Journal of Multivariate Analysis 159 (2017) 151–167

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

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

Víctor H. Lachos [a,*], Edgar J. López Moreno [a], Kun Chen [b], Celso Rômulo Barbosa Cabral [c]

[a] *Departamento de Estatística, Universidade Estadual de Campinas, Brazil*
[b] *Department of Statistics, University of Connecticut, USA*
[c] *Departamento de Estatística, Universidade Federal de Amazonas, Brazil*

ARTICLE INFO

Article history:
Received 11 August 2016
Available online 25 May 2017

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

ABSTRACT

Finite mixture models have been widely used for the modeling and analysis of data from a heterogeneous population. Moreover, data of this kind can be subject 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 datasets are reported to demonstrate the effectiveness of the proposed methodology. The proposed algorithm and methods are implemented in the new R package `CensMixReg`.

1. Introduction

The occurrence of censored data due to limit of detection (LOD) is a common problem in many fields, e.g., 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, 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 [39]. In econometrics, the study of the labor force participation of married women is usually conducted under the censored Tobit model [12]. 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 [3].

The proportion of censored data in these studies may be substantial, 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

* Correspondence to: Departamento de Estatística, IMECC, Universidade Estadual de Campinas, CEP 13083-859, Campinas, São Paulo, Brazil
E-mail address: hlachos@ime.unicamp.br (V.H. Lachos).

http://dx.doi.org/10.1016/j.jmva.2017.05.005
0047-259X/

severe bias in statistical estimation. In the past few decades, several alternative approaches have been developed to handle censored data. Vaida and Liu [39] proposed an exact Expectation–Maximization (EM) algorithm for maximum likelihood (ML) estimation in mixed effects models for censored data, which uses closed-form expressions at the E-step. Further, Matos et al. [29] developed diagnostic measures for assessing local influence in these models. Militino and Ugarte [34] developed an EM algorithm for conducting ML estimation in censored spatial data. De Oliveira [14] 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 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. [3] (see also [28]) proposed the Student-t censored regression model. Garay et al. [16] (see also [30]) advocated the use of the multivariate Student-t distribution in the context of censored regression models, where a simple and efficient EM-type algorithm for iteratively computing ML estimates of the parameters was also presented. Castro et al. [10] proposed a likelihood-based estimation for a multivariate Tobit confirmatory factor analysis model using the Student-t distribution. More recently, Wang et al. [44] proposed a multivariate extension of the models of Garay et al. [16] and Matos et al. [30], 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 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 mixtures of censored models, Karlsson and Laitila [22] proposed an EM algorithm to estimate the parameters, and compared their method with those proposed in [11,36,37]. In a multivariate setting, He [19] proposed a Gaussian mixture model to approximate flexibly 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 undoubtedly very flexible, but the problems related to the simultaneous occurrence of skewness, anomaly observations and multimodality remain. Even when modeling using normal mixtures, overestimation of the number of components (i.e., 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.

In this article, we propose a robust mixture model for censored data based on the multivariate Student-t distribution so that the FM-tMC model is defined and a fully likelihood-based approach is carried out, including the implementation of an exact EM-type algorithm for the ML estimation. Like Matos et al. [30], we show that the E-step reduces to computing the first two moments of a truncated multivariate Student-t distribution. The likelihood function is easily computed as a byproduct of the E-step and is used for monitoring convergence and for model selection. The methodology addressed in this paper is implemented in the `R` package `CensMixReg`.

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. [16] and the related ML estimation. In Section 3, we introduce 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 with 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

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

where Γ is the standard gamma function and $\delta(\mathbf{y}) = (\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ is the squared Mahalanobis distance. The notation adopted for the vector with Student-t distribution is $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. The cumulative distribution function (cdf) is denoted by $T_p(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. Moreover, as $\nu \to \infty$, $\mathbf{Y}$ 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 scale mixture of a normal random vector and a positive random variable, i.e.,

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

where $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, i.e., a normal distribution with zero-mean vector and covariance matrix $\boldsymbol{\Sigma}$, independent of U, which is a random variable with gamma distribution $\mathcal{G}(\nu/2, \nu/2)$, where $\mathcal{G}(a, b)$ denotes a gamma distribution with a/b mean.

Let $\mathbb{A}$ be a Borel set in $\mathbb{R}^p$. We say that the random vector $\mathbf{Y}$ has a truncated Student-t distribution on $\mathbb{A}$ when $\mathbf{Y}$ has the same distribution as $\mathbf{Y}|(\mathbf{Y} \in \mathbb{A})$. In this case, the pdf of $\mathbf{Y}$ is given by

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{t_p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{P(\mathbf{Y} \in \mathbb{A})} \mathbf{1}_{\mathbb{A}}(\mathbf{y}),$$

where $\mathbf{1}_{\mathbb{A}}$ is the indicator function of $\mathbb{A}$. We use the notation $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$. If $\mathbb{A}$ has the form

$$\mathbb{A} = \{(x_1, \ldots, x_p) \in \mathbb{R}^p : x_1 \leq a_1, \ldots, x_p \leq a_p\}, \tag{2}$$

then we use the notation $(\mathbf{Y} \in \mathbb{A}) = (\mathbf{Y} \leq \mathbf{a})$, where $\mathbf{a} = (a_1, \ldots, a_p)^\top$. Analogously we define $(\mathbf{Y} \geq \mathbf{a})$. Then we say that the distribution of $\mathbf{Y}$ is truncated from above and truncated from below, respectively.

The following properties of the multivariate Student-t and truncated Student-t distributions are useful for the implementation of the EM-algorithm; see Sections 2.4 and 3.1. The proof of Proposition 4 is given in Ho et al. [20]. We start with the marginal-conditional decomposition of a Student-t random vector.

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 $dim(\mathbf{Y}_1) = p_1$, $dim(\mathbf{Y}_2) = p_2$, $p_1 + p_2 = p$. Let*

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

be 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*

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

where

$$\widetilde{\boldsymbol{\Sigma}}_{22.1} = \left(\frac{\nu + \delta_1}{\nu + p_1}\right) \boldsymbol{\Sigma}_{22.1}, \qquad \delta_1 = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{y}_1 - \boldsymbol{\mu}_1), \qquad \boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12},$$

and $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{y}_1 - \boldsymbol{\mu}_1)$.

A proof of this proposition was given by Arellano-Valle and Bolfarine [2].

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

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

where

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

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

Observe that Proposition 2 depends on formulas for $\mathrm{E}(\mathbf{W})$ and $\mathrm{E}(\mathbf{W}\mathbf{W}^\top)$, where $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$. Closed form expressions for these expectations were obtained recently by Ho et al. [20]; they depend on the cdf of the multivariate Student-t distribution. The computation uses existing functions for the cumulative t-distribution, for which the `pmvt` function of the R library `mvtnorm` [18] can be used.

Having established a formula on 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 partitions of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{a}$ ($\mathbf{a}^{y_1}$, $\mathbf{a}^{y_2}$) and $\mathbb{A}$ ($\mathbb{A}^{y_1}$, $\mathbb{A}^{y_2}$). Then, under the notation of Proposition 1,*

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

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

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

Proofs of Propositions 2 and 3 were given by Matos et al. [30]. The following proposition establishes a relationships between the expectation and covariance of $\mathbf{Y}$ and $\mathbf{W}$. The proof of this result is given in Ho et al. [20].

Proposition 4. *Let* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}^*)$, *with* $\mathbb{A}^* = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{a}^* < \mathbf{y} \leq \mathbf{b}^*\}$, *where* $\mathbf{a}^* = (a_1^*, \ldots, a_p^*)^\top$ *and* $\mathbf{b}^* = (b_1^*, \ldots, b_p^*)^\top$. *Suppose that* $\sigma_{ii} > 0$ *for all* $i \in \{1, \ldots, p\}$ *and let* $\boldsymbol{\Lambda} = \text{diag}(\sigma_{11}, \ldots, \sigma_{pp})$. *If* $\mathbf{R} = \boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}$, *we have* $\mathbf{W} = \boldsymbol{\Lambda}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \sim Tt_p(\mathbf{0}, \mathbf{R}, \nu; \mathbb{A})$, *where* $\mathbb{A} = \{\mathbf{w} \in \mathbb{R}^p : \mathbf{a} < \mathbf{w} \leq \mathbf{b}\}$, $\mathbf{a} = \boldsymbol{\Lambda}^{-1}(\mathbf{a}^* - \boldsymbol{\mu})$ *and* $\mathbf{b} = \boldsymbol{\Lambda}^{-1}(\mathbf{b}^* - \boldsymbol{\mu})$. *Therefore,*

$$\text{E}(\mathbf{Y}) = \boldsymbol{\mu} + \boldsymbol{\Lambda}\text{E}(\mathbf{W}), \qquad \text{E}(\mathbf{Y}\mathbf{Y}^\top) = \boldsymbol{\mu}\boldsymbol{\mu}^\top + \boldsymbol{\Lambda}\text{E}(\mathbf{W})\boldsymbol{\mu}^\top + \boldsymbol{\mu}\text{E}(\mathbf{W}^\top)\boldsymbol{\Lambda} + \boldsymbol{\Lambda}\text{E}(\mathbf{W}\mathbf{W}^\top)\boldsymbol{\Lambda}^\top,$$

where $\text{E}(\mathbf{W})$ *and* $\text{E}(\mathbf{W}\mathbf{W}^\top)$ *are given in* [20].

