On moments of folded and truncated multivariate Student-t distributions based on recurrence relations

Christian E. Galarza[1], Tsung-I Lin[2,3], Wan-Lun Wang[4] and Víctor H. Lachos[5]

[1] *Facultad de Ciencias Naturales y Matemáticas, Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador*

[2] *Institute of Statistics, National Chung Hsing University, Taichung 402, Taiwan*

[3] *Department of Public Health, China Medical University, Taichung 404, Taiwan*

[4] *Department of Statistics, Graduate Institute of Statistics and Actuarial Science, Feng Chia University, Taichung 40724, Taiwan*

[5] *Departament of Statistics, University of Connecticut, CT 06269, U.S.A.*

Abstract: The use of the first two moments of the truncated multivariate Student-t distribution has attracted increasing attention from a wide range of applications. This paper develops recurrence relations for integrals that involve the density of multivariate Student-t distributions. The proposed techniques allow for fast computation of arbitrary-order product moments of folded and truncated multivariate Student-t distributions and offer explicit expressions of low-order moments of folded and truncated multivariate Student-t distributions. A real data example containing positive censored responses is applied to illustrate the effectiveness and importance of the proposed methods. An `R MomTrunc` package is developed and publicly available on the CRAN repository.

Key words and phrases: EM algorithm; Folded multivariate Student-t distribution; Product moments; Truncated multivariate normal distribution; Truncated multivariate Student-t distribution.

1. Introduction

The multivariate Student-t (MVT) distribution has played over the past decades a crucial role in statistical analysis, offering a more viable alternative with respect to real-world data. In particular, it has a harmonizing parameter (called the degrees of freedom) to control the thickness of tails and it includes the multivariate normal (MVN) distribution as a limiting case. Both the MVT and the MVN are members of the general family of elliptically symmetric distributions whose properties have been widely studied (Fang et al., 1990). Some recent applications in the areas such as spatial models (De Bastiani et al., 2015), linear mixed effects models (Pinheiro et al., 2001; Savalli et al., 2006), multivariate linear mixed effects models (Wang and Fan, 2011; Wang and Lin, 2014), mixture modelling (Peel and McLachlan, 2000), missing data imputation (Wang et al., 2017) and Bayesian statistical modeling (Fonseca et al., 2008; Wang and Lin, 2015), have been broadly studied.

On the other hand, for many applications on simulations or experimental studies, researches often generate a large number of datasets with values restricted to fixed intervals. For example, variables such as pH, grades, viral load in HIV studies and humidity in environmental studies, have upper and lower bounds due to detection limits, and the support of their densities is restricted to some given intervals. Thus, the necessity of studying the truncated distributions along with their properties arises naturally. In this context, there has been a growing interest in evaluating the moments of truncated distributions. For instance, Tallis (1961) provided the formulae for the first two moments of truncated multivariate normal (TMVN) distributions. Lien (1985) gave the expressions for the moments of truncated bivariate log-normal distributions with applications to testing the Houthakker effect in future markets. Jawitz (2004) derived the truncated moments of several continuous univariate distributions commonly applied to hydrologic problems. Kim (2008) provided analytical formulae for moments of the truncated univariate Student-t distribution in a recursive form. Flecher et al. (2010) obtained expressions for the moments of truncated skew-normal distributions (Azzalini, 1985) and applied the results to model the relative humidity data. Genç (2013) studied the moments of a doubly truncated member of the symmetrical class of normal/independent distributions and their applications to the actuarial data. Ho et al. (2012) presented a general formula to compute the first two moments of the truncated multivariate Student-t (TMVT) distribution based on the moment generating function (MGF) of the TMVN by expressing a TMVT random variable as a TMVN scale mixture variable. Relying on the TMVT moments, Lin and Wang (2017), Wang et al. (2018) and Lin and Wang (2020) conducted several simulation experiments to show the robustness of censored t models against outlying observations. Arismendi (2013) provided explicit expressions for computing arbitrary-order product moments of the TMVN distribution by using the MGF. However, the calculation of this approach relies on differentiation of the MGF and can be prohibitively time consuming. Instead of differentiating the MGF of the TMVN distribution, Kan and Robotti (2017) recently presented recurrence relations for integrals that involve directly the density of the MVN distribution for computing arbitrary order product moments of the TMVN distribution. These recursions offer fast computation of the moments of folded (FMVN) and TMVN distributions, which require evaluating p-dimensional integrals that involve MVN densities. Explicit expressions for some low order moments of FMVN and TMVN distributions are presented. Recently, another MGF-based approach was proposed by Roozegar et al. (2020) who derived explicit expressions for the first two moments of a truncated random vector with mean-variance mixture of normal distribution, which in particular includes the MVT distribution as a special case. Although some proposals to calculate the first two moments of the truncated Student-t distribution (Kim, 2008; Ho et al., 2012) have been recently published so far, to the best of our knowledge, there is no attempt on studying the product moments of folded multivariate Student-t (FMVT) and TMVT distributions. In this paper, we develop recurrence relations for integrals involving the density of MVT distributions based on the idea of Kan and Robotti (2017). The proposed recursions allow fast computation of the product moments of the FMVT and TMVT distributions. The proposed new methodology can be implemented in the R package `MomTrunc` (Galarza et al., 2020) available on CRAN repository.

The rest of this paper is organized as follows. In Section 2, we define the notation and briefly discuss some preliminary results related to the MVT, TMVT and FMVT distributions. Section 3 presents a recurrence formula of an integral for evaluating product moments of the FMVT and TMVT distributions. Explicit expressions for the first two moments of the FMVT and TMVT distributions are also presented. In Section 4, the usefulness of the proposed method is illustrated through a real data example containing censored responses. Some concluding remarks and implications for future research are given in Section 5. Technical details and additional information are provided in the Online Supplement.

2. Preliminaries

We begin our exposition by defining the notation and presenting some basic concepts which are used throughout our methodological developments. As is usual in probability theory and its applications, we denote a random variable by an upper-case letter and its realization by the correspondent lower case and use boldface letters for vectors and matrices. Let $\mathbf{I}_p$ represent a $p \times p$ identity matrix, $\mathbf{A}^\top$ be the transpose of $\mathbf{A}$, and $|\mathbf{X}| = (|X_1|\ldots,|X_p|)^\top$ denote the absolute value of each entry of vector $\mathbf{X}$. For multiple integrals, we use the shorthand notation

$$\int_{\mathbf{a}}^{\mathbf{b}} f(\mathbf{x})d\mathbf{x} = \int_{a_1}^{b_1} \cdots \int_{a_p}^{b_p} f(x_1, \ldots, x_p)\mathrm{d}x_p \ldots \mathrm{d}x_1,$$

where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbf{b} = (b_1, \ldots, b_p)^\top$. For two p-dimensional random vectors $\mathbf{x} = (x_1, \ldots, x_p)^\top$ and $\boldsymbol{\kappa} = (k_1, \ldots, k_p)^\top$, let $\mathbf{x}^{\boldsymbol{\kappa}}$ stand for $(x_1^{\kappa_1}, x_2^{\kappa_2}, \ldots, x_p^{\kappa_p})$. General results to compute the probability of a random vector lying in a hypercube are summarized in the following results

Lemma 1. *Let $\mathbf{X}$ be a p-variate random vector with joint probability density function (pdf) $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ and cumulative distribution function (cdf) $F_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$. Let $\mathbb{A}$ be a Borel set in $\mathbb{R}^p$ of the form*

$$\mathbb{A} = \{(x_1, \ldots, x_p)^\top \in \mathbb{R}^p : \ a_1 \le x_1 \le b_1, \ldots, a_p \le x_p \le b_p\} = \{\mathbf{x} \in \mathbb{R}^p : \mathbf{a} \le \mathbf{x} \le \mathbf{b}\}. \tag{2.1}$$

Then

$$\mathbb{P}(\mathbf{X} \in \mathbb{A}) = \sum_{\mathbf{s} \in S(\mathbf{a},\mathbf{b})} (-1)^{n_s} F_{\mathbf{X}}(\mathbf{s};\boldsymbol{\theta}),$$

where $S(\mathbf{a}, \mathbf{b}) = \{\mathbf{s} : \ \mathbf{s} = (s_1, \ldots, s_p)^\top$, with $s_i = \{a_i, b_i\}$, $i = 1, \ldots, p\}$ and $n_s = \sum_{i=1}^p \mathbb{1}(s_i = a_i)$ with $\mathbb{1}(\cdot)$ being the indicator function.

Proof. Based on the inclusion-exclusion principle, the probability $\mathbb{P}(\mathbf{X} \in \mathbb{A}) = \mathbb{P}(\mathbf{a} \le \mathbf{X} \le \mathbf{b})$ can be computed by summing the 2^p terms corresponding to the $\mathbf{s}$ elements in the solution space of $S(\mathbf{a}, \mathbf{b})$, where the term signs depend on the number of a's elements in the vector $\mathbf{s}$, i.e., n_s. □

Theorem 1. *Let $\mathbf{X}$ be a p-variate random vector with joint pdf $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ and joint cdf $F_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$. If $\mathbf{Y} = |\mathbf{X}|$, then the joint pdf and cdf of $\mathbf{Y}$ that follows a folded distribution are given, respectively, by*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s} \in S(p)} f_{\mathbf{X}}(\boldsymbol{\Lambda}_s\mathbf{y};\boldsymbol{\theta}), \quad \textit{for } \mathbf{y} \ge \mathbf{0},$$

and

$$F_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s} \in S(p)} \pi_s F_{\mathbf{X}}(\boldsymbol{\Lambda}_s\mathbf{y};\boldsymbol{\theta}),$$

where $S(p) = \{\mathbf{s} : \mathbf{s} = (s_1, \ldots, s_p)^\top$, with $s_i = \pm 1, i = 1, \ldots, p\}$, $\boldsymbol{\Lambda}_s = Diag(\mathbf{s})$ and $\pi_s = \prod_{i=1}^p s_i$.

Proof. The distribution function $F_{\mathbf{Y}}(\mathbf{y})$ can be calculated as a particular case of Lemma 1, when $\mathbf{a} = -\mathbf{y}$ and $\mathbf{b} = \mathbf{y}$. The proof ends by differentiating $F_{\mathbf{Y}}(\mathbf{y})$ with respect to each entry of $\mathbf{y}$ to obtain $f_{\mathbf{Y}}(\mathbf{y})$. □

Corollary 1. *If* $\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Psi})$ *belongs to the location-scale family of distributions with location and scale parameters* $\boldsymbol{\xi}$ *and* $\boldsymbol{\Psi}$ *respectively, then* $\mathbf{Z}_s = \boldsymbol{\Lambda}_s\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{z}; \boldsymbol{\Lambda}_s\boldsymbol{\xi}, \boldsymbol{\Lambda}_s\boldsymbol{\Psi}\boldsymbol{\Lambda}_s)$ *and consequently the joint pdf and cdf of* $\mathbf{Y} = |\mathbf{X}|$ *are given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s}\in S(p)} f_{\mathbf{X}}(\mathbf{y}; \boldsymbol{\Lambda}_s\boldsymbol{\xi}, \boldsymbol{\Lambda}_s\boldsymbol{\Psi}\boldsymbol{\Lambda}_s), \quad \textit{for } \mathbf{y} \geq \mathbf{0},$$

and

$$F_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s}\in S(p)} \pi_s F_{\mathbf{X}}(\boldsymbol{\Lambda}_s\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Psi}).$$

Corollary 1 generalizes the results of Chakraborty and Chatterjee (2013) for the FMVN case to all distributions belong to the multivariate location-scale family.

Corollary 2. *Under the same conditions of Corollary 1, we have that*

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}] = \sum_{\mathbf{s}\in S(p)} \mathbb{E}[\mathbf{Z}_s^{+\boldsymbol{\kappa}}],$$

where $\mathbf{X}^{+} = \mathbf{X}\cdot\mathbb{1}(\mathbf{X} > \mathbf{0})$.

