Stochastic Models in Probability and Statistics
Vol. 1, No. 2, 2024, pp. 211–228. Research Paper

A two-piece scale mixture normal measurement error models for replicated data

Fereshteh Kahrari*

Department of Statistical Sciences, University of Padova, Italy
Email(s): fereshteh.kahrari@unipd.it

Abstract. We develop a new class of flexible replicated measurement error models (RMEM) based on the normal two-piece scale mixture (TP-SMN) family to model the distribution of the latent variable. In the proposed approach, the replicated observations are jointly modeled by a mixture of two components from a scale mixture skew-normal (SMSN) density. The flexibility of this class can enable the simultaneous accommodation of skewness, outliers, and multimodality. The proposed connection between the unobserved covariates and the response facilitates the construction of an EM-type algorithm to perform maximum likelihood estimation. The effectiveness of the maximum likelihood estimations is studied through the simulation studies. Also, the method is applied to analyze continuing survey data on food intake by individuals on diet habits.

Keywords: ECM algorithm; Equation error; Replicated measurement error model; Two-piece scale mixture normal.

1 Introduction

A broad coverage of the measurement error model (MEM) has been extensively studied in the literature; see, for example, Carroll et al. (2006), Cheng and Van Ness (1999), Fuller (1987), Gustafson (2004) and Buonaccorsi (2010). These studies assumed that the distribution of the random errors and the unobserved covariate is Gaussian, which is sometimes not feasible as the model is sensitive to skewness, outliers, and unobserved heterogeneity in the data. In the context of MEMs, this can lead to some particular problems for these types of models. First, a non-identifiability problem in normal MEMs (Reiersol (1950)) can occur because one can not establish a single relationship between the parameters of the jointly distribution of the observations and the parameters of the model, consequently, no consistent estimator of the key

*Corresponding author

Received: 15 October 2024 / Accepted: 7 December 2024
DOI: 10.22067/smps.2024.46110

 https://smps.um.ac.ir

parameter, i.e., slope parameter, exists. To avoid this difficulty we may use prior knowledge of the error variances or make some assumptions on the error variances in advance, while such prior knowledge/assumptions are usually not easy to justify. Fortunately, the non-identifiability problem can be overcome in the case of replicated measurement error models (RMEMs) as the error variances can be estimated separately or together with other parameters. In this case, we refer the reader to Chan and Mak (1979), Isogawa (1985), and Lin et al. (2004). Second, recently, Lin and Cao (2013) considered a new replicated structural MEM in which the replicated observations jointly follow scale mixtures of normal (SMN) distributions. The scale mixture normal replicated measurement error model (SMN-RMEM) provides an appealing robust alternative to the usual model based on normal distributions, which also takes advantage of SMN distributions to accommodate extreme and outlying observations. Cao et al. (2015) considered models containing both error-prone covariates and predictors measured without errors, under multivariate SMN-RMEM, providing appealing robust and adaptable alternatives to the usual Gaussian assumptions. Third, the normality assumption and even SMN distributions, however, may be violated when a data set contains asymmetric outcomes. Cao et al. (2018) developed a new RMEM in the class of scale mixtures of skew-normal (SMSN) distributions. There can still be problems related to the simultaneous occurrence of skewness, discrepant observations, and multimodality.

Unlike other constructions in RMEMs, our model is formulated by replacing the normal assumption of the classical formulations by a more flexible class of distributions, called two-piece scale mixture normal (TP-SMN) distributions for one of the components, while retaining SMN for the others. Specific distributions in this family, including univariate versions of two-piece normal (TP-N), two-piece t (TP-T) with ν degrees of freedom, two-piece slash (TP-SL) and two-piece contaminated normal (TP-CN), are examined for the unobserved value of the covariates. In this approach, latent covariates and random observational errors are jointly modeled by a two-component mixture of SMSN densities. However, we view our construction as based on the TP-SMN assumption, as highlighted by the title, while the two-component SMSN likelihood is a mathematical implication of the assumption. Moreover, unlike a similar model previously proposed by Cao et al. (2018), the SMSN and TP-SMN families have different properties (for example, different tail behavior). In addition, numerical stability might also be a property that could be used to motivate our proposed model for the explanatory variable. This means that estimating the skewness parameter in the TP-SMN class of distributions is often easier, and more stable than estimating the shape parameter in the SMSN family (which controls the asymmetry, tails, location of the mode, and spread of the density).

The rest of this paper is organized as follows. Section 2 gives a brief description of the TP-SMN and SMSN distributions. Section 3, the replicated structural measurement error model with TP-SMN distributions (TP-SMN-RMEM) is defined, Proposition 3.1 represents the main result for this section. In Section 4, the advantages in terms of efficiency and robustness of MLEs of model (4) based on the ECM technique studied by Meng and Rubin (1993) are considered. A closed form expression is also obtained for the asymptotic covariance matrix of the ML estimators. In Section 5, the performance of the model and the importance of equation error are examined via simulation studies. Section 6 applies the model to analyze the inner relationship between saturated fat and caloric intake in CSFII data. Some conclusions are given in Section 7.

2 Asymmetric heavy-tailed distributions

2.1 Two-piece scale mixture normal (TP-SMN) distributions

Following Andrews and Mallows (1974) the pdf of random variable $X \sim SMN(\mu,\sigma,\boldsymbol{\nu})$ is denoted as,

$$f_{SMN}(x;\mu,\sigma,\boldsymbol{\nu}) = \int_0^{\infty} \phi\left(x;\mu,\kappa(u)\sigma^2\right) dH(u;\boldsymbol{\nu}), \quad y \in \mathbb{R},$$

where $\phi(.;\mu,\sigma^2)$ represents the pdf of $N(\mu,\sigma^2)$ distribution, $H(.;\boldsymbol{\nu})$ is the cdf of the scale mixing random variable U which is indexed by parameter $\boldsymbol{\nu}$. Also, $X \sim SMN(\mu,\sigma,\boldsymbol{\nu})$ has the stochastic representation given by

$$X \stackrel{d}{=} \mu + \sigma\kappa^{1/2}(U)W, \quad x \in \mathbf{R},$$

where W follows the standard normal distribution and is assumed to be independent of U.

The TP-SMN family is an analogy and alternative to the SMSN family, which contains the light/heavy-tailed and symmetry/asymmetry members including the TP-N, TP-t, TP-SL, and TP-CN distributions.

Definition 1. *Following a general two-piece distribution from Arellano-Valle et al. (2005), the pdf of random variable $Y \sim TP-SMN(\mu,\sigma,\gamma,\boldsymbol{\nu})$ for $y \in \mathbb{R}$ can be defined as,*

$$f_{TP-SMN}(y;\mu,\sigma,\gamma,\boldsymbol{\nu}) = \begin{cases} 2(1-\gamma) f_{SMN}(y;\mu,\sigma(1-\gamma),\boldsymbol{\nu}), & I_{(-\infty,\mu]}(y) \\ 2\gamma f_{SMN}(y;\mu,\sigma\gamma,\boldsymbol{\nu}), & I_{[\mu,\infty)}(y), \end{cases} \tag{1}$$

where $\gamma \in (0,1)$ is the slant parameter, $I_A(x)$ denotes the indicator function of the set A.

Maleki and Mahmoudi (2017) studied the MLE problem for the parameters of the TP-SMN family using an EM-type algorithm. Barkhordar et al. (2022) proposed and examined the performance of a Bayesian approach for a homoscedastic nonlinear regression (NLR) model assuming errors with TP-SMN distributions. Maleki et al. (2019c) and Hoseinzadeh et al. (2021) examined the performance of the TP-SMN family in the context of NLR models (TP-SMN-NLR) using an EM-type algorithm to obtain MLEs for the parameters. Zarei et al. (2022) developed a general class of robust mixture regression model based on two-piece scale mixtures of normal distributions (TP-SMN) . For more details of stochastic representations, statistical inferences, and applications of the TP-SMN family, see Maleki et al. (2019d), Moravveji et al. (2019), Arellano-Valle et al. (2020), and Ghasami et al. (2020).

Corollary 1. *Let $Y \sim TP-SMN(\mu,\sigma,\gamma,\boldsymbol{\nu})$, then Y has a stochastic representation given by,*

i. $Y \stackrel{d}{=} \mu + \sigma\kappa^{1/2}(U)W_\gamma V$, where V is a continuous random variable with density function $2\phi(v)I_{[0,\infty)}(v)$, the standard half-normal density, and W_γ is an independent discrete random variable with probability function

$$p(w;\gamma) = \gamma^{(1+s)/2}(1-\gamma)^{(1-s)/2} I_{\{-1,1\}}(s), \tag{2}$$

with $s = sign(w)$. Equivalently, if $Y \stackrel{d}{=} \mu + \sigma W_\gamma |X|$, where $X \sim SMN(0,1,\boldsymbol{\nu})$ and is independent of W_γ, then $Y \sim TP-SMN(\mu,\sigma,\gamma,\boldsymbol{\nu})$.

2.2 Examples of the TP-SMN distributions

In this section, we present some members of the TP-SMN family which are examined for the unobserved value of the covariates. We consider the case where $\kappa(u) = 1/u$.

Two-piece normal distribution Two-piece normal, TP-N(μ, σ, γ), distribution is obtained from Eq. (1) when $P(U = 1) = 1$.

$$f(y;\mu,\sigma,\gamma,\nu) = \begin{cases} 2(1-\gamma)\phi(y;\mu,\sigma^2(1-\gamma)^2), & I_{(-\infty,\mu]}(y), \\ 2\gamma\phi(y;\mu,\sigma^2\gamma^2), & I_{[\mu,\infty)}(y). \end{cases}$$

Arellano-Valle et al. (2020) studied the main properties of the TP-N distribution.

