Article

Multivariate Skew t -Distribution: Asymptotics for Parameter Estimators and Extension to Skew t -Copula

Tõnu Kollo , Meelis Käärik and Anne Selart *

Institute of Mathematics and Statistics, University of Tartu, Narva mnt 18, 51009 Tartu, Estonia; tonu.kollo@ut.ee (T.K.); meelis.kaarik@ut.ee (M.K.)

* Correspondence: anne.selart@ut.ee

Abstract: Symmetric elliptical distributions have been intensively used in data modeling and robustness studies. The area of applications was considerably widened after transforming elliptical distributions into the skew elliptical ones that preserve several good properties of the corresponding symmetric distributions and increase possibilities of data modeling. We consider three-parameter p -variate skew t -distribution where p -vector μ is the location parameter, $\Sigma : p \times p$ is the positive definite scale parameter, p -vector α is the skewness or shape parameter, and the number of degrees of freedom ν is fixed. Special attention is paid to the two-parameter distribution when $\mu = \mathbf{0}$ that is useful for construction of the skew t -copula. Expressions of the parameters are presented through the moments and parameter estimates are found by the method of moments. Asymptotic normality is established for the estimators of Σ and α . Convergence to the asymptotic distributions is examined in simulation experiments.

Keywords: asymptotic normality; inverse chi-distribution; multivariate cumulants; multivariate moments; skew normal distribution; skew t -copula; skew t -distribution

MSC: 62E20; 62H10; 62H12

Citation: Kollo, T.; Käärik, M.; Selart, A. Multivariate Skew t -Distribution: Asymptotics for Parameter Estimators and Extension to Skew t -Copula. *Symmetry* **2021**, *13*, 1059. <https://doi.org/10.3390/sym13061059>

Academic Editor: Nicola Maria Rinaldo Loperfido

Received: 20 May 2021
Accepted: 8 June 2021
Published: 11 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The idea of constructing multivariate skewed distributions from symmetric elliptical distributions has attracted many statisticians over last 25 years. This generalization preserves several good properties of corresponding symmetric distributions and in the same time increases possibilities of data modeling. In 1996, Azzalini and Dalla Valle introduced multivariate skew normal distribution [1]. Their idea of transforming multivariate normal distribution into a skewed distribution has been carried over to several other distributions and many generalizations have been made since 1996 (see for instance, Theodossiou [2], Arellano-Valle and Genton [3]). From these developments multivariate skew t -distribution has become most frequently used [4]. This distribution became an important tool in many data modeling applications because of heavier tail area than the normal and skew normal distributions (see, for instance, Lee and McLachan [5], Marchenko [6], Parisi and Liseo [7], Fernandez and Steel [8], Jones [9]). A review of applications in finance and insurance mathematics is given in Adcock et al. [10]. Multivariate skew t -distribution has also been successfully used as a data model in environmental studies and modelling coastal floodings (Ghizzolini et al. [11], Thompson and Shen [12]). Several other applications of skew elliptical distributions can be found in the collective monograph edited by Genton [13] and in the book by Azzalini and Capitanio [14] which includes a comprehensive list of references on the topic. Another useful property that makes multivariate t -distribution and skew t -distribution attractive in financial and actuarial applications is the possibility to take tail dependence between marginals into account. Tail dependence of multivariate t - and skew t -distributions has been intensively studied in recent years, see Fung and Seneta [15] or Joe and Li [16], for example. In Kollo et al. [17] it is shown that the upper

tail dependence coefficient of the skew t -distribution can be several times bigger than the corresponding value of multivariate t -distribution. Moments and inferential aspects related to the multivariate skew t -distribution have been also actively studied in Padoan [18], Galarza et al. [19], Zhou and He [20], Lee and McLachlan [5], DiCiccio and Monti [21], for instance. A construction based on the skew t -distribution—skew t -copula—has made this distribution even more attractive in applications where correlated differently distributed marginals are joined into a multivariate distribution. Demarta and McNeil introduced skew t -copula based on multivariate generalized hyperbolic distribution [22], Kollo and Pettere [23] defined skew t -copula on the basis of the multivariate skew t -distribution following Azzalini and Capitanio [4].

The paper is organized as follows. In Section 2, we introduce necessary notation and notions from matrix algebra. In Section 3, skew t -distribution is defined through skew normal and inverse χ -distribution and the first four central moments are found via the moments of skew normal and inverse χ -distribution. The asymptotic normality for parameter estimators in the two-parameter case is also established in this section. In Section 4, obtained analytical results are illustrated by simulations for different scenarios and the speed of convergence to asymptotic distribution is also addressed. In Section 5, multivariate skew t -copula is defined and some properties are discussed. Technical proofs are presented in Appendix A. A comparison of density contours of Gaussian, skew normal and skew t -copula is presented in Supplementary Materials.

2. Notation and Preliminaries

To derive asymptotic normal distribution for the parameter estimators of skew t -distribution, we use matrix technique. Necessary notions and notation are presented in what follows.

Let us first denote the moments of a random vector $\mathbf{x}$ in the following matrix representation:

$$m_k(\mathbf{x}) = E\left[\underbrace{\mathbf{x} \otimes \mathbf{x}' \otimes \mathbf{x} \otimes \mathbf{x}' \otimes \dots}_{k \text{ times}}\right] = E[\mathbf{x}\mathbf{x}' \otimes \mathbf{x}\mathbf{x}' \otimes \dots].$$

The corresponding central moment is

$$\bar{m}_k(\mathbf{x}) = m_k(\mathbf{x} - E\mathbf{x}).$$

We use notions *vec-operator*, *commutation matrix*, *Kronecker product*, *matrix derivative* and their properties. For deeper insight into this technique we refer to Magnus and Neudecker [24], or Kollo and von Rosen [25]. If $\mathbf{A}$ is a $p \times q$ matrix, $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_q)$, the pq column vector $\text{vec } \mathbf{A}$ is formed by stacking the columns $\mathbf{a}_i$ under each other; that is

$$\text{vec } \mathbf{A} = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_q \end{pmatrix}.$$

We denote a $p \times p$ identity matrix by $\mathbf{I}_p$ and a $pq \times pq$ commutation matrix by $\mathbf{K}_{p,q}$. Commutation matrix $\mathbf{K}_{p,q}$ is an orthogonal $pq \times pq$ partitioned matrix consisting of $q \times p$ blocks with the property

$$\mathbf{K}_{p,q} \text{vec } \mathbf{A} = \text{vec } \mathbf{A}'$$

where $\mathbf{A}$ is a $p \times q$ -matrix. The Kronecker product $\mathbf{A} \otimes \mathbf{B}$ of a $p \times q$ matrix $\mathbf{A}$ and an $r \times s$ matrix $\mathbf{B}$ is a $pr \times qs$ partitioned matrix consisting of the $r \times s$ blocks

$$\mathbf{A} \otimes \mathbf{B} = [a_{ij}\mathbf{B}], \quad i = 1, \dots, p; \quad j = 1, \dots, q.$$

The following properties of the notions above are repeatedly used. It is assumed that the dimensions of the matrices match with the operations.

1. $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$;

2. $\mathbf{ab}' = \mathbf{a} \otimes \mathbf{b}' = \mathbf{b}' \otimes \mathbf{a}$;
3. $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec } \mathbf{B}$;
4. $\mathbf{K}_{p,1} = \mathbf{I}_p, \quad \mathbf{K}_{1,q} = \mathbf{I}_q$;
5. for $\mathbf{A} : p \times q$ and $\mathbf{B} : r \times s, \mathbf{A} \otimes \mathbf{B} = \mathbf{K}_{p,r}(\mathbf{B} \otimes \mathbf{A})\mathbf{K}_{s,q}$.

Further, when deriving asymptotic distributions of the estimators of Σ and α in Section 3, we use the matrix derivative where partial derivatives $\frac{\partial y_{kl}}{\partial x_{ij}}$ are ordered by the Kronecker product (see Magnus and Neudecker, [24])

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \frac{d}{d\text{vec}'\mathbf{X}} \otimes \text{vec } \mathbf{Y},$$

with $\mathbf{X} : p \times q$ and $\mathbf{Y} = \mathbf{Y}(\mathbf{X}) : r \times s$. It follows from the definition that

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \frac{d\text{vec } \mathbf{Y}}{d\text{vec}'\mathbf{X}} = \frac{d\text{vec } \mathbf{Y}}{d\mathbf{X}} = \frac{d\mathbf{Y}}{d\text{vec}'\mathbf{X}}.$$

The main properties of the matrix derivative are listed below.

1. $\frac{d\mathbf{X}}{d\mathbf{X}} = \mathbf{I}_{pq}$;
2. $\frac{d(\mathbf{Y}+\mathbf{Z})}{d\mathbf{X}} = \frac{d\mathbf{Y}}{d\mathbf{X}} + \frac{d\mathbf{Z}}{d\mathbf{X}}, \mathbf{Z} = \mathbf{Z}(\mathbf{X}) : r \times s$;
3. $\frac{d\mathbf{AXB}}{d\mathbf{X}} = \mathbf{B}' \otimes \mathbf{A}$;
4. If $\mathbf{Z} = \mathbf{Z}(\mathbf{Y}), \mathbf{Y} = \mathbf{Y}(\mathbf{X})$ then $\frac{d\mathbf{Z}}{d\mathbf{X}} = \frac{d\mathbf{Z}}{d\mathbf{Y}} \frac{d\mathbf{Y}}{d\mathbf{X}}$;
5. If $\mathbf{W} = \mathbf{YZ}, \mathbf{Z} = \mathbf{Z}(\mathbf{X}) : s \times t$ then $\frac{d\mathbf{W}}{d\mathbf{X}} = (\mathbf{Z}' \otimes \mathbf{I}_r) \frac{d\mathbf{Y}}{d\mathbf{X}} + (\mathbf{I}_t \otimes \mathbf{Y}) \frac{d\mathbf{Z}}{d\mathbf{X}}$;
6. If $\mathbf{X} : p \times p$ then $\frac{d\mathbf{X}^{-1}}{d\mathbf{X}} = -(\mathbf{X}')^{-1} \otimes \mathbf{X}^{-1}$;
7. If $\mathbf{X} \in \mathbb{R}^{p \times p}$ then $\frac{d|\mathbf{X}|}{d\mathbf{X}} = |\mathbf{X}| \text{vec}(\mathbf{X}^{-1})'$, where $|\cdot|$ denotes the determinant;
8. If $\mathbf{W} = \mathbf{Y} \otimes \mathbf{Z}, \mathbf{Z} = \mathbf{Z}(\mathbf{X}) : m \times n$, then $\frac{d\mathbf{W}}{d\mathbf{X}} = (\mathbf{I}_s \otimes \mathbf{K}_{r,n} \otimes \mathbf{I}_m) \left[(\mathbf{I}_{rs} \otimes \text{vec } \mathbf{Z}) \frac{d\mathbf{Y}}{d\mathbf{X}} + (\text{vec } \mathbf{Y} \otimes \mathbf{I}_{mn}) \frac{d\mathbf{Z}}{d\mathbf{X}} \right]$.
9. If $y = y(\mathbf{X}), \mathbf{Z} = \mathbf{Z}(\mathbf{X}) : m \times n$, then $\frac{d(y\mathbf{Z})}{d\mathbf{X}} = \text{vec } \mathbf{Z} \frac{dy}{d\mathbf{X}} + y \frac{d\mathbf{Z}}{d\mathbf{X}}$.

