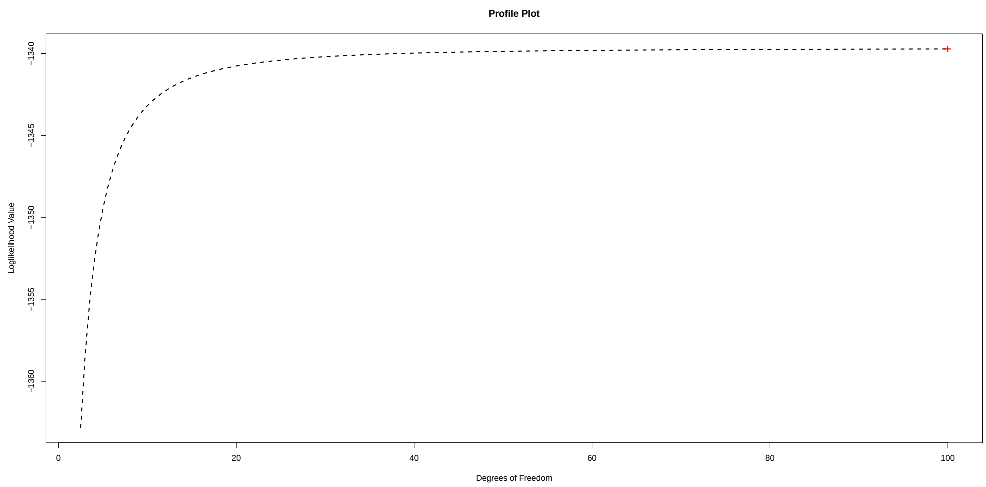

Figure 2: Profile plot of degrees of freedom ν over $\mathcal{B} = [2.5, 100]$.

Univariate histograms of the examination scores in the four subjects are shown in Figure 1. The plots appear to suggest that normality is a reasonable assumption, albeit slight skewness is noticeable in those for "Vectors" and "Statistics". In addition, potential outliers may be present, as evidenced by the histograms' long tails. Taking all this into account, we model the joint distribution of X and $\mathbf{Y} = (Y_1, Y_2, Y_3)^\top$ as a multivariate t-distribution with mean $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$, and degrees of freedom ν, with PDF given by

$$f_{X,\mathbf{Y}}(x, \mathbf{y}) \quad = \quad \frac{\Gamma\left(\frac{\nu+4}{2}\right)}{(\pi\nu)^2\Gamma\left(\frac{\nu}{2}\right)}|\boldsymbol{\Sigma}|^{-1/2}\left\{1 + \frac{1}{\nu}\left(\begin{pmatrix} x \\ \mathbf{y}\end{pmatrix} - \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\begin{pmatrix} x \\ \mathbf{y}\end{pmatrix} - \boldsymbol{\mu}\right)\right\}^{-(\nu+4)/2}.$$

Note that the above multivariate t-distribution belongs to the family of multivariate elliptically-contoured distributions. It provides a robust alternative to multivariate normality while approximating the latter for large degrees of freedom.

To estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we adopt Liu and Rubin's (1995) EM algorithm and use profile likelihood estimation for ν. Specifically, we fix ν in some gridded interval $\mathcal{B} \subset (2, \infty)$, and use the EM algorithm to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, assuming ν is known. The MLEs $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\Sigma}}$ are those estimates corresponding to $\widetilde{\nu}$, obtained as the point where the maximum of the profile plot occurs. A profile plot of ν is shown in Figure 2; it is clear that the plot is increasing, an indication that normality is a tenable assumption (since the multivariate t-distribution approaches the multivariate normal distribution for large ν). We take $\widetilde{\nu} = 100$, the largest possible value of ν in $\mathcal{B} = [2.5, 100]$; in fact, we can take $\widetilde{\nu}$ to be a value larger than 100, but the effect on the estimates will be negligible. At $\widetilde{\nu} = 100$, the MLEs $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\Sigma}}$ are

$$\widetilde{\boldsymbol{\mu}} = \left(\begin{array}{c} 42.26 \\ \hline 50.72 \\ 50.67 \\ 46.78 \end{array}\right) \quad \text{and} \quad \widetilde{\boldsymbol{\Sigma}} = \left(\begin{array}{c|ccc} 299.14 & 99.33 & 121.78 & 156.06 \\ \hline & 172.16 & 84.97 & 95.16 \\ & & 112.34 & 111.89 \\ & & & 221.1 \end{array}\right). \tag{40}$$

