Finite mixture modeling of censored data using the multivariate Student-t distribution

Victor H. Lachos[a] *, Edgar J. López Moreno[a] and Kun Chen[b]

[a] *Departamento de Estatística, Universidade Estadual de Campinas, Brazil*

[b] *Department of Statistics, University of Connecticut, USA*

Abstract

Finite mixture models have been widely used for the modelling and analysis of data from a heterogeneous population. Moreover, these kind of data can be subjected to some upper and/or lower detection limits because of the restriction of experimental apparatus. Another complication arises when measures of each population depart significantly from normality, for instance, in the presence of heavy tails or atypical observations. For such data structures, we propose a robust model for censored data based on finite mixtures of multivariate Student-t distributions. This approach allows us to model data with great flexibility, accommodating multimodality, heavy tails and also skewness depending on the structure of the mixture components. We develop an analytically simple yet efficient EM-type algorithm for conducting maximum likelihood estimation of the parameters. The algorithm has closed-form expressions at the E-step, that rely on formulas for the mean and variance of the multivariate truncated Student-t distributions. Further, a general information-based method for approximating the asymptotic covariance matrix of the estimators is also presented. Results obtained from the analysis of both simulated and real data sets are reported to demonstrate the effectiveness of the proposed methodology. The proposed algorithm and methods are implemented in the new R package `CensMixReg`.

Keywords: Censored data, Detection limit, EM-type algorithms, Finite mixture models, Multivariate Student-t.

1 Introduction

The occurrence of censored data due to limit of detection (LOD) is a common problem in many fields, like econometrics, geostatistics, clinical trials, medical surveys, environmental analysis, among others. For example, environmental monitoring of different variables often involves left-censored observations falling below the minimum LOD of the instruments used to quantify them. In AIDS research, the viral load measures may be subject to some upper and lower detection limits,

*Corresponding author. Address for correspondence: Departamento de Estatística, Rua Sérgio Buarque de Holanda, 651, Cidade Universitária Zeferino Vaz, Campinas, São Paulo, Brazil. CEP 13083-859. e-mail: `hlachos@ime.unicamp.br`

below or above which they are not quantifiable. As a result, the viral load responses are either left or right censored depending on the diagnostic assays used (Vaida & Liu, 2009). In econometrics, the study of the labor force participation of married women is usually conducted under the censored Tobit model (see, for instance, Chib, 1992). In this case, the observed response is the wage rate, which is typically considered as censored below zero, i.e., for working women, positive values for the wage rates are registered, whereas for non-working women, the observed wage rate is zero (Arellano-Valle *et al.*, 2012).

The proportion of censored data in these studies may not be small, so the use of crude/ad hoc methods, such as substituting a threshold value or some arbitrary point like a midpoint between zero and cutoff for detection, might lead to severe bias in statistical estimation. In the past few decades, several alternative approaches have been developed to handle the censored data. Vaida & Liu (2009) proposed an exact EM algorithm for maximum likelihood (ML) estimation in mixed effects models for censored data (LMEC/NLMEC), which uses closed-form expressions at the E-step. Further, Matos *et al.* (2013a) developed diagnostic measures for assessing the local influence in LMEC/NLMEC models. Militino & Ugarte (1999) developed an EM algorithm for conducting ML estimation in censored spatial data. De Oliveira (2005) adopted a Bayesian approach to make inference and prediction with spatially correlated censored observations. For mathematical tractability, a normal distribution was assumed for modeling the censored data. However, it is well-known that real-world phenomena are not always in agreement with this assumption, often producing data from a distribution with heavier tails, skewness or multimodality. Hence, from a practical perspective, there is a need to seek an appropriate theoretical model that avoids data transformations, yet preserves a robust and convenient Gaussian-like framework.

Many extensions of the classic multivariate Gaussian censored (nMC) model have been proposed to broaden the applicability of linear regression analysis to situations where the Gaussian error assumption may be inadequate. For instance, Arellano-Valle *et al.* (2012) (see also, Massuia *et al.*, 2015) proposed the Student-t censored regression model. Garay *et al.* (2014) (see also, Matos *et al.*, 2013b) advocated the use of the multivariate Student-t distribution in the context of censored regression (tMC) models, where a simple and efficient EM-type algorithm for iteratively computing ML estimates of the parameters was also presented. Castro *et al.* (2015) proposed a likelihood- based estimation for a multivariate Tobit confirmatory factor analysis model using the Student-t distribution (t-TCFA model). More recently, Wang *et al.* (2016) proposed a multivariate extension of a model introduced by Garay *et al.* (2014) and Matos *et al.* (2013b) for analyzing multi-outcome longitudinal data with censored observations where they established a feasible EM algorithm that admits closed-form expressions at E-steps and tractable solutions at M-steps. They demonstrated its robustness aspects against outliers through extensive simulations. A common drawback of these proposals is that they are not appropriate when the observed data exhibit, for instance, multimodality, heavy tails and skewness, simultaneously.

In the context of finite mixture of censored models, Karlsson & Laitila (2014) proposed an EM algorithm to estimate the parameters, and compared their method with those proposed by Powell (1984), Powell (1986) and Caudill (2012). In a multivariate setting, He (2013) proposed a Gaussian mixture model to flexibly approximate the underlying distribution of the observed data due to its good approximation capability and generation mechanism, where to cope with the censored data, an EM algorithm in a multivariate setting was developed. These methods are doubtlessly very flexible, but the problems related to the simultaneous occurrence of skewness, anomaly observations and multimodality can remain. Even when modeling using normal mixtures, overestimation of the number of components (that is, the number of densities in the mixture of the random error) necessary to capture the asymmetric and/or heavy-tailed nature of each subpopulation can occur. Thus in this article we propose a robust mixture model for censored data based on the multivariate Student-t distribution (FM-tMC model) by extending the mixture of normal mixtures proposed by He (2013). More specifically, our objectives are: (i) to propose a multivariate mixture model for censored data (and associated likelihood inference) based on the mixtures of multivariate Student-t distribution (ii) to implement and evaluate the proposed method computationally; and (iii) to apply these results to the analysis of a real-life dataset.

The remainder of the paper is organized as follows. In Section 2, we briefly discuss some preliminary results related to the truncated multivariate Student-t distribution and some of its key properties. In addition, we present the tMC model proposed by Garay *et al.* (2014) and the related ML estimation. In Section 3, we present the robust FM-tMC model, including the EM algorithm for ML estimation, and derive the empirical information matrix analytically to obtain the standard errors. In Sections 4 and 5, numerical examples using both simulated and real data are given to illustrate the performance of the proposed method. Finally, some concluding remarks are presented in Section 6.

2 The multivariate Student-t censored regression model

2.1 Preliminaries

In this section, we present some useful results associated to the p-variate Student-t distribution that will be needed for implementing the EM algorithm for ML estimation. We start with the probability density function (*pdf*) of a Student-t random vector $\mathbf{Y} \in \mathbb{R}^p$ with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and ν degrees of freedom. Its *pdf* is given by

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

where $\Gamma(\cdot)$ is the standard gamma function and $Q(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ is the Mahalanobis distance. The notation adopted for the Student-t *pdf* is $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$.

The cumulative distribution function (*cdf*) is denoted by $T_p(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. It is important to stress that if $\nu > 1$, the mean of $\mathbf{y}$ is $\boldsymbol{\mu}$ and if $\nu > 2$, the covariance matrix is given by $\nu(\nu - 2)^{-1}\boldsymbol{\Sigma}$. Moreover, as ν tends to infinity, $\mathbf{Y}$ converges in distribution to a multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

An important property of the random vector $\mathbf{Y}$ is that it can be written as a function of a normal random vector and a positive random variable, *i.e*,

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

where $\mathbf{Z}$ is a normal random vector, with zero-mean vector and covariance $\boldsymbol{\Sigma}$, independent of U, which is a positive random variable with a gamma distribution Gamma$(\nu/2, \nu/2)$[1].

[1] Gamma(a, b) denotes a gamma distribution with a/b mean.

The distribution of $\mathbf{Y}$ constrained to lie within the right-truncated hyperplane

$$\mathbb{A} = \{\mathbf{y} \in \mathbb{R}^p \mid \mathbf{y} \leq \mathbf{a}\}, \tag{2}$$

where $\mathbf{y} = (y_1, \ldots, y_p)^\top$ and $\mathbf{a} = (a_1, \ldots, a_p)^\top$, is a truncated Student-t distribution, denoted by $Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, with *pdf* given by $f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \dfrac{t_p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{T_p(\mathbf{a} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}\mathbb{I}_{\mathbb{A}}(\mathbf{y})$, where $\mathbb{I}_{\mathbb{A}}(\cdot)$ is the indicator function of $\mathbb{A}$.

As was mentioned at the beginning of this section, the following properties of the multivariate Student-t and truncated Student-t distributions are useful for the implementation of the EM-algorithm. We start with the marginal-conditional decomposition of a Student-t random vector. Details of the proofs are provided in Arellano-Valle & Bolfarine (1995).

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

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

(ii) the conditional cdf of $\mathbf{Y}_2 \mid \mathbf{Y}_1 = \mathbf{y}_1$ is given by

$$P(\mathbf{Y}_2 \leq \mathbf{y}_2 \mid \mathbf{Y}_1 = \mathbf{y}_1) = T_{p_2}\left(\mathbf{y}_2 \mid \boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1\right),$$

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

The following results provide the truncated moments of a Student-t random vector. The proofs of Proposition 2 and 3 are given in Matos *et al.* (2013b). The proof of Proposition 4 is given in Ho *et al.* (2012).

Proposition 2 *If $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ with $\mathbb{A}$ as in (2), then the k-th moment of $\mathbf{Y}$, $k = 0, 1, 2$, is*

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

where $c_p(\nu, r) = \left(\dfrac{\nu + p}{\nu}\right)^r \left(\dfrac{\Gamma((p + \nu)/2)\Gamma((\nu + 2r)/2)}{\Gamma(\nu/2)\Gamma((p + \nu + 2r)/2)}\right)$, $\delta = (\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$, $\mathbf{a} = (a_1, \ldots, a_p)^\top$, $\boldsymbol{\Sigma}^* = \dfrac{\nu}{\nu + 2r}\boldsymbol{\Sigma}$, $\mathbf{Y}^{(0)} = 1$, $\mathbf{Y}^{(1)} = \mathbf{Y}$, $\mathbf{Y}^{(2)} = \mathbf{Y}\mathbf{Y}^\top$, and $\nu + 2r > 0$.

Having established a formula for the k-order moments of $\mathbf{Y}$, we now present a result on conditional moments of the partition of $\mathbf{Y}$.

Proposition 3 *Let $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ with $\mathbb{A}$ as in (2). 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 partition of the parameters*

$\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{a}$ ($\mathbf{a}^{y_1}, \mathbf{a}^{y_2}$) and $\mathbb{A}$ ($\mathbb{A}^{y_1}, \mathbb{A}^{y_2}$). Then, under the notation of Proposition 1, the conditional k-th moment of $\mathbf{Y}_2$ is

$$E\left[\left(\frac{\nu+p}{\nu+\delta}\right)^r \mathbf{Y}_2^{(k)} \mid \mathbf{Y}_1\right] = \frac{d_p(p_1,\nu,r)}{(\nu+\delta_1)^r}\frac{T_{p_2}(\mathbf{a}^{y_2} \mid \boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu+p_1+2r)}{T_{p_2}(\mathbf{a}^{y_2} \mid \boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu+p_1)} E_{\mathbf{W}}\left[\mathbf{W}^{(k)}\right],$$

where $\mathbf{W} \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu+p_1+2r; \mathbb{A}^{y_2})$, $\delta = (\mathbf{Y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})$, $\delta_1 = (\mathbf{Y}_1 - \boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Y}_1-\boldsymbol{\mu}_1)$, $\mathbf{a}^{y_2} = (a_1,\ldots,a_{p_2})^\top$, $\widetilde{\boldsymbol{\Sigma}}^*_{22.1} = \left(\dfrac{\nu+\delta_1}{\nu+2r+p_1}\right)\boldsymbol{\Sigma}_{22.1}$, $\nu+p_1+2r>0$ and

$$d_p(p_1,\nu,r) = (\nu+p)^r\left(\frac{\Gamma((p+\nu)/2)\Gamma((p_1+\nu+2r)/2)}{\Gamma((p_1+\nu)/2)\Gamma((p+\nu+2r)/2)}\right).$$

In the following Proposition, we establish relationships between the expectation and covariance of $\mathbf{Y}$ and $\mathbf{W}$.