Proof. It follows from basic probability theory that

$$\begin{aligned}\int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}\mathbf{y} &= \sum_{\mathbf{s}\in S(p)} \int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} f_{\mathbf{X}}(\mathbf{y}; \boldsymbol{\Lambda}_s\boldsymbol{\xi}, \boldsymbol{\Lambda}_s\boldsymbol{\Psi}\boldsymbol{\Lambda}_s)\mathrm{d}\mathbf{y}\\ &= \sum_{\mathbf{s}\in S(p)} \int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} f_{\mathbf{Z}_s}(\mathbf{y})\mathrm{d}\mathbf{y}\\ &= \sum_{\mathbf{s}\in S(p)} \mathbb{E}[\mathbf{Z}_s^{+\boldsymbol{\kappa}}].\end{aligned}$$

□

2.1 The MVT and FMVT distributions and main properties

A random variable $\mathbf{X}$ having a p-variate MVT distribution with location vector $\boldsymbol{\mu}$, positive-definite scale-covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν, denoted by $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, has the pdf:

$$t_p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}} \nu^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{\delta(\mathbf{x})}{\nu}\right)^{-(p+\nu)/2},$$

where $\Gamma(\cdot)$ is the standard gamma function and $\delta(\mathbf{x}) = (\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ is the Mahalanobis distance. Let $L_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ represent

$$L_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \int_{\mathbf{a}}^{\mathbf{b}} t_p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu})\mathrm{d}\mathbf{x},$$

where $\mathbf{a} = (a_1,\ldots,a_p)^\top$ and $\mathbf{b} = (b_1,\ldots,b_p)^\top$. The cdf of $\mathbf{X}$ is denoted as

$$T_p(\mathbf{b}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \int_{-\boldsymbol{\infty}}^{\mathbf{b}} t_p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)d\mathbf{x} = L_p(-\boldsymbol{\infty},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu).$$

In light of Lemma 1, we have

$$L_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \sum_{\mathbf{s}\in S(\mathbf{a},\mathbf{b})} (-1)^{n_s} T_p(\mathbf{s}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu),$$

where $S(\mathbf{a},\mathbf{b})$ and n_s are defined as in Lemma 1.

It is known that as $\nu\to\infty$, $\mathbf{X}$ converges in distribution to a multivariate normal with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, denoted by $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. An important property of the random vector $\mathbf{X}$ is that it can be written as a scale mixture of the MVN random vector coupled with a positive random variable, i.e.,

$$\mathbf{X} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z},$$

where $\mathbf{Z}\sim N_p(\mathbf{0},\boldsymbol{\Sigma})$, and is independent of $U\sim \mathrm{Gamma}(\nu/2,\nu/2)$, where $\mathrm{Gamma}(a,b)$ denotes a gamma distribution with mean a/b.

The following properties of the MVT distribution are useful for our theoretical developments. We start with the marginal-conditional decomposition of a MVT random vector. The proof of the following propositions can be found in Arellano-Valle and Bolfarine (1995).

Proposition 1. *Let* $\mathbf{X}\sim t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ *partitioned as* $\mathbf{X}^\top = (\mathbf{X}_1^\top,\mathbf{X}_2^\top)^\top$ *with* $dim(\mathbf{X}_1) = p_1$, $dim(\mathbf{Y}_2) = p_2$, *where* $p_1+p_2=p$. *Let* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top,\boldsymbol{\mu}_2^\top)^\top$ *and* $\boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$ *be the corresponding partitions of* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$. *Then, we have*

(i) $\mathbf{X}_1\sim t_{p_1}(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\nu)$*; and*

(ii) The conditional distribution of $\mathbf{X}_2\mid(\mathbf{X}_1=\mathbf{x}_1)$ *is given by*

$$\mathbf{X}_2\mid(\mathbf{X}_1=\mathbf{x}_1)\sim t_{p_2}\left(\mathbf{y}_2\mid\boldsymbol{\mu}_{2.1},\widetilde{\boldsymbol{\Sigma}}_{22.1},\nu+p_1\right),$$

where $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ *and* $\widetilde{\boldsymbol{\Sigma}}_{22.1} = \left(\dfrac{\nu+\delta_1}{\nu+p_1}\right)\boldsymbol{\Sigma}_{22.1}$ *with* $\delta_1 = (\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ *and* $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Proposition 2. *Let* $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. *Then for any fixed vector* $\mathbf{b} \in \mathbb{R}^m$ *and matrix* $\mathbf{A} \in \mathbb{R}^{m\times p}$ *of full rank we get*

$$\mathbf{V} = \mathbf{b} + \mathbf{A}\mathbf{X} \sim t_m(\mathbf{b} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top, \nu).$$

We are interested in computing $E\left[|X_1|^{k_1}\ldots|X_p|^{k_p}\right]$ and $\mathbb{E}[X_1^{k_1}\ldots X_p^{k_p}|a_i < X_i < b_i, i = 1,\ldots,p]$ for any nonnegative integer values $k_i = 0, 1, 2, \ldots$, where the former is the moment of a FMVT distribution $|\mathbf{X}| = (|X_1|,\ldots,|X_p|)^\top$, and the later is the moment of a TMVT distribution, with X_i truncated at the lower limit a_i and upper limit b_i, $i = 1,\ldots,p$. Remark that some of the $a_i's$ can be $-\infty$ and some of the $b_i's$ can be $+\infty$ in the second expression. When all the b_i's are ∞, the distribution is called the lower TMVT, and when all the a_i's are $-\infty$, the distribution is called the upper TMVT.

2.2 The TMVT distribution and main properties

A p-dimensional random vector $\mathbf{Y}$ is said to follow a doubly truncated Student-t distribution with location vector $\boldsymbol{\mu}$, scale-covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν over the truncation region $\mathbb{A}$ defined in (2.1), denoted by $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, if it has the pdf:

$$Tt_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}, \ \mathbf{a} \le \mathbf{y} \le \mathbf{b}.$$

The cdf of $\mathbf{Y}$ evaluated at the region $\mathbf{a} \le \mathbf{y} \le \mathbf{b}$ is

$$TT_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{1}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}\int_{\mathbf{a}}^{\mathbf{y}} t_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)d\mathbf{x} = \frac{L_p(\mathbf{a}, \mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}.$$

The following propositions are related to the marginal and conditional moments of the first two moments of the TMVT distributions under a double truncation. The proof is similar to those given in Matos et al. (2013). In what follows, we shall use the notation $\delta(\mathbf{Y}) \equiv \delta(\mathbf{Y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\mathbf{Y} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$ to stand for the Mahalanobis distance of $\mathbf{Y}$ with respect to the points with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, and $\mathbf{Y}^{(0)} = 1$, $\mathbf{Y}^{(1)} = \mathbf{Y}$ and $\mathbf{Y}^{(2)} = \mathbf{Y}\mathbf{Y}^\top$.

Proposition 3. *If* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$ *then it holds that*

$$\mathbb{E}\left[\left(\frac{\nu+p}{\nu+\delta(\mathbf{Y})}\right)^r \mathbf{Y}^{(k)}\right] = c_p(\nu, r)\frac{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu+2r)}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}\mathbb{E}[\mathbf{W}^{(k)}],$$

where $k = 0, 1, 2,$

$$c_p(\nu, r) = \left(\frac{\nu + p}{\nu}\right)^r \frac{\Gamma\left(\frac{p+\nu}{2}\right)\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{p+\nu+2r}{2}\right)},$$

$\boldsymbol{\Sigma}^* = \nu\boldsymbol{\Sigma}/(\nu + 2r)$ *and* $\nu + 2r > 0$, *with* $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r; (\mathbf{a}, \mathbf{b}))$.

Notice that Proposition 3 depends on formulas for $\mathbb{E}[\mathbf{W}]$ and $\mathbb{E}[\mathbf{W}\mathbf{W}^\top]$, where $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r; (\mathbf{a}, \mathbf{b}))$. Having established the formula on the k-order moment of $\mathbf{Y}$, we provide an explicit formula for the conditional moments with respect to a two-component partition of $\mathbf{Y}$.

Proposition 4. *Let* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$. *Consider the partition* $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)$ *with* $\dim(\mathbf{Y}_1) = p_1$, $\dim(\mathbf{Y}_2) = p_2$, $p_1 + p_2 = p$, *and the corresponding partitions of* $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2)^\top$, $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2)^\top$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^\top$ *and* $\boldsymbol{\Sigma}$ *(as in Proposition 1). Then, for* $k = 0, 1, 2$, *we have that*

$$\mathbb{E}\left[\left(\frac{\nu + p}{\nu + \delta(\mathbf{Y})}\right)^r \mathbf{Y}_2^{(k)} \mid \mathbf{Y}_1\right] = \frac{d_p(p_1, \nu, r)}{(\nu + \delta(\mathbf{Y}_1))^r} \frac{L_{p_2}(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu + p_1 + 2r)}{L_{p_2}(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1)} \mathbb{E}[\mathbf{W}_2^{(k)}],$$

for $\nu + p_1 + 2r > 0$, *with* $\delta(\mathbf{Y}_1) = \delta(\mathbf{Y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$,

$$\widetilde{\boldsymbol{\Sigma}}^*_{22.1} = \left(\frac{\nu + \delta_1}{\nu + 2r + p_1}\right)\boldsymbol{\Sigma}_{22.1} \quad \textit{and} \quad d_p(p_1, \nu, r) = (\nu + p)^r \frac{\Gamma\left(\frac{p+\nu}{2}\right)\Gamma\left(\frac{p_1+\nu+2r}{2}\right)}{\Gamma\left(\frac{p_1+\nu}{2}\right)\Gamma\left(\frac{p+\nu+2r}{2}\right)},$$

where $\boldsymbol{\mu}_{2.1}$ *and* $\boldsymbol{\Sigma}_{22.1}$ *as defined in proposition 1. Moreover,* $\mathbf{W}_2 \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu + p_1 + 2r; [\mathbf{a}_2, \mathbf{b}_2])$.

3. The recurrence relation for the multivariate Student-t integral

Let $\mathbf{a}_{(i)}$ be a vector $\mathbf{a}$ with its ith element being removed. For a matrix $\boldsymbol{\Delta}$, we let $\boldsymbol{\Delta}_{i(j)}$ stand for the ith row of $\boldsymbol{\Delta}$ with its jth element being removed. Similarly, $\boldsymbol{\Delta}_{(i)(j)}$ stands for the matrix $\boldsymbol{\Delta}$ with its ith row and jth columns being removed. Besides, let $\mathbf{e}_i$ denote a $p \times 1$ vector with its ith element equalling one and zero otherwise.

The integral that we are interested in evaluating is

$$\mathrm{F}^p_{\boldsymbol{\kappa}}(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{x}^{\boldsymbol{\kappa}} t_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)\mathrm{d}\mathbf{x}.$$

The boundary condition is obviously $\mathrm{F}^p_0(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. The recurrence relation for the normal case has been recently presented by Kan and Robotti (2017). When $p = 1$, the use of integration by parts straightforwardly leads to

$$\mathrm{F}^1_0(a, b; \mu, \sigma^2, \nu) \quad = \quad T_1(b|\mu, \sigma^2, \nu) - T_1(a|\mu, \sigma^2, \nu),$$

$$\begin{aligned}\mathrm{F}^1_{k+1}(a,b;\mu,\sigma^2,\nu) &= \mu\mathrm{F}^1_k(a,b;\mu,\sigma^2,\nu)+\tfrac{k\nu\sigma^2}{(\nu-2)}\mathrm{F}^1_{k-1}(a,b;\mu,\tfrac{\nu}{\nu-2}\sigma^2,\nu-2) \\ &\quad+\tfrac{\nu\sigma^2}{(\nu-2)}[a^k t_1(a|\mu,\tfrac{\nu}{\nu-2}\sigma^2,\nu-2)-b^k t_1(b|\mu,\tfrac{\nu}{\nu-2}\sigma^2,\nu-2)],\ (k\geq 0).\end{aligned}\tag{3.2}$$

When $p>1$, we need a similar recurrence relation in order to compute $\mathrm{F}^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ which is presented in the following Theorem.

