Multivariate unified skew-*t* distributions and their properties

Kesen Wang [a],[*], Maicon J. Karling [a], Reinaldo B. Arellano-Valle [b], Marc G. Genton [a]

[a] *Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia*
[b] *Departamento de Estadística, Pontificia Universidad Catolica de Chile, Santiago, 7820436, Chile*

ARTICLE INFO

ABSTRACT

The unified skew-*t* (SUT) is a flexible parametric multivariate distribution that accounts for skewness and heavy tails in the data. A few of its properties can be found scattered in the literature or in a parameterization that does not follow the original one for unified skew-normal (SUN) distributions, yet a systematic study is lacking. In this work, explicit properties of the multivariate SUT distribution are presented, such as its stochastic representations, moments, SUN-scale mixture representation, linear transformation, additivity, marginal distribution, canonical form, quadratic form, conditional distribution, change of latent dimensions, Mardia measures of multivariate skewness and kurtosis, and non-identifiability issue. These results are given in a parameterization that reduces to the original SUN distribution as a sub-model, hence facilitating the use of the SUT for applications. Several models based on the SUT distribution are provided for illustration.

1. Introduction

Multivariate distributions beyond the classical Gaussian framework are needed to model modern datasets. For instance, Ogundimu and Hutton [32] studied skew-normal sample selection models to address non-Gaussian settings in econometrics data. Arellano-Valle et al. [8] investigated scale and shape mixtures of multivariate skew-normal distributions to more flexibly analyze a wind dataset. Adcock and Azzalini [2] provided a comprehensive overview of the developments of skew-elliptical models and their applications. The need for non-Gaussian models, allowing to control skewness and kurtosis in the distribution of data, has grown rapidly over the past few years. Recently, Loperfido [29] explored the skewness of normal mixtures. Karlsson et al. [26] studied the classic time series vector autoregressive model under skewed and heavy-tailed settings. Kiss et al. [27] put non-Gaussian models into real-world forecasting applications.

Also to this end, Arellano-Valle et al. [6] proposed a selection approach to obtain multivariate distributions in a unified way while being very flexible in terms of controlling skewness and kurtosis features. Specifically, given two random vectors $\mathbf{U}_0 \in \mathbb{R}^m$ and $\mathbf{U}_1 \in \mathbb{R}^d$, and a subset $C \subset \mathbb{R}^m$, they coined the distribution of the random vector $\mathbf{Z} = (\mathbf{U}_1 | \mathbf{U}_0 \in C)$ a selection distribution. In this way, the cumulative distribution function (cdf) of $\mathbf{Z}$ can be easily computed as

$$F_{\mathbf{Z}}(\mathbf{z}) = \Pr(\mathbf{U}_1 \leq \mathbf{z} | \mathbf{U}_0 \in C) = \frac{\Pr(\mathbf{U}_0 \in C, \mathbf{U}_1 \leq \mathbf{z})}{\Pr(\mathbf{U}_0 \in C)}, \quad \mathbf{z} \in \mathbb{R}^d. \tag{1}$$

In the absolutely continuous case, the probability density function (pdf) of $\mathbf{Z}$ is then

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{U}_1 | \mathbf{U}_0 \in C}(\mathbf{z}) = f_{\mathbf{U}_1}(\mathbf{z}) \frac{\Pr(\mathbf{U}_0 \in C | \mathbf{U}_1 = \mathbf{z})}{\Pr(\mathbf{U}_0 \in C)}, \quad \mathbf{z} \in \mathbb{R}^d, \tag{2}$$

where $f_{\mathbf{U}_1}(\mathbf{z})$ is the pdf of $\mathbf{U}_1$. Note that the selection pdf (2) may also be motivated by

$$\Pr(\mathbf{U}_0 \in C) = \mathrm{E}\{\Pr(\mathbf{U}_0 \in C | \mathbf{U}_1)\} = \int_C \Pr(\mathbf{U}_0 \in C | \mathbf{U}_1 = \mathbf{z}) f_{\mathbf{U}_1}(\mathbf{z}) \mathrm{d}\mathbf{z}.$$

One of the best-known examples of a selection distribution is the multivariate unified skew-normal (SUN) distribution, studied by Arellano-Valle and Azzalini [4], that can account for skewness in the data. For this, they first defined the selection random vector $\mathbf{Z} = (\mathbf{U}_1 | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0})$ with $\boldsymbol{\tau} \in \mathbb{R}^m$ being a vector of truncation parameters and assuming that $\mathbf{U}_0$ and $\mathbf{U}_1$ have a multivariate normal joint distribution with zero mean and positive-definite correlation matrix $\bar{\boldsymbol{\Omega}}^*$ ($\bar{\boldsymbol{\Omega}}^* > 0$ hereinafter); that is,

$$\begin{pmatrix} \mathbf{U}_0 \\ \mathbf{U}_1 \end{pmatrix} \sim \mathcal{N}_{m+d}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \bar{\boldsymbol{\Omega}}^* = \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & \boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta} & \bar{\boldsymbol{\Omega}} \end{pmatrix} \right), \tag{3}$$

where $\bar{\boldsymbol{\Gamma}}$ and $\bar{\boldsymbol{\Omega}}$ are the correlation matrices of $\mathbf{U}_0$ and $\mathbf{U}_1$, respectively, and $\boldsymbol{\Delta}$ is the correlation matrix between $\mathbf{U}_0$ and $\mathbf{U}_1$. Then they defined the SUN distribution as the distribution of $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}$, with $\boldsymbol{\omega} > 0$ being a diagonal $d \times d$ scale matrix, denoted by $\mathbf{Y} \sim \mathcal{SUN}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$, with $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$, and with pdf

$$f_{\mathbf{Y}}(\mathbf{y}) = |\boldsymbol{\omega}^{-1}| f_{\mathbf{Z}}\{\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})\} = \phi_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \frac{\Phi_m\left\{\boldsymbol{\tau} + \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}\right\}}{\Phi_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}})}, \quad \mathbf{y} \in \mathbb{R}^d,$$

where $\phi_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega})$ is the pdf of $\mathcal{N}_d(\boldsymbol{\xi}, \boldsymbol{\Omega})$ and $\Phi_m(\cdot; \boldsymbol{\Psi})$ is the cdf of $\mathcal{N}_m(\mathbf{0}, \boldsymbol{\Psi})$; and with cdf

$$F_{\mathbf{Y}}(\mathbf{y}) = \Pr\{\mathbf{Z} \le \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})\} = \frac{\Phi_{d+m}(\mathbf{y}_* - \boldsymbol{\xi}_*; \boldsymbol{\Omega}_*)}{\Phi_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}})}, \quad \mathbf{y} \in \mathbb{R}^d,$$

where

$$\mathbf{y}_* = \begin{pmatrix} \boldsymbol{\tau} \\ \mathbf{y} \end{pmatrix}, \quad \boldsymbol{\xi}_* = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\xi} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Omega}_* = \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & -\boldsymbol{\Delta}^\top\boldsymbol{\omega} \\ -\boldsymbol{\omega}\boldsymbol{\Delta} & \boldsymbol{\Omega} \end{pmatrix}. \tag{4}$$

Another important example of a selection distribution is the multivariate unified skew-t (SUT) distribution, a flexible parametric family that can account for both skewness and heavy tails in the data. We start with a formal definition of the SUT distribution from the aforementioned selection approach. For this, we denote by $\mathcal{T}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ the d-dimensional multivariate t distribution with location vector $\boldsymbol{\xi} \in \mathbb{R}^d$, $d \times d$ dispersion matrix $\boldsymbol{\Omega}$ and degrees-of-freedom parameter $\nu > 0$; also its pdf and cdf are denoted by $t_d(\cdot; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ and $T_d(\cdot - \boldsymbol{\xi}; \boldsymbol{\Omega}, \nu)$, respectively.

Definition 1. A random vector $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}$, where $\mathbf{Z} = (\mathbf{U}_1 | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0})$, with $\boldsymbol{\tau} \in \mathbb{R}^m$ and

$$\begin{pmatrix} \mathbf{U}_0 \\ \mathbf{U}_1 \end{pmatrix} \sim \mathcal{T}_{m+d}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \bar{\boldsymbol{\Omega}}^* = \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & \boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta} & \bar{\boldsymbol{\Omega}} \end{pmatrix}, \nu \right), \tag{5}$$

is said to have a multivariate unified skew-t (SUT) distribution with location vector $\boldsymbol{\xi} \in \mathbb{R}^d$, $d \times d$ dispersion matrix $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$, $d \times m$ skewness matrix $\boldsymbol{\Delta}$, latent truncation vector $\boldsymbol{\tau} \in \mathbb{R}^m$, $m \times m$ latent correlation matrix $\bar{\boldsymbol{\Gamma}}$, and degrees-of-freedom parameter $\nu > 0$. In brief, $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$, where m is the latent dimension.

As in the SUN case, if $\bar{\boldsymbol{\Omega}}_* > 0$ then the selection approach used to define the SUT distribution allows to easily obtain the pdf and cdf of $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$ by applying (5) in (2) and (1), respectively, as follows:

$$f_{\mathbf{Y}}(\mathbf{y}) = t_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) \frac{T_m\left[\alpha_{\nu, Q_{\mathbf{y}}}^{-1/2}\left\{ \boldsymbol{\tau} + \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}) \right\}; \bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}, \nu + d \right]}{T_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}}, \nu)}, \quad \mathbf{y} \in \mathbb{R}^d, \tag{6}$$

where $\alpha_{\nu, Q_{\mathbf{y}}} = \{\nu + Q_{\mathbf{y}}\}/(\nu + d)$ with $Q_{\mathbf{y}} = (\mathbf{y} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})$; and

$$F_{\mathbf{Y}}(\mathbf{y}) = \frac{T_{d+m}(\mathbf{y}_* - \boldsymbol{\xi}_*; \boldsymbol{\Omega}_*, \nu)}{T_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}}, \nu)}, \quad \mathbf{y} \in \mathbb{R}^d, \tag{7}$$

with $\mathbf{y}_*$, $\boldsymbol{\xi}_*$ and $\boldsymbol{\Omega}_*$ defined as in (4). In the pdf (6) of the SUT, the factor $\alpha_{\nu, Q_{\mathbf{y}}}$ has to be included, whereas it does not arise in the case of the SUN. This originates from the conditional pdf generator for the multivariate elliptical distributions, including the multivariate t distribution as a particular case, specified in Fang et al. [19].

If $\nu \to \infty$ in Definition 1, then (5) reduces to (3) and, therefore, the SUT becomes the SUN distribution. If instead $m = 1$ in Definition 1, then $\mathbf{Y} \sim \mathcal{EST}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta}, \boldsymbol{\tau}, \nu)$, the extended skew-$t$ (EST) distribution introduced by Arellano-Valle and Genton [10], and if in addition $\nu \to \infty$, then $\mathbf{Y} \sim \mathcal{ESN}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta}, \boldsymbol{\tau})$, the extended skew-normal (ESN) distribution. Finally, if $m = 1$ and $\boldsymbol{\tau} = \mathbf{0}$ in Definition 1, then $\mathbf{Y} \sim \mathcal{ST}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta}, \nu)$, the skew-$t$ (ST) distribution in the form introduced by Azzalini and Capitanio [13], and if in addition $\nu \to \infty$, then $\mathbf{Y} \sim \mathcal{SN}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta})$, the skew-normal (SN) distribution of Azzalini and Dalla Valle [15]. The multivariate t distribution, $\mathbf{Y} \sim \mathcal{T}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$, is recovered by setting $\boldsymbol{\tau} = \mathbf{0}$ and $\boldsymbol{\Delta} = \mathbf{0}$ in Definition 1, from which we obtain the multivariate normal distribution, $\mathbf{Y} \sim \mathcal{N}_d(\boldsymbol{\xi}, \boldsymbol{\Omega})$, by taking the limit when $\nu \to \infty$.

Notice that one interesting property of the pdf and cdf of the SUT is that unlike the SUN distribution, for which $\boldsymbol{\Delta} = \mathbf{0}$ is sufficient to recover its corresponding elliptically symmetric parent distribution, the SUT requires both $\boldsymbol{\Delta} = \mathbf{0}$ and $\boldsymbol{\tau} = \mathbf{0}$. Indeed, when $\boldsymbol{\Delta} = \mathbf{0}$ and $\boldsymbol{\tau} = \mathbf{0}$, the pdf of $\mathbf{Y}$ becomes:

$$f_{\mathbf{Y}}(\mathbf{y}) = t_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) \frac{T_m(\mathbf{0}; \bar{\boldsymbol{\Gamma}}, \nu + d)}{T_m(\mathbf{0}; \bar{\boldsymbol{\Gamma}}, \nu)} = t_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu), \quad \mathbf{y} \in \mathbb{R}^d,$$

(a) $\mathcal{SUN}_{2,1} \equiv \mathcal{SN}_2$ (b) $\mathcal{SUN}_{2,2}$ (c) $\mathcal{SUN}_{2,3}$

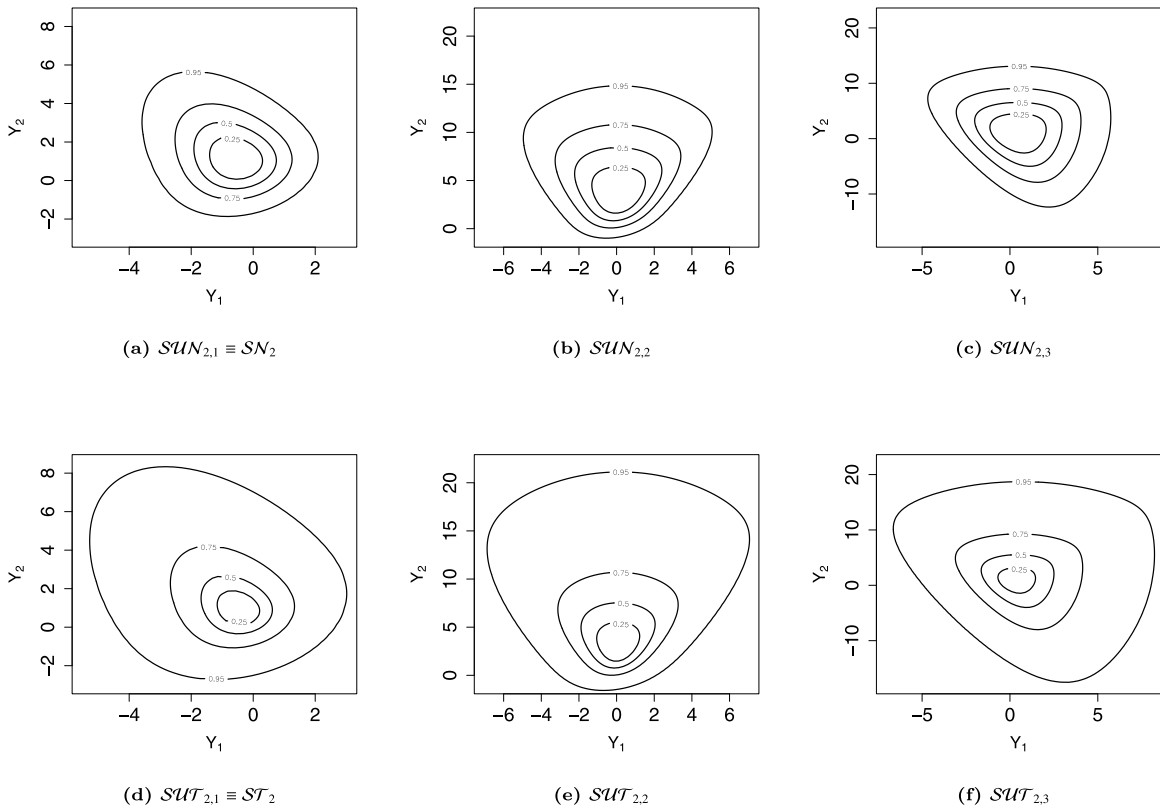(d) $\mathcal{SUT}_{2,1} \equiv \mathcal{ST}_2$ (e) $\mathcal{SUT}_{2,2}$ (f) $\mathcal{SUT}_{2,3}$

Fig. 1. Contours of the pdfs of the bivariate ($d = 2$) SUN and SUT (at $\nu = 5$) distributions with the same parameter specifications with $(\boldsymbol{\Omega}, \boldsymbol{\Delta}, \bar{\boldsymbol{\Gamma}})$ specified to make the distribution skewed in the direction $(-1, 2)^\top$ ($m = 1$), $\{(-1, 2)^\top, (1, 2)^\top\}$ ($m = 2$), and $\{(-1, 2)^\top, (1, 2)^\top, (1, -6)^\top\}$ ($m = 3$), respectively.

since $T_m(\mathbf{0}; \bar{\boldsymbol{\Gamma}}, \nu + d) = T_m(\mathbf{0}; \bar{\boldsymbol{\Gamma}}, \nu) = \boldsymbol{\Phi}_m(\mathbf{0}; \bar{\boldsymbol{\Gamma}})$ according to the properties proved in Fang et al. [19] and detailed later in Section 3.2. Therefore, the cdf of $\mathbf{Y}$ is then $T_d(\mathbf{y} - \boldsymbol{\xi}; \boldsymbol{\Omega}, \nu)$.

