Shahid Bahonar University of Kerman
Journal of Mahani Mathematical Research
Print ISSN: 2251-7952 Online ISSN: 2645-4505

PARSIMONIOUS MIXTURE OF MEAN-MIXTURE OF NORMAL DISTRIBUTIONS WITH MISSING DATA

F. Hashemi ✉ and S. Darijani

Special issue Dedicated to memory of professor Mahbanoo Tata
Article type: Research Article
(Received: 08 December 2023, Received in revised form 21 February 2024)
(Accepted: 28 April 2024, Published Online: 29 April 2024)

Abstract. Clustering multivariate data based on mixture distributions is a usual method to characterize groups and label data sets. Mixture models have recently been received considerable attention to accommodate asymmetric and missing data via exploiting skewed and heavy-tailed distributions. In this paper, a mixture of multivariate mean-mixture of normal distributions is considered for handling missing data. The EM-type algorithms are carried out to determine maximum likelihood of parameters estimations. We analyzed the real data sets and conducted simulation studies to demonstrate the superiority of the proposed methodology.

Keywords: EM-type algorithms, Finite mixture model, MMN distribution, Missing data, Skew distribution.
2020 MSC: Primary 62H10, 62H12, 62D10.

1. Introduction

Model-based clustering is an important sub-field of pattern recognition for modeling data with a complex structure. Among several statistical methods, finite mixture (FM) of multivariate distribution is mostly used to identify the pattern of sophisticated and high-dimensional data. Classification based on FM models can divide datasets into some groups such that the similarity within and between groups are maximum and minimum, respectively. Specifically, a p-variate random vector $\boldsymbol{Y}$ is said to have an FM model if it takes the probability density function (pdf) as

$$f(\boldsymbol{y};\boldsymbol{\Theta}) = \sum_{i=1}^{g} \pi_i \phi(\boldsymbol{y}, \boldsymbol{\theta}_i), \tag{1}$$

where $\pi_i > 0$ are mixing proportions with limitation $\pi_1 + \cdots + \pi_g = 1$, $\boldsymbol{\Theta} = (\pi_1, \ldots, \pi_{g-1}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g)$ and $\phi(\cdot, \boldsymbol{\theta}_i)$ denotes the pdf of ith mixing component. Although the initial FM model [6] assumed the multivariate normal distribution as a mixture component, the interest of skewed and heavily-tailed distributions

✉ farzane.hashemi1367@yahoo.com, ORCID: 0000-0002-9260-4631
https://doi.org/10.22103/jmmr.2024.22642.1549

Publisher: Shahid Bahonar University of Kerman
How to cite: F. Hashemi, S. Darijani, *Parsimonious mixture of mean-mixture of normal distributions with missing data*, J. Mahani Math. Res. 2024; 13(3): 33 - 54.

have increasingly been considered as a robust framework for data analysis. For more study about FM models, we can refer to [12, 21, 22, 27], a few recently published contributions.

The parsimonious mixture of multivariate t distributions as well as missing data was proposed in [15] that is applicable for the symmetric and tick-tail data as a rival for the original finite mixture model. In order to extend the normally-based FM model, two important issues are always in interest: (i) theoretical convenience and mathematical properties, and (ii) the flexibility and robustness in analyzing strongly skewed and missing data. In this regard, the family of skew-normal (SN) (Azzalini [4]) distributions is one of distributions that used in constructing an efficient finite mixture models for analyzing skewed data with missing information. For instance, Lin and Lin [16] considered a version of multivariate SN distribution that provides more flexibility in clustering data with strong skewness features. The finite mixture of skew-t distributions introduced in [30] as an extension of the finite mixture of SN distributions for handling heavily-tailed asymmetric with missing information datasets.

The mean mixture of normal (MMN) distribution is a class of skew distributions that introduced by Negarestani et al. [25] as an alternative to the baseline SN distribution. Based on [25], the MMN distribution has inherited mathematical and computational aspects and a flexible platform of statistical analysis. Recently, Hashemi et al. [13] extended the original factor analyzer model based on MMN distribution and showed that the class of MMN distributions outperforms the skew-t distribution whenever the degree of freedom increases. Simulation results of Hashemi et al. [13] showed that the speed of convergence of MMN distribution on the data is much faster than some existing models. Naderi et al. [23] used MMN distributions for proposing a new matrix-variate model and showed that the class of MMN distributions offers different orientation compared to the family of mean-variance mixture of normal distributions.

In this paper, we are devoted to contrive a skew extension of finite mixture models based on the MMN distribution for analysis skewed and missing data, referred to as the FM-MMN model henceforth. The problem of missing data is ubiquitous in many scientific fields and should be addressed appropriately before engaging learning algorithms. Statistical methods of simply deleting subjects with missing values can cause seriously biased estimates and clustering and subsequently leads to distorted inference. A variety of traditional and modern techniques to deal with missing data are available, see Little and Rubin [17] for a comprehensive overview. Therefore, the clustering of multivariate data with missing information is one of the main reasons introduced FM-MMN model. Another the main reasons is reduce free parameters model by introducing 14 parsimonious structures via eigen-decomposition [9, 19]. To address the presence of missing values, two auxiliary indicator matrices are incorporated for methodological and theoretical developments in a more efficient manner. Based on a convenient hierarchical representation of the proposed model, we develop a computational feasible EM-based algorithm [11] to estimate maximum likelihood with almost all closed-form expressions for the E- and M-steps.

The paper layout is as follows. In section 2, we recall the MMN distribution with some preliminary theoretical properties and four special cases. Section 3 addresses the estimation and computational framework of the MMN mixture. For parameter estimation, we extend an EM-type algorithm of the Expectation Conditional Maximization either (ECME), represented by [18]. In section 4 and 5, the proposed methodology is demonstrated with extensive application of a real data set and simulation studies.

2. Model formulation

2.1. **The formulation of MMN distribution.** MMN formulation and some properties of its class of distributions are reviewed in this section to describe the motivation and parameter estimation better. Let us consider the p-dimensional random vector $\boldsymbol{Y} \in \mathbb{R}^p$ following an MMN distribution and denoted it by $\boldsymbol{Y} \sim \mathcal{MMN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ with mean vector $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$ and shape parameter $\boldsymbol{\lambda}$. Considering mixing cumulative distribution function (cdf) $H(\cdot;\boldsymbol{\nu})$, the pdf of an MMN distribution as

$$f_{\text{MMN}}(\boldsymbol{y};\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \phi(\boldsymbol{y};\boldsymbol{\mu}+\boldsymbol{\lambda}w,\boldsymbol{\Sigma})h(w;\boldsymbol{\nu})\,dw, \quad \boldsymbol{y} \in \mathbb{R}^p, \tag{2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu})$. Then, a p-variate random vector $\boldsymbol{Y}$ is said to have a MMN distribution if it can be generated through the linear stochastic representation

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\lambda}W + \boldsymbol{Z}, \qquad \boldsymbol{Z} \perp W, \tag{3}$$

where $\boldsymbol{Z}$ follows the p-variate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, W is an arbitrary random variable with cdf $H(\cdot;\boldsymbol{\nu})$ and the symbol '$\perp$' denotes the independence of two random variables.

2.2. **Special case of the MMN distribution.**

- Convolution with truncated normal distribution. Let W be a random variable in (3) with the truncated standard normal distribution lying within a truncated interval $(0,\infty)$, $W \sim \mathcal{TN}(0,1;(0,\infty))$. Therefore, the pdf of restricted skew-normal (rSN) distribution can be obtained as

$$f_{\text{rSN}}(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda}) = 2\phi_p(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Omega})\Phi\left(\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}{\sqrt{1-\boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}}}\right), \quad \boldsymbol{y} \in \mathbb{R}^p, \tag{4}$$

  where $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top$, and $\Phi(\cdot)$ is the cdf of the univariate standard normal distribution, $\mathcal{N}(0,1)$. The notation $\boldsymbol{Y} \sim r\mathcal{SN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ will be used if $\boldsymbol{Y}$ has pdf (4).

An extension of rSN distribution can be postulated by considering $W \sim \mathcal{TN}(0,1;(-\nu,\infty))$. This leads to obtain the pdf of Y, called ErSN and denoted by $\boldsymbol{Y} \sim \mathcal{E}r\mathcal{SN}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)$, as

$$f_{\text{ErSN}}(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu) = \frac{\phi_p(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Omega})}{\Phi(\nu)}\Phi\left(\frac{\nu+\boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}{\sqrt{1-\boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}}}\right), \quad \boldsymbol{y}\in\mathbb{R}^p. \tag{5}$$

- Convolution with the two-piece normal distribution. Let W be a random variable in (3) with the pdf

$$f(w;\nu) = \phi(\frac{w}{1+\nu})I_{[0,\infty)}(w) + \phi(\frac{w}{1-\nu})I_{(-\infty,0)}(w), \qquad \nu\in(-1,1), \tag{6}$$

where $I_A(\cdot)$ denotes the indicator function of set A and $\phi(\cdot)$ is the pdf of $\mathcal{N}(0,1)$. Therefore, the pdf of two-piece mixing rSN (TPrSN) distribution, denoted by $\boldsymbol{Y} \sim \mathcal{TP}r\mathcal{SN}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)$, can be obtained as

$$\begin{aligned} f_{\text{TPrSN}}(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu) &= \frac{(1+\nu)}{2}f_{\text{rSN}}(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},(1+\nu)\boldsymbol{\lambda}) \\ &+ \frac{(1-\nu)}{2}f_{\text{rSN}}(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},(\nu-1)\boldsymbol{\lambda}), \quad \boldsymbol{y}\in\mathbb{R}^p, \;\; \nu\in(-1,1). \end{aligned} \tag{7}$$

As expected, a mixture of two rSN distributions formulated pdf (7).

- Convolution with a exponential distribution. Let W be a random variable in (3) with the exponential distribution of mean $1/\nu$, denoted by $\mathcal{E}(\nu)$. Therefore, the pdf of exponential MMN (MMNE) distribution, denoted by $\boldsymbol{Y} \sim \mathcal{MMNE}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)$, can be obtained as

$$f_{\text{MMNE}}(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu) = \frac{\nu\sqrt{2\pi}}{\tau}\exp\left\{\frac{A_E^2}{2}\right\}\phi_p(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi(A_E), \quad \boldsymbol{y}\in\mathbb{R}^p, \tag{8}$$

where $\tau^2 = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}$, and $A_E = \tau^{-1}\big[\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})-\nu\big]$.

- Convolution with a mixture of exponential and half-normal distribution. Let W be a random variable in (3) with the pdf

$$f(w;\nu) = \nu_1\nu_2\exp\{-\nu_2 w\} + 2(1-\nu_1)\phi(w), \quad w,\nu_2>0, \quad 0<\nu_1<1. \tag{9}$$

Therefore, the pdf of half-normal exponentiated MMN (MMNEH) distribution, denoted by $\boldsymbol{Y} \sim \mathcal{MMNEH}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)$, can be obtained as

$$\begin{aligned} f_{\text{MMNEH}}(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\nu}) =& \nu_1\nu_2\frac{\sqrt{2\pi}}{\tau}\exp\left\{\frac{A_{EH}^2}{2}\right\}\phi_p(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi(A_{EH}) \\ &+ (1-\nu_1)f_{\text{rSN}}(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda}), \quad \boldsymbol{y}\in\mathbb{R}^p, \end{aligned}$$

where $A_{EH} = \tau^{-1}\big[\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})-\nu_2\big]$. As can be expected, the pdf above is formulated by a mixture of the MMNE and rSN distributions.