Proposition 4 *Let $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu},\Sigma,\nu;\mathbb{A}^*)$, with $\mathbb{A}^* = \{\mathbf{y} \in \mathbb{R}^p \mid \mathbf{a}^* < \mathbf{y} \leq \mathbf{b}^*\}$, where $\mathbf{a}^* = (a_1^*,\ldots,a_p^*)^\top$, $\mathbf{b}^* = (b_1^*,\ldots,b_p^*)^\top$, $\Sigma = \Lambda\mathbf{R}\Lambda$ and $\Lambda = Diag\,(\sigma_{11},\ldots,\sigma_{pp})$ is a $p\times p$ diagonal matrix with each diagonal element being positive. We have that $\mathbf{W} = \boldsymbol{\Lambda}^{-1}(\mathbf{Y}-\boldsymbol{\mu}) \sim Tt_p(\mathbf{0},\mathbf{R},\nu;\mathbb{A})$, where $\mathbf{a} = \Lambda^{-1}(\mathbf{a}^*-\boldsymbol{\mu})$ and $\mathbf{b} = \boldsymbol{\Lambda}^{-1}(\mathbf{b}^*-\boldsymbol{\mu})$. Therefore,*

$$\begin{aligned} E[\mathbf{Y}] &= \boldsymbol{\mu} + \boldsymbol{\Lambda}E[\mathbf{W}] \\ E[\mathbf{Y}\mathbf{Y}^\top] &= \boldsymbol{\mu}\boldsymbol{\mu}^\top + \Lambda E[\mathbf{W}]\boldsymbol{\mu}^\top + \mu E[\mathbf{W}^\top]\boldsymbol{\Lambda} + \boldsymbol{\Lambda}E[\mathbf{W}\mathbf{W}^\top]\boldsymbol{\Lambda}^\top, \end{aligned}$$

where $E[\mathbf{W}]$ and $E[\mathbf{W}\mathbf{W}^\top]$ are given in Ho et al. *(2012).*

2.2 The statistical model

Now we present the robust multivariate t model for censored data. Let us write

$$\mathbf{Y}_i \sim t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu), \quad i=1,\ldots,n, \tag{3}$$

where $\mathbf{Y}_i = (Y_{i1},\ldots,Y_{ip})^\top$ is a $p\times 1$ vector of responses for sample unit i, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_p)^\top$ and the dispersion matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$ depends on unknown and reduced parameter vector $\boldsymbol{\alpha}$. We assume $\mathbf{Y}_i$, $i=1,\ldots,n$, are independent and identically distributed. We consider the approach proposed by Vaida & Liu (2009) and Matos *et al.* (2013b) to model the censored responses. Thus, the observed data for the ith subject is given by $(\mathbf{V}_i,\mathbf{C}_i)$, where $\mathbf{V}_i$ represents the vector of uncensored reading or censoring level and $\mathbf{C}_i$ the vector of censoring indicators. In other words,

$$Y_{il} \leq V_{il} \quad \text{if} \quad C_{il}=1, \text{ and } Y_{il}=V_{il} \quad \text{if} \quad C_{il}=0, \tag{4}$$

$i=1,\ldots,n$, $l=1,\ldots,p$, so that, (3) along with (4) defines the Student-t censored model for multivariate responses (tMC model) . Notice that a left censoring structure causes a right truncation of the distribution, since we only know that the true observation y_{il} is less than or equal to the observed quantity V_{il}. Moreover, the right censored problem can be represented by a left censored problem by simultaneously transforming the response y_{il} and censoring level V_{il} to $-Y_{il}$ and $-V_{il}$.

2.3 The likelihood function

To obtain the likelihood function of the tMC model, first we treat separately the observed and censored components of $\mathbf{y}_i$, *i.e.*, $\mathbf{y}_i = (\mathbf{y}_i^{o\top}, \mathbf{y}_i^{c\top})^\top$, with $C_{il} = 0$ for all elements in $\mathbf{y}_i^o$, and $C_{il} = 1$ for all elements in $\mathbf{y}_i^c$. Accordingly, we write $\mathbf{V}_i = vec(\mathbf{V}_i^o, \mathbf{V}_i^c)$, where $vec(\cdot)$ denotes the function which stacks vectors or matrices of the same number of columns, with $\mathbf{\Sigma}_i = \mathbf{\Sigma}_i(\boldsymbol{\alpha}) = \begin{pmatrix} \mathbf{\Sigma}_i^{oo} \mathbf{\Sigma}_i^{oc} \\ \mathbf{\Sigma}_i^{co} \mathbf{\Sigma}_i^{cc} \end{pmatrix}$ and $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_i^{o\top}, \boldsymbol{\mu}_i^{c\top})^\top$. Then, using Proposition 1, we have that $\mathbf{y}_i^o \sim t_{p_i^o}(\boldsymbol{\mu}_i^o, \mathbf{\Sigma}_i^{oo}, \nu)$ and $\mathbf{y}_i^c \mid \mathbf{y}_i^o \sim t_{p_i^c}(\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + p_i^o)$, where

$$\boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \mathbf{\Sigma}_i^{co}\mathbf{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o), \;\; \mathbf{S}_i^{co} = \left(\frac{\nu + Q(\mathbf{y}_i^o)}{\nu + p_i^o}\right)\mathbf{\Sigma}_i^{cc.o}, \tag{5}$$

with $\mathbf{\Sigma}_i^{cc.o} = \mathbf{\Sigma}_i^{cc} - \mathbf{\Sigma}_i^{co}\mathbf{\Sigma}_i^{oo-1}\mathbf{\Sigma}_i^{oc}$ and $Q(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)^\top\mathbf{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)$. Therefore, the likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\alpha}^\top, \nu)^\top$ for subject i is given by

$$\begin{aligned} L_i(\boldsymbol{\theta} \mid \mathbf{y}) &= f(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i^c \leq \mathbf{V}_i^c \mid \mathbf{y}_i^o = \mathbf{V}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o = \mathbf{V}_i^o \mid \boldsymbol{\theta}) \\ &= T_{p_i^c}(\mathbf{V}_i^c \mid \boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + p_i^o) t_{p_i^o}(\mathbf{V}_i^o \mid \boldsymbol{\mu}_i^o, \mathbf{\Sigma}_i^{oo}, \nu) = L_i. \end{aligned} \tag{6}$$

Straightforwardly, the log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta} \mid \mathbf{y}) = \sum_{i=1}^n \log L_i$. It is important to note that this function can be computed at each step of the EM-type algorithm without additional computational burden since the L_i's have already been computed at the E-step. We assume that the degrees of freedom parameter of the Student-t distribution is fixed. For choosing the most appropriate value of this parameter, we will use the log-likelihood profile (Lange *et al.*, 1989; Meza *et al.*, 2011). This assumption is based on the work by Lucas (1997), in which the author showed that the protection against outliers is preserved only if the degrees of freedom parameter is fixed. Consequently, the parameter vector for the tMC model is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$.

2.4 Parameter estimation via the EM algorithm

We describe in detail how to carry out ML estimation for the proposed tMC model. The EM algorithm, originally proposed by Dempster *et al.* (1977), is a very popular iterative optimization strategy commonly used to obtain ML estimates for incomplete data problems. This algorithm has many attractive features such as the numerical stability and the simplicity of implementation and its memory requirements are quite reasonable (Couvreur, 1996).

In order to propose the EM algorithm for our tMC model, firstly we define $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ such that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the i-th subject. Now, we treat $\mathbf{u}$ and $\mathbf{y}$ as hypothetical missing data, and augmenting with the observed data $\mathbf{V}, \mathbf{C}$ corresponding to the censoring mechanism. Consequently, we set the complete-data vector as $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{u}^\top)^\top$. The complete data log-likelihood function is given by

$$\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c),$$

where

$$\ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c) = -\frac{1}{2}\left[n \log |\boldsymbol{\Sigma}| + u_i(\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right] + h(u_i \mid \nu) + c,$$

with c being a constant that does not depend on $\boldsymbol{\theta}$ and $h(u_i \mid \nu)$ being the Gamma$(\nu/2, \nu/2)$ *pdf*. Finally, the EM algorithm for the tMC model can be summarized through the following two steps.

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

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = E\left[\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}), \tag{7}$$

where

$$Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = Q_i(\boldsymbol{\mu}, \boldsymbol{\alpha} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}\Big([\widehat{u\mathbf{y}^2_i}^{(k)} - 2\widehat{u\mathbf{y}}_i^{(k)}\boldsymbol{\mu}^\top + \widehat{u}_i^{(k)}\boldsymbol{\mu}\boldsymbol{\mu}^\top]\boldsymbol{\Sigma}^{-1}\Big).$$

with $\widehat{u\mathbf{y}}_i^{(k)} = E[U_i\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{u\mathbf{y}^2_i}^{(k)} = E[U_i\mathbf{Y}_i\mathbf{Y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ and $\widehat{u}_i^{(k)} = E[U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$. Note that, since ν is fixed, there is no need to obtain $E\left[h(U_i \mid \nu) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\right]$.

M-step:
In this step, $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$ is conditionally maximized with respect to $\boldsymbol{\theta}$ and a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$ is obtained. Specifically, we have that

$$\widehat{\boldsymbol{\mu}}^{(k+1)} = \left(\sum_{i=1}^{n} \widehat{u}_i^{(k)}\right)^{-1} \sum_{i=1}^{n} \widehat{u\mathbf{y}}_i^{(k)}, \tag{8}$$

$$\widehat{\Sigma}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{u\mathbf{y}^2_i}^{(k)} - 2\widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\mu}}^{(k+1)} + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\mu}}^{(k+1)}\widehat{\boldsymbol{\mu}}^{(k+1)\top}\right]. \tag{9}$$

The algorithm is iterated until a suitable convergence rule is satisfied. In this case, we adopt the distance involving two successive evaluations of the log-likelihood defined in (6), that is, $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$ as a convergence criterion. It is important to stress that from equations (8)-(9), the E-step reduces to the computation of $\widehat{u\mathbf{y}^2_i}$, $\widehat{u\mathbf{y}}_i$, and $\widehat{u}_i$. These expected values can be determined in closed form, using Propositions 1-4, as follows:

1. If the subject i has only censored components, from Proposition 2

$$\begin{aligned}
\widehat{u\mathbf{y}^2_i}^{(k)} &= E\left[U_i\mathbf{Y}_i\mathbf{Y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \frac{T_p(\mathbf{V}_i \mid \widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{*(k)}, \nu+2)}{T_p(\mathbf{V}_i \mid \widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{(k)}, \nu)} E\left[\mathbf{W}_i\mathbf{W}_i^\top\right], \\
\widehat{u\mathbf{y}}_i^{(k)} &= E\left[U_i\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \frac{T_p(\mathbf{V}_i \mid \widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{*(k)}, \nu+2)}{T_p(\mathbf{V}_i \mid \widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{(k)}, \nu)} E\left[\mathbf{W}_i\right], \\
\widehat{u}_i^{(k)} &= E\left[U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \frac{T_p(\mathbf{V}_i \mid \widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{*(k)}, \nu+2)}{T_p(\mathbf{V}_i \mid \widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{(k)}, \nu)},
\end{aligned}$$

   where $\mathbf{W}_i \sim Tt_p(\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{*(k)}, \nu+2; \mathbb{A}_i)$, $\widehat{\boldsymbol{\Sigma}}^{*(k)} = \dfrac{\nu}{\nu+2}\widehat{\boldsymbol{\Sigma}}^{(k)}$, and $\mathbb{A}_i = \{\mathbf{W}_i \in \mathbb{R}^p \mid \mathbf{w}_i \leq \mathbf{V}_i\}$ where $\mathbf{w}_i = (w_{i1}, \ldots, w_{in_i})^\top$ and $\mathbf{V}_i = (V_{i1}, \ldots, V_{in_i})^\top$.