3. Multivariate Skew t_v -Distribution

3.1. Definition

Multivariate skew t -distribution has been defined in different ways, see Kotz and Nadarajah [26] (Chapter 5). Kshirsagar [27] was the first to introduce a noncentral multivariate t -distribution, but the density expression was complicated and the distribution was not used in applications. From more recent definitions we refer to Fang et al. [28], but as the marginals in their definition have different degrees of freedom, the distribution had problems with practical use. Gupta [29] defined the multivariate skew t -distribution in a similar way like Azzalini and Capitanio [4], but their definition had several advantages and has become widely used in applications (for example, it is possible to use explicit expression of the moment generating function of skew normal distribution for calculating moments of skew t -distribution). We follow the approach by Azzalini and Capitanio [4].

A p -vector $\mathbf{x}$ has multivariate skew t -distribution with ν degrees of freedom if it is the ratio of a p -variate skew normal vector and a chi-distributed random variable with ν degrees of freedom. We have slightly modified their definition of the skew normal vector and use the covariance matrix as a parameter instead of the correlation matrix. Let $\mathbf{y}$ be a p -variate skew normal vector with density

$$f(\mathbf{y}) = 2f_{N_p(\mathbf{0}, \Sigma)}(\mathbf{y})\Phi(\alpha'\mathbf{y}), \quad (1)$$

where $f_{N_p(\mathbf{0}, \Sigma)}(\mathbf{y})$ is the density function (density) of the p -variate normal distribution $N_p(\mathbf{0}, \Sigma)$ and $\Phi(\cdot)$ is the cumulative distribution function (distribution function) of $N(0, 1)$.

Denote the distribution of vector $\mathbf{y}$ with density (1) by $\mathbf{y} \sim SN_p(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$, where $\boldsymbol{\Sigma} : p \times p$ is the scale parameter and $\boldsymbol{\alpha} \in \mathbb{R}^p$ is the shape or skewing parameter.

Definition 1. Let $\mathbf{y}$ be skew normally distributed with density (1) and let Z be χ -distributed with ν degrees of freedom, independent from $\mathbf{y}$. Then

$$\mathbf{x} = \boldsymbol{\mu} + \sqrt{\nu} \frac{\mathbf{y}}{Z} \quad (2)$$

is skew t_ν -distributed, $\mathbf{x} \sim St_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$.

The density function of $\mathbf{x} \sim St_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$ is of the form (a modification of Azzalini and Capitanio [4])

$$f_{\mathbf{x}}(\mathbf{x}) = 2f_{\nu,p}(\mathbf{x})F_{p+\nu}\left(\boldsymbol{\alpha}'(\mathbf{x} - \boldsymbol{\mu})\sqrt{\frac{p+\nu}{Q+\nu}}\right),$$

where $Q = (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$, F_k is the distribution function of the univariate t -distribution with k degrees of freedom and $f_{\nu,p}$ is the density function of the p -variate t -distribution with ν degrees of freedom and scale matrix $\boldsymbol{\Sigma}$

$$f_{\nu,p}(\mathbf{x}) = \frac{\Gamma(\frac{\nu+p}{2})}{(\nu\pi)^{p/2}\Gamma(\frac{\nu}{2})|\boldsymbol{\Sigma}|^{1/2}} \left[1 + \frac{(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})}{\nu}\right]^{-\frac{\nu+p}{2}}.$$

3.2. Moments

By Formula (2), a skew t -distributed $\mathbf{x}$ is constructed using independent skew normal random vector $\mathbf{y}$ and χ -distributed Z . Because of the independence of $\mathbf{y}$ and Z , the moments of $\mathbf{x}$ can be found as products of the moments of the skew normal and inverse χ -distribution. The k -th moment, central moment and cumulant of a skew-normal random vector $\mathbf{y}$ are denoted by $m_k(\mathbf{y})$, $\bar{m}_k(\mathbf{y})$ and $c_k(\mathbf{y})$, respectively. Explicit expressions of the raw moments, central moments and cumulants of $\mathbf{y}$ can be obtained by differentiating the moment generating function

$$M(\mathbf{t}) = 2e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}\Phi\left(\frac{\boldsymbol{\alpha}'\boldsymbol{\Sigma}\mathbf{t}}{\sqrt{1 + \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}}}\right)$$

and the cumulant generating function

$$C(\mathbf{t}) = \ln M(\mathbf{t}).$$

The first four cumulants for a skew-normal random vector $\mathbf{y}$ are (Kollo et al. [30])

$$c_1(\mathbf{y}) = E\mathbf{y} = \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}, \quad c_2(\mathbf{y}) = D\mathbf{y} = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\delta}\boldsymbol{\delta}', \quad (3)$$

$$c_3(\mathbf{y}) = \bar{m}_3(\mathbf{y}) = \sqrt{\frac{2}{\pi}}\left(\frac{4}{\pi} - 1\right)\boldsymbol{\delta} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\delta}, \quad (4)$$

$$c_4(\mathbf{y}) = \frac{8}{\pi}\left(1 - \frac{3}{\pi}\right)\boldsymbol{\delta} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\delta} \otimes \boldsymbol{\delta}',$$

where $\boldsymbol{\delta}$ equals

$$\boldsymbol{\delta} = \frac{\boldsymbol{\Sigma}\boldsymbol{\alpha}}{\sqrt{1 + \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}}}. \quad (5)$$

From general relation between the fourth order cumulants and central moments we have (see, e.g., Kollo and von Rosen [25] (p. 187))

$$\bar{m}_4(\mathbf{y}) = c_4(\mathbf{y}) + (\mathbf{I}_{p^2} + K_{p,p})(D\mathbf{y} \otimes D\mathbf{y}) + \text{vec}(D\mathbf{y})\text{vec}'(D\mathbf{y}). \quad (6)$$

After some algebra, we get the expression for α from (5):

$$\alpha = \frac{\Sigma^{-1}\delta}{\sqrt{1 - \delta'\Sigma^{-1}\delta}}. \quad (7)$$

The quantity under square root in (7) influences the stability of parameter estimates and will be used later in a related discussion. Thus, we add a notation for it:

$$\eta = 1 - \delta'\Sigma^{-1}\delta. \quad (8)$$

Note also that each component δ_i in vector $\delta = (\delta_1, \dots, \delta_p)'$ is directly connected to the corresponding marginal distribution, whereas this property does not hold for vector $\alpha = (\alpha_1, \dots, \alpha_p)'$. For a more comprehensive discussion on parametrizations of skew-normal vector see Käärik et al. [31].

To derive the central moments of the multivariate skew t_ν -distribution, we also need the moments of the skew normally distributed vector $\mathbf{y}$

$$m_2(\mathbf{y}) = \Sigma; \quad (9)$$

$$m_3(\mathbf{y}) = \sqrt{\frac{2}{\pi}}(\text{vec } \Sigma \delta' + \delta \otimes \Sigma + \Sigma \otimes \delta - \delta \delta' \otimes \delta). \quad (10)$$

Recall that a random variable $V_\nu = \frac{1}{Z}$ is inverse χ_ν -distributed if its probability density function is (Lee [32] (p. 378))

$$f(x) = \frac{1}{2^{\nu/2-1}\Gamma(\frac{\nu}{2})} x^{-\nu-1} \exp\left(-\frac{1}{2x^2}\right).$$

The moments of the inverse χ_ν -distribution can be obtained by integration.

Lemma 1. *The moments of the inverse χ_ν -distribution are:*

$$EV_\nu = \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)}, \quad \nu > 1;$$

$$EV_\nu^k = \frac{\Gamma\left(\frac{\nu-k}{2}\right)}{2^{k/2}\Gamma\left(\frac{\nu}{2}\right)}, \quad \nu > k.$$

Proof. By standard integration technique, we can derive the expectation:

$$\begin{aligned} EV_\nu &= \int_0^\infty x \frac{1}{2^{\nu/2-1}\Gamma(\frac{\nu}{2})} x^{-\nu-1} \exp\left(-\frac{1}{2x^2}\right) dx \\ &= \frac{1}{\Gamma(\frac{\nu}{2})} \int_0^\infty \frac{\Gamma(\frac{\nu-1}{2})}{2^{\frac{\nu-1}{2}-1}\sqrt{2}\Gamma(\frac{\nu-1}{2})} x^{-\nu} \exp\left(-\frac{1}{2x^2}\right) dx = \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{2}}. \end{aligned}$$

Following the same approach, the k -th order moment can be derived as follows

$$\begin{aligned}
EV_\nu^k &= \int_0^\infty x^k \frac{1}{2^{\nu/2-1}\Gamma(\frac{\nu}{2})} x^{-\nu-1} \exp\left(-\frac{1}{2x^2}\right) dx \\
&= \frac{2^{\frac{\nu-1}{2}-1}\Gamma(\frac{\nu-1}{2})}{2^{\frac{\nu}{2}-1}\Gamma(\frac{\nu}{2})} \int_0^\infty \frac{1}{2^{\frac{\nu-1}{2}-1}\Gamma(\frac{\nu-1}{2})} x^{-(\nu-1)-1} x^{k-1} \exp\left(-\frac{1}{2x^2}\right) dx \\
&= \prod_{i=0}^{k-1} EV_{\nu-i} = \frac{\Gamma(\frac{\nu-k}{2})}{2^{\frac{k}{2}}\Gamma(\frac{\nu}{2})}, \quad \nu > k.
\end{aligned}$$

□

The expressions of the second, third and fourth moment are special cases of this result:

$$\begin{aligned}
EV_\nu^2 &= \frac{1}{\nu-2}, \quad \nu > 2; \\
EV_\nu^3 &= \frac{EV_\nu}{\nu-3}, \quad \nu > 3; \\
EV_\nu^4 &= \frac{1}{(\nu-2)(\nu-4)}, \quad \nu > 4.
\end{aligned}$$

After some elementary algebra we get expressions of the corresponding central moments

$$\begin{aligned}
\bar{m}_3(V_\nu) &= EV_\nu \left(\frac{7-2\nu}{(\nu-3)(\nu-2)} + 2(EV_\nu)^2 \right); \\
\bar{m}_4(V_\nu) &= \frac{1}{(\nu-2)(\nu-4)} + 2 \frac{\nu-5}{(\nu-2)(\nu-3)} (EV_\nu)^2 - 3(EV_\nu)^4.
\end{aligned}$$

To keep the notation simpler, we denote $V = V_\nu$.