We present in Fig. 1 the pdf contours of the bivariate SUN and SUT (at $\nu = 5$) distributions with the same parameter specifications. Fig. 1 indicates that the SUN and SUT distributions are flexible in terms of the direction of skewness. With desired specifications of $(\boldsymbol{\Omega}, \boldsymbol{\Delta}, \bar{\boldsymbol{\Gamma}})$, it is practical to enable the contours to be skewed simultaneously toward any possible directions as shown in Figs. 1(a)–1(f). In particular, by the convolution representation of SUT (see Section 2.2) and of SUN [4], the column vectors of $\boldsymbol{\omega}\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}$ dictate the directions of skewness. In addition, compared with SUN, the SUT possesses more weights on the tails, which is the desirable heavy-tail property. The central bulks of the SUN, on the other hand, contain more weight than the SUT.

The formal properties of the aforementioned sub-models of Definition 1 have been investigated in detail in the literature. For example, the properties of the SUN distribution can be found in Arellano-Valle and Azzalini [4], Gupta et al. [24], Arellano-Valle and Azzalini [5], and Wang et al. [34], whereas the properties of the EST distribution are described in Arellano-Valle and Genton [10]. Moreover, book-length accounts of the properties of the SN and ST distributions can be found in Azzalini and Capitanio [14] and Genton [21]; see also the recent review by Azzalini [12] and references therein.

A few of the properties of the multivariate SUT distribution can be found scattered in the literature. In principle, these properties can be derived from those of the unified skew-elliptical (SUE) distributions studied by Arellano-Valle and Genton [11] by plugging the Student's t generator in these results, but this is a cumbersome task. Moreover, the parameterization used by Arellano-Valle and Genton [11] does not follow the original one for the SUN distributions introduced by Arellano-Valle and Azzalini [4]. Hence, a systematic study of the SUT distribution and its properties in the original parameterization of the SUN distribution is lacking.

In this work, explicit properties of the multivariate SUT distribution are presented, such as its stochastic representations, moments, SUN-scale mixture representation, linear transformation, additivity, marginal distribution, canonical form, quadratic form, conditional distribution, change of latent dimensions, Mardia measures of multivariate skewness and kurtosis, and non-identifiability issue. These results are given in the parameterization of Definition 1 that reduces to the original SUN distribution as a sub-model, hence facilitating the use of the SUT for applications. Several models based on the SUT distribution are provided for illustration.

The remainder of this paper is structured as follows. Section 2 discusses stochastic representation methods for the construction of the SUT with different possible parameterizations, the moments of the SUT, and relates the SUT to the SUN. Section 3 describes some of the SUT's main properties, including linear transformation, additivity, marginal distribution, canonical form, and quadratic forms. Section 4 provides the SUT's conditional distributions, as well as change of latent dimensions. Section 5 describes the Mardia

measures of multivariate skewness and kurtosis of the SUT. Lastly, Section 6 discusses a non-identifiability problem of the SUT with its remedies, and some identifiable sub-models. The paper ends with a discussion in Section 7.

2. Stochastic representations and moments

2.1. Several selection representations

As discussed in Arellano-Valle et al. [6], the same selection vector can be represented in several equivalent ways. Therefore, various equivalent representations exist for a random vector following a SUT distribution, $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$. In this section, we detail several possible choices.

From Definition 1, we notice that $\mathbf{Y} = \boldsymbol{\xi} + \mathbf{Z}^*$, where $\mathbf{Z}^* = \boldsymbol{\omega}\mathbf{Z} = (\boldsymbol{\omega}\mathbf{U}_1 | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0}) = (\mathbf{U}_1^* | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0})$, with $\mathbf{U}_1^* = \boldsymbol{\omega}\mathbf{U}_1$. Hence, the multivariate t distribution specified in Eq. (5) above can be reformulated as

$$\begin{pmatrix} \mathbf{U}_0 \\ \mathbf{U}_1^* \end{pmatrix} \sim \mathcal{T}_{m+d}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & \boldsymbol{\Delta}^\top\boldsymbol{\omega} \\ \boldsymbol{\omega}\boldsymbol{\Delta} & \boldsymbol{\Omega} \end{pmatrix}, \nu \right). \tag{8}$$

In addition, the additive constants can be integrated into the multivariate distribution to arrive at a more concise form with

$$\begin{pmatrix} \tilde{\mathbf{U}}_0 \\ \tilde{\mathbf{U}}_1 \end{pmatrix} \sim \mathcal{T}_{m+d}\left(\begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{\xi} \end{pmatrix}, \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & \boldsymbol{\Delta}^\top\boldsymbol{\omega} \\ \boldsymbol{\omega}\boldsymbol{\Delta} & \boldsymbol{\Omega} \end{pmatrix}, \nu \right), \tag{9}$$

and $\mathbf{Y} = (\tilde{\mathbf{U}}_1 | \tilde{\mathbf{U}}_0 > \mathbf{0})$. Some other formulations are also available from the SUE case in Arellano-Valle and Genton [11] with various linear transformations of the multivariate t distribution and the given conditions. In particular, the random vector $\mathbf{Z}$ can be reformulated as $(\mathbf{U}_1 | \mathbf{U}_0^* < \boldsymbol{\Lambda}\mathbf{U}_1 + \boldsymbol{\tau})$ with $\boldsymbol{\Lambda} = \boldsymbol{\Delta}^\top\bar{\boldsymbol{\Omega}}^{-1}$ and

$$\begin{pmatrix} \mathbf{U}_0^* \\ \mathbf{U}_1 \end{pmatrix} \sim \mathcal{T}_{m+d}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & \bar{\boldsymbol{\Omega}} \end{pmatrix}, \nu \right). \tag{10}$$

The equivalence relationship originates in the given condition $\mathbf{U}_0^* < \boldsymbol{\Lambda}\mathbf{U}_1 + \boldsymbol{\tau}$, which can also be expressed as $\mathbf{U}_0^* - \boldsymbol{\Lambda}\mathbf{U}_1 < \boldsymbol{\tau}$. Notice that (5) can be retrieved by setting $\mathbf{U}_0 = \boldsymbol{\Lambda}\mathbf{U}_1 - \mathbf{U}_0^*$ and applying the idempotent transformation $\mathbf{A} = \begin{pmatrix} -\mathbf{I}_m & \boldsymbol{\Lambda} \\ \mathbf{0} & \mathbf{I}_d \end{pmatrix}$ to $\begin{pmatrix} \mathbf{U}_0^* \\ \mathbf{U}_1 \end{pmatrix}$. Reciprocally, to reach (10) from (5), we can set $\mathbf{U}_0^* = \boldsymbol{\Lambda}\mathbf{U}_1 - \mathbf{U}_0$ and the joint distribution

$$\begin{pmatrix} \mathbf{U}_0^* \\ \mathbf{U}_1 \end{pmatrix} = \mathbf{A}\begin{pmatrix} \mathbf{U}_0 \\ \mathbf{U}_1 \end{pmatrix},$$

which still follows a multivariate t distribution with mean $\mathbf{0}$, covariance

$$\mathbf{A}\begin{pmatrix} \bar{\boldsymbol{\Gamma}} & \boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta} & \bar{\boldsymbol{\Omega}} \end{pmatrix}\mathbf{A}^\top = \begin{pmatrix} \bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & \bar{\boldsymbol{\Omega}} \end{pmatrix},$$

and degrees-of-freedom ν as specified in (10).

Another approach to arrive at the same result is by directly considering the representation $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}$ where $\mathbf{Z} = (\mathbf{U}_1 | \mathbf{U}_0^* < \boldsymbol{\Lambda}\mathbf{U}_1 + \boldsymbol{\tau})$. Then $\mathbf{Y} = \boldsymbol{\xi} + (\mathbf{U}_1^* | \mathbf{U}_0^* < \boldsymbol{\Lambda}\boldsymbol{\omega}^{-1}\mathbf{U}_1^* + \boldsymbol{\tau})$. Now if we set $\mathbf{U}_0 = \boldsymbol{\Lambda}\boldsymbol{\omega}^{-1}\mathbf{U}_1^* - \mathbf{U}_0^*$, then the joint distribution of $(\mathbf{U}_0^\top, \mathbf{U}_1^{*\top})^\top$ is exactly the same as specified in (8) after applying the corresponding linear transformation:

$$\begin{pmatrix} -\mathbf{I}_m & \boldsymbol{\Lambda}\boldsymbol{\omega}^{-1} \\ \mathbf{0} & \mathbf{I}_d \end{pmatrix} \text{ to } \begin{pmatrix} \mathbf{U}_0^* \\ \mathbf{U}_1^* \end{pmatrix}.$$

Although we have summarized numerous feasible settings of the conditioning mechanism for the generation of a SUT random vector, we will only apply the setting characterized in (8) and (9) to derive the properties of the SUT in the rest of this work due to consistency and simplicity, except for its quadratic form because we will have to impose the "uncorrelated" assumption of $\mathbf{U}_0^*$ and $\mathbf{U}_1$ implied in (10). A summary of the constructions for selection distributions, of which the SUT is a special case, can be found in Section 4 of Arellano-Valle et al. [6].

2.2. Convolution mechanism

In addition to the selection representation, there is an equivalent convolution representation. Analogous to the SUN distribution, the SUT can be represented by the convolution of a multivariate t random vector and a truncated multivariate t random vector.

Proposition 1. *Let* $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$. *Then* $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}$, *with* $\mathbf{Z} = \boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\mathbf{U}_* + \sqrt{(\nu + Q_{\mathbf{U}_*})/(\nu + m)} \cdot \mathbf{W}_*$, *where* $Q_{\mathbf{U}_*} = \mathbf{U}_*^\top\bar{\boldsymbol{\Gamma}}^{-1}\mathbf{U}_*$, $\mathbf{U}_* = (\mathbf{U}_0 | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0})$, $\mathbf{U}_0 \sim \mathcal{T}_m(\mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)$, $\mathbf{W}_* \sim \mathcal{T}_d(\mathbf{0}, \bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Delta}^\top, \nu + m)$, *and* $\mathbf{U}_0$ *and* $\mathbf{W}_*$ *are independent random vectors.*

Proof. The proof of this result is well known; see Eq. (9) of Arellano-Valle and Genton [10]. □

2.3. Mean and variance

With the convolution mechanism of the SUT, it is feasible to construct semi-explicit forms of the moments. To start, we propose the following lemma.

Lemma 1. *Let $Q_{\mathbf{U}_*} = \mathbf{U}_*^\top \bar{\boldsymbol{\Gamma}}^{-1} \mathbf{U}_*$, $\mathbf{U}_* = (\mathbf{U}_0 | \mathbf{U}_0 + \tau > \mathbf{0})$, and $\mathbf{U}_0 \sim \mathcal{T}_m(\mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)$. Then:*

$$\mathrm{E}\left\{ \left(\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m} \right)^{k/2} h(\mathbf{U}_*) \right\} = \frac{M_k(\tau, \bar{\boldsymbol{\Gamma}}, \nu)}{T_m(\tau; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)} \left(\frac{\nu}{\nu + m} \right)^{k/2} \frac{c(\nu, m)}{c(\nu - k, m)},$$

where

$$c(\upsilon, r) = \frac{\Gamma\{(\upsilon + r)/2\}}{\Gamma(\upsilon/2)(\pi \upsilon)^{r/2}}, \quad \upsilon, r > 0,$$

$$M_k(\tau, \bar{\boldsymbol{\Gamma}}, \nu) = \int_{\mathbf{u}+\tau>\mathbf{0}} h(\mathbf{u}) t_m\left(\sqrt{\frac{\nu - k}{\nu}} \mathbf{u}; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - k \right) \mathrm{d}\mathbf{u} = \left(\frac{\nu}{\nu - k} \right)^{m/2} T_m(\tau_k; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - k) \mathrm{E}\left\{ h\left(\sqrt{\frac{\nu}{\nu - k}} \mathbf{U}_k \right) \right\},$$

with $\mathbf{U}_k = (\mathbf{W}_k | \mathbf{W}_k + \tau_k > \mathbf{0})$, $\mathbf{W}_k \sim \mathcal{T}_m(\mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - k)$, and $\tau_k = \sqrt{(\nu - k)/\nu} \cdot \tau$ for $0 < k < \nu$.

Proof. First of all, we use the result that

$$\left(\frac{\nu + Q_{\mathbf{u}}}{\nu + m} \right)^{k/2} t_m(\mathbf{u}; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu) = \left(\frac{\nu}{\nu + m} \right)^{k/2} \frac{c(\nu, m)}{c(\nu - k, m)} t_m\left(\sqrt{\frac{\nu - k}{\nu}} \mathbf{u}; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - k \right),$$

where $\nu > k > 0$. Next, with the fact that $\mathbf{U}_*$ has pdf $g(\mathbf{u}) = t_m(\mathbf{u}; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu) / T_m(\tau; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)$ for $\mathbf{u} + \tau > \mathbf{0}$, we have:

$$\begin{aligned}
\mathrm{E}\left\{ \left(\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m} \right)^{k/2} h(\mathbf{U}_*) \right\} &= \int_{\mathbf{u}+\tau>\mathbf{0}} \left(\frac{\nu + Q_{\mathbf{u}}}{\nu + m} \right)^{k/2} h(\mathbf{u}) \, g(\mathbf{u}) \mathrm{d}\mathbf{u} \\
&= \frac{1}{T_m(\tau; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)} \int_{\mathbf{u}+\tau>\mathbf{0}} \left(\frac{\nu + Q_{\mathbf{u}}}{\nu + m} \right)^{k/2} h(\mathbf{u}) \, t_m(\mathbf{u}; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu) \mathrm{d}\mathbf{u} \\
&= \frac{1}{T_m(\tau; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)} \left(\frac{\nu}{\nu + m} \right)^{k/2} \frac{c(\nu, m)}{c(\nu - k, m)} \int_{\mathbf{u}+\tau>\mathbf{0}} h(\mathbf{u}) \, t_m\left(\sqrt{\frac{\nu - k}{\nu}} \mathbf{u}; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - k \right) \mathrm{d}\mathbf{u}.
\end{aligned}$$

Thus, the proof is done by the change of variable $\mathbf{v} = \sqrt{(\nu - k)/\nu} \cdot \mathbf{u}$ and using the fact that the truncated vector $\mathbf{U}_k$ has pdf given by $g(\mathbf{u}) = t_m(\mathbf{u}; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - k) / T_m(\tau_k; \mathbf{0}, \bar{\boldsymbol{\Gamma}}; \nu - k)$ for $\mathbf{u} + \tau_k > \mathbf{0}$, $\tau_k = \sqrt{(\nu - k)/\nu} \cdot \tau$, and $\nu - k > 0$. $\square$

Some specific results that emerge from Lemma 1 are the following:
1. If $h(\mathbf{u}) = 1$, then $M_k(\tau, \bar{\boldsymbol{\Gamma}}, \nu) = \{\nu/(\nu - k)\}^{m/2} T_m(\tau_k; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - k)$, and so

$$\mathrm{E}\left\{ \left(\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m} \right)^{k/2} \right\} = \frac{T_m(\tau_k; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - k)}{T_m(\tau; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)} \left(\frac{\nu}{\nu + m} \right)^{k/2} \frac{c(\nu, m)}{c(\nu - k, m)} \left(\frac{\nu}{\nu - k} \right)^{m/2}.$$

In particular, for $k = 2$:

$$\eta(Q_{\mathbf{U}_*}) = \mathrm{E}\left(\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m} \right) = \frac{T_m(\tau_k; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - 2)}{T_m(\tau; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)} \left(\frac{\nu}{\nu - 2} \right) \left(\frac{\nu + m - 2}{\nu + m} \right) \tag{11}$$

$$\implies \mathrm{E}\left(Q_{\mathbf{U}_*} \right) = \frac{T_m(\tau_k; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - 2)}{T_m(\tau; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)} \frac{m\nu}{\nu - 2}. \tag{12}$$

The result in (12) gives the expectation of the quadratic form of the truncated multivariate t random vector.
2. If $k = 2$ and $h(\mathbf{u}) = \mathbf{u}$, then

$$\mathrm{E}\left\{ \left(\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m} \right) \mathbf{U}_* \right\} = \frac{T_m(\tau_k; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - 2)}{T_m(\tau; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)} \left(\frac{\nu}{\nu - 2} \right)^{3/2} \left(\frac{\nu + m - 2}{\nu + m} \right) \mathrm{E}\left(\mathbf{U}_k \right).$$

3. Similarly, if $k = 2$ and $h(\mathbf{u}) = \mathbf{u}\mathbf{u}^\top$, then

$$\mathrm{E}\left\{ \left(\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m} \right) \mathbf{U}_* \mathbf{U}_*^\top \right\} = \frac{T_m(\tau_k; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu - 2)}{T_m(\tau; \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)} \left(\frac{\nu}{\nu - 2} \right)^{2} \left(\frac{\nu + m - 2}{\nu + m} \right) \mathrm{E}\left(\mathbf{U}_k \mathbf{U}_k^\top \right).$$

Now Lemma 1 and these specific cases enable the derivations of the semi-explicit forms of the mean and variance for the SUT, as well as the expectations $\mathrm{E}(V_*)$, $\mathrm{E}(V_* \mathbf{U}_*)$ and $\mathrm{E}(V_* \mathbf{U}_* \mathbf{U}_*^\top)$, where $V_* = (\nu + Q_{\mathbf{U}_*})/(\nu + m)$, required in Section 2.4.

