Multivariate Measurement Error Models Based on Student-t Distribution under Censored Responses

Larissa A. Matos\textsuperscript{a,}\textsuperscript{*}, Luis M. Castro\textsuperscript{b}, Celso R. B. Cabral\textsuperscript{c} and Víctor H. Lachos\textsuperscript{d}

\textsuperscript{a} Department of Statistics, Campinas State University, Campinas, São Paulo, Brazil
\textsuperscript{b} Department of Statistics, Pontificia Universidad Católica de Chile, Chile
\textsuperscript{c} Department of Statistics, Universidade Federal do Amazonas, Amazonas, Brazil
\textsuperscript{d} Department of Statistics, University of Connecticut, Storrs, CT 06269, U.S.A

Abstract
Measurement error models constitute a wide class of models, that include linear and nonlinear regression models. They are very useful to model many real life phenomena, particularly in the medical and biological areas. The great advantage of these models is that, in some sense, they can be represented as mixed effects models, allowing to us the implementation of well-known techniques, like the EM-algorithm for the parameter estimation. In this paper, we consider a class of multivariate measurement error models where the observed response and/or covariate are not fully observed, \textit{i.e.}, the observations are subject to certain threshold values below or above which the measurements are not quantifiable. Consequently, these observations are considered censored. We assume a Student-t distribution for the unobserved true values of the mismeasured covariate and the error term of the model, providing a robust alternative for parameter estimation. Our approach relies on a likelihood-based inference using an EM-type algorithm. The proposed method is illustrated through some simulation studies and the analysis of an AIDS clinical trial dataset.

Key words: Censored responses; EM algorithm; Measurement error models; Student-t distribution.

1. Introduction
Measurement error – hereafter ME – models (also known as error-in-variables models) are defined as regression models where the covariates cannot be measured/observed directly, or are measured with a substantial error. From a practical point of view, such models are very useful because they take into account some notions of randomness inherent to the covariates. For example, in AIDS studies, linear and nonlinear mixed effects models are typically considered to study the relation between the viral load (HIV-1 RNA) measures and CD4\textsuperscript{+} (T helper cells) cell count. However, as pointed out by many authors (see for instance Wu (2010); Bandyopadhyay et al. (2015), among others), this covariate is measured (in general) with substantial error.

From the AIDS research point of view, many strategies to model longitudinal viral load data with measurement error on the covariates have been proposed in the statistical literature. Particularly, we recommend the review of the work of Liang \textit{et al.} (2003), proposing a mixed effects varying-coefficient model under measurement error to study the relationship between the HIV-1 RNA copies with the CD4 cell counts, and more recently Huang \& Dagne (2012a,b); Huang \textit{et al.} (2012); Dagne \& Huang (2013), where a variety of nonlinear mixed effects models under skewness

\textsuperscript{*}Address for correspondence: Larissa Avila Matos, Departamento de Estatística, IMECC, Universidade Estadual de Campinas, CEP 13083-859, Campinas, São Paulo, Brazil. E-mail: larissa.amatos@gmail.com.
and measurement error, among others features, are proposed. It is important to stress that all of these proposals follows the Bayesian paradigm.

A wide variety of proposals exist in the statistical literature trying to deal with the presence of ME in multivariate data. For example, Carrol et al. (1997) proposed a generalized linear mixed ME model and Buonaccorsi et al. (2000) (see also Dumitrescu, 2010) studied estimation of the variance components in a linear mixed effect model with ME in a time varying covariate. Zhang et al. (2011) introduced a multivariate ME model including the presence of zero inflation. Recently, Abarin et al. (2014) proposed a method of moments for the parameter estimation in the linear mixed effect with ME model. Moreover, Cabral et al. (2014) studied a multivariate ME model using finite mixtures of skew Student-t distributions. A comprehensive review of ME models can be found in the books of Fuller (1987), Cheng & Van Ness (1999), Carroll et al. (2006) and Buonaccorsi (2010).

Although many models for multivariate data consider the existence of mismeasured covariates, many of them do not consider censored observations or detection limits for the response variable. This aspect is relevant, since in many studies the observed response is subject to maximum/minimum detection limits. For that reason, clearly there is a need for a new methodology that takes into account censored responses in multivariate data and mismeasured covariates at the same time. We propose an approach where the random observational errors and the unobserved latent variable are jointly modeled by a Student-t distribution, which has heavier tails than the normal one. Besides this, our estimation approach relies on an exact EM-type algorithm, providing explicit expressions for the E and M steps, obtaining as byproduct the asymptotic covariance of the maximum likelihood estimates. To illustrate the applicability of the method, we analyze a real dataset obtained from an AIDS clinical trial.

The paper is organized as follows. Section 2 presents some results about the multivariate Student-t distribution, focusing on its truncated version. Section 3 proposes the ME model for censored multivariate responses under the Student-t distribution. Sections 4 and 5 present the maximum likelihood estimates. To illustrate the applicability of the method, we analyze a real dataset obtained from an AIDS clinical trial.

2. The multivariate Student-t distribution and truncated related ones

We say that the random vector \( Y : p \times 1 \) has a Student-t distribution with location vector \( \mu \), dispersion matrix \( \Sigma \) and \( \nu \) degrees of freedom, when its probability density function (pdf) is given by

\[
t_p(y|\mu, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi^{p/2} \nu^{-p/2} |\Sigma|^{-1/2}} \left(1 + \frac{1}{\nu} (y-\mu)^\top \Sigma^{-1} (y-\mu)\right)^{-\left(\nu+p\right)/2},
\]

where \( \nu \) is a positive real number, not necessarily integer, \( \Gamma(\cdot) \) is the standard gamma function and

\[
d_\Sigma(y, \mu) = (y - \mu)^\top \Sigma^{-1} (y - \mu),
\]

is the Mahalanobis distance. The cumulative distribution function (cdf) of \( Y \) is denoted by \( T_p(\cdot | \mu, \Sigma, \nu) \). If \( \nu > 1 \), \( \mu \) is the mean of \( Y \), and if \( \nu > 2 \), \( \nu(\nu-2)^{-1} \Sigma \) is its covariance matrix. We use the notation \( Y \sim t_p(\mu, \Sigma, \nu) \).

It is possible to show that \( Y \) admits the stochastic representation

\[
Y = \mu + U^{-1/2} Z, \quad Z \sim N_p(0, \Sigma), \quad U \sim \text{Gamma}(\nu/2, \nu/2),
\]

where \( Z \) and \( U \) are independent, and Gamma(\( a, b \)) denotes the gamma distribution with mean \( ab \). As \( \nu \) tends to infinity, \( U \) converges to one with probability one and \( Y \) is approximately distributed as a \( N_p(\mu, \Sigma) \) distribution. From this representation we can easily deduce that an affine transformation \( AY + b \) has a \( t_q(\overline{A}\mu + b, A\Sigma A^\top, \nu) \) distribution, where \( A \) is a \( q \times p \) matrix.
and \( b \) is a \( q \)-dimensional vector. For a reference with extensive material regarding the multivariate Student-t distribution (see Kotz & Nadarajah, 2004).

The following result shows that the Student-t family of distributions is closed under marginalization and conditioning.

**Proposition 1.** Let \( Y \sim t_p(\mu, \Sigma, \nu) \). Consider the partition \( Y = (Y_1^T, Y_2^T)^T \), with \( Y_1 : p_1 \times 1 \) and \( Y_2 : p_2 \times 1 \). Accordingly, consider the partitions \( \mu = (\mu_1^T, \mu_2^T)^T \) and \( \Sigma = (\Sigma_{ij}) \), \( i,j = 1,2 \).

Then

(i) \( Y_1 \sim t_{p_1}(\mu_1, \Sigma_{11}, \nu) \),

(ii) \( Y_2 | Y_1 = y_1 \sim t_{p_2}(\mu_{2,1}, \Sigma_{22,1}, \nu + p_1) \),

where

\[
\mu_{2,1} = \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (y_1 - \mu_1), \quad \Sigma_{22,1} = \frac{\nu + d \Sigma_{11}}{\nu + p_1} \Sigma_{22,1}, \quad \text{and} \quad \Sigma_{22,1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}.
\]

**Proof 1.** The proof can be found in Matos et al. (2013), Proposition 1.

Let \( Y \sim t_p(\mu, \Sigma, \nu) \) and \( \mathbb{D} \) be a Borel set in \( \mathbb{R}^p \). We say that the random vector \( Z \) has a truncated Student-t distribution on \( \mathbb{D} \) when \( Z \) has the same distribution as \( Y | (Y \in \mathbb{D}) \). In this case, the pdf of \( Z \) is given by

\[
T_p(z|\mu, \Sigma, \nu; \mathbb{D}) = \frac{t_p(z|\mu, \Sigma, \nu)}{P(Y \in \mathbb{D})} I_{\mathbb{D}}(z),
\]

where \( I_{\mathbb{D}}(\cdot) \) is the indicator function of \( \mathbb{D} \), that is, \( I_{\mathbb{D}}(z) = 1 \) if \( z \in \mathbb{D} \) and \( I_{\mathbb{D}}(z) = 0 \) otherwise. We use the notation \( Z \sim T_p(\mu, \Sigma, \nu; \mathbb{D}) \). If \( \mathbb{D} \) has the form

\[
\mathbb{D} = \{(x_1, \ldots, x_p) \in \mathbb{R}^p; \ x_1 \leq d_1, \ldots, x_p \leq d_p\}, \quad (3)
\]

then we use the notation \( (Y \in \mathbb{D}) = (Y \leq \mathbf{d}) \), where \( \mathbf{d} = (d_1, \ldots, d_p)^T \). In this case, \( P(Y \leq \mathbf{d}) = T_p(d|\mu, \Sigma, \nu) \). Notice that we can have \( d_i = +\infty \), \( i = 1,\ldots,p \).