In order to derive the asymptotic normality for parameter estimators of St_ν , we need the first 4 central moments of the distribution. Let us start with general expressions for central moments $\bar{m}_3(\mathbf{x})$ and $\bar{m}_4(\mathbf{x})$.

Lemma 2. Let $\mathbf{x}$ be a p -dimensional random vector with finite fourth order moments. Then the third and fourth central moment of $\mathbf{x}$ can be expressed as follows:

$$\begin{aligned}
\bar{m}_3(\mathbf{x}) &= m_3(\mathbf{x}) - (\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(m_2(\mathbf{x}) \otimes \mathbf{E}\mathbf{x}) - \text{vec } m_2(\mathbf{x})\mathbf{E}\mathbf{x}' + 2\mathbf{E}\mathbf{x}\mathbf{E}\mathbf{x}' \otimes \mathbf{E}\mathbf{x}, \\
\bar{m}_4(\mathbf{x}) &= m_4(\mathbf{x}) - \mathbf{C}(\mathbf{x})(\mathbf{I}_{p^2} + \mathbf{K}_{p,p}) - (\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(\mathbf{C}(\mathbf{x}))' - 3\mathbf{E}\mathbf{x}\mathbf{E}\mathbf{x}' \otimes \mathbf{E}\mathbf{x}\mathbf{E}\mathbf{x}' \\
&\quad + (\mathbf{I}_{p^2} + \mathbf{K}_{p,p})\mathbf{A}(\mathbf{x}) + \mathbf{A}(\mathbf{x})\mathbf{K}_{p,p} + \mathbf{B}(\mathbf{x}) + (\mathbf{A}(\mathbf{x}) + \mathbf{B}(\mathbf{x}))',
\end{aligned}$$

where

$$\mathbf{A}(\mathbf{x}) = \mathbf{E}\mathbf{x}' \otimes m_2(\mathbf{x}) \otimes \mathbf{E}\mathbf{x}; \quad \mathbf{B}(\mathbf{x}) = \mathbf{E}\mathbf{x}' \otimes \text{vec } m_2(\mathbf{x}) \otimes \mathbf{E}\mathbf{x}'; \quad \mathbf{C}(\mathbf{x}) = m_3(\mathbf{x}) \otimes \mathbf{E}\mathbf{x}'.$$

Proof of the lemma is presented in Appendix A.1. Applying this lemma to skew t -distributed $\mathbf{x}$, the next result follows.

Lemma 3. Let $\mathbf{x} \sim St_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$. Then its mean $\mathbf{E}\mathbf{x}$, covariance matrix $D\mathbf{x}$, and central moments $\bar{m}_3(\mathbf{x})$ and $\bar{m}_4(\mathbf{x})$ can be expressed through the moments of inverse χ_ν -distributed V and $\mathbf{y} \sim SN_p(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$ in the following form:

$$\mathbf{E}\mathbf{x} = \boldsymbol{\mu} + \sqrt{\nu}\mathbf{E}V\mathbf{E}\mathbf{y}; \quad (11)$$

$$D\mathbf{x} = \nu m_2(V)m_2(\mathbf{y}) - \mathbf{E}\mathbf{x}\mathbf{E}\mathbf{x}' = \nu(m_2(V)D\mathbf{y} + D\mathbf{V}\mathbf{E}\mathbf{y}\mathbf{E}\mathbf{y}'); \quad (12)$$

$$\begin{aligned}\bar{m}_3(\mathbf{x}) = & \nu^{3/2} \{ EV^3 \bar{m}_3(\mathbf{y}) - 2(EV^3 - (EV)^3)(E\mathbf{y}E\mathbf{y}' \otimes E\mathbf{y}) \\ & + (EV^3 - EVEV^2)[(\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(m_2(\mathbf{y}) \otimes E\mathbf{y}) + \text{vec } m_2(\mathbf{y})E\mathbf{y}'] \};\end{aligned}\quad (13)$$

$$\begin{aligned}\bar{m}_4(\mathbf{x}) = & \nu^2 \{ EV^4 \bar{m}_4(\mathbf{y}) + (EV^4 - EVEV^3)(\mathbf{C}(\mathbf{y})(\mathbf{I}_{p^2} + \mathbf{K}_{p,p}) + (\mathbf{I}_{p^2} + \mathbf{K}_{p,p})\mathbf{C}(\mathbf{y})') \\ & + 3(EV^4 - (EV)^4)E\mathbf{y}E\mathbf{y}' \otimes E\mathbf{y}E\mathbf{y}' \\ & - (EV^4 - EV^2(EV)^2)[(\mathbf{I}_{p^2} + \mathbf{K}_{p,p})\mathbf{A}(\mathbf{y}) + \mathbf{A}(\mathbf{y})\mathbf{K}_{p,p} + \mathbf{B}(\mathbf{y}) + (\mathbf{A}(\mathbf{y}) + \mathbf{B}(\mathbf{y}))'] \}.\end{aligned}\quad (14)$$

Proof is straightforward and follows from Lemma 2 after substituting in the required moments of skew normal distribution from Formulas (9) and (10) and the moments of inverse χ -distribution given in Lemma 1.

3.3. Point Estimates

Let us now focus on the case when $\mu = \mathbf{0}$ and estimate the parameters of the two-parameter distribution $St_\nu(\Sigma, \alpha)$ using the method of moments.

Let us have a sample of size n , $n > p$, from $St_\nu(\Sigma, \alpha)$ with sample mean $\bar{\mathbf{x}}$ and unbiased sample covariance matrix $\mathbf{S}$. Now, the estimates for parameters Σ and α through the moments of V and $\mathbf{y}$ can be obtained from Expressions (11) and (12):

$$E\mathbf{x} = \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \sqrt{\frac{\nu}{\pi}} \delta,$$

$$D\mathbf{x} = \nu \left[\frac{1}{\nu-2} \Sigma - \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 \frac{1}{\pi} \delta \delta' \right],$$

where δ is defined in (5). From here we get the following estimates

$$\hat{\Sigma} = \frac{\nu-2}{\nu} (\mathbf{S} + \bar{\mathbf{x}}\bar{\mathbf{x}}'); \quad (15)$$

$$\hat{\alpha} = \frac{\frac{\nu}{\nu-2} (\mathbf{S} + \bar{\mathbf{x}}\bar{\mathbf{x}}')^{-1} \bar{\mathbf{x}}}{\sqrt{\frac{\nu}{\pi} \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 - \frac{\nu}{\nu-2} \bar{\mathbf{x}}' (\mathbf{S} + \bar{\mathbf{x}}\bar{\mathbf{x}}')^{-1} \bar{\mathbf{x}}}}. \quad (16)$$

3.4. Asymptotic Normality

As the estimators depend on both, the sample mean and the sample covariance matrix, we need their joint distribution to find the asymptotic distributions of the parameter estimators. In general, the following convergence in distribution takes place for $\begin{pmatrix} \bar{\mathbf{x}} \\ \text{vec } \mathbf{S} \end{pmatrix}$ if $\bar{m}_4(\mathbf{x})$ exists (Parring [33]):

$$\sqrt{n} \left[\begin{pmatrix} \bar{\mathbf{x}} \\ \text{vec } \mathbf{S} \end{pmatrix} - E \begin{pmatrix} \bar{\mathbf{x}} \\ \text{vec } \mathbf{S} \end{pmatrix} \right] \xrightarrow{D} N(\mathbf{0}, \mathbf{D}_{as}),$$

where

$$\mathbf{D}_{as} = \begin{pmatrix} D\mathbf{x} & \vdots & \bar{m}_3(\mathbf{x})' \\ \bar{m}_3(\mathbf{x}) & \vdots & \bar{m}_4(\mathbf{x}) - \text{vec}(D\mathbf{x})\text{vec}'(D\mathbf{x}) \end{pmatrix}.$$

For the asymptotic covariance matrix of the joint distribution of sample mean and covariance matrix from skew t -distribution, $\mathbf{D}_{as}^{St}$, we get the expressions of necessary central moments from Equations (12)–(14). Derivation of asymptotic normality for estimators of parameters is based on Theorem 1 by Anderson [34] (pp. 132–133).

Theorem 1. Assume that for $\{\mathbf{x}_n\}$

$$\sqrt{n}(\mathbf{x}_n - \mathbf{a}) \xrightarrow{D} N_p(\mathbf{0}, \Sigma)$$

and

$$\mathbf{x}_n \xrightarrow{P} \mathbf{a},$$

if $n \rightarrow \infty$. Let the function $\mathbf{g}(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}^q$ have continuous partial derivatives in a neighborhood of $\mathbf{a}$. Then, if $n \rightarrow \infty$

$$\sqrt{n}(\mathbf{g}(\mathbf{x}_n) - \mathbf{g}(\mathbf{a})) \xrightarrow{D} N_q(\mathbf{0}, \xi \Sigma \xi'),$$

where

$$\xi = \left. \frac{d\mathbf{g}(\mathbf{x})}{d\mathbf{x}} \right|_{\mathbf{x}=\mathbf{a}} \neq \mathbf{0}$$

is the matrix derivative.

The vector $(\bar{\mathbf{x}}', \text{vec}' \mathbf{S})'$ satisfies assumptions of $\mathbf{x}_n$ and we consider $\hat{\Sigma}$ and $\hat{\alpha}$ as the functions of $(\bar{\mathbf{x}}', \text{vec}' \mathbf{S})'$ and derive asymptotic normality for the estimators. For $\hat{\Sigma}$ the following statement holds.

Theorem 2. Let us have a p -vector $\mathbf{x} \sim St_\nu(\Sigma, \alpha)$, where ν is fixed. Then, for the estimator $\hat{\Sigma}$ in (15), the following convergence takes place, if $n \rightarrow \infty$

$$\sqrt{n} \text{vec}(\hat{\Sigma} - \Sigma) \xrightarrow{D} N_{p^2}(\mathbf{0}, \mathbf{D}_{as}^\Sigma),$$

where

$$\mathbf{D}_{as}^\Sigma = \left. \frac{d(\frac{\nu-2}{\nu}(\mathbf{S} + \bar{\mathbf{x}}\bar{\mathbf{x}}'))}{d(\bar{\mathbf{x}}', \text{vec}' \mathbf{S})} \right|_{\mathbf{E} \begin{pmatrix} \bar{\mathbf{x}} \\ \text{vec} \mathbf{S} \end{pmatrix}} \mathbf{D}_{as}^{St} \left(\frac{d(\frac{\nu-2}{\nu}(\mathbf{S} + \bar{\mathbf{x}}\bar{\mathbf{x}}'))}{d(\bar{\mathbf{x}}', \text{vec}' \mathbf{S})} \right)' \Big|_{\mathbf{E} \begin{pmatrix} \bar{\mathbf{x}} \\ \text{vec} \mathbf{S} \end{pmatrix}},$$

the matrix $\mathbf{D}_{as}^{St}$ is the asymptotic covariance matrix of $\begin{pmatrix} \bar{\mathbf{x}} \\ \text{vec} \mathbf{S} \end{pmatrix}$ and the derivative

$$\left. \frac{d(\frac{\nu-2}{\nu}(\mathbf{S} + \bar{\mathbf{x}}\bar{\mathbf{x}}'))}{d(\bar{\mathbf{x}}', \text{vec}' \mathbf{S})} \right|_{\mathbf{E} \begin{pmatrix} \bar{\mathbf{x}} \\ \text{vec} \mathbf{S} \end{pmatrix}} = \frac{\nu-2}{\nu} \left((\mathbf{I}_{p^2} + \mathbf{K}_{p,p}) (\mathbf{I}_p \otimes \sqrt{\frac{2}{\pi}} \boldsymbol{\delta}) ; \mathbf{I}_{p^2} \right).$$

Proof. To find the matrix derivative, we can rely on the corresponding expression for the skew normal distribution, derived by Kollo et al. [30] (Proposition 2) as the expression of the derivative for skew t -distribution differs only by a constant.