3. FM-MMN models with incomplete data

3.1. **The general model.** Let $\boldsymbol{Y}_j$ be a p-dimensional random vector of the jth individual for $j = 1, \ldots, n$ from a FM-MMN distributions. The pdf $\boldsymbol{Y}_j$s can be formulated as

$$f(\boldsymbol{y}_j; \boldsymbol{\Theta}) = \sum_{i=1}^{g} \pi_i f_{MMN_p}(\boldsymbol{y}_j; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\lambda}_i, \boldsymbol{\nu}_i), \quad i = 1, \ldots, g, \tag{10}$$

where π_i's are mixing probabilities sum of which equals one, $\boldsymbol{\Theta}$ denotes the model parameters, and $f_{MMN_p}(\boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ is the MMN density defined in (2). Define the latent membership-indicator vectors, $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$, where $\boldsymbol{Z}_j = (Z_{1j}, \ldots, Z_{gj})$ with $Z_{ij} = 1$ if jth observation belongs to the ith component and 0 otherwise. FM-MMN models can be represented hierarchically at the following three levels:

$$\begin{aligned} \boldsymbol{Y}_j|(w_j, z_{ij} = 1) &\sim \mathcal{N}_p(\boldsymbol{\mu}_i + \boldsymbol{\lambda}_i w_j, \boldsymbol{\Sigma}_i), \\ W_j|(z_{ij} = 1) &\sim h(w_j|\boldsymbol{\nu}_i), \\ \boldsymbol{Z}_j &\sim \mathcal{M}(1; \pi_1, \ldots, \pi_g). \end{aligned} \tag{11}$$

This paper examines the case where uncontrolled nonresponses cause missing values. For the FM-MMN model to be formulated with incomplete data, we consider two matrices $\boldsymbol{O}_j$ $(p_j^o \times p)$ and $\boldsymbol{M}_j((p - p_j^o) \times p)$ for partition $\boldsymbol{Y}_j(p \times 1)$. Therefore, we have used these two matrices $\boldsymbol{Y}_j^o = \boldsymbol{O}_j\boldsymbol{Y}_j$ and $\boldsymbol{Y}_j^m = \boldsymbol{M}_j\boldsymbol{Y}_j$, observed and missing parts of $\boldsymbol{Y}_j$, respectively.

The following proposition presents some significant consequences, which help to obtain the Q-function of the ECME algorithm.

Proposition 3.1. *If the conditions of relation* (11) *hold then:*

(a) *The marginal distribution of the observed component $\boldsymbol{Y}_j^o$ is given $z_{ij} = 1$*

$$\boldsymbol{Y}_j^o|(z_{ij} = 1) \sim \mathcal{MMN}_{p_j^o}(\boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, \boldsymbol{\lambda}_{ij}^o, \boldsymbol{\nu}_i), \tag{12}$$

where $\boldsymbol{\mu}_{ij}^o = \boldsymbol{O}_j\boldsymbol{\mu}_i$, $\boldsymbol{\lambda}_{ij}^o = \boldsymbol{O}_j\boldsymbol{\lambda}_i$ and $\boldsymbol{\Sigma}_{ij}^{oo} = \boldsymbol{O}_j\Sigma_i\boldsymbol{O}_j^\top$.

(b) *The conditional distribution of $\boldsymbol{Y}_j^o$ given w_j and $z_{ij} = 1$ is*

$$\boldsymbol{Y}_j^o \mid (w_j, z_{ij} = 1) \sim \mathcal{N}_{p_j^o}(\boldsymbol{\mu}_{ij}^o + w_j\boldsymbol{\lambda}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}),$$

(c) *The conditional distribution of $\boldsymbol{Y}_j^m$ given $\boldsymbol{y}_j^o$, w_j and $z_{ij} = 1$ is*

$$\boldsymbol{Y}_j^m \mid (\boldsymbol{y}_j^o, w_j, z_{ij} = 1) \sim \mathcal{N}_{p-p_j^o}(\boldsymbol{\varphi}_{ij}^{m.o}, \boldsymbol{\Sigma}_{ij}^{mm.o}),$$

where $\boldsymbol{\varphi}_j^{m.o} = \boldsymbol{M}_j\left[\boldsymbol{\mu}_i + w_j\boldsymbol{\lambda}_i + \boldsymbol{\Sigma}_i\boldsymbol{S}_{ij}^{oo}(\boldsymbol{y}_j - \boldsymbol{\mu}_i - w_j\boldsymbol{\lambda}_i)\right]$, $\boldsymbol{\Sigma}_{ij}^{mm.o} = \boldsymbol{M}_j\left(\boldsymbol{I}_p - \boldsymbol{\Sigma}_i\boldsymbol{S}_{ij}^{oo}\right)\boldsymbol{\Sigma}_i\boldsymbol{M}_j^\top$, and $\boldsymbol{S}_{ij}^{oo} = \boldsymbol{O}_j^\top\left(\boldsymbol{O}_j\boldsymbol{\Sigma}_i\boldsymbol{O}_j^\top\right)^{-1}\boldsymbol{O}_j$.

Proposition 3.2. (a) *Let* $\boldsymbol{Y}_j^o \sim \mathcal{E}r\mathcal{SN}_{p_j^o}(\boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, \boldsymbol{\lambda}_{ij}^o, \nu_i)$ *and* $W \sim \mathcal{TN}(0,1;(-\nu_i,\infty))$. *Then,* W_j *conditionally on* $\boldsymbol{Y}_j^o = \boldsymbol{y}_j^o$ *and* $Z_{ij}=1$*, denoted by* $W_{\boldsymbol{y}_j^o}$*, follows* $W_{\boldsymbol{y}_j^o} \sim \mathcal{TN}\left(\xi, \omega^2; (0,\infty)\right)$*, where* $\xi = \nu_i + \boldsymbol{\lambda}_{ij}^{o\top} \boldsymbol{\Omega}_{ij}^{oo-1}(\boldsymbol{y}_j^o - \boldsymbol{\mu}_{ij}^o)$ *and* $\omega^2 = 1 - \boldsymbol{\lambda}_{ij}^{o\top}\boldsymbol{\Omega}_{ij}^{oo-1}\boldsymbol{\lambda}_{ij}^o$ *for* $\boldsymbol{\Omega}_{ij}^{oo} = \boldsymbol{\Sigma}_{ij}^{oo} + \boldsymbol{\lambda}_{ij}^o\boldsymbol{\lambda}_{ij}^{o\top}$.

(b) *Let* $\boldsymbol{Y}_j^o \sim \mathcal{TP}r\mathcal{SN}_{p_j^o}(\boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, \boldsymbol{\lambda}_{ij}^o, \nu_i)$ *and* W_j *have a density in* (6). *Then,* $W_{\boldsymbol{y}_j^o}$ *has pdf*

$$f_{W_{\boldsymbol{y}_j^o}}(w) = \pi(\boldsymbol{y}_j^o)\frac{\phi\left(w;\xi_1,\omega_1^2\right)}{\Phi(\xi_1/\omega_1)} I_{[0,\infty)}(w) + (1-\pi(\boldsymbol{y}_j^o))\frac{\phi\left(w;\xi_2,\omega_2^2\right)}{\Phi(-\xi_2/\omega_2)} I_{(-\infty,0)}(w),$$

where $\xi_1 = (1+\nu_i)^2\boldsymbol{\lambda}_{ij}^{o\top}\boldsymbol{\Omega}_{1ij}^{oo\,-1}(\boldsymbol{y}_j^o - \boldsymbol{\mu}_{ij}^o)$ *and* $\omega_1^2 = (1+\nu_i)^2 - (1+\nu_i)^4\boldsymbol{\lambda}_{ij}^{o\top}\boldsymbol{\Omega}_{1ij}^{oo\,-1}\boldsymbol{\lambda}_{ij}^o$, $\boldsymbol{\Omega}_{1j}^{oo} = \boldsymbol{\Sigma}_{ij}^{oo} + (1+\nu_i)^2\boldsymbol{\lambda}_{ij}^o\boldsymbol{\lambda}_{ij}^{o\top}$, $\xi_2 = (1-\nu_i)^2\boldsymbol{\lambda}_{ij}^{o\top}\boldsymbol{\Omega}_{2ij}^{oo\,-1}(\boldsymbol{y}_j^o - \boldsymbol{\mu}_{ij}^o)$ *and* $\omega_2^2 = (1-\nu_i)^2 - (1-\nu_i)^4\boldsymbol{\lambda}_{ij}^{o\top}\boldsymbol{\Omega}_{1ij}^{oo\,-1}\boldsymbol{\lambda}_{ij}^o$, $\boldsymbol{\Omega}_{2j}^{oo} = \boldsymbol{\Sigma}_{ij}^{oo} + (1-\nu_i)^2\boldsymbol{\lambda}_{ij}^o\boldsymbol{\lambda}_{ij}^{o\top}$*, and*

$$\pi(\boldsymbol{y}_j^o) = \frac{(1+\nu_i) f_{rSN}(\boldsymbol{y}_j^o;\boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, (1+\nu)\boldsymbol{\lambda}_{ij}^o)}{2 f_{TPrSN}(\boldsymbol{y}_j^o;\boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, \boldsymbol{\lambda}_{ij}^o, \nu_i)}.$$

Furthermore, for any $\boldsymbol{y}_j^o \in \mathbb{R}^{p_j^o}$*, and* $k = 1, 2, \ldots,$

$$E(W_{\boldsymbol{y}_j^o}^k) = \pi(\boldsymbol{y}_j^o) E\left(V_1^k\right) + (1-\pi(\boldsymbol{y}_j^o)) E(V_2^k),$$

where $V_1 \sim \mathcal{TN}\left(\xi_1, \omega_1^2; (0,\infty)\right)$, $V_2 \sim \mathcal{TN}\left(\xi_2,\omega_2^2;(-\infty,0)\right)$.

(c) *Let* $\boldsymbol{Y}_j^o \sim \mathcal{MMNE}_{p_j^o}(\boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, \boldsymbol{\lambda}_{ij}^o, \nu_i)$ *and* $W_j \sim \mathcal{E}(\nu_i)$. *Then,* $W_{\boldsymbol{y}_j^o} \sim \mathcal{TN}\left(A_{Eij}^o \tau_{ij}^{o\,-1}, \tau_{ij}^{o\,-2}; (0,\infty)\right)$ *where* $\tau_{ij}^{o\,2} = \boldsymbol{\lambda}_{ij}^{o\top}\boldsymbol{\Sigma}_{ij}^{oo-1}\boldsymbol{\lambda}_{ij}^o$ *and* $A_{Eij}^o = \tau_i^{o-1}\left[\boldsymbol{\lambda}_{ij}^{o\top}\boldsymbol{\Sigma}_{ij}^{oo-1}(\boldsymbol{y}_j^o - \boldsymbol{\mu}_{ij}^o) - \nu_i\right]$.