2.2. The statistical model

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

$$\mathbf{Y}_1, \ldots, \mathbf{Y}_n \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu), \tag{3}$$

where for each $i \in \{1, \ldots, n\}$, $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ 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 an unknown and reduced parameter vector $\boldsymbol{\alpha}$. We assume that $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent and identically distributed. Using the representation (1), we have that the distribution of $\mathbf{Y}_i$ can be written hierarchically as

$$\mathbf{Y}_i \mid U_i = u_i \overset{\text{ind.}}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, u_i^{-1}\boldsymbol{\Sigma}), \qquad U_i \overset{\text{ind.}}{\sim} \mathcal{G}(\nu/2, \nu/2), \tag{4}$$

where $\overset{\text{ind.}}{\sim}$ denotes independence of random variables.

We consider the approach proposed by Matos at al. [30] (see also [39]) to model the censored responses. Thus, the observed data for the ith subject is given by $(\mathbf{V}_i, \mathbf{C}_i)$, where $\mathbf{V}_i = (V_{i1}, \ldots, V_{ip})^\top$ represents the vector of uncensored readings or censoring levels and $\mathbf{C}_i = (C_{i1}, \ldots, C_{ip})^\top$ is the vector of censoring indicators. In other words,

$$Y_{ik} \leq V_{ik} \quad \text{if } C_{ik} = 1 \quad \text{and} \quad Y_{ik} = V_{ik} \quad \text{if } C_{ik} = 0, \tag{5}$$

for all $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, p\}$, i.e., $C_{ik} = 1$ if Y_{ik} is left censored. Thus, (3) along with (5) defines the Student-t censored model for multivariate responses (hereafter, the tMC model). Notice that a left censoring structure causes truncation from below of the distribution, since we only know that the true observation Y_{ik} is less than or equal to the observed quantity V_{ik}. Moreover, the right censored problem can be represented by a left censored problem by simultaneously transforming the response Y_{ik} and censoring level V_{ik} to $-Y_{ik}$ and $-V_{ik}$.

2.3. The likelihood function

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, where $\mathbf{y}_i$ is a realization of $\mathbf{Y}_i \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. 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 $C_{ik} = 0$ 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 $\mathbf{V}_i = \text{vec}(\mathbf{V}_i^o, \mathbf{V}_i^c)$, where $\text{vec}(\cdot)$ denotes the function which stacks vectors or matrices of the same number of columns, with

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

Then, using Proposition 1, we have that $\mathbf{Y}_i^o \sim t_{p_i^o}(\boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \nu)$ 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

$$\boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o), \qquad \mathbf{S}_i^{co} = \left\{\frac{\nu + \delta(\mathbf{y}_i^o)}{\nu + p_i^o}\right\}\boldsymbol{\Sigma}_i^{cc.o}, \tag{6}$$

$$\boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}\boldsymbol{\Sigma}_i^{oc} \quad \text{and} \quad \delta(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)^\top(\boldsymbol{\Sigma}_i^{oo})^{-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o). \tag{7}$$

Therefore, the likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\alpha}^\top, \nu)^\top$ for subject i is given by

$$\begin{aligned} L_i(\boldsymbol{\theta} \mid \mathbf{V}_i, \mathbf{C}_i) &= f(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i^c \leq \mathbf{V}_i^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o \mid \boldsymbol{\theta}) \\ &= T_{p_i^c}(\mathbf{V}_i^c \mid \boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + p_i^o) t_{p_i^o}(\mathbf{y}_i^o \mid \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \nu) \equiv L_i. \end{aligned} \tag{8}$$

Obviously, the log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{C}) = \ln L_1 + \cdots + \ln L_n$, where $\mathbf{V} = \text{vec}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = \text{vec}(\mathbf{C}_1, \ldots, \mathbf{C}_n)$. It is important to note that this function can be computed at each step of the EM-type algorithm (which will be derived in Section 2.4) 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 ν is fixed. For choosing the most appropriate value of this parameter, we will use the log-likelihood profile [24,33]. This assumption is based on the work of Lucas [27], 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 tMC model. The EM algorithm, originally proposed by Dempster et al. [13], 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 [31].

From (4), the complete data log-likelihood function is given by

$$\ell_c(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_{ic}(\boldsymbol{\theta}),$$

where

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

where c is a constant that does not depend on $\boldsymbol{\theta}$ and $h(u_i \mid \nu)$ is 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 estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$ at the kth step of the algorithm, the E-step provides the conditional expectation of the complete data log-likelihood function

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{E}\left\{\ell_c(\boldsymbol{\theta}) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\right\} = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}),$$

where

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

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

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

$$\widehat{\boldsymbol{\mu}}^{(k+1)} = \left(\sum_{i=1}^{n}\widehat{u}_i^{(k)}\right)^{-1}\sum_{i=1}^{n}\widehat{u\mathbf{y}}_i^{(k)}, \tag{9}$$

$$\widehat{\boldsymbol{\Sigma}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\mu}}^{(k+1)\top} - \widehat{\boldsymbol{\mu}}^{(k+1)}(\widehat{u\mathbf{y}}_i^{(k)})^\top + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\mu}}^{(k+1)}\widehat{\boldsymbol{\mu}}^{(k+1)\top}\right\}. \tag{10}$$

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 (8), i.e., $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)} \mid \mathbf{y})/\ell(\widehat{\boldsymbol{\theta}}^{(k)} \mid \mathbf{y}) - 1|$ as a convergence criterion.

It is important to stress that, from Eqs. (9) and (10), the E-step reduces to the computation of $\widehat{u\mathbf{y}_i^2}$, $\widehat{u\mathbf{y}}_i$, and $\widehat{u}_i$. To compute these expected values, first observe that they can be written in terms of $\mathrm{E}(U_i \mid \mathbf{Y}_i)$, where $\mathbf{Y}_i \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$—see the definition of U_i in (4).

For example, we have that $\widehat{u}_i = \mathrm{E}\{\mathrm{E}(U_i \mid \mathbf{Y}_i) \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\}$. It is straightforward to prove that $\mathrm{E}(U_i \mid \mathbf{Y}_i) = (\nu + p)/(\nu + \delta)$, where $\delta = (\mathbf{Y}_i - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \boldsymbol{\mu})$. Then, we can use Propositions 2 and 3 to obtain closed form expressions as follows:

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

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

   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 subject i has only censored components, from Proposition 2 (with $r = 1$),

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

   where

$$\widehat{\varphi}^{(k)}(\mathbf{V}_i) = \frac{T_p(\mathbf{V}_i \mid \widehat{\boldsymbol{\mu}}^{(k)}, \quad \widehat{\boldsymbol{\Sigma}}^{*(k)}, \nu + 2)}{T_p(\mathbf{V}_i \mid \widehat{\boldsymbol{\mu}}^{(k)}, \quad \widehat{\boldsymbol{\Sigma}}^{(k)}, \nu)}, \qquad \widehat{\mathbf{w}}_i^{c(k)} = \mathrm{E}(\mathbf{W}_i \mid \widehat{\boldsymbol{\theta}}^{(k)}), \qquad \widehat{\mathbf{w}}_i^{2^{c(k)}} = \mathrm{E}(\mathbf{W}_i\mathbf{W}_i^\top \mid \widehat{\boldsymbol{\theta}}^{(k)}),$$

$$\mathbf{W}_i \sim Tt_p(\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{*(k)}, \nu+2; \mathbb{A}_i), \qquad \widehat{\boldsymbol{\Sigma}}^{*(k)} = \frac{\nu}{\nu+2}\widehat{\boldsymbol{\Sigma}}^{(k)},$$

and $\mathbb{A}_i = \{\mathbf{w}_i \in \mathbb{R}^p : \mathbf{w}_i \leq \mathbf{V}_i\}$. To compute $\mathrm{E}(\mathbf{W}_i)$ and $\mathrm{E}(\mathbf{W}_i\mathbf{W}_i^\top)$ we use Proposition 4.