Proposition 2. *Let $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \tau, \bar{\boldsymbol{\Gamma}}, \nu)$. Then:*

$$\mathrm{E}(\mathbf{Y}) = \boldsymbol{\xi} + \boldsymbol{\omega} \boldsymbol{\Delta} \bar{\boldsymbol{\Gamma}}^{-1} \mathrm{E}(\mathbf{U}_*), \quad \nu > 1,$$

$$\text{Var}(\mathbf{Y}) = \boldsymbol{\omega}\{\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\text{Var}(\mathbf{U}_*)\bar{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Delta}^{\top} + \eta(Q_{\mathbf{U}_*})(\bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}\boldsymbol{\Delta}^{\top})\}\boldsymbol{\omega}, \quad \nu > 2,$$

where $Q_{\mathbf{U}_*}$ is the same as in Proposition 1 and $\eta(Q_{\mathbf{U}_*})$ is given in (11).

Proof. Following from Proposition 1, $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}$, where $\mathbf{Z} = \boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\mathbf{U}_* + \sqrt{(\nu + Q_{\mathbf{U}_*})/(\nu + m)} \cdot \mathbf{W}_*$. Then, $\text{E}(\mathbf{Y}) = \text{E}\left\{\boldsymbol{\xi} + \boldsymbol{\omega}\left(\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\mathbf{U}_* + \sqrt{(\nu + Q_{\mathbf{U}_*})/(\nu + m)} \cdot \mathbf{W}_*\right)\right\} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\text{E}(\mathbf{U}_*)$ if $\nu > 1$, indicating that $\text{E}(\mathbf{Z}) = \boldsymbol{\omega}\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\text{E}(\mathbf{U}_*)$. The variance is $\text{Var}(\mathbf{Y}) = \text{Var}(\boldsymbol{\omega}\mathbf{Z}) = \boldsymbol{\omega}\text{Var}(\mathbf{Z})\boldsymbol{\omega}$, where

$$\text{Var}(\mathbf{Z}) = \boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\text{Var}(\mathbf{U}_*)\bar{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Delta}^{\top} + \text{Var}\left(\sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\mathbf{W}_*\right) + \boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\text{Cov}\left(\mathbf{U}_*, \sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\mathbf{W}_*\right) + \text{Cov}\left(\sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\mathbf{W}_*, \mathbf{U}_*\right)\bar{\boldsymbol{\Gamma}}\boldsymbol{\Delta}^{\top}.$$

Here,

$$\begin{aligned}
\text{Cov}\left(\mathbf{U}_*, \sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\mathbf{W}_*\right) &= \text{E}\left(\mathbf{U}_*\sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\mathbf{W}_*^{\top}\right) - \text{E}\left(\mathbf{U}_*\right)\text{E}\left(\sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\mathbf{W}_*\right)^{\top} \\
&= \text{E}\left(\mathbf{U}_*\sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\right)\text{E}(\mathbf{W}_*)^{\top} - \text{E}(\mathbf{U}_*)\text{E}\left(\sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\right)\text{E}(\mathbf{W}_*)^{\top} \\
&= \mathbf{0}.
\end{aligned}$$

Now, if $\nu > 2$, we then have

$$\begin{aligned}
\text{Var}\left(\sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\mathbf{W}_*\right) &= \text{E}\left(\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}\mathbf{W}_*\mathbf{W}_*^{\top}\right) - \text{E}\left(\sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\mathbf{W}_*\right)\text{E}\left(\sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\mathbf{W}_*\right)^{\top} \\
&= \text{E}\left(\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}\right)\text{E}(\mathbf{W}_*\mathbf{W}_*^{\top}) = \eta(Q_{\mathbf{U}_*})(\bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Delta}^{\top}),
\end{aligned}$$

where $\eta(Q_{\mathbf{U}_*}) = \text{E}\left\{(\nu + Q_{\mathbf{U}_*})/(\nu + m)\right\}$ is computed by letting $k = 2$ in Lemma 1 as indicated in (11). $\square$

The mean vector $\text{E}(\mathbf{U}_*)$ and covariance matrix $\text{Var}(\mathbf{U}_*)$ for the truncated multivariate t distribution of $\mathbf{U}_* = (\mathbf{U}_0|\mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0})$, with $\mathbf{U}_0 \sim \mathcal{T}_m(\mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)$, can be computed numerically using methods similar to Arellano-Valle et al. [7]. Two R packages allow for computations of the first and second moments of the SUT random vector, mnormt [16] and MomTrunc [20].

2.4. Higher-order moments

Explicit higher-order moments of the non-shifted SUT random vector, $\boldsymbol{\omega}\mathbf{Z}$, can be computed in the same way that the moments for the SUN are computed (see details in Proposition A.4 in the Appendix of Arellano-Valle and Azzalini [5]).

Proposition 3. Let $\mathbf{X} = \boldsymbol{\omega}\mathbf{Z} \sim \mathcal{SUT}_{d,m}(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$ and hence, $\mathbf{X} = \mathbf{A}\mathbf{U} + \mathbf{B}\mathbf{V}$ with

$$\mathbf{A} = \boldsymbol{\omega}\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}, \quad \mathbf{U} = \mathbf{U}_*, \quad \mathbf{B} = \boldsymbol{\omega}, \quad \mathbf{V} = V_*^{1/2}\mathbf{W}_*, \quad V_* = \frac{\nu + Q_{\mathbf{U}_*}}{\nu + m},$$

where $(\mathbf{U}_*, V_*)$ and $\mathbf{W}_*$ are independent and defined in Proposition 1 along with $Q_{\mathbf{U}_*}$. Then:

$$\begin{aligned}
\text{E}(\mathbf{X} \otimes \mathbf{X}\mathbf{X}^{\top}) =& (\mathbf{A} \otimes \mathbf{A})\mu_3(\mathbf{U})\mathbf{A}^{\top} + (\mathbf{I}_{d^2} + \mathbf{K}_d)(\mathbf{A} \otimes \mathbf{B})\left\{\text{E}(V_*\mathbf{U}) \otimes \mu_2(\mathbf{W}_*)\right\}\mathbf{B}^{\top} + (\mathbf{B} \otimes \mathbf{B})\text{vec}\{\mu_2(\mathbf{W}_*)\}\text{E}(V_*\mathbf{U}^{\top})\mathbf{A}^{\top}, \\
\text{E}(\mathbf{X}\mathbf{X}^{\top} \otimes \mathbf{X}\mathbf{X}^{\top}) =& (\mathbf{A} \otimes \mathbf{A})\mu_4(\mathbf{U})(\mathbf{A} \otimes \mathbf{A})^{\top} \\
& + (\mathbf{A} \otimes \mathbf{B})\{\text{E}(V_*\mathbf{U}\mathbf{U}^{\top}) \otimes \mu_2(\mathbf{W}_*)\}(\mathbf{A}^{\top} \otimes \mathbf{B}^{\top}) + (\mathbf{A} \otimes \mathbf{B})\{\text{E}(V_*\mathbf{U}\mathbf{U}^{\top}) \otimes \mu_2(\mathbf{W}_*)\}(\mathbf{A}^{\top} \otimes \mathbf{B}^{\top})\mathbf{K}_d \\
& + \mathbf{K}_d(\mathbf{A} \otimes \mathbf{B})\{\text{E}(V_*\mathbf{U}\mathbf{U}^{\top}) \otimes \mu_2(\mathbf{W}_*)\}(\mathbf{A}^{\top} \otimes \mathbf{B}^{\top}) + \mathbf{K}_d(\mathbf{A} \otimes \mathbf{B})\{\text{E}(V_*\mathbf{U}\mathbf{U}^{\top}) \otimes \mu_2(\mathbf{W}_*)\}(\mathbf{A}^{\top} \otimes \mathbf{B}^{\top})\mathbf{K}_d \\
& + (\mathbf{A} \otimes \mathbf{A})\text{vec}\{\text{E}(V_*\mathbf{U}\mathbf{U}^{\top})\}\text{vec}\{\mu_2(\mathbf{W}_*)\}^{\top}(\mathbf{B}^{\top} \otimes \mathbf{B}^{\top}) + (\mathbf{B} \otimes \mathbf{B})\text{vec}\{\mu_2(\mathbf{W}_*)\}\text{vec}\{\text{E}(V_*\mathbf{U}\mathbf{U}^{\top})\}^{\top}(\mathbf{A}^{\top} \otimes \mathbf{A}^{\top}) \\
& + (\mathbf{B} \otimes \mathbf{B})\mu_2(V_*)\mu_4(\mathbf{W}_*)(\mathbf{B}^{\top} \otimes \mathbf{B}^{\top}),
\end{aligned}$$

where $\mu_k(\cdot)$ represents the kth multivariate moment of the indicated variable and $\mathbf{K}_d$ is the commutation matrix of order $d \times d$.

Proof. Since $\mathbf{X} = \mathbf{A}\mathbf{U} + \mathbf{B}\mathbf{V}$, the results above can be obtained by directly applying the formula obtained in Proposition A.4 in Arellano-Valle and Azzalini [5]. In this way, although in our case the vectors $\mathbf{U} = \mathbf{U}_*$ and $\mathbf{V} = V_*^{1/2}\mathbf{W}_*$ are not independent since V_* depends on $\mathbf{U}_*$ through $Q_{\mathbf{U}_*}$, the moments up to the fourth order can be calculated directly from the proof of Proposition A.4 in Arellano-Valle and Azzalini [5].

In fact, for those expectations involving the function V_* of the truncated vector $\mathbf{U} = \mathbf{U}_*$, we must first use Lemma 1 and then the procedures proposed in the literature for the calculation of truncated moments under the multivariate t distribution (see specific cases 1, 2 and 3 after Lemma 1). The moments that involve the variable $\mathbf{V} = \mathbf{W}_*$ are obtained in a similar way to what has been indicated in Section 2.5 with $\mathbf{W}_* = V^{-1/2}\mathbf{Z}_0$, where $V \sim \mathcal{G}amma(\nu/2, \nu/2)$ and $\mathbf{Z}_0 \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Omega})$, and they are independent. $\square$

We calculate only the third and fourth moments because the first and second can be obtained directly through the mean and variance computed in Section 2.3. For the shifted SUT random vector, $\mathbf{Y}$, the detailed calculations of the moments are provided in Equations (A.5)-(A.8) in the Appendix of Arellano-Valle and Azzalini [5].

2.5. SUN-scale mixtures

2.5.1. SUN-scale mixture representation

Gupta [22] has shown that the skew-t (ST) distribution can be represented as a skew-normal (SN) scale mixture. Here, we show that the SUT distribution has a similar scale mixture representation concerning the SUN distribution under the condition that the vector of truncation parameters $\boldsymbol{\tau} = \mathbf{0}$.

Proposition 4. *Let* $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)$. *Then,*

$$\mathbf{Y} = \boldsymbol{\xi} + V^{-1/2}\mathbf{Z}_0, \tag{13}$$

where $V \sim \mathcal{G}amma(\nu/2, \nu/2)$ *and* $\mathbf{Z}_0 \sim \mathcal{SUN}_{d,m}(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \mathbf{0}, \bar{\boldsymbol{\Gamma}})$ *are independent.*

Proof. Similar to the SUT case, we refer to the stochastic representation for the SUN random vector, $\mathbf{Z}_0 \overset{d}{=} (\boldsymbol{\omega}\mathbf{U}_1 | \mathbf{U}_0 > \mathbf{0})$, where according to the definition of SUN distribution from Arellano-Valle and Azzalini [4], we have that $(\mathbf{U}_0^\top, \mathbf{U}_1^\top)^\top$ follows the distribution in (3). Then:

$$\begin{aligned} \boldsymbol{\xi} + V^{-1/2}\mathbf{Z}_0 &= \boldsymbol{\xi} + V^{-1/2}\left(\boldsymbol{\omega}\mathbf{U}_1 | \mathbf{U}_0 > \mathbf{0}\right) = \boldsymbol{\xi} + (V^{-1/2}\boldsymbol{\omega}\mathbf{U}_1 | \mathbf{U}_0 > \mathbf{0}) \\ &= \boldsymbol{\xi} + (V^{-1/2}\tilde{\mathbf{U}} | V^{-1/2}\mathbf{U}_0 > \mathbf{0}), \end{aligned} \tag{14}$$

where in (14) we applied the independence assumption between V and $\mathbf{Z}_0$; the independence assumption is extended to $(V, \tilde{\mathbf{U}})$ and $(V, \mathbf{U}_0)$ for the conditioning mechanism. Also, we used the fact that $V^{-1/2} > 0$ with probability 1. Lastly, we have

$$V^{-1/2}\begin{pmatrix} \mathbf{U}_0 \\ \tilde{\mathbf{U}} \end{pmatrix} \sim \mathcal{T}_{m+d}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & \boldsymbol{\Delta}^\top\boldsymbol{\omega} \\ \boldsymbol{\omega}\boldsymbol{\Delta} & \boldsymbol{\Omega} \end{pmatrix}, \nu \right).$$

Therefore, (14) yields a SUT random vector $\mathbf{Y}$ according to the stochastic representation (8) in Section 2. $\quad\square$

An immediate consequence of the SUN-scale mixture property is that it enables retrieval of the SUN random vector through conditioning.

Corollary 1. *Let* $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)$ *and consider the SUN-scale representation presented in Proposition 4. Then, for any* $\upsilon > 0$,

$$(\mathbf{Y} | V = \upsilon) \sim \mathcal{SUN}_{d,m}\left(\boldsymbol{\xi}, \upsilon^{-1}\boldsymbol{\Omega}, \upsilon^{-1/2}\boldsymbol{\Delta}, \mathbf{0}, \bar{\boldsymbol{\Gamma}}\right).$$

Proof. Using the representation in (13), we have

$$(\mathbf{Y} | V = \upsilon) = \left(\boldsymbol{\xi} + V^{-1/2}\mathbf{Z}_0 | V = \upsilon\right) = \boldsymbol{\xi} + \upsilon^{-1/2}\mathbf{Z}_0,$$

which follows a $\mathcal{SUN}_{d,m}\left(\boldsymbol{\xi}, \upsilon^{-1}\boldsymbol{\Omega}, \upsilon^{-1/2}\boldsymbol{\Delta}, \mathbf{0}, \bar{\boldsymbol{\Gamma}}\right)$ distribution. $\quad\square$

Finally, we would like to remark that, according to the selection approach used in the definition of the SUT, the restriction $\boldsymbol{\tau} = \mathbf{0}$ is imperative in Proposition 4 and Corollary 1 because we need the random vector $V^{-1/2}\mathbf{U}_0$, as indicated in (14), in the given condition of the selection representation to formulate a multivariate t distribution. If $\boldsymbol{\tau} \neq \mathbf{0}$, we would have a multivariate t random vector plus a scaled Gamma random vector, which is inconsistent with the setting of the selection representation of a SUT distribution.

2.5.2. SUN-scale mixture-based methods for the moments

Semi-explicit moments can be derived through the SUN-scale mixture representation obtained in Section 2.5.1.