(g) *Let* $\boldsymbol{Y}_j^o \sim \mathcal{MMNEH}_{p_j^o}(\boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, \boldsymbol{\lambda}_{ij}^o, \boldsymbol{\nu}_i)$ *with* $\boldsymbol{\nu}_i = (\nu_{1i}, \nu_{2j})$ *and* W_j *have a density in* (9). *Then,* $W_{\boldsymbol{y}_j^o}$ *has pdf*

$$f_{W_{\boldsymbol{y}_j^o}}(w) = \pi(\boldsymbol{y}_j^o)\frac{\phi\left(w; A_{EHij}^o\tau_{ij}^{o\,-1}, \tau_{ij}^{o\,-2}\right)}{\Phi(A_{EHij}^o)} + (1-\pi(\boldsymbol{y}_j^o))\frac{\phi\left(w;\xi,\omega^2\right)}{\Phi(\xi/\omega)},$$

where

$$\pi(\boldsymbol{y}_j^o) = \frac{\nu_{1i}\nu_{2j}\sqrt{2\pi}}{2\tau_{ij}^o f_{MMNEH_{p_j^o}}(\boldsymbol{\mu}_j^o, \boldsymbol{\Sigma}_j^{oo}, \boldsymbol{\lambda}_j^o, \nu)}\phi_p\left(\boldsymbol{y};\boldsymbol{\mu}_j^o, \boldsymbol{\Sigma}_j^{oo}\right) \times \exp\left(\frac{A_{EHij}^o}{2}\right)\Phi(A_{EHij}^o),$$

where $A_{EHij}^o == \tau_{ij}^{o\,-1}\left[\boldsymbol{\lambda}_{ij}^{o\top}\boldsymbol{\Sigma}_{ij}^{oo-1}(\boldsymbol{y}_j^o - \boldsymbol{\mu}_{ij}^o) - \nu_{2i}\right]$. *Furthermore, for any* $\boldsymbol{y}_j^o \in \mathbb{R}^{p_j^o}$*, and* $k = 1, 2, \ldots,$

$$E(W_{\boldsymbol{y}}^k) = \pi(\boldsymbol{y}) E\left(V_1^k\right) + (1-\pi(\boldsymbol{y})) E(V_2^k),$$

where $V_1 \sim \mathcal{TN}\left(A_{EHij}^o\tau_{ij}^{o\,-1}, \tau_{ij}^{o\,-2}; (0,\infty)\right)$, $V_2 \sim \mathcal{TN}\left(\xi,\omega^2;(0,\infty)\right)$.

3.2. **Parameter estimation.** In this section, the parameter estimation of the FM-MMN is carried out via an ECME algorithm. [18] considered the ECME algorithm as an extension of the expectation maximization algorithm [20]. Several properties of this algorithm include stable features, implementation simplicity, and monotone convergence. To simplify notation, we denote the complete data by $\boldsymbol{y}_c = (\boldsymbol{y}^o, \boldsymbol{y}^m, \boldsymbol{U}, \boldsymbol{W}, \boldsymbol{Z})$, where $\boldsymbol{y}^o = (\boldsymbol{y}_1^o, \dots, \boldsymbol{y}_n^o)$, $\boldsymbol{y}^m = (\boldsymbol{y}_1^m, \dots, \boldsymbol{y}_n^m)$, $\boldsymbol{W} = (W_1, \dots, W_n)$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n)$. From (11), the log-likelihood function of $\boldsymbol{\Theta}$ for $\boldsymbol{y}_c$, without considering constant terms, is

$$\begin{aligned}\ell(\boldsymbol{\Theta}|\boldsymbol{y}_c) = &\sum_{i=1}^{g}\sum_{j=1}^{n} z_{ij}\left[\log \pi_i + \log h(w_j|\boldsymbol{\nu}_i) - \frac{1}{2}\log \boldsymbol{\Sigma}_i\right] \\ &- \frac{1}{2}\mathtt{tr}\Bigg\{\sum_{i=1}^{g}\boldsymbol{\Sigma}_i^{-1}\sum_{j=1}^{n} z_{ij}\Big[(\boldsymbol{y}_j - \boldsymbol{\mu}_i)(\boldsymbol{y}_j - \boldsymbol{\mu}_i)^\top \\ &- w_j(\boldsymbol{y}_j - \boldsymbol{\mu}_i)\boldsymbol{\lambda}_i^\top - w_j\boldsymbol{\lambda}_i(\boldsymbol{y}_j - \boldsymbol{\mu}_i)^\top + w_j^2\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i^2\Big]\Bigg\}. \end{aligned} \tag{13}$$

We then have the following conditional expectations.

$$\begin{aligned}\hat{z}_{ij}^{(k)} =& \frac{\hat{\boldsymbol{\pi}}_i^{(k)} f_{\mathrm{MMN}_{p_j^o}}(\boldsymbol{y}_j^o; \boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, \boldsymbol{\lambda}_{ij}^o, \boldsymbol{\nu}_i)}{\sum_{h=1}^{g}\hat{\boldsymbol{\pi}}_h^{(k)} f_{\mathrm{MMN}_{p_j^o}}(\boldsymbol{y}_j^o; \boldsymbol{\mu}_{hj}^o, \boldsymbol{\Sigma}_{hj}^{oo}, \boldsymbol{\lambda}_{hj}^o, \boldsymbol{\nu}_h)}, \\ \hat{w}_{rij}^{(k)} =& E(W_j^r \mid \boldsymbol{y}_j^o, z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}), \quad \hat{\boldsymbol{\Psi}}_{ij}^{(k)} = E(\log h(W_j;\nu_i) \mid \boldsymbol{y}_j^o, z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}),\end{aligned} \tag{14}$$

for $j = 1, \dots, n$ and $i = 1, \dots, g$. Evaluations can be conducted using previous Propositions 3.2.The ECME algorithm used to obtain the ML estimate of the FM-MMN distributions iterates the below E- and CM-steps as follows.

- E-step: Calculate

$$\begin{aligned}Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}) = &\sum_{i=1}^{g}\hat{n}_i^{(k)}\log \pi_i + \sum_{i=1}^{g}\sum_{j=1}^{n}\hat{z}_{ij}^{(k)}\widehat{\boldsymbol{\Psi}}_{ij}^{(k)} \\ &- \frac{1}{2}\sum_{i=1}^{g}\hat{n}_i^{(k)}\log \boldsymbol{\Sigma}_i - \frac{1}{2}\sum_{i=1}^{g}\mathtt{tr}\left(\widehat{\boldsymbol{\Upsilon}}_{ij}^{(k)}\right),\end{aligned} \tag{15}$$

where $\hat{n}_i^{(k)} = \sum_{j=1}^{n}\hat{z}_{ij}^{(k)}$, $\widehat{\boldsymbol{\Upsilon}}_{ij}^{(k)} = E\Big(\big(\boldsymbol{Y}_j - \boldsymbol{\mu}_i - W_j\boldsymbol{\lambda}_i\big)\big(\boldsymbol{Y}_j - \boldsymbol{\mu}_i - W_j\boldsymbol{\lambda}_i\big)^\top \mid \boldsymbol{y}_j^o, z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\Big)$, and the necessary conditional expectations obtained by Proposition 3.1 and 3.2 are

$$
\begin{aligned}
\hat{z}_{ij}^{(k)} =& \frac{\hat{\boldsymbol{\pi}}_i^{(k)} f_{\mathrm{MMN}_{p_j^o}}(\boldsymbol{y}_j^o; \boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, \boldsymbol{\lambda}_{ij}^o, \boldsymbol{\nu}_i)}{\sum_{h=1}^g \hat{\boldsymbol{\pi}}_h^{(k)} f_{\mathrm{MMN}_{p_j^o}}(\boldsymbol{y}_j^o; \boldsymbol{\mu}_{hj}^o, \boldsymbol{\Sigma}_{hj}^{oo}, \boldsymbol{\lambda}_{hj}^o, \boldsymbol{\nu}_h)}, \\
\hat{w}_{rij}^{(k)} =& E(W_j^r \mid \boldsymbol{y}_j^o, z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}), \\
\hat{\boldsymbol{\Psi}}_{ij}^{(k)} =& E(\log h(W_j; \nu_i) \mid \boldsymbol{y}_j^o, z_{ij} = 1, \hat{\boldsymbol{\Theta}}^{(k)}),
\end{aligned} \tag{16}
$$

- CM-step 1: Maximizing (15) over π_i, $\boldsymbol{\mu}_i$, $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Sigma}_i$ leads to the following CM estimators:

$$
\begin{aligned}
\hat{\boldsymbol{\pi}}_i^{(k+1)} =& \frac{\hat{n}_i^{(k)}}{n}, \\
\hat{\boldsymbol{\mu}}_i^{(k+1)} =& \frac{\sum_{j=1}^n \hat{z}_{ij}^{(k)} \hat{\boldsymbol{q}}_{ij}^{oo^{(k)}} - \hat{\boldsymbol{\Sigma}}_i^{(k)} \sum_{j=1}^n \hat{z}_{ij}^{(k)} \hat{w}_{1ij}^{(k)} \hat{\boldsymbol{S}}_{ij}^{oo^{(k)}} \hat{\boldsymbol{\lambda}}_i^{(k)}}{\hat{n}_i^{(k)}}, \\
\hat{\boldsymbol{\lambda}}_i^{(k+1)} =& \frac{\sum_{j=1}^n \left[\hat{z}_{ij}^{(k)} \hat{w}_{2ij}^{(k)} \hat{\boldsymbol{E}}_{ij}^{oo^{(k)}} + \hat{z}_{ij}^{(k)} \hat{w}_{1ij}^{(k)} (\hat{\boldsymbol{q}}_{ij}^{oo^{(k)}} - \hat{\boldsymbol{\mu}}_i^{(k+1)}) \right]}{\sum_{j=1}^n \hat{z}_{ij}^{(k)} \hat{w}_{2ij}^{(k)}}, \\
\hat{\boldsymbol{\Sigma}}_i^{(k+1)} =& \frac{\sum_{j=1}^n \hat{z}_{ij}^{(k)} \hat{\boldsymbol{\Upsilon}}_{ij}^{(k+1)}}{\hat{n}_i^{(k)}},
\end{aligned} \tag{17}
$$

where $\hat{\boldsymbol{q}}_{ij}^{oo^{(k)}} = \hat{\boldsymbol{\mu}}_i^{(k)} + \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\boldsymbol{S}}_{ij}^{oo^{(k)}} (\boldsymbol{y}_j - \hat{\boldsymbol{\mu}}_i^{(k)})$, $\hat{\boldsymbol{E}}_j^{oo^{(k)}} = \left(\boldsymbol{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\boldsymbol{S}}_{ij}^{oo^{(k)}} \right) \hat{\boldsymbol{\lambda}}_i^{(k)}$ and