3. If the subject i has censored and uncensored components, then from Proposition 3 with $r = 1$ and $k = 0$, 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

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

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

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

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 (7), $\widehat{\mathbf{w}}_i^{c(k)}$ and $\widehat{\mathbf{w}}_i^{2c(k)}$ are defined as before, $\mathbf{W}_i \sim Tt_{p_i^c}(\widehat{\boldsymbol{\mu}}_i^{co(k)}, \widetilde{\mathbf{S}}_i^{co(k)}, \nu+p_i^o+2; \mathbb{A}_i^c)$ and $\mathbb{A}_i^c$ is defined as in (2), with the vector with censoring levels for the ith subject replacing $\mathbf{a}$. Again, to compute $\mathrm{E}(\mathbf{W}_i)$ and $\mathrm{E}(\mathbf{W}_i\mathbf{W}_i^\top)$ we use Proposition 4.

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 follow a mixture of multivariate Student-t distributions:

$$\mathbf{Y}_i \sim \sum_{j=1}^{G} \pi_j\, t_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \nu_j), \tag{11}$$

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 (11) is also defined as: let Z_{ij} be a latent class variable such that

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

Thus, given $Z_{ij} = 1$, the response $\mathbf{Y}_i$ follows a multivariate Student-t distribution

$$\mathbf{Y}_i \sim t_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \nu_j), \quad i \in \{1, \ldots, n\},\ j \in \{1, \ldots, G\}. \tag{12}$$

Now, suppose $\Pr(Z_{ij} = 1) = \pi_j$, then the density of $\mathbf{Y}_i$, without observing Z_{ij}, is

$$f(\mathbf{y}_i \mid \boldsymbol{\theta}) = \sum_{j=1}^{G} \pi_j\, t_p(\mathbf{y}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \boldsymbol{\nu}_j), \tag{13}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_G^\top)^\top$, with $\boldsymbol{\theta}_j = (\pi_j, \boldsymbol{\mu}_j^\top, \boldsymbol{\Sigma}_j, \nu_j)^\top$. The model (13) is the regression model based on the mixture of Student-t distributions studied, for instance, by Peel and McLachlan [35]. Concerning the parameter $\nu_1, \ldots, \nu_G$, for computational convenience we assume that $\nu = \nu_1 = \cdots = \nu_G$. This strategy works very well in the empirical studies that we have conducted and greatly simplifies the optimization problem.

Consider the partitions

$$\boldsymbol{\Sigma}_j = \begin{pmatrix} \boldsymbol{\Sigma}_{ij}^{oo} & \boldsymbol{\Sigma}_{ij}^{oc} \\ \boldsymbol{\Sigma}_{ij}^{co} & \boldsymbol{\Sigma}_{ij}^{cc} \end{pmatrix}, \qquad \boldsymbol{\mu}_j = (\boldsymbol{\mu}_{ij}^{o\top}, \boldsymbol{\mu}_{ij}^{c\top})^\top.$$

Following Karlsson and Laitila [22], we define the mixture model for censored data as a mixture of the tMC models given in (8), viz.

$$f(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}) = \sum_{j=1}^{G} \pi_j f_{ij}(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}), \tag{14}$$

with

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

where, for each component j, $\boldsymbol{\mu}_{ij}^{co}$ and $\mathbf{S}_{ij}^{co}$ are defined like (6) and (7), respectively. The model defined in (14) will be called the FM-tMC model. Thus, the log-likelihood function given the observed data $(\mathbf{V}, \mathbf{C})$ is given by

$$\ell(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{C}) = \sum_{i=1}^{n} \ln\{f(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta})\}.$$

3.1. Maximum likelihood estimation via the EM algorithm

In this section, we present an EM algorithm for the ML estimation of the FM-tMC model. To do so, we present the FM-tMC model in an incomplete-data framework, using the results presented in Section 2.2. We recall that the likelihood associated to finite mixtures of Student-t distributions may be unbounded, as shown by [9]. Using a straightforward extension of their argument, it can be shown that the likelihood may be unbounded in the FM-tMC case as well. Despite this, following [32, p. 41], we shall henceforth refer to the solution provided by the EM algorithm as the ML estimate even in situations where it may not globally maximize the likelihood.

Using the representation of the Student-t distribution as a scale mixture given in (4), it follows that the complete data log-likelihood function is $\ell_c(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_{ic}(\boldsymbol{\theta})$, where, for each $i \in \{1, \ldots, n\}$,

$$\ell_{ic}(\boldsymbol{\theta}) = c + \sum_{j=1}^{G} z_{ij} \ln \pi_j - \frac{1}{2} \sum_{j=1}^{G} z_{ij} \ln(|\boldsymbol{\Sigma}_j|) - \frac{1}{2} \sum_{j=1}^{G} z_{ij} u_i (\mathbf{y}_i - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_j) + \sum_{j=1}^{G} z_{ij} \ln h(u_i \mid \nu), \tag{15}$$

and c is a constant which is independent of the parameter vector $\boldsymbol{\theta}$.

For each $j \in \{1, \ldots, G\}$, let $\widehat{\boldsymbol{\theta}}_j^{(k)} = (\widehat{\pi}_j^{(k)}, \widehat{\boldsymbol{\Sigma}}_j^{(k)}, \widehat{\boldsymbol{\mu}}_j^{(k)})^\top$, and let $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\theta}}_1^{(k)\top}, \ldots, \widehat{\boldsymbol{\theta}}_G^{(k)\top})^\top$ be the estimate 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

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

where

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

By using known properties of conditional expectation, we obtain

$$\begin{aligned} \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)}) &= \frac{\widehat{\pi}_j^{(k)} f_{ij}(\mathbf{V}_i \mid \mathbf{C}_i, \widehat{\boldsymbol{\theta}}_j^{(k)})}{\sum_{j=1}^{G} \widehat{\pi}_j^{(k)} f_{ij}(\mathbf{V}_i \mid \mathbf{C}_i, \widehat{\boldsymbol{\theta}}_j^{(k)})}, \\ \mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)}) \mathrm{E}(U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1), \\ \mathcal{E}_{1ij}(\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)}) \mathrm{E}(U_i \mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1) \text{ and} \\ \mathcal{E}_{2ij}(\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)}) \mathrm{E}(U_i \mathbf{Y}_i \mathbf{Y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1). \end{aligned} \tag{16}$$

The conditional expectations $\mathrm{E}(U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1)$, $\mathrm{E}(U_i \mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1)$, $\mathrm{E}(U_i \mathbf{Y}_i \mathbf{Y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}, Z_{ij} = 1)$ can be directly obtained from the expressions of $\widehat{u}_i^{(k)}$, $\widehat{u\mathbf{y}}_i^{(k)}$, and $\widehat{u\mathbf{y}_i^2}^{(k)}$, respectively, given in Section 2.4. Thus, we have closed form expressions 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 in the FM-tMC model.

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, compute $\mathcal{E}_{sij}(\widehat{\boldsymbol{\theta}}^{(k)})$ for all $s \in \{0, 1, 2\}$ and $\mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)})$ for all $i \in \{1, \ldots, n\}, j \in \{1, \ldots, G\}$.

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:

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

$$\widehat{\boldsymbol{\Sigma}}_j^{(k+1)} = \left\{\sum_{i=1}^{n} \mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}}^{(k)})\right\}^{-1} \sum_{i=1}^{n} \left\{\mathcal{E}_{2ij}(\widehat{\boldsymbol{\theta}}^{(k)}) - \boldsymbol{\mu}_j^{(k+1)}\mathcal{E}_{1ij}^{\top}(\widehat{\boldsymbol{\theta}}^{(k)}) - \mathcal{E}_{1ij}(\widehat{\boldsymbol{\theta}}^{(k)})\boldsymbol{\mu}_j^{(k+1)\top} + \mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}}^{(k)})\boldsymbol{\mu}_j^{(k+1)}\boldsymbol{\mu}_j^{(k+1)\top}\right\},$$

for all $j \in \{1, \ldots, G\}$.