The following propositions are crucial to obtain the expectations in the E step of the EM type algorithm, which will be used to compute maximum likelihood estimates of the parameters in the model proposed in this work. We will use the notations \( Z(0) = 1, Z(1) = Z \) and \( Z(2) = ZZ^T \).

**Proposition 2.** Let \( Z \sim T_p(\mu, \Sigma, \nu; \mathbb{D}) \), where \( \mathbb{D} \) is as in (3). Then, for \( k = 0, 1, 2 \),

\[
E \left[ \left( \frac{\nu + p}{\nu + d \Sigma^*(Z, \mu)} \right)^r Z(k) \right] = c_p(\nu,r) \frac{T_p(d|\mu, \Sigma^*, \nu + 2r)}{T_p(d|\mu, \Sigma, \nu)} E[Y(k)], \quad (4)
\]

where \( \nu + 2r > 0 \) and

\[
Y \sim T_p(\mu, \Sigma^*, \nu + 2r; \mathbb{D}), \quad (5)
\]

\[
\Sigma^* = \frac{\nu}{\nu + 2r} \Sigma, \quad c_p(\nu,r) = \left( \frac{\nu + p}{\nu} \right)^r \left( \frac{\Gamma((p + \nu)/2) \Gamma((\nu + 2r)/2)}{\Gamma(\nu/2) \Gamma((p + \nu + 2r)/2)} \right).
\]

**Proof 2.** The proof can be found in Matos et al. (2013), Proposition 2.

Observe that the computation of the expectation on the left-hand side of (4) is reduced to the computation of the moments of the truncated Student-t distribution in (5). These moments are available in closed form in Ho et al. (2012) and the implementations were done using the \( R \) package `TTmoment()`, available on CRAN.
Proposition 3. Let $Z \sim \text{Tt}_p(\mu, \Sigma, \nu; \mathbb{D})$, where $\mathbb{D}$ is as in (3). Consider the partition $Z = (Z^1, Z^2)^T$, with $Z^1 : p_1 \times 1$ and $Z^2 : p_2 \times 1$. Accordingly, consider the partitions $\mu = (\mu^1, \mu^2)^T$ and $\Sigma = (\Sigma_{ij})$, $i, j = 1, 2$. Then,

$$
E \left[ \left( \frac{\nu + p}{\nu + d_{\Sigma}(Z, \mu)} \right)^r Z^{(k)} | Z_1 = z_1 \right] = \frac{h_p(p_1, \nu, r)}{(\nu + d_{\Sigma}(z_1, \mu))^r} \times \frac{T_{p_1}(d_2 | \mu_{2.1}, \Sigma_{22.1}, \nu + p + 2r)}{T_{p_2}(d_2 | \mu_{2.1}, \Sigma_{22.1}, \nu + p)} E[Y^{(k)}],
$$

where $\nu + p + 2r > 0$, $d_2 = (d_{p+1,1}, \ldots, d_p)^T$,

$$
Y \sim \text{Tt}_{p_2}(\mu_{2.1}, \Sigma_{22.1}^*, \nu + p + 2r; \mathbb{D}_2),
$$

$$
\mathbb{D}_2 = \{(x_{p_i+1}, \ldots, x_p) \in \mathbb{R}^{p_2}; \; x_{p_i+1} \leq d_{p_i+1}, \ldots, x_p \leq d_p\},
$$

$$
\Sigma_{22.1}^* = \frac{\nu + d_{\Sigma}(z_1, \mu)}{\nu + p + 2r} \Sigma_{22.1},
$$

$$
h_p(p_1, \nu, r) = (\nu + p)^5 \left( \frac{(p + \nu)/2 \Gamma((p + \nu + 2r)/2)}{\Gamma((p + \nu)/2) \Gamma((p + \nu + 2r)/2)} \right),
$$

$\mu_{2.1}$, $\Sigma_{22.1}$ and $\Sigma_{22.1}^*$ are given in Proposition 1.

Proof 3. The proof can be found in Matos et al. (2013), Proposition 3.

3. Model specification

Let $Y_i = (Y_{i1}, \ldots, Y_{ir})^T$ be the vector of responses for the $i$th experimental unit, where $Y_{ij}$ is the $j$th observed response of unit $i$ (for $i = 1, \ldots, n$ and $j = 1, \ldots, r$). Let $X_i$ be the $i$th observed value of the covariate and $z_i$ be the unobserved (true) covariate value for unit $i$. Following Barnett (1969), the multivariate ME model is formulated as

$$
X_i = z_i + \xi_i
$$

(6)

and

$$
Y_i = \alpha + \beta x_i + e_i,
$$

(7)

where $\alpha = (\alpha_1, \ldots, \alpha_r)^T$ is a vector of regression parameters, $\beta = (\beta_1, \ldots, \beta_r)^T$ are vectors of regression parameters, $\xi_i = (\xi_i, e_i)^T$ and $Z_i = (X_i, Y_i)^T = (z_{i1}, \ldots, z_{ip})^T$. Then, equations (6) and (7) imply

$$
Z_i = a + bx_i + e_i = a + Br_i, \; i = 1, \ldots, n,
$$

(8)

where $a = (0, \alpha^T)^T$ and $b = (1, \beta^T)^T$ are $p \times 1$ vectors, with $p = r + 1$, $B = [b; I_p]$ is a $p \times (p + 1)$ matrix and $r_i = (x_i, e_i^T)^T$. Since the first component of $a$ is equal to zero and the first component of $b$ is equal to one, then the model is identifiable as discussed in Galea et al. (2005) (see also Vidal & Castro, 2010). Thus, from equation (8), the distribution of $Z_i$ becomes specified once the distribution of $r_i$ is specified. Usually, a normality assumption is made, such that

$$
r_i \overset{iid}{\sim} N_1+p \left( \begin{pmatrix} \mu_x \\ 0_p \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0_p^T \\ 0_p & \Omega \end{pmatrix} \right), \; i = 1, \ldots, n,
$$

(9)

where $0_p = (0, \ldots, 0)^T : p \times 1$, $\Omega = \text{diag}(\omega_1^2, \ldots, \omega_p^2)$ and $iid$ denotes independent and identically distributed random vectors. Marginally, we have that $x_i \overset{iid}{\sim} N(\mu_x, \sigma_x^2)$ and $e_i \overset{iid}{\sim} N(0, \Omega)$ are independent for all $i = 1, \ldots, n$. For more details see, for example, (Fuller, 1987, Sec. 4.1).

However, even though Gaussian assumption is mostly reasonable, it may lack robustness in parameter estimation under presence of heavy tails and outliers (see Pinheiro et al., 2001). Note
that, in this context, robustness is defined as the capability of a statistical model to be unaffected by outliers and/or influential observations.

Although in the statistical literature there exists several methods for data transformation in order to achieve normality, they present some problems (see for example Azzalini & Capitanio, 1999). Hence, an appropriate theoretical but ‘robust’ framework that avoids data transformation is desirable.

Following these ideas, we propose to replace assumption (9) by

$$
\mathbf{r}_i = \begin{bmatrix} x_i \\ \epsilon_i \end{bmatrix} \overset{\text{iid}}{\sim} t_{1+p} \left( \begin{pmatrix} \mu_x \\ \mathbf{0}_p \end{pmatrix}, \begin{pmatrix} \sigma^2_x \\ \mathbf{0}_p \\ \mathbf{0}_p \end{pmatrix}, \nu \right), \quad i = 1, \ldots, n. \tag{10}
$$

By (2), this formulation implies

$$
\begin{align*}
\begin{bmatrix} x_i \\ \epsilon_i \end{bmatrix} | U_i = u_i & \sim N_{1+p} \left( \begin{pmatrix} \mu_x \\ \mathbf{0}_p \end{pmatrix}, u_i^{-1} \begin{pmatrix} \sigma^2_x \\ \mathbf{0}_p \\ \mathbf{0}_p \end{pmatrix} \right), \\
U_i & \sim \text{Gamma} \left( \frac{\nu}{2}, \frac{\nu}{2} \right),
\end{align*}
$$

for $i = 1, \ldots, n$. Consequently,

$$
\begin{align*}
x_i | U_i = u_i & \overset{\text{iid}}{\sim} N(\mu_x, u_i^{-1}\sigma^2_x) \quad \text{and,} \\
\epsilon_i | U_i = u_i & \overset{\text{iid}}{\sim} N_p(0, u_i^{-1}\mathbf{I}). \tag{11, 12}
\end{align*}
$$

Besides this, $\epsilon_i$ and $x_i$ have Student-t marginal distributions, with $\epsilon_i \sim t_p(0, \Omega, \nu)$ and $x_i \sim t(\mu_x, \sigma^2_x, \nu)$.