Theorem 2. *For $p\geq 1$, $i=1,\ldots,p$ and $\nu>2$,*

$$F^p_{\boldsymbol{\kappa}+\mathbf{e}_i}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)=\mu_i F^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)+\tfrac{\nu}{\nu-2}\mathbf{e}_i^\top\boldsymbol{\Sigma}\mathbf{c}_{\boldsymbol{\kappa}},\tag{3.3}$$

where $\mathbf{c}_{\boldsymbol{\kappa}}$ is a $p\times 1$ vector with the jth element being

$$\begin{aligned}c_{\boldsymbol{\kappa},j} &= k_j F^p_{\boldsymbol{\kappa}-\mathbf{e}_j}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2) \\ &\quad+a_j^{k_j}t_1(a_j|\mu_j,\sigma_j^{*2},\nu-2)F^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}^{\mathbf{a}}_j,\tilde{\delta}^{\mathbf{a}}_j\tilde{\boldsymbol{\Sigma}}_j,\nu-1) \\ &\quad-b_j^{k_j}t_1(b_j|\mu_j,\sigma_j^{*2},\nu-2)F^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}^{\mathbf{b}}_j,\tilde{\delta}^{\mathbf{b}}_j\tilde{\boldsymbol{\Sigma}}_j,\nu-1),\end{aligned}\tag{3.4}$$

and

$$\begin{aligned}\tilde{\boldsymbol{\Sigma}}_j &= \boldsymbol{\Sigma}^*_{(j)(j)}-\frac{1}{\sigma_j^{*2}}\boldsymbol{\Sigma}^*_{(j)j}\boldsymbol{\Sigma}^*_{j(j)},\ \tilde{\delta}^{\mathbf{a}}_j=\frac{\nu-2+\frac{(a_j-\mu_j)^2}{\sigma_j^{*2}}}{\nu-1},\ \tilde{\delta}^{\mathbf{b}}_j=\frac{\nu-2+\frac{(b_j-\mu_j)^2}{\sigma_j^{*2}}}{\nu-1}, \\ \tilde{\boldsymbol{\mu}}^{\mathbf{a}}_j &= \boldsymbol{\mu}_{(j)}+\tfrac{(a_j-\mu_j)}{\sigma_j^{*2}}\boldsymbol{\Sigma}^*_{(j)j},\ \ \tilde{\boldsymbol{\mu}}^{\mathbf{b}}_j=\boldsymbol{\mu}_{(j)}+\tfrac{(b_j-\mu_j)}{\sigma_j^{*2}}\boldsymbol{\Sigma}^*_{(j)j},\ \boldsymbol{\Sigma}^*=\tfrac{\nu}{\nu-2}\boldsymbol{\Sigma}\quad \textit{and}\quad \sigma_j^{*2}=\frac{\nu}{\nu-2}\sigma_j^2.\end{aligned}$$

When $k_j=0$, the first term in (3.4) vanishes. When $a_j=-\infty$ and $k_j\leq\nu-2$, the second term vanishes, and when $b_j=+\infty$ and $k_j\leq\nu-2$, the third term vanishes.

Proof. After taking the first derivative of the MVT density, we have

$$-\frac{\partial t_p(\mathbf{x}|\boldsymbol{\mu},\frac{\nu}{\nu-2}\boldsymbol{\Sigma},\nu-2)}{\partial\mathbf{x}}=\frac{\nu-2}{\nu}t_p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}).$$

Multiplying each element on both sides by $\mathbf{x}^{\boldsymbol{\kappa}}$ and integrating $\mathbf{x}$ from $\mathbf{a}$ to $\mathbf{b}$, we have

$$\mathbf{c}_{\boldsymbol{\kappa}}=\tfrac{\nu-2}{\nu}\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathrm{F}^p_{\boldsymbol{\kappa}+\mathbf{e}_1} & - & \mu_1\mathrm{F}^p_{\boldsymbol{\kappa}}\\ \mathrm{F}^p_{\boldsymbol{\kappa}+\mathbf{e}_2} & - & \mu_2\mathrm{F}^p_{\boldsymbol{\kappa}}\\ & \vdots & \\ \mathrm{F}^p_{\boldsymbol{\kappa}+\mathbf{e}_p} & - & \mu_p\mathrm{F}^p_{\boldsymbol{\kappa}}.\end{bmatrix}.$$

Using integration by parts, the jth element of the left-hand side is

$$c_{\boldsymbol{\kappa},j}=-\int_{\mathbf{a}_{(j)}}^{\mathbf{b}_{(j)}}\mathbf{x}^{\boldsymbol{\kappa}}t_p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2)|_{x_j=a_j}^{b_j}\mathrm{d}\mathbf{x}_{(j)}+\int_{\mathbf{a}}^{\mathbf{b}}k_j\mathbf{x}^{\boldsymbol{\kappa}-\mathbf{e}_j}t_p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2)\mathrm{d}\mathbf{x}.\tag{3.5}$$

Using the fact that

$$\begin{aligned} t_p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2)|_{x_j=a_j} &= t_1(a_j|\mu_j,\sigma_j^{*2},\nu-2)t_{p-1}(\mathbf{x}_{(j)}|\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\delta}_j^{\mathbf{a}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1) \quad \text{and} \\ t_p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2)|_{x_j=b_j} &= t_1(b_j|\mu_j,\sigma_j^{*2},\nu-2)t_{p-1}(\mathbf{x}_{(j)}|\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\delta}_j^{\mathbf{b}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1), \end{aligned}$$

we get

$$\begin{aligned} c_{\boldsymbol{\kappa},j} &= k_j \mathrm{F}^p_{\boldsymbol{\kappa}-\mathbf{e}_j}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2) \\ &\quad + a_j^{k_j} t_1(a_j|\mu_j,\sigma_j^{*2},\nu-2)\mathrm{F}^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\delta}_j^{\mathbf{a}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1) \\ &\quad - b_j^{k_j} t_1(b_j|\mu_j,\sigma_j^{*2},\nu-2)\mathrm{F}^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\delta}_j^{\mathbf{b}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1). \end{aligned}$$

When $k_j=0$, the last integral in (3.5) is equal to zero, and the first term of $c_{\boldsymbol{\kappa},j}$ vanishes. When $a_j\to-\infty$ and $k_j\le\nu-2$, $a_j^{k_j}t_1(a_j|\mu_j,\sigma_j^{*2},\nu-2)\to 0$, so the second term of $c_{\boldsymbol{\kappa},j}$ vanishes. Similarly when $b_j\to\infty$ the third term of $c_{\boldsymbol{\kappa},j}$ vanishes. Finally, the desired result is obtained by multiplying $\frac{\nu}{\nu-2}\boldsymbol{\Sigma}$ on both sides of (3.3). □

As a consequence, $\mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}}]$ always exists for $\sum_{j=1}^p \kappa_j<\nu$. When all $a_i's$ are $-\infty$ or all $b_i's$ are $+\infty$, the length of the recurrence relation is reduced to $2p+1$ rather than the original $3p+1$. When all $a_i's$ are $-\infty$ and all $b_i's$ are $+\infty$, we have

$$\mathrm{F}^p_{\boldsymbol{\kappa}}(-\infty,+\infty;\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)=\mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}}],\quad \mathbf{X}\sim t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu),$$

and the recursive relation of length $(p+1)$ is

$$\mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}+\mathbf{e}_i}]=\mu_i E[\mathbf{X}^{\boldsymbol{\kappa}}]+\sum_{j=1}^p \sigma^*_{ij}k_j\mathbb{E}[\mathbf{y}^{\boldsymbol{\kappa}-\mathbf{e}_i}],\ \mathbf{y}\sim t_p(\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2),\ i=1,\ldots,p.$$

Another special case of interest occurs when $a_i=0$ and $b_i=+\infty$, $i=1,\ldots,p$. In this scenario, we have

$$I^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)=\mathrm{F}^p_{\boldsymbol{\kappa}}(\mathbf{0},+\boldsymbol{\infty};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu).$$

The recurrence relation for $I^p_{\boldsymbol{\kappa}}$ can be written as

$$I^p_{\boldsymbol{\kappa}+\mathbf{e}_i}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)=\mu_i I^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)+\sum_{j=1}^p \sigma^*_{ij}d_{\kappa,j},\ i=1,\ldots,p,$$

where

$$d_{\kappa,j}=\begin{cases} k_j I^p_{\boldsymbol{\kappa}-\mathbf{e}_i}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) & \text{for } k_j>0,\\ t_1(0|\mu_j,\sigma_j^{*2},\nu-2)I^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\tilde{\boldsymbol{\mu}}_j,\tilde{\delta}_j\tilde{\boldsymbol{\Sigma}}_j,\nu-1) & \text{for } k_j=0, \end{cases}$$

with

$$\tilde{\boldsymbol{\mu}}_j = \boldsymbol{\mu}_{(j)} - \frac{\mu_j}{\sigma_j^{*2}}\boldsymbol{\Sigma}^*_{(j)j},\ \tilde{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}^*_{(j)(j)} - \frac{1}{\sigma_j^{*2}}\boldsymbol{\Sigma}^*_{(j)j}\boldsymbol{\Sigma}^*_{j(j)},\ \text{and } \tilde{\delta}_j = \frac{\nu - 2 + \frac{\mu_j^2}{\sigma_j^{*2}}}{\nu - 1}.$$

3.1 The first two moments of the doubly TMVT distribution

Let $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $\mathbf{Y} = \mathbf{X} \mid (\mathbf{a} \leq \mathbf{X} \leq \mathbf{b}) \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$. It follows that

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}] = \frac{1}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)} \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{x}^{\boldsymbol{\kappa}} t_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) d\mathbf{x} = \frac{\mathrm{F}^p_{\boldsymbol{\kappa}}(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}.$$

Furthermore, let $\mathrm{F}^p_{\boldsymbol{\kappa}} \equiv \mathrm{F}^p_{\boldsymbol{\kappa}}(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $L \equiv L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ for simplicity. In light of Theorem 2, it is straightforward to see that

$$\mathbb{E}[Y_i] = \frac{F^p_{\mathbf{e}_i}}{L} = \mu_i + \frac{1}{L}\mathbf{e}_i^\top \boldsymbol{\Sigma}^* \mathbf{c}_{\mathbf{0}} \quad \text{and} \quad \mathbb{E}[Y_i Y_j] = \frac{F^p_{\mathbf{e}_i + \mathbf{e}_j}}{L} = \mu_j \mathbb{E}[Y_i] + \frac{1}{L}\mathbf{e}_j^\top \boldsymbol{\Sigma}^* \mathbf{c}_{\mathbf{e}_i}, \tag{3.6}$$

where $\mathbf{c}_{\mathbf{0}} = \mathbf{c}_{\mathbf{a}} - \mathbf{c}_{\mathbf{b}}$, with

$$c_{\mathbf{a}} = \Big[t_1(a_j|\mu_j, \sigma_j^{*2}, \nu - 2)\, L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}^{\mathbf{a}}_j, \tilde{\delta}^{\mathbf{a}}_j \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)\Big]_{j=1}^p,$$
$$c_{\mathbf{b}} = \Big[t_1(b_j|\mu_j, \sigma_j^{*2}, \nu - 2)\, L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}^{\mathbf{b}}_j, \tilde{\delta}^{\mathbf{b}}_j \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)\Big]_{j=1}^p,$$

and

$$c_{\mathbf{e}_i} = \Big[\mathbf{e}_{ij} L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu - 2) + a_j t_1(a_j|\mu_j, \sigma_j^{*2}, \nu - 2)\mathrm{F}^{p-1}_{\mathbf{e}_{i(j)}}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}^{\mathbf{a}}_j, \tilde{\delta}^{\mathbf{a}}_j \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$$
$$-\ b_j t_1(b_j|\mu_j, \sigma_j^{*2}, \nu - 2)\mathrm{F}^{p-1}_{\mathbf{e}_{i(j)}}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}^{\mathbf{b}}_j, \tilde{\delta}^{\mathbf{b}}_j \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)\Big]_{j=1}^p,$$

where

$$\begin{aligned} c_{\mathbf{e}_i i} &= L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu - 2) + a_j c_{\mathbf{a}i} - b_j c_{\mathbf{b}i}, \\ c_{\mathbf{e}_i j} &\overset{i \neq j}{=} a_j c_{\mathbf{a}i} \mathbb{E}[(\mathbf{X}_{(j)} \mid X_j = a_j) \mid \mathbf{a}_{(j)} \leq \mathbf{X}_{(j)} \leq \mathbf{b}_{(j)}] \\ &\quad -b_j c_{\mathbf{b}i} \mathbb{E}[(\mathbf{X}_{(j)} \mid X_j = b_j) \mid \mathbf{a}_{(j)} \leq \mathbf{X}_{(j)} \leq \mathbf{b}_{(j)}]. \end{aligned}$$

The last equality is obtained by noting that

$$\mathbb{P}(\mathbf{a}_{(j)} \leq \mathbf{X}_{(j)} \leq \mathbf{b}_{(j)}|X_j = a_j) = L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}^{\mathbf{a}}_j, \tilde{\delta}^{\mathbf{a}}_j \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$$

and $\mathbb{P}(\mathbf{a}_{(j)} \leq \mathbf{X}_{(j)} \leq \mathbf{b}_{(j)}|X_j = b_j) = L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}^{\mathbf{b}}_j, \tilde{\delta}^{\mathbf{b}}_j \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$.