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:

(i) Partition the data (censoring levels replacing the censored observations) into G groups using the K-means clustering algorithm [8].
(ii) Compute the proportion of data points belonging to the same cluster j, say $\pi_j^{(0)}$, $j \in \{1, \ldots, G\}$. This gives the initial value for π_j.
(iii) For each group j, compute the initial values $\boldsymbol{\mu}_j^{(0)}$, $\boldsymbol{\Sigma}_j^{(0)}$ 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) [1], the Bayesian information criterion (BIC) [38] and the efficient determination criterion (EDC) [6]. In a similar form of AIC and BIC, 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. [8] and Cabral et al. [9]. 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 the ML estimates of the parameters in the FM-tMC model 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 (14). In this work we use the alternative method suggested by Basford et al. [7], which consists of approximating the inverse of the covariance matrix by

$$\mathbf{I}_o(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i^{\top}, \quad \text{where } \widehat{\mathbf{s}}_i = \mathrm{E}\left\{\frac{\partial \ell_{ic}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\middle|\mathbf{V}, \mathbf{C}\right\}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \tag{17}$$

where $\ell_{ic}(\boldsymbol{\theta})$ is given in (15) and

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

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:

$$\begin{aligned}
\widehat{s}_{i,\boldsymbol{\mu}_j} &= \widehat{\boldsymbol{\Sigma}}_j^{-1}\{\mathcal{E}_{1ij}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\mu}}_j\},\\
\widehat{s}_{i,\pi_j} &= \frac{\mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}})}{\widehat{\pi}_j} - \frac{\mathcal{Z}_{iG}(\widehat{\boldsymbol{\theta}})}{\widehat{\pi}_G},\\
\widehat{s}_{i,\alpha_{jr}} &= -\frac{1}{2}\mathrm{tr}\left\{\mathcal{Z}_{ij}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_j^{-1}\frac{\partial\boldsymbol{\Sigma}_j}{\partial\alpha_{jr}} - \psi(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_j^{-1}\frac{\partial\boldsymbol{\Sigma}_j}{\partial\alpha_{jr}}\widehat{\boldsymbol{\Sigma}}_j^{-1}\right\},
\end{aligned} \tag{18}$$

where $\psi(\widehat{\boldsymbol{\theta}}) = \{\mathcal{E}_{2ij}(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mu}}_j\mathcal{E}_{1ij}^{\top}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{1ij}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\mu}}_j^{\top} + \mathcal{E}_{0ij}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\mu}}_j\widehat{\boldsymbol{\mu}}_j^{\top}\}$ 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 $j \in \{1, \ldots, G\}$, since ν is assumed to be known.

The information-based approximation (17) is asymptotically applicable. However, it may be not reliable if the sample size is small. The bootstrap approach [15] is a viable alternative to obtain more accurate standard error estimates, however it requires enormous amounts of computing power. As a future research direction, for multivariate Student-t mixture models it is possible to provide more accurate information-based standard errors based on the recent work proposed by Wang and Lin [43].

4. Simulation studies

In order to study the performance of our proposed method, we present three simulation studies. The first one investigates 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.

Table 1
Simulated data: parameter estimation. Mean, standard deviations (Std) for the 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 Section 3.3.

Censored	Measure	Parameter										
		μ_{11}	μ_{12}	$\sigma_{1,11}$	$\sigma_{1,12}$	$\sigma_{1,22}$	π_1	μ_{21}	μ_{22}	$\sigma_{2,11}$	$\sigma_{2,12}$	$\sigma_{2,22}$
$n = 100$												
	True	(−5)	(−4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	−4.95	−3.94	2.90	0.98	4.41	0.65	1.92	2.91	2.06	1.07	3.59
5%	Std.	0.65	0.69	0.88	0.76	1.20	0.05	0.75	0.83	0.89	0.80	1.45
	IM Std	0.39	0.56	1.00	0.94	1.83	0.08	0.44	0.65	1.01	1.06	2.05
	COV	94%	94%	90%	92%	91%	99%	92%	92%	99%	97%	99%
	True	(−5)	(−4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	−5.03	−4.00	2.80	0.88	4.46	0.66	2.04	3.03	1.90	0.93	3.41
10%	Std.	0.34	0.39	0.78	0.64	1.18	0.05	0.38	0.47	0.85	0.76	1.40
	IM Std	0.25	0.32	0.77	0.61	1.17	0.08	0.30	0.40	0.76	0.70	1.31
	COV	94%	94%	89%	91%	92%	99%	94%	93%	99%	96%	99%
	True	(−5)	(−4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	−5.01	−3.91	2.81	1.00	5.02	0.70	2.31	3.35	1.62	0.52	2.83
30%	Std.	0.26	0.36	1.01	0.81	2.00	0.05	0.38	0.42	0.95	0.63	1.29
	IM Std	0.25	0.34	0.91	0.68	1.41	0.09	0.31	0.40	0.70	0.59	1.20
	COV	94%	93%	89%	92%	90%	98%	90%	92%	99%	94%	98%
$n = 400$												
	True	(−5)	(−4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	−5.02	−4.00	2.84	0.89	4.38	0.65	2.01	3.01	1.94	1.00	3.47
5%	Std.	0.12	0.16	0.38	0.31	0.54	0.03	0.15	0.19	0.37	0.34	0.63
	IM Std	0.13	0.16	0.38	0.30	0.55	0.04	0.15	0.20	0.36	0.34	0.64
	COV	95%	95%	92%	91%	92%	99%	94%	96%	99%	96%	99%
	True	(−5)	(−4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	−5.03	−3.98	2.76	0.88	4.55	0.66	2.09	3.08	1.78	0.89	3.27
10%	Std.	0.13	0.16	0.37	0.32	0.58	0.03	0.15	0.19	0.35	0.32	0.67
	IM Std	0.12	0.16	0.38	0.30	0.58	0.04	0.15	0.19	0.34	0.31	0.61
	COV	92%	94%	90%	90%	95%	99%	92%	93%	99%	94%	99%
	True	(−5)	(−4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	−5.00	−3.91	2.81	0.95	4.99	0.70	2.27	3.30	1.65	0.58	2.85
30%	Std.	0.33	0.43	1.05	0.70	1.79	0.05	0.43	0.50	0.98	0.67	1.49
	IM Std	0.25	0.34	0.91	0.67	1.41	0.09	0.31	0.40	0.73	0.61	1.21
	COV	91%	93%	90%	92%	90%	98%	90%	90%	99%	93%	99%
$n = 1000$												
	True	(−5)	(−4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	−5.02	−4.00	2.81	0.87	4.38	0.65	2.02	3.02	1.93	0.98	3.43
5%	Std.	0.08	0.10	0.24	0.19	0.34	0.02	0.09	0.12	0.23	0.22	0.40
	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	93%	95%	93%	92%	92%	100%	95%	96%	99%	92%	99%
	True	(−5)	(−4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	−5.03	−3.97	2.75	0.87	4.52	0.67	2.10	3.07	1.73	0.84	3.26
10%	Std.	0.08	0.10	0.23	0.19	0.36	0.02	0.10	0.13	0.21	0.19	0.39
	IM Std	0.08	0.10	0.24	0.19	0.36	0.03	0.09	0.12	0.21	0.19	0.38
	COV	92%	93%	92%	91%	95%	99%	90%	93%	99%	91%	99%
	True	(−5)	(−4)	(3)	(1)	(4.5)	(0.65)	(2)	(3)	(2)	(1)	(3.5)
	Mean	−5.01	−3.94	2.73	0.91	4.81	0.71	2.30	3.35	1.56	0.48	2.66
30%	Std.	0.08	0.11	0.28	0.22	0.52	0.02	0.11	0.13	0.26	0.16	0.36
	IM Std	0.08	0.10	0.28	0.20	0.43	0.03	0.10	0.12	0.20	0.17	0.33
	COV	94%	91%	90%	91%	90%	98%	90%	92%	98%	92%	98%

Parameter estimation

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 artificially generated data from the model (14), with several censoring proportion settings (5%, 10%, 30%). We generated 500 Monte Carlo (MC) samples of size $n = 100, 400, 1000$. We consider small and different variances with the following parameter setup:

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

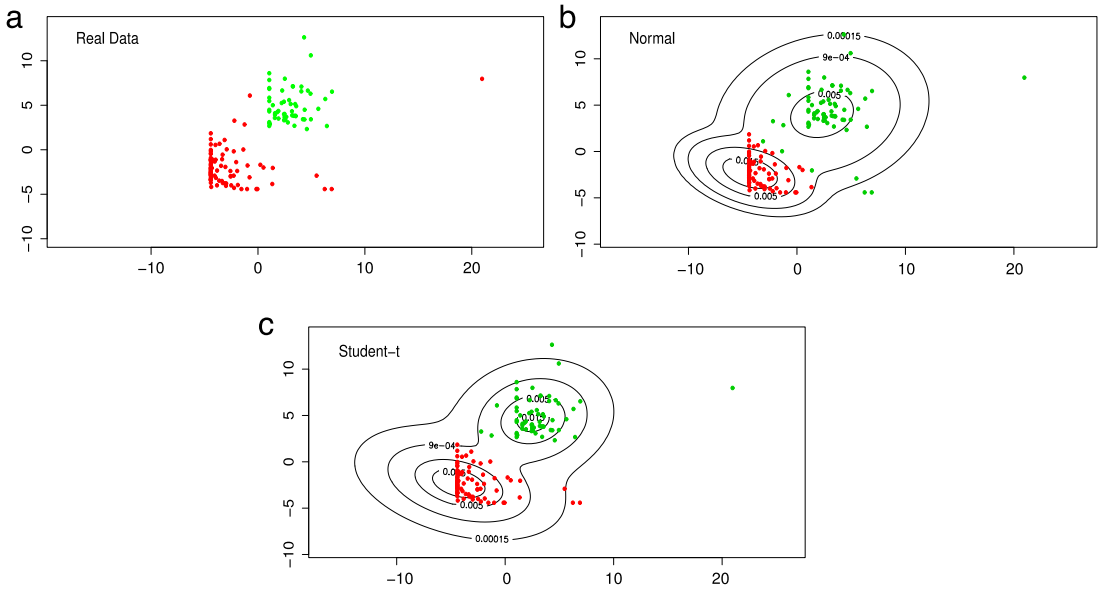

Fig. 1. Simulated data from a mixture of two skew-t models ($n = 150$, 15% of censoring): Clustering—scenario I. (a) Scatter plot for one simulated sample along with the original group (green and red colors) and the respective density contours: (b) FM-nMC fit and allocations; (c) FM-tMC fit and allocations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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

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, with few exceptions, like in the case of scale parameters with censoring level 30%. Thus, in most of the cases, the estimates seem to be insensitive to the variation of the censoring level. In general, the results suggest that the proposed FM-tMC model produces satisfactory estimates, as expected. We also see 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.

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 [9], 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 MC samples of size $n = 60, 150, 500$ with 5%, 15% and 30% of censoring under the following scenarios: (I) scenario 1 (Fig. 1): a mixture of two skew-t models [4], and (II) scenario 2 (Fig. 2): a mixture of two skew-slash distributions [41]. The parameter values were chosen to present a considerable proportion of outliers and skewness pattern. Figs. 1 and 2 show simulated samples of size $n = 150$ with 15% censoring, with the respective density contours.

We proceed with clustering ignoring the known true classification. Following the method proposed by Liu and Lin [26], 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 model was fitted using the algorithm described in 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 the mth sample, $m \in \{1, \ldots, 300\}$, we computed the correct classification rate, denoted by CCR_m. Then we obtained the average $ACCR = \sum_{m=1}^{300} CCR_m/300$.

Figs. 1 and 2 show the allocations in each group. Tables 2 and 3 show the $ACCR$ values. The results are compared with that for the FM-nMC model, which is a mixture of normal multivariate censored models. We can see that the FM-tMC model produces an improvement in the outright clustering, showing the robustness of this model to discrepant observations as well as to censored distributions which seem to occur quite often in practice.

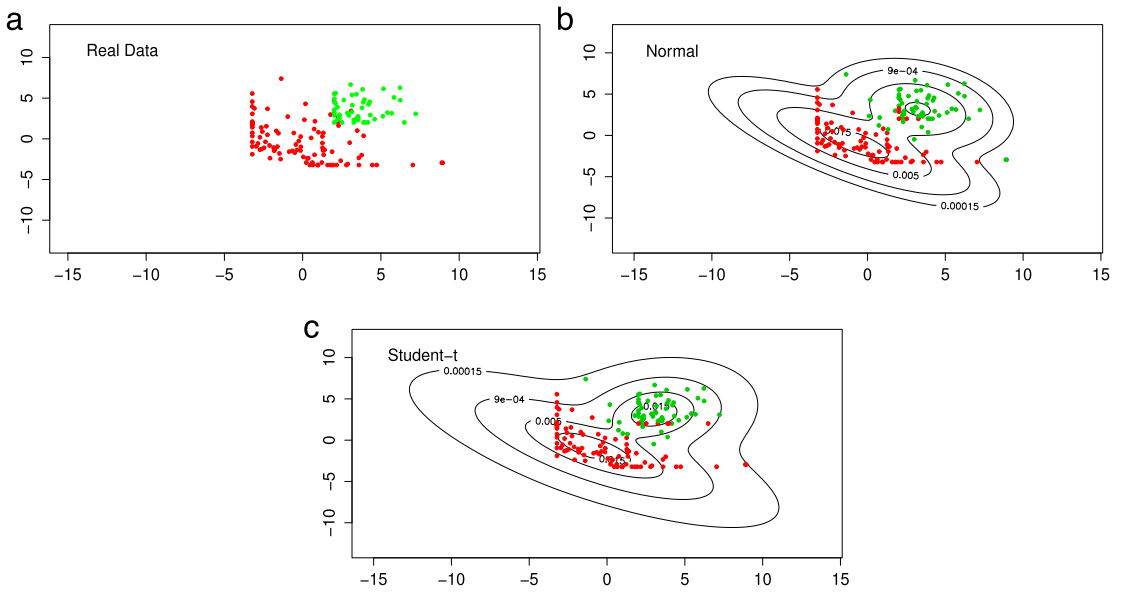

Fig. 2. Simulated data from a mixture of two skew-slash models ($n = 150$, 15% of censoring): Clustering—scenario II. (a) Scatter plot for one simulated sample along with the original group (green and red colors) and the respective density contours: (b) FM-nMC fit and allocations; (c) FM-tMC fit and allocations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Simulated data from a mixture of two skew-t (Scenario I) models ($n = 60, 150, 500$): Clustering. MC mean of right allocation rates for fitted FM-tMCR and FM-nMCR models.

n	5%		15%		30%	
	FM-nMCR	FM-tMCR	FM-nMCR	FM-tMCR	FM-nMCR	FM-tMCR
60	0.863	0.931	0.847	0.921	0.792	0.899
150	0.895	0.957	0.852	0.956	0.833	0.944
500	0.914	0.963	0.875	0.962	0.846	0.951

Table 3
Simulated data from a mixture of two skew-slash (Scenario II) models ($n = 60, 150, 500$): Clustering. MC mean of right allocation rates for fitted FM-tMCR and FM-nMCR models.

n	5%		15%		30%	
	FM-nMCR	FM-tMCR	FM-nMCR	FM-tMCR	FM-nMCR	FM-tMCR
60	0.627	0.683	0.601	0.718	0.536	0.772
150	0.771	0.816	0.788	0.795	0.759	0.784
500	0.794	0.850	0.828	0.837	0.785	0.797

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

$$Bias(\boldsymbol{\theta}_i) = \frac{1}{M}\sum_{m=1}^{M}|\widehat{\boldsymbol{\theta}}_i^{(m)} - \boldsymbol{\theta}_i| \quad \text{and} \quad MSE(\boldsymbol{\theta}_i) = \frac{1}{M}\sum_{m=1}^{M}(\widehat{\boldsymbol{\theta}}_i^{(m)} - \boldsymbol{\theta}_i)^2, \tag{19}$$

where M is the number of MC samples, $\widehat{\boldsymbol{\theta}}_i^{(m)}$ is the ML estimate of the parameter $\boldsymbol{\theta}_i$ for the mth sample. Six different sample sizes ($n = 100, 200, 300, 400, 600, 1000$) were considered.

For each sample size, we generated 500 MC 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:

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

Table 4
VDEQ data. Model selection criteria for various FM-tMC and FM-nMC models. Values in bold correspond to the best model.