Since for each $i$, $\epsilon_i$ and $x_i$ are indexed by the same scale mixing factor $U_i$, they are not independent in general. The independence corresponds to the case where $U_i = 1$ (normal case). However, conditional on $U_i$, $\epsilon_i$ and $x_i$ are independent for each $i = 1, \ldots, n$, which implies that $\epsilon_i$ and $x_i$ are not correlated, since $\text{Cov}(\epsilon_i, x_i) = \mathbb{E}[\epsilon_i x_i | U_i] = 0$. By (8), $Z_i$ is an affine transformation of $\mathbf{r}_i$. Thus, its distribution is given by

$$
Z_i \sim t_p(\mu_z, \Sigma_z, \nu), \quad i = 1, \ldots, n, \tag{13}
$$

where

$$
\mu_z = \mathbf{a} + \mathbf{b} \mu_x \quad \text{and} \quad \Sigma_z = \sigma^2_z \mathbf{b} \mathbf{b}^\top + \Omega. \tag{14}
$$

As mentioned earlier, our model considers censored observations. Following Matos et al. (2013), we consider the case in which the response $Z_{ij}$ is not fully observed for all $i, j$. What we truly observe, for each $i = 1, \ldots, n$, is the random vector $V_i = (V_{i1}, \ldots, V_{ip})^\top$, such that $V_{ij} = \max\{Z_{ij}, \kappa_{ij}\}$, where $\kappa_{ij}$ is a censoring level, that is,

$$
V_{ij} = \begin{cases} Z_{ij} & \text{if} \quad Z_{ij} > \kappa_{ij} \\
\kappa_{ij} & \text{if} \quad Z_{ij} \leq \kappa_{ij}. \end{cases} \tag{15}
$$

The model defined by Equations (6), (7) along with (10) and (15) is named the Student-t Censored Measurement Error Model – hereafter t-MEC model.

3.1. The likelihood function

In this section we present the likelihood function, which will be used in the model selection computations to compare fitted models.

First, let us partition $\mathbf{Z}_i$ into the observed and censored components, namely, $\mathbf{Z}_i = \text{vec}(\mathbf{Z}_i^{\circ}, \mathbf{Z}_i^{\circ\circ})$, where $\mathbf{Z}_i^{\circ} : p_o \times 1$ corresponds to the former case, $\mathbf{Z}_i^{\circ\circ} : p_c \times 1$ corresponds to the latter and $\text{vec}(\cdot)$ denotes the function which stacks vectors or matrices of the same number of columns. Accordingly, let us consider $\mathbf{V}_i = \text{vec}(\mathbf{V}_i^{\circ}, \mathbf{V}_i^{\circ\circ})$ and, recalling that $\mathbf{Z}_i \sim t_p(\mu_z, \Sigma_z, \nu)$, see (13), $\mathbf{Z}_i^{\circ} = \text{vec}(\mu_z^\circ, \Sigma_z^\circ)$ and $\Sigma_z^\circ = \frac{\Sigma_z^{(o)} \Sigma_z^{(co)}}{\Sigma_z^{(o)} + \Sigma_z^{(co)}}$). $\kappa^c_i$ is the vector with the corresponding censoring levels for $\mathbf{Z}_i^{\circ\circ}$. By Proposition 1, we have

$$
\begin{align*}
\mathbf{Z}_i^{\circ} & \sim t_{p_o}(\mu_z^\circ, \Sigma_z^{(o)}, \nu) \quad \text{and} \quad \mathbf{Z}_i^{\circ\circ} | \mathbf{Z}_i^{\circ} = \mathbf{Z}_i^{\circ\circ} \sim t_{p_c}(\mu_z^{co}, \Sigma_z^{co}, \nu + p_o), \tag{16}
\end{align*}
$$
where
\begin{align}
\mu_z^\theta &= \mu_z^c + \Sigma_z^\theta (\Sigma_z^{oo})^{-1}(z_i - \mu_z^c), \\
S_z^\theta &= \frac{\nu + d\Sigma_z^{oo}(\mu_z^c, \mu_z^c)}{\nu + p} \Sigma_z^{cc,o}, \\
\Sigma_z^{cc,o} &= \Sigma_z^{cc} - \Sigma_z^{co} \Sigma_z^{oo}^{-1} \Sigma_z^{oc}.
\end{align}

The observed sample for the experimental unit \(i\) is \(\{z_i^c, \kappa_i^c\}\). The associated likelihood for \(\theta = (\alpha, \beta, \omega, \mu_z, \sigma_z^2)^T\) is
\[L_i(\theta) = P(V_i^c = \kappa_i^c| Z_i^c = z_i^c) f(z_i^c),\]
where \(f(\cdot)\) is the marginal density of \(Z_i^c\) and \(\omega = (\omega_i^2, \ldots, \omega_i^2)\). But \(V_i^c = \kappa_i^c\) if and only if \(Z_i^c \leq \kappa_i^c\). By (16), we obtain
\[L_i(\theta) = t_{p_\nu}(\kappa_i^c| \mu_z^c, S_z^\theta, \nu + p_\nu) t_{p_\nu}(z_i^c| \mu_z^c, S_z^{co}, \nu).
\]
The log-likelihood associated with the whole sample is
\[\ell(\theta) = \sum_{i=1}^n \log L_i(\theta).
\]

4. The ECM algorithm

In this section, we describe how the t-MEC model can be fitted by using the ECM algorithm (Meng & Rubin, 1993). This algorithm considers a simple modification of the traditional EM algorithm initially proposed by Dempster et al. (1977) and is an efficient tool to obtain the maximum likelihood estimates under a missing data framework.

The t-MEC model can be formulated in a flexible hierarchical representation that is useful for theoretical derivations. It is easily obtained through Equations (8), (11) and (12) and is given by
\begin{align}
Z_i | x_i, U_i &= u_i \sim \mathcal{N}(\mu_z, u_i^{-1} \Omega), \\
x_i | U_i &= u_i \sim \mathcal{N}(\mu_z, u_i^{-1} \sigma_z^2), \\
U_i & \sim \text{Gamma}(\nu/2, \nu/2), \quad i = 1, \ldots, n.
\end{align}

Following the suggestions of Lange et al. (1989) and Lucas (1997), who pointed out difficulties in estimating \(\nu\) due to problems of unbounded and local maxima in the likelihood function, we consider the value of \(\nu\) to be known.

Now, we enunciate two important results that will be useful in the E step of the EM algorithm.

**Proposition 4.** Consider the hierarchical representation of the t-MEC model given in (21)–(23). Then,
\[x_i|U_i = u_i, Z_i = z_i \sim \mathcal{N}\left(\frac{\mu_z + \sigma_z^2 b' \Omega^{-1} (z_i - a)}{1 + \sigma_z^2 b' \Omega^{-1} b}, \frac{\sigma_z^2}{u_i (1 + \sigma_z^2 b' \Omega^{-1} b)}\right).
\]

**Proof 4.** The proof follows directly from the relation \(f(x_i|u_i, z_i) \propto f(z_i|x_i, u_i) f(x_i|u_i)\), where \(f(\cdot)\) denotes a generic pdf.

