A flexible factor analysis based on the class of mean-mixture of normal distributions

Farzane Hashemi[a,b], Mehrdad Naderi[b,*], Ahad Jamalizadeh[c], Andriette Bekker[b]

[a] *Department of Statistics, Faculty of Mathematical Sciences, University of Kashan, Kashan, Iran*
[b] *Department of Statistics, Faculty of Natural & Agricultural Sciences, University of Pretoria, Pretoria, South Africa*
[c] *Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran*

Abstract

Factor analysis is a statistical technique for data reduction and structure detection that traditionally relies on the normality assumption for factors. However, due to the presence of non-normal features such as asymmetry and heavy tails in many practical situations, the first two moments cannot adequately explain the factors. An extension of the factor analysis model is introduced by assuming a generalization of the multivariate restricted skew-normal distribution for the vector of unobserved factors. An efficient and computationally tractable EM-type algorithm is adopted for computing the maximum likelihood estimates by presenting a hierarchical representation of the proposed model. Finally, the efficiency and advantages of the proposed novel methodology are demonstrated through both simulated and real benchmark datasets.

Keywords: Mean-mixture of normal distribution, EM-type algorithm, Factor analysis, Skewness and kurtosis.
MSC codes: 62H12, 62H25.

1. Introduction

Factor analysis (FA), originally proposed in the seminal paper of Spearman (1904), is a widely acknowledged statistical technique that not only aims to reduce the dimensions of data, but also to identify the underlying structure of the data. Generally, the FA model is a generalization of the principal component analysis with an additional appealing scaling invariance property. This means that any change in the scales of the response variables only leads to scale change in the corresponding row of the factor loadings matrix. Theoretically, the FA model relaxed the assumption with respect to the normality distribution of factors and errors. Specifically, let $\{\boldsymbol{Y}_i\}_{i=1}^n$ be a set of n independent and identically distributed (*iid*) random vectors followed by a p-dimensional continuous distribution. The FA model can then be formulated as

$$\boldsymbol{Y}_j = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{U}_j + \boldsymbol{\varepsilon}_j, \qquad \boldsymbol{U}_j \overset{iid}{\sim} \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{I}_q), \quad \boldsymbol{\varepsilon}_j \overset{iid}{\sim} \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{D}), \qquad \boldsymbol{U}_j \perp \boldsymbol{\varepsilon}_j, \tag{1}$$

where $\mathcal{N}_p(\boldsymbol{\xi}, \boldsymbol{\Sigma})$ denotes the p-variate normal distribution with mean vector $\boldsymbol{\xi}$ and covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{I}_q$ is the identity matrix of dimension q, and the symbol '$\perp$' denotes the independence of two random variables. Furthermore, $\boldsymbol{\mu} \in \mathbb{R}^p$ is a location vector, $\boldsymbol{B} \in \mathbb{R}^{p\times q}$ is the matrix of factor loadings, $\boldsymbol{U}_j \in \mathbb{R}^q$ with $q < p$ being the latent variables called *common factors*, $\boldsymbol{\varepsilon}_j \in \mathbb{R}^p$ denote the model errors called *specific factors*, and $\boldsymbol{D}$ is a positive diagonal matrix, say $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{d})$ where $\boldsymbol{d} = (\sigma_1^2, \ldots, \sigma_p^2)$. It can be seen from (1) that $E(\boldsymbol{U}_j) = \boldsymbol{0}$, $\mathrm{cov}(\boldsymbol{U}_j) = \boldsymbol{I}_q$ and $\mathrm{cov}(\boldsymbol{Y}_j) = \boldsymbol{B}\boldsymbol{B}^\top + \boldsymbol{D}$.