Two-piece t distribution The pdf of a two-piece t distribution with ν degree of freedom, TP-t$(\mu, \sigma, \gamma, \nu)$ say, is derived from Eq. (1) by taking U to be distributed as Gamma$(\nu/2, \nu/2)$, $\nu > 0$.

$$f(y;\mu,\sigma,\gamma,\nu) = \begin{cases} 2\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\sigma}\left(1+\frac{1}{\nu}(\frac{y-\mu}{\sigma(1-\gamma)})^2\right)^{-\frac{\nu+1}{2}}, & I_{(-\infty,\mu]}(y), \\ 2\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\sigma}\left(1+\frac{1}{\nu}(\frac{y-\mu}{\sigma\gamma})^2\right)^{-\frac{\nu+1}{2}}, & I_{[\mu,\infty)}(y). \end{cases}$$

Two-piece slash distribution Two-piece slash distribution, denoted by TP-SL$(\mu, \sigma, \gamma, \nu)$, arises when U has $Beta(\nu, 1)$ distribution. The TP-SL pdf is then given by

$$f(y;\mu,\sigma,\gamma,\nu) = \begin{cases} 2\nu(1-\gamma)\int_0^1 u^{\nu-1}\phi(y;\mu,u^{-1}\sigma^2(1-\gamma)^2)du, & I_{(-\infty,\mu]}(y), \\ 2\nu\gamma\int_0^1 u^{\nu-1}\phi(y;\mu,u^{-1}\sigma^2\gamma^2)du, & I_{[\mu,\infty)}(y). \end{cases}$$

Two-piece contaminated normal distribution Another member of the TP-SMN family is known as the two-piece contaminated normal distribution, and denoted by $TP-CN(\mu, \sigma, \gamma, \nu_1, \nu_2)$, arises when U is a discrete random variable with probability function

$$h(u;\nu_1,\nu_2) = \nu_1\mathbb{I}_{(u=\nu_2)} + (1-\nu_1)\mathbb{I}_{(u=1)}, \text{ where } 0 < \nu_i < 1 \text{ for } i = 1,2.$$

In this case, the random variable Y has pdf given by

$$f(y;\mu,\sigma,\gamma,\nu) = \begin{cases} 2(1-\gamma)(1-\nu_1)\phi(y;\mu,\sigma^2(1-\gamma)^2), & I_{(-\infty,\mu]}(y), \\ 2\gamma\nu_1\phi(y;\mu,\nu_2^{-1}\sigma^2\gamma^2), & I_{[\mu,\infty)}(y). \end{cases}$$

2.3 Scale mixture skew-normal distributions

Following Branco and Dey (2001) a random vector $\mathbf{Y}$ has a multivariate scale mixture skew-normal distribution (SMSN) with location vector $\boldsymbol{\mu}$, positive definite scale matrix $\boldsymbol{\Omega}$ and skewness/shape vector $\boldsymbol{\lambda}$, denoted by $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \boldsymbol{\nu})$, if its pdf is given by

$$f_{SMSN}(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda},\boldsymbol{\nu}) = \int_0^\infty 2\phi_p(\mathbf{y};\boldsymbol{\mu},\kappa(u)\boldsymbol{\Omega})\,\Phi\left(\kappa^{-1/2}(u)A\right)dH(u;\boldsymbol{\nu}),$$

where $A = \boldsymbol{\lambda}^\top \boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})$. The SMSN random vector $\mathbf{Y}$ can be introduced as the location-scale transformation, following the stochastic representation given by

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa^{1/2}(U)\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{X}, \tag{3}$$

where following Arellano-Valle and Genton (2005), the random vector $\mathbf{X}$ has a multivariate skew-normal (SN) distribution denoted by $\mathbf{X} \sim SN_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda})$ with the pdf given by

$$f_{SN}(\mathbf{x}; \mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda}) = 2\phi_p(\mathbf{x}; \mathbf{I}_p, \boldsymbol{\lambda})\Phi(\boldsymbol{\lambda}^\top \mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^p.$$

From (3) it follows straightforward that $\mathbf{Y}|U = u \sim SN_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Omega}, \boldsymbol{\lambda})$. Thus, using Eq.(3), it can be shown that the mean vector and variance-covariance matrix of $\mathbf{Y}$ are given, respectively, by

$$\begin{aligned} E(\mathbf{Y}) &= \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\mathrm{E}\left(\kappa^{\frac{1}{2}}(U)\right)\boldsymbol{\Delta}, \\ V(\mathbf{Y}) &= \mathrm{E}(\kappa(U))\boldsymbol{\Omega} - \frac{2}{\pi}\mathrm{E}\left(\kappa^{\frac{1}{2}}(U)\right)^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^\top. \end{aligned}$$

where $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$, and $\boldsymbol{\Delta} = \boldsymbol{\Omega}^{1/2}\boldsymbol{\delta}$.

3 The TP-SMN structural RMEM

The proposed model can be described below. Suppose that we observe (X_t, Y_t), as surrogates of true (latent) unobserved variables (x_t, y_t), plus additive measurement errors (δ_t, ξ_t), that is, they satisfy an incomplete linear relationship $y_t = \alpha + \beta x_t + e_t$, in which the equation error e_t means that in some situations, the true variables are not perfectly related if factors other than x_t are responsible for the variation in y_t. Note that the equation error does not need to exist, which means that $e_t = 0$. In this case, the model is known as a no-equation-error model. Suppose that x_t and y_t are observed p and q times (respectively) to make replicated observations $X_t^{(i)}$ and $Y_t^{(j)}$.

$$\begin{cases} X_t^{(i)} = x_t + \delta_t^{(i)}, & i = 1, \ldots, p, \\ Y_t^{(j)} = y_t + \xi_t^{(j)}, & j = 1, \ldots, q, \\ y_t = \alpha + \beta x_t + e_t, & t = 1, \ldots, n. \end{cases} \tag{4}$$

By introducing the vectors $\mathbf{Z}_t = (\mathbf{X}_t^\top, \mathbf{Y}_t^\top)^\top$, where $\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(p)})^\top$ and $\mathbf{Y}_t = (Y_t^{(1)}, \ldots, Y_t^{(q)})^\top$, and $\boldsymbol{\epsilon}_t = (\delta_t^{(1)}, \ldots, \delta_t^{(p)}, e_t + \xi_t^{(1)}, \ldots, e_t + \xi_t^{(q)})^\top$, $\boldsymbol{a} = (\mathbf{0}_p^\top, \alpha\mathbf{1}_q^\top)^\top$, $\boldsymbol{b} = (\mathbf{1}_p^\top, \beta\mathbf{1}_q^\top)^\top$. Model (4) can be written compactly as

$$\mathbf{Z}_t = \boldsymbol{a} + \boldsymbol{b}x_t + \boldsymbol{\epsilon}_t. \tag{5}$$

Thus, from Eq. (5) the distribution of $\mathbf{Z}_t$ becomes specified once the joint distribution of $\mathbf{r}_t = (x_t, \boldsymbol{\epsilon}_t{}^\top)^\top$ is specified. To obtain robust estimation of the parameters, we propose to replace the normal assumption by

$$
\begin{aligned}
&U_t \overset{iid}{\sim} H(u;\boldsymbol{\nu}), \quad t=1,\ldots,n,\\
&x_t|U_t=u_t \overset{iid}{\sim} TP-N(\mu_x,u_t^{-1/2}\sigma_x,\gamma_x),\ i.e.,\ x_t \overset{iid}{\sim} TP-SMN(\mu_x,\sigma_x,\gamma_x,\boldsymbol{\nu}),\\
&\delta_t^{(i)}|U_t=u_t \overset{iid}{\sim} N(0,u_t^{-1}\phi_\delta),\ i.e.,\ \delta_t^{(i)} \overset{iid}{\sim} SMN(0,\phi_\delta,\boldsymbol{\nu}),\\
&\xi_t^{(j)}|U_t=u_t \overset{iid}{\sim} N(0,u_t^{-1}\phi_\xi),\ i.e.,\ \xi_t^{(j)} \overset{iid}{\sim} SMN(0,\phi_\xi,\boldsymbol{\nu}),\\
&e_t|U_t=u_t \overset{iid}{\sim} N(0,u_t^{-1}\phi_e),\ i.e.,\ e_t \overset{iid}{\sim} SMN(0,\phi_e,\boldsymbol{\nu}),
\end{aligned} \tag{6}
$$

which we call the two-piece scale mixture replicated measurement error model (TP-SMN-RMEM). From (5) and (6), since for each t, x_t and $\boldsymbol{\epsilon}_t$ are indexed by the same scale mixing factor U_t, they are not independent. However, x_t and $\boldsymbol{\epsilon}_t$ are conditionally independent, which implies $\mathrm{cov}(x_t,\boldsymbol{\epsilon}_t|U_t)=0$. Model (4) can be specified equivalently in the hierarchical form as

$$
\begin{aligned}
&\mathbf{Z}_t|x_t,U_t=u_t \overset{iid}{\sim} N_m(\boldsymbol{a}+\boldsymbol{b}x_t,u_t^{-1}\boldsymbol{\Sigma}),\\
&x_t|U_t=u_t \overset{iid}{\sim} TP-N(\mu_x,u_t^{-1/2}\sigma_x,\gamma_x),\\
&U_t \overset{iid}{\sim} H(u_t;\boldsymbol{\nu}),
\end{aligned} \tag{7}
$$

where $m=p+q$, $\boldsymbol{c}=(\mathbf{0}_p^\top,\mathbf{1}_q^\top)^\top$, $\mathbf{D}(.)$ denotes a diagonal matrix and $\boldsymbol{\Sigma}$ is a $m\times m$ block matrix as follows

$$
\boldsymbol{\Sigma}=\mathbf{D}(\boldsymbol{\phi})+\phi_e\mathbf{c}\mathbf{c}^\top=\begin{bmatrix}\boldsymbol{\Sigma}_\delta & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{e,\xi}\end{bmatrix}, \tag{8}
$$

with $\boldsymbol{\Sigma}_\delta=\phi_\delta\mathbf{I}_p$ and $\boldsymbol{\Sigma}_{e,\xi}=\phi_\xi\mathbf{I}_q+\phi_e\mathbf{1}_q\mathbf{1}_q{}^\top$.