**Proposition 5.** For the t-MEC model,
\[E(U_i|Z_i = z_i) = \frac{p + \nu}{d\Sigma_z(z_i, \mu_z) + \nu}.
\]

**Proof 5.** To prove this result, recall that \(Z_i \sim t_p(\mu_z, \Sigma_z, \nu)\), which implies \(Z_i|U_i = u_i \sim N_p(\mu_z, u_i^{-1} \Sigma_z)\) and \(U_i \sim \text{Gamma}(\nu/2, \nu/2)\) – see (2) –. Using the relation \(f(u_i|z_i) \propto f(z_i|u_i) f(u_i)\), we can prove that \(U_i|Z_i = z_i \sim \text{Gamma}\left(\frac{p + \nu}{2}, \frac{1}{2} d\Sigma_z(z_i, \mu_z) + \nu\right)\), and the result follows.
4.1. The E Step

Let \( Z = (Z_1^T, \ldots, Z_n^T)^T \), \( x = (x_1, \ldots, x_n)^T \) and \( u = (u_1, \ldots, u_n)^T \). Let \( \theta \) be the vector with all the parameters in the model. Apart from constants which do not depend on \( \theta \), the complete log-likelihood associated with the complete data \( Z_c = \{Z, x, u\} \) is given by

\[
\ell_c(\theta | Z_c) = -\frac{n}{2} \sum_{i=1}^{p} \log \omega_j^2 - \frac{1}{2} \sum_{i=1}^{n} u_i (Z_i - a - bx_i)^T \Omega^{-1} (Z_i - a - bx_i) \\
- \frac{n}{2} \log \sigma_x^2 - \frac{1}{2 \sigma_x^2} \sum_{i=1}^{n} u_i (x_i - \mu_x)^2.
\]

Suppose that at the \( k \)th stage of the algorithm we obtain an estimate \( \hat{\theta}^{(k)} \) of \( \theta \). The E step consists of the computation of the conditional expectation

\[
Q(\theta | \hat{\theta}^{(k)}) = E_{\theta^{(k)}} \left[ \ell_c(\theta | Z_c) | V \right],
\]

where \( E_{\theta^{(k)}} \) means that the expectation is being affected using \( \hat{\theta}^{(k)} \) as the true parameter value and \( V = (V_1^T, \ldots, V_m^T)^T \). The M step consists of maximizing \( Q(\cdot | \hat{\theta}^{(k)}) \) in \( \theta \). To do so, first observe that the function \( Q(\cdot | \hat{\theta}^{(k)}) \) can be decomposed into

\[
Q \left( \theta | \hat{\theta}^{(k)} \right) = Q_1 \left( \alpha, \beta, \omega | \hat{\theta}^{(k)} \right) + Q_2 \left( \mu_x, \sigma_x^2 | \hat{\theta}^{(k)} \right),
\]

where \( \omega = (\omega_1^2, \ldots, \omega_p^2)^T \),

\[
Q_1 \left( \alpha, \beta, \omega | \hat{\theta}^{(k)} \right) =
E_{\theta^{(k)}} \left[ -\frac{n}{2} \sum_{i=1}^{p} \log \omega_j^2 - \frac{1}{2} \sum_{i=1}^{n} u_i (Z_i - a - bx_i)^T \Omega^{-1} (Z_i - a - bx_i) | V \right]
\]

and

\[
Q_2 \left( \mu_x, \sigma_x^2 | \hat{\theta}^{(k)} \right) = E_{\theta^{(k)}} \left[ -\frac{n}{2} \log \sigma_x^2 - \frac{1}{2 \sigma_x^2} \sum_{i=1}^{n} u_i (x_i - \mu_x)^2 | V \right].
\]

Given this decomposition, we can reduce the problem to the maximization of two independent functions, searching for critical points of \( Q_1(\cdot | \hat{\theta}^{(k)}) \) and \( Q_2(\cdot | \hat{\theta}^{(k)}) \) separately. Expanding the expressions of \( Q_1(\cdot | \hat{\theta}^{(k)}) \) and \( Q_2(\cdot | \hat{\theta}^{(k)}) \) and taking expectations, it follows that

\[
Q_1 \left( \alpha, \beta, \omega | \hat{\theta}^{(k)} \right) = -\frac{n}{2} \sum_{i=1}^{p} \log \omega_j^2 - \frac{1}{2} \sum_{i=1}^{n} \left\{ \text{tr} \left( \Omega^{-1} u \hat{z}_i^2 \right) - 2a^T \Omega^{-1} u \hat{z}_i \\
- 2u^T \hat{x} \Omega^{-1} b + a^T \Omega^{-1} a \hat{u}_i + 2a^T \Omega^{-1} b \hat{u}_i + b^T \Omega^{-1} b \hat{u}_i^2 \right\},
\]

\[
Q_2(\mu_x, \sigma_x^2 | \hat{\theta}^{(k)}) = -\frac{n}{2} \log \sigma_x^2 - \frac{1}{2 \sigma_x^2} \sum_{i=1}^{n} \left\{ u \hat{x}_i^2 - 2 \mu_x u \hat{x}_i + \mu_x^2 \hat{u}_i \right\},
\]

where \( \text{tr}(\cdot) \) denotes the trace of a matrix,

\[
\hat{u} \hat{x}_i^2 = E[U_i, Z_i^T | V_i], \quad \hat{u} \hat{z}_i = E[U_i, Z_i | V_i], \\
\hat{u}_i = E[U_i | V_i], \quad \hat{u} \hat{z}_i = E[U_i, x_i^T | V_i], \\
\hat{u} \hat{x}_i = E[U_i, x_i | V_i], \quad \hat{u} \hat{x}_i^2 = E[U_i, x_i^2 | V_i],
\]

7
and we have omitted the subscript $\hat{\theta}^{(k)}$ to simplify the notation. To obtain expressions for these expectations, we will use a result from probability theory called the tower property of conditional expectation: if $X$ and $Y$ are arbitrary random vectors and $f(\cdot)$ is a measurable function, then $\mathbb{E}[\mathbb{E}(X|Y)] = \mathbb{E}[X|f(Y)]$. For a proof, see (Ash, 2000, Theo. 5.5.10). Now, observe that, by (15), $V_i$ is a function of $Z_i$. Then, by this property, we can write

$$
\hat{w}_i^2 = \mathbb{E}\{E[U_i | Z_i, Z_i^0] | V_i]\}, \quad \hat{u}_i^2 = \mathbb{E}\{E[U_i | Z_i] | V_i]\}, \quad \hat{u}_i = \mathbb{E}\{E[U_i | Z_i] | V_i}\}. \quad (26)
$$

Proposition 5 gives the conditional expectation $\mathbb{E}[U_i | Z_i]$ and, from this result and formulas (26) we obtain the following expressions for $\hat{u}_i$, $\hat{u}_i$ and $\hat{w}_i^2$ (as we will see soon, all expectations involved in the E step are written as functions of these), considering three different cases:

(i) Individual $i$ does not have censored components. In this case, $V_i = Z_i$ – see Equation (15) –. Thus,

$$
\hat{u}_i = \mathbb{E}[E[U_i | Z_i] | Z_i \leq \kappa_i] = \mathbb{E}\left[\frac{p + \nu}{d_{\Sigma_z}(Z_i, \mu_z) + \nu} | Z_i \leq \kappa_i\right].
$$

By (13) and the definition of a truncated Student-t distribution, we have that $Z_i | (Z_i \leq \kappa_i) \sim \text{T}_p(\mu_z, \Sigma_z, \nu; \mathbb{D}_i)$, where $\mathbb{D}_i$ is like in (3) with $d = \kappa_i$. Using $r = 1$ and $k = 0$ in Proposition 2, we get

$$
\hat{u}_i = \frac{T_p(\kappa_i | \mu_z, \Sigma_z^*, \nu + 2r)}{T_p(\kappa_i | \mu_z, \Sigma_z, \nu)},
$$

where $\Sigma_z^* = (\nu/(\nu + 2))\Sigma_z$. Using $r = 1$ and $k = 1$ in Proposition 2, we obtain

$$
\hat{u}_i = \frac{T_p(\kappa_i | \mu_z, \Sigma_z^*, \nu + 2r)}{T_p(\kappa_i | \mu_z, \Sigma_z, \nu)} \mathbb{E}[Y_i],
$$

where $Y_i \sim \text{T}_p(\mu_z, \Sigma_z^*, \nu + 2; \mathbb{D}_i)$. Finally, $r = 1$ and $k = 2$ in Proposition 2 imply

$$
\hat{w}_i^2 = \frac{T_p(\kappa_i | \mu_z, \Sigma_z^*, \nu + 2r)}{T_p(\kappa_i | \mu_z, \Sigma_z, \nu)} \mathbb{E}[Y_i Y_i^\top].
$$

(ii) Individual $i$ has only censored components. By Equation (15), this fact occurs if and only if $Z_i \leq \kappa_i$, where $\kappa_i$ is the vector with the censoring levels for individual $i$. Thus,

$$
\hat{u}_i = \mathbb{E}[E[U_i | Z_i] | Z_i \leq \kappa_i] = \mathbb{E}\left[\frac{p + \nu}{d_{\Sigma_z}(Z_i, \mu_z) + \nu} | Z_i \leq \kappa_i\right].
$$