$$
\begin{aligned}
\hat{\boldsymbol{\Upsilon}}_{ij}^{(k+1)} =& \left(\hat{\boldsymbol{q}}_{ij}^{oo^{(k)}} - \hat{\boldsymbol{\mu}}_i^{(k+1)} \right) \left(\hat{\boldsymbol{q}}_{ij}^{oo^{(k)}} - \hat{\boldsymbol{\mu}}_i^{(k+1)} \right)^\top + \left(\boldsymbol{I}_p - \hat{\boldsymbol{\Sigma}}_i^{(k)} \hat{\boldsymbol{S}}_{ij}^{oo^{(k)}} \right) \hat{\boldsymbol{\Sigma}}_i^{(k)} \\
&+ \hat{w}_{2ij}^{(k)} \left(\hat{\boldsymbol{E}}_{ij}^{oo^{(k)}} - \hat{\boldsymbol{\lambda}}_i^{(k+1)} \right) \left(\hat{\boldsymbol{E}}_{ij}^{oo^{(k)}} - \hat{\boldsymbol{\lambda}}_i^{(k+1)} \right) \\
&+ \hat{w}_{1ij}^{(k)} \left(\hat{\boldsymbol{q}}_{ij}^{oo^{(k)}} - \hat{\boldsymbol{\mu}}_i^{(k+1)} \right) \left(\hat{\boldsymbol{E}}_{ij}^{oo^{(k)}} - \hat{\boldsymbol{\lambda}}_i^{(k+1)} \right)^\top \\
&+ \hat{w}_{1ij}^{(k)} \left(\hat{\boldsymbol{E}}_{ij}^{oo^{(k)}} - \hat{\boldsymbol{\lambda}}_i^{(k+1)} \right) \left(\hat{\boldsymbol{q}}_{ij}^{oo^{(k)}} - \hat{\boldsymbol{\mu}}_i^{(k+1)} \right)^\top.
\end{aligned}
$$

- CML-step 2: In the light of (11), when the $\boldsymbol{\nu}_i$s are assumed to be unequal, the updated estimate of $\boldsymbol{\nu}_i$ is obtained as

$$
\hat{\boldsymbol{\nu}}_i^{(k+1)} = \arg\max_{\boldsymbol{\nu}_i} \sum_{j=1}^n \hat{z}_{ij}^{(k)} f_{\mathrm{MMN}_{p_j^o}} \left(\boldsymbol{y}_j^o; \hat{\boldsymbol{\mu}}_{ij}^{o^{(k+1)}}, \hat{\boldsymbol{\Sigma}}_{ij}^{oo^{(k+1)}}, \hat{\boldsymbol{\lambda}}_{ij}^{o^{(k+1)}}, \boldsymbol{\nu}_i \right), \tag{18}
$$

where $f_{\mathrm{MMN}_{p_j^o}} \left(\boldsymbol{y}_j^o; \cdots \right)$ is the MMN pdf as defined in (2). In the case where the $\boldsymbol{\nu}_i$s are constrained to be identical, say, we update $\boldsymbol{\nu}_1 = \cdots = \boldsymbol{\nu}_g = \boldsymbol{\nu}$ by maximizing the constrained actual observed log-likelihood function, namely

$$\hat{\nu}^{(k+1)} = \arg\max_{\boldsymbol{\nu}} \sum_{i=1}^{g}\sum_{j=1}^{n} \left\{ \hat{\pi}_i^{(k+1)} \log f_{\mathrm{MMN}_{p_j^o}} \left(\boldsymbol{y}_j^o; \hat{\boldsymbol{\mu}}_{ij}^{o^{(k+1)}}, \hat{\boldsymbol{\Sigma}}_{ij}^{oo^{(k+1)}}, \hat{\boldsymbol{\lambda}}_{ij}^{o^{(k+1)}}, \boldsymbol{\nu}_i \right) \right\},$$

where $\hat{\boldsymbol{\mu}}_{ij}^{o^{(k+1)}}$, $\hat{\boldsymbol{\Sigma}}_{ij}^{oo^{(k+1)}}$, and $\hat{\boldsymbol{\lambda}}_{ij}^{o^{(k+1)}}$ are $\hat{\boldsymbol{\mu}}_{ij}^{o}$, $\hat{\boldsymbol{\Sigma}}_{ij}^{oo}$, and $\hat{\boldsymbol{\lambda}}_{ij}^{o}$ in Proposition 3.1, respectively, evaluated at the current estimates at the start of the $(k+1)$th iteration.

In short, the update of $\boldsymbol{\nu}_i$ for the special cases of the FM-MMN can be obtained as

$$\hat{\nu}_i^{(k+1)} = \begin{cases} \text{Free of parameter} & \text{for rSN;} \\ \text{Obtained by (18)} & \text{for ErSN;} \\ 1/\hat{\bar{w}}_{1i} & \text{for MMNE;} \\ \text{Obtained by (18)} & \text{for TPrSN;} \end{cases}$$

where $\hat{\bar{w}}_{1i} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{w}_{1ij}^{(k)}}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}$ and $\hat{\bar{w}}_{2i} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{w}_{2ij}^{(k)}}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}$.

Remark 3.3. For the update estimates of $\boldsymbol{\nu}_i = (\nu_{1i}, \nu_{2i})$, we can circumvent this difficulty by introducing a binary variable V_{ij} as the second source of missing data into the representation (11) as

$$\begin{aligned} \boldsymbol{Y}_j^o|(W_j = w_j, Z_{ij} = 1, V_{ij} = 1) &\sim \mathcal{N}_{p_j^o}(\boldsymbol{\mu}_{ij}^o + w_j \boldsymbol{\lambda}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}), \\ W_j|(Z_{ij} = 1, V_{ij} = 1) &\sim \mathcal{E}(\nu_{2i}), \\ \boldsymbol{V}_j|(Z_{ij} = 1) &\sim \mathcal{B}(1, \nu_{1i}), \\ \boldsymbol{Z}_j &\sim \mathcal{M}(1; \pi_1, \dots, \pi_g). \end{aligned} \tag{19}$$

where $\mathcal{B}(1, \nu_{1i})$ denotes the Bernoulli trail with ν_{1i} probability. In this part, $V_{ij} = 1$ if $\boldsymbol{y}_j$ in group g is taken from the MMNEH distribution and $V_{ij} = 0$ if $\boldsymbol{x}_j^o$ in group g is generated by the rMSN model. According to representation (19), the update of parameter ν_i through the ECME algorithm can be computed as

$$\hat{\nu}_{1i}^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{v}_{ij}^{(k)}}{\hat{n}_i^{(k)}} \quad \text{and} \quad \hat{\nu}_{2i}^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{v}_{ij}^{(k)}}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{w}_{1ij}^{(k)} \hat{v}_{ij}^{(k)}}, \tag{20}$$

where

$$\hat{v}_{ij}^{(k)} = \frac{\nu_{1i}\nu_{2i}\sqrt{2\pi}\Phi(A_{ij}^o)}{2\tau_{ij}^o f_{\mathrm{MMNEH}}(\boldsymbol{y}_j^o; \boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo}, \boldsymbol{\lambda}_{ij}^o, \boldsymbol{\nu}_i)} \phi_p \left(\boldsymbol{y}_j^o; \boldsymbol{\mu}_{ij}^o, \boldsymbol{\Sigma}_{ij}^{oo} \right) \exp\left(\frac{A_{ij}^{o^2}}{2} \right). \tag{21}$$

3.3. **Parsimonious versions of the general model.** For the FM-MMN models, the covariance matrices of latent factors ($\boldsymbol{\Sigma}_i$) have $p(p+1)/2$ free parameters. By increasing the number of components, the model fitting the process could be affected by over-parameterization. For this potential problem, the parsimony in the multivariate normal and non-normal mixture models was introduced by [6]. They consider the eigenvalue decomposition parameterization on the component covariance matrices

$$\boldsymbol{\Sigma}_i = \eta_i \boldsymbol{\Gamma}_i \boldsymbol{\Delta}_i \boldsymbol{\Gamma}_i^\top, \tag{22}$$

where $\eta_i = |\boldsymbol{\Sigma}_i|^{1/p}$ known as the constant of proportionality, $\boldsymbol{\Delta}_i$ is a diagonal matrix of eigenvalues of $\boldsymbol{\Sigma}_i$ sorted from highest to lowest $|\boldsymbol{\Delta}_i| = 1$, and $\boldsymbol{\Gamma}_i$ is the orthogonal matrix consisting of eigenvectors of $\boldsymbol{\Sigma}_i$ that are ordered according to their eigenvalues. By applying various constraints to the eigenvalue decomposition of the covariance matrix (22), [9] introduced clustering using fourteen Gaussian parsimonious mixture models. The results are summarized in Table 1 in the supplemental file of this paper. Bear in mind that parameterization in (22) has different geometric interpretations. Indeed, η_i represents the volume of the ith cluster while $\boldsymbol{\Delta}_i$ and $\boldsymbol{\Gamma}_i$ determine the shape and orientation of the ith cluster, respectively.

3.4. **Predicting missing information.** To calculate predicting factor scores and missing information, denote the ML estimates by $\hat{\boldsymbol{\Theta}} = (\hat{\pi}_i, \hat{\mu}_i, \hat{\boldsymbol{\Sigma}}_i, \hat{\boldsymbol{\lambda}}_i, \hat{\boldsymbol{\nu}}_i)$, as a by-product of our ECME algorithm and Proposition 3.1 part (c), conditional imputation is used to estimate missing values as

$$\hat{\boldsymbol{y}}_j^m = \boldsymbol{M}_j \left[\hat{\boldsymbol{\mu}}_i + \hat{w}_{1ij} \hat{\boldsymbol{\lambda}}_i + \hat{\boldsymbol{\Sigma}}_i \hat{\boldsymbol{S}}_{ij}^{oo} (\boldsymbol{y}_j - \hat{\boldsymbol{\mu}}_i - w_j \hat{\boldsymbol{\lambda}}_i) \right], \tag{23}$$

where $\hat{\boldsymbol{S}}_{ij}^{oo} = \boldsymbol{O}_j^\top \left(\boldsymbol{O}_j \hat{\boldsymbol{\Sigma}}_i \boldsymbol{O}_j^\top \right)^{-1} \boldsymbol{O}_j$. We use the mean squared deviation (MSD) as a measure of the difference between the true value $\boldsymbol{y}_j^m$ and the imputed value $\hat{\boldsymbol{y}}_j^m$. MSD can be calculated as follows:

$$\text{MSD} = \frac{1}{n^*} \sum_{j=1}^{n} (\boldsymbol{y}_j^m - \hat{\boldsymbol{y}}_j^m)^\top (\boldsymbol{y}_j^m - \hat{\boldsymbol{y}}_j^m), \tag{24}$$

where n^* is the number of missing items.