Algorithm 1: Mean vector for $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$

mean$(\mathbf{a}, \mathbf{b}, \boldsymbol{\theta})$
$L \leftarrow L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$; $\boldsymbol{c}_{\mathbf{a}} \leftarrow \mathbf{0}$; $\boldsymbol{c}_{\mathbf{b}} \leftarrow \mathbf{0}$;
for $j = 1 : p$ **do**
  $\boldsymbol{\theta}_j^{\mathbf{a}} \leftarrow (\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}}, \tilde{\delta}_j^{\mathbf{a}} \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$; $\boldsymbol{\theta}_j^{\mathbf{b}} \leftarrow (\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}}, \tilde{\delta}_j^{\mathbf{b}} \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$;
  if $a_j \neq \infty$ **then**
    $\boldsymbol{c}_{\mathbf{a}}(j) \leftarrow t_1(a_j | \mu_j, \sigma_j^{*2}, \nu - 2) L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^{\mathbf{a}}, \tilde{\delta}_j^{\mathbf{a}} \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$;
  end
  if $b_j \neq \infty$ **then**
    $\boldsymbol{c}_{\mathbf{b}}(j) \leftarrow t_1(b_j | \mu_j, \sigma_j^{*2}, \nu - 2) L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^{\mathbf{b}}, \tilde{\delta}_j^{\mathbf{b}} \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$;
  end
end
$\boldsymbol{\xi} \leftarrow \boldsymbol{\mu} + \frac{\nu}{\nu-2} \boldsymbol{\Sigma} (\boldsymbol{c}_{\mathbf{a}} - \boldsymbol{c}_{\mathbf{b}}) / L$;
return $\boldsymbol{\xi}$;

Let $\mathbf{C} = (\mathbf{c}_{\mathbf{e}_1}, \mathbf{c}_{\mathbf{e}_2}, \ldots, \mathbf{c}_{\mathbf{e}_p})$. From expressions in (3.6), we can note that for $\mathbb{E}[Y_i]$, $\mathbf{c}_{\mathbf{0}}$ does not depend on i and, for $\mathbb{E}[Y_i Y_j]$, $\mathbf{c}_{\mathbf{e}_i}$ does not depend on j. Then, it is easy to establish the mean vector and variance-covariance matrix for $\mathbf{Y}$, respectively, given by

$$\boldsymbol{\xi} = \mathbb{E}[\mathbf{Y}] = \boldsymbol{\mu} + \frac{1}{L} \boldsymbol{\Sigma}^* \boldsymbol{c}_{\mathbf{0}}, \tag{3.7}$$

$$\boldsymbol{\Psi} = \mathrm{cov}[\mathbf{Y}] = \frac{1}{L} \boldsymbol{\Sigma}^* (\mathbf{C} - \boldsymbol{c}_{\mathbf{0}} \boldsymbol{\xi}^\top), \tag{3.8}$$

as well as the second moment $\mathbb{E}[\mathbf{Y}\mathbf{Y}^\top] = \boldsymbol{\mu}\boldsymbol{\xi}^\top + \frac{1}{L} \mathbf{C} \boldsymbol{\Sigma}^*$.

Methods for computing the mean and variance-covariance matrix of $\mathbf{Y}$ are summarized in Algorithms 1 and 2. Note that, to calculate the variance-covariance matrix $\boldsymbol{\Psi}$ in Algorithm 2, it is necessary to compute $2p$ $(p-1)$-variate mean vectors (lines 8 and 13) through Algorithm 1. This schema leads to only $1 + 2p$ necessary integrals to compute the mean and additional $1 + 2p + 4p(p-1)$ integrals for the variance-covariance matrix. It is noteworthy to mention that (i) probabilities between lines 7 and 12 in Algorithm 2 can be recycled from the mean$(\mathbf{a}, \mathbf{b}, \boldsymbol{\theta})$ function, and (ii) $\mathbf{C}$ is not symmetric, however both of its (i,j)-th and (j,i)-th elements $c_{\mathbf{e}_j i}$ and $c_{\mathbf{e}_i j}$ depends on probabilities of the form $\mathbb{P}\left(\mathbf{a}_{(i,j)} \leq \mathbf{X}_{(i,j)} \leq \mathbf{b}_{(i,j)} \mid (X_i, X_j) = (x_i, x_j)\right)$, with $(x_i, x_j) \in \{a_i, b_i\} \times \{a_j, b_j\}$. This leads to an optimal schema with a maximum total of $2(1 + p^2)$ integrals to compute the mean and the variance-covariance matrix in the case that the distribution is doubly truncated. Lastly, we remark that this recurrence is limited to work when the degrees of freedom is greater than 3 due to the computation of $\boldsymbol{\Sigma}^{**} = \nu^* \boldsymbol{\Sigma} / (\nu^* - 2)$ when

Algorithm 2: Mean vector and variance-covariance matrix for $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$

meanvar$(\mathbf{a}, \mathbf{b}, \boldsymbol{\theta})$
$L \leftarrow L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$; $L^* \leftarrow L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu - 2)$;
$\mathbf{W_a} \leftarrow \mathbf{0}_{p\times p}$; $\mathbf{W_b} \leftarrow \mathbf{0}_{p\times p}$;
for $j = 1 : p$ **do**
  $\boldsymbol{\theta}_j^{\mathbf{a}} \leftarrow (\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}}, \tilde{\delta}_j^{\mathbf{a}} \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$; $\boldsymbol{\theta}_j^{\mathbf{b}} \leftarrow (\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}}, \tilde{\delta}_j^{\mathbf{b}} \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$;
  if $a_j \neq \infty$ **then**
    $\boldsymbol{c}_{\mathbf{a}}(j) \leftarrow t_1(a_j|\mu_j, \sigma_j^{*2}, \nu - 2) L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^{\mathbf{a}}, \tilde{\delta}_j^{\mathbf{a}} \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$;
    $\mathbf{W_a}(-j, j) \leftarrow \text{mean}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}, \boldsymbol{\theta}_{(j)}^{\mathbf{a}})$;
    $\mathbf{W_a}(j, j) \leftarrow \mathbf{a}(j)$;
  end
  if $b_j \neq \infty$ **then**
    $\boldsymbol{c}_{\mathbf{b}}(j) \leftarrow t_1(b_j|\mu_j, \sigma_j^{*2}, \nu - 2) L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^{\mathbf{b}}, \tilde{\delta}_j^{\mathbf{b}} \tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$;
    $\mathbf{W_b}(-j, j) \leftarrow \text{mean}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}, \boldsymbol{\theta}_{(j)}^{\mathbf{b}})$;
    $\mathbf{W_b}(j, j) \leftarrow \mathbf{b}(j)$;
  end
end
$\boldsymbol{\xi} \leftarrow \boldsymbol{\mu} + \boldsymbol{\Sigma}^*(\boldsymbol{c}_{\mathbf{a}} - \boldsymbol{c}_{\mathbf{b}})/L$;
$\boldsymbol{\Psi} \leftarrow (L^* \text{diag}(p) + \mathbf{W_a} \text{diag}(\mathbf{c_a}) - \mathbf{W_b} \text{diag}(\mathbf{c_b}))\boldsymbol{\Sigma}^*/L$;
return $\boldsymbol{\xi}$, $\boldsymbol{\Psi}$;

$\nu^* = \nu - 1$. For instance, Ho et al. (2012) devised a general formula for computing the first two moments the TMVT distribution based on the moment generating function but their method is limited to $\nu > 4$.

3.2 The first two moments of the TMVT distribution when a non-truncated partition exists

We describe a trick for fast computation of the first two moments of the TMVT distribution when there are double infinite limits. Consider the partition $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ such that $dim(\mathbf{X}_1) = p_1$, $dim(\mathbf{X}_2) = p_2$, where $p_1 + p_2 = p$. Using the law of total expectations, we have

$$\mathbb{E}[\mathbf{X}] = \mathbb{E}\begin{bmatrix} \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2] \\ \mathbf{X}_2 \end{bmatrix}$$

and

$$\mathrm{cov}[\mathbf{X}] = \begin{bmatrix} \mathbb{E}[\mathrm{cov}[\mathbf{X}_1|\mathbf{X}_2]] + \mathrm{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] & \mathrm{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2], \mathbf{X}_2] \\ \mathrm{cov}[\mathbf{X}_2, \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] & \mathrm{cov}[\mathbf{X}_2] \end{bmatrix}.$$

Let p_1 be the number of pairs in $[\mathbf{a}, \mathbf{b}]$ that are both infinite. We consider the partition $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ in which upper and lower truncation limits associated with $\mathbf{X}_1$ are both infinite, but at least one of the truncation limits associated with $\mathbf{X}_2$ is finite. Let

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \mathbf{a} = (\mathbf{a}_1^\top, \mathbf{a}_2^\top)^\top \text{ and } \mathbf{b} = (\mathbf{b}_1^\top, \mathbf{b}_2^\top)^\top$$

be corresponding partitions of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{a}$ and $\mathbf{b}$. Since $\mathbf{a}_1 = -\boldsymbol{\infty}$ and $\mathbf{b}_1 = \boldsymbol{\infty}$, it follows that

$$\begin{aligned} \mathbf{X}_1|\mathbf{X}_2 &\sim t_{p_1}\big(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})\frac{(\nu + \delta(\mathbf{X}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}))}{(\nu + p_2)}, \nu + p_2\big), \\ \mathbf{X}_2 &\sim \mathrm{Tt}_{p_2}\big(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, \nu; [\mathbf{a}_2, \mathbf{b}_2]\big). \end{aligned}$$

This leads to

$$\mathbb{E}[\mathbf{X}] = \mathbb{E}\begin{bmatrix} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2) \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{\xi}_2 - \boldsymbol{\mu}_2) \\ \boldsymbol{\xi}_2 \end{bmatrix}. \tag{3.9}$$

On the other hand, we have $\mathrm{cov}[\mathbf{X}_2, \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] = \mathrm{cov}[\mathbf{X}_2, \mathbf{X}_2\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}] = \boldsymbol{\Psi}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, $\mathrm{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Psi}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ and $\mathbb{E}[\mathrm{cov}[\mathbf{X}_1|\mathbf{X}_2]] = \omega_{1.2}(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$ where $\boldsymbol{\xi}_2 = \mathbb{E}[\mathbf{X}_2]$, $\boldsymbol{\Psi}_{22} = \mathrm{cov}[\mathbf{X}_2]$ and

$$\omega_{1.2} = \mathbb{E}\left(\frac{\nu + \delta(\mathbf{X}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})}{\nu + p_2 - 2}\right) = \left(\frac{\nu}{\nu - 2}\right)\frac{L_p(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}^*, \nu - 2)}{L_p(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, \nu)}, \tag{3.10}$$

with $\boldsymbol{\Sigma}_{22}^* = \nu\boldsymbol{\Sigma}_{22}/(\nu - 2)$. The last expression follows from Proposition 3. Finally,

$$\mathrm{cov}[\mathbf{X}] = \begin{bmatrix} \omega_{1.2}\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\big(\omega_{1.2}\mathbf{I}_{p_2} - \boldsymbol{\Psi}_{22}\boldsymbol{\Sigma}_{22}^{-1}\big)\boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Psi}_{22} \\ \boldsymbol{\Psi}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{\Psi}_{22} \end{bmatrix}, \tag{3.11}$$

where $\boldsymbol{\xi}_2$ and $\boldsymbol{\Psi}_{22}$ are the mean vector and variance-covariance matrix of the TMVT distribution, which can be computed by using (3.7) and (3.8), respectively.

In general, marginal distributions of a TMVT distribution are not TMVT, however this holds for $\mathbf{X}_2$ due to the particular case $\mathbf{a}_1 = -\boldsymbol{\infty}$ and $\mathbf{b}_1 = \boldsymbol{\infty}$. Note that obtaining (3.10) does not require the computation of additional integrals given that probabilities

$L_p(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}^*_{22}, \nu - 2)$ and $L_p(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, \nu)$ are involved in the calculation of $\boldsymbol{\xi}_2$ and $\boldsymbol{\Psi}_{22}$ (see Algorithm 2, Line 2).