Proposition 5. *Let* $\mathbf{Y} = V^{-1/2}\mathbf{Z}_0 \sim \mathcal{SUT}_{d,m}(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)$, *where* $V \sim \mathcal{G}amma(\nu/2, \nu/2)$ *and* $\mathbf{Z}_0 \sim \mathcal{SUN}_{d,m}(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \mathbf{0}, \bar{\boldsymbol{\Gamma}})$ *are independent. Then,* $\mu_i(\mathbf{Y}) = \mu_i(V^{-1/2})\mu_i(\mathbf{Z}_0), i \in \{1, \dots, 4\}$ *exist in semi-explicit forms, provided that* $\mu_i(V^{-1/2})$ *is finite.*

Proof. In the above setting:

$$\mu_1(\mathbf{Y}) = \mathrm{E}(\mathbf{Y}) = \mathrm{E}(V^{-1/2}\mathbf{Z}_0) = \mathrm{E}(V^{-1/2})\mathrm{E}(\mathbf{Z}_0) = M_1\mu_1(\mathbf{Z}_0).$$

Here $\mu_i(\mathbf{Z}_0)$, $i = 1, \dots, 4$, indicate the multivariate moments of the SUN random vector $\mathbf{Z}_0$ with explicit forms up to the fourth order as computed in Arellano-Valle and Azzalini [5]. The existence and computation of $M_1 = \mathrm{E}(V^{-1/2})$ and the corresponding higher-order moments M_2, M_3, and M_4 are well known; see, e.g., Rohatgi and Ehsanes Saleh [33], p. 212. Indeed, let $X \sim \mathcal{G}amma(\alpha, \beta)$. Then,

the moments for X are $\mathrm{E}(X^k) = \beta^k \Gamma(k+\alpha)/\Gamma(\alpha)$, $k + \alpha > 0$. The variance of the SUT random vector $\mathbf{Y}$ can be constructed through the second moments:

$$\mu_2(\mathbf{Y}) = \mathrm{E}(\mathbf{Y}\mathbf{Y}^\top) = \mathrm{E}(V^{-1}\mathbf{Z}_0\mathbf{Z}_0^\top) = \mathrm{E}(V^{-1})\mathrm{E}(\mathbf{Z}_0\mathbf{Z}_0^\top) = M_2\mu_2(\mathbf{Z}_0).$$

Therefore, $\mathrm{Var}(\mathbf{Y}) = \mu_2(\mathbf{Y}) - \mu_1(\mathbf{Y})\mu_1(\mathbf{Y})^\top$. Higher-order moments of $\mathbf{Y}$ can also be obtained using similar approaches. For instance, the third moment is

$$\mu_3(\mathbf{Y}) = \mathrm{E}(\mathbf{Y} \otimes \mathbf{Y}\mathbf{Y}^\top) = \mathrm{E}(V^{-3/2}\mathbf{Z}_0 \otimes \mathbf{Z}_0\mathbf{Z}_0^\top) = \mathrm{E}(V^{-3/2})\mathrm{E}(\mathbf{Z}_0 \otimes \mathbf{Z}_0\mathbf{Z}_0^\top) = M_3\mu_3(\mathbf{Z}_0),$$

and the fourth moment is

$$\mu_4(\mathbf{Y}) = \mathrm{E}(\mathbf{Y}\mathbf{Y}^\top \otimes \mathbf{Y}\mathbf{Y}^\top) = \mathrm{E}(V^{-2}\mathbf{Z}_0\mathbf{Z}_0^\top \otimes \mathbf{Z}_0\mathbf{Z}_0^\top) = \mathrm{E}(V^{-2})\mathrm{E}(\mathbf{Z}_0\mathbf{Z}_0^\top \otimes \mathbf{Z}_0\mathbf{Z}_0^\top) = M_4\mu_4(\mathbf{Z}_0). \quad \square$$

For all of the moments computed above, we set $\xi = \mathbf{0}$ to avoid the tedious polynomial expansions involved in the higher-order moments so that we can lay down the foundations for the moments of the shifted random vector. Therefore, the condition $\tilde{\xi} = \mathbf{0}$ is not a restriction. The moments can be exactly computed by recognizing that the shifted random vector is $\tilde{\mathbf{Y}} = \xi + \mathbf{Y} \sim \mathcal{SUT}_{d,m}(\xi, \Omega, \Delta, \mathbf{0}, \bar{\Gamma}, \nu)$ and by applying the expansions used in Equations (A.5)-(A.8) in Arellano-Valle and Azzalini [5] together with the moments computed above.

3. Linear transformations and quadratic forms

3.1. Linear transformations

To start, we describe the properties of linear transformations of the SUT random vector.

Proposition 6. *Let* $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, \nu)$. *The linear transformation* $\mathbf{Y_A} = \mathbf{A}\mathbf{Y} + \mathbf{b}$, *where* $\mathbf{A} \in \mathbb{R}^{n \times d}$ *with rank* $n \leq d$ *and* $\mathbf{b} \in \mathbb{R}^n$, *follows a* $\mathcal{SUT}_{n,m}(\xi_\mathbf{A}, \Omega_\mathbf{A}, \Delta_\mathbf{A}, \tau, \bar{\Gamma}, \nu)$ *with* $\xi_\mathbf{A} = \mathbf{A}\xi + \mathbf{b}$, $\Omega_\mathbf{A} = \mathbf{A}\Omega\mathbf{A}^\top$, *and* $\Delta_\mathbf{A} = \omega_\mathbf{A}^{-1}\mathbf{A}\omega\Delta$, *where* $\omega_\mathbf{A} = \mathrm{diag}(\Omega_\mathbf{A})^{1/2}$.

Proof. Following from the assumption in the proposition,

$$\mathbf{Y_A} = \mathbf{A}\mathbf{Y} + \mathbf{b} = \mathbf{A}(\xi + \omega\mathbf{Z}_0) + \mathbf{b} = \xi_\mathbf{A} + \mathbf{A}\omega\mathbf{Z}_0 = \xi_\mathbf{A} + (\mathbf{U}_1^\mathbf{A}|\mathbf{U}_0 + \tau > \mathbf{0}), \tag{15}$$

where $\mathbf{U}_1^\mathbf{A} = \mathbf{A}\omega\mathbf{U}_1$. From (15), the random vector

$$\begin{pmatrix} \mathbf{U}_0 \\ \mathbf{U}_1^\mathbf{A} \end{pmatrix} \sim \mathcal{T}_{m+n}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \bar{\Gamma} & \Delta^\top\omega\mathbf{A}^\top \\ \mathbf{A}\omega\Delta & \mathbf{A}\Omega\mathbf{A}^\top \end{pmatrix}, \nu \right) = \mathcal{T}_{m+n}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \bar{\Gamma} & \Delta_\mathbf{A}^\top\omega_\mathbf{A} \\ \omega_\mathbf{A}\Delta_\mathbf{A} & \Omega_\mathbf{A} \end{pmatrix}, \nu \right).$$

Again, by the stochastic representation (8), $\mathbf{Y_A}$ follows a SUT distribution with transformed parameters as specified above. $\quad \square$

Similar to the SUN distribution, the latent parameters τ and $\bar{\Gamma}$ of a SUT distribution remain unchanged under linear transformations. Moreover, the degrees-of-freedom parameter remains unchanged too. With the general linear transformation property, we can formulate a series of propositions involving additivity, marginal distributions, and canonical form.

3.1.1. Marginal distribution

In this section, we show that the marginals of a SUT distribution still follow a SUT distribution. For this purpose, we assume that the random vector $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, \nu)$ can be partitioned as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim \mathcal{SUT}_{d_1+d_2,m}\left(\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}, \tau, \bar{\Gamma}, \nu \right), \tag{16}$$

where $\mathbf{Y}_i \in \mathbb{R}^{d_i}$, $\xi_i \in \mathbb{R}^{d_i}$, $\Omega_{ij} \in \mathbb{R}^{d_i \times d_j}$, $\omega_i = \mathrm{diag}(\Omega_{ii})^{1/2}$, $\bar{\Omega}_{ij} = \omega_i^{-1}\Omega_{ij}\omega_j^{-1}$ and $\Delta_i \in \mathbb{R}^{d_i \times m}$, for $i, j \in \{1, 2\}$, such that $d_1 + d_2 = d$.

Proposition 7. *Let* $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, \nu)$. *Then, the marginal random vector* $\mathbf{Y}_i \sim \mathcal{SUT}_{d_i,m}(\xi_i, \Omega_{ii}, \Delta_i, \tau, \bar{\Gamma}, \nu)$, *for* $i \in \{1, 2\}$.

Proof. Consider the marginal random vector $\mathbf{Y}_1 = \mathbf{A}_1\mathbf{Y}$, where $\mathbf{A}_1 = (\mathbf{I}_{d_1}, \mathbf{0})$ denotes the projection matrix on the first d_1 coordinates. After applying the properties of linear transformations derived in Proposition 6, it follows that $\mathbf{Y}_1 \sim \mathcal{SUT}_{d_1,m}(\xi_1, \Omega_{11}, \Delta_1, \tau, \bar{\Gamma}, \nu)$ because $\mathbf{A}_1\xi = \xi_1$, $\mathbf{A}_1\Omega\mathbf{A}_1^\top = \Omega_{11}$, and $\omega_{\mathbf{A}_1}^{-1}\mathbf{A}_1\omega\Delta = \omega_1^{-1}\omega_1\Delta_1 = \Delta_1$, where $\bar{\Omega}_{11} = \omega_1^{-1}\Omega_1\omega_1^{-1}$. Analogously, one obtains the distribution of $\mathbf{Y}_2 = \mathbf{A}_2\mathbf{Y} \sim \mathcal{SUT}_{d_2,m}(\xi_2, \Omega_{22}, \Delta_2, \tau, \bar{\Gamma}, \nu)$, with $\mathbf{A}_2 = (\mathbf{0}, \mathbf{I}_{d_2})$. $\quad \square$

Note from Proposition 1 that each marginal can be represented stochastically as

$$\mathbf{Y}_i = \xi_i + \omega_i\mathbf{Z}_i, \quad \mathbf{Z}_i = \Delta_i\bar{\Gamma}^{-1}\mathbf{U}_* + \sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}}\mathbf{W}_{*i}, \quad i \in \{1, 2\},$$

where $\mathbf{W}_{*i} = \mathbf{A}_i \mathbf{W}_* \sim \mathcal{T}_{d_i}(\mathbf{0}, \bar{\mathbf{\Omega}}_{ii} - \mathbf{\Delta}_i \bar{\mathbf{\Gamma}}^{-1} \mathbf{\Delta}_i^\top, \nu)$ is independent of $\mathbf{U}_*$. From this result it is clear that if $\mathbf{\Delta}_i = \mathbf{0}$, then $\mathbf{Y}_i$ will have a symmetric distribution, which becomes a multivariate t when $\tau = \mathbf{0}$. Moreover, provided that the required moments exist, we have:

$$\mathrm{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) = \boldsymbol{\omega}_1 \mathrm{Cov}(\mathbf{Z}_1, \mathbf{Z}_2) \boldsymbol{\omega}_2 = \boldsymbol{\omega}_1 \mathrm{Cov}\left(\mathbf{\Delta}_1 \bar{\mathbf{\Gamma}}^{-1} \mathbf{U}_* + \sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}} \mathbf{W}_{*1}, \mathbf{\Delta}_2 \bar{\mathbf{\Gamma}}^{-1} \mathbf{U}_* + \sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}} \mathbf{W}_{*2} \right) \boldsymbol{\omega}_2$$

$$= \boldsymbol{\omega}_1 \mathbf{\Delta}_1 \bar{\mathbf{\Gamma}}^{-1} \mathrm{Var}(\mathbf{U}_*) \bar{\mathbf{\Gamma}}^{-1} \mathbf{\Delta}_2^\top \boldsymbol{\omega}_2 + \eta(Q_{\mathbf{U}_*}) \boldsymbol{\omega}_1 \mathrm{Cov}(\mathbf{W}_{*1}, \mathbf{W}_{*2}) \boldsymbol{\omega}_2,$$

where $\eta(Q_{\mathbf{U}_*})$ is defined in (11). We can then conclude that $\mathrm{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) = \mathbf{0}$ if any of the following conditions are met: (i) $\mathbf{\Delta}_1 = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{W}_{*1}, \mathbf{W}_{*2}) = \mathbf{0}$; or (ii) $\mathbf{\Delta}_2 = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{W}_{*1}, \mathbf{W}_{*2}) = \mathbf{0}$. Since $\mathrm{Cov}(\mathbf{W}_{*1}, \mathbf{W}_{*2}) = \frac{\nu}{\nu-2} \left(\bar{\mathbf{\Omega}}_{12} - \mathbf{\Delta}_1 \bar{\mathbf{\Gamma}}^{-1} \mathbf{\Delta}_2^\top \right)$, it follows that $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are uncorrelated under the condition that $\bar{\mathbf{\Omega}}_{12} = \mathbf{0}$ and either $\mathbf{\Delta}_1 = \mathbf{0}$ or $\mathbf{\Delta}_2 = \mathbf{0}$.

3.1.2. Additivity

A specific case of the linear transformation is the additivity of the marginals $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$, where $\mathbf{Y}_1 \in \mathbb{R}^d$ and $\mathbf{Y}_2 \in \mathbb{R}^d$. We have the following proposition.

Proposition 8. *Let $\mathbf{Y}_1 \in \mathbb{R}^d$ and $\mathbf{Y}_2 \in \mathbb{R}^d$ be two random vectors with joint distribution:*

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim SU\mathcal{T}_{d+d,m} \left(\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \begin{pmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{pmatrix}, \begin{pmatrix} \mathbf{\Delta}_1 \\ \mathbf{\Delta}_2 \end{pmatrix}, \tau, \bar{\mathbf{\Gamma}}, \nu \right). \tag{17}$$

Then $\mathbf{Y}_1 + \mathbf{Y}_2$ is $SU\mathcal{T}_{d,m}(\xi_+, \mathbf{\Omega}_+, \mathbf{\Delta}_+, \tau, \bar{\mathbf{\Gamma}}, \nu)$ with $\xi_+ = \xi_1 + \xi_2$, $\mathbf{\Omega}_+ = \mathbf{\Omega}_{11} + \mathbf{\Omega}_{22} + \mathbf{\Omega}_{12} + \mathbf{\Omega}_{21}$, $\boldsymbol{\omega}_+ = \mathrm{diag}(\mathbf{\Omega}_+)^{1/2}$, $\mathbf{\Delta}_+ = \boldsymbol{\omega}_+^{-1} \left(\boldsymbol{\omega}_1 \mathbf{\Delta}_1 + \boldsymbol{\omega}_2 \mathbf{\Delta}_2 \right)$ where $\mathbf{\Omega}_+ = \boldsymbol{\omega}_+ \bar{\mathbf{\Omega}}_+ \boldsymbol{\omega}_+$, $\mathbf{\Omega}_{11} = \boldsymbol{\omega}_1 \bar{\mathbf{\Omega}}_{11} \boldsymbol{\omega}_1$, and $\mathbf{\Omega}_{22} = \boldsymbol{\omega}_2 \bar{\mathbf{\Omega}}_{22} \boldsymbol{\omega}_2$.

Proof. The derivation for the additivity is through the properties of linear transformations demonstrated in Proposition 6. If we let $\mathbf{A} = (\mathbf{I}_d, \mathbf{I}_d)$ and $\mathbf{b} = \mathbf{0}$, then

$$\xi_+ = \mathbf{A} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \xi_1 + \xi_2, \quad \mathbf{\Omega}_+ = \mathbf{A} \begin{pmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{pmatrix} \mathbf{A}^\top = \mathbf{\Omega}_{11} + \mathbf{\Omega}_{22} + \mathbf{\Omega}_{12} + \mathbf{\Omega}_{21},$$

$$\mathbf{\Delta}_+ = \boldsymbol{\omega}_+^{-1} \mathbf{A} \begin{pmatrix} \boldsymbol{\omega}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\omega}_2 \end{pmatrix} \begin{pmatrix} \mathbf{\Delta}_1 \\ \mathbf{\Delta}_2 \end{pmatrix} = \boldsymbol{\omega}_+^{-1} \left(\boldsymbol{\omega}_1 \mathbf{\Delta}_1 + \boldsymbol{\omega}_2 \mathbf{\Delta}_2 \right). \quad \square$$

Therefore, the sum of two SUT random vectors still follows a SUT distribution provided that their joint distribution has the structure indicated in (17). However, unlike the SUN distribution, in which the sum of any two independent SUN random vectors are closed under the same distribution, the assumption that $(\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ follows a SUT distribution is needed [11]. Thus, only the sum of the marginals of a SUT random vector is closed under the same distribution. Some conditions to have a null correlation between $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are given in the previous subsection. For instance, this fact holds if $\mathbf{\Omega}_{12} = \mathbf{0}$ and $\mathbf{\Delta}_1 = \mathbf{0}$ or $\mathbf{\Delta}_2 = \mathbf{0}$.