Not surprisingly, these estimates are comparatively numerically quite close to the following corresponding usual MLEs $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ assuming a multivariate normal joint distribution for X and $\mathbf{Y}$:

$$\widehat{\boldsymbol{\mu}} = \begin{pmatrix} 42.31 \\ \hline 50.59 \\ 50.6 \\ 46.68 \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}} = \left(\begin{array}{c|ccc} 297.76 & 99.01 & 121.87 & 155.54 \\ \hline & 172.84 & 85.16 & 94.67 \\ & & 112.89 & 112.11 \\ & & & 220.38 \end{array}\right). \tag{41}$$

We thus proceed assuming joint normality of X and $\mathbf{Y}$ in the sequel. Proposition 3 gives the joint CDF of X and $Y_{(3)} = max(Y_1, Y_2, Y_3)$ as

$$F_{X,Y_{(3)}}(x,y) = \sum_{i=1}^{3} \pi_i F_{2,2}^{\phi(4)}(x,y;\boldsymbol{\Theta}_i), \tag{42}$$

where $\pi_i = \Phi_2(\boldsymbol{\eta}_i;\boldsymbol{\Gamma}_i)$ and $\boldsymbol{\Theta}_i = \{\boldsymbol{\xi}_i, \boldsymbol{\eta}_i, \boldsymbol{\Omega}_i, \boldsymbol{\Gamma}_i, \boldsymbol{\Lambda}_i\}$, with $\boldsymbol{\xi}_i = (\mu_x, \mu_y, i)^\top$, $\boldsymbol{\eta}_i = \mu_{y,i}\mathbf{1}_2 - \boldsymbol{\mu}_{y,-i}$, $\boldsymbol{\Gamma}_i = \sigma_{yy,ii}\mathbf{1}_2\mathbf{1}_2^\top + \boldsymbol{\Sigma}_{yy,-i-i} - \mathbf{1}_2\boldsymbol{\sigma}_{yy,-ii}^\top - \boldsymbol{\sigma}_{yy,-ii}\mathbf{1}_2^\top$,

$$\boldsymbol{\Omega}_i = \begin{pmatrix} \sigma_{xx} & \sigma_{yx,i} \\ & \sigma_{yy,ii} \end{pmatrix}, \quad \boldsymbol{\Lambda}_i = \begin{pmatrix} \sigma_{yx,i}\mathbf{1}_2^\top - \boldsymbol{\sigma}_{yx,-i}^\top \\ \sigma_{yy,ii}\mathbf{1}_2^\top - \boldsymbol{\sigma}_{yy,-ii}^\top \end{pmatrix}.$$

Using MLEs $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ in (41), we get the plug-in MLE $\widehat{F}_{X,Y_{(3)}}(\cdot)$ as

$$\widehat{F}_{X,Y_{(3)}}(x,y) = \sum_{i=1}^{3} \widehat{\pi}_i \widehat{F}_{2,2}^{\phi(4)}(x,y;\widehat{\boldsymbol{\Theta}}_i), \tag{43}$$

where $\widehat{\pi}_1 = 0.43$, $\widehat{\pi}_2 = 0.32$, $\widehat{\pi}_3 = 0.25$, and

$$\widehat{\boldsymbol{\Theta}}_1 = \left\{ \begin{pmatrix} 42.31 \\ 50.59 \end{pmatrix}, \begin{pmatrix} -0.01 \\ 3.91 \end{pmatrix}, \begin{pmatrix} 297.76 & 99.01 \\ & 172.84 \end{pmatrix}, \begin{pmatrix} 115.41 & 105.12 \\ & 203.88 \end{pmatrix}, \begin{pmatrix} -22.86 & 56.53 \\ 87.68 & 78.17 \end{pmatrix} \right\},$$

$$\widehat{\boldsymbol{\Theta}}_2 = \left\{ \begin{pmatrix} 42.31 \\ 50.6 \end{pmatrix}, \begin{pmatrix} 0.01 \\ 3.92 \end{pmatrix}, \begin{pmatrix} 297.76 & 121.87 \\ & 172.84 \end{pmatrix}, \begin{pmatrix} 115.41 & 10.29 \\ & 109.05 \end{pmatrix}, \begin{pmatrix} 22.86 & -33.67 \\ 27.73 & 0.78 \end{pmatrix} \right\},$$

$$\widehat{\boldsymbol{\Theta}}_3 = \left\{ \begin{pmatrix} 42.31 \\ 46.68 \end{pmatrix}, \begin{pmatrix} -3.91 \\ -3.92 \end{pmatrix}, \begin{pmatrix} 297.76 & 155.54 \\ & 220.38 \end{pmatrix}, \begin{pmatrix} 203.88 & 98.76 \\ & 109.05 \end{pmatrix}, \begin{pmatrix} 56.53 & 33.67 \\ 127.71 & 108.27 \end{pmatrix} \right\}.$$