It is important to emphasize that $\mathbb{E}[\mathbf{X}]$ and $\mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ exist if and only if $\nu + p_2 > 1$ and $\nu + p_2 > 2$, respectively. This is equivalent to say that, (3.9) exists if at least one dimension containing a finite limit exists. Besides, (3.11) exists if at least two dimensions containing a finite limit exist.

As can be seen, we can use equations (3.9) and (3.11) to deal with doubly infinite limits, where truncated moments are computed only over a p_2-variate partition, avoiding some unnecessary integrals and saving a significant computational cost.

3.3 A numerical illustration

In order to illustrate our method, we performed a simple simulation study to show how Monte Carlo(MC) estimators for the mean vector and variance-covariance matrix elements converge to the real values calculated by our method.

We consider a 5-variate t distribution $\mathbf{X} \sim t_5(\mathbf{0}, \boldsymbol{\Sigma}, 4)$, where $\boldsymbol{\Sigma}$ is a positive-definite matrix such that its diagonal elements are equal to one, then $\sigma_{ii} = 1$, $i = 1, \ldots, p$, and off-diagonal elements $\sigma_{ij} = \sigma_i\sigma_j$ for $i \neq j = 1, \ldots, p$, with $\boldsymbol{\sigma} = (-0.4, -0.7, 1, 0.7, 0.4)^\top$.

Let $\mathbf{Y} \overset{d}{=} \mathbf{X} \mid (\mathbf{a} \leq \mathbf{X} \leq \mathbf{b})$ be a TMVT random variable with lower and upper truncation limits $\mathbf{a} = (-\infty, -\infty, -\infty, -3, -3)^\top$ and $\mathbf{b} = (\infty, \infty, 1, 1, \infty)^\top$. Note that the first two dimensions are not truncated, while the other three are upper, interval and lower truncated, respectively. Hence, we can write $\mathbf{a} = (-\boldsymbol{\infty}_2, \mathbf{a}_2)$ and $\mathbf{b} = (\boldsymbol{\infty}_2, \mathbf{b}_2)$, with $\mathbf{a}_2 = (-\infty, -3, -3)$ and $\mathbf{b}_2 = (1, 1, \infty)$. Consider the partitions $\mathbf{X}_1 = (X_1, X_2)^\top$ and $\mathbf{X}_2 = (X_3, X_4, X_5)^\top$. In order to compute $\boldsymbol{\xi} = \mathbb{E}[\mathbf{Y}]$ and $\boldsymbol{\Omega} = \mathrm{cov}[\mathbf{Y}]$, we use relations (3.9) and (3.11) given in Subsection 3.2 because of a non-truncated partition $\mathbf{X}_1$. The computed true values are

$$\boldsymbol{\xi} = \left(\begin{array}{c} 0.167 \\ 0.292 \\ \hline \mathbf{-0.417} \\ \mathbf{-0.397} \\ \mathbf{-0.110} \end{array}\right) \quad \text{and} \quad \boldsymbol{\Omega} = \left(\begin{array}{cc|ccc} 1.355 & & & & \\ 0.224 & 1.137 & & & \\ \hline -0.321 & -0.561 & \mathbf{0.802} & & \\ -0.166 & -0.290 & \mathbf{0.414} & \mathbf{0.698} & \\ -0.101 & -0.177 & \mathbf{0.253} & \mathbf{0.131} & \mathbf{1.165} \end{array}\right).$$

In this scenario, lower partitions of $\boldsymbol{\xi}$ and $\boldsymbol{\Omega}$ (values in bold) correspond to $\boldsymbol{\xi}_2 = \mathbb{E}[\mathbf{X}_2 \mid \mathbf{a}_2 \leq \mathbf{X}_2 \leq \mathbf{b}_2]$ and $\boldsymbol{\Omega}_{22} = \mathrm{cov}[\mathbf{X}_2 \mid \mathbf{a}_2 \leq \mathbf{X}_2 \leq \mathbf{b}_2]$ due to $\mathbb{P}(\mathbf{a} \leq \mathbf{X} \leq \mathbf{b}) = \mathbb{P}(\mathbf{a}_2 \leq \mathbf{X}_2 \leq \mathbf{b}_2)$, which are computed using our recurrence-based formulas (3.7) and (3.8), while the reminder are computed using basic algebra where no integrals are needed.

Finally, we performed a MC simulation where 10000 realizations of $\mathbf{Y}$ were generated.

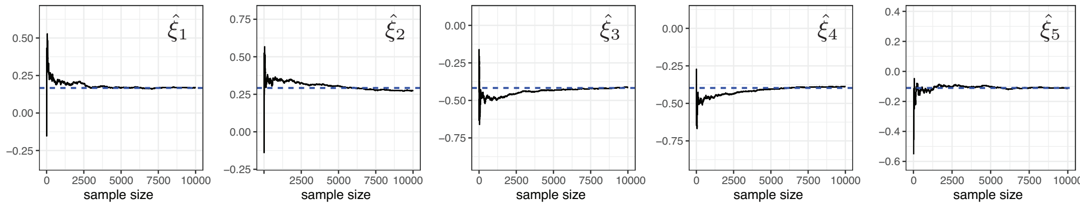

Figure 3.1: MC estimates for the elements of $\boldsymbol{\xi} = \mathbb{E}[\mathbf{Y}]$.

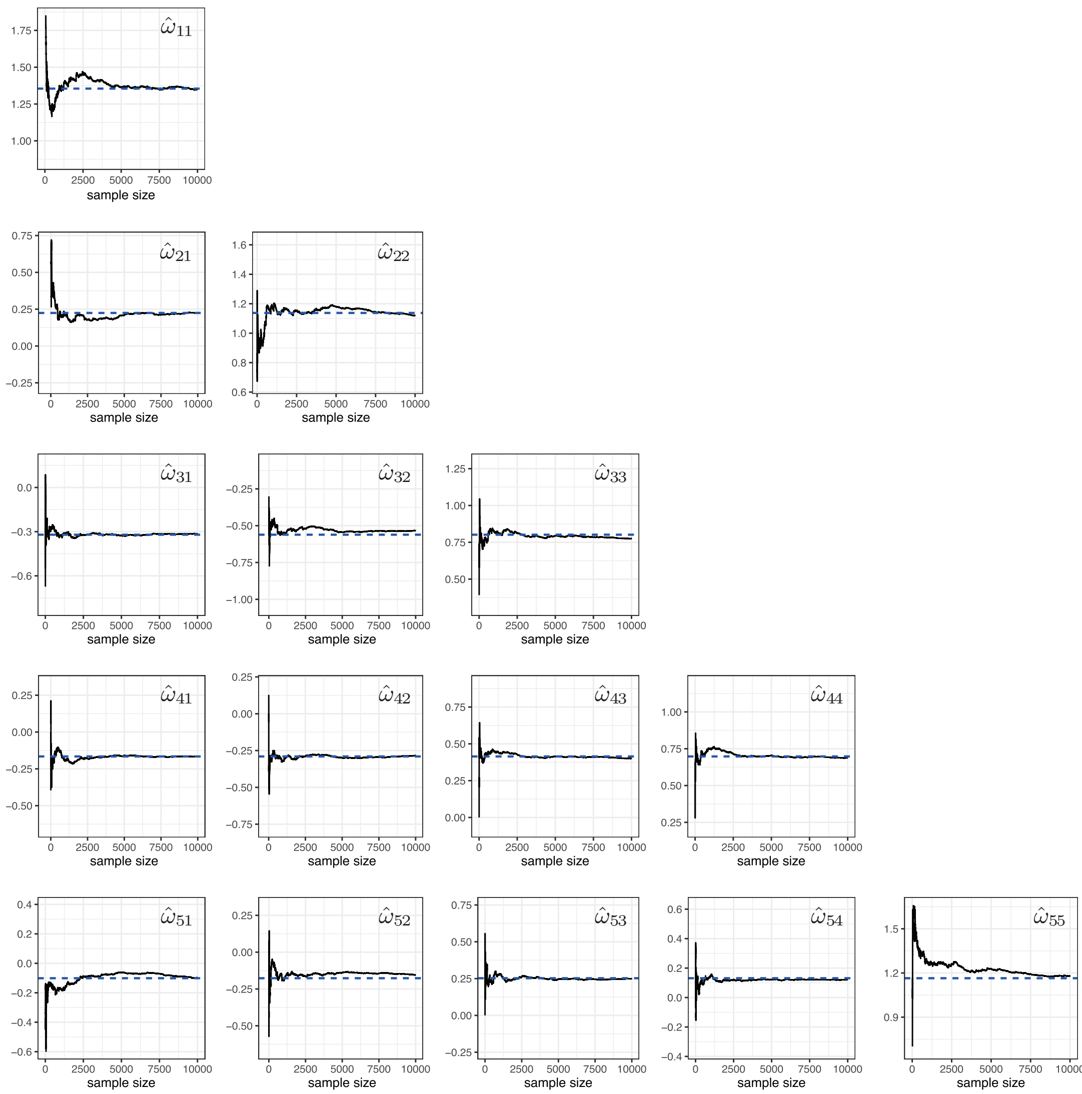

Figure 3.2: MC estimates for the distinct elements of $\boldsymbol{\Omega} = \text{cov}[\mathbf{Y}]$.

At each iteration, we compute the sample mean and the sample variance-covariance matrix. Figures 3.1 and 3.2 shows the evolution trace of the MC estimates for the distinct elements of $\boldsymbol{\xi}$ and $\boldsymbol{\Omega}$ denoted by $\hat{\xi}_i$ and $\hat{\omega}_{ij}$ for $i, j = 1, \ldots, p$, with $i \neq j$. True computed values are depicted as blue dashed lines. Note that even with 10000 MC simulations there exist slight variation in the chains for some elements as depicted in Figure 3.2.

Remark: (computational time) To compute the first two moments of $\mathbf{Y}$ using expressions $\boldsymbol{\xi}_2$ and $\boldsymbol{\Omega}_{22}$ stated in Subsection 3.2, the result is 1.2 times faster than considering the full vector $\mathbf{Y}$ with the non-truncated partition. Even though integrals involving infinite values are faster to evaluate, the number of integrals required increases exponentially as the dimension p increases. For instance, considering a vector of dimension $p = 20$, where 15 (75%) of its dimensions are non-truncated, the difference on time to compute the first two moments using expressions (3.9) and (3.11) is 10 times faster than using the crude method, which certainly is a significant difference.

3.4 Folded Multivariate Student-t distribution

Let $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, we now turn our attention to discuss the computation of any arbitrary order moment of $|\mathbf{X}|$. First, we established the following corollary.

Corollary 3. *If* $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *then* $\mathbf{Z}_s = \boldsymbol{\Lambda}_s\mathbf{X} \sim t_p(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu)$ *and consequently the joint* pdf, cdf *and the* $\boldsymbol{\kappa}$*th raw moment of* $\mathbf{Y} = |\mathbf{X}|$ *are, respectively, given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s} \in S(p)} t_p(\mathbf{y}|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu),$$

$$F_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s} \in S(p)} \pi_s T_p(\mathbf{y}_s|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$$

and

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}] = \sum_{\mathbf{s} \in S(p)} I^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu),$$

where $\mathbf{y}_s = \boldsymbol{\Lambda}_s\mathbf{y}$, $\boldsymbol{\mu}_s = \boldsymbol{\Lambda}_s\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_s = \boldsymbol{\Lambda}_s\boldsymbol{\Sigma}\boldsymbol{\Lambda}_s$ *and* $I^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu) = \int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} t_p(\mathbf{y}|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu) d\mathbf{y}$.