However, uncorrelation of the marginals is not guaranteed for finite values of ν if $\mathbf{\Delta}_1 \neq \mathbf{0}$ and $\mathbf{\Delta}_2 \neq \mathbf{0}$, because the truncation operation introduces correlation between the corresponding marginals through $\mathbf{U}_*$. In fact, consider $(\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ partitioned as in (17) and, additionally, assume that $\mathbf{\Omega}_{12} = \mathbf{0}$, $\mathbf{\Delta}_1 = (\mathbf{\Delta}_{11} \quad \mathbf{0})$, $\mathbf{\Delta}_2 = (\mathbf{0} \quad \mathbf{\Delta}_{22})$ and $\bar{\mathbf{\Gamma}} = \mathrm{diag}(\bar{\mathbf{\Gamma}}_{11}, \bar{\mathbf{\Gamma}}_{22})$, with respective latent dimensions m_1 and m_2 ($m = m_1 + m_2$), and $\tau = \mathbf{0}$. So, by Proposition 4, we have $\mathbf{Y}_i = \xi_i + V^{-1/2} \mathbf{Z}_i$, where $V \sim \mathcal{G}amma(\nu/2, \nu/2)$ and $\mathbf{Z}_i \sim SU\mathcal{N}_{d,m_i}(\mathbf{0}, \mathbf{\Omega}_{ii}, \mathbf{\Delta}_{ii}, \mathbf{0}, \bar{\mathbf{\Gamma}}_{ii})$, for $i \in \{1, 2\}$. Here V is independent with respect to each $\mathbf{Z}_i$ and, by Appendix B in Arellano-Valle and Azzalini [4], $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are independent as well. Thus, we end up with

$$\mathrm{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) = (M_2 - M_1^2) \mathrm{E}(\mathbf{Z}_1) \mathrm{E}(\mathbf{Z}_2),$$

where $M_1 = \mathrm{E}(V^{-1/2}) = \sqrt{(\nu/2)} \, \Gamma\{(\nu - 1)/2\}/\Gamma(\nu/2)$ and $M_2 = \mathrm{E}(V^{-1}) = \nu/(\nu - 2)$. By equation (8) in Arellano-Valle and Azzalini [5], $\mathrm{E}(\mathbf{Z}_i)$ only vanishes if $\mathbf{\Delta}_{ii} = \mathbf{0}$. Moreover, $\lim_{\nu \to \infty}(M_2 - M_1^2) = 0$, so that $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are uncorrelated only in the limiting case, i.e., when $\nu \to \infty$, which brings us back to the SUN case.

3.1.3. Canonical form

An interesting result based on Proposition 1 is the formulation of the canonical form [17] of a SUT random vector. The canonical form moves all of the skewness of a SUT random vector toward the first component, leaving the remaining components symmetric.

Proposition 9. *Let $\mathbf{Y} \sim SU\mathcal{T}_{d,m}(\xi, \mathbf{\Omega}, \mathbf{\Delta}, \tau, \bar{\mathbf{\Gamma}}, \nu)$ be partitioned as $(\mathrm{Y}_1, \mathbf{Y}_2^\top)^\top$ with $m \geq 2$ and $d \geq m + 2$, where $\mathrm{Y}_1 \in \mathbb{R}$ and $\mathbf{Y}_2 \in \mathbb{R}^{d-1}$. Then, the representation in Proposition 1 with $\mathbf{\Delta} = (\mathbf{\Delta}_1^\top, \mathbf{\Delta}_2^\top)^\top$ and $\mathbf{\Delta}_2 = \mathbf{0}$ results in the canonical form, where $\mathbf{\Delta}_1 \in \mathbb{R}^m$ and $\mathbf{\Delta}_2 \in \mathbb{R}^{(d-1) \times m}$.*

Proof. Using the results from Proposition 1, we have $\mathbf{Y} = \xi + \boldsymbol{\omega} \left\{ \mathbf{B}_0 \mathbf{U}_* + \sqrt{(\nu + Q_{\mathbf{U}_*})/(\nu + m)} \cdot \mathbf{W}_* \right\}$ and

$$\begin{pmatrix} \mathrm{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \xi + \boldsymbol{\omega} \left\{ \begin{pmatrix} \mathbf{\Delta}_1 \bar{\mathbf{\Gamma}}^{-1} \\ \mathbf{0} \end{pmatrix} \mathbf{U}_* + \sqrt{\frac{\nu + Q_{\mathbf{U}_*}}{\nu + m}} \mathbf{W}_* \right\}. \tag{18}$$

In (18), the truncated t random vector is only distributed to the first component Y_1 and the quantity is $\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\mathbf{U}_*$. The $(d-1)\times m$ zero matrix $\boldsymbol{\Delta}_2$ annihilates the skewness in $\mathbf{Y}_2$. Therefore, Y_1 is a univariate SUT random variable and $\mathbf{Y}_2 \in \mathbb{R}^{d-1}$ is a symmetric multivariate random vector of the form $\sqrt{(\nu+Q_{\mathbf{U}_*})/(\nu+m)}\cdot\mathbf{W}_{*2}$, where $\mathbf{W}_* = (W_{*1}, \mathbf{W}_{*2}^\top)^\top$. As mentioned in the previous subsection, we note that $\sqrt{(\nu+Q_{\mathbf{U}_*})/(\nu+m)}\cdot\mathbf{W}_{*2}$ will be a multivariate t when $\boldsymbol{\tau}=\mathbf{0}$. $\quad\square$

With Propositions 6 and 9, it is possible to construct the canonical (linear) transformation matrix $\mathbf{A} = \mathbf{C} = (\mathbf{C}_1^\top, \mathbf{C}_2^\top)^\top$ for the regular SUT random vector $\mathbf{Y}$, where $\mathbf{C}_1 \in \mathbb{R}^d$ and $\mathbf{C}_2 \in \mathbb{R}^{(d-1)\times d}$ such that $\mathbf{C}_2\boldsymbol{\Delta}=\mathbf{0}$, so that $\mathbf{CY}$ has the canonical form. Note that the canonical transformation matrix $\mathbf{C}$ requires the column space of $\boldsymbol{\Delta}$ to be a subspace of the null space of $\mathbf{C}_2$. Hence, the canonical transformation does not always exist. This result is different from the SN and ST families, for which the canonical form always exists [17], and which, setting $m=1$, is equivalent to $\mathcal{SUN}_{d,1}$ or $\mathcal{SUT}_{d,1}$. The assured existence of the canonical transformation for the SN family is due to the fact that it is always possible to construct a transformation matrix with one-dimensional null space. In the SUT case ($m \geq 2$), the matrix $\boldsymbol{\Delta} \in \mathbb{R}^{d\times m}$. Hence, $\text{rank}(\boldsymbol{\Delta}) \leq \min(d,m)$. The null space for $\mathbf{C}_2 \in \mathbb{R}^{(d-1)\times d}$ is at most $(d-2)$-dimensional. Hence, we need $m \leq d-2$ in order for the canonical transformation to exist. Thus, for example, there is no canonical form for $d=3$ and $m=2$. The canonical form and transformation for the SUN are obtained when $\nu \to \infty$.

3.2. Quadratic forms

Arellano-Valle and Genton [9] have systematically explored the conditions for which the invariance property of the quadratic form holds for selection distributions. The main motivation was to find the condition that makes the two random vectors in the selection representation uncorrelated and therefore, the conditional cumulative distribution on the latent selection variable is identical to the unconditional cumulative distribution of the component given identical support. Following this approach, in this section we compute the quadratic form of the SUT random vector and explore the required conditions that establish the invariance property for the SUT.

According to Arellano-Valle et al. [6], one of the main properties of selection distributions is: if $\mathbf{Z} = (\mathbf{U}_1 | \mathbf{U}_0 \in C)$ then $g(\mathbf{Z}) = (g(\mathbf{U}_1) | \mathbf{U}_0 \in C)$ for every measurable function $g : \mathbb{R}^d \to \mathbb{R}^p$. It follows that, for example, if the distribution of $\mathbf{U}_1$ is closed under linear transformations then the distribution of $\mathbf{Z}$ is also closed. Note that if $g(\mathbf{U}_1)$ and $\mathbf{U}_0$ are uncorrelated (which can happen when $(\mathbf{U}_0, \mathbf{U}_1)$ is symmetric about the origin and $g(\cdot)$ is an even function), then we may have that $\Pr(\mathbf{U}_0 \in C | g(\mathbf{U}_1) = \mathbf{x}) = \Pr(\mathbf{U}_0 \in C)$.

With the aforementioned framework and property, we provide detailed steps to calculate the pdf of quadratic forms in SUT random vectors.

Proposition 10. *Let* $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$. *The quadratic form* $Q_{\mathbf{Y}} = (\mathbf{Y}-\boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1}(\mathbf{Y}-\boldsymbol{\xi})$ *has pdf*

$$f_{Q_{\mathbf{Y}}}(v) = f_{Q_{\mathbf{U}_1}}(v)\frac{\Pr(\mathbf{U}_0^* < \boldsymbol{\Lambda}\mathbf{U}_1 + \boldsymbol{\tau} | Q_{\mathbf{U}_1} = v)}{\Pr(\mathbf{U}_0^* < \boldsymbol{\Lambda}\mathbf{U}_1 + \boldsymbol{\tau})},$$

where $Q_{\mathbf{U}_1} = \mathbf{U}_1^\top \bar{\boldsymbol{\Omega}}^{-1}\mathbf{U}_1$ *with pdf* $f_{Q_{\mathbf{U}_1}}(v)$ *and* $\boldsymbol{\Lambda} = \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}$.

Proof. Consider the stochastic representation $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}_0$, where $\mathbf{Z}_0 = (\mathbf{U}_1 | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0})$ and note that $Q_{\mathbf{Y}} = Q_{\mathbf{Z}_0} = \mathbf{Z}_0^\top \bar{\boldsymbol{\Omega}}^{-1}\mathbf{Z}_0$. Then:

$$Q_{\mathbf{Z}_0} = \left(\mathbf{U}_1^\top \bar{\boldsymbol{\Omega}}^{-1}\mathbf{U}_1 | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0}\right) = \left(\mathbf{U}_1^\top \bar{\boldsymbol{\Omega}}^{-1}\mathbf{U}_1 | \mathbf{U}_0^* < \boldsymbol{\Lambda}\mathbf{U}_1 + \boldsymbol{\tau}\right) = \left(Q_{\mathbf{U}_1} | \mathbf{U}_0^* < \boldsymbol{\Lambda}\mathbf{U}_1 + \boldsymbol{\tau}\right),$$

where $\mathbf{U}_0^* = \boldsymbol{\Lambda}\mathbf{U}_1 - \mathbf{U}_0$, which is uncorrelated with $\mathbf{U}_1$ (see (10)). Hence, the pdf of the quadratic form has the form stated. $\quad\square$

To compute the conditional probability $\Pr(\mathbf{U}_0^* < \boldsymbol{\Lambda}\mathbf{U}_1 + \boldsymbol{\tau} | Q_{\mathbf{U}_1} = v)$, we need two well-known results from Fang et al. [19]. First, if $\mathbf{X}_1$ and $\mathbf{X}_2$ are jointly elliptical and uncorrelated, that is $(\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top \sim \mathcal{EC}_{d_1+d_2}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h)$ with location parameter $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$, dispersion matrix $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})$, and density generator function h, then $(\mathbf{X}_1 | \mathbf{X}_2) = (\mathbf{X}_1 | Q_{\mathbf{X}_2})$ and $(\mathbf{X}_2 | \mathbf{X}_1) = (\mathbf{X}_2 | Q_{\mathbf{X}_1})$, where $Q_{\mathbf{X}_1} = \mathbf{X}_1^\top \boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1$ and $Q_{\mathbf{X}_2} = \mathbf{X}_2^\top \boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2$. Second, it is well known that $\mathbf{W}_1 = Q_{\mathbf{X}_1}^{-1/2}\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{X}_1$ is a uniform random vector on the unit sphere of $\mathbb{R}^{d_1}$ (denoted usually by $\mathbf{U}^{(d_1)}$) and that it is independent of $Q_{\mathbf{X}_1}$ and of $Q_{\mathbf{X}_1}^{-1/2}\mathbf{X}_2$. Applying these two results, we have:

$$\Pr(\mathbf{U}_0^* < \boldsymbol{\Lambda}\mathbf{U}_1 + \boldsymbol{\tau} | Q_{\mathbf{U}_1} = v) = \text{E}_{\mathbf{W}_1}\{\Pr(\mathbf{T}_* < \bar{\boldsymbol{\Lambda}}\mathbf{W}_1 + \boldsymbol{\tau}_v | \mathbf{W}_1, Q_{\mathbf{U}_1} = v)\},$$

where $\mathbf{T}_* = Q_{\mathbf{U}_1}^{-1/2}\mathbf{U}_0^*$, $\bar{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}\bar{\boldsymbol{\Omega}}^{1/2}$, $\boldsymbol{\tau}_v = v^{-1/2}\boldsymbol{\tau}$, $\boldsymbol{\Upsilon} = \bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}$, and $\mathbf{W}_1 = Q_{\mathbf{U}_1}^{-1/2}\bar{\boldsymbol{\Omega}}^{-1/2}\mathbf{U}_1$. Note that, by the first result, $(\mathbf{U}_0^* | Q_{\mathbf{U}_1} = v) \sim \mathcal{T}_m(\mathbf{0}, \alpha_{v,v}\boldsymbol{\Upsilon}, \nu+d)$. Therefore, $(\mathbf{T}_* | Q_{\mathbf{U}_1} = v) \sim \mathcal{T}_m(\mathbf{0}, v^{-1}\alpha_{v,v}\boldsymbol{\Upsilon}, \nu+d)$ and so $\Pr(\mathbf{U}_0^* < \boldsymbol{\Lambda}\mathbf{U}_1 + \boldsymbol{\tau} | Q_{\mathbf{U}_1} = v) = \text{E}_{\mathbf{W}_1}\{T_m(\bar{\boldsymbol{\Lambda}}\mathbf{W}_1 + \boldsymbol{\tau}_v; v^{-1}\alpha_{v,v}\boldsymbol{\Upsilon}, \nu+d)\}$. Consequently,

$$f_{Q_{\mathbf{Y}}}(v) = f_{Q_{\mathbf{U}_1}}(v)\frac{\text{E}_{\mathbf{W}_1}\{T_m(\bar{\boldsymbol{\Lambda}}\mathbf{W}_1 + \boldsymbol{\tau}_v; v^{-1}\alpha_{v,v}\boldsymbol{\Upsilon}, \nu+d)\}}{T_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}}, \nu)}.$$

In the pdf above, $Q_{\mathbf{U}_1}$ does not have a closed form because, even though the square of a t random variable follows an F-distribution, the sum of F random variables does not yield a random variable with an explicit pdf. Moreover, because $\text{E}_{\mathbf{W}_1}\{T_m(\bar{\boldsymbol{\Lambda}}\mathbf{W}_1 + \boldsymbol{\tau}_v; v^{-1}\alpha_{v,v}\boldsymbol{\Upsilon}, \nu+d)\}$ and $T_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}}, \nu)$ must be computed numerically, the computation of $f_{Q_{\mathbf{Y}}}(v)$ has to rely on numerical methods.

Another interesting point is that if $\bar{\boldsymbol{\Lambda}} = \mathbf{0}$ and $\boldsymbol{\tau}_v = \mathbf{0}$, hence $\boldsymbol{\Delta} = \mathbf{0}$ and $\boldsymbol{\tau} = \mathbf{0}$, then $\text{E}_{\mathbf{W}_1}\{T_m(\bar{\boldsymbol{\Lambda}}\mathbf{W}_1 + \boldsymbol{\tau}_v; v^{-1}\alpha_{v,v}\boldsymbol{\Upsilon}, \nu+d)\} = T_m(\mathbf{0}; v^{-1}\alpha_{v,v}\bar{\boldsymbol{\Gamma}}, \nu+d) = T_m(\mathbf{0}; \alpha_{v,v}\bar{\boldsymbol{\Gamma}}, \nu+d) = \Pr(\mathbf{U}_0^* < \mathbf{0} | \mathbf{U}_1)$. Furthermore, according to Fang et al. [19], if the marginals of the joint

elliptical distribution are uncorrelated then $\Pr(\mathbf{U}_0^* < \mathbf{0} | \mathbf{U}_1) = \Pr(\mathbf{U}_0^* < \mathbf{0}) = \Phi_m(\mathbf{0}; \mathbf{0}, \bar{\mathbf{\Gamma}})$. Therefore, $f_{Q_\mathbf{Y}}(v) = f_{Q_{\mathbf{U}_1}}(v)$. However, when $\mathbf{\Delta} = \mathbf{0}$ and $\tau = \mathbf{0}$, the random vector $\mathbf{Y}$ has a multivariate t distribution and, therefore, is no longer skewed.

This result can also be directly obtained from the $f_{Q_\mathbf{Y}}(v)$ in Proposition 10. In particular, $\Pr(\mathbf{U}_0^* < \mathbf{\Lambda}\mathbf{U}_1 + \tau | Q_{\mathbf{U}_1} = v) = \Pr(\mathbf{U}_0^* < \mathbf{\Lambda}\mathbf{U}_1 + \tau)$ if $\mathrm{Cov}(\mathbf{U}_1, \mathbf{U}_0^* - \mathbf{\Lambda}\mathbf{U}_1) = \mathbf{0}$ and $\tau = \mathbf{0}$. We can observe that $\mathrm{Cov}(\mathbf{U}_1, \mathbf{U}_0^* - \mathbf{\Lambda}\mathbf{U}_1) = \mathrm{Cov}(\mathbf{U}_1, \mathbf{U}_0^*) - \mathrm{Cov}(\mathbf{U}_1, \mathbf{\Lambda}\mathbf{U}_1) = -\mathbf{\Delta} = \mathbf{0}$.