2. If the subject i has only non-censored components, then,

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

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

3. If the subject i has censored and uncensored components, then from Proposition 3 and given that $\{\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i\}$, $\{\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \mathbf{y}_i^o\}$, and $\{\mathbf{Y}_i^c \mid \mathbf{V}_i, \mathbf{C}_i, \mathbf{y}_i^o\}$ are equivalent processes, we have that

$$\begin{aligned}
\widehat{u\mathbf{y}^2_i}^{(k)} &= E\left[U_i\mathbf{Y}_i\mathbf{Y}_i^\top \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \begin{pmatrix} \mathbf{y}_i^o\mathbf{y}_i^{o\top}\widehat{u}_i^{(k)} & \widehat{u}_i^{(k)}\mathbf{y}_i^o\widehat{\mathbf{w}}_i^{c(k)\top} \\ \widehat{u}_i^{(k)}\widehat{\mathbf{w}}_i^{c(k)}\mathbf{y}_i^{o\top} & \widehat{u}_i^{(k)}\widehat{\mathbf{w^2}}_i^{c(k)} \end{pmatrix}, \\
\widehat{u\mathbf{y}}_i^{(k)} &= E\left[U_i\mathbf{Y}_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] = vec(\mathbf{y}_i^o\widehat{u}_i^{(k)}, \widehat{u}_i^{(k)}\widehat{\mathbf{w}}_i^{c(k)}), \\
\widehat{u}_i^{(k)} &= E\left[U_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \left(\frac{p_i^o+\nu}{\nu+\widehat{Q}^{(k)}(\mathbf{y}_i^o)}\right)\frac{T_{n_i}(\mathbf{V}_i \mid \widehat{\boldsymbol{\mu}}_i^{co(k)}, \widetilde{\mathbf{S}}_i^{co(k)}, \nu+p_i^o+2)}{T_{n_i}(\mathbf{V}_i \mid \widehat{\boldsymbol{\mu}}_i^{co(k)}, \widetilde{\mathbf{S}}_i^{co(k)}, \nu+p_i^o)},
\end{aligned}$$

   where $\widetilde{\mathbf{S}}_i^{co(k)} = \left(\dfrac{\nu+\widehat{Q}^{(k)}(\mathbf{y}_i^o)}{\nu+2+p_i^o}\right)\widehat{\boldsymbol{\Sigma}}_i^{cc.o(k)}$, $\widehat{\mathbf{w}}_i^{c(k)} = E\left[\mathbf{W}_i \mid \widehat{\boldsymbol{\theta}}^{(k)}\right]$ and $\widehat{\mathbf{w^2}}_i^{c(k)} = E\left[\mathbf{W}_i\mathbf{W}_i^\top \mid \widehat{\boldsymbol{\theta}}^{(k)}\right]$, with $\widehat{Q}^{(k)}(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \widehat{\boldsymbol{\mu}}_i^{o(k)})^\top \widehat{\boldsymbol{\Sigma}_i^{oo(k)}}^{-1(k)}(\mathbf{y}_i^o - \widehat{\boldsymbol{\mu}}_i^{o(k)})$, $\mathbf{W}_i \sim Tt_{p_i^c}(\widehat{\boldsymbol{\mu}}_i^{co(k)}, \widetilde{\mathbf{S}}_i^{co(k)}, \nu+p_i^o+2; \mathbb{A}_i^c)$ and $\boldsymbol{\Sigma}_i^{cc.o}$, $\boldsymbol{\mu}_i^{co}$, and $\mathbf{S}_i^{co}$ are as in (14).

As was mentioned in Proposition 4, formulas for $E[\mathbf{W}]$ and $E[\mathbf{W}\mathbf{W}^\top]$, where $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, can be found in Ho *et al.* (2012). For the computation of multivariate Student-t *cdf* we used the $pmvt$ function of the $mvtnorm$ package(Genz *et al.*, 2008) from R software.

3 The FM-tMC model

Ignoring censoring for the moment, we consider a more general and robust framework for the multivariate response variable $\mathbf{Y}_i$ of the model defined in (3), which is assumed to follows a mixture of multivariate Student-t distributions:

$$\mathbf{Y}_i \sim \sum_{j=1}^{G} \pi_j \, t_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \nu_j), \tag{10}$$

where π_j are weights adding to 1 and G is the number of groups, also called components in mixture models. The mixture regression model considered in (10) is also defined as: let Z_i be a latent class variable such that given $Z_i = j$, the response $\mathbf{y}_i$ follows a multivariate Student-t distribution

$$\mathbf{Y}_i \sim t_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \nu_j), \;\; i = 1, \ldots, n, \;\; j = 1, \ldots, G. \tag{11}$$

Now, suppose $P(Z_i = j) = \pi_j$, then the density of $\mathbf{y}_i$, without observing Z_i, is

$$f(\mathbf{y}_i \mid \boldsymbol{\theta}) = \sum_{j=1}^{G} \pi_j \, t_p(\mathbf{y}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \boldsymbol{\nu}_j), \tag{12}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_G^\top)^\top$, with $\boldsymbol{\theta}_j = (\pi_j, \boldsymbol{\mu}_j^\top \boldsymbol{\alpha}_j^\top, \nu_j)^\top$. The model (12) is the regression model based on the mixture of Student-t distributions, studied, for instance, by Peel & McLachlan (2000). Concerning the parameter ν_j, $j = 1, \ldots, G$, for computational convenience we assume that $\nu = \nu_1 = \nu_2 = \ldots, = \nu_G$. This strategy works very well in the empirical studies that we have conducted and greatly simplifies the optimization problem.

Following Karlsson & Laitila (2014), the mixture model for censored data can be formulated in a similar way to the model defined in (12) as:

$$f(\mathbf{y}_i \mid \boldsymbol{\theta}) = \sum_{j=1}^{G} \pi_j f_{ij}(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}), \tag{13}$$

where

$$f_{ij}(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}) \;\; = \;\; T_{p_i^c}(\mathbf{V}_i^c \mid \boldsymbol{\mu}_{ij}^{co}, \mathbf{S}_{ij}^{co}, \nu + p_i^o) t_{p_i^o}(\mathbf{V}_i^o \mid \boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, \nu),$$

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

with $\boldsymbol{\Sigma}_{ij}^{cc.o} = \boldsymbol{\Sigma}_{ij}^{cc} - \boldsymbol{\Sigma}_{ij}^{co} \boldsymbol{\Sigma}_{ij}^{oo-1} \boldsymbol{\Sigma}_{ij}^{oc}$ and $Q_{ij}(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \boldsymbol{\mu}_{ij}^o)^\top \boldsymbol{\Sigma}_i^{oo-1} (\mathbf{y}_i^o - \boldsymbol{\mu}_{ij}^o)$. The model defined in (13) will be called the FM-tMC model. Thus, the log-likelihood function given the observed data $\mathbf{y}$, is given by

$$\ell(\boldsymbol{\theta} \mid \mathbf{y}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta} \mid \mathbf{y}) = \sum_{i=1}^{n} \log(f(\mathbf{y}_i \mid \boldsymbol{\theta})).$$

3.1 Maximum likelihood estimation via EM algorithm

In this section, we present an EM algorithm for the ML estimation of the FM-tMC model defined in (13). To explore the EM algorithm, we present the FM-tMC model in an incomplete-data framework, using the results presented in Section 2.

In order to simplify notations, algebra and future interpretations, it is appropriate to deal with a random vector $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iG})^\top$ instead of the random variable Z_i, where

$$Z_{ij} = \begin{cases} 1, & \text{if the ith observation is from the } j\text{th component;} \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, under this approach the random vector $\mathbf{Z}$ has multinomial distribution considering a withdrawal into G categories, with probabilities $p_1, \ldots, p_G$, i.e.,

$$P(\mathbf{Z}_i = \mathbf{z}_i) = \pi_1^{z_{i1}} \pi_2^{z_{i2}} \ldots \pi_G^{z_{iG}},$$

where $\sum_{j=1}^{G} \pi_j = 1$, such that

$$\mathbf{Y}_i \mid Z_{ij} = 1 \overset{\text{ind.}}{\sim} t_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \nu).$$

For the vector $\mathbf{Z}_i$ we will use the notation $\mathbf{Z}_i \overset{\text{iid.}}{\sim} Multinomial(1, p_1, \ldots, p_g)$. Observe that $Z_{ij} = 1$ if and only if $Z_i = j$. Thus, from (1), the set-up defined above can be written hierarchically as

$$\begin{aligned} \mathbf{Y}_i \mid U_i = u_i, Z_{ij} = 1 &\overset{\text{ind.}}{\sim} N(\boldsymbol{\mu}_j, u_i^{-1}\boldsymbol{\Sigma}_j), && (14)\\ U_i \mid Z_{ij} = 1 &\overset{\text{ind.}}{\sim} Gamma(\nu/2, \nu/2), && (15)\\ \mathbf{Z}_i &\overset{\text{iid.}}{\sim} Multinomial(1, p_1, \ldots, p_g), && (16) \end{aligned}$$

for $i = 1, \ldots, n$, all independent. For censored data, let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ such that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the i-th subject and $\mathbf{z} = (\mathbf{z}_1^\top, \ldots, \mathbf{z}_n^\top)^\top$. Then, under the hierarchical representation (14)–(16), it follows that the complete log-likelihood function associated with $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{V}^\top, \mathbf{C}^\top, \mathbf{u}^\top, \mathbf{z}^\top)^\top$ is

$$\begin{aligned} \ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) &= c + \sum_{i=1}^{n}\sum_{j=1}^{G} z_{ij} \log \pi_j - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{G} z_{ij} \log\left(|\boldsymbol{\Sigma}_j|\right) \\ &\quad - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{G} z_{ij} u_i (\mathbf{y}_i - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_j) + \sum_{i=1}^{n}\sum_{j=1}^{G} z_{ij} \log h(u_i \mid \nu), \quad (17) \end{aligned}$$

where c is a constant that is independent of the parameter vector $\boldsymbol{\theta}$.
Letting $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\theta}}_1^{(k)\top}, \ldots, \widehat{\boldsymbol{\theta}}_G^{(k)\top})^\top$, with $\widehat{\boldsymbol{\theta}}_j^{(k)} = (\widehat{\pi}_j^{(k)}, \widehat{\boldsymbol{\Sigma}}_j^{(k)}, \widehat{\boldsymbol{\mu}}_j^{(k)})^\top$, $j = 1, \ldots, G$, the estimates of $\boldsymbol{\theta}$ at the kth iteration. It follows, after some simple algebra, that the conditional expectation of the

complete log-likelihood function has the form

$$
\begin{aligned}
Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) &= c + \sum_{i=1}^{n}\sum_{j=1}^{G} \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)}) \log \pi_j - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{G} \mathcal{Z}_{ij}(\boldsymbol{\theta}^{(k)}) \log\left(|\boldsymbol{\Sigma}_j|\right) \\
&- \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{G} \operatorname{tr}\left[\left(\mathcal{E}_{2ij}(\widehat{\boldsymbol{\theta}}^{(k)}) - 2\boldsymbol{\mu}_j \mathcal{E}_{1ij}^{\top}(\widehat{\boldsymbol{\theta}}^{(k)}) + \mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}}^{(k)})\boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\top}\right)\boldsymbol{\Sigma}_j^{-1}\right], \quad (18)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathrm{E}[Z_{ij}U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}], \;\; \mathcal{E}_{1ij}(\widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{E}[Z_{ij}U_i\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}], \\
\mathcal{E}_{2ij}(\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathrm{E}[Z_{ij}U_i\mathbf{Y}_i\mathbf{Y}_i^{\top} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] \;\text{ and }\; \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{E}[Z_{ij} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}].
\end{aligned}
$$

By using known properties of conditional expectation, we obtain

$$
\mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)}) = \frac{\widehat{\pi}_j^{(k)} f_{ij}(\mathbf{V}_i \mid \mathbf{C}_i, \widehat{\boldsymbol{\theta}}_j^{(k)})}{\sum_{j=1}^{G} \widehat{\pi}_j^{(k)} f_{ij}(\mathbf{V}_i \mid \mathbf{C}_i, \widehat{\boldsymbol{\theta}}_j^{(k)})}, \tag{19}
$$