Criteria	FM-tMC				FM-nMC	
	$\nu = 3$		$\nu = 4$			
	$G = 2$	$G = 3$	$G = 2$	$G = 3$	$G = 2$	$G = 3$
Log-likelihood	**−1493.04**	−1543.89	−1507.51	−1547.42	−1650.72	−1638.15
AIC	**3038.08**	3151.77	3067.02	3158.84	3353.43	3340.31
BIC	**3121.67**	3254.65	3150.61	3261.72	3437.02	3443.18
EDC	**3056.62**	3174.59	3085.56	3181.66	3371.97	3363.12

The results are depicted in Figs. 3–5, respectively. In general, we can see a pattern of convergence to zero of the (Bias) and MSE when n increases, independent of the censoring pattern (the exceptions are for some scale parameters when the censoring rate is 30%). 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 offer desirable asymptotic properties.

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 and Johnson [21], 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 [40]. The limit of detection is 0.1 μg/L for Cu and Pb, 1.0 mg/L for Zn, 0.5 mg/L for Ca and 1.0 mg/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. Fig. 6 shows the histogram of the concentration levels of each trace metal and all together.

We can see that most of the distributions associated with the individual metals have heavy tails, two or more modes and are skewed to the 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 Section 3.2) as shown in Table 4. Note that, as expected, the FM-tMC model performs significantly better than the FM-nMC model, also, it can be seen that the model with two components and 3 degrees of freedom fits the data best. This finding can be also confirmed from Fig. 7 where the profile 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 covariance matrices to be equal in order to reduce the number of parameters to be estimated.

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

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

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 5. 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, in Table 6, we have the estimates of $\boldsymbol{\Sigma}$ under the FM-tMC and FM-nMC ($\widehat{\boldsymbol{\Sigma}}_t$ and $\widehat{\boldsymbol{\Sigma}}_N$, respectively). Also, we have the respective standard errors of the estimates of the variance components under the FM-tMC model (SE_t), which are less than those under the FM-nMC model (SE_N), indicating that the FM-tMC model produces more precise estimates.

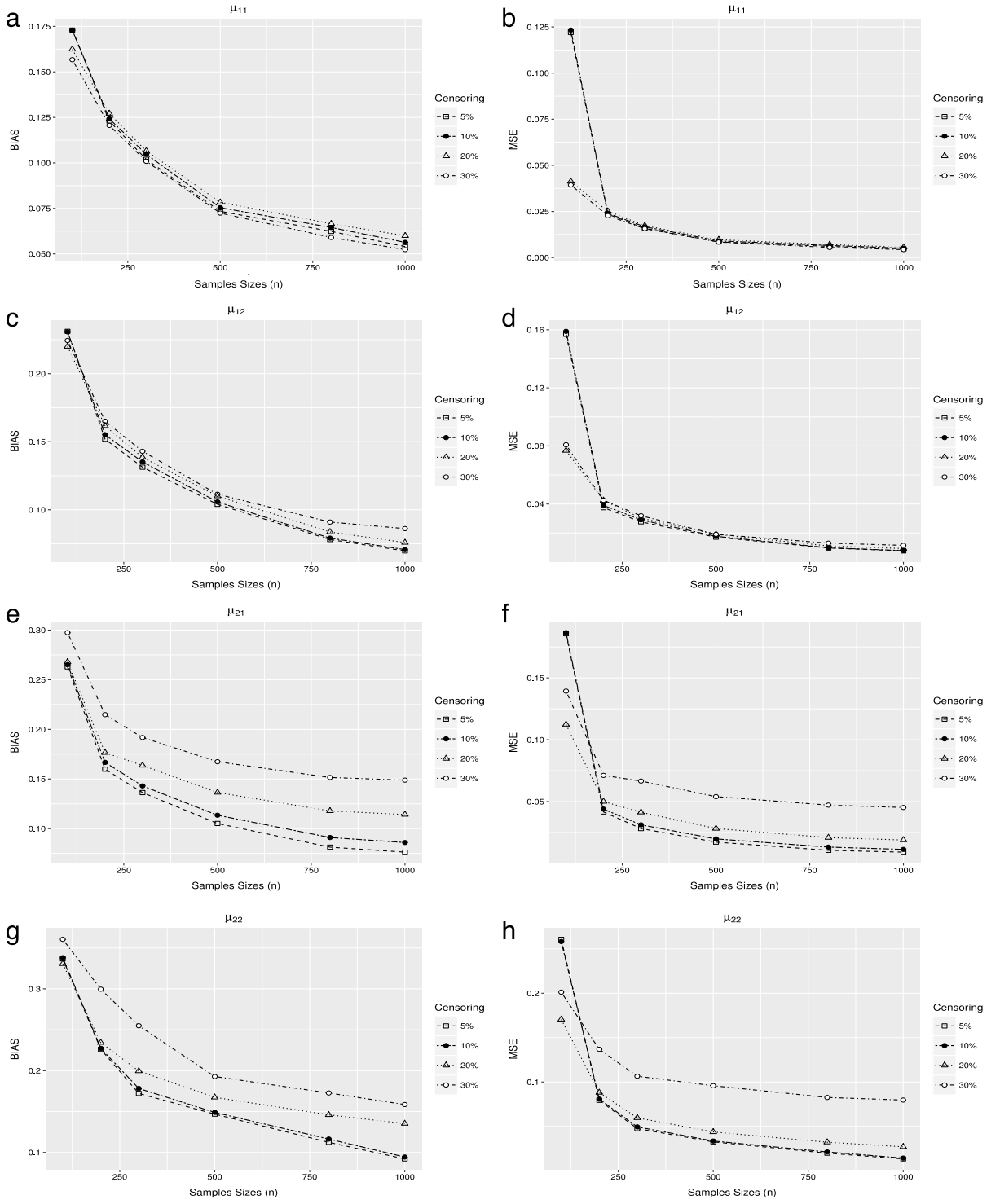

Fig. 3. Simulated data: Asymptotic properties. Bias (first column) and MSE (second column) of the estimates of μ_{11} (a, b), μ_{12} (c, d), μ_{21} (e, f) and μ_{22} (g, h) under the FM-tMC model with different levels of censoring (5%, 10%, 20%, 30%).