3.5. **Notes on Implementation.** Like any EM-type algorithm, if the ECME algorithm is given reasonable parameter estimates, convergence may be sped up or made more manageable. When the raw data contains missing values, after filling in the missing values of kth variable with the mean of the corresponding column regardless of the missing values, for specifying. By using `R` function "`kmeans( )`", we classified datasets into g groups and used the number of data points belonging to the same cluster i division to a number of data points. For each group, create the initial value $\hat{\boldsymbol{\mu}}_i^{(0)}$ and $\hat{\boldsymbol{\Sigma}}_i^{(0)}$ as the mean and covariance of the data, respectively. Following [29], we considered $\hat{\boldsymbol{\lambda}}_i^{(0)} = \mathbf{0}$ and $\hat{\nu}_i^{(0)} = 1$ for ErSN and MMNE distributions, $\hat{\nu}_i^{(0)} = 0.5$ for TPrSN distribution, and $(\nu_1, \nu_2) = (0.5, 1)$ for MMNEH distribution to near-normality.

The Bayesian Information Criterion (BIC) [28] is used to select the number of classes and factors. We calculate as

$$\text{BIC} = -2\ell_{\max} + m \log n,$$

where m is the number of free parameters, and $\ell_{\max}$ is the maximized log-likelihood value.

As an alternative measure, the integrated completed likelihood (ICL) [7] is also suitable for measuring the number of clusters, defined as

$$\text{ICL} = \text{BIC} + 2\text{ENT}(\hat{\boldsymbol{z}}),$$

where $\text{ENT}(\hat{\boldsymbol{z}}) = -\sum_{i=1}^{g}\sum_{j=1}^{n}\hat{z}_{ij}\log\hat{z}_{ij}$. Models with smaller BIC values are generally better fitted. To measure the ability of clustering agreement, we employ the misclassification rate (MCR), the correct classification rate (CCR), and adjusted Rand index (ARI, [14]), which compensates for the shortcomings of the AR due to chance agreement. The model with the highest CCR and ARI score provides the most reliable classification accuracy.

For stopping rule of the ECME algorithm, the algorithm can be continued to reach $\ell(\hat{\boldsymbol{\Theta}}^{(k)};\boldsymbol{y}) - \ell(\hat{\boldsymbol{\Theta}}^{(k-1)};\boldsymbol{y}) < \epsilon$ at iteration (k), where $\ell(\hat{\boldsymbol{\Theta}}^{(k)};\boldsymbol{y})$ is the log-likelihood value evaluated with parameters estimation and ϵ is a pre-specified tolerance. When $k_{\max} = 2000$ iterations have been performed, or when the log-likelihood difference between two successive iterations is less than $\epsilon = 10^{-5}$, the algorithm terminates.

4. **Hepatitis disease data**

We applied our method on Hepatitis Data as an example, which consists of 20 attributes with missing information: 6 are continuous, 14 are categorical data . There are $n = 205$ samples in the data but no known grouping labels of the Automobile data. The data have been previously used by [10]. Before fitting, each attribute is normalized to have zero mean and unit variance.

Under the fact of unknown grouping labels for this data, at first, we fit 14 covariance structures the 6 continuous attributes of the data with $g = 2$ by finite mixture normal distribution (FM-N), finite mixture t distribution (FM-T), finite mixture restricted skew normal distribution (FM-rSN), and FM-MMN models and report the best BIC in Table 1. Table 1 summarizes the maximum likelihood results, containing the number of parameters and the value of BIC criteria. To compare clustering performance in the mentioned models, the classification settlement obtained by CCR and ARI are also presented in Table 1. The best value for each model is highlighted. According to this table, the VVE structure in FM-MMNE is the best model to fit this dataset (BIC=1851.787) and the best classification accuracy for this dataset is CCR = 0.854. Based on Figure 1, the FM-T, FM-rSN, and FM-MMN models via FM-N provide somewhat different imputations. Therefore, the FM-MMN models via FM-N has a different performance for analyzing missing values, but the FM-T via FM-N has similar performance. Thus, we are interested that displayed the scatters and contours plot of marginal bivariate fitted FM-MMN models. By using Figure 2, the marginal contours look skewed and have bi-modal and long

Table 1. Estimating the performance of finite mixture models for $g = 2$ fitted the Hepatitis disease data.

Model		EII	VII	EEI	VEI	EVI	VVI	EEE	VEE	EVE	EEV	VVE	VEV	EVV	VVV
FM-N	nu. par.	14	15	19	20	24	25	34	35	39	49	50	40	54	55
	BIC	2239.443	2129.149	2248.113	2124.229	2110.241	2018.718	2312.079	2166.883	2144.606	2175.037	2102.112	**2050.155**	2190.422	2096.033
	CCR	0.690	0.639	0.794	0.671	**0.806**	0.665	0.748	0.729	0.768	0.748	0.671	0.677	0.748	0.671
	ARI	0.027	0.043	0.111	0.113	**0.312**	0.104	0.009	0.152	0.206	0.147	0.096	0.113	0.174	0.091
FM-T	nu. par.	16	17	21	22	26	27	36	37	41	51	52	42	56	57
	BIC	2150.633	2115.213	2132.524	2096.465	2058.110	2014.366	2172.947	2149.686	2097.031	2118.803	2102.452	**2054.736**	2133.293	2099.617
	CCR	0.742	0.626	0.781	0.606	**0.813**	0.626	0.768	0.684	0.732	0.735	0.723	0.671	0.723	0.671
	ARI	0.225	0.055	0.294	0.034	**0.355**	0.057	0.168	0.106	0.272	0.154	0.150	0.091	0.143	0.091
FM-rSN	nu. par.	26	27	31	32	36	37	46	47	41	61	62	52	66	67
	BIC	2145.523	2027.592	2083.025	1956.276	1907.485	**1888.259**	1978.419	2018.791	1973.846	1973.253	1909.528	1906.201	1910.482	1923.768
	CCR	0.755	0.613	**0.826**	0.723	0.677	0.658	0.794	0.806	0.761	0.748	0.703	0.652	0.748	0.716
	ARI	0.240	0.046	**0.341**	0.193	0.107	0.085	0.278	0.271	0.157	0.157	0.118	0.075	0.189	0.134
FM-ErSN	nu. par.	27	28	32	33	37	38	47	48	42	62	63	53	67	68
	BIC	2131.461	2083.483	1972.711	1997.626	1897.207	**1888.577**	2210.075	1987.353	1916.076	1942.718	1901.420	1908.176	1912.664	1889.950
	CCR	0.748	0.632	0.742	0.710	0.677	0.658	**0.800**	0.768	0.690	0.794	0.768	0.677	0.774	0.716
	ARI	0.229	0.063	0.217	0.170	0.107	0.085	**0.321**	0.206	0.118	0.254	0.228	0.107	0.233	0.141
FM-MMNE	nu. par.	27	28	32	33	37	38	47	48	42	62	63	53	67	68
	BIC	2150.047	2122.769	2050.394	1951.412	1878.431	1880.659	2213.743	1950.240	1861.305	1888.999	**1851.787**	1896.631	1879.966	1854.898
	CCR	0.826	0.761	0.794	0.671	0.645	0.645	0.826	0.768	0.794	0.787	**0.854**	0.665	0.800	0.781
	ARI	0.257	0.255	0.312	0.113	0.075	0.075	0.349	0.214	0.291	0.250	**0.370**	0.100	0.284	0.252
FM-MMNEH	nu. par.	28	29	33	34	38	39	48	49	43	63	64	54	68	69
	BIC	2114.719	2136.396	2024.350	1943.354	1880.431	1882.728	2203.031	1921.970	**1881.079**	1933.135	1921.718	1896.068	1918.814	1906.826
	CCR	0.755	0.774	0.794	0.748	0.645	0.645	**0.826**	0.781	0.677	0.761	0.710	0.671	0.671	0.677
	ARI	0.243	0.233	0.262	0.217	0.075	0.075	**0.356**	0.220	0.118	0.195	0.112	0.108	0.102	0.098
FM-TPrSN	nu. par.	27	28	32	33	37	38	47	48	42	62	63	53	67	68
	BIC	2147.533	2102.035	1991.359	2001.968	1925.336	**1892.201**	2222.080	1961.113	1942.107	1961.563	1912.485	1908.932	1955.524	1914.688
	CCR	0.755	0.632	0.755	0.594	0.677	0.671	**0.813**	0.716	0.697	0.729	0.665	0.677	0.703	0.665
	ARI	0.240	0.063	0.246	0.029	0.103	0.099	**0.339**	0.152	0.126	0.152	0.085	0.107	0.134	0.085

tails, which indicates that the FM-MMN models may be more appropriate for capturing this dataset.

For evaluating the performance of the asymmetric FM-MMN models in revealing group structures in data, the confusion matrices of each data set are presented in Figure 3. From Figure 3, it is clear that the asymmetric FM-MMN models yield excellent clustering performance in the case studies considered.

5. Simulation study

In this section, simulation studies were used to investigate the parameters estimation's asymptotic properties according to ECME algorithm and the introduced models' robustness in dealing with non-normal and thick-tailed data.

5.1. Asymptotic properties of parameters estimation.

In this experiment, we obtain the performance of the proposed ML estimations based on ECME algorithm to regain the true parameters of four specific cases of the FM-MMN model. We consider non-normal data with $n = 100, 200, 400, 800$ that are generated from four special cases of FM-MMN with two components ($g = 2$) and structure VEE. The common true parameters for these models are

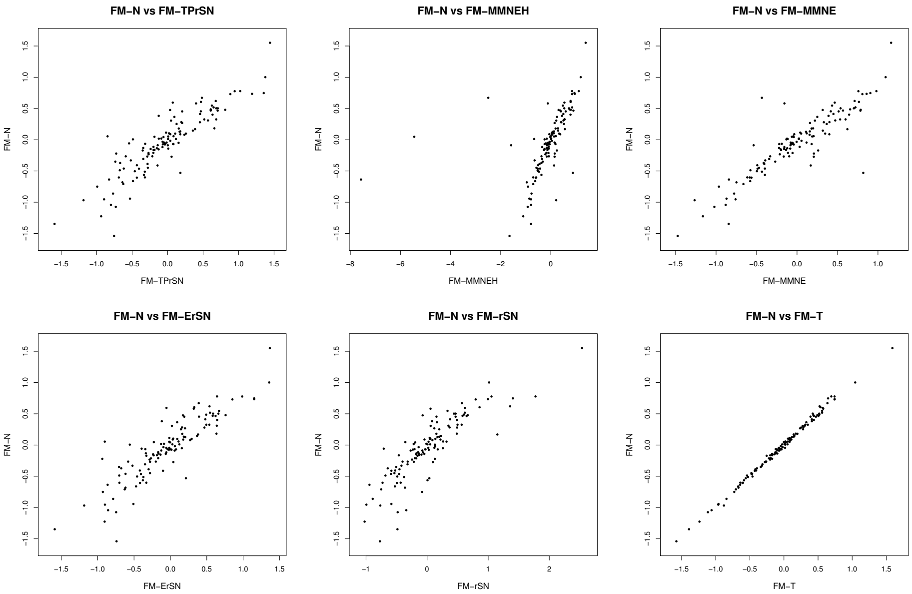

FIGURE 1. Scatter plots of imputed missing values using the FM-T and FM-MMN models via FM-N for the Hepatitis disease data.

as follows

$$\pi_1 = 0.4, \quad \pi_2 = 0.6, \quad \omega_1 = 0.25, \quad \omega_2 = 0.30,$$
$$\boldsymbol{\mu}_1 = (\mu_{11}, \mu_{12}, \mu_{13})^\top = (-1, -1, -1)^\top, \quad \boldsymbol{\mu}_2 = (\mu_{21}, \mu_{22}, \mu_{23})^\top = (2, 2, 2)^\top,$$
$$\boldsymbol{\lambda}_1 = (\lambda_{11}, \lambda_{12}, \lambda_{13})^\top = (2, 1, 1)^\top, \quad \boldsymbol{\lambda}_1 = (\lambda_{21}, \lambda_{22}, \lambda_{23})^\top = (2, -1, 2)^\top,$$
$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}^\top = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}^\top = \begin{pmatrix} 0.67 & 0.33 & 0.65 \\ 0.33 & 2.90 & 0.42 \\ 0.65 & 0.42 & 1.20 \end{pmatrix}^\top,$$

such that $\boldsymbol{\Sigma}_1 = \omega_1\boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}^\top$, $\boldsymbol{\Sigma}_2 = \omega_2\boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}^\top$. In addition, other parameters for FM-MMN are $(\nu_1, \nu_2) = (2.2, 0.25)$ for FM-ErSN, $(\nu_1, \nu_2) = (4, 1.5)$ for FM-MMNE, $(\nu_{11}, \nu_{12}) = (3, 1)$ and $(\nu_{21}, \nu_{22}) = (0.25, 0.30)$ for FM-MMNEH, and $(\nu_1, \nu_2) = (0.10, 0.0.86)$ for FM-TPrSN. We generated missing data with 10% level of missingness rate based on missing at random (MAR) mechanism. For each replication, four special cases and missingness rate, we fitted VEE structure FM-MMN to generated datasets.