$\mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}}^{(k)}) = \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)})E[U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1]$, $\mathcal{E}_{1ij}(\widehat{\boldsymbol{\theta}}^{(k)}) = \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)})E[U_i\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1]$ and $\mathcal{E}_{2ij}(\widehat{\boldsymbol{\theta}}^{(k)}) = \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)})E[U_i\mathbf{Y}_i\mathbf{Y}_i^{\top} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1]$, where the conditional expectations

$$
E[U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1],\; E[U_i\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1],\; E[U_i\mathbf{Y}_i\mathbf{Y}_i^{\top} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1], \tag{20}
$$

can be directly obtained from the expressions $\widehat{u}_i^{(k)}$, $\widehat{u\mathbf{y}}_i^{(k)}$, and $\widehat{u\mathbf{y}^2}_i^{(k)}$, respectively, given in Subsection 2.3. Thus, we have closed form expression for all the quantities involved in the E-step of the algorithm. Next, we describe the EM algorithm for maximum likelihood estimation of the parameters of the FM-tMC model.
E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, compute $\mathcal{E}_{sij}(\widehat{\boldsymbol{\theta}}^{(k)})$, $s = 0, 1, 2$ and $\mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)})$ for $i = 1, \ldots, n$, $j = 1, \ldots, G$.
M-step: Update $\widehat{\boldsymbol{\theta}}^{(k+1)}$ by maximizing $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$ over $\boldsymbol{\theta}$, which leads to the following closed form expressions:

$$
\begin{aligned}
\widehat{\pi}_j^{(k+1)} &= \frac{\sum_{i=1}^{n} \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)})}{n}, & (21) \\
\widehat{\boldsymbol{\mu}}_j^{(k+1)} &= \left(\sum_{i=1}^{n} \mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}}^{(k)})\right)^{-1} \sum_{i=1}^{n}\left(\mathcal{E}_{1ij}(\widehat{\boldsymbol{\theta}}^{(k)})\right), & (22) \\
\widehat{\boldsymbol{\Sigma}}_j^{(k+1)} &= \left(\sum_{i=1}^{n} \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)})\right)^{-1} \sum_{i=1}^{n}\left[\left(\mathcal{E}_{2ij}(\widehat{\boldsymbol{\theta}}^{(k)}) - 2\boldsymbol{\mu}_j^{(k+1)}\mathcal{E}_{1ij}^{\top}(\widehat{\boldsymbol{\theta}}^{(k)}) + \mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}}^{(k)})\boldsymbol{\mu}_j^{(k+1)}\boldsymbol{\mu}_j^{\top((k+1))}\right)\right], & (23)
\end{aligned}
$$

where $j = 1, \dots, G$.

It is well known that mixture models can provide a multimodal log-likelihood function. In this sense, the method of maximum likelihood estimation through EM algorithm may not give global solutions if the starting values are far from the real parameter values. Thus, the choice of starting values for the EM algorithm in the mixture context plays a big role in parameter estimation. In our examples and simulation studies, we consider the following procedure for the FM-tMC model:

- Partition the observation into G groups using the K-means clustering algorithm (Basso *et al.*, 2010). In this case, the censored values are considered as observed observations;
- Compute the proportion of data points belonging to the same cluster j, say $p_j^{(0)}$, $j = 1, \dots, G$. This is the initial value for p_j;
- For each group j, compute the initial values $\boldsymbol{\mu}_j^{(0)}$, $(\boldsymbol{\Sigma}_j^2)^{(0)}$ using the method of moments estimators.

3.2 Model selection

Because there is no universal criterion for mixture model selection, we chose three criteria to compare the models considered in this work, namely, the Akaike information criterion (AIC) (Akaike, 1974), the Bayesian information criterion (BIC) (Schwarz, 1978) and the efficient determination criterion (EDC) (Bai *et al.*, 1989). Like the more popular AIC and BIC criteria, EDC has the form

$$-2\ell(\widehat{\boldsymbol{\theta}}) + \rho c_n,$$

where $\ell(\boldsymbol{\theta})$ is the actual log-likelihood, ρ is the number of free parameters that has to be estimated in the model and the penalty term c_n is a convenient sequence of positive numbers. Here, we use $c_n = 0.2\sqrt{n}$, a proposal that was considered in Basso *et al.* (2010) and Cabral *et al.* (2012). We have $c_n = 2$ for AIC, $c_n = \log n$ for BIC, where n is the sample size.

3.3 Provision of standard errors

A simple way of obtaining the standard errors of ML estimates of mixture model parameters is to approximate the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$ by the inverse of the observed information matrix. Let $\mathbf{I}_o(\boldsymbol{\theta}) = -\partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ be the observed information matrix, where $\ell(\boldsymbol{\theta})$ is the observed log-likelihood function in (13). In this work we use the alternative method suggested by Basford *et al.* (1997), which consists of approximating the inverse of the covariance matrix by

$$\mathbf{I}_o(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^\top, \quad \text{where} \quad \widehat{\mathbf{s}}_i = \mathrm{E}\left[\frac{\partial \ell_{ic}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mid \mathbf{y}_{\mathrm{c}}\right]\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \tag{24}$$

where $\ell_{ic}(\boldsymbol{\theta})$ is given in (17) and

$$\widehat{\mathbf{s}}_i = (\widehat{s}_{i,\boldsymbol{\mu}_1}, \dots, \widehat{s}_{i,\boldsymbol{\mu}_G}, \widehat{s}_{i,\boldsymbol{\alpha}_1}, \dots, \widehat{s}_{i,\boldsymbol{\alpha}_G}, \widehat{s}_{i,\pi_1}, \dots, \widehat{s}_{i,\pi_{G-1}})^\top.$$

Expressions for the elements $\widehat{s}_{i,\boldsymbol{\mu}_j}, \widehat{s}_{i,\boldsymbol{\alpha}_j}, \widehat{s}_{i,\pi_j}$ are given in the following:

$$
\begin{aligned}
\widehat{s}_{i,\boldsymbol{\mu}_j} &= \widehat{\boldsymbol{\Sigma}}_j^{-1}(\mathcal{E}_{1ij}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\mu}}_j),\\
\widehat{s}_{i,\pi_j} &= \frac{\mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}})}{\widehat{\pi}_j} - \frac{\mathcal{Z}_{iG}(\widehat{\boldsymbol{\theta}})}{\widehat{\pi}_G},\\
\widehat{s}_{i,\alpha_{jr}} &= -\frac{1}{2}\mathrm{tr}\left[\mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_j^{-1}\frac{\partial \boldsymbol{\Sigma}_j}{\partial \alpha_{jr}} - \psi(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_j^{-1}\frac{\partial \boldsymbol{\Sigma}_j}{\partial \alpha_{jr}}\widehat{\boldsymbol{\Sigma}}_j^{-1}\right]
\end{aligned}
\tag{25}
$$

where $\psi(\widehat{\boldsymbol{\theta}}) = \left(\mathcal{E}_{2ij}(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mu}}_j\mathcal{E}_{1ij}^{\top}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{1ij}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\mu}}_j^{\top} + \mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\mu}}_j\widehat{\boldsymbol{\mu}}_j^{\top}\right)$ and α_{jr} denotes the rth element of $\boldsymbol{\alpha}_j$. It is important to stress that in our analysis we focus solely on comparing the SE of $\boldsymbol{\mu}_j$, $\boldsymbol{\alpha}_j$ and π_j, with $j = 1, \ldots, G$, since that ν is assumed to be known.

The information-based approximation (24) is asymptotically applicable. However, it is less reliable unless the sample size is sufficiently large. It is common practice to perform the parametric bootstrap approach (Efron & Tibshirani, 1986) to obtain more accurate standard error estimates. However, we do not employ the bootstrap approach, since it requires enormous amounts of computing power.

4 Simulation studies

In order to study the performance of our proposed method, we present three simulation studies. The first one shows the parameter recovery, that is, if we can estimate the true parameter values accurately by using the proposed EM algorithm. The second one investigates the ability of the FM-tMC model to cluster observations. Finally, the third one shows the asymptotic behavior of the EM estimates for the proposed model.

Parameter recovery

In this section, we consider one scenario for simulation in order to verify if we can estimate the true parameter values accurately by using the proposed EM algorithm. This is the first step to ensure that the estimation procedure works satisfactorily. We fit data that were artificially generated from the model (13) and several censoring proportion settings $(5\%, 10\%, 30\%)$. We generated 500 Monte Carlo samples of size $n = 100, 400, 1000$. We consider small and different variances with the following parameter setup:

$$
0.65\ t_2\left(\begin{bmatrix}-5\\-4\end{bmatrix}, \begin{bmatrix}3 & 1\\1 & 4.5\end{bmatrix}, 4\right) + 0.35\ t_2\left(\begin{bmatrix}2\\3\end{bmatrix}, \begin{bmatrix}2 & 1\\1 & 3.5\end{bmatrix}, 4\right).
$$

The average values (Mean) and standard deviations (Std) of the estimates across the 500 Monte Carlo samples were computed. Also were computed the average (IM Std) values of the approximate standard errors of the estimates obtained through the method described in Subsection 3.3 and the percentage of coverage of the resulting 95% confidence intervals (COV) assuming asymptotic normality.

$n=100$												
		Parameter										
Censored	Measure	μ_{11}	μ_{12}	$\sigma_{1,11}$	$\sigma_{1,12}$	$\sigma_{1,22}$	π_1	μ_{21}	μ_{22}	$\sigma_{2,11}$	$\sigma_{2,12}$	$\sigma_{2,22}$
	True	(-5)	(-4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	-5.00	-3.99	2.88	0.89	4.45	0.65	1.99	2.98	1.99	1.11	3.66
5%	Std.	0.50	0.52	0.78	0.65	1.14	0.05	0.55	0.62	0.77	0.77	1.43
	IM Std	0.26	0.32	0.78	0.63	1.16	0.08	0.30	0.41	0.78	0.74	1.38
	COV	93%	95%	91%	92%	92%	100%	93%	95%	100%	96%	100%
	True	(-5)	(-4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	-5.03	-3.93	2.81	0.92	4.60	0.66	2.06	3.05	1.81	0.90	3.43
10%	Std.	0.38	0.44	0.80	0.63	1.21	0.05	0.44	0.52	0.75	0.70	1.50
	IM Std	0.25	0.32	0.77	0.62	1.20	0.08	0.30	0.40	0.73	0.68	1.35
	COV	94%	93%	88%	92%	92%	100%	92%	92%	100%	95%	100%
	True	(-5)	(-4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	-5.01	-3.92	2.76	0.92	4.86	0.70	2.24	3.28	1.74	0.65	2.91
30%	Std.	0.42	0.48	1.02	0.69	1.72	0.05	0.49	0.53	1.06	0.73	1.43
	IM Std	0.25	0.33	0.90	0.66	1.39	0.09	0.32	0.40	0.76	0.63	1.22
	COV	92%	95%	85%	92%	89%	100%	80%	84%	100%	92%	100%
$n=400$												
		Parameter										
Censored	Measure	μ_{11}	μ_{12}	$\sigma_{1,11}$	$\sigma_{1,12}$	$\sigma_{1,22}$	π_1	μ_{21}	μ_{22}	$\sigma_{2,11}$	$\sigma_{2,12}$	$\sigma_{2,22}$
	True	(-5)	(-4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	-5.02	-4.00	2.85	0.90	4.41	0.65	2.02	3.02	1.94	0.98	3.45
5%	Std.	0.13	0.16	0.37	0.31	0.52	0.03	0.15	0.20	0.39	0.36	0.63
	IM Std	0.13	0.16	0.38	0.30	0.56	0.04	0.15	0.20	0.36	0.34	0.63
	COV	94%	93%	91%	91%	95%	100%	92%	94%	99%	95%	99%
	True	(-5)	(-4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	-5.03	-3.97	2.77	0.89	4.55	0.66	2.10	3.07	1.74	0.87	3.25
10%	Std.	0.13	0.16	0.38	0.31	0.59	0.02	0.15	0.20	0.36	0.31	0.61
	IM Std	0.12	0.16	0.38	0.30	0.58	0.04	0.14	0.19	0.33	0.31	0.60
	COV	94%	94%	85%	90%	96%	100%	87%	92%	100%	94%	100%
	True	(-5)	(-4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	-5.01	-3.92	2.74	0.90	4.85	0.71	2.30	3.35	1.53	0.48	2.72
30%	Std.	0.13	0.18	0.45	0.32	0.87	0.03	0.17	0.21	0.39	0.27	0.58
	IM Std	0.12	0.17	0.44	0.32	0.68	0.04	0.15	0.20	0.32	0.27	0.55
	COV	93%	90%	85%	93%	88%	88%	86%	89%	100%	90%	100%
$n=1000$												
		Parameter										
Censored	Measure	μ_{11}	μ_{12}	$\sigma_{1,11}$	$\sigma_{1,12}$	$\sigma_{1,22}$	π_1	μ_{21}	μ_{22}	$\sigma_{2,11}$	$\sigma_{2,12}$	$\sigma_{2,22}$
	True	(-5)	(-4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	-5.02	-4.00	2.85	0.88	4.40	0.65	2.02	3.02	1.91	0.97	3.41
5%	Std.	0.08	0.10	0.23	0.19	0.34	0.02	0.09	0.13	0.22	0.21	0.38
	IM Std	0.08	0.10	0.24	0.19	0.35	0.03	0.09	0.12	0.22	0.21	0.39
	COV	94%	95%	88%	90%	93%	100%	94%	94%	99%	93%	100%
	True	(-5)	(-4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	-5.03	-3.97	2.75	0.87	4.55	0.66	2.10	3.07	1.75	0.86	3.26
10%	Std.	0.08	0.10	0.23	0.20	0.38	0.02	0.09	0.13	0.23	0.20	0.40
	IM Std	0.08	0.10	0.24	0.19	0.37	0.03	0.09	0.12	0.21	0.19	0.38
	COV	93%	93%	80%	86%	93%	100%	79%	90%	100%	91%	99%
	True	(-5)	(-4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	-5.00	-3.94	2.72	0.90	4.80	0.71	2.30	3.35	1.58	0.50	2.67
30%	Std.	0.08	0.11	0.29	0.20	0.53	0.02	0.11	0.13	0.24	0.17	0.37
	IM Std	0.08	0.10	0.27	0.20	0.43	0.03	0.10	0.12	0.21	0.17	0.33
	COV	94%	92%	77%	90%	86%	86%	87%	93%	100%	88%	100%