In this case, we have that $Z_i | (Z_i \leq \kappa_i) \sim \text{T}_p(\mu_z, \Sigma_z, \nu; \mathbb{D}_i^c)$, with

$$
\mathbb{D}_i^c = \{(x_1, \ldots, x_p) \in \mathbb{R}^p; \ x_i \leq \kappa_i, \ i \in \mathcal{C}\}, \quad (27)
$$

where $\mathcal{C}$ is the set of indices for the censored components. Consequently, we make $d_i = +\infty$ for $i \notin \mathcal{C}$ in (3). Thus, $\hat{u}_i$ can be calculated using Proposition 3, with $Z_i^c$ and $Z_i^0$ playing the role of $Z_1$ and $Z_2$, respectively, taking $r = 1$ and $k = 0$. Then, we get

$$
\hat{u}_i = \frac{p^c + \nu}{\nu + d_{\Sigma_z^c}(Z_i^c, \mu_z^c)} \frac{T_p(\kappa_i^{(c)} | \mu_z^c, \Sigma_z^c, \nu + p_o + 2)}{T_p(\kappa_i^{(c)} | \mu_z^c, \Sigma_z^c, \nu + p_o)},
$$

$$
\hat{w}_i^2 = \mathbb{E}[E[U_i | Z_i, Z_i^0] | V_i].
$$

8
where \( p^o \) and \( p^c \) are the dimensions of the vectors \( \hat{Z}_i^o \) and \( Z_i^c \), respectively, \( \nu + p^o + 2 > 0, \)
\[
\hat{S}_{z^{co}}^o = \frac{\nu + d_{\Sigma^o}(Z^o, \mu^o)}{\nu + p^o + 2} \Sigma_{cc},
\]
\( \mu^o_z, \hat{S}_{z^{co}}^o \) and \( \Sigma_{cc}^{co} \) are given in (17), (18) and (19), respectively. Regarding \( \hat{u}\hat{Z}_i \), we have that
\[
\hat{u}\hat{Z}_i = E \left[ \frac{p + \nu}{d_{\Sigma}(Z_i, \mu_z)} \text{vec}(Z_i) \bigg| Z_i = z_i^o, Z_i^c \leq \kappa_i^c \right]
\]
\[
= \text{vec} \left( E \left[ \frac{p + \nu}{d_{\Sigma}(Z_i, \mu_z)} z_i^o \bigg| Z_i = z_i^o, Z_i^c \leq \kappa_i^c \right] \right) E \left[ \frac{p + \nu}{d_{\Sigma}(Z_i, \mu_z)} Z_i^c \bigg| Z_i = z_i^o, Z_i^c \leq \kappa_i^c \right]
\]
\[
= \text{vec}(\hat{u}_i Z_i^o, E[Y_i]),
\]
where
\[
Y_i \sim T_{p^c}(\mu_z^o, \hat{S}_{z^{co}}^o, \nu + p^o + 2; D_i^c).
\]

Finally, to compute \( \hat{u}\hat{Z}_i \), observe that
\[
\hat{u}\hat{Z}_i^2 = E \left[ \frac{p + \nu}{d_{\Sigma}(Z_i, \mu_z)} \left( \begin{array}{c} \hat{Z}_i^o \hat{Z}_i^o \top \\ \hat{Z}_i^o \hat{Z}_i^c \top \\ \hat{Z}_i^c \hat{Z}_i^c \top \end{array} \right) \bigg| \hat{Z}_i = z_i^o, Z_i^c \leq \kappa_i^c \right]
\]
\[
= \left( \begin{array}{c} \hat{u}_i \hat{Z}_i^o \hat{Z}_i^o \top \\ \hat{u}_i E[Y_i] \hat{Z}_i^c \top \\ \hat{u}_i E[Y_i] \hat{Z}_i^c \top \end{array} \right)
\]
where \( Y_i \) is as in (28).

Regarding the remaining expectations, we have
\[
E[x_i | U_i, V_i = v_i] = \int \int x_i u_i \pi(x_i, u_i | v_i) dx_i du_i
\]
\[
= \int x_i \pi(x_i | u_i, v_i) dx_i \int u_i \pi(u_i | v_i) du_i
\]
\[
= E[x_i | U_i = u_i, V_i = v_i] E[U_i | V_i = v_i].
\]

By the tower property, we have
\[
E[x_i | U_i, V_i] = E[ E(x_i | U_i, Z_i) | U_i, V_i].
\]

Consequently,
\[
\hat{u} \hat{x}_i = E[U_i | V_i] = E \left[ \frac{\mu_x + \sigma^2 b^\top \Omega^{-1} (Z_i - a)}{1 + \sigma^2 b^\top \Omega^{-1} b} | U_i, V_i \right] E[U_i | V_i]
\]
\[
= \frac{\mu_x E[U_i | V_i] + \sigma^2 b^\top \Omega^{-1} E[Z_i | U_i, V_i] E[U_i | V_i] - a E[U_i | V_i]}{1 + \sigma^2 b^\top \Omega^{-1} b}
\]
\[
= \mu_x \hat{u}_i + \sigma^2 b^\top \Omega^{-1} \hat{Z}_i - \sigma^2 b^\top \Omega^{-1} a \hat{u}_i
\]
\[
= \mu_x \hat{u}_i + \varphi(\hat{u}\hat{Z}_i - \mu_z \hat{u}_i),
\]
where
\[
\varphi = \frac{\sigma^2 b^\top \Omega^{-1}}{1 + \sigma^2 b^\top \Omega^{-1} b}
\]
and the equality in (30) is obtained by proving that \( E[Z_i | U_i, V_i] E[U_i | V_i] = E[U_i Z_i | V_i] = \hat{u}\hat{Z}_i \), which can be done following the same paths that led to (29), replacing \( x_i \) with \( Z_i \).
In a similar fashion, we get
\[
\hat{u}_i^2 = \Lambda + \mu_x^2 \hat{u}_i + 2\varphi [\hat{u} \hat{z}_i - \mu_x \hat{u}_i] + \varphi \left[ \hat{u} \hat{z}_i^2 - \hat{u} \hat{z}_i \mu_x^\top - \mu_x \hat{z}_i \mu_x^\top + \mu_x \mu_x^\top \hat{u}_i \right] \varphi^\top, \quad \text{and}
\]
\[
\hat{u}_i \hat{z}_i = \mu_x \hat{z}_i + \varphi \left[ \hat{u} \hat{z}_i - \mu_x \hat{z}_i \right],
\]
with
\[
\Lambda = \frac{\sigma^2}{1 + \sigma^2 \mathbf{b} \cdot \mathbf{\Omega}^{-1} \mathbf{b}}. \tag{32}
\]

4.2. The CM Step

Given the current estimate \( \hat{\theta} = \hat{\theta}^{(k)} \) at the \( k \)th stage, the CM-step of the ECM algorithm consists of the conditional maximization of the \( Q \) function given in (24). More precisely, ECM replaces each M-step of the EM algorithm by a sequence of \( S \) conditional maximization steps, called CM-steps, each of which maximizes the \( Q \) function over \( \theta \) but with some vector function of \( \theta \), \((g_1(\theta), \ldots, g_S(\theta))\) say, fixed at its previous value. In our case, for example, we first maximize conditionally the function \( Q_1(\alpha, \beta, \omega | \theta^{(k)}) \) in (25) over \( \alpha \) fixing the values \( \beta = \hat{\beta}^{(k)} \) and \( \omega = \hat{\omega}^{(k)} \). Then we maximize \( Q_1(\alpha, \beta, \omega | \theta^{(k)}) \) over \( \beta \) fixing the values \( \alpha = \hat{\alpha}^{(k+1)} \) and \( \omega = \hat{\omega}^{(k)} \) and so on. We get the following closed expressions:

\[
\hat{\alpha}^{(k+1)} = \zhat_u - \pi_u \hat{\beta}^{(k)},
\]
\[
\hat{\beta}^{(k+1)} = n \hat{u}_i \sum_{i=1}^n \hat{u} \hat{z}_i \hat{u}_i^{(k)} - \sum_{i=1}^n \hat{u} \hat{z}_i \hat{u}_i \sum_{i=1}^n \hat{u} \hat{z}_i \hat{u}_i^{(k)} - \sum_{i=1}^n \hat{u} \hat{z}_i \hat{u}_i^{(k)} (\sum_{i=1}^n \hat{u} \hat{z}_i \hat{u}_i)^2),
\]
\[
\hat{\omega}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n (\hat{u} \hat{z}_i \hat{u}_i^{(k)} - 2 \hat{u} \hat{z}_i \hat{u}_i \hat{u}_i^{(k)} + \hat{u} \hat{z}_i \hat{u}_i^{(k)}),
\]
\[
\hat{\omega}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \left( \hat{u} \hat{z}_i \hat{u}_i^{(k)} - 2 \hat{u} \hat{z}_i \hat{u}_i \hat{u}_i^{(k)} + \hat{u} \hat{z}_i \hat{u}_i^{(k)} \right),
\]
\[
\hat{\omega}^{(k+1)} = \pi_u \hat{\beta}^{(k)},
\]
\[
\hat{\omega}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n (\hat{u} \hat{z}_i \hat{u}_i^{(k)} - 2 \hat{u} \hat{z}_i \hat{u}_i \hat{u}_i^{(k)} + \hat{u} \hat{z}_i \hat{u}_i^{(k)}),
\]
\[
\hat{\omega}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n (\hat{u} \hat{z}_i \hat{u}_i^{(k)} - 2 \hat{u} \hat{z}_i \hat{u}_i \hat{u}_i^{(k)} + \hat{u} \hat{z}_i \hat{u}_i^{(k)}),
\]

where \( \hat{z}_u^{(k)} = \sum_{i=1}^n \hat{u} \hat{z}_i \hat{u}_i^{(k)}, \pi_u^{(k)} = \sum_{i=1}^n \hat{u}_i \hat{z}_i \hat{u}_i^{(k)} \) and \( \hat{u}_i^{(k)} = \frac{1}{n} \sum_{i=1}^n \hat{u}_i \hat{u}_i^{(k)} \), with \( \hat{u} \hat{z}_i \hat{u}_i^{(k)} = (\hat{u} \hat{z}_i \hat{u}_i, \ldots, \hat{u} \hat{z}_i \hat{u}_i)^\top \)

4.3. Practical issues from computational aspects

4.3.1. Initialization

As other optimization techniques, the ECM algorithm may suffer from convergent difficulties such as singularity of component covariance matrices (for example in the case of mixed effects models) or undetermined local maximum. To deal with these situations, we consider to try many different initial values, selecting the solution that provides the highest value of the likelihood function. In our case, and for obtaining different sets of initial values, we performed multiple simple regression model using the R function \( \text{lm}() \).
4.3.2. Determining degrees of freedom

As was mentioned above, Lucas (1997) (see also Lange et al., 1989) showed that the protection against outliers is preserved only if the degrees of freedom parameter are fixed. Consequently, we assume the degrees of freedom for the Student-t to be fixed, using a model selection procedure based on the profile log-likelihood to choose the most appropriate value of \( \nu \).