The multivariate normality assumption for the factors of the model (1) provides a mathematically as well as computationally tractable method to investigate the complex correlations between the variables under consideration (Basilevsky, 1994). However, the robustness of the model against atypical observations is often criticized in relation to real-world problems (Montanari and Viroli, 2010; Hashemi et al., 2020; Liu and Lin, 2015; Lin et al., 2015). In

*Corresponding author
Email address: m.naderi@up.ac.za (Mehrdad Naderi)

this regard, the interest in skew distributions provide a platform for a robust extension of the FA model. For instance, Montanari and Viroli (2010) proposed a factor model characterized by skew-normally (Azzalini, 1985) distributed factors. Liu and Lin (2015) postulated the restricted multivariate skew-normal (rMSN) FA model (called the rSNFA model) for accommodating incomplete or missing data. Due to its appealing properties and proven proficiency, the rMSN distribution has been employed in a vast number of scientific applications. However, a major drawback of the rMSN distribution is that it is sensitive in the presence of extreme outliers. To accommodate for presence of outliers in the skew-normal type FA models, Lin et al. (2015) proposed a new generalization of the rSNFA and student-t FA (tFA; McLachlan et al. (2007)) models by assuming the restricted multivariate skew-t (rMST) distribution for the vector of unobserved factors and errors, referred to as the rSTFA model. The rMST and rMSN distributions (Pyne et al., 2009) are equivalent to the classical versions, proposed by Azzalini and Capitanio (2003) and Azzalini (1985), after appropriate re-parameterization. The rMSN model belongs to the class of mean-mixture of normal (MMN) distributions. Recently, Negarestani et al. (2019) extended the MMN method to obtain models that not only have an equal number of parameters, but are also more flexible than the rMSN or rMST distributions. Specifically, a p-dimensional random vector $\boldsymbol{X}$ is said to have an MMN distribution if it can be generated through the linear stochastic relationship

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\lambda} W + \boldsymbol{Z}, \qquad \boldsymbol{Z} \perp W, \tag{2}$$

where $\boldsymbol{Z} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, and W is an arbitrary random variable. It is obvious that model (2) assumes that the mean is not fixed for all members of the population. The MMN model can be reduced to symmetric distribution if W is a symmetrically distributed model. However, a more flexible and skew-type one can be obtained based on the assumption that W in (2) follows any asymmetric distribution, preferably a positive support model such as the truncated-normal, exponential and gamma distributions. Alternatively, the MMN distribution might also belong to the class of skew-elliptical models (Azzalini and Capitanio, 1999) if, for example, one considers that W follows the truncated-normal model. Proposing any non-elliptical as well as non-symmetric distribution (e.g. the exponential and gamma models) for the mixing random variable W, (2) would lead to a skew non-elliptically contoured distribution. By introducing two new special cases of the MMN model, Negarestani et al. (2019) showed that the new model could take a wider range of skewness and kurtosis than the rMSN, rMST and skew-t-normal (Ho et al., 2011) distributions. They showed that the MMN model inherits the log-concavity property from the rMSN distribution, and that it is infinitely divisible, unlike the rMSN model. The infinite divisibility enables investigators to study the central limit theorem based on the underlying distribution.

With respect to the mentioned properties of Negarestani et al. (2019), the objective of this paper is to propose a new factor model by assuming the MMN distribution for the factors. The proposed hierarchical representation enables the development of an expectation-maximization (EM; Dempster et al. (1977)) type algorithm for computing the maximum likelihood (ML) estimates of parameters. In the rMST-based models, especially rSTFA, it is known that the rMSN-based models are obtained as the degree of freedom tends to infinity. A simulation study in Section 4 shows that the proposed model outperforms both rSNFA and rSTFA models when the degree of freedom increases. The mathematical and computational efficiency of the presented methodology, namely the finite sample properties and outperformance in dealing with the highly skewed data, are also verified. Finally, two real-world datasets provide a comparative analysis of the performance of the new factor model compared with some existing FA models.

The layout of the paper is as follows. In Section 2, the MMN model formulation and some of its characteristics are presented. Section 3 presents the formulation of the MMN factor analysis (MMNFA) model along with its parameter estimation. Three simulation scenarios are conducted in Section 4 to investigate the performance of the model and to study the finite sample properties of the proposed EM-based estimators. The usefulness of the proposed method is illustrated in Section 5 by analyzing two real datasets. Finally, discussion and suggestions for future work follow. Some technical details and additional information are provided in the Online Supplement.

2. The multivariate MMN distribution: review and some properties

2.1. General formulation

For the sake of notation, let $\phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the probability density function (PDF) of $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\Phi(\cdot)$ be the cumulative distribution function of the univariate standard normal distribution (CDF). Following Negarestani et al.