Table 1: Simulated data: Parameter recovery. Mean, standard deviations (Std) for EM estimates and percentage of coverage (COV) based on 500 samples from the FM-tMC model. IM Std indicates the average of the approximate standard errors of the estimates obtained through the method described in Subsection 3.3.

The results are presented in Table 1. The estimates of the parameters are close to the true values of the parameters and become closer as the sample size increases. Moreover, the estimates are less sensitive to the variation of the censoring level. In general, the results suggest that the proposed FM-tMC model produced satisfactory estimates, as expected. We also see this from Table 1 that the estimation method of the standard errors provides relatively close results (Std and IM Std), indicating that the proposed asymptotic approximation for the variances of the ML estimates is reliable. This can also be seen in the coverage parameters (COV), since in general a confidence interval above 90% coverage is maintained for each parameter. In a similar way, we analyze other scenario: large but equal variances, and the results are presented in Table 6 given in the Appendix.

Clustering

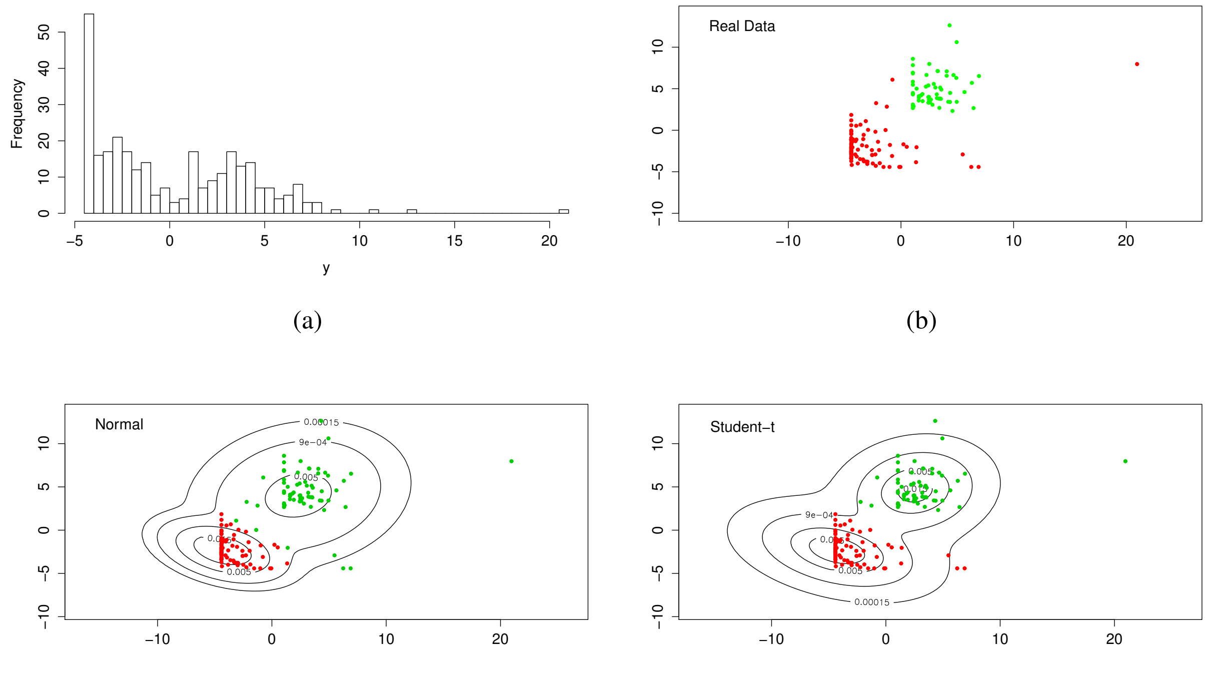
In this section, we illustrate the ability of the FM-tMC model to fit data with a mixture structure generated from a different family of distributions, such as the skew-normal independent (SNI) family of distributions (Cabral *et al.*, 2012), and we also investigate the ability of the FM-tMC model to cluster observations, that is, to allocate them into groups of observations that are similar in some sense. We know that each data point belongs to one of G components in a heterogeneous population, but we do not know how to discriminate between them. Modeling by mixture models allows clustering of the data in terms of the estimated (posterior) probability that a single point belongs to a given group.

We generated 300 Monte Carlo samples with 15% of censoring under the following scenarios: (I) scenario 1 (Figure 1): a mixture of two skew-t models (Azzalini & Genton, 2008), and (II) scenario 2 (Figure 2): a mixture of two skew-slash (Wang & Genton, 2006) distributions. The parameter values were chosen to present a considerable proportion of outliers and skewness pattern. It can be seen from Figures 1 and 2 that the groups are poorly separated. Furthermore, note that although we have a two components mixture, the histogram may not to be clearly bimodal.

For each sample of size $n = 150$, we proceed with clustering ignoring the known true classification. Following the method proposed by Liu & Lin (2014), to assess the quality of the classification function of each mixture model, an index measure was used in the current study, called correct classification rate **CCR**, which is based on the posterior probability assigned to each subject. The FM-tMC were fitted using the algorithm described in the Section 3.1 in order to obtain the estimate of the posterior probability that an observation $\mathbf{Y}_i$ belongs to the jth component of the mixture, i.e. $\mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)})$. For sample $l,\ l = 1, ..., 300$, we compute the number of correct allocations $(CCRs)$ divided by the sample size n, that is, $ACCR = \frac{1}{300}\sum_{l=1}^{300} CCR_l$.

Fitted model	ACCR	
	Scenario I	Scenario II
FM-nMC	0.87	0.74
FM-tMC	0.96	0.81

Table 2: Simulated data from a mixture of two skew-t (Scenario I) and two skew-slash (Scenario II) models $(n = 150)$: Clustering. Monte Carlo mean of right allocation rates for fitted FM-tMC and FM-nMC models



(a) (b)

(c) (d)

Figure 1: Simulated data from a mixture of two skew-t models $(n = 150)$: Clustering. **(a)** Histogram for the simulated sample - scenario I, **(b)** Scatter plot for one simulated sample along with the original group (green and red colors) and the the respective density contours: **(c)** FM-nMC fit and **(d)** FM-tMC fit.

Table 2 shows the mean value of the correct allocation rates $ACCR$, where larger values indicate better classification results. Comparing with the results for the FM-nMC model, we can see that modeling using the FM-tMC model represents an improvement in the outright clustering and has a better performance, showing the robustness of the FM-tMC model to discrepant observations as well as to censored distributions which seem to occur quite often in practice.

Asymptotic properties

In this simulation study, we analyze the absolute bias (Bias) and mean square error (MSE) of the estimates obtained from the FM-tMC model through the proposed EM algorithm. These measures are defined by

$$Bias(\boldsymbol{\theta}_i) = \frac{1}{M}\sum_{j=1}^{M} |\widehat{\boldsymbol{\theta}}_i^{(j)} - \boldsymbol{\theta}_i| \quad \text{and} \quad MSE(\boldsymbol{\theta}_i) = \frac{1}{M}\sum_{j=1}^{M} (\widehat{\boldsymbol{\theta}}_i^{(j)} - \boldsymbol{\theta}_i)^2, \tag{26}$$

where $\widehat{\boldsymbol{\theta}}_i^{(j)}$ is the ML estimate of the parameter $\boldsymbol{\theta}_i$ for the jth sample. Six different sample sizes $(n = 100, 200, 300, 400, 600, 1000)$ are considered.

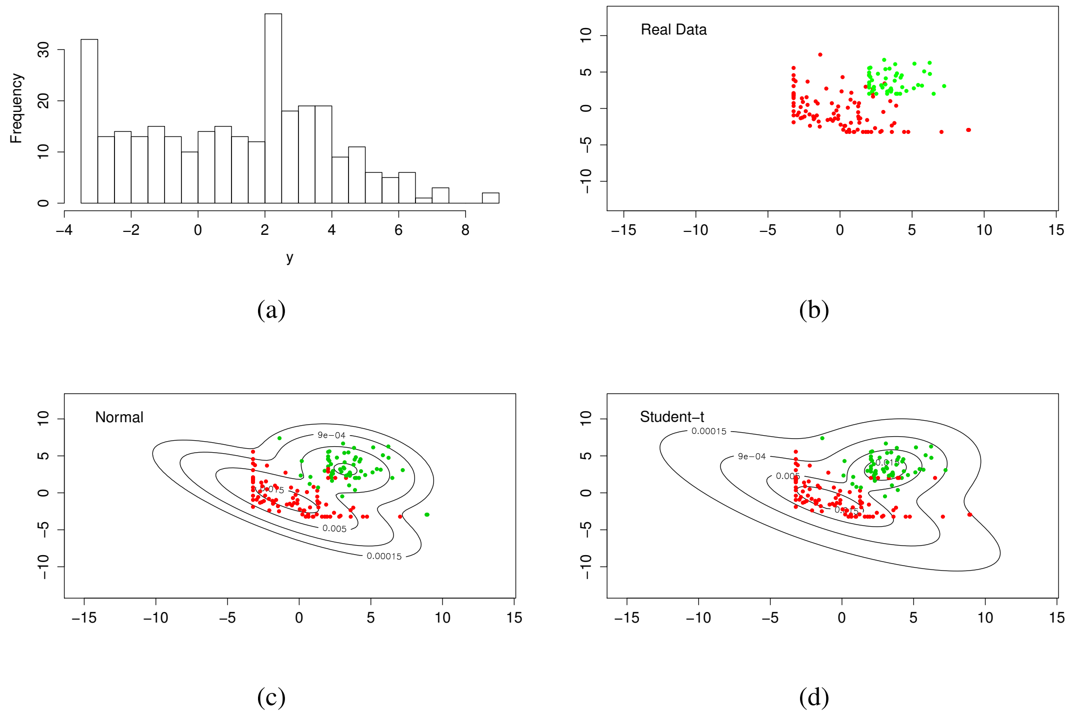
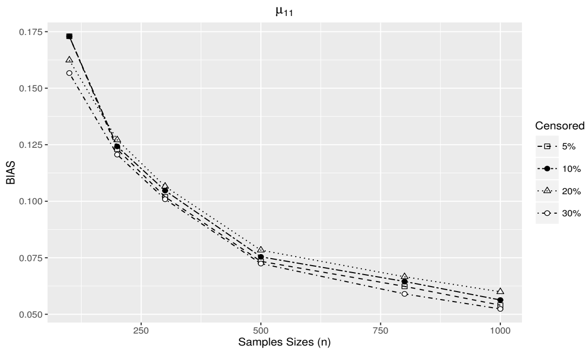


Figure 2: Simulated data from a mixture of two skew-slash models ($n = 150$): Clustering. **(a)** Histogram for the simulated sample - scenario I, **(b)** Scatter plot for one simulated sample along with the original group (green and red colors) and the the respective density contours: **(c)** FM-nMC fit and **(d)** FM-tMC fit.