Proof. The proof follows straightforwardly from the definition of probability theory and basic matrix algebra and thus is omitted. □

Thus, product moments of $\mathbf{Y}$ can be easily calculated using $I^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu)$. In light of Corollary 2, we have the mean vector $\boldsymbol{\xi}$ and variance-covariance matrix $\boldsymbol{\Psi}$ of $\mathbf{Y}$, calculated as

$$\boldsymbol{\xi} = \sum_{\mathbf{s}\in S(p)} \mathbb{E}[\mathbf{Z}_s^+], \quad \text{and} \quad \boldsymbol{\Psi} = \sum_{\mathbf{s}\in S(p)} \mathbb{E}\left[\mathbf{Z}_s^+\mathbf{Z}_s^{+\top}\right] - \boldsymbol{\xi}\boldsymbol{\xi}^\top,$$

respectively, where $\mathbf{Z}_s^+$ is the positive component of $\mathbf{Z}_s = \boldsymbol{\Lambda}_s\mathbf{X} \sim t_p(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu)$ from Corollary 1. Note that there are 2^p times more integrals to be calculated as compared to the non-folded case. Specifically, $(1+p)2^p$ integrals are required for the mean vector, and additional $(1+2p+2p^2)2^p$ integrals for the variance-covariance matrix. For the univariate case, explicit expressions for the first four raw moments of $Y = |X|$, where $X \sim t(\mu, \sigma^2, \nu)$, based on (3.2) and (3.4) can be obtained as

$$\begin{aligned}
\mathbb{E}[Y] &= \mu[1 - 2T_1(0|\mu, \sigma^2, \nu)] + 2\sigma^{*2}t_1(0|\mu, \sigma^{*2}, \nu-2),\\
\mathbb{E}[Y^2] &= \mu^2 + \sigma^{*2},\\
\mathbb{E}[Y^3] &= \mu^2\mathbb{E}[Y] + 3\mu\sigma^{*2}[1 - 2T_1(0|\mu, \sigma^{*2}, \nu-2)] + \tfrac{4(\nu-2)\sigma^{*4}}{\nu-4}t_1(0|\mu, \tfrac{\nu}{\nu-4}\sigma^2, \nu-4),\\
\mathbb{E}[Y^4] &= \mu^4 + 6\mu^2\sigma^{*2} + \tfrac{3(\nu-2)}{\nu-4}\sigma^{*4}.
\end{aligned}$$

Illustrative results via the implementation of `R` package `MomTrunc` are presented in the Appendix B.

4. Application in MVT interval censored responses

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ be a $p \times 1$ response vector for the ith sample unit, for $i \in \{1, \ldots, n\}$, and consider the set of random samples (independent and identically distributed):

$$\mathbf{Y}_1, \ldots, \mathbf{Y}_n \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu), \tag{4.12}$$

where the location vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$ and the dispersion matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$ depend on an unknown and reduced parameter vector $\boldsymbol{\alpha}$. However, the response vector $\mathbf{Y}_i$ may not be fully observed due to censoring, so we define $(\mathbf{V}_i, \mathbf{C}_i)$ the observed data for the ith sample, where $\mathbf{V}_i = (V_{i1}, \ldots, V_{ip})^\top$ represents either an uncensored observation $(V_{ik} = V_{0i})$ or the interval censoring level $(V_{ik} \in [V_{1ik}, V_{2ik}])$, and $\mathbf{C}_i = (C_{i1}, \ldots, C_{ip})^\top$ is the vector of censoring indicators, satisfying

$$C_{ik} = \begin{cases} 1 & \text{if } V_{1ik} \le Y_{ik} \le V_{2ik},\\ 0 & \text{if } Y_{ik} = V_{0i}, \end{cases} \tag{4.13}$$

for all $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, p\}$, i.e., $C_{ik} = 1$ if Y_{ik} is located within a specific interval. In this case, (4.12) along with (4.13) defines the multivariate Student-t interval censored model (hereafter, the MVT-IC model). Notice that the left censoring causes truncation from the lower limit of the support of the distribution, since we only know that the true observation Y_{ik} is less than or equal to the observed quantity V_{2ik}. This situation has been studied by Lachos et al. (2017). Missing observations can be handled by considering $V_{1ik} = -\infty$ and $V_{2ik} = +\infty$.

4.1 The likelihood function

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^\top$ is a realization of $\mathbf{Y}_i \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. To obtain the likelihood function of the MVT-IC model, we firstly treat observed and censored components of $\mathbf{y}_i$, separately, i.e., $\mathbf{y}_i = (\mathbf{y}_i^{o\top}, \mathbf{y}_i^{c\top})^\top$, where $C_{ik} = 0$ for all elements in the p_i^o-dimensional vector $\mathbf{y}_i^o$, and $C_{ik} = 1$ for all elements in the p_i^c-dimensional vector $\mathbf{y}_i^c$. Accordingly, we write $\mathbf{V}_i = \mathrm{vec}(\mathbf{V}_i^o, \mathbf{V}_i^c)$, where $\mathbf{V}_i^c = (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c)$ with

$$\boldsymbol{\mu}_i = (\boldsymbol{\mu}_i^{o\top}, \boldsymbol{\mu}_i^{c\top})^\top \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \begin{pmatrix} \boldsymbol{\Sigma}_i^{oo} \boldsymbol{\Sigma}_i^{oc} \\ \boldsymbol{\Sigma}_i^{co} \boldsymbol{\Sigma}_i^{cc} \end{pmatrix}.$$

Then, using Proposition 1, we have that $\mathbf{Y}_i^o \sim t_{p_i^o}(\boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \nu)$ and $\mathbf{Y}_i^c \mid \mathbf{Y}_i^o = \mathbf{y}_i^o \sim t_{p_i^c}(\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + p_i^o)$, where

$$\boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co} \boldsymbol{\Sigma}_i^{oo-1} (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o), \quad \mathbf{S}_i^{co} = \left\{ \frac{\nu + \delta(\mathbf{y}_i^o)}{\nu + p_i^o} \right\} \boldsymbol{\Sigma}_i^{cc.o},$$

$$\boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co} (\boldsymbol{\Sigma}_i^{oo})^{-1} \boldsymbol{\Sigma}_i^{oc} \quad \text{and} \quad \delta(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)^\top (\boldsymbol{\Sigma}_i^{oo})^{-1} (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o). \tag{4.14}$$

Let $\mathbf{V} = \mathrm{vec}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = \mathrm{vec}(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ denote the observed data. Therefore, the log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\alpha}^\top, \nu)^\top$, where $\boldsymbol{\alpha} = \mathrm{vech}(\boldsymbol{\Sigma})$, for the observed data $(\mathbf{V}, \mathbf{C})$ is

$$\ell(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{C}) = \sum_{i=1}^{n} \ln L_i, \tag{4.15}$$

where L_i represents the likelihood function of $\boldsymbol{\theta}$ for the ith sample, given by

$$\begin{aligned} L_i \equiv & L_i(\boldsymbol{\theta} \mid \mathbf{V}_i, \mathbf{C}_i) = f(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}) = f(\mathbf{V}_{1i}^c \le \mathbf{y}_i^c \le \mathbf{V}_{2i}^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o \mid \boldsymbol{\theta}) \\ & = L_{p_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + p_i^o) t_{p_i^o}(\mathbf{y}_i^o \mid \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \nu). \end{aligned}$$

4.2 Parameter estimation via the EM algorithm

We describe how to carry out ML estimation for the MVT-IC model. The EM algorithm, originally proposed by Dempster et al. (1977), is a very popular iterative optimization strategy and commonly used to carry out ML estimation for the models with incomplete-data problems. This algorithm has many attractive features such as numerical stability, simplicity of implementation and quite reasonable memory requirements (McLachlan and Krishnan, 2008).

By the essential property of MVT distribution, we write $\mathbf{Y}_i|(U_i = u_i) \sim N_p(\boldsymbol{\mu}, u_i^{-1}\boldsymbol{\Sigma})$ and $u_i \sim \text{Gamma}(\nu/2, \nu/2)$. Fixing the value of ν, the complete-data log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by $\ell_c(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_{ic}(\boldsymbol{\theta})$, where the individual complete-data log-likelihood is

$$\ell_{ic}(\boldsymbol{\theta}) = -\frac{1}{2}\big\{\ln|\boldsymbol{\Sigma}| + u_i(\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\big\} + \ln h(u_i \mid \nu) + c,$$

where c is a constant irrelevant of $\boldsymbol{\theta}$ and $h(u_i \mid \nu)$ is the pdf of $\text{Gamma}(\nu/2, \nu/2)$. In summary, the EM algorithm for the MVT-IC model can be adopted as follows:

E-step: Given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\hat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{(k)})$ at the kth step, the E-step provides the conditional expectation of the complete-data log-likelihood function:

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = \mathbb{E}\Big[\ell_c(\boldsymbol{\theta}) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\Big] = \sum_{i=1}^n Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}),$$

where

$$Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\text{tr}\left[\left\{\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\boldsymbol{\mu}^\top - \boldsymbol{\mu}(\widehat{u\mathbf{y}}_i^{(k)})^\top + \widehat{u}_i^{(k)}\boldsymbol{\mu}\boldsymbol{\mu}^\top\right\}\boldsymbol{\Sigma}^{-1}\right],$$

with $\widehat{u\mathbf{y}}_i^{(k)} = \mathbb{E}[U_i\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{u\mathbf{y}_i^2}^{(k)} = \mathbb{E}[U_i\mathbf{Y}_i\mathbf{Y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ and $\widehat{u}_i^{(k)} = \mathbb{E}[U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ which are collected in Appendix A. Note that, since ν is fixed, the calculation of $\mathbb{E}[\ln h(U_i \mid \nu) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}]$ is unnecessary.

M-step: Conditionally maximizing $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$ with respect to each entry of $\boldsymbol{\theta}$, we update the estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\hat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{(k)})$ by

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}^{(k+1)} &= \left(\sum_{i=1}^n \widehat{u}_i^{(k)}\right)^{-1} \sum_{i=1}^n \widehat{u\mathbf{y}}_i^{(k)}, \\
\widehat{\boldsymbol{\Sigma}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^n \left\{\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\mu}}^{(k+1)\top} - \widehat{\boldsymbol{\mu}}^{(k+1)}(\widehat{u\mathbf{y}}_i^{(k)})^\top + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\mu}}^{(k+1)}\widehat{\boldsymbol{\mu}}^{(k+1)\top}\right\}.
\end{aligned}$$

After the M-step, we will update the parameter $\boldsymbol{\nu}$ by maximizing the marginal log-likelihood function of $\mathbf{y}$, that is,

$$\widehat{\nu}^{(k+1)} = \arg\max_{\nu} \sum_{i=1}^n \log f(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}; \nu).$$

The algorithm is iterated until a suitable convergence rule is satisfied. In the later analysis, the algorithm is terminated when the difference between two successive evaluations of the log-likelihood defined in (4.15) is less than a tolerance, i.e.,

$\ell(\widehat{\boldsymbol{\theta}}^{(k+1)} \mid \mathbf{V}, \mathbf{C}) - \ell(\widehat{\boldsymbol{\theta}}^{(k)} \mid \mathbf{V}, \mathbf{C}) < \epsilon$, for example, $\epsilon = 10^{-6}$.

4.3 Application: concentration levels data

We applied our method to the dataset consisting of concentration levels of certain dissolved trace metals in freshwater streams across the Commonwealth of Virginia. The data were provided by Virginia Department of Environment Quality (VDEQ) and were previously analyzed by Hoffman and Johnson (2015), who proposed a pseudo-likelihood approach for estimating parameters of multivariate normal and log-normal models. It is very important to determine the quality of Virginia's water resources across the state to guide their safe use. The methodology adopted must neither underestimate nor overestimate the levels of contamination, as otherwise, the results can compromise public health, environmental safety or can unfairly restrict local industry.

Specifically, this dataset consists of the concentration levels of the dissolved trace metals copper (Cu), lead (Pb), zinc (Zn), calcium (Ca) and magnesium (Mg) from 184 independent randomly selected sites in freshwater streams across Virginia. The Cu, Pb, and Zn concentrations are reported in μg/L of water, whereas Ca and Mg concentrations are suitably reported in mg/L of water. Since the measurements are taken at different times, the presence of multiple limit of detection values is possible for each trace metal (VDEQ, 2019). The limit of detection is 0.1μg/L for Cu and Pb, 1.0mg/L for Zn, 0.5mg/L for Ca and 1.0mg/L for Mg.

The percentages of left-censored values are 2.7% for Ca, 4.9% for Cu, 9.8% for Mg, which are small in comparison to 78.3% for Pb and 38.6% for Zn. Also note that 17.9% of the streams had 0 non-detected trace metals, 39.1% had 1, 37.0% had 2, 3.8% had 3, 1.1% had 4 and 1.1% had 5.