(2019), let W in (2) have the PDF $h(\cdot;\boldsymbol{\nu})$, parameterized by a vector parameter $\boldsymbol{\nu}$. Therefore, by the hierarchical representation

$$\boldsymbol{X}|W = w \sim \mathcal{N}_p(\boldsymbol{\mu} + \boldsymbol{\lambda} w, \boldsymbol{\Sigma}), \qquad W \sim h(w;\boldsymbol{\nu}),$$

$\boldsymbol{X}$ has the MMN distribution with PDF

$$f_{\mathrm{MMN}}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\boldsymbol{\nu}) = \int_{-\infty}^{\infty} \phi(\boldsymbol{x};\boldsymbol{\mu} + \boldsymbol{\lambda} w, \boldsymbol{\Sigma}) h(w;\boldsymbol{\nu})\, dw, \quad \boldsymbol{x} \in \mathbb{R}^p. \tag{3}$$

The notation $\boldsymbol{X} \sim \mathcal{MMN}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda}; h(w;\boldsymbol{\nu}))$ will be used to indicate that $\boldsymbol{X}$ has PDF (3). The mean, covariance matrix and moment generating function of $\boldsymbol{X}$ are respectively

$$\begin{aligned} E(\boldsymbol{X}) =& \boldsymbol{\mu} + E(W)\boldsymbol{\lambda}, \quad \text{if} \quad E(|W|) < \infty, \quad \operatorname{cov}(\boldsymbol{X}) = \boldsymbol{\Sigma} + \operatorname{Var}(W)\boldsymbol{\lambda}\boldsymbol{\lambda}^\top, \quad \text{if} \quad E(W^2) < \infty, \\ M_{\boldsymbol{X}}(\boldsymbol{t};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda}) =& \exp\left(\boldsymbol{t}^\top \boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right) M_W(\boldsymbol{t}^\top \boldsymbol{\lambda}), \end{aligned} \tag{4}$$

where $M_W(\cdot)$ denotes the moment generating function of W.

Theorem 1. The MMN distribution is closed under linear transformation, i.e. if $\boldsymbol{X} \sim \mathcal{MMN}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda}; h(w;\boldsymbol{\nu}))$, then for any full rank matrix $\boldsymbol{L} \in \mathbb{R}^{q\times p}$, $1 \le q \le p$, the random vector $\boldsymbol{L}\boldsymbol{X}$ is distributed by $\mathcal{MMN}_q(\boldsymbol{L}\boldsymbol{\mu}, \boldsymbol{L}\boldsymbol{\Sigma}\boldsymbol{L}^\top, \boldsymbol{L}\boldsymbol{\lambda}; h(w;\boldsymbol{\nu}))$.

Proof. The proof follows by applying the $\boldsymbol{L}\boldsymbol{X}$ transformation to the moment generating function (4). □

Theorem 2. If $\boldsymbol{X} \sim \mathcal{MMN}_q(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1,\boldsymbol{\lambda}; h(w;\boldsymbol{\nu}))$ and $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$, then for any matrix $\boldsymbol{A}$ of dimension $p \times q$ follows that

$$\boldsymbol{A}\boldsymbol{X} + \boldsymbol{Y} \sim \mathcal{MMN}_p(\boldsymbol{A}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{A}\boldsymbol{\Sigma}_1\boldsymbol{A}^\top + \boldsymbol{\Sigma}_2, \boldsymbol{\lambda}; h(w;\boldsymbol{\nu})).$$

2.2. *Special cases*

In this section, three distributions of the MMN family are presented.

- **The rMSN distribution**: Let the mixing variable W follow the truncated standard normal distribution lying within a truncated interval $(0,\infty)$, denoted by $W \sim \mathcal{TN}(0,1;(0,\infty))$. Then, the PDF of a p-dimensional random vector $\boldsymbol{X}$ following the rMSN distribution is given by

$$f_{\mathrm{rMSN}}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda}) = 2\phi_p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Omega})\Phi\left(\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\sqrt{1-\boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}}}\right), \tag{5}$$

where $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top$. The rMSN distribution, denoted by $r\mathcal{MSN}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda})$, has been used extensively, see Lee and McLachlan (2013) and Lin et al. (2016) to name a few.