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

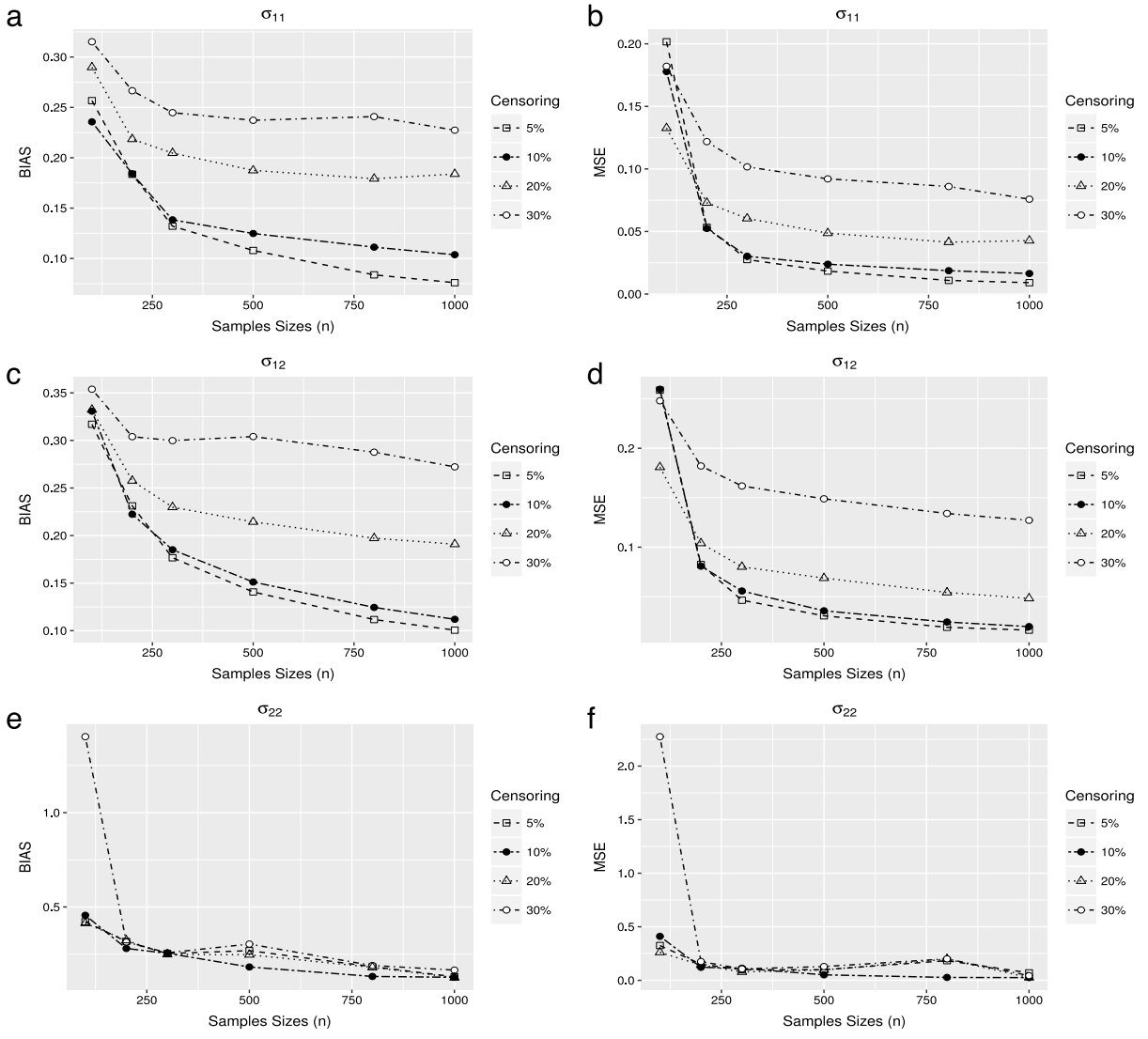

Fig. 4. Simulated data: Asymptotic properties. Bias (first column) and MSE (second column) of the estimates of σ_{11} (a, b), σ_{12} (c, d) and σ_{22} (c, d) under the FM-tMC model with different levels of censoring (5%, 10%, 20%, 30%).

Table 5
VDEQ data. Estimation (Est) and standard errors (SE) for parameters under the FM-nMC and FM-tMC models.

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

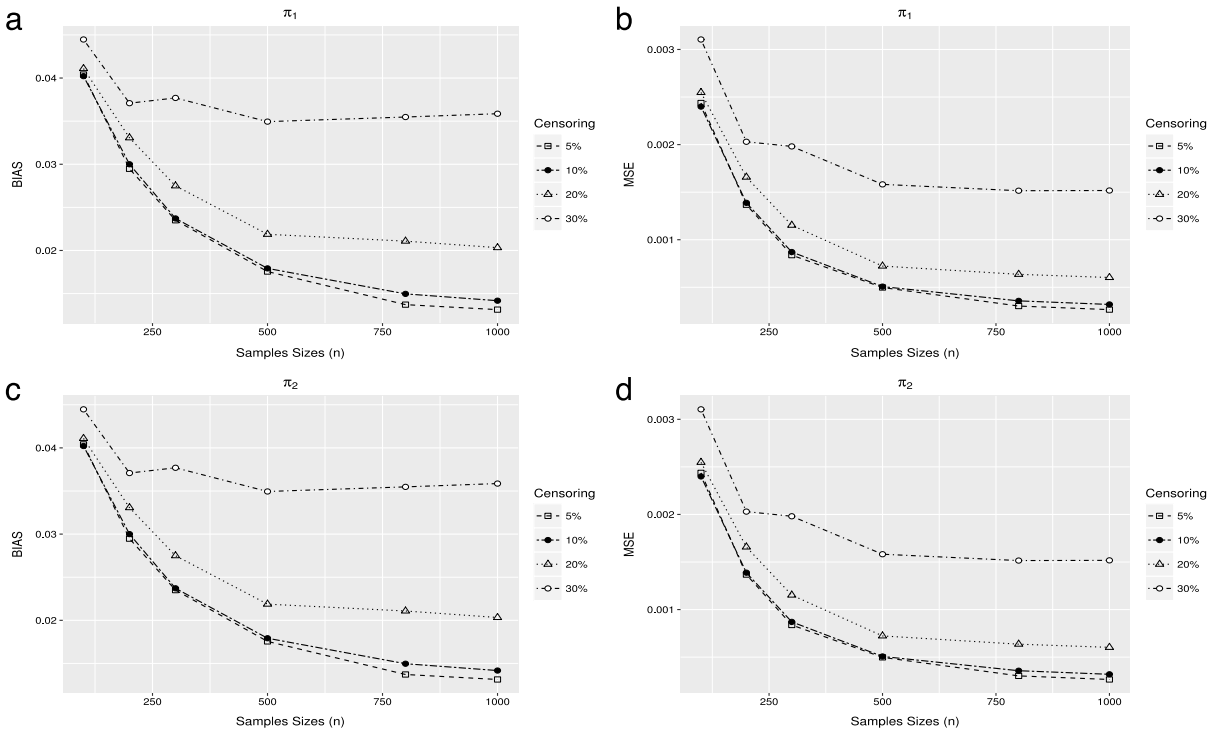

Fig. 5. Simulated data: Asymptotic properties. Bias (first column) and MSE (second column) of the estimates of π_1 (a, b) and π_2 (c, d) under the FM-tMC model with different levels of censoring (5%, 10%, 20%, 30%).

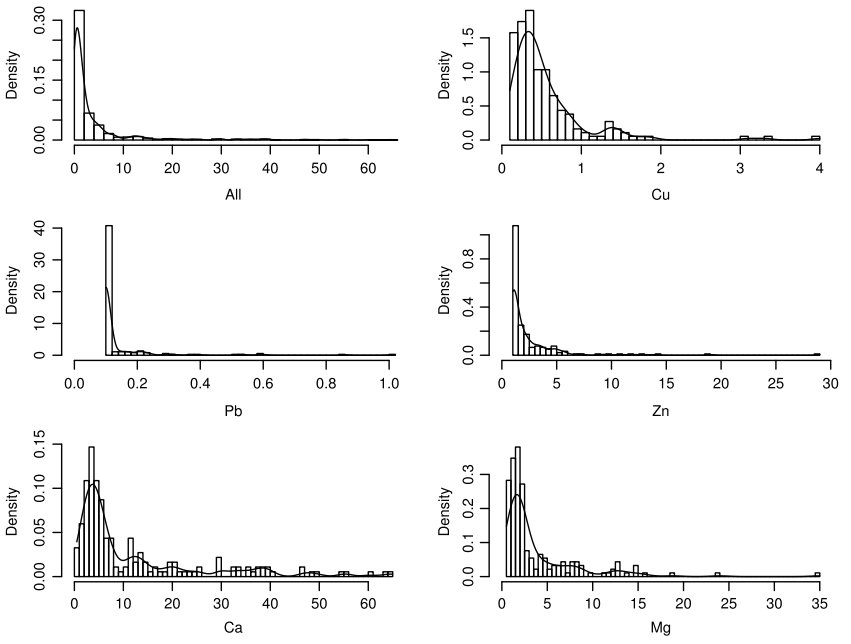

Fig. 6. VDEQ data. Histogram of the dissolved trace metals.

mixture of Gaussian components [11,19,22]. 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 vector and covariance matrix of the multivariate truncated Student-t distributions [20]. The proposed EM algorithm was implemented as part of the R package `CensMixReg` and is

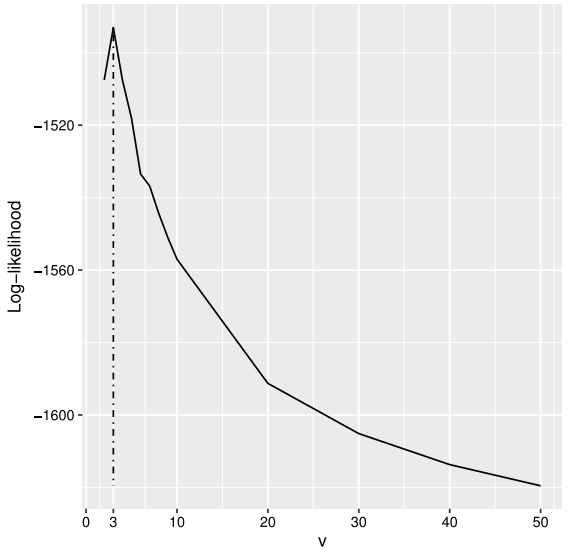

Fig. 7. VDEQ data. Plot of the profile log-likelihood of the degrees of freedom ν.

Table 6
Concentration levels. Covariance matrices estimates under the FM-nMC model ($\widehat{\boldsymbol{\Sigma}}_N$) and under the FM-tMC model ($\widehat{\boldsymbol{\Sigma}}_t$), standard errors under the FM-nMC model (SE_N) and under the FM-tMC model (SE_t).