The following proposition represents the main result for this section. The vector of parameters is denoted by $\boldsymbol{\theta}=(\alpha,\beta,\mu_x,\sigma_x,\gamma_x,\phi_e,\phi_\delta,\phi_\xi)^\top$.

Proposition 1. *Under the hierarchical representation defined by (7), the observed random vectors* $(\mathbf{z}_1,\ldots,\mathbf{z}_n)$ *are drawn independently from the common distribution given by*

$$
\begin{aligned}
f(\mathbf{z}_t;\boldsymbol{\theta}) &= \gamma_x f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x},\boldsymbol{\nu})\\
&\quad +(1-\gamma_x)f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{1-\gamma_x},-\boldsymbol{\lambda}_{1-\gamma_x},\boldsymbol{\nu}), \qquad \mathbf{z}_t\in\mathbb{R}^m,
\end{aligned} \tag{9}
$$

where

$$
\begin{aligned}
&\boldsymbol{\mu}=\boldsymbol{a}+\boldsymbol{b}\mu_x,\\
&\boldsymbol{\Omega}_{\gamma_x}=\boldsymbol{\Sigma}+\gamma_x{}^2\sigma_x^2\mathbf{b}\mathbf{b}^\top,\\
&\boldsymbol{\lambda}_{\gamma_x}=\frac{\gamma_x\sigma_x}{\sqrt{1-\sigma_x^2\gamma_x^2\mathbf{b}^\top\boldsymbol{\Omega}_{\gamma_x}^{-1}\mathbf{b}}}\boldsymbol{\Omega}_{\gamma_x}^{-1/2}\mathbf{b}.
\end{aligned}
$$

Similarly, $\boldsymbol{\Omega}_{1-\gamma_x}$ *and* $\boldsymbol{\lambda}_{1-\gamma_x}$ *could be defined by using* $1-\gamma_x$ *instead of* γ_x.

Using the well-known Sherman-Morrison formula

$$\mathbf{\Omega}_{\gamma_x}^{-1} = \left(\mathbf{\Sigma} + \sigma_x^2\gamma_x^2\mathbf{b}\mathbf{b}^\top\right)^{-1} = \mathbf{\Sigma}^{-1} - \frac{\gamma_x^2\sigma_x^2\mathbf{\Sigma}^{-1}\mathbf{b}\mathbf{b}^\top\mathbf{\Sigma}^{-1}}{1+\sigma_x^2\gamma_x^2\mathbf{b}^\top\mathbf{\Sigma}^{-1}\mathbf{b}}.$$

Since we assume that $x_t \sim TP-SMN(\mu_x,\sigma_x,\gamma_x,\boldsymbol{\nu})$, based on Corollary 1, the stochastic representation of x_t is $x_t \stackrel{iid}{=} \mu_x + U_t^{-1/2} w_t v_t$ for $w_t \in \{\gamma_x, -(1-\gamma_x)\}$, which is a random variable with probability function $p(w_t;\gamma_x)$ given by (2), and $v_t \sim HN(0,\sigma_x^2;(0,\infty))$, and $U_t \sim H(.;\boldsymbol{\nu})$ then we have the equivalent formulation of our model.

Proposition 2. *The structural TP-SMN-RMEM defined by* (7) *has a hierarchical representation as follows,*

$$\begin{aligned}
&\mathbf{z}_t|v_t,w_t,u_t \stackrel{ind.}{\sim} N_m(\mathbf{a}+\mathbf{b}\mu_x+\mathbf{b}v_t w_t, u_t^{-1}\mathbf{\Sigma}),\\
&v_t|u_t \stackrel{i.d.d.}{\sim} HN(0,u_t^{-1}\sigma_x^2;(0,\infty)),\\
&w_t \stackrel{i.i.d.}{\sim} p(.;\gamma_x),\\
&U_t \stackrel{i.i.d.}{\sim} H(.;\boldsymbol{\nu}),
\end{aligned} \tag{10}$$

where $v_t|u_t$ and w_t are independent variables, for $t = 1,\dots,n$.

As a first consequence of Proposition 2, we compute the first two moments of the observation variables $\mathbf{z}_t$. Using Corollary 1, equation (5) and the hierarchical representation given by (7), we obtain

$$\begin{aligned}
\mathrm{E}(\mathbf{Z}_t) &= \mathrm{E}(\mathrm{E}(\mathbf{Z}_t|x_t)) = \mathbf{a}+\mathbf{b}\mu_x+\mathbf{b}\sigma_x(2\gamma_x-1)\sqrt{\frac{2}{\pi}}\mathrm{E}\left(U_t^{-1/2}\right),\\
\mathrm{var}(\mathbf{Z}_t) &= \mathrm{var}(\mathrm{E}(\mathbf{Z}_t|x_t)) + \mathrm{E}(\mathrm{var}(\mathbf{Z}_t|x_t))\\
&= \mathrm{E}\left(U_t^{-1}\right)\left(\mathbf{\Sigma}+[\gamma_x^3+(1-\gamma_x)^3]\sigma_x^2\mathbf{b}\mathbf{b}^\top\right)\\
&\quad -\frac{2}{\pi}\mathrm{E}\left(U_t^{-1/2}\right)^2(2\gamma_x-1)^2\sigma_x^2\mathbf{b}\mathbf{b}^\top.
\end{aligned}$$

Denoting the t-th complete dataset by $\mathbf{z}_{ct} = \{\mathbf{z}_t, v_t, w_t, u_t\}$, according to

$$\begin{aligned}
f(\mathbf{z}_t,v_t,w_,u_t;\boldsymbol{\theta}) &= f(\mathbf{z}_t|v_t,w_t,u_t;\boldsymbol{\theta})f(v_t|u_t;\boldsymbol{\theta})p(w_t;\gamma_x)h(u_t;\boldsymbol{\nu})\\
&= \phi_m(\mathbf{z}_t;\mathbf{a}+\mathbf{b}\mu_x+\mathbf{b}v_t w_t, u_t^{-1}\mathbf{\Sigma})\\
&\quad \times 2\phi(v_t;0,u_t^{-1}\sigma_x^2)\times\gamma_x^{(1+s_t)/2}(1-\gamma_x)^{(1-s_t)/2}I_{\{-1,1\}}(s_t)h(u_t;\boldsymbol{\nu}),
\end{aligned} \tag{11}$$

which yields the following proposition.

Proposition 3. *Let us represent the random vector $\mathbf{z}_t \sim TP-SMN-RMEM$ as* (9)*, the structural TP-SMN-RMEM defined by Proposition 2 and the joint distribution of $(\mathbf{z}_t,v_t,w_t,u_t)$ admits* (11)*,* *thus following that,*

$$(i)\quad p(w_t|\mathbf{z}_t;\boldsymbol{\theta}) = \begin{cases} \pi_{\gamma_x t} & if \quad w_t = \gamma_x \\ \pi_{(1-\gamma_x)t} & if \quad w_t = -(1-\gamma_x), \end{cases}$$

$$(ii)\quad f(v_t|w_t,u_t,\mathbf{z}_t;\boldsymbol{\theta}) = \frac{\phi(v_t;\mu_{w_t},\sigma^2_{w_t}(u_t))}{\Phi(\tau_{w_t}(u_t))} I_{\{v_t\}}(0,\infty),$$

$$(iii)\quad f(v_t|u_t,\mathbf{z}_t;\boldsymbol{\theta}) = \pi_{\gamma_x t}(u_t)\frac{\phi(v_t;\mu_{\gamma_x t},\sigma^2_{\gamma_x t}(u_t))}{\Phi(\tau_{\gamma_x t}(u_t))} I_{\{v_t\}}(0,\infty) + \pi_{(1-\gamma_x)t}(u_t)\frac{\phi(v_t;\mu_{(1-\gamma_x)t},\sigma^2_{(1-\gamma_x)t}(u_t))}{\Phi(\tau_{-(1-\gamma_x)t}(u_t))} I_{\{v_t\}}(0,\infty),$$

$$(iv)\quad f(u_t|w_t,\mathbf{z}_t;\boldsymbol{\theta}) = \frac{\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{w_t})\Phi(\tau_{w_t}(u_t))h(u_t;\boldsymbol{\nu})}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{w_t},\boldsymbol{\lambda}_{w_t},\boldsymbol{\nu})},$$