4.4. Imputation of censored components

Let \( \mathbf{Z}_i^{(c)} \) be the true (partially or completely unobserved) response vector for the censored components of the \( i \)th unit. We define a predictor for \( \mathbf{Z}_i^{(c)} \) as

\[
\tilde{\mathbf{Z}}_i^{(c)} = E[\mathbf{Z}_i | \mathbf{V}_i = \mathbf{v}_i].
\]

We have the following particular cases:

1. If unit \( i \) has only censored components then, if we make \( r = 0 \) and \( k = 1 \) in Proposition 2, we get

\[
\tilde{\mathbf{Z}}_i^{(c)} = E[\mathbf{Y}_i], \quad \text{with} \quad \mathbf{Y}_i \sim T_p(\tilde{\mu}_z, \tilde{\Sigma}_z, \nu; \mathcal{D}_i),
\]

where \( \tilde{\mu}_z \) and \( \tilde{\Sigma}_z \) are the EM estimates of \( \mu_z \) and \( \Sigma_z \), respectively, and \( \mathcal{D}_i \) is as in (3) with \( \mathbf{d} = \mathbf{k}_i \), where \( \mathbf{k}_i \) is the vector with censoring levels for unit \( i \).

2. Unit \( i \) has uncensored and censored components. In this case, we partition the vector \( \mathbf{Z}_i \) as \( \mathbf{Z}_i = \text{vec}(\mathbf{Z}_i^o, \mathbf{Z}_i^c) \). Components are censored if and only if \( \mathbf{Z}_i^c \leq \mathbf{\kappa}_i^c \), such that

\[
\tilde{\mathbf{Z}}_i^{(c)} = E[\text{vec}(\mathbf{Z}_i^o, \mathbf{Z}_i^c) | \mathbf{Z}_i^o = \mathbf{z}_i^o, \mathbf{Z}_i^c \leq \mathbf{\kappa}_i^c] = \text{vec}(\mathbf{z}_i^o, \tilde{\mathbf{S}}_i^c),
\]

where, by Proposition 3 with \( r = 0 \) and \( k = 1 \),

\[
\tilde{\mathbf{y}}_i^c = E[\mathbf{Y}_i], \quad \text{with} \quad \mathbf{Y}_i \sim T_p(\mu_{z^c}, \mathbf{S}_{z^c}^c, \nu + p^c; \mathcal{D}_i^c),
\]

where \( \mu_{z^c} \) and \( \mathbf{S}_{z^c}^c \) are given in (17) and (18), respectively, and \( \mathcal{D}_i^c \) is given in (27).

4.5. Estimation of \( x_i \)

Following Lin & Lee (2006), Ho et al. (2012) and recently Castro et al. (2015), we consider the conditional mean to estimate the unobserved latent covariate. Using the tower property and Proposition 4, we have that an estimator for \( x_i \) can be obtained through

\[
\tilde{x}_i = E[x_i | \mathbf{V}_i] = E[E(x_i | U_i, \mathbf{Z}_i) | \mathbf{V}_i] = E \left[ \frac{\mu_x + \sigma_x^2 \mathbf{b}' \Omega^{-1} (\mathbf{Z}_i - \mathbf{a})}{1 + \sigma_x^2 \mathbf{b}' \Omega^{-1} \mathbf{b}} | \mathbf{V}_i \right] = \mu_x + \varphi(\tilde{\mathbf{Z}}_i - \mathbf{a} - \mathbf{b}\mu_x),
\]

where \( \varphi \) is given in (31) and \( \tilde{\mathbf{Z}}_i = E[\mathbf{Z}_i | \mathbf{V}_i] \). Observe that, if individual \( i \) does not have censored components, then \( \tilde{\mathbf{Z}}_i \) is the first moment of a \( t_p(\mathbf{\mu}_z, \mathbf{\Sigma}_z, \nu) \) distribution. If all its components are censored, then \( E[\mathbf{Z}_i | \mathbf{V}_i] = E[\mathbf{Z}_i | \mathbf{Z}_i \leq \mathbf{\kappa}_i] \), which can be computed using Proposition 2 with \( r = 0 \) and \( k = 1 \). Finally, if it has censored and uncensored components, then \( E[\mathbf{Z}_i | \mathbf{V}_i] = \text{vec}(\mathbf{Z}_i^o, \tilde{\mathbf{y}}_i^c) \), see (33). The parameter values in (34) must be replaced with the respective EM estimates.

Moreover, the conditional covariance matrix of \( x_i \) given \( \mathbf{V}_i \) is

\[
\text{Var}[x_i | \mathbf{V}_i] = E[x_i^2 | \mathbf{V}_i] - (E[x_i | \mathbf{V}_i])^2.
\]

By the tower property and Proposition 4, we have

\[
E[x_i^2 | \mathbf{V}_i] = E\{E(E(x_i^2 | U_i, \mathbf{Z}_i) | \mathbf{Z}_i) | \mathbf{V}_i\} = E \left( E \left[ \frac{\sigma_x^2}{U_i(1 + \sigma_x^2 \mathbf{b}' \Omega^{-1} \mathbf{b})} | \mathbf{Z}_i \right] | \mathbf{V}_i \right) + E \left[ \left( \frac{\mu_x + \sigma_x^2 \mathbf{b}' \Omega^{-1} (\mathbf{Z}_i - \mathbf{a})}{1 + \sigma_x^2 \mathbf{b}' \Omega^{-1} \mathbf{b}} \right)^2 | \mathbf{V}_i \right].
\]
It is easy to show that \( E[U_i^{-1} | Z_i] = (d_{\Sigma_i}(Z_i, \mu_z) + \nu)/(p + \nu - 2) \) − recall that \( U_i | Z_i = z_i \sim \text{Gamma}(\frac{p+\nu}{2}, \frac{1}{2}(d_{\Sigma_i}(Z_i, \mu_z) + \nu)) \), see the result after Proposition 5. After some lengthy algebra, we can prove that
\[
\text{Var}[x_i | V_i] = \Lambda E \left[ \frac{d_{\Sigma_i}(Z_i, \mu_z) + \nu}{p + \nu - 2} | V_i \right] + \Lambda^2 b^T \Omega^{-1} \text{Var}[Z_i | V_i] \Omega^{-1} b, \tag{35}
\]
where \( \Lambda \) is given in (32) and
\[
\text{Var}[Z_i | V_i] = E[Z_i Z_i^T | V_i] - E[Z_i | V_i] E[Z_i | V_i]^T.
\]

If individual \( i \) has only uncensored components, then expression (35) can be computed using the moments of the \( t_p(\mu_z, \Sigma_z, \nu) \) distribution using Proposition 2: it is enough to make \( d_1 = \cdots = d_p = +\infty, r = 1 \) and \( k = 0 \) to obtain the first expectation in (35) and \( r = 0, k = 1 \) (\( k = 2 \)) to obtain the other one. If the components are all censored, we again use Proposition 2, but now considering the moments of a \( T_{r_p}(\mu, \Sigma, \nu; D_i) \) distribution. Finally, if there are censored and uncensored components, then the expectations can be computed through Proposition 3, using the partition \( Z_i = \text{vec}(Z_i^\xi, Z_i^c) \). Besides this, the parameter values in (35) must be replaced with the respective EM estimates.