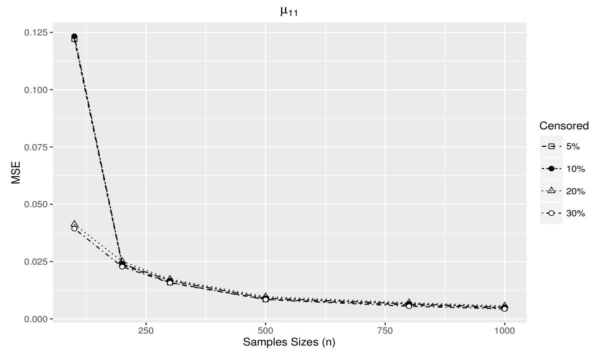
For each sample size, we generate 500 Monte Carlo samples with $5\%, 10\%, 20\%, 30\%$ of censoring proportion. Using the EM algorithm, the absolute bias and mean squared error for each parameter over the 500 datasets were computed. The parameter setup is as follows:

$$0.35\ t_2\left(\begin{bmatrix}2\\3\end{bmatrix},\begin{bmatrix}2 & 1\\1 & 3.5\end{bmatrix},4\right)+0.65\ t_2\left(\begin{bmatrix}-5\\-4\end{bmatrix},\begin{bmatrix}2 & 1\\1 & 3.5\end{bmatrix},4\right),$$

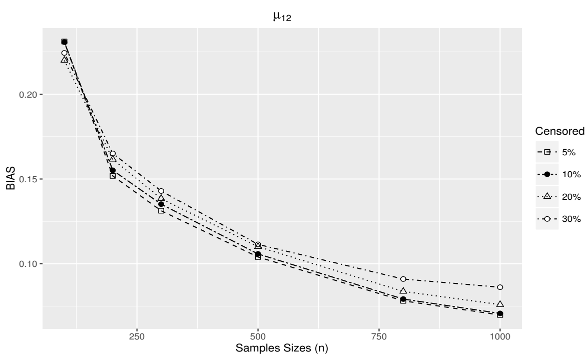
The results for the estimates of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\pi}$ are given in Figures 3, 4 and 5, respectively. We can see a pattern of convergence to zero of the (Bias) and MSE when n increases, independent of the censoring pattern. As a general rule, we can say that Bias and MSE approach to zero when the sample size increases, indicating that the estimates based on the proposed EM-type algorithm under the FM-tMC model do admit desirable asymptotic properties.



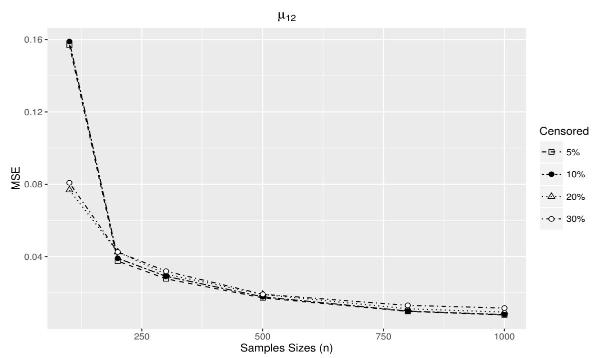
(a)



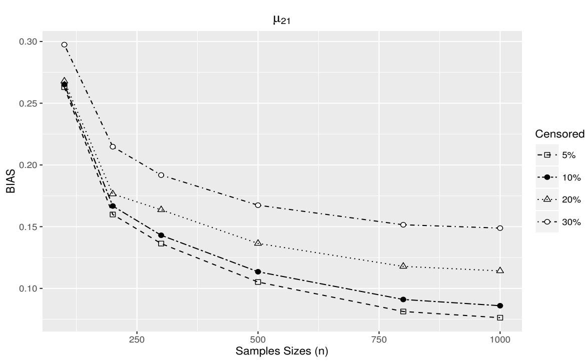
(b)



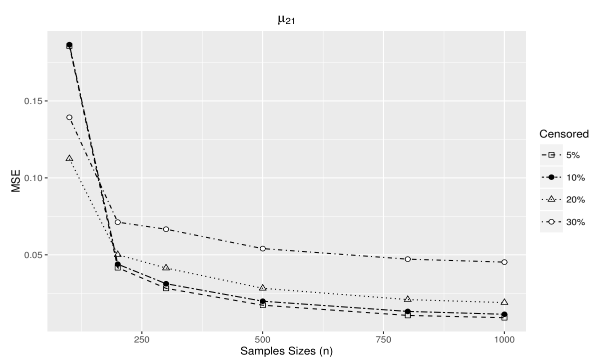
(c)



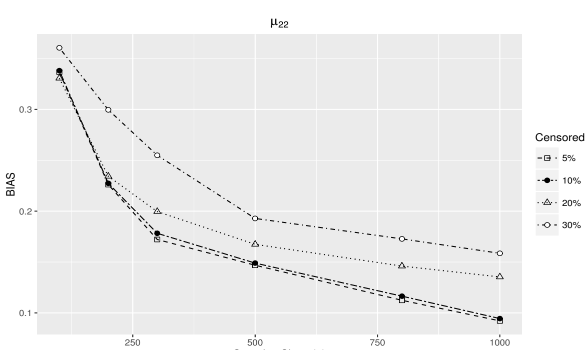
(d)



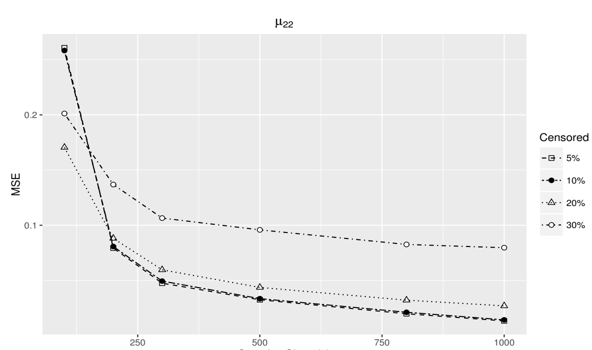
(e)



(f)

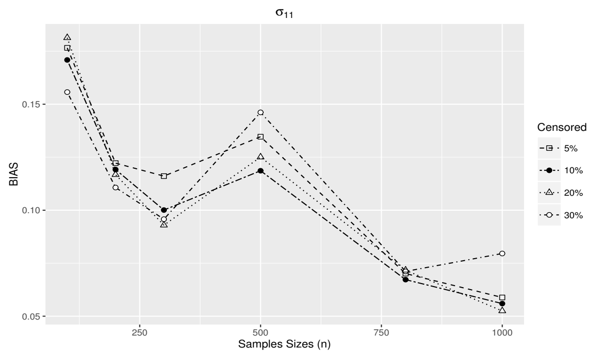


(g)

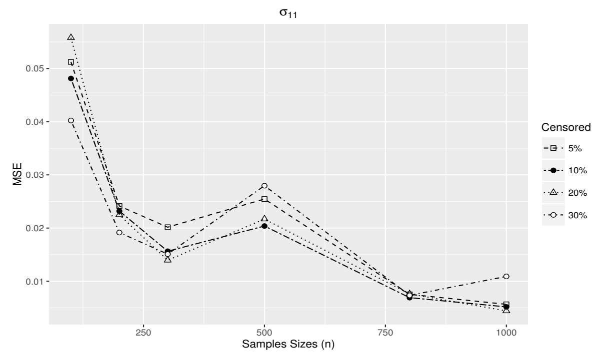


(h)

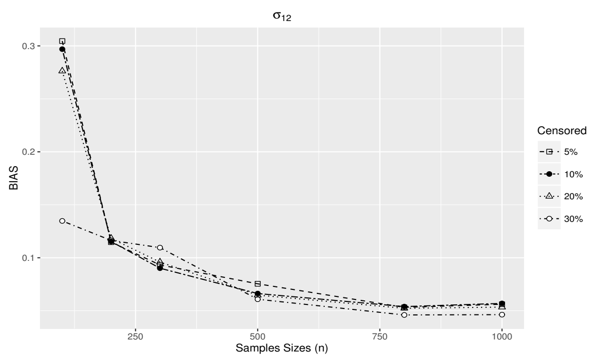
Figure 3: Simulated data: Asymptotic properties. Bias (first column) and MSE (second column) of **(a, b)** for μ_{11}, **(c, d)** for μ_{12}, **(e, f)** for μ_{21} and **(g, h)** for μ_{22} estimate under FM-tMC model with different levels of censoring $(5\%, 10\%, 20\%, 30\%)$



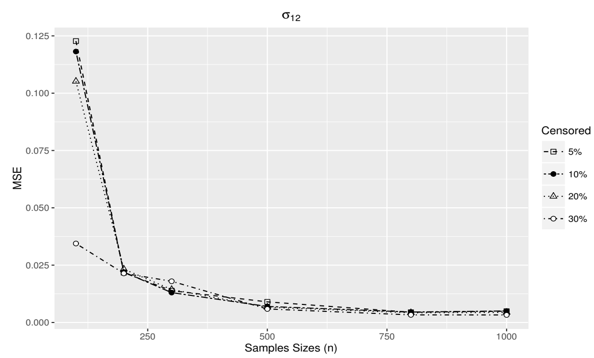
(a)



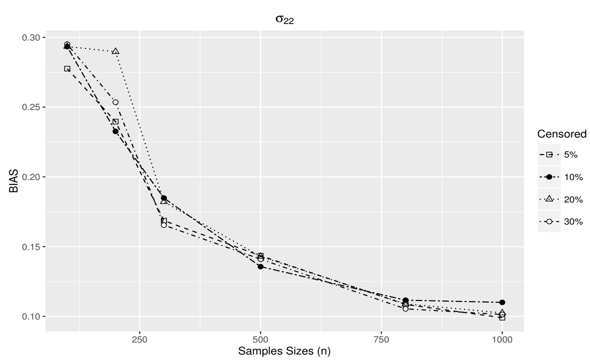
(b)



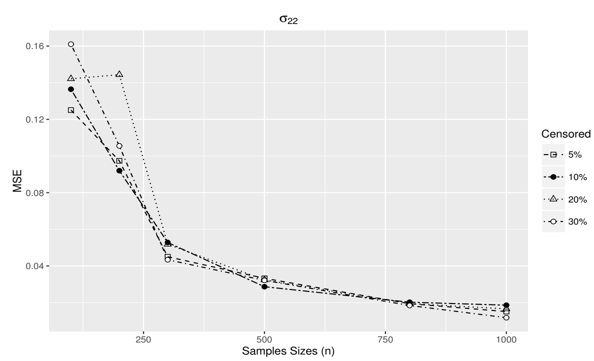
(c)



(d)

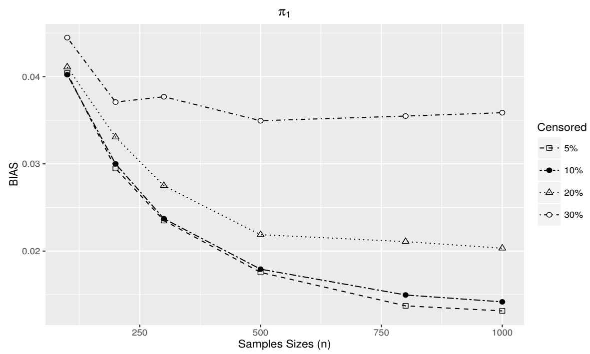


(e)

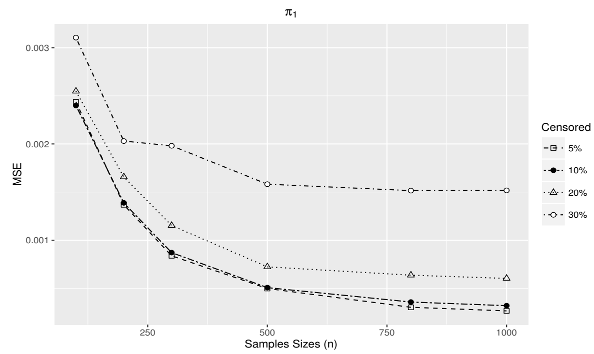


(f)