In addition, when $\tau = \mathbf{0}$ and $m \geq 2$, another way to study the distribution of $Q_\mathbf{Z} = \mathbf{Z}^\top \bar{\mathbf{\Omega}}^{-1} \mathbf{Z}$, where $\mathbf{Z} = (\mathbf{U}_1 | \mathbf{U}_0 > \mathbf{0})$, is to use the representation $Q_\mathbf{Z} = V^{-1} Q_{\mathbf{Z}_0}$, where V is independent of $Q_{\mathbf{Z}_0} = \mathbf{Z}_0^\top \bar{\mathbf{\Omega}}^{-1} \mathbf{Z}_0$, with $\mathbf{Z}_0 = (\mathbf{W}_1 | \mathbf{W}_0 > \mathbf{0})$ and $(\mathbf{W}_0, \mathbf{W}_1)$ being the respective (centered) normal variables. Thus, since $\bar{\mathbf{W}}_1 = \mathbf{W}_1 - \mathbf{\Lambda}\mathbf{W}_0$, where $\mathbf{\Lambda} = \mathbf{\Delta}\bar{\mathbf{\Gamma}}^{-1}$, is independent of $\mathbf{W}_0$, we have $Q_{\mathbf{Z}_0} = (\bar{\mathbf{W}}_1 + \mathbf{\Lambda}|\mathbf{W}_0|)^\top \bar{\mathbf{\Omega}}^{-1}(\bar{\mathbf{W}}_1 + \mathbf{\Lambda}|\mathbf{W}_0|)$, which has χ_d^2 distribution (and so $Q_\mathbf{Z}$ has a Fisher type of distribution for $m \geq 2$) iff $\mathbf{\Lambda} = \mathbf{\Delta}\bar{\mathbf{\Gamma}}^{-1} = \mathbf{0}$. When $m = 1$, we know that $Q_{Z_0} \sim \chi_1^2$ [14].

4. Conditional distributions

In this section, we show that the conditionals of a SUT distribution still follow a SUT distribution, and that a form of conditioning allows to change the latent dimension.

4.1. Conditional distribution

Proposition 11. Let $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top \sim \mathcal{SUT}_{d,m}(\xi, \mathbf{\Omega}, \mathbf{\Delta}, \tau, \bar{\mathbf{\Gamma}}, v)$ be partitioned as in (16). Then:

$$(\mathbf{Y}_2 | \mathbf{Y}_1 = \mathbf{y}_1) \sim \mathcal{SUT}_{d_2,m}(\xi_{2\cdot 1}, \alpha_{v,Q_{\mathbf{y}_1}} \mathbf{\Omega}_{2\cdot 1}, \mathbf{\Delta}_{2\cdot 1}, \alpha_{v,Q_{\mathbf{y}_1}}^{-1/2} \tau_{2\cdot 1}, \bar{\mathbf{\Gamma}}_{2\cdot 1}, v + d_1), \quad \mathbf{y}_1 \in \mathbb{R}^{d_1},$$

where $\xi_{2\cdot 1} = \xi_2 + \mathbf{\Omega}_{21}\mathbf{\Omega}_{11}^{-1}(\mathbf{y}_1 - \xi_1)$, $\alpha_{v,Q_{\mathbf{y}_1}} = \{v + Q_{\mathbf{y}_1}\}/\{v + d_1\}$, $Q_{\mathbf{y}_1} = (\mathbf{y}_1 - \xi_1)^\top \mathbf{\Omega}_{11}^{-1}(\mathbf{y}_1 - \xi_1)$, $\mathbf{\Omega}_{2\cdot 1} = \mathbf{\Omega}_{22} - \mathbf{\Omega}_{21}\mathbf{\Omega}_{11}^{-1}\mathbf{\Omega}_{12}$, $\mathbf{\Delta}_{2\cdot 1} = \omega_{2\cdot 1}^{-1}(\omega_2 \mathbf{\Delta}_2 - \mathbf{\Omega}_{21}\mathbf{\Omega}_{11}^{-1}\omega_1 \mathbf{\Delta}_1)\gamma_{2\cdot 1}^{-1}$, $\omega_{2\cdot 1} = \mathrm{diag}(\mathbf{\Omega}_{2\cdot 1})^{1/2}$, $\tau_{2\cdot 1} = \gamma_{2\cdot 1}^{-1}\{\tau + \mathbf{\Delta}_1^\top \bar{\mathbf{\Omega}}_{11}^{-1}\omega_1^{-1}(\mathbf{y}_1 - \xi_1)\}$, $\bar{\mathbf{\Gamma}}_{2\cdot 1} = \gamma_{2\cdot 1}^{-1}\mathbf{\Gamma}_{2\cdot 1}\gamma_{2\cdot 1}^{-1}$, $\mathbf{\Gamma}_{2\cdot 1} = \bar{\mathbf{\Gamma}} - \mathbf{\Delta}_1^\top \bar{\mathbf{\Omega}}_{11}^{-1}\mathbf{\Delta}_1$ and $\gamma_{2\cdot 1} = \mathrm{diag}(\mathbf{\Gamma}_{2\cdot 1})^{1/2}$.

Proof. The proof follows the same reasoning as the one given for Proposition 3.2 in Arellano-Valle and Genton [11] in the particular case where the Student's t density generator function is considered and the parameterization in Definition 1 is used. $\square$

It is worth noting that the degrees-of-freedom is increased after conditioning by the dimension of the conditioning vector, hence making the resulting SUT distribution closer to the SUN.

4.2. Changing latent dimensions

We investigate ways to change the dimension of the latent variables.

Proposition 12. Let $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top \sim \mathcal{SUT}_{d,m}(\xi, \mathbf{\Omega}, \mathbf{\Delta}, \tau, \bar{\mathbf{\Gamma}}, v)$ with parameters partitioned as in (16). Then,

$$(\mathbf{Y}_2 | \mathbf{Y}_1 > \mathbf{0}) \sim \mathcal{SUT}_{d_2, d_1+m}(\xi_2, \mathbf{\Omega}_{22}, \mathbf{\Delta}_{2\neg 1}, \tau_{2\neg 1}, \bar{\mathbf{\Gamma}}_{2\neg 1}, v),$$

where

$$\mathbf{\Delta}_{2\neg 1} = \begin{pmatrix} \mathbf{\Delta}_2 & \bar{\mathbf{\Omega}}_{21} \end{pmatrix}, \quad \bar{\mathbf{\Gamma}}_{2\neg 1} = \begin{pmatrix} \bar{\mathbf{\Gamma}} & \mathbf{\Delta}_1^\top \\ \mathbf{\Delta}_1 & \bar{\mathbf{\Omega}}_{11} \end{pmatrix}, \quad \tau_{2\neg 1} = \begin{pmatrix} \tau \\ \omega_1^{-1}\xi_1 \end{pmatrix}.$$

Proof. Firstly, note that, from Proposition 7 it follows that $\mathbf{Y}_2 \sim \mathcal{SUT}_{d_2,m}(\xi_2, \mathbf{\Omega}_{22}, \mathbf{\Delta}_2, \tau, \bar{\mathbf{\Gamma}}, v)$ and $-\mathbf{Y}_1 \sim \mathcal{SUT}_{d_1,m}(-\xi_1, \mathbf{\Omega}_{11}, -\mathbf{\Delta}_1, \tau, \bar{\mathbf{\Gamma}}, v)$. Secondly, from Proposition 11 (exchanging the subscripts 1 and 2) we obtain that:

$$(-\mathbf{Y}_1 | \mathbf{Y}_2 = \mathbf{y}_2) \sim \mathcal{SUT}_{d_1,m}(-\xi_{1\cdot 2}, \alpha_{v,Q_{\mathbf{y}_2}} \mathbf{\Omega}_{1\cdot 2}, -\mathbf{\Delta}_{1\cdot 2}, \alpha_{v,Q_{\mathbf{y}_2}}^{-1/2} \tau_{1\cdot 2}, \bar{\mathbf{\Gamma}}_{1\cdot 2}, v + d_2).$$

Hence:

$$f_{\mathbf{Y}_2}(\mathbf{y}_2) = t_{d_2}(\mathbf{y}_2; \xi_2, \mathbf{\Omega}_{22}, v) \frac{T_m[\alpha_{v,Q_{\mathbf{y}_2}}^{-1/2}\{\tau + \mathbf{\Delta}_2^\top \bar{\mathbf{\Omega}}_{22}^{-1}\omega_2^{-1}(\mathbf{y}_2 - \xi_2)\}; \bar{\mathbf{\Gamma}} - \mathbf{\Delta}_2^\top \bar{\mathbf{\Omega}}_{22}^{-1}\mathbf{\Delta}_2, v + d_2]}{T_m(\tau; \bar{\mathbf{\Gamma}}, v)},$$

$$\Pr(-\mathbf{Y}_1 \leq \mathbf{0} | \mathbf{Y}_2 = \mathbf{y}_2) = \frac{T_{d_1+m}\left\{ \begin{pmatrix} \alpha_{v,Q_{\mathbf{y}_2}}^{-1/2}\tau_{1\cdot 2} \\ \xi_{1\cdot 2} \end{pmatrix}; \begin{pmatrix} \bar{\mathbf{\Gamma}}_{1\cdot 2} & \mathbf{\Delta}_{1\cdot 2}^\top \alpha_{v,Q_{\mathbf{y}_2}}^{1/2}\omega_{1\cdot 2} \\ \omega_{1\cdot 2}\alpha_{v,Q_{\mathbf{y}_2}}^{1/2}\mathbf{\Delta}_{1\cdot 2} & \alpha_{v,Q_{\mathbf{y}_2}}\mathbf{\Omega}_{1\cdot 2} \end{pmatrix}, v + d_2 \right\}}{T_m(\alpha_{v,Q_{\mathbf{y}_2}}^{-1/2}\tau_{1\cdot 2}; \bar{\mathbf{\Gamma}}_{1\cdot 2}, v + d_2)},$$

$$\Pr(-\mathbf{Y}_1 \leq \mathbf{0}) = \frac{T_{d_1+m}\left\{ \begin{pmatrix} \tau \\ \xi_1 \end{pmatrix}; \begin{pmatrix} \bar{\mathbf{\Gamma}} & \mathbf{\Delta}_1^\top \omega_1 \\ \omega_1 \mathbf{\Delta}_1 & \mathbf{\Omega}_{11} \end{pmatrix}, v \right\}}{T_m(\tau; \bar{\mathbf{\Gamma}}, v)}.$$

Then, after some simplifications and a few algebraic manipulations, the pdf of $(\mathbf{Y}_2|\mathbf{Y}_1 > \mathbf{0})$ can be computed at $\mathbf{y}_2 \in \mathbb{R}^{d_2}$ as follows:

$$f_{\mathbf{Y}_2|\mathbf{Y}_1>\mathbf{0}}(\mathbf{y}_2) = f_{\mathbf{Y}_2}(\mathbf{y}_2)\frac{\Pr(-\mathbf{Y}_1 \le \mathbf{0}|\mathbf{Y}_2 = \mathbf{y}_2)}{\Pr(-\mathbf{Y}_1 \le \mathbf{0})}$$

$$= t_{d_2}(\mathbf{y}_2; \boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, \nu)\frac{T_{d_1+m}\left\{\begin{pmatrix} \alpha_{\nu,Q_{\mathbf{y}_2}}^{-1/2}\boldsymbol{\tau}_{1\cdot 2} \\ \boldsymbol{\xi}_{1\cdot 2} \end{pmatrix}; \begin{pmatrix} \bar{\boldsymbol{\Gamma}}_{1\cdot 2} & \boldsymbol{\Delta}_{1\cdot 2}^\top \alpha_{\nu,Q_{\mathbf{y}_2}}^{1/2}\boldsymbol{\omega}_{1\cdot 2} \\ \boldsymbol{\omega}_{1\cdot 2}\alpha_{\nu,Q_{\mathbf{y}_2}}^{1/2}\boldsymbol{\Delta}_{1\cdot 2} & \alpha_{\nu,Q_{\mathbf{y}_2}}\boldsymbol{\Omega}_{1\cdot 2} \end{pmatrix}, \nu+d_2\right\}}{T_{d_1+m}\left\{\begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{\xi}_1 \end{pmatrix}; \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & \boldsymbol{\Delta}_1^\top\boldsymbol{\omega}_1 \\ \boldsymbol{\omega}_1\boldsymbol{\Delta}_1 & \boldsymbol{\Omega}_{11} \end{pmatrix}, \nu\right\}}$$

$$= t_{d_2}(\mathbf{y}_2; \boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, \nu)\frac{T_{d_1+m}\left[\alpha_{\nu,Q_{\mathbf{y}_2}}^{-1/2}\{\boldsymbol{\tau}_{2\neg 1} + \boldsymbol{\Delta}_{2\neg 1}^\top\bar{\boldsymbol{\Omega}}_{22}^{-1}\boldsymbol{\omega}_2^{-1}(\mathbf{y}_2-\boldsymbol{\xi}_2)\}; \bar{\boldsymbol{\Gamma}}_{2\neg 1} - \boldsymbol{\Delta}_{2\neg 1}^\top\bar{\boldsymbol{\Omega}}_{22}^{-1}\boldsymbol{\Delta}_{2\neg 1}, \nu+d_2\right]}{T_{d_1+m}\left(\boldsymbol{\tau}_{2\neg 1}; \bar{\boldsymbol{\Gamma}}_{2\neg 1}, \nu\right)},$$

which, according to (6), is the pdf of a $\mathcal{SUT}_{d_2,d_1+m}(\boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, \boldsymbol{\Delta}_{2\neg 1}, \boldsymbol{\tau}_{2\neg 1}, \bar{\boldsymbol{\Gamma}}_{2\neg 1}, \nu)$ distribution. $\square$

Next, we explore the possibility of having redundant latent dimensions.

Lemma 2. *We have*

$$T_{d+m}(\mathbf{y}_* - \boldsymbol{\xi}_*; \boldsymbol{\Omega}_*, \nu) = T_d(\mathbf{y} - \boldsymbol{\xi}; \boldsymbol{\Omega}, \nu)T_m(\mathbf{0}; \bar{\boldsymbol{\Gamma}}, \nu), \tag{19}$$

where $\mathbf{y}_* = (\mathbf{0}^\top, \mathbf{y}^\top)^\top$, $\boldsymbol{\xi}_* = (\mathbf{0}^\top, \boldsymbol{\xi}^\top)^\top$, *and* $\boldsymbol{\Omega}_* = \text{diag}(\bar{\boldsymbol{\Gamma}}, \boldsymbol{\Omega})$.

Proof. On the one hand, if $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \mathbf{0}, \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)$, then $\mathbf{Y}$ is elliptically distributed. So, it holds that $\Pr(\mathbf{Y} \le \mathbf{y}) = T_d(\mathbf{y}-\boldsymbol{\xi}; \boldsymbol{\Omega}, \nu)$. On the other hand, from (7), we have $\Pr(\mathbf{Y} \le \mathbf{y}) = T_{d+m}(\mathbf{y}_* - \boldsymbol{\xi}_*; \boldsymbol{\Omega}_*, \nu)/T_m(\mathbf{0}; \bar{\boldsymbol{\Gamma}}, \nu)$. Thus, (19) follows by equating the two previous identities. $\square$

Remark 1. Eq. (19) holds more generally. Indeed, let $(\mathbf{X}_0, \mathbf{X}_1) = R\mathbf{U}^{(m+d)} = (R_0\mathbf{U}^{(m)}, R_1\mathbf{U}^{(d)})$ be an $(m+d)$-dimensional spherical random vector, where $R = \sqrt{R_0^2 + R_1^2}$ (the radial variable) and $\mathbf{U}^{(m+d)}$ (the uniform vector on the unit sphere) are independent; also (R_0, R_1), $\mathbf{U}^{(m)}$ and $\mathbf{U}^{(d)}$ are independent. Hence:

$$\Pr(\mathbf{X}_0 \le \mathbf{0}, \mathbf{X}_1 \le \mathbf{y}) = \Pr(R_0\mathbf{U}^{(m)} \le \mathbf{0}, R_1\mathbf{U}^{(d)} \le \mathbf{y}) = \Pr(\mathbf{U}^{(m)} \le \mathbf{0}, R_1\mathbf{U}^{(d)} \le \mathbf{y}) = \Pr(\mathbf{U}^{(m)} \le \mathbf{0})\Pr(R_1\mathbf{U}^{(d)} \le \mathbf{y})$$

$$= \Pr(R_0\mathbf{U}^{(m)} \le \mathbf{0})\Pr(R_1\mathbf{U}^{(d)} \le \mathbf{y}) = \Pr(\mathbf{X}_0 \le \mathbf{0})\Pr(\mathbf{X}_1 \le \mathbf{y}).$$

It then follows for the uncorrelated elliptically contoured case defined by $\mathbf{Y}_0 = \boldsymbol{\Omega}_0^{1/2}\mathbf{X}_0$ and $\mathbf{Y}_1 = \boldsymbol{\Omega}_1^{1/2}\mathbf{X}_1$, i.e., $(\mathbf{Y}_0^\top, \mathbf{Y}_1^\top)^\top \sim \mathcal{EC}_{m+d}(\mathbf{0}, \text{diag}(\boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1), h^{(m+d)})$ where $h^{(m+d)}$ is the density generator, that also $\Pr(\mathbf{Y}_0 \le \mathbf{0}, \mathbf{Y}_1 \le \mathbf{y}) = \Pr(\mathbf{Y}_0 \le \mathbf{0})\Pr(\mathbf{Y}_1 \le \mathbf{y})$.