- **Convolution with the exponential distribution**: The p-variate exponentiated MMN (MMNE) distribution, say $\boldsymbol{X} \sim \mathcal{MMNE}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda})$, is derived from (2) by taking W as a standard exponential distribution, $\mathcal{E}(1)$. Using (3), the PDF of $\boldsymbol{X}$ can be obtained as

$$f_{\mathrm{MMNE}}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda}) = \frac{\sqrt{2\pi}}{\delta}\exp\left(\frac{A^2}{2}\right)\phi_p\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}\right)\Phi(A), \quad \boldsymbol{x} \in \mathbb{R}^p, \tag{6}$$

where $\delta^2 = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}$ and $A = \delta^{-1}\left[\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - 1\right]$. It is interesting to note that the number of parameters of the MMNE model is equal to that of the rMSN distribution. The mean, covariance matrix and moment generating function of the MMNE distribution, obtained by (4), are

$$E(\boldsymbol{X}) = \boldsymbol{\mu} + \boldsymbol{\lambda}, \quad \operatorname{cov}(\boldsymbol{X}) = \boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top, \quad M_{\boldsymbol{X}}(\boldsymbol{t};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda}) = \frac{\exp\left(\boldsymbol{t}^\top\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t}\right)}{1-\boldsymbol{t}^\top\boldsymbol{\lambda}}, \quad \forall_t \ \ \boldsymbol{t}^\top\boldsymbol{\lambda} \neq 1.$$

Proposition 1. The PDF of the multivariate MMNE distribution is log-concave.

Proof. The proposition is obtained immediately through the properties of the log-concave function, i.e. the class of log-concave functions is closed under multiplication. □

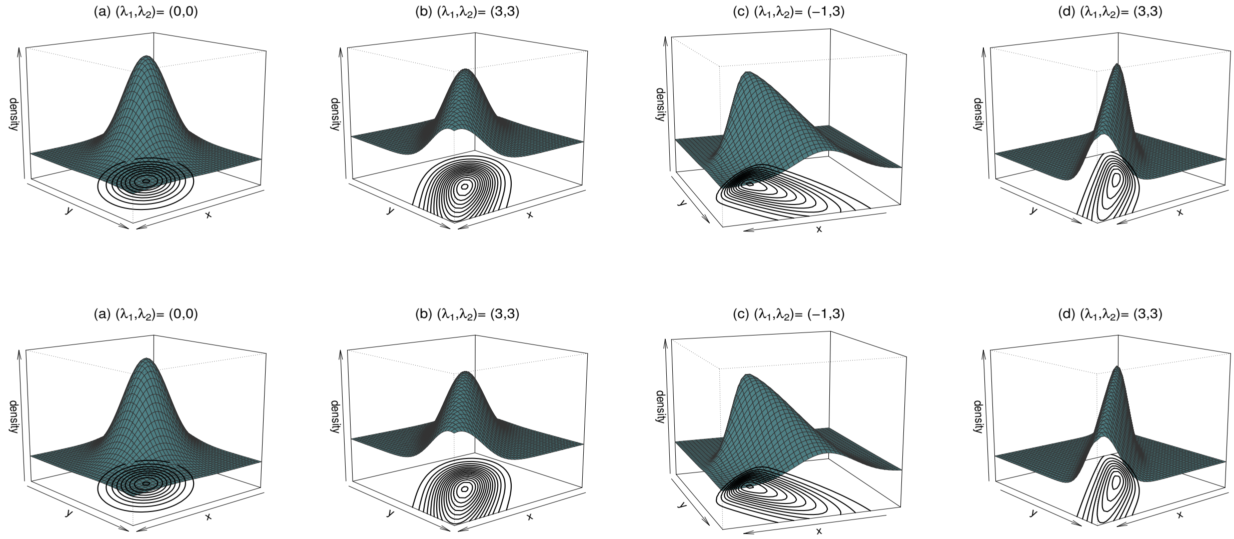

Figure 1: The perspective density and contour plots of the MMNE (upper panel) and MMNEH (with $\nu = 0.15$; lower panel) distributions for various settings of parameters (the two first panels from left for $\mathbf{\Sigma}_1$ and from right for $\mathbf{\Sigma}_2$).