$$(v)\quad f(u_t|\mathbf{z}_t;\boldsymbol{\theta}) = \frac{\gamma_x\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{\gamma_x})\Phi(\tau_{\gamma_x t}(u_t))h(u_t;\boldsymbol{\nu})}{f(\mathbf{z}_t;\boldsymbol{\theta})} + \frac{(1-\gamma_x)\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{1-\gamma_x})\Phi(\tau_{-(1-\gamma_x)t}(u_t))h(u_t;\boldsymbol{\nu})}{f(\mathbf{z}_t;\boldsymbol{\theta})},$$

with $w_t = \gamma_x, -(1-\gamma_x)$, for $t = 1,\ldots,n$, leading to

$$\mu_{w_t} = \frac{\sigma_x^2 w_t \mathbf{b}^\top\boldsymbol{\Sigma}^{-1}(\mathbf{z}_t-\boldsymbol{\mu})}{1+\sigma_x^2 w_t^2\mathbf{b}^\top\boldsymbol{\Sigma}^{-1}\mathbf{b}},\ \sigma^2_{w_t}(u_t) = \frac{u_t^{-1}\sigma_x^2}{1+\sigma_x^2 w_t^2\mathbf{b}^\top\boldsymbol{\Sigma}^{-1}\mathbf{b}}, \tau_{w_t}(u_t) = \frac{\mu_{w_t}}{\sigma_{w_t}(u_t)},$$

$$\pi_{\gamma_x t} = \frac{\gamma_x f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x},\boldsymbol{\nu})}{f(\mathbf{z}_t;\boldsymbol{\theta})},$$

$$\pi_{(1-\gamma_x)t} = \frac{(1-\gamma_x)f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{1-\gamma_x},-\boldsymbol{\lambda}_{1-\gamma_x},\boldsymbol{\nu})}{f(\mathbf{z}_t;\boldsymbol{\theta})},$$

$$\pi_{\gamma_x t}(u_t) = \frac{\gamma_x\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{\gamma_x})\Phi(\tau_{\gamma_x t}(u_t))}{f(\mathbf{z}_t|u_t,\boldsymbol{\theta})},$$

$$\pi_{(1-\gamma_x)t}(u_t) = \frac{(1-\gamma_x)\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{1-\gamma_x})\Phi(\tau_{-(1-\gamma_x)t}(u_t))}{f(\mathbf{z}_t|u_t,\boldsymbol{\theta})},$$

$$f(\mathbf{z}_t|u_t,\boldsymbol{\theta}) = \gamma_x\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{\gamma_x})\Phi(\tau_{\gamma_x t}(u_t)) + (1-\gamma_x)\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{1-\gamma_x})\Phi(\tau_{-(1-\gamma_x)t}(u_t)).$$

4 Maximum likelihood estimation via the EM algorithm

The advantages in terms of efficiency and robustness of MLEs of model (4) based on the ECM technique studied by Meng and Rubin (1993) are considered. We start by considering the hierarchical representation (10) and the fact that

$$p(w_t;\gamma_x) = \gamma_x^{(1+s_t)/2}(1-\gamma_x)^{(1-s_t)/2} I_{\{-1,1\}}(s_t).$$

Denoting the estimate of $\boldsymbol{\theta}$ at the k-th iteration by $\widehat{\boldsymbol{\theta}}^{(k)}$ and $\mathbf{z}_c = \{\mathbf{z}, \mathbf{v}, \mathbf{w}, \mathbf{u}\}$ be the complete dataset of model (4), where $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$, $\mathbf{v} = \{v_1, \ldots, v_n\}$, $\mathbf{w} = \{w_1, \ldots, w_n\}$, and $\mathbf{u} = \{u_1, \ldots, u_n\}$. With $\kappa(u_t) = 1/u_t$ and using the equation given by (11), the log-likelihood function for $\boldsymbol{\theta}$ based on the t-th complete data, $\mathbf{z}_{ct}$, is in the following form

$$\ell(\boldsymbol{\theta}|\mathbf{z}_c) = \sum_{t=1}^{n} \ell(\boldsymbol{\theta}|\mathbf{z}_{ct}) = \sum_{t=1}^{n} \left(\ell_{\mathbf{z}_t|v_t,w_t,u_t} + \ell_{v_t|u_t} + \ell_{w_t}\right),$$

where

$$\begin{aligned}
\ell_{\mathbf{z}_t|v_t,w_t,u_t} &= -\frac{1}{2} log(|2\pi u_t^{-1}\boldsymbol{\Sigma}|) - \frac{u_t}{2}(\mathbf{z}_t - \mathbf{a})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z}_t - \mathbf{a}) \\
&\quad + u_t(\mathbf{z}_t - \mathbf{a})^\top \boldsymbol{\Sigma}^{-1}\mathbf{b}\mu_x + uwv_t(\mathbf{z}_t - \mathbf{a})^\top \boldsymbol{\Sigma}^{-1}\mathbf{b} \\
&\quad - uwv_t \mathbf{b}^\top \boldsymbol{\Sigma}^{-1}\mathbf{b}\mu_x - \frac{u_t}{2}\mathbf{b}^\top \boldsymbol{\Sigma}^{-1}\mathbf{b}\mu_x^2 - \frac{uw^2v^2{}_t}{2}\mathbf{b}^\top \boldsymbol{\Sigma}^{-1}\mathbf{b}, \\
\ell_{v_t|u_t} &= -\frac{1}{2} log(2\pi u_t^{-1}\sigma_x^2) - \frac{u_t}{2}\frac{v_t^2}{\sigma_x^2}, \\
\ell_{w_t} &= \frac{1}{2}(1+s_t) log\gamma_x + \frac{1}{2}(1-s_t) log(1-\gamma_x).
\end{aligned}$$

Obviously, from equation (8), we have that $|\boldsymbol{\Sigma}| = \phi_\delta^p \phi_\xi^{q-1}(\phi_\xi + q\phi_e)$, and

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}^{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{22} \end{bmatrix} \tag{12}$$

in which $\boldsymbol{\Sigma}^{11} = \frac{1}{\phi_\delta}\mathbf{I}_p$ and $\boldsymbol{\Sigma}^{22} = \frac{1}{\phi_\xi}\mathbf{I}_q - \frac{\phi_e}{\phi_\xi(\phi_\xi + q\phi_e)}\mathbf{1}_q\mathbf{1}_q{}^\top$. The constants that are independent of $\boldsymbol{\theta}$ in the above expressions can be ignored. The EM algorithm is constructed as follows. Given the current value $\widehat{\boldsymbol{\theta}}^{(k)}$ of $\boldsymbol{\theta}$, the E-step of the EM algorithm calculates the Q-function defined by $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{t=1}^{n} Q_t(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$

$$\begin{aligned}
Q_t(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{p}{2}\log\phi_\delta - \frac{q-1}{2}\log\phi_\xi - \frac{1}{2}\log(\phi_\xi + q\phi_e), \\
&\quad - \frac{1}{2}\widehat{u}_t^{(k)}\left[\frac{1}{\phi_\delta}\sum_{i=1}^{p} X_t^{2(i)} + \frac{1}{\phi_\xi}\sum_{j=1}^{q}(Y_t^{(j)} - \alpha)^2 - \frac{q^2\phi_e}{\phi_\xi(\phi_\xi + q\phi_e)}(\bar{Y}_t - \alpha)^2\right] \\
&\quad + \widehat{u}_t^{(k)}\mu_x\left[\frac{p}{\phi_\delta}\bar{X}_t + \frac{q\beta}{(\phi_\xi + q\phi_e)}(\bar{Y}_t - \alpha)\right] \\
&\quad + \widehat{uwv}_t^{(k)}\left[\frac{p}{\phi_\delta}\bar{X}_t + \frac{q\beta}{(\phi_\xi + q\phi_e)}(\bar{Y}_t - \alpha)\right] \\
&\quad - \widehat{uwv}_t^{(k)}\mu_x\left[\frac{p}{\phi_\delta} + \frac{q\beta^2}{(\phi_\xi + q\phi_e)}\right] \\
&\quad - \frac{1}{2}\widehat{u}_t^{(k)}\mu_x^2\left[\frac{p}{\phi_\delta} + \frac{q\beta^2}{(\phi_\xi + q\phi_e)}\right] \\
&\quad - \frac{1}{2}\widehat{uw^2v^2{}_t}^{(k)}\left[\frac{p}{\phi_\delta} + \frac{q\beta^2}{(\phi_\xi + q\phi_e)}\right] \\
&\quad - \log\sigma_x - \frac{1}{2\sigma_x^2}\widehat{uv^2{}_t}^{(k)} + \frac{1}{2}(1 + \widehat{s}_t^{(k)})\log\gamma_x + \frac{1}{2}(1 - \widehat{s}_t^{(k)})\log(1-\gamma_x),
\end{aligned} \tag{13}$$

in which $\mathbf{Y}_t = (\mathbf{Y}_t^{(1)}, \ldots, \mathbf{Y}_t^{(q)})^\top$, $\boldsymbol{\mu}_y = (\alpha + \beta\mu_x)\mathbf{1}_q$, $\boldsymbol{\beta} = \beta\mathbf{1}_q$. Furthermore, to obtain the Q-function, we first need to calculate the following conditional expectations, which must be evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$.
E-step:

$$
\begin{aligned}
\widehat{s}_t^{(k)} &= E_{\widehat{\boldsymbol{\theta}}^{(k)}}\{s_t|\mathbf{Z}_t = \mathbf{z}_t\}, \\
\widehat{u}_t^{(k)} &= E_{\widehat{\boldsymbol{\theta}}^{(k)}}\{U_t|\mathbf{Z}_t = \mathbf{z}_t\}, \\
\widehat{uv^2}_t^{(k)} &= E_{\widehat{\boldsymbol{\theta}}^{(k)}}\left\{U_t v_t^2|\mathbf{Z}_t = \mathbf{z}_t\right\}, \\
\widehat{uwv}_t^{(k)} &= E_{\widehat{\boldsymbol{\theta}}^{(k)}}\{U_t w_t v_t|\mathbf{Z}_t = \mathbf{z}_t\}, \\
\widehat{uw^2v^2}_t^{(k)} &= E_{\widehat{\boldsymbol{\theta}}^{(k)}}\left\{U_t w_t^2 v_t^2|\mathbf{Z}_t = \mathbf{z}_t\right\},
\end{aligned}
$$

where $E_{\widehat{\boldsymbol{\theta}}^{(k)}}$ means that the expectation is computed at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$. The derivation of these moments will be useful in the implementation of the EM algorithm.