The result in (19) can be used to show the following.

Proposition 13. *If* $\mathbf{Y} \sim \mathcal{SUT}_{d,m_1+m_2}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$ *with* $\boldsymbol{\Delta} = (\mathbf{0}, \boldsymbol{\Delta}_2)$, $\boldsymbol{\tau} = (\mathbf{0}^\top, \boldsymbol{\tau}_2^\top)^\top$, *and* $\bar{\boldsymbol{\Gamma}} = \text{diag}(\bar{\boldsymbol{\Gamma}}_{11}, \bar{\boldsymbol{\Gamma}}_{22})$, *then* $\mathbf{Y} \sim \mathcal{SUT}_{d,m_2}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}_2, \boldsymbol{\tau}_2, \bar{\boldsymbol{\Gamma}}_{22}, \nu)$.

Proof. We have for the cdf of $\mathbf{Y}$:

$$F_{\mathbf{Y}}(\mathbf{y}) = \frac{T_{d+m_1+m_2}(\mathbf{y}^* - \boldsymbol{\xi}^*; \boldsymbol{\Omega}^*, \nu)}{T_{m_1+m_2}\{(\mathbf{0}^\top, \boldsymbol{\tau}_2^\top)^\top; \bar{\boldsymbol{\Gamma}}, \nu\}}, \quad \mathbf{y}^* = (\mathbf{0}^\top, \boldsymbol{\tau}_2^\top, \mathbf{y}^\top)^\top, \quad \boldsymbol{\xi}^* = (\mathbf{0}^\top, \mathbf{0}^\top, \boldsymbol{\xi}^\top)^\top, \quad \boldsymbol{\Omega}^* = \begin{pmatrix} \bar{\boldsymbol{\Gamma}}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \bar{\boldsymbol{\Gamma}}_{22} & -\boldsymbol{\Delta}_2^\top\boldsymbol{\omega} \\ \mathbf{0} & -\boldsymbol{\omega}\boldsymbol{\Delta}_2 & \boldsymbol{\Omega} \end{pmatrix},$$

$$= \frac{T_{m_1}(\mathbf{0}; \bar{\boldsymbol{\Gamma}}_{11}, \nu)\, T_{d+m_2}(\mathbf{y}^- - \boldsymbol{\xi}^-; \boldsymbol{\Omega}^-, \nu)}{T_{m_1}(\mathbf{0}; \bar{\boldsymbol{\Gamma}}_{11}, \nu)\, T_{m_2}(\boldsymbol{\tau}_2; \bar{\boldsymbol{\Gamma}}_{22}, \nu)}, \quad \mathbf{y}^- = (\boldsymbol{\tau}_2^\top, \mathbf{y}^\top)^\top, \quad \boldsymbol{\xi}^- = (\mathbf{0}^\top, \boldsymbol{\xi}^\top)^\top, \quad \boldsymbol{\Omega}^- = \begin{pmatrix} \bar{\boldsymbol{\Gamma}}_{22} & -\boldsymbol{\Delta}_2^\top\boldsymbol{\omega} \\ -\boldsymbol{\omega}\boldsymbol{\Delta}_2 & \boldsymbol{\Omega} \end{pmatrix},$$

$$= \frac{T_{d+m_2}(\mathbf{y}^- - \boldsymbol{\xi}^-; \boldsymbol{\Omega}^-, \nu)}{T_{m_2}(\boldsymbol{\tau}_2; \bar{\boldsymbol{\Gamma}}_{22}, \nu)},$$

which is the cdf of a $\mathcal{SUT}_{d,m_2}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}_2, \boldsymbol{\tau}_2, \bar{\boldsymbol{\Gamma}}_{22}, \nu)$ distribution. $\square$

Consequently, following from this property, if the random vector $\mathbf{Y} \sim \mathcal{SUT}_{d,m_1+\cdots+m_n}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$, where $m = m_1+\cdots+m_n$, $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_n)$, $\boldsymbol{\tau} = (\boldsymbol{\tau}_1^\top, \ldots, \boldsymbol{\tau}_n^\top)^\top$, and $\bar{\boldsymbol{\Gamma}} = \text{diag}(\bar{\boldsymbol{\Gamma}}_1, \ldots, \bar{\boldsymbol{\Gamma}}_n)$, we can construct a latent dimension reduction matrix $\mathbf{R}_i$, $i = 2, \ldots, n$, for the dimension m_i by solving the equation $\mathbf{R}_i\boldsymbol{\Delta}_i = \mathbf{0}$, provided that $\boldsymbol{\tau}_i = \mathbf{0}$. As a result, $\mathbf{R}_i\mathbf{Y} \sim \mathcal{SUT}_{d,m-m_i}(\mathbf{R}_i\boldsymbol{\xi}, \mathbf{R}_i\boldsymbol{\Omega}\mathbf{R}_i^\top, \mathbf{R}_i\boldsymbol{\Delta}_{-i}, \boldsymbol{\tau}_{-i}, \bar{\boldsymbol{\Gamma}}_{-i}, \nu)$ with the negative index indicating the removal of the ith element. Note that i can be any number between 1 and n because the SUT distribution is non-identifiable with respect to the order of its latent variables [34].
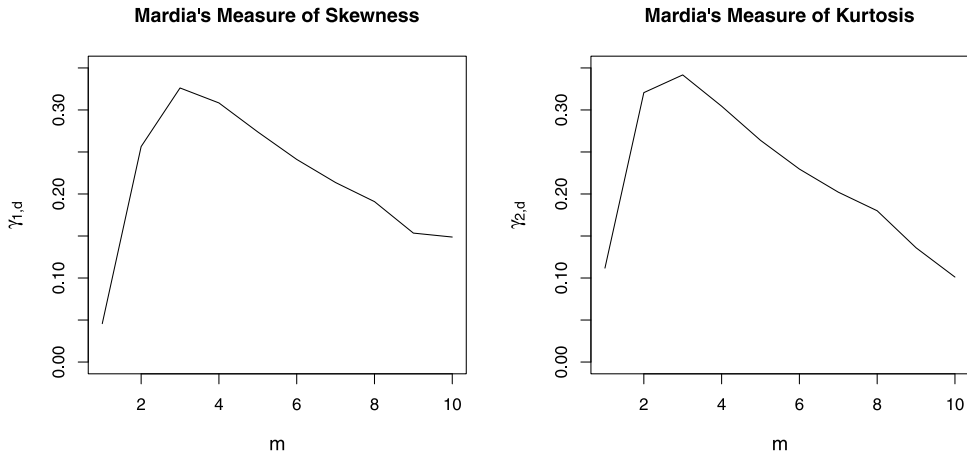
Mardia's Measure of Skewness **Mardia's Measure of Kurtosis**

Fig. 2. Mardia's measures of multivariate skewness $\gamma_{1,d}$ and kurtosis $\gamma_{2,d}$ against latent dimension m with $(\boldsymbol{\Omega}, \boldsymbol{\Delta}, \bar{\boldsymbol{\Gamma}})$ specified to skew the distribution in the direction $(1, 1)^{\top}$. Here $\boldsymbol{\xi}$ and $\boldsymbol{\tau}$ are set as $\mathbf{0}$, and $\nu = 5$.

5. Mardia's measures of multivariate skewness and kurtosis

5.1. Computation

Mardia's measures of multivariate skewness and kurtosis [30] can also be computed exactly. Following from the previous setting, we consider $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$ and denote $\mathrm{Var}(\mathbf{Y}) = \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\top}$ and $\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu} = \boldsymbol{\xi} + \mathrm{E}(\mathbf{U}_1^* | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0}) = \boldsymbol{\xi} + \boldsymbol{\mu}_0$, where $\mathbf{U}_1^* = \omega \mathbf{U}_1$. We first need to standardize the random vector $\mathbf{Y}$ to compute the skewness and kurtosis measure. In particular, we let $\mathbf{Z} = \mathbf{L}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \sim \mathcal{SUT}_{d,m}(-\mathbf{L}^{-1}\tilde{\boldsymbol{\mu}}_0, \boldsymbol{\Omega}_{\mathbf{L}}, \boldsymbol{\Delta}_{\mathbf{L}}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$, where $\boldsymbol{\Omega}_{\mathbf{L}} = \mathbf{L}^{-1}\boldsymbol{\Omega}(\mathbf{L}^{-1})^{\top}$, $\boldsymbol{\Delta}_{\mathbf{L}} = \omega_{\mathbf{L}}^{-1}\mathbf{L}^{-1}\omega\boldsymbol{\Delta}$ and $\omega_{\mathbf{L}} = \mathrm{diag}(\boldsymbol{\Omega}_{\mathbf{L}})^{1/2}$. As a result, we have that $\mathrm{E}(\mathbf{Z}) = \mathbf{0}$ and $\mathrm{Var}(\mathbf{Z}) = \mathbf{I}_d$.

According to Kollo and Srivastava [28], the Mardia measures of multivariate skewness and kurtosis of the standardized random vector $\mathbf{Z}$ can then be computed using trace operation on the third and fourth moments:

$$\beta_{1,d} = \mathrm{tr}\{\mu_3(\mathbf{Z})^{\top}\mu_3(\mathbf{Z})\} = \mathrm{vec}\{\mu_3(\mathbf{Z})\}^{\top}\mathrm{vec}\{\mu_3(\mathbf{Z})\},$$

$$\beta_{2,d} = \mathrm{tr}\{\mu_4(\mathbf{Z})\}.$$

Here $\mu_3(\mathbf{Z})$ and $\mu_4(\mathbf{Z})$ can be computed using the convolution-based method described in Section 2.4. One point to notice is the displacement in the kurtosis measure when dealing with high dimensions. The non-shifted measure $\gamma_{2,d}$ can be adjusted by equation (2.9) in Mardia [31]. Overall, $\gamma_{1,d}$ and $\gamma_{2,d}$ are invariant with respect to location and scale. Consequently, it is sufficient to assume that $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$ as indicated in Arellano-Valle and Azzalini [5] for the computations.

5.2. Visualization

To demonstrate the effectiveness of the two measures, we visualize $\beta_{1,d} = \gamma_{1,d}$ and $\gamma_{2,d}$ with increasing latent dimensions m in Fig. 2. We impose skewness to the distribution along the direction $(1, 1)^{\top}$ to see the variations. Fig. 2 indicates that the two measures increase first and then decrease. The reason behind such a behavior is the asymptotic linear dependence in increasing latent dimensions. We articulate the rationales in the following proposition.

Proposition 14. *Let* $\mathbf{Y} = (Y_1, \dots, Y_d)^{\top} \sim \mathcal{SUT}_{d,m}(\mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$, *where* $\nu > 2$ *and* $\boldsymbol{\Delta} = (\delta_1, \dots, \delta_m)$. *Then, if* $m \to \infty$ *with an infinite number of* $\delta_k \neq \mathbf{0} \in \mathbb{R}^d$, $k \in \{1, \dots, m\}$, *the correlation between* Y_i *and* Y_j, $\rho_{i,j} \to \pm 1$ *or* 0 *for* $i \neq j$ *and* $i, j \in \{1, \dots, d\}$, *indicating that* Y_i *and* Y_j *exhibit asymptotic linearity or uncorrelation as* m *increases.*

Proof. To ease the proof, we need to introduce a change of parameterization. In particular, $\boldsymbol{\Omega} = \boldsymbol{\Psi} + \mathbf{H}\bar{\boldsymbol{\Gamma}}\mathbf{H}^{\top}$ and $\omega\boldsymbol{\Delta} = \mathbf{H}\bar{\boldsymbol{\Gamma}}$, where $\boldsymbol{\Psi}$ is also a covariance matrix. By Proposition 2, $\mathrm{Var}(\mathbf{Y}) = \omega\{\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\mathrm{Var}(\mathbf{U}_*)\bar{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Delta}^{\top} + \eta(Q_{\mathbf{U}_*})(\bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}\boldsymbol{\Delta}^{\top})\}\omega$. By re-arranging the terms and plugging in the restrictions, we can have the following quantity:

$$\mathrm{Var}(\mathbf{Y}) = \mathbf{H}\mathrm{Var}(\mathbf{U}_*)\mathbf{H}^{\top} + \eta(Q_{\mathbf{U}_*})\boldsymbol{\Psi} = \mathbf{H}\mathbf{L}\mathbf{L}^{\top}\mathbf{H}^{\top} + \eta(Q_{\mathbf{U}_*})\boldsymbol{\Psi}, \quad \mathrm{Var}(\mathbf{U}_*) = \mathbf{L}\mathbf{L}^{\top}$$

$$= \mathbf{H}_{\mathbf{L}}\mathbf{H}_{\mathbf{L}}^{\top} + \eta(Q_{\mathbf{U}_*})\boldsymbol{\Psi} = \sum_{k=1}^{m}\mathbf{h}_{\mathbf{L}k}\mathbf{h}_{\mathbf{L}k}^{\top} + \tilde{\boldsymbol{\Psi}}, \quad \mathbf{H}_{\mathbf{L}} = \mathbf{H}\mathbf{L}, \tilde{\boldsymbol{\Psi}} = \eta(Q_{\mathbf{U}_*})\boldsymbol{\Psi}.$$

Now we have that:

$$\rho_{i,j} = \frac{\tilde{\Psi}_{i,j} + \sum_{k=1}^{m} h_{k,i} h_{k,j}}{\sqrt{\tilde{\Psi}_{i,i} + \sum_{k=1}^{m} h_{k,i}^2} \sqrt{\tilde{\Psi}_{j,j} + \sum_{k=1}^{m} h_{k,j}^2}} = \frac{\tilde{\Psi}_{i,j} + \sum_{k=1}^{m} h_{k,i} h_{k,j}}{\sqrt{\tilde{\Psi}_{i,i} \tilde{\Psi}_{j,j} + \sum_{k=1}^{m} h_{k,i}^2 \tilde{\Psi}_{j,j} + \sum_{k=1}^{m} h_{k,j}^2 \tilde{\Psi}_{i,i} + \sum_{k=1}^{m} \sum_{w=1}^{m} h_{k,i}^2 h_{w,j}^2}},$$

where $h_{k,i}$ and $h_{k,j}$ denote the respective ith and jth component of $\mathbf{h}_{\mathbf{L}k}$. In the denominator, we claim that $\sum_{k=1}^{m} \sum_{w=1}^{m} h_{k,i}^2 h_{w,j}^2$ is the dominating term as $m \to \infty$. Here are the details:

$$\lim_{m \to \infty} \frac{\sum_{k=1}^{m} h_{k,i}^2}{\sum_{k=1}^{m} \sum_{w=1}^{m} h_{k,i}^2 h_{w,j}^2} \leq \lim_{m \to \infty} \frac{m \cdot \max(\{h_{k,i}\}_{i=1,\ldots,m})}{m^2 \cdot \min(\{h_{k,i}^2 h_{w,j}^2\}_{k,w=1,\ldots,m})} = 0,$$

$$\lim_{m \to \infty} \frac{\sum_{k=1}^{m} h_{k,i}^2}{\sum_{k=1}^{m} \sum_{w=1}^{m} h_{k,i}^2 h_{w,j}^2} = 0, \quad \sum_{k=1}^{m} h_{k,i}^2 \geq 0, \quad \sum_{k=1}^{m} \sum_{w=1}^{m} h_{k,i}^2 h_{w,j}^2 \geq 0.$$

The same result for comparing with $\sum_{k=1}^{m} h_{k,j}^2$ can be replicated by replacing i with j in the above derivation. Therefore, we only need to focus on the increments (decrements) $\sum_{k=1}^{m} h_{k,i} h_{k,j}$ and $\sqrt{\sum_{k=1}^{m} \sum_{w=1}^{m} h_{k,i}^2 h_{w,j}^2}$. By the well-known Cauchy–Schwarz inequality, we have $-1 \leq \sum_{k=1}^{m} h_{k,i} h_{k,j} / \sqrt{\sum_{k=1}^{m} \sum_{w=1}^{m} h_{k,i}^2 h_{w,j}^2} \leq 1$ and the ratio is equal to ± 1 only if $h_{k,i} = \beta_{i,j} h_{k,j}$ with $\beta_{i,j} \in \mathbb{R}$, indicating that the $\mathbf{h}_{\mathbf{L}k}$s are pointing to either the same direction or infinitely many times along and against the same unit vector. Note that this situation excludes the directions of the vectors in the planar or hyperplanar subspaces formulated by the main axes, which are special cases that we will explore later. Now, if the rate of the increments (decrements) are equal, then $\lim_{m \to \infty} \rho_{i,j} = \pm 1$. Otherwise, $\lim_{m \to \infty} \rho_{i,j} = 0$ because the increment (decrement) rate in the denominator is higher. This is the case when $\mathbf{h}_{\mathbf{L}k}$s point to infinitely many directions.