For $\mathbf{y} \sim SN_p(\Sigma, \alpha)$, the derivative of $\hat{\Sigma}$ is:

$$\left((\mathbf{I}_{p^2} + \mathbf{K}_{p,p}) (\mathbf{I}_p \otimes \sqrt{\frac{2}{\pi}} \boldsymbol{\delta}) ; \mathbf{I}_{p^2} \right).$$

From here, the derivative of $\hat{\Sigma}$ for $\mathbf{x} \sim St_\nu(\Sigma, \alpha)$ is:

$$\left. \frac{d\hat{\Sigma}}{d(\bar{\mathbf{x}}', \text{vec}' \mathbf{S})} \right|_{\mathbf{E} \begin{pmatrix} \bar{\mathbf{x}} \\ \text{vec} \mathbf{S} \end{pmatrix}} = \frac{\nu-2}{\nu} \left((\mathbf{I}_{p^2} + \mathbf{K}_{p,p}) (\mathbf{I}_p \otimes \sqrt{\frac{2}{\pi}} \boldsymbol{\delta}) ; \mathbf{I}_{p^2} \right).$$

This completes the proof. $\square$

The proof of asymptotic normality of estimator for parameter α ,

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{D} N_p(\mathbf{0}, \mathbf{D}_{as}^\alpha),$$

is similar. In order to shorten the notation from (16), we define

$$M = \frac{\nu}{\pi} \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 - \frac{\nu}{\nu-2} \bar{\mathbf{x}}' (\mathbf{S} + \bar{\mathbf{x}}\bar{\mathbf{x}}')^{-1} \bar{\mathbf{x}}. \quad (17)$$

This allows us to express the estimator $\hat{\boldsymbol{\alpha}}$ as

$$\hat{\boldsymbol{\alpha}} = \frac{\hat{\boldsymbol{\Sigma}}^{-1} \bar{\mathbf{x}}}{\sqrt{M}}. \quad (18)$$

Theorem 3. Let us have a p -vector $\mathbf{x} \sim St_\nu(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$, where ν is fixed. Then, for the estimator $\hat{\boldsymbol{\alpha}}$ in (18), the following convergence takes place, if $n \rightarrow \infty$

$$\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{D} N_p(\mathbf{0}, \mathbf{D}_{as}^{\boldsymbol{\alpha}}),$$

where

$$\mathbf{D}_{as}^{\boldsymbol{\alpha}} = \left. \frac{d\hat{\boldsymbol{\alpha}}}{d(\bar{\mathbf{x}}', \text{vec}' \mathbf{S})} \right|_{\mathbf{E} \begin{pmatrix} \bar{\mathbf{x}} \\ \text{vec} \mathbf{S} \end{pmatrix}} \mathbf{D}_{as}^{St} \left(\left. \frac{d\hat{\boldsymbol{\alpha}}}{d(\bar{\mathbf{x}}', \text{vec}' \mathbf{S})} \right)' \right|_{\mathbf{E} \begin{pmatrix} \bar{\mathbf{x}} \\ \text{vec} \mathbf{S} \end{pmatrix}}, \quad (19)$$

matrix $\mathbf{D}_{as}^{St}$ is the asymptotic covariance matrix and the derivative

$$\frac{d\hat{\boldsymbol{\alpha}}}{d(\bar{\mathbf{x}}', \text{vec}' \mathbf{S})} = \left(\frac{d\hat{\boldsymbol{\alpha}}}{d\bar{\mathbf{x}}'}; \frac{d\hat{\boldsymbol{\alpha}}}{d\text{vec} \mathbf{S}} \right),$$

where

$$\frac{d\hat{\boldsymbol{\alpha}}}{d\bar{\mathbf{x}}} = M^{-\frac{1}{2}} \hat{\boldsymbol{\Sigma}}^{-1} \left[\left(1 - \frac{\nu-2}{\nu} \bar{\mathbf{x}}' \hat{\boldsymbol{\Sigma}}^{-1} \bar{\mathbf{x}} \right) \mathbf{I}_p + M^{-1} \left(\left(1 - \frac{\nu-2}{\nu} (\bar{\mathbf{x}}' \hat{\boldsymbol{\Sigma}}^{-1} \bar{\mathbf{x}} + M) \right) \bar{\mathbf{x}} \bar{\mathbf{x}}' \hat{\boldsymbol{\Sigma}}^{-1} \right) \right].$$

and

$$\frac{d\hat{\boldsymbol{\alpha}}}{d\text{vec} \mathbf{S}} = -\frac{\nu-2}{\nu} M^{-\frac{1}{2}} \left[(\bar{\mathbf{x}}' \hat{\boldsymbol{\Sigma}}^{-1} \otimes \hat{\boldsymbol{\Sigma}}^{-1}) + \frac{1}{2} M^{-1} (\hat{\boldsymbol{\Sigma}}^{-1} \bar{\mathbf{x}} \bar{\mathbf{x}}' \hat{\boldsymbol{\Sigma}}^{-1} \otimes \bar{\mathbf{x}}' \hat{\boldsymbol{\Sigma}}^{-1}) \right].$$

The derivatives in (19) for the asymptotic covariance matrix $\mathbf{D}_{as}^{\boldsymbol{\alpha}}$ are obtained when the estimators $\bar{\mathbf{x}}$ and $\hat{\boldsymbol{\Sigma}}$ are replaced by their expected values.

Derivations in the proof of Theorem 3 are more technical than in the proof of Theorem 2 and therefore the proof is presented in Appendix A.2.

4. Simulations and Discussion

To illustrate theoretical results, we consider a bivariate skew t -distribution $\mathbf{x} \sim St_\nu(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$ and study the convergence speed of the estimates using several parameter sets. First we fix two values of η in (8), since, by the construction, this quantity has effect on the stability of the moments-based estimate for $\boldsymbol{\alpha}$. Namely, values of η close to zero make the moment-based estimates unstable (see, e.g., Formula (7)). Taking this into account, we choose one “problematic” value of η , $\eta \approx 0.03$, and one “safe” value, $\eta \approx 0.3$. Different values for $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}$ are chosen to obtain given values of η . In both cases, two $\boldsymbol{\alpha}$ values of different magnitude are considered: small value is set to $(0.5, 1)'$ and moderate to $(3, 2)'$. Furthermore, the off-diagonal values in matrix parameter $\boldsymbol{\Sigma}$ are pre-set to achieve low or high absolute value of $\sigma_{12} / \sqrt{\sigma_{11}\sigma_{22}}$ with both signs. Thus, altogether 16 sets of parameters are used, the exact values can be found in Appendix A.3, Table A1. The number of degrees of freedom ν is taken equal to 4 in all cases considered.

With all parameter sets, the data is simulated using sample sizes n from 100 to 10^6 (using sequence for the powers of 10 from 2 to 6 with 50 equally spaced points) and $k = 3000$ repetitions for each sample size. The mean absolute percentage deviation between the estimate and the theoretical asymptotic value (MAPD) is used as an indicator of convergence. Given k generated samples, MAPD is calculated as

$$\text{MAPD}(\theta) = \frac{1}{k} \sum_{j=1}^k \left| \frac{\hat{\theta}_j - \theta}{\theta} \right|,$$

where θ is the theoretical asymptotic value of estimate and $\hat{\theta}_j$ is the parameter estimate from the j th sample.

It must be noted that depending on the sample size and the value of η there are occasions when it is not possible to find an eligible estimate for α since the sample value for η is negative. To achieve the planned 3000 estimates for parameter sets with $\eta \approx 0.03$ and small sample sizes there is up to 2.75 times more samples needed altogether. For parameter sets with $\eta \approx 0.3$ the number of failed estimates is considerably smaller. Figure A1 in Appendix A.4 shows the ratio of the number of failed samples to the number of planned estimates from our simulations.

Figure 1 shows resulting MAPD values for parameters depending on the logarithm of sample size. Red tones represent $\eta \approx 0.03$ group and blue tones $\eta \approx 0.3$. Lighter tone is for smaller α and dark for moderate size α . Line types represent different sign and magnitude combinations of $\sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$.

As one can see, the convergence for diagonal elements of Σ is fastest and shows no remarkable differences for different parameter sets. However, MAPD for $\hat{\sigma}_{11}$ and $\hat{\sigma}_{22}$ is slightly bigger for sample sizes between 100 to 1000 if η is small.

For the off-diagonal element of Σ , the convergence speed varies with different parameter sets. If the absolute value of $\sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$ is big, the pattern is actually similar to the convergence of diagonal elements, but for small values, the convergence is slower and MAPD can be close to or exceed value 1 even with sample sizes near 1000.

Most recognizable distinction is seen with the MAPD values of the estimates of α and it is due to the η value. If η is small ($\eta \approx 0.03$), the MAPD value retains almost constantly high level with sample sizes from 100 up to 10,000, then starts to decrease with larger samples, but is still bigger than the MAPD value for the group with $\eta \approx 0.3$ with sample sizes near 10^6 . For $\eta \approx 0.3$, the convergence is faster and reaches smaller values.

Another aspect is the stability of estimates: with smaller η ($\eta \approx 0.03$), the estimates are also more unstable for sample sizes under 10^4 than the estimates corresponding to $\eta \approx 0.3$. Other variations (low and moderate α elements and different $\sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$ values) do not seem to have an effect on stability of the estimates.