We propose a MVT-IC model to fit the data, now with dimension $p = 5$, that is, $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i5}) \sim t_5(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. For the sake of comparison, we also fit a multivariate MVN-IC model which can be consider as a limiting case when $\nu \to \infty$.

As concentration levels are strictly positive measures, to guarantee this, we consider an interval-censoring analysis by setting all lower limit of detection equal to 0 for all trace metals. Also, due to different scales for each trace metal, we standardize the dataset to have zero mean and variance equal to one as in Wang et al. (2019). The above mentioned work considered this study as a left-censoring problem without taking in account the possibility of predicting negative concentration levels for trace metals. For instance, we can see from Figure 4.4 that Pb censored concentrations take values on the small interval

$[0, 0.1]$.

The ML estimates of model parameters were obtained using the EM algorithm described in Subsection 4.2. The estimated mean of the trace metals, degrees of freedom ν and the log-likelihood as well as AIC are shown in Table 4.1. Here, we can see that estimated mean values are quite similar under both models. Notice that the estimated value of ν is fairly small, indicating a lack of adequacy of the normal assumption for the VDEQ data. This finding can also be confirmed from Figure 4.3 where the profile log-likelihood values are depicted for a grid of values of ν. The AIC value for our MVT-IC model is lower than that for the MVN-IC model as expected.

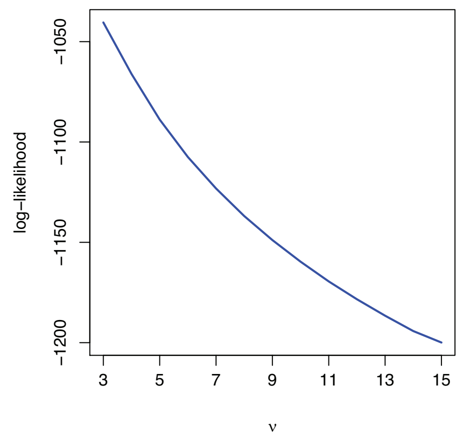

Figure 4.3: VDEQ data. Plot of the profile log-likelihood of the degrees of freedom ν.

Figure 4.4 shows histograms and pair-wise scatter plots for the concentration levels study. From histograms (diagonal of the matrix plot), we can see how censored observations (taking values over the dashed lines) are distributed to the left (blue bins) after fitting our proposed model, while gray bins represent complete observed points. On the other hand, scatter plots (off-diagonal of the matrix plot) show complete observed (black) points and predicted observations using the multivariate SN-C model (blue triangles).

Finally, with the aim of validating the proposed censored model, we compare the

Table 4.1: VDQE data. Estimated mean and ML estimate for ν and model criteria.

Model	Cu	Pb	Zn	Ca	Mg	ν	$\ell(\hat{\boldsymbol{\theta}}\mid\mathbf{Y})$	AIC
Normal	0.556	0.099	2.314	12.083	3.814	-	-1351.60	2743.19
Student-t	0.557	0.102	2.329	12.084	3.817	3	-1040.21	2120.42

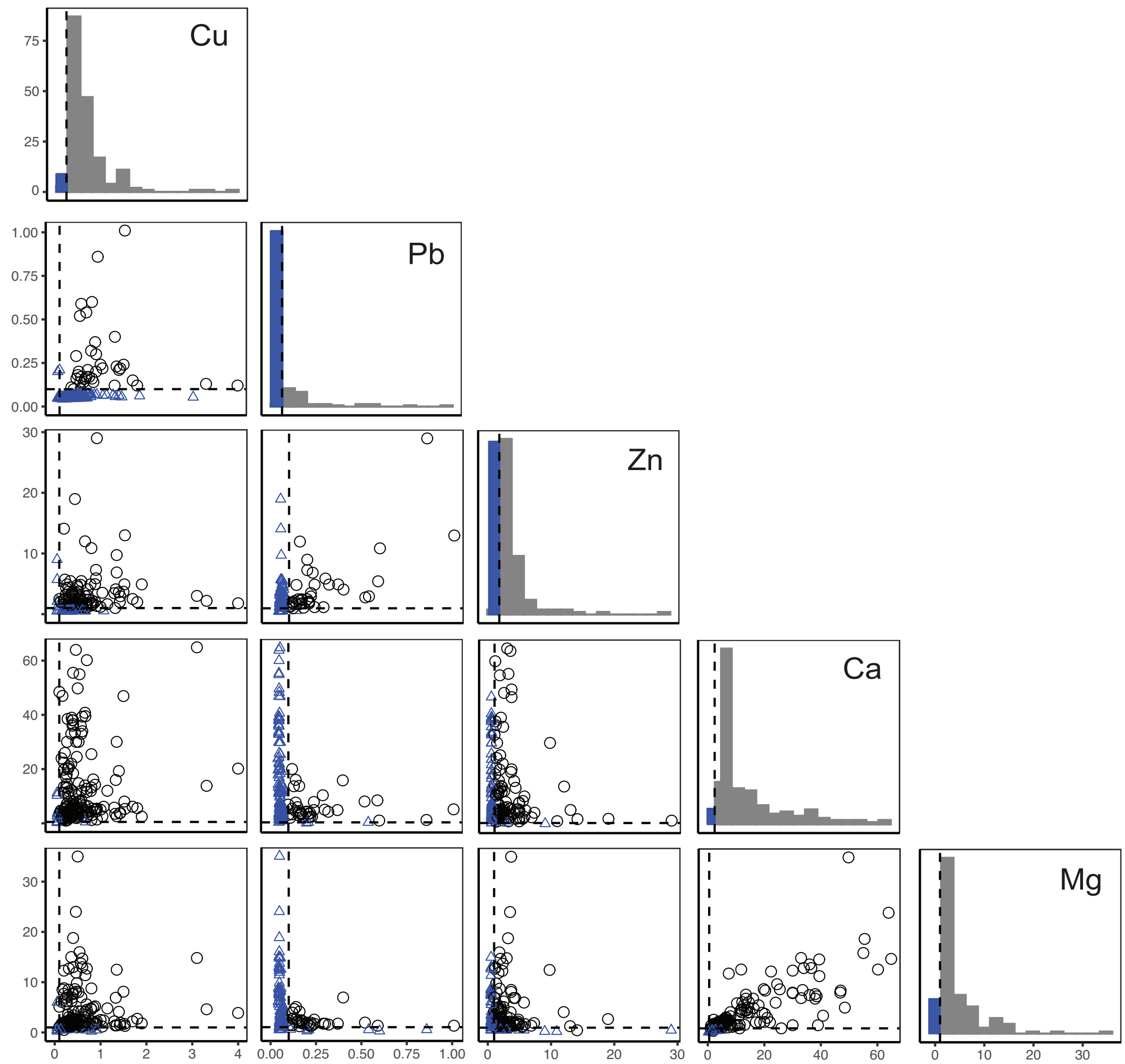

Figure 4.4: VDEQ data. Histograms (diagonal) and pair-wise scatter plots (lower-triangle) for the concentration levels study. Complete observed points are represented in black points (gray bins) and T predicted observations in blue triangles (bins). Limit of detection are represented in dashed lines.

correlation matrices of the data by considering 5 strategies:

(a) *Original:* original data

(b) *Omitting:* zeros are not considered

(c) *Manipulating:* multiplying the limit of detection by the factor 0.75

(d) MVN-IC model

(e) MVT-IC model

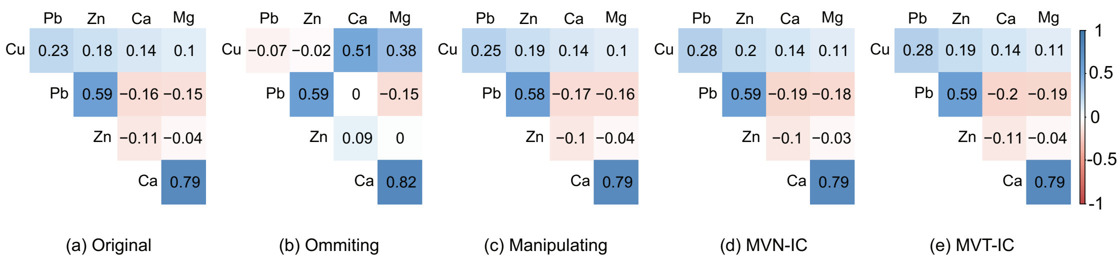

Figure 4.5: VDQE data. Correlation matrices of concentration levels for 5 different strategies.

From the results depicted in Figure 4.5, we find that the correlation matrices for the MVN-IC and MVT-IC models are similar. Based on the AIC, we consider the second one as a reference. We can get very decent results for this study by using (a) original data or (c) even manipulating the data, with both tending to underestimate correlations. Omitting (b) is by far the worst strategy. For example, the correlation between the Pb and Cu is poorly estimated to the point that they have the sign changed. Similar problems arise for correlations between Zn and other three elements. Given the large number of censored observations, omitting leads to loss of information (as is the case of the correlation between Ca and Pb, as well as between Ca and Mg, where correlation was estimated to be zero).

5. Conclusions

In this paper, we have developed recurrence relations for integrals that involve the density of MVT distributions. These recursions allow fast computation of arbitrary-order product moments of TMVT and FMVT distributions. Explicit expressions for the first two moments of the TMVT and FMVT are also provided. For the reader who is interested in real-world applications, we have shown the practicability of our results through a real data example that contains positive censored observations. Our methods can be also applied in the context of missing observations (Lin et al., 2009). The proposed methodology can been implemented by `R MomTrunc` package, which is available on CRAN repository.

We conjecture that our method can be extended to more complicated cases such as the multivariate skew-t distribution (Azzalini and Capitanio, 2003) and scale mixtures of normal distributions (Branco and Dey, 2001). Also, our censored/missing data model can be extended for the generalized hyperbolic distribution (which includes the Student-t as a particular case), where its first two moments are given explicitly in (Roozegar et al., 2020). An in-depth investigation of such extensions is beyond the scope of the present paper, but it is an interesting topic for further research.

Conflict of interest statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgment

We would like to thank the Associate Editor and two reviewers for their constructive comments, which helped to improve this paper substantially. C. Galarza acknowledges support from FAPESP-Brazil (Grant 2015/17110-9 and Grant 2018/11580-1). T.I. Lin and W.L. Wang would like to acknowledge the support of the Ministry of Science and Technology of Taiwan under Grant Nos. MOST 109-2118-M-005-005-MY3 and MOST 107-2628-M-035-001-MY3, respectively.

Bibliography

Arellano-Valle, R.B., Bolfarine, H., 1995. On some characterizations of the t-distribution. Statistics & Probability Letters 25, 79–85.

Arismendi, J.C., 2013. Multivariate truncated moments. Journal of Multivariate Analysis 117, 41–75.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12, 171–178.

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65, 2, 367–389.

Branco, M. D., Dey, K., 2001. A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis 79, 99–113.

Chakraborty, A.K., Chatterjee, M., 2013. On multivariate folded normal distribution. Sankhya B 75, 1–15.

De Bastiani, F., de Aquino Cysneiros, A.H.M., Uribe-Opazo, M.A., Galea, M., 2015. Influence diagnostics in elliptical spatial linear models. Test 24, 322–340.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 39, 1–38.

Fang, K.T., Kotz, S., Ng, K.W., 1990. Symmetric multivariate and related distribuitions. Chapman & Hall, London.

Flecher, C., Allard, D., Naveau, P., 2010. Truncated skew-normal distributions: moments, estimation by weighted moments and application to climatic data. Metron 68, 331–345.

Fonseca, T.C., Ferreira, M.A., Migon, H.S., 2008. Objective bayesian analysis for the student-t regression model. Biometrika 95, 325–333.

Galarza, C.E., Kan, R., Lachos, V.H., 2020. MomTrunc: Moments of Folded and Doubly Truncated Multivariate Distributions. R package version 5.69. URL: `https://CRAN.R-project.org/package=MomTrunc`.

Genç, A.İ., 2013. Moments of truncated normal/independent distributions. Statistical Papers 54, 741–764.

Ho, H.J., Lin, T.I., Chen, H.Y., Wang, W.L., 2012. Some results on the truncated multivariate t distribution. Journal of Statistical Planning and Inference 142, 25–40.