Proposition 4. *The structural TP-SMN-RMEM defined by Proposition 2 and the joint distribution of* $(\mathbf{z}_t, v_t, w_t, u_t)$ *admits (11), thus following that*

$$
\begin{aligned}
&(i) \quad E\{s_t|\mathbf{z}_t;\boldsymbol{\theta}\} = \pi_{\gamma_x t} - \pi_{(1-\gamma_x)t}, \\
&(ii) \quad E\{v_t^2 U_t|\mathbf{z}_t;\boldsymbol{\theta}\} = \pi_{\gamma_x t}\left\{\frac{\chi_t(\gamma_x)\mu_{\gamma_x t}^2 + \zeta_t(\gamma_x)\mu_{\gamma_x t}\sigma_{\gamma_x t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x},\boldsymbol{\nu})} + \sigma_{\gamma_x t}^2\right\} \\
&\qquad + \pi_{(1-\gamma_x)t}\left\{\frac{\chi_t(1-\gamma_x)\mu_{(1-\gamma_x)t}^2 + \zeta_t(1-\gamma_x)\mu_{(1-\gamma_x)t}\sigma_{(1-\gamma_x)t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{1-\gamma_x},-\boldsymbol{\lambda}_{1-\gamma_x},\boldsymbol{\nu})} + \sigma_{(1-\gamma_x)t}^2\right\}, \\
&(iii) \quad E\{v_t w_t U_t|\mathbf{z}_t;\boldsymbol{\theta}\} = \gamma_x\pi_{\gamma_x t}\left\{\frac{\chi_t(\gamma_x)\mu_{\gamma_x t} + \zeta_t(\gamma_x)\sigma_{\gamma_x t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x},\boldsymbol{\nu})}\right\} \\
&\qquad - (1-\gamma_x)\pi_{(1-\gamma_x)t}\left\{\frac{\chi_t(1-\gamma_x)\mu_{(1-\gamma_x)t} + \zeta_t(1-\gamma_x)\sigma_{(1-\gamma_x)t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{1-\gamma_x},-\boldsymbol{\lambda}_{1-\gamma_x},\boldsymbol{\nu})}\right\}, \\
&(iv) \quad E\{v_t^2 w_t^2 U_t|\mathbf{z}_t;\boldsymbol{\theta}\} = \gamma_x^2\pi_{\gamma_x t}\left\{\frac{\chi_t(\gamma_x)\mu_{\gamma_x t}^2 + \zeta_t(\gamma_x)\mu_{\gamma_x t}\sigma_{\gamma_x t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x},\boldsymbol{\nu})} + \sigma_{\gamma_x t}^2\right\} \\
&\qquad + (1-\gamma_x)^2\pi_{(1-\gamma_x)t}\left\{\frac{\chi_t(1-\gamma_x)\mu_{(1-\gamma_x)t}^2 + \zeta_t(1-\gamma_x)\mu_{(1-\gamma_x)t}\sigma_{(1-\gamma_x)t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{1-\gamma_x},-\boldsymbol{\lambda}_{1-\gamma_x},\boldsymbol{\nu})} + \sigma_{(1-\gamma_x)t}^2\right\},
\end{aligned}
$$

where $t = 1, \ldots, n$, $s_t = sign(w_t) \in \{-1, 1\}$ *and*

$$
\chi_t(\gamma_x) = \begin{cases} f_{SN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x}) & \text{if } U_t \stackrel{d}{=} 1, \\ 2t_m(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\nu)\frac{\nu+m}{\nu+d_{\gamma_x}}T(\sqrt{\frac{\nu+m+2}{\nu+d_{\gamma_x}}}A_{\gamma_x};\nu+m+2), & \text{if } U_t \sim \Gamma(\nu/2,\nu/2), \\ 2f_{SL}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\nu)\frac{2\nu+m}{d_{\gamma_x}}\frac{P_1(\frac{m+2\nu+2}{2},\frac{d_{\gamma_x}}{2})}{P_1(\frac{m+2\nu}{2},\frac{d_{\gamma_x}}{2})}E\left(\Phi(S_{\gamma_x}^{1/2}A_{\gamma_x})\right), & \text{if } U_t \sim Beta(\nu,1), \end{cases}
$$

Also, for

$$
h(u_t;\nu_1,\nu_2) = \begin{cases} \nu_2 & \nu_1 \\ 1 & 1-\nu_1, \end{cases}
$$

we have

$$
\chi_t(\gamma_x) = 2\left\{\nu_1\nu_2\phi_m(\mathbf{z}_t;\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Omega}_{\gamma_x})\Phi(\nu_2^{1/2}A_{\gamma_x}) + (1-\nu_1)\phi_m(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x})\Phi(A_{\gamma_x})\right\},
$$

$$\zeta_t(\gamma_x)=\begin{cases}\phi_m(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x})\phi(A_{\gamma_x}), & \text{if } U_t \overset{d}{=} 1,\\ \frac{\sigma_{\gamma_x t}}{\sigma_x}\frac{\Gamma(\frac{m+\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu+2}{2})}\{\nu+d_{\mathbf{z}_t}(\boldsymbol{\mu},\boldsymbol{\Sigma})\}^{-\frac{1}{2}}t_m(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Sigma},\nu), & \text{if } U_t\sim\Gamma(\nu/2,\nu/2),\\ \frac{\sigma_{\gamma_x t}}{\sigma_x}\frac{\Gamma(\frac{m+2\nu+1}{2})}{\sqrt{2\pi}\Gamma(\frac{2\nu+m}{2})}\{\frac{2}{d_{\mathbf{z}_t}(\boldsymbol{\mu},\boldsymbol{\Sigma})}\}^{\frac{1}{2}} & \\ \times\frac{P_1(\frac{2\nu+m+1}{2},\frac{d_{\mathbf{z}_t}(\boldsymbol{\mu},\boldsymbol{\Sigma})}{2})}{P_1(\frac{2\nu+m}{2},\frac{d_{\mathbf{z}_t}(\boldsymbol{\mu},\boldsymbol{\Sigma})}{2})}f_{SL}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) & \text{if } U_t\sim Beta(\nu,1).\end{cases}$$

Also, for

$$h(u_t;\nu_1,\nu_2)=\begin{cases}\nu_2 & \nu_1\\ 1 & 1-\nu_1\end{cases}$$

we have

$$\zeta_t(\gamma_x)=\nu_1\sqrt{\nu_2}\phi_m(\mathbf{z}_t;\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Omega}_{\gamma_x})\phi(\sqrt{\nu_2}A_{\gamma_x}),$$

where

$$\sigma^2_{\gamma_x t}=\frac{\sigma_x^2}{1+\sigma_x^2\gamma_x^2\mathbf{b}^\top\boldsymbol{\Sigma}^{-1}\mathbf{b}},\qquad d_{\mathbf{z}_t}(\boldsymbol{\mu},\boldsymbol{\Sigma})=(\mathbf{z}_t-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{z}_t-\boldsymbol{\mu}).$$

Furthermore, $\pi_{\gamma_x t}$*,* $\mu_{\gamma_x t}$ *are given by Proposition 3, and the distribution of the variable* S_s *is given above. Also, note that we should use* $-\boldsymbol{\lambda}_{1-\gamma_x}$ *wherever it is needed.*

The M-step consists of maximization of $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$. For this, we use the faster extension of the original EM, the ECME algorithm, by replacing the M-step with a sequence of conditional maximization (CM) steps. CM-steps then conditionally maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$, obtaining a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as follows:

CM-step Maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$. This is accomplished by the following steps.