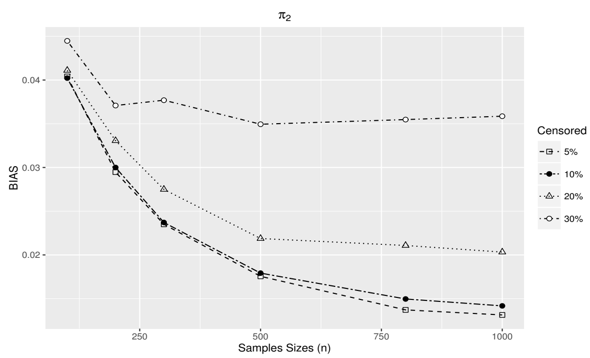
Figure 4: Simulated data: Asymptotic properties. Bias (first column) and MSE (second column) of **(a, b)** for σ_{11}, **(c, d)** for σ_{12} and **(e, f)** for σ_{22} estimate under FM-tMC model with different levels of censoring $(5\%, 10\%, 20\%, 30\%)$



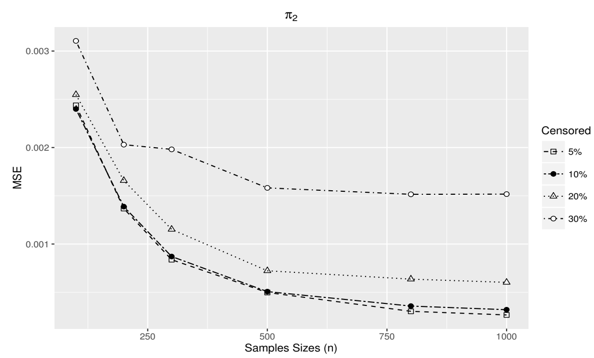
(a)



(b)



(c)



(d)

Figure 5: Simulated data: Asymptotic properties. Bias (first column) and MSE (second column) of **(a, b)** for π_1 and **(c, d)** for π_2 estimate under FM-tMC model with different levels of censoring $(5\%, 10\%, 20\%, 30\%)$

5 Application

We consider a dataset consisting of concentration levels of certain dissolved trace metals in freshwater streams across the Commonwealth of Virginia. The Virginia Department of Environment Quality (VDEQ) provided the data used in this application, and these data were previously analyzed by Hoffman & Johnson (2015), where they 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, 2003). 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. Figure 6 shows the histogram of the concentration levels of each trace metal and all together. We

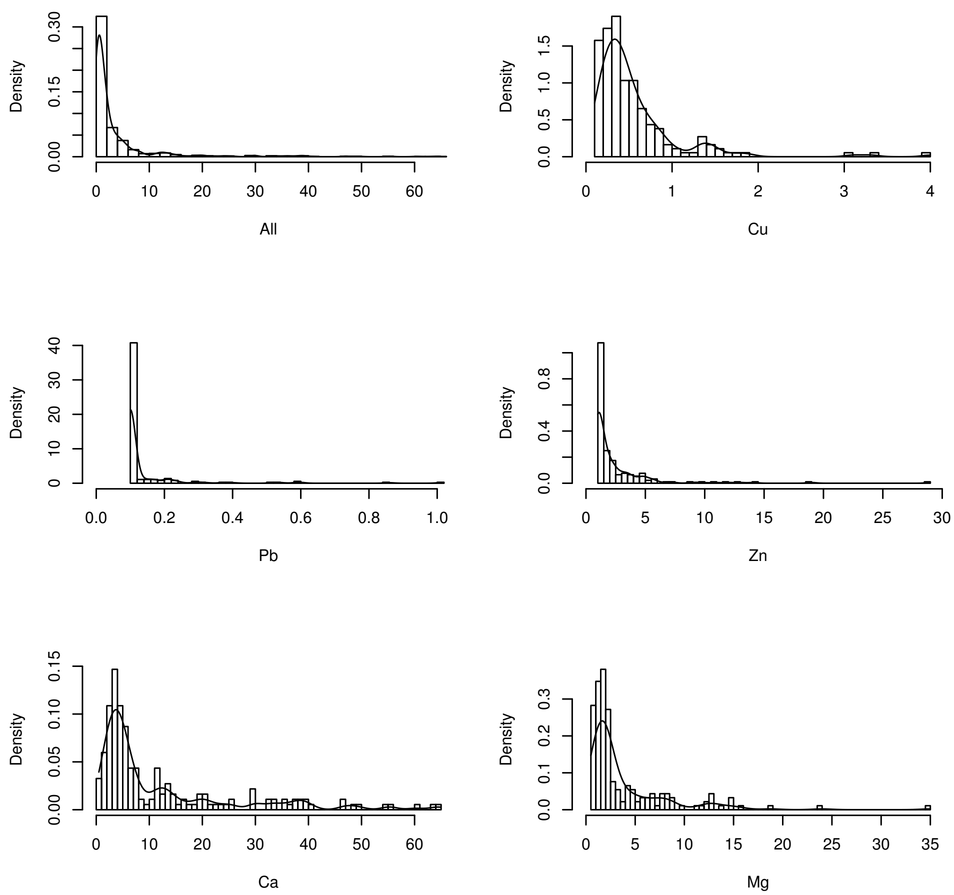


Figure 6: VDEQ data. Histogram of the dissolved trace metals.

can see that most of the distributions associated with the individual metals have heavy tails, two or more modes and skews to right. Because of these empirical evidences, we propose to fit a FM-tMC model. The number of groups of the model is chosen according to the information criteria (see Subsection 3.2) as shown in Table 3. It can be seen that the model with two components and 3 degrees of freedom fits the data best. This finding can be also appreciated from Figure 7 where the log-likelihood values are depicted for a grid of values of ν. Notice also that the estimated value of ν is fairly small, indicating a lack of adequacy of the normal assumption for the VDEQ data. We considered the variance-covariance to be equal in order to reduce the number of parameters to be estimated (parsimonious model).

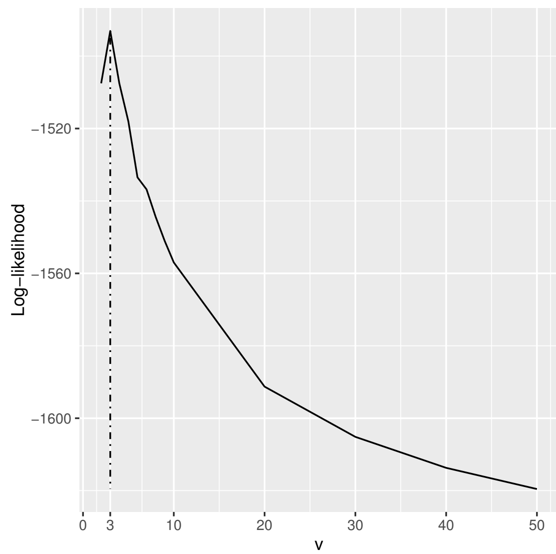


Figure 7: VDEQ data. Plot of the profile log-likelihood of the degrees of freedom ν

Thus, we arrive at the following model for the VDEQ data: $f(\mathbf{y}_i \mid \mathbf{\Theta}) = \sum_{j=1}^{2} \pi_j t_5(\mathbf{y}_i \mid \boldsymbol{\mu}_j, \mathbf{\Sigma}, 3)$, where

$$\boldsymbol{\mu}_j = (\mu_{j1}, \mu_{j2}, \mu_{j3}, \mu_{j4}, \mu_{j5})^\top,\ j = 1, 2, \quad \text{and} \quad \mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ & & & \sigma_{44} & \sigma_{45} \\ & & & & \sigma_{55} \end{bmatrix}$$

Criteria	$\nu = 3$		$\nu = 4$	
	$G = 2$	$G = 3$	$G = 2$	$G = 3$
Log-likelihood	**-1493.04**	-1543.89	-1507.51	-1547.42
AIC	**3038.08**	3151.77	3067.02	3158.84
BIC	**3121.67**	3254.65	3150.61	3261.72
EDC	**3056.62**	3174.59	3085.56	3181.66

Table 3: VDEQ data. Selection criteria for various FM-tMC models.

The ML estimates of the parameters were obtained using the EM algorithm described in Section 3. The results of the EM algorithm are shown in Table 4. This table shows that the estimates (Est) of μ_1 and μ_2 for the FM-nMC and FM-tMC models are close. However, the standard errors (SE) of μ_1 and μ_2 are smaller than those under the normal counterpart, indicating that the FM-tMC model seems to produce more precise estimates.

Similarly, we have that the ML estimates and standard errors (SE) for the variance components under the FM-tMC model (Σ_t) are lower than those under the FM-nMC model (Σ_N). Moreover, we can see that the SE under the FM-tMC model (SE_t) are less than those under the FM-nMC

Parameter	FM-nMC		FM-tMC	
	Est	SE	Est	SE
$\boldsymbol{\mu}_{11}$	0.54	0.07	0.42	0.02
$\boldsymbol{\mu}_{12}$	-0.03	0.03	0.04	0.01
$\boldsymbol{\mu}_{13}$	1.49	0.48	1.20	0.15
$\boldsymbol{\mu}_{14}$	6.65	0.85	4.84	0.43
$\boldsymbol{\mu}_{15}$	2.33	0.47	1.96	0.16
$\boldsymbol{\mu}_{21}$	0.57	0.29	0.43	0.24
$\boldsymbol{\mu}_{22}$	-0.47	2.17	-0.26	0.51
$\boldsymbol{\mu}_{23}$	-0.02	1.91	-0.22	0.89
$\boldsymbol{\mu}_{24}$	39.91	1.17	34.18	1.45
$\boldsymbol{\mu}_{25}$	10.33	0.52	6.89	0.56
π_1	0.84	0.07	0.86	0.08

Table 4: VDEQ data. Estimation (Est) and standard errors (SE) for parameters under the FM-nMC and FM-tMC models.

model (SE_N), indicating that the FM-tMC model produces more precise estimates.

$$\boldsymbol{\Sigma}_N = \begin{bmatrix} 0.25 & 0.04 & 0.06 & 0.54 & 0.45 \\ & 15.79 & 0.96 & -0.40 & 1.38 \\ & & 46.04 & 0.30 & -0.40 \\ & & & 1.68 & 16.09 \\ & & & & 13.21 \end{bmatrix}, \quad \boldsymbol{\Sigma}_t = \begin{bmatrix} 0.04 & 0.01 & 0.01 & 0.09 & 0.04 \\ & 1.58 & 0.17 & -0.10 & 0.04 \\ & & 10.28 & 0.07 & -0.04 \\ & & & 0.08 & 3.46 \\ & & & & 1.46 \end{bmatrix}$$

$$SE_N = \begin{bmatrix} 0.03 & 0.02 & 0.31 & 0.31 & 0.19 \\ & 0.01 & 0.06 & 0.21 & 0.13 \\ & & 1.08 & 3.20 & 1.31 \\ & & & 2.71 & 1.24 \\ & & & & 0.60 \end{bmatrix}, \quad SE_t = \begin{bmatrix} 0.01 & 0.00 & 0.02 & 0.06 & 0.02 \\ & 0.00 & 0.01 & 0.03 & 0.01 \\ & & 0.24 & 0.43 & 0.16 \\ & & & 1.31 & 0.46 \\ & & & & 0.18 \end{bmatrix}$$

Concentration levels. Variance-Covariance estimates under the FM-nMC model ($\widehat{\boldsymbol{\Sigma}}_N$), variance estimates under the FM-tMC model ($\widehat{\boldsymbol{\Sigma}}_t$), SE obtained under the FM-nMC model (SE_N) and SE obtained under the FM-tMC model (SE_t).

Table 5 compares the fit of the two mixture models using the model selection criteria discussed in Subsection 3.2. Note that, as expected, the FM-tMC model performs significantly better than the FM-nMC model.

Criteria	FM-nMC	FM-tMC
Log-likelihood	-1650.72	**-1493.04**
AIC	3353.43	**3038.08**
BIC	3437.02	**3121.67**
EDC	3371.97	**3056.62**

Table 5: VDEQ data. Model selection criteria. Values in bold correspond to the best model.

6 Conclusions

In this paper, a novel approach to analyze correlated censored data has been developed based on the use of finite mixtures of multivariate Student-t distributions. This approach generalizes several previously proposed solutions, such as, the finite mixture of Gaussian components (Caudill, 2012; Karlsson & Laitila, 2014; He, 2013). A simple and efficient EM-type algorithm was developed, which has closed-form expressions at the E-step and relies on formulas for the mean and variance of the multivariate truncated Student-t distributions (Ho *et al.*, 2012). The proposed EM algorithm was implemented as part of the R package `CensMixReg` and is available for download at the CRAN repository. The experimental results and the analysis of a real dataset provide support for the usefulness and effectiveness of our proposal.