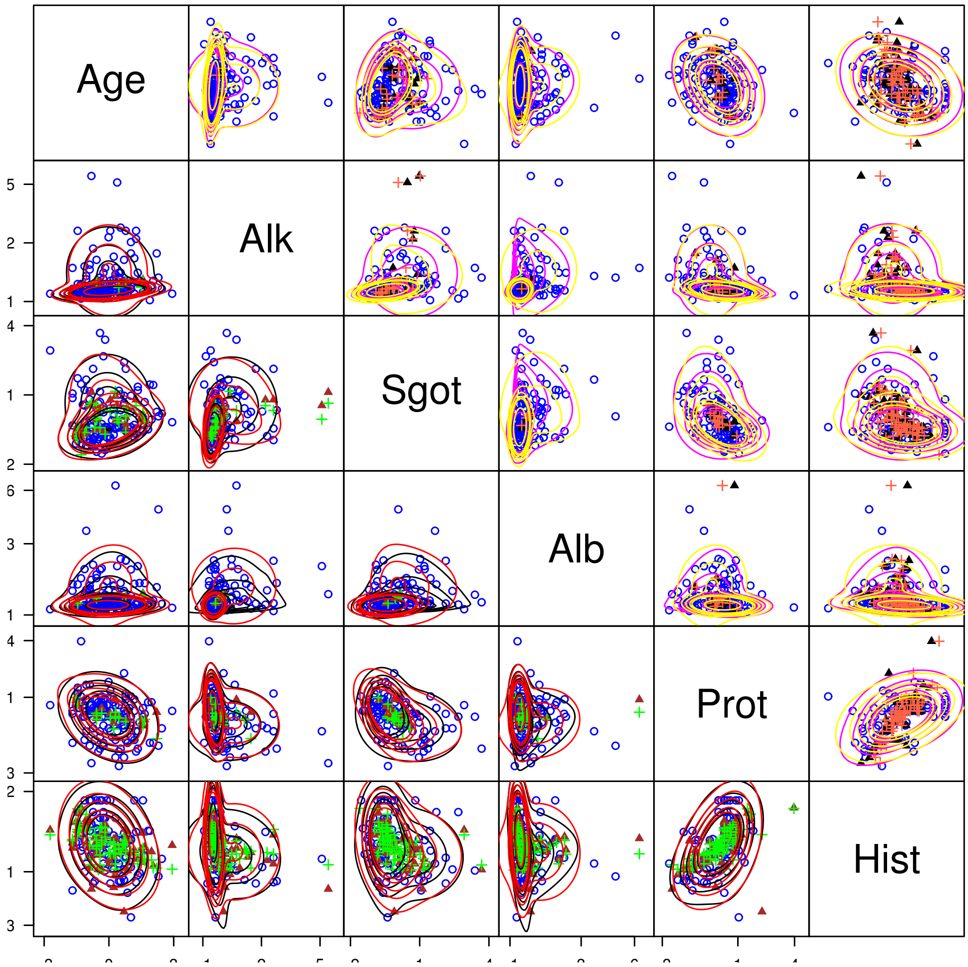

FIGURE 2. Scatter-contour plots in the Hepatitis disease data based on fitted FM-MMN models (FM-MMNE pink line, FM-TPrSN yellow line, FM-MMNEH black line and FM-ErSN red line).

To measure the estimation accuracy, we compute the relative absolute bias (RBias) and the root mean squared error (RMSE):

$$\text{RBias} = \frac{1}{100}\sum_{i=1}^{100} |\hat{\theta}_i - \theta_{true}| \quad \text{and} \quad \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{100}(\hat{\theta}_i - \bar{\hat{\theta}})^2}{100}},$$

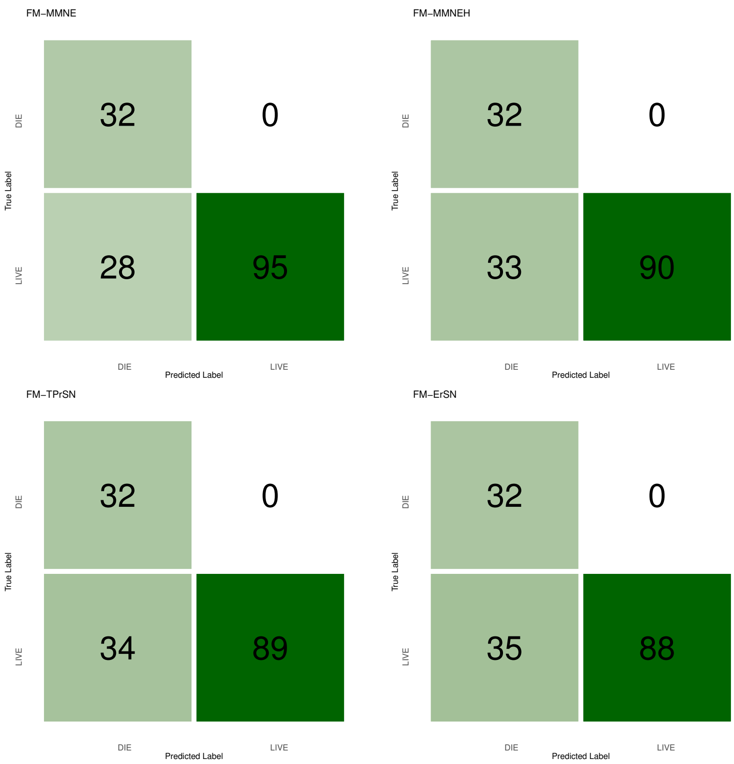

FIGURE 3. The confusion matrices for clustering the Hepatitis disease data using the FM-MMN models.

where $\hat{\theta}_i$ denotes the ML estimate of a specific parameter at the ith replication and θ_{true} is its true value.

Figures 4 and 5 show the means of RBias and RMSE for every parameter in each FM-MMN model with 10% missing rate. It should be mentioned that the assumed parameter i differs for each model. Based on Figures 4 and 5, the simulation experiment's results obviously show the effectiveness of the proposed ECME algorithm in parameter recovery of all the considered sub-models.

5.2. **The performance of the proposed model via thick-tailed data.** In this simulation study, we examine the performance of the proposed model based on heavy-tailed data in terms of clustering quality. These generated data set are simulated from a three-component FM-MMN model that the convolutional variable W in (3) has a Birnbaum-Saunders distribution with shape parameter 1 and scale parameter α. This model, assigned to as FM-MMNBS, is not considered in previous section because the pdf and conditional expectations

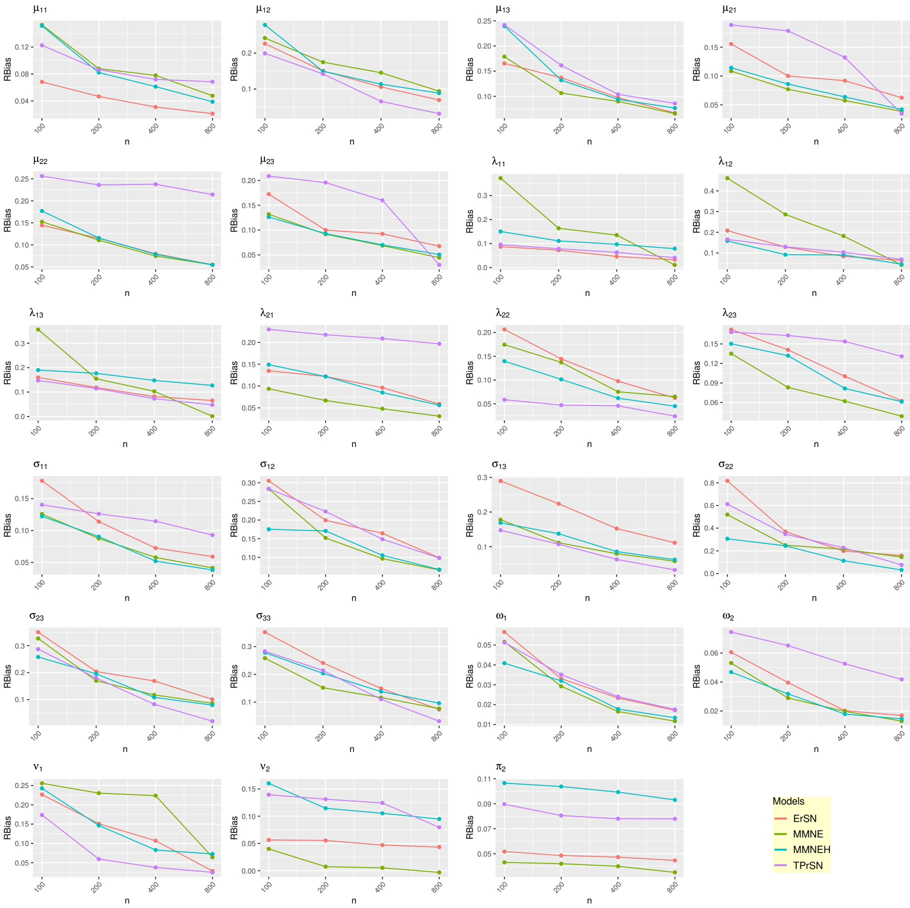

FIGURE 4. Simulation results for assessing the consistency of parameter estimates (RBias) as a function of sample size and 10% missing data.