Now we assume without loss of generality that $h_{k,i} \neq 0, \forall i \in z \subset \{1, \ldots, d\}$, and $h_{k,j} = 0, \forall j \in \{1, \ldots, d\} - z, \forall k = 1, \ldots, m$. Then, all $\mathbf{h}_{\mathbf{L}k} \in \mathbb{R}^{|j|} \subset \mathbb{R}^d$, and:

$$\lim_{m \to \infty} \rho_{i,j} = \lim_{m \to \infty} \frac{\tilde{\Psi}_{i,j} + \sum_{k=1}^{m} h_{k,i} h_{k,j}}{\sqrt{\tilde{\Psi}_{i,i} + \sum_{k=1}^{m} h_{k,i}^2} \sqrt{\tilde{\Psi}_{j,j} + \sum_{k=1}^{m} h_{k,j}^2}} = \lim_{m \to \infty} \frac{\tilde{\Psi}_{i,j}}{\sqrt{\tilde{\Psi}_{i,i} + \sum_{k=1}^{m} h_{k,i}^2} \sqrt{\tilde{\Psi}_{j,j}}} = 0.$$

This result has a rather straightforward interpretation: we inflate the variability of Y_is to ∞ and leave it unchanged for the remaining Y_js. The Y_is can still exhibit asymptotic linearity or uncorrelation given the above-mentioned choices of $\mathbf{h}_{\mathbf{L}k-0} \in \mathbb{R}^{|j|}$, where -0 denotes the removal of the 0 terms. □

By Proposition 14, the SUT will show either asymptotic linearity or uncorrelation (symmetry) in the bivariate case as m increases, explaining the humped shapes of Mardia's measures in Fig. 2. Per the results, we advise against using excessively large latent dimensions for the SUT. The asymptotic linearity and uncorrelation in the latent dimensions also hold for the SUN distribution because it is a particular case of the SUT distribution.

Another noteworthy proposition we would like to make is the following.

Proposition 15. *Let $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\xi, \Omega, \Delta, \mathbf{0}, \bar{\Gamma}, \nu)$. If $\bar{\Gamma}$ is an equi-correlation matrix with $\bar{\Gamma}_{i,j} = \rho \approx 1, \forall i \neq j$, then the corresponding $\lim_{\rho \to 1} \mathbf{Y} \equiv \mathbf{Y}_{ST} \sim \mathcal{ST}_d$.*

Proof. By the convolution representation in Proposition 1, $\mathbf{Y} = \xi + \omega \left\{ \Delta \bar{\Gamma}^{-1} \mathbf{U}_* + \sqrt{(\nu + Q_{\mathbf{U}_*})/(\nu + m)} \cdot \mathbf{W}_* \right\}$ with $\mathbf{U}_* = (\mathbf{U}_0 | \mathbf{U}_0 + \tau > \mathbf{0})$, where $\mathbf{U}_0 \sim \mathcal{T}_m(\mathbf{0}, \bar{\Gamma}, \nu)$. Now, if $\rho \approx 1$, then $\mathbf{U}_{*i} \approx \mathbf{U}_{*j}, \forall i \neq j$, where $i, j \in \{1, \ldots, m\}$, $\Delta \bar{\Gamma}^{-1} \mathbf{U}_* \approx (\sum_{i=1}^{m} \mathbf{H}_i) \mathbf{U}_{*1}$, where $\mathbf{H} = (\mathbf{H}_1, \ldots, \mathbf{H}_m) = \Delta \bar{\Gamma}^{-1}$. This is exactly the ST distribution, for which the direction of skewness is $\sum_{i=1}^{m} \mathbf{H}_i$. □

The same argument can be applied to the SUN and SN distributions. Hence, we recommend to use the SN or ST distributions directly in case of strong latent correlations.

6. Non-identifiability

Wang et al. [34] have demonstrated that the SUN distribution is non-identifiable subject to permutations, $\mathcal{P}(m) = \{\mathbf{P} \in \mathbb{R}^{m \times m} | \mathbf{P}\mathbf{P}^\top = \mathbf{P}^\top \mathbf{P} = \mathbf{I}_m$ and $\mathbf{P}\mathbf{1}_m = \mathbf{1}_m\}$, of its latent variables if $m > 1$. In particular, the random vector $\mathbf{X} \sim \mathcal{SUN}_{d,m}(\xi, \Omega, \Delta, \tau, \bar{\Gamma})$ is equal in distribution to the random vector $\mathbf{X}_p \sim \mathcal{SUN}_{d,m}(\xi, \Omega, \Delta_p, \tau_p, \bar{\Gamma}_p)$, where $\Delta_p = \Delta \mathbf{P}^\top$, $\tau_p = \mathbf{P}\tau$, and $\bar{\Gamma}_p = \mathbf{P}\bar{\Gamma}\mathbf{P}^\top$, $\mathbf{P} \in \mathcal{P}(m)$. Therefore, two sets of parameters yield identical probability densities for the same realization. Furthermore, Wang et al. [34] have shown that the non-identifiability also holds for the unified skew-elliptical (SUE) class and even more generally for selection distributions [6]. Consequently, the SUT is no exception.

Non-identifiability is problematic, especially for parameter inference, because the resulting optimization curve could possess multiple peaks and mislead the optimization algorithms to erroneous stopping values, rendering the SUT family non-applicable. One possible approach to address this issue includes ranking the components τ_i, $i = 1, \ldots, m$, with a strictly increasing (decreasing) order. However, this approach does not cover the case $\tau = \mathbf{0}$.

Some sub-models discussed in Wang et al. [34] can also be applied in the SUT case. For instance, it is possible to eradicate the ordering flexibility of Δ by imposing $\Delta = \delta \Omega^{1/2}$ or other similar relationships between Δ and Ω. Another path involves making $\bar{\Gamma}$ an equi-correlation matrix and Δ having identical entries. Therefore, here is a list of some identifiable SUT distributions:

(1) $\mathcal{SUT}_{d,m}(\xi, \Omega, \delta\mathbf{1}_m^\top, \tau\mathbf{1}_m, (1 - \rho)\mathbf{I}_m + \rho\mathbf{1}_m\mathbf{1}_m^\top, \nu)$, where $\tau = \tau\mathbf{1}_m$ ($\tau \in \mathbb{R}$), $\Delta = \delta\mathbf{1}_m^\top$ ($\delta \in \mathbb{R}^d$) and $\bar{\Gamma} = (1 - \rho)\mathbf{I}_m + \rho\mathbf{1}_m\mathbf{1}_m^\top$, with $\rho \in (-\frac{1}{m-1}, 1)$. Moreover, by Proposition 15, when $\rho \approx 1$, it is preferable to opt for the ST distribution;

(2) $\mathcal{SUT}_{d,m}(\xi, \Omega, \Delta, \alpha\mathbf{1}_m + \beta\mathbf{j}_m, \bar{\Gamma}, \nu)$, where $\tau = \alpha\mathbf{1}_m + \beta\mathbf{j}_m$ ($\alpha, \beta \in \mathbb{R}, \beta \neq 0$) and $\mathbf{j}_m = (1, 2, \ldots, m)^\top$;

(3) $\mathcal{SUT}_{d,d}(\xi, \omega^2\bar{\Omega}, \omega\delta(1 + \delta^2)^{-1/2}\bar{\Omega}, \mathbf{0}, \bar{\Omega}, \nu)$, where $\omega \in \mathbb{R}$ and $\delta \in \mathbb{R}$;

(4) $\mathcal{SUT}_{d,d}(\xi, \Omega, \delta\Omega^{1/2}, \mathbf{0}, \mathbf{I}_d, \nu)$, where $\delta \in \mathbb{R}$.

In addition, these sub-models can be combined to formulate various new identifiable cases.

7. Discussion

In this work, we comprehensively explored the properties of the SUT distribution. The SUT generalizes the ST distribution proposed in Azzalini and Capitanio [13] so that the latent variables can have specified mean τ and correlation $\bar{\Gamma}$. We derived stochastic representations and a SUN-scale mixture method to construct the SUT random variable. Moreover, we described numerous formal probabilistic properties, such as linear transformations, marginals, and conditionals, among many others. In addition, the SUT can also be viewed as a generalization of the SUN, which is retrieved by letting $\nu \to \infty$. We have provided possible solutions to the non-identifiability issue associated with the SUT, rendering the distribution applicable in practice.

Although the EM algorithm can provide inference for SN parameters (see, e.g., Abe et al. [1] and references therein), well-developed inference mechanisms for the SUN and the SUT distributions are challenging and currently lacking. Gupta and Aziz [23] applied the method of weighted moments on a particular case of the SUN with $\tau = \mathbf{0}$, $\Delta = \delta\Omega^{1/2}$, and $\bar{\Gamma} = \mathbf{I}_m$ because such detailed specification can significantly simplify the computation of the moments. Nonetheless, the resulting estimates exhibit numerical instability, although unbiased. In addition, the method has only been tested up to bivariate data. Therefore, a more general inference scheme is needed. Only after the successful development of a proper inference algorithm for the SUT distribution can it be applied to real datasets.

Recently, the SUN distribution has been heralded as a significant breakthrough in the field of Bayesian inference. Durante [18] and Anceschi et al. [3] proved that the multivariate linear regression, probit, and tobit models, among others, admit conjugate SUN priors with the posterior distribution remaining in the very same class. This phenomenon allows for a significant reduction in computational cost to perform Bayesian posterior inference via closed-form solutions, thus providing a competitive alternative to well-established sampling methods. Subsequently, Karling et al. [25] have shown that the SUT family is also conjugate to the likelihood induced by general multivariate regression models arising from fully observed, censored, or dichotomized ST distributions. The truncation of the data on those two latter frameworks leads to hidden layers that the SUT distribution can model, winding up as a natural candidate due to the inherent conjugacy properties. Namely, the closeness under linear transformations, conditioning, and marginalization, such as introduced in Propositions 6, 7, 11, and 12 above, play a major role in the work by Karling et al. [25], showcasing the fundamental properties of the SUT distribution. Other important modeling schemes comprehending the SUN and SUT distributions might benefit from the properties shown in the present work and may be of fundamental help in the future. Anceschi et al. [3] have demonstrated several computational approaches to perform Bayesian inference based on SUN conjugacy, such as analytical methods (based on computing high-dimensional Gaussian probabilities), sampling-based methods (e.g., data augmentation Gibbs samplers), and deterministic approximation-based methods (e.g., variational Bayes and expectation-propagation). Their extension to the SUT framework is the topic of ongoing research to be reported elsewhere.

Acknowledgments

References

[1] T. Abe, H. Fujisawa, T. Kawashima, C. Ley, EM algorithm using overparameterization for the multivariate skew-normal distribution, Econom. Stat. 19 (2021) 151–168.

[2] C. Adcock, A. Azzalini, A selective overview of skew-elliptical and related distributions and of their applications, Symmetry 12 (1) (2020) 118.

[3] N. Anceschi, A. Fasano, D. Durante, G. Zanella, Bayesian conjugacy in probit, tobit, multinomial probit and extensions: a review and new results, J. Amer. Statist. Assoc. 118 (542) (2023) 1451–1469.

[4] R.B. Arellano-Valle, A. Azzalini, On the unification of families of skew-normal distributions, Scand. J. Stat. 33 (3) (2006) 561–574.

[5] R.B. Arellano-Valle, A. Azzalini, Some properties of the unified skew-normal distribution, Statist. Papers 63 (2) (2022) 461–487.

[6] R.B. Arellano-Valle, M.D. Branco, M.G. Genton, A unified view on skewed distributions arising from selections, Canad. J. Statist. 34 (4) (2006) 581–601.

[7] R.B. Arellano-Valle, L.M. Castro, R.H. Loschi, Change point detection in the skew-normal model parameters, Comm. Statist. Theory Methods 42 (4) (2013) 603–618.

[8] R.B. Arellano-Valle, C.S. Ferreira, M.G. Genton, Scale and shape mixtures of multivariate skew-normal distributions, J. Multivariate Anal. 166 (2018) 98–110.

[9] R.B. Arellano-Valle, M.G. Genton, An invariance property of quadratic forms in random vectors with a selection distribution, with application to sample variogram and covariogram estimators, Ann. Inst. Statist. Math. 62 (2) (2010) 363–381.

[10] R.B. Arellano-Valle, M.G. Genton, Multivariate extended skew-t distributions and related families, Metron 68 (3) (2010) 201–234.

[11] R.B. Arellano-Valle, M.G. Genton, Multivariate unified skew-elliptical distributions, Chil. J. Stat. 1 (1) (2010) 17–33.

[12] A. Azzalini, An overview on the progeny of the skew-normal family—a personal perspective, J. Multivariate Anal. 188 (2022) Paper No. 104851.

[13] A. Azzalini, A. Capitanio, Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, J. R. Stat. Soc. Ser. B Stat. Methodol. 65 (2) (2003) 367–389.

[14] A. Azzalini, A. Capitanio, The skew-normal and related families, in: Institute of Mathematical Statistics (IMS) Monographs, vol. 3, Cambridge University Press, Cambridge, 2014, p. viii+262.

[15] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (4) (1996) 715–726.

[16] A. Azzalini, A. Genz, Package "mnormt" : The multivariate normal and t distributions, and their truncated versions (version 2.0.1), 2020, https://cran.r-project.org/package=mnormt.

[17] A. Capitanio, On the canonical form of scale mixtures of skew-normal distributions, Statistica 80 (2) (2020) 145–160.

[18] D. Durante, Conjugate Bayes for probit regression via unified skew-normal distributions, Biometrika 106 (4) (2019) 765–779.

[19] K. Fang, S. Kotz, K. Ng, Symmetric Multivariate and Related Distributions, Chapman and Hall, 1990.

[20] C.E. Galarza, K. Raymond, V.H. Lachos, Package "MomTrunc": Moments of folded and doubly truncated multivariate distributions, 2020, https://CRAN.R-project.org/package=MomTrunc/.

[21] M.G. Genton (Ed.), Skew-Elliptical Distributions And Their Applications: A Journey Beyond Normality, Chapman & Hall/CRC, Boca Raton, FL, 2004, xviii+396.

[22] A.K. Gupta, Multivariate skew t-distribution, Statistics 37 (4) (2003) 359–363.

[23] A.K. Gupta, M.A. Aziz, Estimation of parameters of the unified skew normal distribution using the method of weighted moments, J. Stat. Theory Pract. 6 (2012) 402–416.

[24] A.K. Gupta, M.A. Aziz, W. Ning, On some properties of the unified skew normal distribution, J. Stat. Theory Pract. 7 (3) (2013) 480–495.

[25] M.J. Karling, D. Durante, M.G. Genton, Conjugacy properties of multivariate unified skew–elliptical distributions, 2024, arXiv:2402.09837.

[26] S. Karlsson, S. Mazur, H. Nguyen, Vector autoregression models with skewness and heavy tails, J. Econom. Dynam. Control 146 (2023) 104580.

[27] T. Kiss, S. Mazur, H. Nguyen, P. Österholm, Modelling the relation between the US real economy and the corporate bond-yield spread in Bayesian VARs with non-Gaussian disturbances, J. Forecast. 42 (2) (2023) 347–368.

[28] T. Kollo, M.S. Srivastava, Estimation and testing of parameters in multivariate Laplace distribution, Comm. Statist. Theory Methods 33 (10) (2005) 2363–2387.

[29] N. Loperfido, The skewness of mean–variance normal mixtures, J. Multivariate Anal. 199 (2024) 105242.

[30] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (3) (1970) 519–530.

[31] K.V. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, Sankhyā (1974) 115–128.

[32] E.O. Ogundimu, J.L. Hutton, A sample selection model with skew-normal distribution, Scand. J. Stat. 43 (1) (2016) 172–190.

[33] V.K. Rohatgi, A.K.M. Ehsanes Saleh, An introduction to probability and statistics, second ed., in: Wiley Series in Probability and Statistics: Texts and References Section, Wiley-Interscience, New York, 2001, p. xviii+716.

[34] K. Wang, R.B. Arellano-Valle, A. Azzalini, M.G. Genton, On the non-identifiability of unified skew-normal distributions, Stat 12 (1) (2023) e597.