Plug-in MLEs of the best predictors of X and of $Y_{(3)}$ in section 3 based on $Y_{(3)}$ and on X, respectively, are as follows:

$$\begin{aligned} \widehat{E(X|Y_{(3)} = y)} = & \frac{1}{\Phi_2\left(\begin{matrix} 0.51y - 25.68, \\ 0.45y - 18.98 \end{matrix}; \begin{pmatrix} 70.93 & 65.47 \\ & 168.53 \end{pmatrix}\right)} \{1.91\phi(0.06y - 3.05) \\ & \times \Phi\left(\frac{-0.02y + 4.72}{10.4}\right) + 1.11\phi(0.03y - 1.47)\Phi\left(\frac{0.33y - 18.31}{6.74}\right)\Bigg\} \\ & + \frac{1}{\Phi_2\left(\begin{matrix} 0.08y - 3.24, \\ -0.11y - 8.7 \end{matrix}; \begin{pmatrix} 113.65 & 12.88 \\ & 105.24 \end{pmatrix}\right)} \{0.55\phi(0.01y - 0.3) \end{aligned}$$

$$
\begin{aligned}
& \Phi\left(\frac{-0.12y+9.07}{10.19}\right)+0.45\phi(-0.01y+0.85)\Phi\left(\frac{0.09y-4.3}{10.59}\right)\Bigg\} \\
& +\frac{1}{\Phi_2\left(\begin{array}{c}0.19y-11.94,\\ 0.11y-8.7\end{array};\left(\begin{array}{cc}193.15 & 92.37\\ & 105.24\end{array}\right)\right)}\{1.73\phi(0.01y-0.86) \\
& \times\Phi\left(\frac{0.02y-2.3}{7.81}\right)+2.21\phi(0.01y-0.8)\Phi\left(\frac{0.09y-4.3}{10.59}\right)\Bigg\} \\
& +16.5y+22.52, \qquad (44)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{E(Y_{(3)}|X=x)} = & -\frac{1}{\Phi_2\left(\begin{array}{c}0.51x-25.68,\\ 0.45x-18.98\end{array};\left(\begin{array}{cc}70.93 & 65.47\\ & 168.53\end{array}\right)\right)}\{3.73\phi(0.06x-3.05) \\
& \times\Phi\left(\frac{-0.02x+4.72}{10.4}\right)-3.35\phi(0.03x-1.47)\Phi\left(\frac{0.33x-18.31}{6.74}\right)\Bigg\} \\
& +\frac{1}{\Phi_2\left(\begin{array}{c}0.08x-3.24,\\ -0.11x-8.7\end{array};\left(\begin{array}{cc}113.65 & 12.88\\ & 105.24\end{array}\right)\right)}\{3.38\phi(0.01x-0.3) \\
& \Phi\left(\frac{-0.12x+9.07}{12.39}\right)+5.82\phi(-0.01x+0.85)\Phi\left(\frac{0.09x-4.3}{10.59}\right)\Bigg\} \\
& +\frac{1}{\Phi_2\left(\begin{array}{c}0.19x-11.94,\\ 0.11x-8.7\end{array};\left(\begin{array}{cc}193.15 & 92.37\\ & 105.24\end{array}\right)\right)}\{1.73\phi(0.01x-0.86) \\
& \times\Phi\left(\frac{0.02x-2.1}{10.23}\right)+2.21\phi(0.01x-0.85)\Phi\left(\frac{0.09x-4.3}{10.59}\right)\Bigg\} \\
& +0.5x+22.52, \qquad (45)
\end{aligned}
$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Note the computational simplicity of (44) and (45); this is due to the simpler form for the joint CDF of X and $Y_{(3)}$ outlined in Proposition 3 containing only $N\binom{N-1}{r-1} = 3\binom{2}{2} = 3$ terms, as opposed to that obtained from Proposition 1 which contains $N! = 3! = 6$ terms.

5. Conclusion

In this paper, we derived general results on distributions of L-statistics and on prediction involving multivariate skew-elliptical distributions, after assuming multivariate elliptical distributions for the data. The results are general enough to include several previous results as special cases (e.g., Jamalizadeh and Balakrishnan, 2009b; Balakrishnan et al., 2011). By considering elliptical distributions, which include normal as well as t-distributions, we provide a robust alternative to conventional formulations based on normality. We used data on student test scores to illustrate the calculation of best predictors based on normality and on the more robust t-distribution. Other elliptical distributions such as the Laplace and slash distributions will be considered in a future work.