Next we compare asymptotic covariance matrices of the estimators found in Theorems 2 and 3 with their estimates from simulated samples. Parameter sets 1 and 9 from Table A1 in Appendix A.3 were used. The former (set 1) corresponds to safe $\eta \approx 0.3$ value and the latter (set 9) to problematic $\eta \approx 0.03$ value and these parameter sets have the same value of α . For both cases $\hat{\alpha}$ and $\hat{\Sigma}$ were calculated 1000 times from samples of size 500. These parameter estimates were used to estimate D_{as}^{α} and D_{as}^{Σ} . For the simulations, the number of degrees of freedom $\nu = 5$ was used, as it is the minimal value for calculating the asymptotic covariance matrices.

Figure 1. Convergence speed of parameter estimates.

For the parameter set 9 with $\eta \approx 0.03$, the theoretical covariance matrix for $\sqrt{n}\hat{\alpha}$ is

$$\mathbf{D}_{as}^{\alpha} = \begin{pmatrix} 280.18 & 503.84 \\ 503.84 & 993.65 \end{pmatrix}$$

and its estimate

$$\widehat{\mathbf{D}}_{as}^{\alpha} = \begin{pmatrix} 230.93 & 372.54 \\ 372.54 & 756.02 \end{pmatrix}.$$

As the theoretical expected value is $\sqrt{n}\alpha \approx (11.2, 22.4)'$ we can say that the variance of estimates is high in this case, but the theoretical and estimated covariance matrices are in agreement, although the theoretical values are higher. For the estimator $\sqrt{n}\text{vec}\hat{\Sigma}$, the results are similar. In this case, the theoretical covariance matrix is

$$\mathbf{D}_{as}^{\Sigma} = \begin{pmatrix} 200.00 & -460.00 & -460.00 & 1213.50 \\ -460.00 & 1291.25 & 1291.25 & -3864.00 \\ -460.00 & 1291.25 & 1291.25 & -3864.00 \\ 1213.50 & -3864.00 & -3864.00 & 14112.00 \end{pmatrix}$$

and the corresponding estimate:

$$\widehat{\mathbf{D}}_{as}^{\Sigma} = \begin{pmatrix} 182.98 & -425.81 & -425.81 & 1044.01 \\ -425.81 & 1191.52 & 1191.52 & -3293.59 \\ -425.81 & 1191.52 & 1191.52 & -3293.59 \\ 1044.01 & -3293.59 & -3293.59 & 11819.03 \end{pmatrix}.$$

For the parameter set with $\eta \approx 0.3$, the same applies: theoretical and estimated covariance matrices are close while theoretical values are higher. Here the theoretical covariance matrix for $\sqrt{n}\hat{\alpha}$ is

$$\mathbf{D}_{as}^{\alpha} = \begin{pmatrix} 12.11 & 13.01 \\ 13.01 & 21.38 \end{pmatrix}$$

and the estimate

$$\widehat{\mathbf{D}}_{as}^{\alpha} = \begin{pmatrix} 9.59 & 9.23 \\ 9.23 & 14.71 \end{pmatrix}.$$

As these two parameter sets have the same value of α , we can see that for the second case the estimates vary much less. For $\eta \approx 0.3$, the theoretical covariance matrix for $\sqrt{n}\text{vec}\hat{\Sigma}$ is

$$\mathbf{D}_{as}^{\Sigma} = \begin{pmatrix} 32.00 & -35.20 & -35.20 & 45.04 \\ -35.20 & 48.20 & 48.20 & -70.40 \\ -35.20 & 48.20 & 48.20 & -70.40 \\ 45.04 & -70.40 & -70.40 & 128.00 \end{pmatrix}$$

and its estimate

$$\widehat{\mathbf{D}}_{as}^{\Sigma} = \begin{pmatrix} 29.14 & -30.29 & -30.29 & 36.03 \\ -30.29 & 39.30 & 39.30 & -54.62 \\ -30.29 & 39.30 & 39.30 & -54.62 \\ 36.03 & -54.62 & -54.62 & 97.27 \end{pmatrix}.$$

Figure 2 presents scatterplots for estimates of α_1 and α_2 with theoretical 95% confidence ellipses. On the left panel ($\eta \approx 0.03$), the estimates of α are more scattered as compared to the right panel, where $\eta \approx 0.3$, as expected. In addition, the placing of points does not correspond to an ellipse on the left panel, indicating the asymptotics does not work with sample size $n = 500$ for distribution with $\eta \approx 0.03$. At the same time the result for $\eta \approx 0.3$ can be considered good.

Figure 2. Comparison of estimates and theoretical confidence region: α .

Similar plots for the elements of $\hat{\Sigma}$ are presented in Figure 3. Here the estimates and ellipses again fit well for $\eta \approx 0.3$ case. For $\eta \approx 0.03$ the fit is also better compared to estimates of α in the previous figure.

Figure 3. Comparison of estimates and theoretical confidence region: Σ .

5. Skew t -Copula

5.1. Definition

The skew t -copula is introduced in Kollo and Pettere [23] (Chapter 15). When the location parameter $\mu = \mathbf{0}$, the copula has the form

$$C_{p,\nu}(u_1, \dots, u_p; \Sigma, \alpha) = F_{p,\nu}(F_{1,\nu}^{-1}(u_1; \sigma_{11}, \alpha_1^*), \dots, F_{1,\nu}^{-1}(u_p; \sigma_{pp}, \alpha_p^*); \Sigma, \alpha),$$

where $F_{1,\nu}^{-1}(u_i; \sigma_{ii}, \alpha_i^*)$ denotes the inverse of the univariate skew $t_{1,\nu}$ -distribution function and $F_{p,\nu}$ is the distribution function of the p -variate skew $t_{p,\nu}$ -distribution. The arguments σ_{ii} are the main diagonal elements of the covariance matrix Σ and α_i^* can be derived by the following simplification of (7):

$$\alpha_i^* = \frac{\sigma_{ii}^{-1} \delta_i}{\sqrt{1 - \sigma_{ii}^{-1} \delta_i^2}}, \quad i = 1, \dots, p. \quad (20)$$

Note that α_i^* -s are not equal to the components α_i of the parameter vector α of the multivariate distribution. The corresponding copula density is

$$c_{p,\nu}(\mathbf{u}; \Sigma, \alpha) = \frac{f_{p,\nu}\left((F_{1,\nu}^{-1}(u_1; \sigma_{11}, \alpha_1^*), \dots, F_{1,\nu}^{-1}(u_p; \sigma_{pp}, \alpha_p^*))'; \Sigma, \alpha\right)}{\prod_{i=1}^p f_{1,\nu}\left(F_{1,\nu}^{-1}(u_i; \sigma_{ii}, \alpha_i^*); \sigma_{ii}, \alpha_i^*\right)},$$

where $f_{p,\nu}(\mathbf{u}; \Sigma, \alpha)$ is the density function of $St_{p,\nu}$ and the inverse functions $F_{1,\nu}^{-1}(u_i; \sigma_{ii}, \alpha_i^*)$ are as before.

In the standard case when $\mu = \mathbf{0}$ and the covariance matrix Σ is equal to the correlation matrix $\mathbf{R}$, the density of the skew t -copula is of the form

$$c_{p,\nu}(\mathbf{u}; \mathbf{R}, \alpha) = \frac{f_{p,\nu}\left((F_{1,\nu}^{-1}(u_1; \alpha_1^*), \dots, F_{1,\nu}^{-1}(u_p; \alpha_p^*))'; \mathbf{R}, \alpha\right)}{\prod_{i=1}^p f_{1,\nu}\left(F_{1,\nu}^{-1}(u_i; \alpha_i^*); \alpha_i^*\right)}, \quad (21)$$

where $F_{1,\nu}(u_i; \alpha_i^*)$, $i = 1, \dots, p$ are the distribution functions of the univariate standard skew $t_{1,\nu}$ -distribution. Density (21) is the most often used version of the skew t -copula density in data analysis.

As the numerator of the copula density consists of the density of multivariate skew t -distribution, the resulting copula is distribution function of a skewed multivariate distribution. When we want to apply skew t -copula in data analysis to join (possibly) differently distributed marginals into a multivariate distribution, we have to take into account the correlations between marginals. Pearson correlation matrix can easily be expressed through the parameter Σ . Unfortunately Pearson correlations are not invariant under copula transformations (inverse distribution functions). We have to switch over to rank correlations. Fortunately for continuous elliptical distributions, Kendall's τ can be expressed through the linear correlation (see Lindskog et al. [35])

$$\tau(X_i, X_j) = \frac{2}{\pi} \arcsin(r_{ij}),$$

so we can switch over from linear correlations to rank correlations and vice versa.

5.2. Examples of Bivariate Skew t -Copula Densities

In the following we present some examples of two-dimensional copula densities to compare Gaussian, skew-normal and skew t -copulas using contour plots. First we present the formulas for the before mentioned copulas in bivariate case. The bivariate Gaussian copula has one parameter ρ and using the general formula for multivariate Gaussian copula from Cherubini et al. [36] (p. 148) we can express bivariate density as

$$c_N(u_1, u_2; \rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2(z_1^2 + z_2^2) - 2\rho z_1 z_2}{2(1-\rho^2)}\right),$$

where $z_i = \Phi^{-1}(u_i)$ and $u_i \in [0, 1]$, $i = 1, 2$. Skew-normal copula has in addition to ρ parameter vector α and following Karik et al. [37] its density in bivariate case is of form

$$c_{SN}(u_1, u_2; \rho, \alpha) = \frac{1}{2\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2(w_1^2 + w_2^2) - 2\rho w_1 w_2}{2(1-\rho^2)}\right) \frac{\Phi(\alpha'(w_1, w_2)')}{\Phi(\alpha_1^* w_1) \Phi(\alpha_2^* w_2)},$$

where α_i^* are given by (20) and $w_i = F^{-1}(u_i; \alpha_i^*)$ denotes inverse of the univariate skew-normal distribution function with parameter α_i^* , as $u_i \in [0, 1]$, $i = 1, 2$. The bivariate skew

t -copula has parameters ρ , α and ν and using the general Formula (21) we get the bivariate density as

$$c_{2,\nu}(u_1, u_2; \rho; \alpha) = \frac{\nu}{4\sqrt{1-\rho^2}} \left(\frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})} \right)^2 \frac{\left(1 + \frac{Q}{(1-\rho^2)\nu}\right)^{-\frac{\nu+2}{2}} F_{2+\nu}(\alpha'(q_1, q_2)' \sqrt{\frac{2+\nu}{Q+\nu}})}{\prod_{i=1}^2 (1 + \frac{q_i^2}{\nu})^{-\frac{\nu+1}{2}} \prod_{i=1}^2 F_{1+\nu}(\alpha_i^* q_i \sqrt{\frac{1+\nu}{q_i^2+\nu}})},$$

where $q_i = F_{1,\nu}^{-1}(u_i; \alpha_i^*)$, $u_i \in [0, 1]$, $i = 1, 2$, denotes inverse of the univariate skew t distribution function with parameter α_i^* , $Q = q_1^2 - 2q_1q_2\rho + q_2^2$ and F_k is the distribution function of the univariate t_k -distribution.