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

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

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. [17] considered the problem of censored linear regression models using scale mixtures of normal distributions (SMN), which contains as a particular case the Student-t distribution. 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. Other extensions of the current work include, for example, a generalization of the FM-tMC model to the multivariate skew-t distribution [9,23,42] or mixture of linear mixed-effects models with censored observations [5].

Missing observations may frequently occur in practice. Some literature related to handling the missing data problem in the context of finite mixtures of multivariate Student-t models under the missing at random (MAR) mechanism can be found, e.g., in Lin [25] and Wang and Lin [42]. In this setup, a natural extension would be to generalize the current approach for analyzing multivariate data with censored responses and missing values simultaneously.

Acknowledgments

This paper was written while Víctor H. Lachos was a visiting professor in the Department of Statistics at the University of Connecticut, and Celso R.B. Cabral was a visiting professor in the Department of Statistics at the Universidade Estadual de Campinas, Brazil. V.H. Lachos acknowledges the support from CNPq-Brazil (Grant 306334/2015-1) and FAPESP-Brazil (Grant 2014/02938-9). Celso R.B. Cabral acknowledges the support from FAPESP (Grant 2015/20922-5) and CNPq-Brazil (Grant 447964/2014-3). Chen's work is partially supported by the US National Science Foundation (DMS-1613295) and the US National Institutes of Health (U01-HL114494).

References

[1] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Control 19 (1974) 716–723.
[2] R.B. Arellano-Valle, H. Bolfarine, On some characterizations of the t-distribution, Statist. Probab. Lett. 25 (1995) 79–85.

[3] R.B. Arellano-Valle, L. Castro, G. González-Farías, K. Muños Gajardo, Student-t censored regression model: Properties and inference, Stat. Methods Appl. 21 (2012) 453–473.
[4] A. Azzalini, M. Genton, Robust likelihood methods based on the skew-t and related distributions, Internat. Statist. Rev. 76 (2008) 1490–1507.
[5] X. Bai, K. Chen, W. Yao, Mixture of linear mixed models using multivariate t distribution, J. Stat. Comput. Simul. 86 (2016) 771–787.
[6] Z. Bai, P. Krishnaiah, L. Zhao, On rates of convergence of efficient detection criteria in signal processing with white noise, IEEE Trans. Inform. Theory 35 (1989) 380–388.
[7] K. Basford, D. Greenway, G. McLachlan, D. Peel, Standard errors of fitted component means of normal mixtures, Comput. Statist. 12 (1997) 1–18.
[8] R.M. Basso, V.H. Lachos, C.R.B. Cabral, P. Ghosh, Robust mixture modeling based on scale mixtures of skew-normal distributions, Comput. Statist. Data Anal. 54 (2010) 2926–2941.
[9] C.R.B. Cabral, V.H. Lachos, M.O. Prates, Multivariate mixture modeling using skew-normal independent distributions, Comput. Statist. Data Anal. 56 (2012) 126–142.
[10] L.M. Castro, D.R. Costa, M.O. Prates, V.H. Lachos, Likelihood-based inference for Tobit confirmatory factor analysis using the multivariate Student-t distribution, Stat. Comput. 25 (2015) 1163–1183.
[11] S.B. Caudill, A partially adaptive estimator for the censored regression model based on a mixture of normal distributions, Stat. Methods Appl. 21 (2012) 121–137.
[12] S. Chib, Bayes inference in the Tobit censored regression model, J. Econometrics 51 (1992) 79–99.
[13] A. Dempster, N. Laird, D. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1977) 1–38.
[14] V. De Oliveira, Bayesian inference and prediction of Gaussian random fields based on censored data, J. Comput. Graph. Statist. 14 (2005) 95–115.
[15] B. Efron, R.J. Tibshirani, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Statist. Sci. (1986) 54–75.
[16] A. Garay, L. Castro, J. Leskow, V.H. Lachos, Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution, Stat. Methods Med. Res. 26 (2017) 542–566.
[17] A.M. Garay, V.H. Lachos, H. Bolfarine, C.R. Cabral, Linear censored regression models with scale mixtures of normal distributions, Statist. Papers 58 (2017) 247–278.
[18] A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, T. Hothorn, mvtnorm: Multivariate Normal and t Distributions, R package version 1.0-5, 2016. URL http://CRAN.R-project.org/package=mvtnorm.
[19] J. He, Mixture model based multivariate statistical analysis of multiply censored environmental data, Adv. Water Resour. 59 (2013) 15–24.
[20] H.J. Ho, T.I. Lin, H.Y. Chen, W.L. Wang, Some results on the truncated multivariate t distribution, J. Statist. Plann. Inference 142 (2012) 25–40.
[21] H. Hoffman, R. Johnson, Pseudo-likelihood estimation of multivariate normal parameters in the presence of left-censored data, J. Agric. Biol. Environ. Stat. 20 (2015) 156–171.
[22] M. Karlsson, T. Laitila, Finite mixture modeling of censored regression models, Statist. Papers 55 (2014) 627–642.
[23] V.H. Lachos, P. Ghosh, R.B. Arellano-Valle, Likelihood based inference for skew-normal independent linear mixed models, Statist. Sinica 20 (2010) 303–322.
[24] K.L. Lange, R.J.A. Little, J.M.G. Taylor, Robust statistical modeling using t distribution, J. Amer. Statist. Assoc. 84 (1989) 881–896.
[25] T.-I. Lin, Learning from incomplete data via parameterized t mixture models through eigenvalue decomposition, Comput. Statist. Data Anal. 71 (2014) 183–195.
[26] M. Liu, T.-I. Lin, A skew-normal mixture regression model, Educ. Psychol. Meas. 74 (2014) 139–162.
[27] A. Lucas, Robustness of the student t based M-estimator, Commun. Stat. - Theory Methods 26 (1997) 1165–1182.
[28] M.B. Massuia, C.R.B. Cabral, L.A. Matos, V.H. Lachos, Influence diagnostics for Student-t censored linear regression models, Statistics 49 (2015) 1074–1094.
[29] L.A. Matos, V.H. Lachos, N. Balakrishnan, F.V. Labra, Influence diagnostics in linear and nonlinear mixed-effects models with censored data, Comput. Statist. Data Anal. 57 (2013) 450–464.
[30] L.A. Matos, M.O. Prates, M.H. Chen, V.H. Lachos, Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution, Statist. Sinica 23 (2013) 1323–1342.
[31] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, socond ed., Wiley, 2008.
[32] G.J. McLachlan, D. Peel, Finite Mixture Models, Wiley, New York, 2000.
[33] C. Meza, F. Osorio, R. De la Cruz, Estimation in nonlinear mixed-effects models using heavy-tailed distributions, Stat. Comput. 22 (2011) 1–19.
[34] A.F. Militino, M.D. Ugarte, Analyzing censored spatial data, Math. Geol. 31 (1999) 551–561.
[35] D. Peel, G.J. McLachlan, Robust mixture modelling using the t distribution, Stat. Comput. 10 (2000) 339–348.
[36] J.L. Powell, Least absolute deviations estimation for the censored regression model, J. Econometrics 25 (1984) 303–325.
[37] J.L. Powell, Symmetrically trimmed least squares estimation for Tobit models, Econometrica 54 (1986) 1435–1460.
[38] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (1978) 461–464.
[39] F. Vaida, L. Liu, Fast implementation for normal mixed effects models with censored response, J. Comput. Graph. Statist. 18 (2009) 797–817.
[40] VDEQ, 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, 2003, pp. 13–16. URL http://www.deq.virginia.gov/Portals/0/DEQ/Water/WaterQualityMonitoring/ProbabilisticMonitoring/report1.pdf.
[41] J. Wang, M.G. Genton, The multivariate skew-slash distribution, J. Statist. Plann. Inference 136 (2006) 209–220.
[42] W.-L. Wang, T.-I. Lin, Robust model-based clustering via mixtures of skew-t distributions with missing information, Adv. Data Anal. Classif. 9 (2015) 423–445.
[43] W.-L. Wang, T.-I. Lin, Maximum likelihood inference for the multivariate t mixture model, J. Multivariate Anal. 149 (2016) 54–64.
[44] W.-L. Wang, T.-I. Lin, V.H. Lachos, Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails, Stat. Methods Med. Res., http://dx.doi.org/10.1177/0962280215620229.