- **Convolution with a mixture of exponential and half-normal distributions**: If the PDF of W in (2) is a mixture of an exponential distribution with mean 2, $\mathcal{E}(2)$, and $\mathcal{TN}(0, 1; (0, +\infty))$ given by

$$f_W(w) = 0.5\nu \exp\big(-0.5w\big) + 2(1-\nu)\phi(w), \quad w > 0, \;\; 0 < \nu < 1, \tag{7}$$

then, the half-normal exponentiated MMN (MMNEH) distribution follows. Denoted by $\boldsymbol{X} \sim \mathcal{MMNEH}_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}, \nu)$, the associated PDF of $\boldsymbol{X}$ obtained by (3) is

$$f_{\text{MMNEH}}(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}, \nu) = \nu \frac{\sqrt{2\pi}}{2\delta} \phi_p\left(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{\Sigma}\right) \exp\left(\frac{A^{*2}}{2}\right) \Phi(A^*) + (1-\nu) f_{\text{rMSN}}(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}), \quad \boldsymbol{x} \in \mathbb{R}^p, \tag{8}$$

where $A^* = \delta^{-1}\left[\boldsymbol{\lambda}^\top \mathbf{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) - 0.5\right]$. It is clearly seen that the MMNEH distribution approaches the rMSN model as ν tends zero. Moreover, the PDF of MMNEH distribution (8) tends to the normal one as both ν and $\boldsymbol{\lambda}$ approach zero. Furthermore, the mean, covariance matrix and moment generating function of the MMNEH distribution are

$$E(\boldsymbol{X}) = \boldsymbol{\mu} + \left(\nu(2 - \sqrt{2/\pi}) + \sqrt{2/\pi}\right)\boldsymbol{\lambda}, \qquad \text{cov}(\boldsymbol{X}) = \mathbf{\Sigma} + \left(7\nu + 1 - \left(\nu(2 - \sqrt{2/\pi}) + \sqrt{2/\pi}\right)^2\right)\boldsymbol{\lambda}\boldsymbol{\lambda}^\top,$$
$$M_{\boldsymbol{X}}(\boldsymbol{t}; \boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}) = \exp\left(\boldsymbol{t}^\top\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top\mathbf{\Sigma}\boldsymbol{t}\right)\left(\frac{\nu}{1 - 2\boldsymbol{t}^\top\boldsymbol{\lambda}} + (1-\nu)\exp\left(\frac{1}{2}(\boldsymbol{t}^\top\boldsymbol{\lambda})^2\right)\Phi(\boldsymbol{t}^\top\boldsymbol{\lambda})\right), \quad \forall_{\boldsymbol{t}} \;\; \boldsymbol{t}^\top\boldsymbol{\lambda} \neq 0.5.$$

Figure 1 illustrates the perspective density plots with added contours for the bivariate MMNE (upper panel) and MMNEH (lower panel) distributions by setting $\boldsymbol{\mu} = (0, 0)$, $\mathbf{\Sigma}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathbf{\Sigma}_2 = \begin{pmatrix} 0.3 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, $\nu = 0.15$, and with different settings of $\boldsymbol{\lambda}$. These plots depict that both MMNE and MMNEH distributions show different degrees of flatness, skewness and kurtosis, depending on the choice of parameters. Figure 2 displays the contour plots of bivariate densities given in (5), (6) and (8), obtained with the solutions of $f(\boldsymbol{x}; \Theta) = c$, for $c = 0.1$ and 0.03, where $\boldsymbol{\mu} = (0, 0)$, $\mathbf{\Sigma} = \mathbf{\Sigma}_2$, $\boldsymbol{\lambda} = (1, \text{-}1)$, and ν takes various choices from (0,1). Here, $f(\boldsymbol{x}; \Theta)$ represents the PDF of the rMSN, MMNE or MMNEH models. Note that the rMSN contour is outside those of the MMNE and MMNEH models for $c = 0.1$, whereas for $c = 0.03$ the contours of the MMNE and MMNEH distributions apparently peak outside the rMSN contour. This behaviour is also seen for large values of ν for the MMNEH contours against the MMNE ones.