CM-step.1: Update $\widehat{\gamma}_x^{(k)},\widehat{\sigma}_x^{2(k)},\widehat{\mu}_x^{(k)}$ by

$$\begin{aligned}\widehat{\gamma}_x^{(k)}&=\frac{1+\widehat{\bar{s}}^{(k)}}{2},\\ \widehat{\sigma}_x^{2(k+1)}&=\frac{1}{n}\sum_{t=1}^n\widehat{uv^2}_t^{(k)},\\ \widehat{\mu}_x^{(k+1)}&=\frac{\sum_{t=1}^n\widehat{u}_t\bar{X}_t-\sum_{t=1}^n\widehat{uwv}_t}{\sum_{t=1}^n\widehat{u}_t},\\ \widehat{\mu}_y^{(k+1)}&=\frac{\sum_{t=1}^n\widehat{u}_t\bar{Y}_t-\sum_{t=1}^n\widehat{uwv}_t\widehat{\beta}^{(k)}}{\sum_{t=1}^n\widehat{u}_t},\end{aligned}$$

where $\mu_y=\alpha+\beta\mu_x$, $\widehat{\bar{s}}^{(k)}=\sum_{t=1}^n\widehat{s}_t^{(k)}/n$, $\bar{X}_t=\sum_{i=1}^p X_t^{(i)}/p$, $\bar{Y}_t=\sum_{i=1}^q Y_t^{(j)}/q$,

CM-step.2: With fixed $\widehat{\mu}_x=\widehat{\mu}_x^{(k+1)}$, $\widehat{\mu}_y=\widehat{\mu}_y^{(k+1)}$

1. Update $\widehat{\beta}^{(k)}$ by

$$\widehat{\beta}^{(k+1)}=\frac{\sum_{t=1}^n\widehat{uwv}_t^{(k)}(\bar{Y}_t-\widehat{\mu}_y^{(k+1)})}{\sum_{t=1}^n\widehat{uw^2v^2}_t^{(k)}}.$$

2. Update $\widehat{\phi}_{\delta}^{(k)}$ by

$$\widehat{\phi}_{\delta}^{(k+1)} = \frac{1}{n}\left\{\sum_{t=1}^{n}\frac{1}{p}\widehat{u}_t^{(k)}\sum_{i=1}^{p}(X_t^{(i)}-\widehat{\mu}_x^{(k+1)})^2 - 2\sum_{t=1}^{n}\widehat{uwv}_t^{(k)}(\bar{X}_t-\widehat{\mu}_x^{(k+1)})\right\} + \frac{1}{n}\sum_{t=1}^{n}\widehat{uw^2v^2}_t^{(k)}$$

3. Update $\widehat{\phi}_{\xi}^{(k)}$ by

$$\widehat{\phi}_{\xi}^{(k+1)} = \frac{1}{n(q-1)}\left\{\sum_{t=1}^{n}\widehat{u}_t^{(k)}\sum_{j=1}^{q}(Y_t^{(j)}-\widehat{\mu}_y^{(k+1)})^2 - q\sum_{t=1}^{n}\widehat{u}_t^{(k)}(\bar{Y}_t-\widehat{\mu}_y^{(k+1)})^2\right\}.$$

4. Update $\widehat{\phi}_{e}^{(k)}$ by

$$\widehat{\phi}_{e}^{(k+1)} = \frac{1}{n(q-1)}\left\{q\sum_{t=1}^{n}\widehat{u}_t^{(k)}(\bar{Y}_t-\widehat{\mu}_y^{(k+1)})^2 - \frac{1}{q}\sum_{t=1}^{n}\widehat{u}_t^{(k)}\sum_{j=1}^{q}(Y_t^{(j)}-\widehat{\mu}_y^{(k+1)})^2\right\} + \frac{1}{n}\sum_{t=1}^{n}\widehat{uw^2v^2}_t^{(k)}\widehat{\beta}^{2(k+1)} - \frac{2}{n}\sum_{t=1}^{n}\widehat{uwv}_t^{(k)}\widehat{\beta}^{(k+1)}(\bar{Y}_t-\widehat{\mu}_y^{(k+1)}).$$

CM-step.3: Update $\widehat{\alpha}^{(k)}$ by

$$\widehat{\alpha}^{(k+1)} = \widehat{\mu}_y^{(k+1)} - \widehat{\beta}^{(k+1)}\widehat{\mu}_x^{(k+1)}.$$

Remark 1. *The iteration of the ECM algorithm is repeated until the difference between two successive log-likelihood values,* $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)};\mathbf{z}_1,\ldots,\mathbf{z}_n)-\ell(\widehat{\boldsymbol{\theta}}^{(k)};\mathbf{z}_1,\ldots,\mathbf{z}_n)|$*, is sufficiently small, say* 10^{-6}*. As initial values for the algorithm, we can use the mean vector* $\bar{\mathbf{z}}$ *for* $\boldsymbol{\mu}$*, the sample variance* $S_{\mathbf{X}}^2=\frac{1}{n}\sum_{t=1}^{n}S_{\mathbf{X}_t}^2$ *with* $S_{\mathbf{X}_t}^2=\frac{1}{p}\sum_{i=1}^{p}(X_t^{(i)}-\bar{X}_t)^2$ *(according to the model given by (4)) for* $\bar{u}(\sigma_x^2+\phi_\delta)$ *with* $\bar{u}=\frac{1}{n}\sum_{i=1}^{n}u_t$*. As for* ϕ_δ*, it can be set equal to some fraction of* σ_x^2*; in our subsequent numerical work, we have started the iterative process with* $\sigma_x^2=\phi_\delta$*. Similarly, the sample variance* $S_{\mathbf{Y}}^2=\frac{1}{n}\sum_{t=1}^{n}S_{\mathbf{Y}_t}^2$ *with* $S_{\mathbf{Y}_t}^2=\frac{1}{q}\sum_{j=1}^{p}(Y_t^{(j)}-\bar{Y}_t)^2$ *for* $\bar{u}\phi_\xi$*. Regarding* γ*, an option would be to start from the sample skewness of the observed explanatory variables* $\mathbf{X}=(\mathbf{X}_1,\ldots,\mathbf{X}_n)^\top$ *with* $\mathbf{X}_t=(X_t^{(1)},\ldots,X_t^{(p)})^\top$*, namely* $g_1(\mathbf{X})=\frac{1}{n}\sum_{t=1}^{n}g_1(\mathbf{X}_t)$*, where* $g_1(\mathbf{X}_t)=\frac{m_{3\mathbf{X}_t}}{m_{2\mathbf{X}_t}}$*, with* $m_{k\mathbf{X}_t}=\frac{1}{p}\sum_{i=1}^{p}(X_t^{(i)}-\bar{X}_t)^k$*, and invert these values to obtain an initial choice of* γ_x*, In part (ii) Corollary 1 the Pearson measure of skewness was calculated for the TP-SMN family. .*

4.1 Asymptotic covariance matrix of MLEs

According to Arellano-Valle et al. (2020) approach, the empirical information matrix can be approximated as

$$\mathbf{I}_{emp}(\boldsymbol{\theta};\mathbf{z}) = \sum_{t=1}^{n}U(\mathbf{z}_t;\boldsymbol{\theta})U^\top(\mathbf{z}_t;\boldsymbol{\theta}) - n\bar{U}(\mathbf{z};\boldsymbol{\theta})\bar{U}^\top(\mathbf{z};\boldsymbol{\theta}),$$

where $\bar{U}(\mathbf{z};\boldsymbol{\theta})=\frac{1}{n}\sum_{t=1}^{n}U(\mathbf{z}_t;\boldsymbol{\theta})$ and $U(\mathbf{z}_t;\boldsymbol{\theta})$ is the empirical scoring function for the subject t. The individual score function can be determined as (Louis (1982))

$$U(\mathbf{z}_t;\boldsymbol{\theta}) = \frac{\partial \log f(\mathbf{z}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \mathrm{E}\left\{\frac{\partial\ell(\boldsymbol{\theta};\mathbf{z}_{ct})}{\partial\boldsymbol{\theta}}|\mathbf{z}_t;\theta\right\},$$

where $\mathbf{z}_{ct} = \{\mathbf{z}_t, w_t, v_t, u_t\}$, is the complete data vector from the tth observation and $\ell(\boldsymbol{\theta};\mathbf{z}_{ct})$ is the corresponding contribution to the log-likelihood function.

$$\mathbf{I}_{emp}(\widehat{\boldsymbol{\theta}};\mathbf{z}) = \sum_{t=1}^{n} \widehat{U}_t \widehat{U}_t^{\top},$$

where $\widehat{U}_t = U(\mathbf{z}_t;\widehat{\boldsymbol{\theta}}) = (\widehat{U}_{t,\alpha}, \widehat{U}_{t,\beta}, \widehat{U}_{t,\mu_x}, \widehat{U}_{t,\sigma_x}, \widehat{U}_{t,\gamma_x}, \widehat{U}_{t,\phi_\delta}, \widehat{U}_{t,\phi_\xi}, \widehat{U}_{t,\phi_e})^{\top}$, with