Hoffman, H.J., Johnson, R.E.. Pseudo-likelihood estimation of multivariate normal parameters in the presence of left-censored data. Journal of Agricultural, Biological, and Environmental Statistics 20, 1, 156–171.

Jawitz, J.W., 2004. Moments of truncated continuous univariate distributions. Advances in Water Resources 27, 269–281.

Kan, R., Robotti, C., 2017. On moments of folded and truncated multivariate normal distributions. Journal of Computational and Graphical Statistics 25, 930–934.

Kim, H.M., 2008. A note on scale mixtures of skew normal distribution. Statistics and Probability Letters 78. 1694-1701.

Lachos, V.H., Moreno, E.J.L., Chen, K., Cabral, C.R.B., 2017. Finite mixture modeling of censored data using the multivariate student-t distribution. Journal of Multivariate Analysis 159. 151-167.

Lien, D.H.D., 1985. Moments of truncated bivariate log-normal distributions. Economics Letters 19, 243–247.

Lin, Tsung-I, Wang, Wan-Lun, 2017. Multivariate-t nonlinear mixed models with application to censored multi-outcome AIDS studies. Biostatistics 18,4, 666–681.

Lin, Tsung-I, Wang, Wan-Lun, 2020. Multivariate-t linear mixed models with censored responses, intermittent missing values and heavy tails. Statistical Methods in Medical Research 29, 5,1288–1304.

Lin, Tsung I, Ho, Hsiu J, Chen, Chiang L, 2009. Analysis of multivariate skew normal models with incomplete data. Journal of Multivariate Analysis 100, 10, 2337–2351.

Little, R.J.A., Rubin, D.B., 1987. Statistical Analysis With Missing Data. John Wiley & Sons, New York.

Matos, L.A., Prates, M.O., Chen, M.H., Lachos, V.H., 2013. Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. Statistica Sinica 23, 1323–1342.

McLachlan, G.J., Krishnan, T., 2008. The EM Algorithm and Extensions. 2 ed., Wiley.

Peel, D., McLachlan, G.J., 2000. Robust mixture modelling using the t distribution. Statistics and Computing 10, 339–348.

Pinheiro, J.C., Liu, C.H., Wu, Y.N., 2001. Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. Journal of Computational and Graphical Statistics 10, 249–276.

Roozegar, R., Balakrishnan, N., Jamalizadeh, A., 2020. On moments of doubly truncated multivariate normal mean-variance mixture distributions with application to multivariate tail conditional expectation. Journal of Multivariate Analysis 177, 104586.

Savalli, C., Paula, G.A., Cysneiros, F.J., 2006. Assessment of variance components in elliptical linear mixed models. Statistical Modelling 6, 59–76.

Tallis, G.M., 1961. The moment generating function of the truncated multi-normal distribution. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 23, 223–229.

VDEQ, 2003. The Quality of Virginia Non-Tidal Streams: First Year Report. Richmond, Virginia. VDEQ Technical Bulletin. WQA/2002-001.

Wang, W.L., Castro, L.M., Lin, T.I., 2017. Automated learning of t factor analysis models with complete and incomplete data. Journal of Multivariate Analysis 161, 157–171.

Wang, W.L., Fan, T.H., 2011. Estimation in multivariate t linear mixed models for multiple longitudinal data. Statistica Sinica 21, 1857–1880.

Wang, W.L., Lin, T.I., 2014. Multivariate t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Statistics in Medicine 33, 3029–3046.

Wang, W.L., Lin, T.I., 2015. Bayesian analysis of multivariate t linear mixed models with missing responses at random. Journal of Statistical Computation and Simulation 85, 3594–3612.

Wang, W.L., Lin, T.I., 2015. Bayesian analysis of multivariate t linear mixed models with missing responses at random. Journal of Statistical Computation and Simulation 85, 3594–3612.

Wang, W.L., Castro, L.M., Lachos, V.H., Lin, T.I., 2019. Model-based clustering of censored data via mixtures of factor analyzers. Computational Statistics & Data Analysis 140, 104–121.

Wang, Wan-Lun, Lin, Tsung-I, Lachos, Victor H., 2018 Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. Statistical Methods in Medical Research 27, 1, 48–64.

Appendix

Appendix A: Details for the expectations in EM algorithm

To compute the required expected values of all latent data, we find that most of them can be written in terms of $\mathbb{E}(U_i \mid \mathbf{Y}_i)$, and thereby we write $\widehat{u}_i = \mathbb{E}\{\mathbb{E}(U_i \mid \mathbf{Y}_i) \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\}$, where $\mathbb{E}(U_i \mid \mathbf{Y}_i) = (\nu + p)/(\nu + \delta)$ with $\delta = (\mathbf{Y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \boldsymbol{\mu})$. Subsequently, we discuss the closed-form expressions of conditional expectations as follows:

1. If the ith subject has only non-censored components, then

$$\widehat{u\mathbf{y}^2_i}^{(k)} = \left\{\frac{\nu + p}{\nu + \widehat{\delta}^{(k)}(\mathbf{y}_i)}\right\}\mathbf{y}_i\mathbf{y}_i^\top, \quad \widehat{u\mathbf{y}}_i^{(k)} = \left\{\frac{\nu + p}{\nu + \widehat{\delta}^{(k)}(\mathbf{y}_i)}\right\}\mathbf{y}_i, \quad \widehat{u}_i^{(k)} = \left\{\frac{\nu + p}{\nu + \widehat{\delta}^{(k)}(\mathbf{y}_i)}\right\},$$

   where $\widehat{\delta}^{(k)}(\mathbf{y}_i) = (\mathbf{y}_i - \widehat{\boldsymbol{\mu}}^{(k)})^\top (\widehat{\boldsymbol{\Sigma}}^{(k)})^{-1}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}^{(k)})$.

2. If the ith subject has only censored components, from Proposition 3 with $r = 1$, we have

$$\widehat{u\mathbf{y}^2_i}^{(k)} = \mathbb{E}[U_i\mathbf{Y}_i\mathbf{Y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\varphi}^{(k)}(\mathbf{V}_i)\widehat{\mathbf{w}}_i^{2c(k)},$$

$$\begin{aligned}\widehat{u\mathbf{y}}_i^{(k)} &= \mathbb{E}[U_i\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\varphi}^{(k)}(\mathbf{V}_i)\widehat{\mathbf{w}}_i^{c(k)},\\ \widehat{u}_i^{(k)} &= \mathbb{E}[U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\varphi}^{(k)}(\mathbf{V}_i),\end{aligned}$$

where

$$\widehat{\varphi}^{(k)}(\mathbf{V}_i) = \frac{L_p(\mathbf{V}_{1i}, \mathbf{V}_{2i}; \widehat{\boldsymbol{\mu}}^{(k)}, \quad \widehat{\boldsymbol{\Sigma}}^{*(k)}, \nu+2)}{L_p(\mathbf{V}_{1i}, \mathbf{V}_{2i}; \widehat{\boldsymbol{\mu}}^{(k)}, \quad \widehat{\boldsymbol{\Sigma}}^{(k)}, \nu)},$$

$$\widehat{\mathbf{w}}_i^{c(k)} = \mathbb{E}[\mathbf{W}_i \mid \widehat{\boldsymbol{\theta}}^{(k)}], \quad \widehat{\mathbf{w}}_i^{2^{c(k)}} = \mathbb{E}[\mathbf{W}_i\mathbf{W}_i^\top \mid \widehat{\boldsymbol{\theta}}^{(k)}], \tag{5.16}$$

with $\mathbf{W}_i \sim Tt_p(\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{*(k)}, \nu+2; (\mathbf{V}_{1i}, \mathbf{V}_{2i}))$ and $\widehat{\boldsymbol{\Sigma}}^{*(k)} = \dfrac{\nu}{\nu+2}\widehat{\boldsymbol{\Sigma}}^{(k)}$. To compute $\mathbb{E}[\mathbf{W}_i]$ and $\mathbb{E}[\mathbf{W}_i\mathbf{W}_i^\top]$ we use the results given in Subsection 3.1.

3. If the ith subject has both censored and uncensored components, then $(\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i)$, $(\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \mathbf{y}_i^o)$, and $(\mathbf{Y}_i^c \mid \mathbf{V}_i, \mathbf{C}_i, \mathbf{y}_i^o)$ are equivalent processes. We obtain

$$\widehat{u\mathbf{y}^2_i}^{(k)} = \mathbb{E}(U_i\mathbf{Y}_i\mathbf{Y}_i^\top \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}) = \begin{pmatrix} \mathbf{y}_i^o\mathbf{y}_i^{o\top}\widehat{u}_i^{(k)} & \widehat{u}_i^{(k)}\mathbf{y}_i^o\widehat{\mathbf{w}}_i^{c(k)\top} \\ \widehat{u}_i^{(k)}\widehat{\mathbf{w}}_i^{c(k)}\mathbf{y}_i^{o\top} & \widehat{u}_i^{(k)}\widehat{\mathbf{w}}_i^{2^{c(k)}} \end{pmatrix},$$

$$\widehat{u\mathbf{y}}_i^{(k)} = \mathbb{E}(U_i\mathbf{Y}_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{vec}(\mathbf{y}_i^o\widehat{u}_i^{(k)}, \widehat{u}_i^{(k)}\widehat{\mathbf{w}}_i^{c(k)}),$$

$$\widehat{u}_i^{(k)} = \mathbb{E}(U_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}) = \left\{\frac{p_i^o+\nu}{\nu+\widehat{\delta}^{(k)}(\mathbf{y}_i^o)}\right\}\frac{L_{p_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \widetilde{\mathbf{S}}_i^{co(k)}, \nu+p_i^o+2)}{L_{p_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \widetilde{\mathbf{S}}_i^{co(k)}, \nu+p_i^o)},$$

where

$$\widetilde{\mathbf{S}}_i^{co(k)} = \left\{\frac{\nu+\widehat{\delta}^{(k)}(\mathbf{y}_i^o)}{\nu+2+p_i^o}\right\}\widehat{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \quad \widehat{\delta}^{(k)}(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \widehat{\boldsymbol{\mu}}_i^{o(k)})^\top(\widehat{\boldsymbol{\Sigma}}_i^{oo(k)})^{-1}(\mathbf{y}_i^o - \widehat{\boldsymbol{\mu}}_i^{o(k)}),$$

$\widehat{\boldsymbol{\Sigma}}_i^{cc.o(k)}$ is defined as in equation (4.22) in the main document, $\widehat{\mathbf{w}}_i^{c(k)}$ and $\widehat{\mathbf{w}}_i^{2^{c(k)}}$ are defined in (5.16) with $\mathbf{W}_i \sim Tt_{p_i^c}(\widehat{\boldsymbol{\mu}}_i^{co(k)}, \widetilde{\mathbf{S}}_i^{co(k)}, \nu+p_i^o+2; (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c))$. Similarly, to compute $\mathbb{E}[\mathbf{W}_i]$ and $\mathbb{E}[\mathbf{W}_i\mathbf{W}_i^\top]$, we use the results given in Subsection 3.1.

Appendix B: Some illustrations using the R MomTrunc package

```
> momentsTMD(kappa=c(2,2,2),lower,upper,mu,Sigma,nu,dist = "t")


Call:
momentsTMD(kappa = c(2, 2, 2), lower, upper, mu,Sigma, dist = "t", nu)


k1 k2 k3    F(k)    E[k]
```

```
1  2  2  2  0.0002  0.0017
2  1  2  2 -0.0003 -0.0021
3  0  2  2  0.0021  0.0172
4  0  1  2 -0.0002 -0.0019
5  0  0  2  0.0161  0.1346
6  0  0  1  0.0089  0.0743
7  0  0  0  0.1194  1.0000


> meanvarTMD(lower,upper,mu,Sigma,nu,dist = "t")

#Comparing results and times with 5000 MC simulations
> means
mean1   mean2   mean3   mean4   mean5   mean6
Proposed    -0.3587 -0.0837 -0.0781  0.2745  0.8097  0.9313
MonteCarlo  -0.3465 -0.0744 -0.0730  0.2912  0.8022  0.9327

> variances
var1    var2    var3    var4    var5    var6
Proposed      0.0807  0.0863  0.1018  0.1340  0.0962  0.1459
MonteCarlo    0.0787  0.0888  0.0992  0.1393  0.0890  0.1464

> times
Proposed  MonteCarlo
3.50       11.89     seconds
```