5. The observed information matrix

Under some general regularity conditions, we follow Lin (2010) to provide an information-based method to obtain the asymptotic covariance of ML estimates of the t-MEC model’s parameters. As defined by Meilijson (1989), the empirical information matrix can be computed as
\[
I_\epsilon(\theta | Z) = \sum_{i=1}^n s(Z_i | \theta) s^T(Z_i | \theta) - \frac{1}{n} S(Z_i | \theta) S^T(Z_i | \theta),
\]
where \( S(Z_i | \theta) = \sum_{i=1}^n s(Z_i | \theta) \) and \( s(Z_i | \theta) \) is the empirical score function for the \( i \)th unit. According to Louis (1982) it is possible to relate the score function of the incomplete data log-likelihood with the conditional expectation of the complete data log-likelihood function. Therefore, the individual score can be determined as
\[
s(Z_i | \theta) = \frac{\partial \log f(Z_i | \theta)}{\partial \theta} = E \left[ \frac{\partial \ell_\text{ic}(\theta | Z_i^c)}{\partial \theta} | V_i, C_i, \theta \right], \tag{36}
\]
where \( \ell_\text{ic}(\theta | Z_i^c) \) is the complete data log-likelihood formed from the single observation \( Z_i, i = 1, \ldots, n \). Using the EM estimates \( \hat{\theta}, S(Z_i | \hat{\theta}) = 0 \), and then (36) is given by
\[
I_\epsilon(\hat{\theta} | Z) = \sum_{i=1}^n \hat{s}_i \hat{s}_i^T, \tag{37}
\]
where \( \hat{s}_i = (\hat{s}_{i,\alpha}, \hat{s}_{i,\beta}, \hat{s}_{i,\omega}, \hat{s}_{i,\mu_x}, \hat{s}_{i,\sigma_x^2})^T \) is a 3p-dimensional vector, with components given by
\[
\begin{align*}
\hat{s}_{i,\alpha} &= (\hat{s}_{i,\alpha_1}, \ldots, \hat{s}_{i,\alpha_r})^T = I_{(p-1)}(\hat{u}_z - \hat{u}_x \hat{a} - \hat{u}_x \hat{b}),
\hat{s}_{i,\beta} &= (\hat{s}_{i,\beta_1}, \ldots, \hat{s}_{i,\beta_k})^T = I_{(p-1)}(\hat{u}_x \hat{z}_i - \hat{u}_x \hat{a} - \hat{u}_x \hat{b}),
\hat{s}_{i,\omega} &= (\hat{s}_{i,\omega_1}, \ldots, \hat{s}_{i,\omega_k})^T = -\frac{1}{2} \Omega^{-1} 1_p + \frac{1}{2} \Omega^{-2} \text{diag}(\hat{a}_i),
\hat{s}_{i,\mu_x} &= \frac{1}{\sigma_x^2}(\hat{u}_x \hat{z}_i - \hat{u}_x \hat{a} \hat{p}_x),
\hat{s}_{i,\sigma_x^2} &= -\frac{1}{2\sigma_x^4} + \frac{1}{2\sigma_x^4}(\hat{u}_x \hat{z}_i - 2 \hat{u}_x \hat{a} \hat{p}_x + \hat{a}_i^2),
\end{align*}
\]
with \( I_{(p)} = [0, I_{p-1}]_{(p-1) \times p}, 1_p = (1, \ldots, 1)^T_{p \times 1} \) and \( \hat{a}_i = \hat{u}_x \hat{z}_i - 2 \hat{u}_x \hat{a} \hat{p}_x + \hat{a}_i \hat{p}_x^2 + \hat{a}_i \hat{a} \hat{p}_x + \hat{u}_x \hat{a} \hat{p}_x^2 + \hat{u}_x \hat{a} \hat{a} = \hat{u}_x \hat{a} \hat{p}_x^2 + \hat{u}_x \hat{a} \hat{a} + \hat{u}_x \hat{a} \hat{a} \hat{p}_x + \hat{u}_x \hat{a} \hat{a} \hat{p}_x^2 \).
6. AIDS A5055 clinical trial

We illustrate the proposed method with a HIV dataset coming from the AIDS A5055 clinical trial (see for details Wang, 2013). This trial involved a total of 44 infected patients with the HIV type 1 (HIV-1).

In AIDS research, the number of ribonucleic acid (RNA) copies (viral load) in blood plasma and its evolutionary trajectories play a prominent role in the diagnosis of HIV-1 disease progression after an ARV treatment regimen (see for example Paxton et al., 1997). Furthermore, the T helper cells (known as CD4⁺) is another immunologic marker frequently used to monitor the progression of the disease.

The dataset analyzed here consists of plasma viral load measurements (in copies per milliliter) and CD4⁺ cell counts measured roughly at 5 different days of follow-up for each patient. For practical reasons, we considered those patients with more than 5 observations, so only 41 patients were analyzed.

In this study, we focus on investigating the longitudinal trajectories for RNA viral load (in log-base-10 scale), denoted by \( \log_{10} \text{RNA} \) and the baseline of CD4⁺ cell counts. In this dataset, the lower detection limit for RNA viral load is 50 copies/milliliter, and therefore 23.9% (49 out of 205) of measurements lie below the limits of assay quantification (left-censored). Figure 1 shows the trajectories of the immunologic response.

![Figure 1: Trajectories of \( \log_{10} \text{RNA} \) for 41 HIV-1 infected patients, dotted line(red) indicates the censoring level (left panel). Plot of the profile log-likelihood of the degrees of freedom \( \nu \) (right panel).](image)

In the spirit of Jiang et al. (2012), and since it is believed that the \( \log_{10} \text{RNA} \) and CD4⁺ cell counts are negatively correlated during treatment, we model the relation between the viral load and the baseline of CD4⁺ using (6)–(7). Let \( \mathbf{Y}_i \) be the vector of viral load and let \( x_i \) be the baseline of CD4⁺ cell count for subject \( i \) measured with error. Then, we propose the following t-MEC model:

\[
X_i = x_i + \xi_i \quad \text{and} \quad \mathbf{Y}_i = \alpha + \beta x_i + \mathbf{e}_i,
\]

with \( i = 1, \ldots, 41 \) and \( p = 5 \). For the t-MEC model, we assumed that the degrees of freedom \( \nu \) is known and we found \( \nu = 4 \) (see Figure 1-left panel), indicating that the normal model is inadequate. In order to compare our approach with a natural competitor, we consider the N-MEC model. The MLE for the parameters of both models, as well as their corresponding standard errors (SE), obtained via the empirical information matrix, are reported in Table 1. This table shows that the estimates of all parameters under the t-MEC and N-MEC models are close. However, the standard errors (SE) of the t-MEC are smaller than those under the N-MEC model, indicating that our proposed model seem to produce more precise estimates.
Table 1: A5055 study. ML and SE for parameter estimates.

<table>
<thead>
<tr>
<th></th>
<th>t-MEC</th>
<th>SE</th>
<th>N-MEC</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>5.8747</td>
<td>1.1656</td>
<td>5.7319</td>
<td>1.9297</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>4.4999</td>
<td>1.0408</td>
<td>5.1879</td>
<td>2.5959</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>4.1651</td>
<td>1.0766</td>
<td>5.0574</td>
<td>2.5451</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>4.0977</td>
<td>1.2586</td>
<td>4.6351</td>
<td>3.2461</td>
</tr>
<tr>
<td>$\alpha_5$</td>
<td>4.0438</td>
<td>1.5317</td>
<td>4.5349</td>
<td>3.3353</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.7129</td>
<td>0.3561</td>
<td>-0.6492</td>
<td>0.5444</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.6831</td>
<td>0.3160</td>
<td>-0.8482</td>
<td>0.7386</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.6897</td>
<td>0.3155</td>
<td>-0.9222</td>
<td>0.7549</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>-0.7736</td>
<td>0.4113</td>
<td>-0.9012</td>
<td>0.9708</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>-0.8679</td>
<td>0.5139</td>
<td>-0.9590</td>
<td>1.0371</td>
</tr>
<tr>
<td>$\mu_x$</td>
<td>2.8694</td>
<td>0.3473</td>
<td>3.0690</td>
<td>0.6008</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.5661</td>
<td>0.5989</td>
<td>0.6718</td>
<td>1.3098</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>1.4655</td>
<td>0.4632</td>
<td>2.5416</td>
<td>0.7429</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.1769</td>
<td>0.0848</td>
<td>0.2580</td>
<td>0.1353</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.0820</td>
<td>0.0528</td>
<td>0.1209</td>
<td>0.1000</td>
</tr>
<tr>
<td>$\phi_5$</td>
<td>0.1691</td>
<td>0.0482</td>
<td>0.2395</td>
<td>0.0862</td>
</tr>
<tr>
<td>$\phi_6$</td>
<td>0.2312</td>
<td>0.1097</td>
<td>0.4700</td>
<td>0.2280</td>
</tr>
</tbody>
</table>

Table 2 compares the fit of the two models using the Akaike information criterion (AIC) and Schwarz information criterion (BIC). Note that, as expected, the t-MEC model outperform the normal one.

<table>
<thead>
<tr>
<th></th>
<th>t-MEC</th>
<th>N-MEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-249.7508</td>
<td>-266.2059</td>
</tr>
<tr>
<td>AIC</td>
<td>535.5016</td>
<td>568.4117</td>
</tr>
<tr>
<td>BIC</td>
<td>598.5976</td>
<td>631.5077</td>
</tr>
</tbody>
</table>

We also have considered a likelihood ratio test (LRT) for the null hypotheses in favor of Gaussian assumption. The resulting LRT statistic is 32.91 with $p$-value less than 0.0001 (highly significant), suggesting that the t-MEC model is more appropriate than the normal one.