Recently, Garay *et al.* (2015) considered the problem of censored linear regression models using scale mixtures of normal distributions (SMN). Therefore, it would be a worthwhile task to investigate the applicability of a likelihood-based treatment in the context of finite mixtures of SMN distributions (FM-SMNC model). It may also be interesting to consider mixture of linear mixed-effects models with censored observations (Bai *et al.*, 2016). Other extensions of the current work include, for example, a generalization of the FM-tMC model to the multivariate skew-t distribution (Lachos *et al.*, 2010; Cabral *et al.*, 2012).

Acknowledgment

The research of V. H. Lachos is partially supported by CNPq-Brazil (Grant 306334/2015-1) and FAPESP-Brazil (Grant 2014/02938-9). Chen's work is partially supported by the U.S. National Science Foundation (DMS-1613295) and the U.S. National Institutes of Health (U01HL114494). E. Lopéz acknowledges support from CAPES-Brazil.

References

Akaike, H. (1974). A new look at the statistical model identification. *Autom Control IEEE Trans*, **19**, 716–723.

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

Arellano-Valle, R. B., Castro, L., González-Farías, G. & Muños Gajardo, K. (2012). Student-t censored regression model: properties and inference. *Statistical Methods and Applications*, **21**(4), 453–473.

Azzalini, A. & Genton, M. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, **76**, 1490–1507.

Bai, X., Chen, K. & Yao, W. (2016). Mixture of linear mixed models using multivariate t distribution.

Bai, Z., Krishnaiah, P. & Zhao, L. (1989). On rates of convergence of efficient detection criteria in signal processing with white noise. *Inform Theory IEEE Trans*, **35**, 380–388.

Basford, K., Greenway, D., McLachlan, G. & Peel, D. (1997). Standard errors of fitted component means of normal mixtures. *Computational Statistics*, **12**, 1–18.

Basso, R. M., Lachos, V. H., Cabral, C. R. B. & Ghosh, P. (2010). Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics & Data Analysis*, **54**(12), 2926–2941.

Cabral, C. R. B., Lachos, V. H. & Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics & Data Analysis*, **56**, 126–142.

Castro, L. M., Costa, D. R., Prates, M. O. & Lachos, V. H. (2015). Likelihood-based inference for Tobit confirmatory factor analysis using the multivariate Student-t distribution. *Statistics and Computing*, **25**(6), 1163–1183.

Caudill, S. B. (2012). A partially adaptive estimator for the censored regression model based on a mixture of normal distributions. *Statistical Methods & Applications*, **21**, 121–137.

Chib, S. (1992). Bayes inference in the Tobit censored regression model. *Journal of Econometrics*, **51**, 79–99.

Couvreur, C. (1996). The EM algorithm: a guided tour. In *Proceedings of the 2d IIEE European Workshop on Computationaly Intensive Methos in Control and Signal Processing*.

De Oliveira, V. (2005). Bayesian inference and prediction of Gaussian random fields based on censored data. *Journal of Computational and Graphical Statistics*, **14**(1), 95–115.

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.

Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, pages 54–75.

Garay, A., Castro, L., Leskow, J. & Lachos, V. H. (2014). Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Statistical Methods in Medical Research*. doi: 10.1177/0962280214551191.

Garay, A. M., Lachos, V. H., Bolfarine, H. & Cabral, C. R. (2015). Linear censored regression models with scale mixtures of normal distributions. *Statistical Papers*. doi: 10.1007/s00362-015-0696-9.

Genz, A., Bretz, F., Hothorn, T., Miwa, T., Mi, X., Leisch, F. & Scheipl, F. (2008). mvtnorm: Multivariate Normal and t Distribution. *R package version 0.9-2, URL http://CRAN.R-project.org/package=mvtnorm.*

He, J. (2013). Mixture model based multivariate statistical analysis of multiply censored environmental data. *Advances in Water Resources*, **59**, 15–24.

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. & Johnson, R. (2015). Pseudo-likelihood estimation of multivariate normal parameters in the presence of left-censored data. *Journal of Agricultural, Biological, and Environmental Statistics*, **20**, 156–171.

Karlsson, M. & Laitila, T. (2014). Finite mixture modeling of censored regression models. *Statistical Papers*, **55**(3), 627–642.

Lachos, V. H., Ghosh, P. & Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, **20**, 303–322.

Lange, K. L., Little, R. J. A. & Taylor, J. M. G. (1989). Robust statistical modeling using t distribution. *Journal of the American Statistical Association*, **84**, 881–896.

Liu, M. & Lin, T.-I. (2014). A skew-normal mixture regression model. *Educational and Psychological Measurement*, **74**, 139–162.

Lucas, A. (1997). Robustness of the student t based M-estimator. *Communications in Statistics-Theory and Methods*, **26**(5), 1165–1182.

Massuia, M. B., Cabral, C. R. B., Matos, L. A. & Lachos, V. H. (2015). Influence diagnostics for Student-t censored linear regression models. *Statistics*, **49**(5), 1074–1094.

Matos, L. A., Lachos, V. H., Balakrishnan, N. & Labra, F. V. (2013a). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistics & Data Analysis*, **57**.

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

Meza, C., Osorio, F. & De la Cruz, R. (2011). Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statistics and Computing*, **22**, 1–19.

Militino, A. F. & Ugarte, M. D. (1999). Analyzing censored spatial data. *Mathematical Geology*, **31**(5), 551–561.

Peel, D. & McLachlan, G. J. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing*, **10**(4), 339–348.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, **25**, 303 – 325.

Powell, J. L. (1986). Symmetrically trimmed least squares estimation for Tobit models. *Econometrica*, **54**, 1435–1460.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

Vaida, F. & Liu, L. (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics*, **18**(4), 797–817.

VDEQ (2003). The quality of virginia non-tidal streams: First year report. *VDEQ Technical Bulletin WQA/2002-2001, Office of Water Quality and Assessments, Virginia Department of Environmental Quality*, pages 13–16. `http://www.deq.virginia.gov/Portals/0/DEQ/Water/WaterQualityMonitoring/ProbabilisticMonitoring/report1.pdf`.

Wang, J. & Genton, M. G. (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference*, **136**(1), 209–220.

Wang, W.-L., Lin, T.-I. & Lachos, V. H. (2016). Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. *Statistical Methods in Medical Research*. doi: 10.1177/0962280215620229.

7 Appendix

A more parsimonious model is achieved by considering $\mathbf{\Sigma}_1 = ... = \mathbf{\Sigma}_G = \mathbf{\Sigma}$, which can be seen as a extension of the FM-tMC model with restricted variance-covariance components (Cabral *et al.*, 2012). We consider large variances and equal with the following parameter setup:

$$0.65\ t_2\left(\begin{bmatrix}-5\\-4\end{bmatrix},\begin{bmatrix}5 & 1\\1 & 6.5\end{bmatrix},4\right)+0.35\ t_2\left(\begin{bmatrix}2\\3\end{bmatrix},\begin{bmatrix}5 & 1\\1 & 6.5\end{bmatrix},4\right)$$

The results are presented in the Table 6. As well as on scenario small variances and different, we have that the results suggest that the proposed FM-tMC model produced satisfactory estimates.

$n = 100$

Censored	Measure	Parameter							
		μ_{11}	μ_{12}	μ_{21}	μ_{22}	σ_{11}	σ_{12}	σ_{22}	π_1
5%	True	(-5)	(-4)	(2)	(3)	(5)	(1)	(6.5)	(0.65)
	Mean	-4.98	-3.95	1.94	2.93	2.26	0.36	3.05	0.64
	Std.	0.73	0.76	0.90	0.96	0.58	0.34	0.74	0.06
	IM Std	0.21	0.25	0.28	0.33	0.35	0.24	0.48	0.09
	COV	77%	76%	68%	71%	2%	31%	2%	100%
10%	True	(-5)	(-4)	(2)	(3)	(5)	(1)	(6.5)	(0.65)
	Mean	-5.00	-3.99	1.94	2.94	2.23	0.33	3.02	0.65
	Std.	0.76	0.78	0.86	0.95	0.55	0.37	0.78	0.05
	IM Std	0.21	0.25	0.29	0.33	0.36	0.24	0.48	0.09
	COV	72%	76%	73%	71%	1%	30%	2%	100%
30%	True	(-5)	(-4)	(2)	(3)	(5)	(1)	(6.5)	(0.65)
	Mean	-5.04	-4.05	2.21	3.20	1.77	0.22	2.80	0.67
	Std.	0.35	0.43	0.74	1.10	0.87	0.57	1.09	0.06
	IM Std	0.19	0.24	0.27	0.34	0.32	0.22	0.49	0.09
	COV	74%	73%	63%	68%	1%	13%	4%	99%

$n = 400$

Censored	Measure	Parameter							
		μ_{11}	μ_{12}	μ_{21}	μ_{22}	σ_{11}	σ_{12}	σ_{22}	π_1
5%	True	(-5)	(-4)	(2)	(3)	(5)	(1)	(6.5)	(0.65)
	Mean	-5.04	-4.03	1.99	2.97	2.32	0.35	3.08	0.64
	Std.	0.17	0.19	0.26	0.28	0.29	0.18	0.35	0.03
	IM Std	0.11	0.12	0.14	0.16	0.18	0.12	0.24	0.04
	COV	76%	79%	72%	72%	1%	2%	1%	99%
10%	True	(-5)	(-4)	(2)	(3)	(5)	(1)	(6.5)	(0.65)
	Mean	-5.07	-4.02	2.03	2.99	2.21	0.31	3.08	0.65
	Std.	0.18	0.20	0.25	0.30	0.27	0.18	0.39	0.03
	IM Std	0.11	0.12	0.14	0.16	0.18	0.12	0.25	0.04
	COV	73%	76%	71%	70%	1%	1%	1%	100%
30%	True	(-5)	(-4)	(2)	(3)	(5)	(1)	(6.5)	(0.65)
	Mean	-5.07	-4.08	2.20	3.25	1.76	0.18	2.78	0.67
	Std.	0.17	0.22	0.44	1.21	0.61	0.41	0.69	0.03
	IM Std	0.09	0.12	0.97	1.15	0.16	0.10	0.24	0.04
	COV	73%	68%	50%	60%	1%	1%	1%	99%

$n = 1000$

Censored	Measure	Parameter							
		μ_{11}	μ_{12}	μ_{21}	μ_{22}	σ_{11}	σ_{12}	σ_{22}	π_1
5%	True	(-5)	(-4)	(2)	(3)	(5)	(1)	(6.5)	(0.65)
	Mean	-5.05	-4.03	1.99	2.97	2.34	0.35	3.11	0.64
	Std.	0.11	0.12	0.16	0.18	0.17	0.10	0.24	0.02
	IM Std	0.07	0.08	0.09	0.10	0.12	0.08	0.15	0.03
	COV	73%	79%	71%	74%	1%	1%	1%	100%
10%	True	(-5)	(-4)	(2)	(3)	(5)	(1)	(6.5)	(0.65)
	Mean	-5.08	-4.02	2.04	2.99	2.20	0.31	3.08	0.65
	Std.	0.11	0.12	0.15	0.17	0.17	0.11	0.22	0.02
	IM Std	0.07	0.08	0.09	0.10	0.11	0.07	0.16	0.03
	COV	68%	78%	69%	76%	1%	1%	1%	100%
30%	True	(-5)	(-4)	(2)	(3)	(5)	(1)	(6.5)	(0.65)
	Mean	-5.08	-4.08	2.20	3.19	1.72	0.18	2.76	0.67
	Std.	0.11	0.13	0.17	0.18	0.16	0.09	0.25	0.02
	IM Std	0.06	0.07	0.08	0.10	0.10	0.06	0.15	0.03
	COV	58%	63%	37%	50%	1%	1%	1%	99%

Table 6: Simulated data: Parameter recovery in the scenario of large but equal variances. Mean, standard deviations (Std) for EM estimates and percentage of coverage (COV) based on 500 samples from the FM-tMC model. IM Std indicates the average of the approximate standard errors of the estimates obtained through the method described in Subsection 3.3.