$$\begin{aligned}
\widehat{U}_{t,\alpha} &= \frac{q}{\widehat{\phi}_\xi + q\widehat{\phi}_e}\left\{\widehat{u}_t(\bar{Y}_t - \widehat{\mu}_y) - \widehat{\beta}\widehat{uwv}_t\right\},\\
\widehat{U}_{t,\beta} &= \frac{q}{\widehat{\phi}_\xi + q\widehat{\phi}_e}\left\{(\bar{Y}_t - \widehat{\alpha})(\widehat{u}_t\widehat{\mu}_x + \widehat{uwv}_t) - \widehat{\beta}(\widehat{u}_t\widehat{\mu}_x^2 + 2\widehat{uwv}_t\widehat{\mu}_x + \widehat{uw^2v^2}_t)\right\},\\
\widehat{U}_{t,\mu_x} &= \frac{p}{\widehat{\phi}_\delta}\left\{\widehat{u}_t(\bar{X}_t - \widehat{\mu}_x) - \widehat{uwv}_t\right\} + \frac{q\widehat{\beta}}{\widehat{\phi}_\xi + q\widehat{\phi}_e}\left\{\widehat{u}_t(\bar{Y}_t - \widehat{\mu}_y) - \widehat{\beta}\widehat{uwv}_t\right\},\\
\widehat{U}_{t,\sigma_x} &= -\frac{1}{\widehat{\sigma}_x} + \frac{1}{\widehat{\sigma}_x^3}\widehat{uv^2}_t,\\
\widehat{U}_{t,\gamma_x} &= \frac{1}{\widehat{\gamma}_x}\left(\frac{1+\widehat{s}_t}{2}\right) - \frac{1}{1-\widehat{\gamma}_x}\left(\frac{1-\widehat{s}_t}{2}\right),\\
\widehat{U}_{t,\phi_\delta} &= \frac{p}{2\widehat{\phi}_\delta^2}\left\{\frac{\widehat{u}_t}{p}\sum_{i=1}^{p}(X_t^{(i)} - \widehat{\mu}_x)^2 - 2(\bar{X}_t - \widehat{\mu}_x)\widehat{uwv}_t + \widehat{uw^2v^2}_t - \widehat{\phi}_\delta\right\},\\
\widehat{U}_{t,\phi_\xi} &= \frac{q}{2(\widehat{\phi}_\xi + q\widehat{\phi}_e)^2}\left\{-q\frac{\widehat{\phi}_e(2\widehat{\phi}_\xi + q\widehat{\phi}_e)}{\widehat{\phi}_\xi^2}(\bar{Y}_t - \widehat{\mu}_y)^2\widehat{u}_t - \widehat{\beta}(\bar{Y}_t - \widehat{\mu}_y)\widehat{uwv}_t + \widehat{\beta}^2\widehat{uw^2v^2}_t\right\}\\
&\quad - \frac{q}{2}\left(\frac{\widehat{\phi}_\xi + (q-1)\widehat{\phi}_e}{\widehat{\phi}_\xi(\widehat{\phi}_\xi + q\widehat{\phi}_e)}\right) + \frac{\widehat{u}_t}{2\widehat{\phi}_\xi^2}\sum_{j=1}^{q}(Y_t^{(j)} - \widehat{\mu}_y)^2,\\
\widehat{U}_{t,\phi_e} &= \frac{q^2}{2(\widehat{\phi}_\xi + q\widehat{\phi}_e)^2}\left\{(\bar{Y}_t - \widehat{\mu}_y)^2\widehat{u}_t - 2\widehat{\beta}(\bar{Y}_t - \widehat{\mu}_y)\widehat{uwv}_t + \widehat{\beta}^2\widehat{uw^2v^2}_t - \frac{1}{q}(\widehat{\phi}_\xi + q\widehat{\phi}_e)\right\}.
\end{aligned}$$

5 Simulation study

The simulation study is considered in this section. It is well known that misspecification of the model's distribution will lead to biases of parameter estimates. The main goal is to confirm the effectiveness and precision of MLEs under TP-SMN distributions. We consider four skew distributions, including TP-N, TP-t with $\nu = 4$, TP-SL with $\nu = 2$, and TP-CN with $\nu_1 = 0.2, \nu_2 = 0.3$, in RMEM with or without equation error, respectively. Under the circumstances of no equation error, we denote the distributions TP-N, TP-t, TP-SL and TP-CN by TP-N_0, TP-t_0, TP-SL_0 and TP-CN_0 respectively. To maintain the general and consistent form of the estimators among different TP-SMN distributions, the degrees of freedom for the TP-SMN model will not be estimated together with the parameters of interest. Hence, we selected some heavy-tailed distributions with fixed degrees of freedom. In the application, to find the best distribution for the data, the degrees of freedom for different distributions, chosen by the Schwarz (1978) information criterion. As usual, we are interested in the regression parameters α and β in the simulations. We first generate data from the model (4) with the number of replicated observations chosen as $p = 4$ and $q = 3$ under the TP-t distribution with $\nu = 4$. Then, we use the EM algorithm to calculate the MLEs of $\boldsymbol{\theta}$ under the TP-N, TP-t, TP-SL and TP-CN distributions with or without equation error, respectively. We used two indicators to evaluate estimates, including sample bias (BIAS) and standard deviation (SD).

Other parameters of model (4) are set as: $\mu_x = 1.5, \sigma_x = 1, \gamma_x = 0.95, \alpha = 2, \beta = 1, \phi_\delta = 1, \phi_\xi = 0.5$. For comparison, we set $\phi_e = 0.5, 1,\ \phi_e = 1.5$ respectively, which shows that the degree of matching between the true covariate and response tends from strong to weak. Based on 1000 simulations, the results are reported in Table 1. As we expected, the BIAS, and the corresponding SD decrease when the sample size increases from 50 to 100 in all scenarios, and the MLE under TP-t ($\nu = 4$) distribution (the true distribution) is the best estimator using any of the two indicators, which fully reflects the effectiveness and accuracy of the ML estimates. Moreover, performance under the TP-CN and TP-t distributions behaves better than those under the TP-N distribution, which may be attributed to their heavy-tailed features. Thus, we encourage the use of the heavy-tailed model when the data show heavy-tailed features, even if the proposed heavy-tailed distribution is not the true one.

Obviously, the estimates with equation error consistently perform better than the estimates without equation error. Especially, when ϕ_e increases from 0.5 to 1.5, the SD ratios without equation error and with equation error become clearly larger. It states that for the data with skewness or heavy-tailed features, ignoring equation error will bring about a serious deviation for statistical inference. The equation error plays an important role in expressing the uncertain relationship between the true covariates and the response.

6 Application

We give an illustrative example using the CSFII data set. This data set has conducted 24-hour recall measures, as well as three additional 24-hour recall phone interviews of 1827 women who were recorded about their daily diet intake (for example, saturated fat, calories, vitamin, etc.). Carroll et al. (2006) have indicated that saturated fat has a great relationship with the risk of breast cancer and other diseases, but the statistical significance of saturated fat disappeared when adding caloric intake to the logistic regression model. Therefore, it is necessary to reveal the inner relationship between saturated fat and caloric intake. In this illustrative example, we take the calorie intake/5000 as x and the saturated fat intake/100 as y (Carroll et al. (2006); Lin and Cao (2013)). Instead of x and y, the nutrition variables X and Y are calculated using four 24-hour recalls and suppose that they follow the model (4) with $p = q = 4$. For the purpose of verifying the existence of skewness and heavy tails in the latent covariate x, Lin and Cao (2013) fit the data using normal RMEM without equation error. They showed that the latent covariate is positively skewed and heavy-tailed. This indicates that a normal model may not offer a good fit. The degrees of freedom for different distributions, chosen by the Schwarz (1978) information criterion, are also reported in Table 2. Using the profile log-likelihood functions for the three models, getting the highest values of the profile log-likelihood, the degrees of freedom are found as $\nu = 4.3$ for TP-t, $\nu = 1.3$ for TP-SL, and $\nu_1 = 0.29, \nu_2 = 0.22$ for TP-CN. We now consider TP-SMN distributions for x_t and SMN distributions for measurement errors δ_t, ξ_t and equation error e_t in the RMEM, that is, the TP-SMN-RMEM proposed in this paper. We calculate the MLEs of the parameter $\boldsymbol{\theta}$ and their standard errors (SE) and also the Akaike information criterion (AIC) (Akaike (1974)) values based on the RMEM model in the above distributions. These estimates are displayed in Table 2. The AIC value under the TP-t distribution is always the smallest, whether there is an equation error or skewness or not, and it gets the smallest value under the TP-t_0 distribution in all situations. Thus, TP-t-RMEM is more suitable for these data. Note that as another potential competitor, AIC values based on the SMSN-RMEM model have added in Table 2 (in parenthesis). Moreover, the heavy-tailed models show smaller SEs than the normal model. It would be better consider skewness in the models for their smaller AIC values. The practical significance of the parameter β can be described as the positive proportion of saturated fat in calories. The skew parameter γ_x is positive, suggesting that these data are skewed to the right. The values of ϕ_e seem very small and only slightly change between different models, indicating that there is an inapparent random relationship between the intake of calories and saturated fats.

Table 1: Performances of estimators under TPt-RMEM with or without equation error

Parameter	n	Estimator	$\phi_e = 0.5$		$\phi_e = 1$		$\phi_e = 1.5$	
			BIAS	SD	BIAS	SD	BIAS	SD
α	50	$TP-N$	-0.123	0.641	-0.108	0.781	-0.145	0.797
		$TP-t$	-0.052	0.390	-0.033	0.492	-0.054	0.501
		$TP-SL$	-0.072	0.483	-0.051	0.587	-0.074	0.598
		$TP-CN$	-0.069	0.472	-0.047	0.568	-0.073	0.595
		$TP-N_0$	-0.946	1.240	-1.997	2.011	-2.924	2.317
		$TP-t_0$	-0.796	0.542	-1.442	1.042	-1.857	1.545
		$TP-SL_0$	-0.803	0.649	-1.508	1.098	-2.152	1.618
		$TP-CN_0$	-0.801	0.631	-1.492	1.087	-2.138	1.609
	100	$TP-N$	-0.056	0.405	-0.061	0.512	-0.067	0.769
		$TP-t$	-0.020	0.254	-0.021	0.316	-0.028	0.531
		$TP-SL$	-0.031	0.297	-0.032	0.312	-0.043	0.552
		$TP-CN$	-0.029	0.289	-0.031	0.308	-0.042	0.548
		$TP-N_0$	-0.880	0.655	-1.884	1.231	-2.625	1.996
		$TP-t_0$	-0.795	0.382	-1.826	0.692	-1.136	1.379
		$TP-SL_0$	-0.827	0.448	-1.835	0.824	-2.021	1.408
		$TP-CN_0$	-0.820	0.439	-1.831	0.808	-2.018	1.402
β	50	$TP-N$	0.047	0.303	0.041	0.356	0.058	0.397
		$TP-t$	0.026	0.196	0.019	0.221	0.037	0.267
		$TP-SL$	0.034	0.207	0.029	0.268	0.041	0.289
		$TP-CN$	0.032	0.201	0.026	0.259	0.039	0.281
		$TP-N_0$	0.488	0.607	0.996	0.927	1.017	1.128
		$TP-t_0$	0.395	0.211	0.884	0.513	0.998	0.885
		$TP-SL_0$	0.407	0.298	0.941	0.584	1.011	0.916
		$TP-CN_0$	0.405	0.281	0.937	0.571	1.009	0.903
	100	$TP-N$	0.021	0.201	0.026	0.217	0.037	0.228
		$TP-t$	0.010	0.131	0.009	0.152	0.011	0.173
		$TP-SL$	0.014	0.142	0.013	0.167	0.018	0.189
		$TP-CN$	0.013	0.136	0.012	0.161	0.017	0.181
		$TP-N_0$	0.428	0.317	0.976	0.627	1.010	0.638
		$TP-t_0$	0.315	0.182	0.824	0.313	0.971	0.585
		$TP-SL_0$	0.385	0.198	0.901	0.384	0.987	0.416
		$TP-CN_0$	0.379	0.192	0.899	0.380	0.979	0.4160