7. Simulation studies

In order to study the performance of our proposed method, we present three simulation studies. The first one shows the asymptotic behavior of the EM estimates for the proposed model. The second one investigates the consequences on parameter inference when the normality assumption is inappropriate. Finally, the third one is designed to investigate the effect of including the censoring component in the model.

7.1. Asymptotic properties

In this simulation study, we analyze the absolute bias (BIAS) and mean square error (MSE) of the regression coefficient estimates obtained from the t-MEC model for six different sample sizes $n$, namely 50, 100, 200, 300, 400 and 600. These measures are defined by

$$\text{BIAS}_k = \frac{1}{M} \sum_{j=1}^{M} |\hat{\theta}_k^{(j)} - \theta_k|$$

and

$$\text{MSE}_k = \frac{1}{M} \sum_{j=1}^{M} (\hat{\theta}_k^{(j)} - \theta_k)^2,$$

where $\hat{\theta}_k^{(j)}$ is the EM estimate of the parameter $\theta_k$, $k = 1, \ldots, 3p$, for the $j$-th sample. The key idea of this simulation is to provide empirical evidence about consistency of the EM estimators.
under the proposed t-MEC model. For each sample size, we generate $M = 100$ datasets with 10% censoring proportion. Using the ECM algorithm, the absolute bias and mean squared error for each parameter over the 100 datasets were computed. The parameter setup (see Section 3), is

$$\alpha = (3, 2, 1, 2)^\top, \quad \beta = (1.5, 1, 1.5, 1)^\top, \quad \mu_x = 4, \quad \sigma^2_x = 2 \quad \text{and} \quad \Omega = \text{diag}(0.5, 0.5, 0.5, 0.5). \quad (39)$$

The degrees of freedom were fixed at the value $\nu = 5$.

The results are presented in Figure 2. From this figure we can see that the MSE tends to zero as the sample size increases. Similar results were obtained after the analysis of the absolute bias (BIAS) as can be seen from Figure 5 in the Appendix. As expected, the proposed ECM algorithm provides ML estimates with good asymptotic properties for the t-MEC model.

7.2. Parameter inference

In this study we investigate the consequences on parameter inference when the normality assumption is inappropriate, as well the ability of some model choice criteria (AIC and BIC) to select the correct model. In addition, we study the effect of different censoring proportions on the EM estimates. For this purpose, we consider a heavy-tail distribution for the random errors. In this context, we generate $M = 100$ datasets coming from a slash distribution with parameter $\nu = 1.5$.
and censoring proportions 0%, 10%, 20% and 30%. The slash distribution arises when we change the distribution of $U$ in (2) to $U \sim \text{Beta}(\nu, 1)$, with $f(u|\nu) = \nu u^{\nu - 1}$, $u \in (0, 1)$, and $\nu > 0$. See Wang & Genton (2006) for details. The parameter values are set as in the previous experimental study.

Table 3: Simulation study 7.2. Summary statistics based on 100 simulated samples from the slash distribution for different levels of censoring (0%, 10%, 20%, 30%).

<table>
<thead>
<tr>
<th>Censoring</th>
<th>Simulated data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_1$</td>
</tr>
<tr>
<td>0%</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>IM SE</td>
</tr>
<tr>
<td></td>
<td>MC CP</td>
</tr>
<tr>
<td></td>
<td>Student-t</td>
</tr>
<tr>
<td></td>
<td>IM SE</td>
</tr>
<tr>
<td></td>
<td>MC CP</td>
</tr>
<tr>
<td>10%</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>IM SE</td>
</tr>
<tr>
<td></td>
<td>MC CP</td>
</tr>
<tr>
<td></td>
<td>Student-t</td>
</tr>
<tr>
<td></td>
<td>IM SE</td>
</tr>
<tr>
<td></td>
<td>MC CP</td>
</tr>
<tr>
<td>20%</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>IM SE</td>
</tr>
<tr>
<td></td>
<td>MC CP</td>
</tr>
<tr>
<td></td>
<td>Student-t</td>
</tr>
<tr>
<td></td>
<td>IM SE</td>
</tr>
<tr>
<td></td>
<td>MC CP</td>
</tr>
<tr>
<td>30%</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>IM SE</td>
</tr>
<tr>
<td></td>
<td>MC CP</td>
</tr>
<tr>
<td></td>
<td>Student-t</td>
</tr>
<tr>
<td></td>
<td>IM SE</td>
</tr>
<tr>
<td></td>
<td>MC CP</td>
</tr>
</tbody>
</table>

For each simulated dataset we fitted the t-MEC (with $\nu = 5$ degrees of freedom) and the N-MEC models. The model selection criteria AIC and BIC as well as the estimates of the model parameters were recorded at each simulation. Summary statistics such as the Monte Carlo mean estimate (MC mean), coverage probability (MC CP) and the approximate standard error obtained through the information-based method (IM SE), discussed in Section 5, for the parameter estimates are presented in Table 3.

From these results we can observe that for all considered levels of censoring, the t-MEC model is chosen as the correct model. Under the t-MEC model, the MC CP for $\alpha$ and $\beta$ are stable, but the MC CP of $\mu_\nu$ is lower than the nominal level (95%). In general, the MC CP values are higher than those obtained under the normal model. Figure 3 shows the MSE for some parameter estimates (the biases are presented in Figure 6 in the Appendix). Note that, the MSE under the t-MEC model is lower than the obtained under the normal, for different levels of censoring.

Regarding the model choice, the t-MEC model was chosen as the best by the two criteria for all samples.

7.3. Censored model

In this section, the main goal is to study the effect of taking into account censored data on the parameter estimates. We generated $M = 100$ samples from the t-MEC model with $\nu = 5$, setting the censoring level at 20%. The other parameter values are set as in (39). For each dataset, we fitted two models: in case 1 we use a naive model, where censored responses are not taken into account. In case 2 we fit a t-MEC model.

Figure 4 shows the box plots corresponding to each parameter estimate considering the $M = 100$ datasets. Note that, the estimates in case 2 are, in general, more precise than those obtained in case 1. It is also possible to note that, in case 2, the variability observed in the estimations is smaller than in case 1, except for some dispersion parameters. We point out that it is important to consider the effect of censoring in data modeling, avoiding ad-hoc methods.

8. Conclusions

In this paper, we introduce the multivariate ME model with censored responses based on the Student-t distribution, the so-called t-MEC model. This model considers the possibility of censor-
ing in the surrogate covariate and the response. Moreover, we assume that the latent unobserved covariate and random observational errors follow a multivariate Student-t distribution, which provides a robust alternative to the usual Gaussian model. For the parameter estimation, an ECM algorithm based on some statistical properties of the multivariate truncated Student-t distribution is developed to obtain ML estimates. Some simulation studies revealed that our proposed method generates less biased estimates of model parameters than the case when the censoring scheme is not taken into account. Moreover, we showed that the use of the Student-t distribution generates better results than the normal one, in the context of the censored ME models for HIV data.

Of course, further extensions of the current work are possible. For example, the proposed method can be naturally extended by considering the family of scale mixtures of normal (SMN) and skew-normal (SMSN) distributions. An efficient estimation procedure to obtain ML estimates of model parameters can be implemented by using a stochastic approximation of the traditional EM (SAEM) algorithm. Other extensions include, a Bayesian treatment via Markov chain Monte Carlo (MCMC) sampling methods in the context of SMN-MEC and SMSN-MEC models (Lachos et al., 2010).

Acknowledgements

The authors are grateful to the Editor, Associate Editor and the referees for their helpful comments on an earlier version of this paper. L. A. Matos acknowledges support from FAPESP-Brazil (Grants 2011/22063-9 and 2015/05385-3). V.H. Lachos and L. A. Matos acknowledges support from FAPESP-Brazil (Grant 2018/05013-7). L.M. Castro acknowledges support from Grant FONDECYT 1170258 from the Chilean government and Iniciativa Científica Milenio - Minecon Núcleo Milenio MiDaS. This paper was written while Celso R. B. Cabral was a visiting professor in the Department of Statistics at the University of Campinas, Brazil. Celso R. B. Cabral acknowledges the support from FAPEAM (Research Support Program - Universal Amazonas), FAPESP (Grant 2015/20922-5) and CNPq-Brazil (Grant 447964/2014-3).
Figure 4: Simulation study 7.3. Boxplots of the parameter estimates. Dotted lines indicate the true parameter value.

References


Appendix

Figure 5: Simulation 7.1. Bias of parameter estimates under the t-MEC model considering 10% of censoring.
Figure 6: Simulation study 7.2. Bias of $\beta$, $\alpha$, $\mu_x$, $\sigma^2_x$ estimates under normal and Student-t models for different levels of censoring (0%, 10%, 20%, 30%).