For a first example, we consider here the case where parameter ρ is zero. Gaussian copula reduces to the product copula when $\rho = 0$. If the corresponding parameter for skew-normal or skew t -copula equals zero, the resulting copula is not a product copula, this can be seen in Figure 4.

Figure 4. Examples of skew-normal and skew t -copulas when $\rho = 0$.

In the figures, contours on the left correspond to skew-normal and on the right to skew t -copulas with different α values (values are given in the header of the figure, note that the same number of degrees of freedom $\nu = 4$ is used for all examples here). Contours are given with step size of 0.2 starting from 0 and ending with 3. Dark blue, yellow and dark red colors represent the lowest, middle and highest contour levels for chosen scale.

If $\rho = 0$ and both components of α are positive, the densities of skew-normal and skew t -copulas tend to show negative dependence that grows stronger if α_i are bigger (see first and second row of Figure 4). If one α_i is negative, the copula represents positive dependence, but the direction of dependence also depends on the magnitude of α_i , i.e., the direction is not so clear for α_i close to zero. When comparing contours in the second and third row we see that adding negative sign to the second component of α results in rotation of the density by 90° to the right. Rotation to the left would be the result if first element of α is negative and the other positive. We note that the two skew copulas give the product copula if in addition to $\rho = 0$ one element in the α vector is also zero.

As a second example, we consider the case when $\rho \neq 0$. For Gaussian copula, the parameter ρ gives the direction and strength of dependence between marginals. For the two skewed copulas, the additional parameter α adds complexity to this relation. We can conclude from the Figures S1–S4. that if both α_i are positive and ρ is negative as given here, the dependence for skewed copulas can be stronger as compared to Gaussian copula. For positive ρ , the opposite may happen: the skew-normal and skew t represent weaker dependence as compared to Gaussian copula or the dependence changes direction, see Figure S2 for $\rho = 0.2$, for example. The effect of parameter α is more notable if its elements have high absolute values.

6. Conclusions

Our main theoretical results, presented in Theorems 2 and 3, assure the asymptotic normality for the moment based parameter estimators of the two-parameter multivariate skew t -distribution. We can make the following observations based on the simulations used to illustrate the analytical results. First, the behaviour of the estimates of α depends on the value of η : if the value of η is close to zero (we used value 0.03), the estimates of α are unstable and the convergence is really slow, MAPD values are greater than 0.5 with sample sizes up to 10^4 . For parameter sets with η value 0.3, the results are more stable and, for the aforementioned sample size, the MAPD values were already less than 0.1. Second, it occurs that the combination of small sample size and low η value can result in ineligible estimates of α . As a third point, we note that the estimates of Σ are stable and the convergence is faster than the convergence of the estimates of α . Furthermore, the value of η does not affect the estimates of Σ . As the two considered values of η were chosen arbitrarily, its effect needs further investigation. Finally we also define skew t -copula, a powerful and flexible tool that can be used to model dependent data with differently distributed marginals in fields like finance, actuarial science and more.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/sym13061059/s1>, Figure S1: Gaussian copula and examples of skew-normal and skew t -copulas when the components of α have small values, Figure S2: Gaussian copula and examples of skew-normal and skew t -copulas when the components of α have relatively high values, Figure S3: Gaussian copula and examples of skew-normal and skew t -copulas when the components of α have different magnitude, Figure S4: Gaussian copula and examples of skew-normal and skew t -copulas when components of α are equal.

Author Contributions: Conceptualization, T.K. and M.K.; methodology, T.K., M.K. and A.S.; simulations and data analysis, A.S.; writing—original draft preparation, T.K.; writing—review and editing, T.K., M.K. and A.S.; visualization, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Estonian Research Council (grant PRG 1197).

Acknowledgments: The authors are thankful to the editor and three anonymous referees for their constructive feedback that certainly improved the quality of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Proof of Lemma 2

Proof. We have to express the third and fourth central moments through raw moments.

$$\begin{aligned}
 \bar{m}_3(\mathbf{x}) &= E[(\mathbf{x} - E\mathbf{x})(\mathbf{x} - E\mathbf{x})' \otimes (\mathbf{x} - E\mathbf{x})] \\
 &= E[(\mathbf{x}\mathbf{x}' - E\mathbf{x}\mathbf{x}' - \mathbf{x}E\mathbf{x} + E\mathbf{x}E\mathbf{x}) \otimes (\mathbf{x} - E\mathbf{x})] \\
 &= E[\mathbf{x}\mathbf{x}' \otimes \mathbf{x} - \mathbf{x}\mathbf{x}' \otimes E\mathbf{x} - E\mathbf{x}\mathbf{x}' \otimes \mathbf{x} + E\mathbf{x}\mathbf{x}' \otimes E\mathbf{x} \\
 &\quad - \mathbf{x}E\mathbf{x}' \otimes \mathbf{x} + \mathbf{x}E\mathbf{x}' \otimes E\mathbf{x} + E\mathbf{x}E\mathbf{x}' \otimes \mathbf{x} - E\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x}] \\
 &= E(\mathbf{x}\mathbf{x}' \otimes \mathbf{x}) - E(\mathbf{x}\mathbf{x}') \otimes E\mathbf{x} - E\mathbf{x}E(\mathbf{x}' \otimes \mathbf{x}) + E\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x} \\
 &\quad - E(\mathbf{x}E\mathbf{x}' \otimes \mathbf{x}) + E\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x} \\
 &= m_3(\mathbf{x}) - m_2(\mathbf{x}) \otimes E\mathbf{x} - E\mathbf{x} \otimes m_2(\mathbf{x}) + 2E\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x} - E\mathbf{x}' \otimes E(\mathbf{x} \otimes \mathbf{x}) \\
 &= m_3(\mathbf{x}) - m_2(\mathbf{x}) \otimes E\mathbf{x} - E\mathbf{x} \otimes m_2(\mathbf{x}) + 2E\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x} - E\mathbf{x}' \otimes \text{vec } m_2(\mathbf{x}).
 \end{aligned}$$

Finally

$$\bar{m}_3(\mathbf{x}) = m_3(\mathbf{x}) - (\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(m_2(\mathbf{x}) \otimes E\mathbf{x}) - \text{vec } m_2(\mathbf{x})E\mathbf{x}' + 2E\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x}.$$

Derivation of the fourth central moment is longer. We can use the intermediate expression from derivation of $\bar{m}_3(\mathbf{x})$ after the third equality sign above

$$\begin{aligned}
 \bar{m}_4(\mathbf{x}) &= E[(\mathbf{x} - E\mathbf{x})(\mathbf{x} - E\mathbf{x})' \otimes (\mathbf{x} - E\mathbf{x})(\mathbf{x} - E\mathbf{x})'] \\
 &= E\{[\mathbf{x}\mathbf{x}' \otimes \mathbf{x} - \mathbf{x}\mathbf{x}' \otimes E\mathbf{x} - E\mathbf{x}\mathbf{x}' \otimes \mathbf{x} + E\mathbf{x}\mathbf{x}' \otimes E\mathbf{x} \\
 &\quad - \mathbf{x}E\mathbf{x}' \otimes \mathbf{x} + \mathbf{x}E\mathbf{x}' \otimes E\mathbf{x} + E\mathbf{x}E\mathbf{x}' \otimes \mathbf{x} - E\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x}] \otimes \mathbf{x}'\} \\
 &\quad - E\{[\mathbf{x}\mathbf{x}' \otimes \mathbf{x} - \mathbf{x}\mathbf{x}' \otimes E\mathbf{x} - E\mathbf{x}\mathbf{x}' \otimes \mathbf{x} + E\mathbf{x}\mathbf{x}' \otimes E\mathbf{x} \\
 &\quad - \mathbf{x}E\mathbf{x}' \otimes \mathbf{x} + \mathbf{x}E\mathbf{x}' \otimes E\mathbf{x} + E\mathbf{x}E\mathbf{x}' \otimes \mathbf{x} - E\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x}] \otimes E\mathbf{x}'\} \\
 &= m_4(\mathbf{x}) - \bar{m}_3(\mathbf{x}) \otimes E\mathbf{x}' - E(\mathbf{x}\mathbf{x}' \otimes E\mathbf{x} \otimes \mathbf{x}') - E(E\mathbf{x}\mathbf{x}' \otimes \mathbf{x} \otimes \mathbf{x}') \\
 &\quad + E(E\mathbf{x}\mathbf{x}' \otimes E\mathbf{x} \otimes \mathbf{x}') - E(\mathbf{x}E\mathbf{x}' \otimes \mathbf{x} \otimes \mathbf{x}') + E(\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x} \otimes \mathbf{x}') \\
 &\quad + E\mathbf{x}E\mathbf{x}' \otimes m_2(\mathbf{x}) - E\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x} \otimes E\mathbf{x}'.
 \end{aligned}$$

The expectations in the last equality can be presented through moments when we use the property of the Kronecker product for vectors $\mathbf{ab}' = \mathbf{a} \otimes \mathbf{b}' = \mathbf{b}' \otimes \mathbf{a}$

$$\begin{aligned}
 E(\mathbf{x}\mathbf{x}' \otimes E\mathbf{x}\mathbf{x}') &= E(\mathbf{x} \otimes \mathbf{x}' \otimes E\mathbf{x} \otimes \mathbf{x}') = E(\mathbf{x}' \otimes \mathbf{x} \otimes \mathbf{x}' \otimes E\mathbf{x}') = m_3'(\mathbf{x}) \otimes E\mathbf{x}; \\
 E(E\mathbf{x}\mathbf{x}' \otimes \mathbf{x}\mathbf{x}') &= E\mathbf{x} \otimes m_3'(\mathbf{x}); \\
 E(E\mathbf{x}\mathbf{x}' \otimes E\mathbf{x}\mathbf{x}') &= E(E\mathbf{x} \otimes \mathbf{x}' \otimes \mathbf{x}' \otimes E\mathbf{x}) = E\mathbf{x} \otimes \text{vec}' m_2(\mathbf{x}) \otimes E\mathbf{x}; \\
 E(\mathbf{x}E\mathbf{x}' \otimes \mathbf{x}\mathbf{x}') &= E(\mathbf{x} \otimes E\mathbf{x}' \otimes \mathbf{x} \otimes \mathbf{x}') = E(E\mathbf{x}' \otimes \mathbf{x} \otimes \mathbf{x}' \otimes \mathbf{x}) = E\mathbf{x}' \otimes m_3(\mathbf{x}); \\
 E(\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x}\mathbf{x}') &= E(\mathbf{x} \otimes E\mathbf{x}' \otimes E\mathbf{x} \otimes \mathbf{x}') = E(E\mathbf{x}' \otimes \mathbf{x} \otimes \mathbf{x}' \otimes E\mathbf{x}') \\
 &= E\mathbf{x}' \otimes m_2(\mathbf{x}) \otimes E\mathbf{x}.
 \end{aligned}$$