Remark 1. It is interesting to emphasize that the class of MMN distributions offers different contour plots comparing to the family of normal mean-variance mixture (NMVM) models (McNeil et al., 2005). To illustrate later, Figure 1

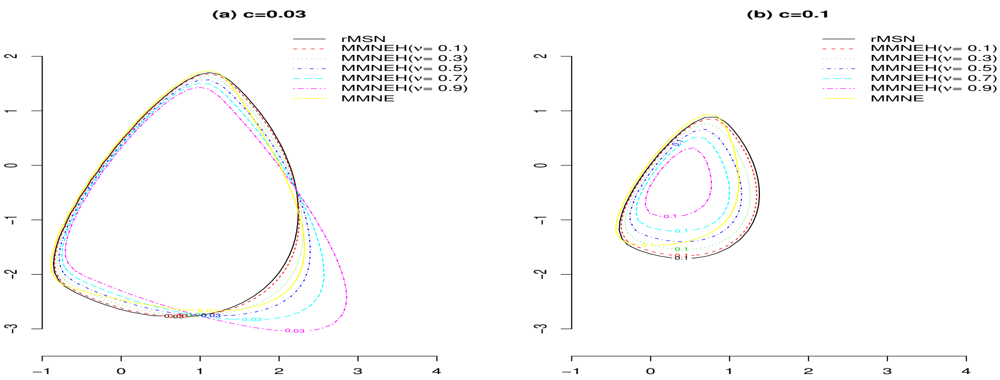

Figure 2: A contour comparison of the rMSN, MMNE and MMNEH (for various choices of ν) distributions by plotting $f(\boldsymbol{x};\Theta) = c$ under two levels (a) $c = 0.03$ and (b) $c = 0.1$.

in the Online Supplement provides the contour plots of three special cases of the MMN and NMVM distributions. Moreover, conditionally on mixing variable $W = w$, note that the NMVM distributions assume that both the variance and mean for all members of the population are not fixed. Details and application of the NMVM model in factor analysis can be found in Murray et al. (2014a); Tortora et al. (2015) and Hashemi et al. (2020), among others. Future development of the current work will therefore be of interest in proposing a scale mixture of the MMN distribution.

Subsequently, some lemmas and theorems are presented that are useful for the calculation of some conditional expectations involved in the proposed EM-type algorithm discussed in the next section.

Lemma 1. If $W \sim \mathcal{TN}\left(\xi, \omega^2; (0,\infty)\right)$, then $E(W) = \xi + \omega\frac{\phi(\xi/\omega)}{\Phi(\xi/\omega)}$, and $E(W^r) = \xi E\left(W^{r-1}\right) + \omega^2(r-1)E\left(W^{r-2}\right)$ for $r = 2, 3, \ldots$.

Lemma 2. Let $\boldsymbol{Y} \sim r\mathcal{MSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ and $W \sim \mathcal{TN}(0, 1; (0,\infty))$. Then, W conditionally on $\boldsymbol{Y} = \boldsymbol{y}$, follows $\mathcal{TN}\left(\xi, \sigma^2; (0,\infty)\right)$, where $\xi = \boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})$ and $\sigma^2 = 1 - \boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$.

Theorem 3. Suppose $\boldsymbol{Y} \sim \mathcal{MMNE}_P(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ and $W \sim \mathcal{E}(1)$. Then, $W|\boldsymbol{Y} = \boldsymbol{y} \sim \mathcal{TN}\left(A\delta^{-1}, \delta^{-2}; (0,\infty)\right)$, where δ and A are defined in (6). Furthermore, for $k = 1, 2, \ldots$,

$$E(W^k|\boldsymbol{Y} = \boldsymbol{y}) = \frac{A}{\delta}E(W^{k-1}|\boldsymbol{Y} = \boldsymbol{y}) + \frac{k-1}{\delta^2}E(W^{k-2}|\boldsymbol{Y} = \boldsymbol{y}),$$

where

$$E(W|\boldsymbol{Y} = \boldsymbol{y}) = \frac{A}{\delta} + \frac{\phi(A)}{\delta\Phi(A)}.$$

Proof. This result follows from Bayes' rule and Lemma 1. □

Theorem 4. Let $\boldsymbol{Y} \sim \mathcal{MMNEH}_P(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ and W have PDF (7). Then, the conditional PDF of W given $\boldsymbol{Y} = \boldsymbol{y}$, is

$$f_{W|\boldsymbol{Y}=\boldsymbol{y}}(w) = \pi(\boldsymbol{y})\frac{\phi\left(w; A^*\delta^{-1}, \delta^{-2}\right)}{\Phi(A^*)} + (1-\pi(\boldsymbol{y}))\frac{\phi\left(w; \vartheta, \sigma^2\right)}{\Phi(\vartheta/\sigma)},$$

where $\vartheta = \boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})$, $\sigma^2 = 1 - \boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$, and

$$\pi(\boldsymbol{y}) = \frac{\nu\sqrt{2\pi}}{2\delta f_{\text{MMNEH}}(\boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)}\phi_p(\boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\exp\left(\frac{A^{*2}}{2}\right)\Phi(A^*).$$