have not closed-form. The true parameters model are as follows

$$\pi_1 = 0.27, \quad \pi_2 = 0.33, \quad \pi_3 = 0.40, \quad \alpha_1 = 0.8, \quad \alpha_2 = 1, \quad \alpha_3 = 1.5,$$
$$\boldsymbol{\mu}_1 = (9, 6)^\top, \quad \boldsymbol{\mu}_2 = (-5, -5)^\top, \quad \boldsymbol{\mu}_3 = (3, 13)^\top, \quad \boldsymbol{\lambda}_1 = (1.7, 1)^\top,$$
$$\boldsymbol{\lambda}_2 = (2, 1.2)^\top, \quad \boldsymbol{\lambda}_3 = (1, -1)^\top, \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 0.15 & -0.40 \\ -0.40 & 0.30 \end{pmatrix}^\top,$$
$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 0.41 & -0.13 \\ -0.13 & 0.16 \end{pmatrix}^\top, \quad \boldsymbol{\Sigma}_3 = \begin{pmatrix} 0.70 & -0.10 \\ -0.10 & 0.11 \end{pmatrix}^\top.$$

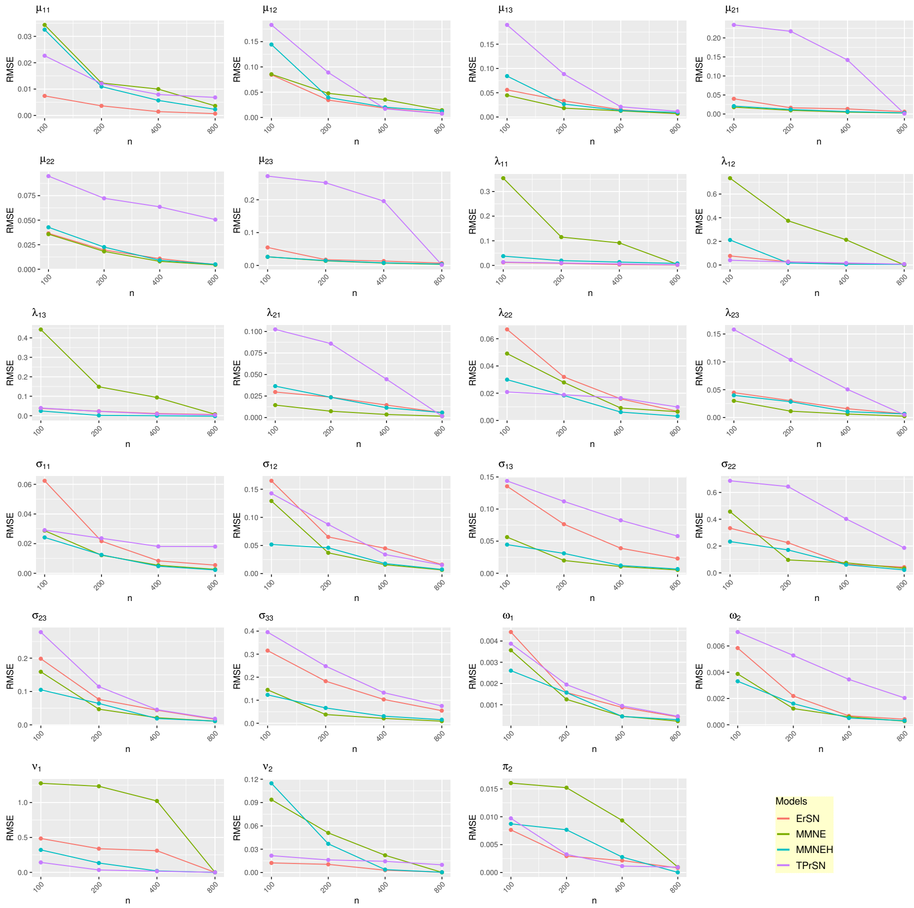

FIGURE 5. Simulation results for assessing the consistency of parameter estimates (RMSE) as a function of sample size and 10% missing data.

We generate missing data in two levels 10% and 30% missingness rates based on MAR mechanism. For the sake of clustering comparison, we fit the full structure FM-N, FM-T, FM-rSN and four special cases of FM-MMN considered with $g = 3$ as well as missingness rates. In addition, we use the MSD to measure for difference between the true value $\boldsymbol{y}_j^m$ and the imputed value $\hat{\boldsymbol{y}}_j^m$.

Table 2 reports the average value of BIC, ARI and MSD for the fourteen constrained and unconstrained variants covariance structures. Focusing on the results Table 2, the average of BIC values supports the outperformance of the FM-MMNEH and FM-MMNE model with the EVE covariance structure in terms of model selection for different missing rates in the strong heavy-tail and

TABLE 2. The average BIC, ARI and MSD over 100 runs, for measuring the proposed model performance in clustering thick-tailed data.

Missing rate	Fitted model		Parsimonious structure													
			EII	EEE	EEI	EEV	EVE	EVI	EVV	VEE	VEI	VEV	VII	VVE	VVI	VVV
10%	FM-N	BIC	2195.617	2132.347	2183.886	1930.801	1932.690	2193.265	1932.661	2131.502	2131.719	2059.704	2144.872	2037.866	2140.292	2044.935
		ARI	0.291	0.239	0.699	0.690	0.291	0.681	0.255	0.270	0.175	0.238	0.241	0.250	0.256	0.291
		MSD	0.422	0.326	0.414	0.262	0.262	0.412	0.256	0.339	0.373	0.276	0.359	0.275	0.376	0.268
	FM-T	BIC	2205.974	2126.872	2195.151	2064.972	1811.121	2204.499	1819.202	2090.044	2131.501	2062.254	2138.771	2053.134	2141.389	2059.937
		ARI	0.284	0.206	0.271	0.699	0.291	0.699	0.215	0.279	0.190	0.238	0.207	0.250	0.238	0.291
		MSD	0.428	0.323	0.412	0.304	0.259	0.411	0.258	0.300	0.377	0.299	0.376	0.275	0.376	0.268
	FM-rSN	BIC	1928.893	1931.198	1934.175	1899.864	1882.205	1927.866	1909.263	1910.699	1917.553	1897.903	1913.762	1897.670	1931.929	1906.768
		ARI	0.896	0.896	0.896	0.896	0.896	0.896	0.896	0.912	0.896	0.896	0.883	0.896	0.882	0.896
		MSD	0.255	0.259	0.256	0.243	0.238	0.254	0.240	0.252	0.244	0.238	0.239	0.233	0.244	0.235
	FM-ErSN	BIC	1923.675	1964.386	1925.582	1898.139	1876.831	1911.201	1907.922	1909.427	1905.104	1923.744	1892.717	1923.870	1915.964	1928.567
		ARI	0.896	0.882	0.896	0.896	0.896	0.896	0.896	0.829	0.829	0.869	0.896	0.869	0.829	0.898
		MSD	0.252	0.268	0.255	0.240	0.232	0.243	0.234	0.227	0.228	0.227	0.239	0.227	0.228	0.229
	FM-MMNE	BIC	1935.522	2051.146	1940.335	1933.918	1889.967	1942.938	1943.624	1972.344	1967.493	1936.994	1955.509	1943.127	1951.282	1944.082
		ARI	0.905	0.868	0.847	0.868	0.890	0.868	0.868	0.863	0.863	0.863	0.877	0.863	0.868	0.863
		MSD	0.356	0.408	0.351	0.361	0.340	0.385	0.359	0.383	0.385	0.360	0.372	0.354	0.384	0.357
	FM-MMNEH	BIC	1857.332	1899.285	1888.350	1836.403	1825.220	1873.259	1846.145	1864.948	1862.165	1846.618	1868.837	1856.547	1867.169	1858.751
		ARI	0.748	0.840	0.901	0.840	0.905	0.885	0.840	0.886	0.901	0.840	0.855	0.868	0.901	0.840
		MSD	0.357	0.373	0.341	0.366	0.350	0.343	0.363	0.344	0.339	0.365	0.355	0.351	0.341	0.361
	FM-TPrSN	BIC	1925.837	1940.009	1925.875	1897.307	1876.330	1902.048	1906.712	1902.574	1898.408	1905.750	1895.404	1905.492	1907.900	1915.452
		ARI	0.896	0.896	0.896	0.896	0.896	0.896	0.896	0.896	0.896	0.896	0.896	0.896	0.896	0.896
		MSD	0.254	0.254	0.256	0.237	0.231	0.243	0.234	0.236	0.238	0.238	0.241	0.234	0.241	0.235
30%	FM-N	BIC	1624.326	1608.665	1622.521	1598.572	1605.903	1629.217	1598.957	1601.829	1609.095	1593.288	1610.138	1596.174	1616.565	1597.642
		ARI	0.549	0.536	0.573	0.543	0.544	0.583	0.543	0.552	0.531	0.543	0.536	0.552	0.541	0.544
		MSD	0.375	0.365	0.378	0.353	0.352	0.379	0.349	0.366	0.384	0.365	0.372	0.353	0.376	0.356
	FM-T	BIC	1588.717	1581.311	1590.016	1577.450	1579.477	1599.094	1575.534	1579.041	1587.094	1579.111	1584.648	1577.432	1595.762	1584.661
		ARI	0.554	0.545	0.545	0.558	0.548	0.593	0.546	0.545	0.544	0.546	0.543	0.548	0.551	0.551
		MSD	0.373	0.360	0.374	0.354	0.349	0.373	0.348	0.356	0.366	0.354	0.365	0.350	0.365	0.350
	FM-rSN	BIC	1582.260	1581.733	1582.627	1582.727	1559.737	1580.108	1589.396	1588.795	1584.405	1586.585	1583.562	1590.684	1587.919	1595.383
		ARI	0.733	0.728	0.733	0.737	0.745	0.753	0.729	0.722	0.722	0.728	0.721	0.720	0.730	0.732
		MSD	0.356	0.356	0.355	0.350	0.350	0.351	0.351	0.357	0.353	0.351	0.352	0.354	0.355	0.351
	FM-ErSN	BIC	1583.647	1650.328	1583.532	1590.530	1567.336	1585.754	1595.993	1588.231	1583.434	1598.389	1582.192	1598.354	1591.955	1603.826
		ARI	0.726	0.714	0.728	0.732	0.730	0.732	0.729	0.722	0.720	0.724	0.725	0.723	0.725	0.729
		MSD	0.352	0.369	0.358	0.354	0.354	0.355	0.355	0.355	0.353	0.352	0.347	0.354	0.355	0.356
	FM-MMNE	BIC	1552.153	1560.035	1552.924	1546.752	1528.225	1551.782	1553.183	1557.787	1556.297	1553.767	1553.968	1559.781	1560.961	1559.585
		ARI	0.794	0.794	0.792	0.769	0.812	0.792	0.805	0.799	0.795	0.798	0.799	0.797	0.796	0.806
		MSD	0.339	0.348	0.340	0.345	0.332	0.338	0.338	0.339	0.339	0.336	0.338	0.336	0.337	0.337
	FM-MMNEH	BIC	1553.121	1568.842	1555.562	1550.258	1531.675	1558.343	1558.961	1567.709	1564.161	1560.773	1559.573	1562.641	1563.756	1568.887
		ARI	0.717	0.726	0.717	0.667	0.735	0.744	0.750	0.744	0.724	0.754	0.730	0.753	0.735	0.745
		MSD	0.363	0.364	0.365	0.357	0.354	0.359	0.357	0.371	0.378	0.378	0.368	0.391	0.370	0.377
	FM-TPrSN	BIC	1573.928	1590.364	1577.367	1581.212	1559.072	1577.360	1588.062	1580.381	1577.954	1589.903	1576.218	1589.040	1585.588	1597.319
		ARI	0.686	0.687	0.693	0.691	0.698	0.698	0.698	0.687	0.692	0.688	0.694	0.699	0.694	0.681
		MSD	0.365	0.369	0.368	0.365	0.366	0.365	0.365	0.370	0.362	0.367	0.364	0.365	0.362	0.369

asymmetric data sets for two levels 10% and 30% missingness rate, respectively. Also, It can be seen that the finite mixture of mean mixture normal, FM-MMNE, FM-ErSN, FM-MMNEH and FM-TPrSN, significantly perform better than the FM-N, FM-T model for estimating the true clustering of data. The clustering of datasets decreases with the increase of the missing rate in many cases, but this amount is not very high in the FM-MMN models. However, comparing the ARI values in Table 2 with the EVE covariance structure, we note that FM-MMNE and FM-MMNEH models for two level missing rate can find more than 90% of true clustering. To charily these results, Figure 6 gives BIC, MSD and ARI for EVE structure for FM-T and four special cases of FM-MMN all 100 replications based on 10% missing rate.