After plugging in the expression of $\bar{m}_3(\mathbf{x})$ we have

$$\begin{aligned}
 \bar{m}_4(\mathbf{x}) &= m_4(\mathbf{x}) - m_3(\mathbf{x}) \otimes E\mathbf{x}' + m_2(\mathbf{x}) \otimes E\mathbf{x}E\mathbf{x}' + E\mathbf{x} \otimes m_2(\mathbf{x}) \otimes E\mathbf{x}' \\
 &\quad + E\mathbf{x}' \otimes \text{vec } m_2(\mathbf{x}) \otimes E\mathbf{x}' - 3E\mathbf{x}E\mathbf{x}' \otimes E\mathbf{x}E\mathbf{x}' - m_3'(\mathbf{x}) \otimes E\mathbf{x} - E\mathbf{x} \otimes m_3'(\mathbf{x}) \\
 &\quad + E\mathbf{x} \otimes \text{vec}' m_2(\mathbf{x}) \otimes E\mathbf{x} - E\mathbf{x}' \otimes m_3(\mathbf{x}) + E\mathbf{x}' \otimes m_2(\mathbf{x}) \otimes E\mathbf{x} \\
 &\quad + E\mathbf{x}E\mathbf{x}' \otimes m_2(\mathbf{x}).
 \end{aligned}$$

Applying notation in Lemma 2

$$\begin{aligned}\mathbf{A}(\mathbf{x}) &= \mathbf{E}\mathbf{x}' \otimes m_2(\mathbf{x}) \otimes \mathbf{E}\mathbf{x}; \\ \mathbf{B}(\mathbf{x}) &= \mathbf{E}\mathbf{x}' \otimes \text{vec } m_2(\mathbf{x}) \otimes \mathbf{E}\mathbf{x}'; \\ \mathbf{C}(\mathbf{x}) &= m_3(\mathbf{x}) \otimes \mathbf{E}\mathbf{x}'\end{aligned}$$

we get the statement after using property 5 of the Kronecker product

$$\begin{aligned}\bar{m}_4(\mathbf{x}) &= m_4(\mathbf{x}) - \mathbf{C}(\mathbf{x})(\mathbf{I}_{p^2} + \mathbf{K}_{p,p}) - (\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(\mathbf{C}(\mathbf{x}))' - 3\mathbf{E}\mathbf{x}\mathbf{E}\mathbf{x}' \otimes \mathbf{E}\mathbf{x}\mathbf{E}\mathbf{x}' \\ &\quad + \mathbf{A}(\mathbf{x}) + \mathbf{B}(\mathbf{x}) + (\mathbf{A}(\mathbf{x}) + \mathbf{B}(\mathbf{x}))' + m_2(\mathbf{x}) \otimes \mathbf{E}\mathbf{x}\mathbf{E}\mathbf{x}' + \mathbf{E}\mathbf{x}\mathbf{E}\mathbf{x}' \otimes m_2(\mathbf{x}).\end{aligned}$$

□

Appendix A.2. Proof of Theorem 3

Proof. Let us start from Expression (18) for $\hat{\mathbf{a}}$:

$$\hat{\mathbf{a}} = \frac{\hat{\Sigma}^{-1}\bar{\mathbf{x}}}{\sqrt{M}},$$

where M is given in (17), which, taking (15) into account, can also be expressed as follows:

$$M = \frac{\nu}{\pi} \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 - \bar{\mathbf{x}}' \hat{\Sigma}^{-1} \bar{\mathbf{x}}.$$

Let us now derive the needed derivatives step by step.

$$\frac{d\hat{\mathbf{a}}}{d\bar{\mathbf{x}}} = \frac{d(\hat{\Sigma}^{-1}\bar{\mathbf{x}}M^{-\frac{1}{2}})}{d\bar{\mathbf{x}}} = M^{-\frac{1}{2}} \frac{d(\hat{\Sigma}^{-1}\bar{\mathbf{x}})}{d\bar{\mathbf{x}}} + \hat{\Sigma}^{-1}\bar{\mathbf{x}} \frac{dM^{-\frac{1}{2}}}{d\bar{\mathbf{x}}}. \quad (\text{A1})$$

Let us find the two derivatives in (A1).

(a) The first derivative can be expressed as

$$\frac{d(\hat{\Sigma}^{-1}\bar{\mathbf{x}})}{d\bar{\mathbf{x}}} = \hat{\Sigma}^{-1} + (\bar{\mathbf{x}}' \otimes \mathbf{I}_p) \frac{d\hat{\Sigma}^{-1}}{d\bar{\mathbf{x}}}. \quad (\text{A2})$$

Now the derivative $\frac{d\hat{\Sigma}^{-1}}{d\bar{\mathbf{x}}}$ can be found as

$$\begin{aligned}\frac{d\hat{\Sigma}^{-1}}{d\bar{\mathbf{x}}} &= \frac{d\hat{\Sigma}^{-1}}{d\hat{\Sigma}} \frac{d\hat{\Sigma}}{d\bar{\mathbf{x}}} = -(\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) \frac{d(\frac{\nu-2}{\nu}(\mathbf{S} + \bar{\mathbf{x}}\bar{\mathbf{x}}'))}{d\bar{\mathbf{x}}} \\ &= -\frac{\nu-2}{\nu} (\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) (\bar{\mathbf{x}} \otimes \mathbf{I}_p + \mathbf{I}_p \otimes \bar{\mathbf{x}}) \\ &= -\frac{\nu-2}{\nu} (\hat{\Sigma}^{-1}\bar{\mathbf{x}} \otimes \hat{\Sigma}^{-1} + \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}\bar{\mathbf{x}}).\end{aligned} \quad (\text{A3})$$

The Equations (A2) and (A3) imply that

$$\begin{aligned}\frac{d(\hat{\Sigma}^{-1}\bar{\mathbf{x}})}{d\bar{\mathbf{x}}} &= \hat{\Sigma}^{-1} - (\bar{\mathbf{x}}' \otimes \mathbf{I}_p) \frac{\nu-2}{\nu} (\hat{\Sigma}^{-1}\bar{\mathbf{x}} \otimes \hat{\Sigma}^{-1} + \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}\bar{\mathbf{x}}) \\ &= \hat{\Sigma}^{-1} - \frac{\nu-2}{\nu} (\bar{\mathbf{x}}' \hat{\Sigma}^{-1} \bar{\mathbf{x}} \hat{\Sigma}^{-1} + \bar{\mathbf{x}}' \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1} \bar{\mathbf{x}}) \\ &= \hat{\Sigma}^{-1} - \frac{\nu-2}{\nu} (\bar{\mathbf{x}}' \hat{\Sigma}^{-1} \bar{\mathbf{x}} \hat{\Sigma}^{-1} + \hat{\Sigma}^{-1} \bar{\mathbf{x}} \bar{\mathbf{x}}' \hat{\Sigma}^{-1}).\end{aligned}$$

(b) The second derivative in (A1) is

$$\begin{aligned}
 \frac{dM^{-\frac{1}{2}}}{d\bar{\mathbf{x}}} &= \frac{dM^{-\frac{1}{2}}}{dM} \frac{dM}{d\bar{\mathbf{x}}} = -\frac{1}{2}M^{-\frac{3}{2}} \frac{d(-\bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}})}{\bar{\mathbf{x}}} \\
 &= \frac{1}{2}M^{-\frac{3}{2}} \left(\bar{\mathbf{x}}'\hat{\Sigma}^{-1} + \bar{\mathbf{x}} \frac{d(\hat{\Sigma}^{-1}\bar{\mathbf{x}})}{d\bar{\mathbf{x}}} \right) \\
 &= \frac{1}{2}M^{-\frac{3}{2}} \left(\bar{\mathbf{x}}'\hat{\Sigma}^{-1} + \bar{\mathbf{x}}'\hat{\Sigma}^{-1} - \bar{\mathbf{x}}' \frac{\nu-2}{\nu} (\bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}}\hat{\Sigma}^{-1} + \hat{\Sigma}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}'\hat{\Sigma}^{-1}) \right) \\
 &= \frac{1}{2}M^{-\frac{3}{2}} \left(2\bar{\mathbf{x}}'\hat{\Sigma}^{-1} - \frac{\nu-2}{\nu} (\bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}}\hat{\Sigma}^{-1} + \hat{\Sigma}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}'\hat{\Sigma}^{-1}) \right) \\
 &= M^{-\frac{3}{2}} \left(\left(1 - \frac{\nu-2}{\nu} \bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}} \right) \bar{\mathbf{x}}'\hat{\Sigma}^{-1} \right).
 \end{aligned}$$

Now, applying the results of parts (a) and (b) into Equation (A1), yields

$$\begin{aligned}
 \frac{d\hat{\alpha}}{d\bar{\mathbf{x}}} &= M^{-\frac{1}{2}} \left(\hat{\Sigma}^{-1} - \frac{\nu-2}{\nu} (\bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}}\hat{\Sigma}^{-1} + \hat{\Sigma}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}'\hat{\Sigma}^{-1}) \right) \\
 &\quad + \hat{\Sigma}^{-1}\bar{\mathbf{x}}M^{-\frac{3}{2}} \left(\left(1 - \frac{\nu-2}{\nu} \bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}} \right) \bar{\mathbf{x}}'\hat{\Sigma}^{-1} \right) \\
 &= M^{-\frac{1}{2}} \left[\left(\hat{\Sigma}^{-1} - \frac{\nu-2}{\nu} (\bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}}\hat{\Sigma}^{-1} + \hat{\Sigma}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}'\hat{\Sigma}^{-1}) \right) \right. \\
 &\quad \left. + M^{-1} \left(\left(1 - \frac{\nu-2}{\nu} \bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}} \right) \hat{\Sigma}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}'\hat{\Sigma}^{-1} \right) \right] \\
 &= M^{-\frac{1}{2}} \left[\left(1 - \frac{\nu-2}{\nu} \bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}} \right) \hat{\Sigma}^{-1} \right. \\
 &\quad \left. + M^{-1} \left(\left(1 - \frac{\nu-2}{\nu} (\bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}} + M) \right) \hat{\Sigma}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}'\hat{\Sigma}^{-1} \right) \right].
 \end{aligned}$$

Taking the derivative of $\hat{\alpha}$ by $\text{vec } \mathbf{S}$ is similar, but simpler. First of all, following Equality (A1), we can write

$$\frac{d\hat{\alpha}}{d\text{vec } \mathbf{S}} = M^{-\frac{1}{2}} \frac{d(\hat{\Sigma}^{-1}\bar{\mathbf{x}})}{d\text{vec } \mathbf{S}} + \hat{\Sigma}^{-1}\bar{\mathbf{x}} \frac{dM^{-\frac{1}{2}}}{d\text{vec } \mathbf{S}}.$$

Let us find the two derivatives in the last sum.