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

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

where $V_1 \sim \mathcal{TN}\left(A^*\delta^{-1}, \delta^{-2}; (0,\infty)\right)$, $V_2 \sim \mathcal{TN}\left(\vartheta, \sigma^2; (0,\infty)\right)$ and

$$E(V_1) = \frac{A^*}{\delta} + \frac{\phi(A^*)}{\delta\Phi(A^*)}, \quad E(V_1^k) = \frac{A^*}{\delta}E(V_1^{k-1}) + \frac{k-1}{\delta^2}E(V_1^{k-2}), \quad k \geq 2,$$
$$E(V_2) = \vartheta + \sigma\frac{\phi(\vartheta/\sigma)}{\Phi(\vartheta/\sigma)}, \quad E(V_2^k) = \vartheta E(V_2^{k-1}) + (k-1)\sigma^2 E(V_2^{k-2}), \quad k \geq 2.$$

Proof. The proof is straightforward. □

The following theorem considers the moment generating function of the quadratic form associated with the special cases of the MMN family of distributions. This quadratic form might be useful for assessing the validity of the underlying distributional assumption.

Theorem 5. The moment generating function of the quadratic form $Q = \boldsymbol{X}^\top \boldsymbol{V}\boldsymbol{X}$ for any symmetric matrix $\boldsymbol{V}$ can be obtained as:

i) If $\boldsymbol{X} \sim \mathcal{MMNE}_p(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$,

$$M_Q(t) = \frac{\sqrt{2\pi}|\boldsymbol{\Psi}|^{1/2}}{\delta|\boldsymbol{\Sigma}|^{1/2}} \exp\left(\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}}{2\delta^4} + 0.5\delta^{-2}\right)\Phi\left(\frac{\delta^{-2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda} - 1}{\sqrt{\delta^2 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}}}\right),$$

ii) If $\boldsymbol{X} \sim \mathcal{MMNEH}_p(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$,

$$M_Q(t) = \frac{\nu\sqrt{2\pi}|\boldsymbol{\Psi}|^{1/2}}{2\delta|\boldsymbol{\Sigma}|^{1/2}} \exp\left(\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}}{2\delta^4} + \frac{1}{8\delta^2}\right)\Phi\left(\frac{\delta^{-2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda} - 2}{\sqrt{\delta^2 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}}}\right) + \frac{(1-\nu)}{|\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1} - 2t\boldsymbol{V})|},$$

where $\boldsymbol{\Psi} = \left(\boldsymbol{\Sigma}^{-1} - 2t\boldsymbol{V} - \delta^{-2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}\right)^{-1}$.

Proof. The proof can be found in Appendix B of the Online Supplement. □

In the following theorems, some conditions are presented under which two linear and/or quadratic forms of the MMN distribution are independent.

Theorem 6. Let $\boldsymbol{X} \sim \mathcal{MMNE}_p(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. For $\boldsymbol{h} \in \mathbb{R}^p$ and $\boldsymbol{V} \in \mathbb{R}^{p\times p}$, the linear form $\boldsymbol{h}^\top\boldsymbol{X}$ and the quadratic form $\boldsymbol{X}^\top\boldsymbol{V}\boldsymbol{X}$ are independent if and only if $\boldsymbol{V}\boldsymbol{\Omega}_1\boldsymbol{h} = \boldsymbol{0}$ and $\boldsymbol{V}\boldsymbol{\Omega}_1\boldsymbol{\alpha}_1 = \boldsymbol{0}$ where $\boldsymbol{\Omega}_1 = \left(\boldsymbol{\Sigma}^{-1} - \delta^{-2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}\right)^{-1}$ and $\boldsymbol{\alpha}_1 = \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}$.