Table 2: Parameter estimations for *CSFII* data

Models	Degree	AIC	Parameter							
			μ_x	α	β	γ_x	σ_x	ϕ_δ	ϕ_ξ	ϕ_e
$TP-N$	/	-18965(-18842)	0.196 (0.007)	-0.060 (0.022)	0.957 (0.067)	0.793 (0.090)	0.017 (0.002)	0.011 (0.0002)	0.014 (0.0003)	0.0012 (0.0006)
$TP-t$	$\nu = 4$	-21957(-21182)	0.200 (0.007)	-0.052 (0.014)	0.914 (0.052)	0.651 (0.085)	0.011 (0.001)	0.007 (0.0002)	0.008 (0.0002)	0.0003 (0.0002)
$TP-SL$	$\nu = 1.2$	-21837.5(-21066)	0.201 (0.009)	-0.051 (0.016)	0.911 (0.049)	0.701 (0.063)	0.006 (0.001)	0.004 (0.001)	0.005 (0.0001)	0.0002 (0.0001)
$TP-CN$	$\nu_1 = 0.4, \nu_2 = 0.2$	-21591.5(-20954)	0.202 (0.008)	-0.049 (0.014)	0.896 (0.043)	0.672 (0.051)	0.007 (0.001)	0.005 (0.0001)	0.006 (0.0001)	0.0003 (0.0001)
$TP-N_0$	/	-19010.5(-18996)	0.199 (0.007)	-0.067 (0.024)	0.974 (0.069)	0.768 (0.082)	0.018 (0.002)	0.011 (0.0003)	0.014 (0.0003)	/
$TP-t_0$	$\nu = 4$	-22041(-21887.5)	0.200 (0.008)	-0.059 (0.013)	0.934 (0.046)	0.642 (0.078)	0.012 (0.001)	0.008 (0.0002)	0.009 (0.0002)	/
$TP-SL_0$	$\nu = 1.2$	-21921.5(-21153)	0.201 (0.009)	-0.055 (0.017)	0.921 (0.036)	0.691 (0.057)	0.005 (0.001)	0.004 (0.0001)	0.004 (0.0002)	/
$TP-CN_0$	$\nu_1 = 0.4, \nu_2 = 0.2$	-21880.5(-21497)	0.202 (0.008)	-0.051 (0.012)	0.901 (0.036)	0.613 (0.037)	0.008 (0.0007)	0.006 (0.0002)	0.006 (0.0002)	/

7 Conclusion

In this paper, we develop an RMEM in the TP-SMN distribution class, called TP-SMN-RMEM, which is suitable for asymmetric, heavy-tailed, and multimodal data. Also, it includes many special cases, such as nonreplicated MEM under SMN distributions (Lachos et al. (2011)), RMEM under normal distribution (Lin et al. (2004)) or SMN distributions (Lin and Cao (2013)). Moreover, in contrast to similar models, especially, considering SMSN-RMEM (proposed by Cao et al. (2018)), what our proposed model offers exactly is that the SMSN and TP-SMN families have different properties (for instance, different tail behavior). In addition, numerical stability might also be a property that could be used to motivate our proposed model for the explanatory variable. This means that, estimating the skewness parameter in the TP-SMN class of distributions is often easier, and more stable, than estimating the shape parameter in the SMSN family (which controls the asymmetry, tails, location of the mode, and spread of the density). We provide explicit expressions for both the EM-type iterative estimates and the corresponding standard errors. Due to the hierarchical structure of the model, the features of skewness of both covariate and response can be captured under these assumptions. The proposed TP-SMN-RMEM models can reduce the negative impact of distribution misspecification and outliers to some extent. In biological, environmental, chemical, medical and other research areas, measurement error, skewness, multimodality, and outliers commonly exist in real data; thus, the robust TP-SMN-RMEM models would be a good choice to fit such data.

Acknowledgement

The author would like to thank Dr. Marcos Oliveira Prats from the Federal University of Minas Gerais for his help with the R Code.

References

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

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, **36**, 99–102.

Arellano-Valle, R. B., Azzalini A., Ferreira C.S. and Santoro, K. (2020). A two-piece normal measurement error model. *Computational Statistics & Data Analysis*, **144**, 106863.

Arellano-Valle, R. B., Gomez, H. and Quintana, F.A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, **128**, 427–443.

Arellano-Valle, R. B. and Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, **96**, 93–116.

Arellano-Valle, R. B., Ozan, S., Bolfarine, H. and Lachos, V. H. (2005). Skew normal measurement error models. *Journal of Multivariate Analysis*, **96**, 265–281.

Barkhordar, Z., Maleki, M., Khodadadi, Z., Wraith, D. and Negahdari, F. (2020). A Bayesian approach on the two-piece scale mixtures of normal homoscedastic nonlinear regression models. *Journal of Applied Statistics*, **9**, 1305–1322.

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

Buonaccorsi. J. P. (2010). *Measurement Error: Models, Methods, and Applications.* Chapman and Hall, Boca Raton.

Cao, C. Z., Lin, J. G., Shi, J. Q., Wang, W. and Zhang, X. Y. (2015). Multivariate measurement error models for replicated data under heavy-tailed distributions. *Journal of Chemometrics*, **29**, 457–466.

Cao, C. Z., Wang, W., Shi, J. Q. and Lin, J. G. (2018). Measurement error models for replicated data under asymmetric heavy-tailed distributions. *Computational Economics*, **52**, 531–553.

Carroll, R. J. Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement error in nonlinear models. A modern perspective (2nd edn).* Chapman and Hall, Boca Raton.

Chan, L. K. and Mak, T. K. (1979). Maximum likelihood estimation of a linear structural relationship with replication. *Journal of the Royal Statistical Society, Series B*, **41**, 263–268.

Cheng, C. L. and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error.* Oxford University Press, London.

Fuller, W. A. (1987). *Measurement Error Models.* Wiley, New York.

Heteroscedastic nonlinear regression models using asymmetric and heavy tailed two-piece distributions. *AStA Advances in Statistical Analysis*, **105**, 451–467.

Isogava, Y. (1985). Estimating a multivariate linear structural relationship with replication. *Journal of the Royal Statistical Society, Series B*, **47**, 211–215.

Gustafson, P. (2004). *Measurement Error and Misclassification in Statistics and Epidemiology.* Chapman and Hall, Boca Raton.

Ghasami, S., Maleki, M., and Khodadadi, Z. (2020). Leptokurtic and platykurtic class of robust symmetrical and asymmetrical time series models. *Journal of Computational and Applied Mathematics*, **376**, 112806.

Lachos, V. H., Angolini, T., and Abanto-Valle, C. A. (2011). On estimation and local influence analysis for measurement errors models under heavy-tailed distributions. *Statistical Papers*, **52**, 567–590.

Lin, N., Bailey, B. A. and He, X. M. (2004). Adjustment of measuring devices with linear models. *Technometrics*, **46**, 127–134.

Lin, J. G. and Cao, C. Z. (2013). On estimation of measurement error models with replication under heavy-tailed distributions. *Computational Statistics*, **28**, 802–829.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **44**, 226–233.

Maleki, M., Barkhordar, Z., Khodadadi, Z., and Wraith, D. (2019c). A robust class of homoscedastic nonlinear regression models. *Journal of Statistical Computation and Simulation*, **89**, 2765–2781

Maleki, M., Contreras-Reyes, J. E., and Mahmoudi, M. R. (2019d). Robust mixture modeling based on two-piece scale mixtures of normal family. *Axioms*, **8**, 38.

Maleki, M., and Mahmoudi, M. R. (2017). Two-piece location-scale distributions based on scale mixtures of normal family. *Communications in Statistics-Theory and Methods*, **46**, 12356–12369.

Meng, X. L., and Rubin, D. B. (1993). Maximum likelihood estimation via the EM algorithm: a general framework. *Biometrika*, **80**, 267–278.

Moravveji, B., Khodadadi, Z., and Maleki, M. (2019). A Bayesian analysis of two-piece distributions based on the scale mixtures of normal family. *Iranian Journal of Science and Technology, Transactions A: Science*, **43**, 991–1001.

Reiersol, O. (1950). Identifiability of a linear relation between variables which are subject to errors. *Econometrica*, **18**, 375–389.

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

Zarei, A., Khodadadi, Z., Maleki, M. and Zare, K. (2022). Robust mixture regression modeling based on two-piece scale mixtures of normal distributions. *Advances in Data Analysis and Classification*, **17**, 181–210.