6. **Conclusion**

We develop a computationally feasible ECME algorithm for estimating the parameters of the FM-MMN model under a missing-information framework as an extended tool to accommodate incomplete data involving asymmetric shapes and heavy tails. Two auxiliary permutation matrices are incorporated in the procedure that can greatly simplify matrix manipulations. The performance of the proposed finite mixture models has been investigated in missingness and prediction applications and model-based clustering using two Monte Carlo

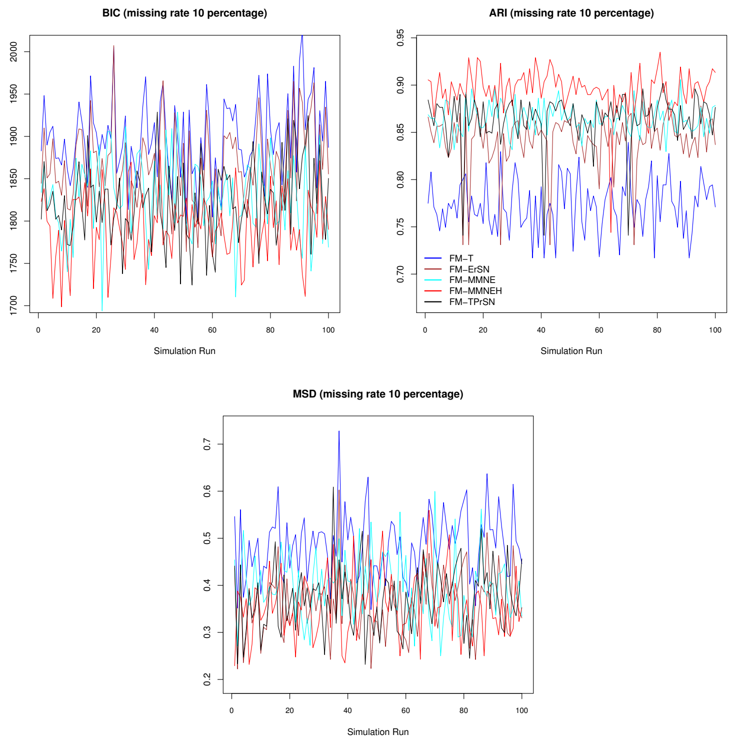

Figure 6. Comparing BIC, ARI and MSD on five mixture models over 100 FM-MMNBS samples with 10% level of missing rate.

simulation studies and a real data set. Numerical result reveal that the FM-MMN model outperforms other competing models on the basis of model fitting and outright clustering when data contain missing values and exhibit non-normal features such as multimodality, asymmetry, and heavy-tailed noises or outliers.

There are a few issues as well as possible modifications related to the proposed methodology deserving further attention. As has been indicated in these distributions, its skew factor analysis based on MMN distributions can be challenged. [5] introduced a finite mixture linear mixed model in which the multivariate t distribution is used for random effects and error distribution. Using the MMN class it will then be of interest to extend the finite mixture linear mixed model to deal with multi-modal, skewed, and heavy-tailed distributed data. We are recently focusing working on these subjects and expect to present the findings in the future papers. The methodology proposed in this paper can facilitate the development of new models for analyzing skewed data in matrix form with censored and/or missing values ([8]). To analyze of high-dimensional data, we can be introduced a new mixture of factor analyzer models based on contaminated mean-mixture of normal distributions proposed by Naderi and Nooghabi [24].

References

[1] Arellano-Valle, RB., Azzalini, A., Ferreira, CS., & Santoro, K. (2020). A two-piece normal measurement error model, Computational Statistics and Data Analysis, 144, 106863. https://doi.org/10.1016/j.csda.2019.106863

[2] Aitken, A. (1925). On Bernoulli's numerical solution of algebraic equations. Proceedings of the Royal Society of Edinburgh, 46, 289 − 305. https://doi.org/10.1017/S0370164600022070

[3] Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In Selected papers of hirotugu akaike, New York, NY: Springer New York, 199 − 213. https://doi.org/10.1007/978-1-4612-1694-0_15

[4] Azzalini, A. (1985). A class of distributions which includes the normal ones. Scandinavian journal of statistics, 171 − 178. https://www.jstor.org/stable/4615982 .

[5] Bai, X., Chen, K., & Yao, W. (2016). Mixture of linear mixed models using multivariate t distribution. Journal of Statistical Computation and Simulation, 86(4), 771 − 787. https://doi.org/10.1080/00949655.2015.1036431

[6] Banfield, JD., & Raftery, AE. (1993). Model-based Gaussian and non-Gaussian clustering. Biometrics, 803 − 821. https://doi.org/10.2307/2532201

[7] Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. IEEE transactions on pattern analysis and machine intelligence, 22(7), 719 − 725. https://doi.org/10.1109/34.865189

[8] Clark, KM., & McNicholas, PD. (2023). Clustering Three-Way Data with Outliers. arXiv preprint arXiv:2310.05288.

[9] Celeux, G., & Govaert, G. (1995). Gaussian parsimonious clustering models. Pattern recognition, 28(5), 781 − 793. https://doi.org/10.1016/0031-3203(94)00125-6

[10] Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. Scientific American, 248(5), 116-131. https://www.jstor.org/stable/24968902

[11] Dempster, AP., Laird, NM., & Rubin, DB. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society: series B (methodological), 39(1), 1 − 22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

[12] Hashemi, F., Naderi, M., & Mashinchi, M. (2019). Clustering right-skewed data stream via Birnbaum–Saunders mixture models: A flexible approach based on fuzzy clustering algorithm. Applied Soft Computing, 82, 105539. https://doi.org/10.1016/j.asoc.2019.105539

[13] Hashemi, F., Naderi, M., Jamalizadeh, A., & Bekker, A. (2021). A flexible factor analysis based on the class of mean-mixture of normal distributions. Computational statistics & data analysis, 157, 107162. https://doi.org/10.1016/j.csda.2020.107162

[14] Hubert, L., & Arabie, P. (1985). Comparing partitions. Journal of classification, 2, 193-218. https://doi.org/10.1007/BF01908075

[15] Lin, TI. (2014). Learning from incomplete data via parameterized t mixture models through eigenvalue decomposition. Computational statistics & data analysis, 71, 183-195. https://doi.org/10.1016/j.csda.2013.02.020

[16] Lin, TC., & Lin, TI. (2010). Supervised learning of multivariate skew normal mixture models with missing information. Computational Statistics, 25, 183 − 201. https://doi.org/10.1007/s00180-009-0169-5

[17] Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data (Vol. 793). John Wiley & Sons.

[18] Liu, C., & Rubin, DB. (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. Biometrika, 81(4), 633 − 648. https://doi.org/10.1093/biomet/81.4.633

[19] McNicholas, PD., & Murphy, TB. (2008). Parsimonious Gaussian mixture models. Statistics and Computing, 18, 285 − 296. https://doi.org/10.1007/s11222-008-9056-0

[20] Meng, XL., & Rubin, DB. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika, 80(2), 267 − 278. https://doi.org/10.1093/biomet/80.2.267

[21] Naderi, M., Hung, WL., Lin, TI., & Jamalizadeh, A. (2019). A novel mixture model using the multivariate normal mean–variance mixture of Birnbaum–Saunders distributions and its application to extrasolar planets. Journal of Multivariate Analysis, 171, 126 − 138. https://doi.org/10.1016/j.jmva.2018.11.015

[22] Naderi, M., Hashemi, F., Bekker, A., & Jamalizadeh, A. (2020). Modeling right-skewed financial data streams: A likelihood inference based on the generalized Birnbaum-Saunders mixture model. Applied Mathematics and Computation, 376, 125109. https://doi.org/10.1016/j.amc.2020.125109

[23] Naderi, M., Bekker, A., Arashi, M., & Jamalizadeh, A. (2020). A theoretical framework for Landsat data modeling based on the matrix variate mean-mixture of normal model. Plos one, 15(4), e0230773. https://doi.org/10.1371/journal.pone.0230773

[24] Naderi, M., & Nooghabi, MJ. (2024). Clustering asymmetrical data with outliers: Parsimonious mixtures of contaminated mean-mixture of normal distributions. Journal of Computational and Applied Mathematics, 437, 115433. https://doi.org/10.1016/j.cam.2023.115433

[25] Negarestani, H., Jamalizadeh, A., Shafiei, S., & Balakrishnan, N. (2019). Mean mixtures of normal distributions: properties, inference and application. Metrika, 82, 501 − 528. https://doi.org/10.1007/s00184-018-0692-x

[26] Rand, WM. (1971). Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association, 66(336), 846 − 850. https://doi.org/10.1080/01621459.1971.10482356

[27] Punzo, A., & McNicholas, PD. (2016). Parsimonious mixtures of multivariate contaminated normal distributions. Biometrical Journal, 58(6), 1506 − 1537. https://doi.org/10.1002/bimj.201500144

[28] Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 461 − 464. https://www.jstor.org/stable/2958889

[29] Sepahdar, A., Madadi, M., Balakrishnan, N., & Jamalizadeh, A. (2022). Parsimonious mixture-of-experts based on mean mixture of multivariate normal distributions. Stat, 11(1), e421. https://doi.org/10.1002/sta4.421

[30] Wang, WL., & Lin, TI. (2015). Robust model-based clustering via mixtures of skew-t distributions with missing information. Advances in Data Analysis and Classification, 9, 423 − 445. https://doi.org/10.1007/s11634-015-0221-y

Farzane Hashemi
Orcid number: 0000-0002-9260-4631
Department of Statistics
University of Kashan
Kashan, Iran
Email address: farzane.hashemi1367@yahoo.com

Saeed Darijani
Orcid number: 0000-0003-2051-5127
Farhangian University Of Kerman
Kerman, Iran
Email address: saeed_darijanir@yahoo.com