(a) The first derivative can be expressed as

$$\frac{d(\hat{\Sigma}^{-1}\bar{\mathbf{x}})}{d\text{vec } \mathbf{S}} = (\bar{\mathbf{x}}' \otimes \mathbf{I}_p) \frac{d\hat{\Sigma}^{-1}}{d\text{vec } \mathbf{S}} = -\frac{\nu-2}{\nu} (\bar{\mathbf{x}}'\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}),$$

(b) The second derivative is

$$\begin{aligned}
 \frac{dM^{-\frac{1}{2}}}{d\text{vec } \mathbf{S}} &= \frac{dM^{-\frac{1}{2}}}{dM} \frac{dM}{d\text{vec } \mathbf{S}} = -\frac{1}{2}M^{-\frac{3}{2}} \frac{d(-\bar{\mathbf{x}}'\hat{\Sigma}^{-1}\bar{\mathbf{x}})}{d\text{vec } \mathbf{S}} \\
 &= -\frac{1}{2}M^{-\frac{3}{2}} (\bar{\mathbf{x}}' \otimes \bar{\mathbf{x}}') (\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) \frac{\nu-2}{\nu} \\
 &= -\frac{\nu-2}{2\nu} M^{-\frac{3}{2}} (\bar{\mathbf{x}}'\hat{\Sigma}^{-1} \otimes \bar{\mathbf{x}}'\hat{\Sigma}^{-1}).
 \end{aligned}$$

The full expression of the derivative $\frac{d\hat{\alpha}}{d\text{vec } \mathbf{S}}$ is

$$\frac{d\hat{\alpha}}{d\text{vec } \mathbf{S}} = -\frac{\nu-2}{\nu} M^{-\frac{1}{2}} \left[(\bar{\mathbf{x}}' \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) + \frac{1}{2} M^{-1} (\hat{\Sigma}^{-1} \bar{\mathbf{x}} \bar{\mathbf{x}}' \hat{\Sigma}^{-1} \otimes \bar{\mathbf{x}}' \hat{\Sigma}^{-1}) \right].$$

□

Appendix A.3. Parameter Sets

Table A1. Parameter sets for simulations.

	η	α_1	α_2	σ_{11}	σ_{22}	σ_{12}	$\sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$
Set 1	0.3030	0.5	1	2.0	4.00	−2.20	−0.78
Set 2	0.3030	0.5	1	1.0	2.20	−0.15	−0.10
Set 3	0.2999	0.5	1	1.5	1.00	0.96	0.78
Set 4	0.3030	0.5	1	3.0	1.35	0.20	0.10
Set 5	0.3049	3.0	2	0.6	1.50	−0.76	−0.80
Set 6	0.3049	3.0	2	0.2	0.30	−0.06	−0.24
Set 7	0.3067	3.0	2	0.1	0.10	0.08	0.80
Set 8	0.3012	3.0	2	0.2	0.10	0.01	0.07
Set 9	0.0305	0.5	1	5.0	42.00	−11.50	−0.79
Set 10	0.0301	0.5	1	28.0	28.00	−2.80	−0.10
Set 11	0.0300	0.5	1	16.0	16.00	12.30	0.77
Set 12	0.0298	0.5	1	15.0	25.00	3.80	0.20
Set 13	0.0301	3.0	2	7.0	22.00	−9.90	−0.80
Set 14	0.0305	3.0	2	3.0	3.00	−0.60	−0.20
Set 15	0.0301	3.0	2	1.1	2.00	1.19	0.80
Set 16	0.0302	3.0	2	2.2	2.00	0.36	0.17

Appendix A.4. Samples with Failed Estimates

Figure A1. Ratio of the number of failed samples to the number of planned estimates from our simulations for all parameter sets.

References

1. Azzalini, A.; Valle, A.D. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726. [\[CrossRef\]](#)
2. Theodossiou, P. Financial data and the skewed generalized t distribution. *Manag. Sci.* **1998**, *44*, 1650–1661. [\[CrossRef\]](#)
3. Arellano-Valle, R.B.; Genton, M.G. Multivariate unified skew-elliptical distributions. *Chil. J. Stat.* **2010**, *1*, 17–31.
4. Azzalini, A.; Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t -distribution. *J. R. Stat. Soc. Ser. (Stat. Methodol.)* **2003**, *65*, 367–389. [\[CrossRef\]](#)
5. Lee, S.X.; McLachlan, G.J. Finite mixtures of multivariate skew t -distributions: some recent and new results. *Stat. Comput.* **2014**, *24*, 181–202. [\[CrossRef\]](#)
6. Marchenko, Y.V. Multivariate Skew-T Distributions in Econometrics and Environmetrics. Ph.D. Thesis, Texas A&M University, College Station, TX, USA, 2010; p. 137.
7. Parisi, A.; Liseo, B. Objective Bayesian analysis for the multivariate skew- t model. *Stat. Methods Appl.* **2018**, *27*, 277–295. [\[CrossRef\]](#)
8. Fernandez, C.; Steel, M. Multivariate Student- t regression models: Pitfalls and inference. *Biometrika* **1999**, *86*, 153–167. [\[CrossRef\]](#)
9. Jones, M.C. On families of distributions with shape parameters. *Int. Stat. Rev.* **2015**, *83*, 175–192. [\[CrossRef\]](#)
10. Adcock, C.; Eling, M.; Loperfido, N. Skewed distributions in finance and actuarial science: A review. *Eur. J. Financ.* **2015**, *21*, 1253–1281. [\[CrossRef\]](#)
11. Ghizzoni, T.; Roth, G.; Rudari, R. Multivariate skew- t approach to the design of accumulation risk scenarios for the flooding hazard. *Adv. Water Resour.* **2010**, *33*, 1243–1255. [\[CrossRef\]](#)
12. Thompson, K.; Shen, Y. Coastal flooding and the multivariate skew- t distribution. In *Skew-Elliptical Distributions and Their Applications: A Journey beyond Normality*; Genton, M.G., Ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2004; Chapter 14, pp. 243–258. [\[CrossRef\]](#)
13. Genton, M.G. (Ed.) *Skew-Elliptical Distributions and Their Applications. A Journey beyond Normality*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2004. [\[CrossRef\]](#)
14. Azzalini, A.; Capitanio, A. *The Skew-Normal and Related Families*; Cambridge University Press: Cambridge, UK, 2014. [\[CrossRef\]](#)
15. Fung, T.; Seneta, E. Tail dependence for two skew t distributions. *Stat. Probab. Lett.* **2010**, *80*, 784–791. [\[CrossRef\]](#)
16. Joe, H.; Li, H. Tail densities of skew-elliptical distributions. *J. Multivar. Anal.* **2019**, *171*, 421–435. [\[CrossRef\]](#)
17. Kollo, T.; Pettere, G.; Valge, M. Tail dependence of skew t -copulas. *Commun. Stat. Simul. Comput.* **2017**, *46*, 1024–1034. [\[CrossRef\]](#)
18. Padoan, S.A. Multivariate extreme models based on underlying skew- t and skew-normal distributions. *J. Multivar. Anal.* **2011**, *102*, 977–991. [\[CrossRef\]](#)
19. Galarza, C.E.; Matos, L.A.; Lachos, V.H. Moments of the doubly truncated selection elliptical distributions with emphasis on the unified multivariate skew- t distribution. *arXiv* **2020**, arXiv:2007.14980.
20. Zhou, T.; He, X. Three-step estimation in linear mixed models with skew- t distributions. *J. Stat. Plan. Inference* **2008**, *138*, 1542–1555. [\[CrossRef\]](#)
21. DiCiccio, T.J.; Monti, A.C. Inferential aspects of the skew exponential power distribution. *J. Am. Stat. Assoc.* **2004**, *99*, 439–450. [\[CrossRef\]](#)
22. Demarta, S.; McNeil, A.J. The t copula and related copulas. *Int. Stat. Rev.* **2005**, *73*, 111–129. [\[CrossRef\]](#)
23. Kollo, T.; Pettere, G. Parameter estimation and application of the multivariate skew t -copula. In *Copula Theory and Its Applications*; Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 289–298. [\[CrossRef\]](#)
24. Magnus, J.R.; Neudecker, H. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd ed.; Wiley: Chichester, UK, 1999.
25. Kollo, T.; von Rosen, D. *Advanced Multivariate Statistics with Matrices*; Springer: Dordrecht, The Netherlands, 2010; [\[CrossRef\]](#)
26. Kotz, S.; Nadarajah, S. *Multivariate t Distributions and Their Applications*; Cambridge University Press: Cambridge, UK, 2004. [\[CrossRef\]](#)
27. Kshirsagar, A.M. Some extensions of the multivariate t -distribution and the multivariate generalization of the distribution of the regression coefficient. *Math. Proc. Camb. Philos. Soc.* **1961**, *57*, 80–85. [\[CrossRef\]](#)
28. Fang, H.B.; Fang, K.T.; Kotz, S. The meta-elliptical distributions with given marginals. *J. Multivar. Anal.* **2002**, *82*, 1–16. [\[CrossRef\]](#)
29. Gupta, A.K. Multivariate skew t -distribution. *Statistics* **2003**, *37*, 359–363. [\[CrossRef\]](#)
30. Kollo, T.; Käärik, M.; Selart, A. Asymptotic normality of estimators for parameters of a multivariate skew-normal distribution. *Commun. Stat. Theory Methods* **2018**, *47*, 3640–3655. [\[CrossRef\]](#)
31. Käärik, M.; Selart, A.; Käärik, E. On Parametrization of multivariate skew-normal distribution. *Commun. Stat. Theory Methods* **2015**, *44*, 1869–1885. [\[CrossRef\]](#)
32. Lee, P.M. *Bayesian Statistics: An Introduction*, 4th ed.; Wiley: Chichester, UK, 2012.
33. Parring, A.M. Computation of asymptotic characteristics of sample functions. *Acta Comment. Univ. Tartu. Math.* **1979**, *492*, 86–90. (In Russian)
34. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*, 3rd ed.; Wiley: New York, NY, USA, 2003.
35. Lindskog, F.; McNeil, A.; Schmock, U. Kendall's tau for elliptical distributions. In *Credit Risk*; Bol, G., Nakhaeizadeh, G., Rachev, S.T., Ridder, T., Vollmer, K.H., Eds.; Physica-Verlag HD: Heidelberg, Germany, 2003; pp. 149–156. [\[CrossRef\]](#)

-
36. Cherubini, U.; Luciano, E.; Vecchiato, W. *Copula Methods in Finance*; John Wiley & Sons: Hoboken, NJ, USA, 2004. [[CrossRef](#)]
 37. Käärik, M.; Selart, A.; Käärik, E. The use of copulas to model conditional expectation for multivariate data. In *ISI World Statistics Congress Proceedings. Bulletin of the ISI 2011 58th WSC, Dublin*; ISI: The Hague, The Netherlands, 2011; pp. 5533–5538.