Acknowledgements

This research is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. The authors are grateful to Beilei Wu for computational assistance.

References

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

[2] Arellano-Valle, R. B. and Genton, M. G. (2010). Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics*, **1**, 17–33.

[3] Arellano-Valle, R. B. and Genton, M. G. (2008). On the exact distribution of the maximum of absolutely dependent random variables. *Statistics and Probability Letters*, **78**, 27–32.

[4] Arellano-Valle, R. B. and Genton, M. G. (2007). On the exact distribution of linear combinations of order statistics from dependent random variables. *Journal of Multivariate Analysis*, **98**, 1876–1894.

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

[6] Balakrishnan, N. (1993). Multivariate normal distributions and multivariate order statistics induced by ordering linear combinations. *Statistics and Probability Letters*, **17**, 343-350.

[7] Balakrishnan, N., Ong, S. H., Habibi, M., and Jamalizadeh, A. (2011). Nonlinear regression in a trivariate elliptical distribution via a linear combination of order statistics. To appear in *Statistics*.

[8] Basu, A. P. and Ghosh, J. K. (1978). Identifiability of the multinormal and other distributions under competing risks model. *Journal of Multivariate Analysis*, **8**, 413–429.

[9] Genc, A. (2006). A note on the distributions of linear functions of two ordered correlated random variables. *Communications in Statistics–Theory & Methods*, **35**, 989–995.

[10] Gupta, S. S. and Pillai, K. C. S. (1965). On linear functions of ordered correlated normal random variables. *Biometrika*, **52**, 367–379.

[11] Jamalizadeh, A. and Balakrishnan, N. (2011). Concomitants of order statistics from multivariate elliptical distributions. To appear in *Journal of Statistical Planning and Inference*.

[12] Jamalizadeh, A. and Balakrishnan, N. (2010). Distributions of order statistics and linear combinations of order statistics from an elliptical distribution as mixtures of skew-elliptical distributons. *Journal of Multivariate Analysis*, **101**, 1412–1427.

[13] Jamalizadeh, A. and Balakrishnan, N. (2009a). Prediction in a trivariate normal distribution via a linear combination of order statistics. *Statistics and Probability Letters*, **79**, 2289–2296.

[14] Jamalizadeh, A. and Balakrishnan, N. (2009b). Order statistics from trivariate normal and t_ν-distributions in terms of generalized skew-normal and skew-t_ν distributions. *Journal of Statistical Planning and Inference*, **139**, 3799–3819.

[15] Jamalizadeh, A. and Balakrishnan, N. (2008). On order statistics from bivariate skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference*, **138**, 4187–4197.

[16] Jamalizadeh, A., Khosravi, M., and Balakrishnan, N. (2009a). Recurrence relations for distributions of a skew-t and a linear combination of order statistics from a bivariate-t. *Computational Statistics and Data Analysis*, **53**, 847–852.

[17] Jamalizadeh, A., Mehrali, Y., and Balakrishnan, N. (2009b). Recurrence relations for bivariate t and extended skew-t distributions and an application to order statistics from bivariate-t. *Computational Statistics and Data Analysis*, **53**, 4018–4027.

[18] Liu, C. and Rubin, D. B. (1995). ML estimation of the multivariate t distribution with unknown degrees of freedom. *Statistica Sinica*, **5**, 19–39.

[19] Loperfido, N. (2008a). A note of skew-elliptical distributions and linear functions of order statistics. *Statistics and Probability Letters*, **78**, 3184–3186.

[20] Loperfido, N. (2008b). Modeling maxima of longitudinal contralateral observations. *Test*, **17**, 370–380.

[21] Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.

[22] Nagaraja, H. N. (1982). A note on linear functions of ordered correlated normal random variables. *Biometrika*, **69**, 284–285.

[23] Olkin, I. and Viana, M. (1995). Correlation analysis of extreme observations from a multivariate normal distribution. *Journal of the American Statistical Association*, **90**, 1373–1379.

[24] Viana, M. (1998). Linear combinations of ordered symmetric observations with applications to visual acuity. In *Handbook of Statistics–Order Statistics:Applications*, N. Balakrishnan and C. R. Rao (eds.), vol. **17**, pp. 513–524, North Holland, Amsterdam.

[25] Wiens, D. P., Beaulieu, P., and Loskot, N. C. (2005). On the exact distribution of the sum of the largest $n-k$ out of n normal random variables with differing means. *Statistics*, **40**, 165–173.