Proof. Proof of the result is provided in Appendix B of the Online Supplement. □

Theorem 7. Let $\boldsymbol{X} \sim \mathcal{MMNEH}_p(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. For any $\boldsymbol{h} \in \mathbb{R}^p$ and symmetric matrix $\boldsymbol{V} \in \mathbb{R}^{p\times p}$, the linear form $\boldsymbol{h}^\top\boldsymbol{X}$ and the quadratic form $\boldsymbol{X}^\top\boldsymbol{V}\boldsymbol{X}$ are independent if and only if $\boldsymbol{V}\boldsymbol{\Omega}_1\boldsymbol{h} = \boldsymbol{0}$, $\boldsymbol{V}\boldsymbol{\Omega}_1\boldsymbol{\alpha}_1 = \boldsymbol{0}$, $\boldsymbol{V}\boldsymbol{\Omega}_2\boldsymbol{h} = \boldsymbol{0}$, $\boldsymbol{V}\boldsymbol{\Omega}_2\boldsymbol{\alpha}_2 = \boldsymbol{0}$ where $\boldsymbol{\Omega}_2 = \boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top$ and $\boldsymbol{\alpha}_2 = \frac{\boldsymbol{\lambda}^\top\boldsymbol{\Omega}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\sqrt{1-\boldsymbol{\lambda}^\top\boldsymbol{\Omega}_2^{-1}\boldsymbol{\lambda}}}$.

Proof. The result can be obtained by following a similar procedure used in Theorem 6. □

Theorem 8. For any symmetric matrix $\boldsymbol{V}_1, \boldsymbol{V}_2 \in \mathbb{R}^{p\times p}$, the quadratic forms $\boldsymbol{X}^\top\boldsymbol{V}_1\boldsymbol{X}$ and $\boldsymbol{X}^\top\boldsymbol{V}_2\boldsymbol{X}$ are independent if and only if:

i) $\boldsymbol{V}_1\boldsymbol{\Omega}_1\boldsymbol{V}_2 = \boldsymbol{0}$, when $\boldsymbol{X} \sim \mathcal{MMNE}_p(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$.

ii) $\boldsymbol{V}_1\boldsymbol{\Omega}_1\boldsymbol{V}_2 = \boldsymbol{0}$ and $\boldsymbol{V}_1\boldsymbol{\Omega}_2\boldsymbol{V}_2 = \boldsymbol{0}$, when $\boldsymbol{X} \sim \mathcal{MMNEH}_p(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$.

Proof. Details of the proof are given in Appendix B of the Online Supplement. □

Theorem 9. Let $\boldsymbol{X} \sim \mathcal{MMNE}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ (or $\boldsymbol{X} \sim \mathcal{MMNEH}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$) and the following partitions

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\boldsymbol{X}_1, \boldsymbol{\mu}_1, \boldsymbol{\lambda}_1 \in \mathbb{R}^q$ and $\boldsymbol{\Sigma}_{11} \in \mathbb{R}^{q\times q}$. Then, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent if and only if two conditions (i) $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$ and (ii) either $\boldsymbol{\lambda}_1 = \boldsymbol{0}$ or $\boldsymbol{\lambda}_2 = \boldsymbol{0}$ hold simultaneously.

Proof. The focus is on the MMNE distribution. The proof of one side is straightforward. Thus, suppose that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent. Then, the moment generating of $\boldsymbol{X}$ can be represented as

$$M_{\boldsymbol{X}}(\boldsymbol{t}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = M_{\boldsymbol{X}_1}(\boldsymbol{t}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\lambda}_1) M_{\boldsymbol{X}_2}(\boldsymbol{t}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, \boldsymbol{\lambda}_2), \quad \forall \boldsymbol{t} = (\boldsymbol{t}_1^\top, \boldsymbol{t}_2^\top)^\top,$$

where $\boldsymbol{t}_1 \in \mathbb{R}^q$ and $\boldsymbol{t}_2 \in \mathbb{R}^{p-q}$, $M_{\boldsymbol{X}}(\cdot;\cdot)$ is defined in (4). Therefore,

$$\exp\left(\boldsymbol{t}_1^\top \boldsymbol{\Sigma}_{12} \boldsymbol{t}_2\right) = \frac{1 - \boldsymbol{t}^\top \boldsymbol{\lambda}}{(1 - \boldsymbol{t}_1^\top \boldsymbol{\lambda}_1)(1 - \boldsymbol{t}_2^\top \boldsymbol{\lambda}_2)}. \tag{9}$$

It is obvious that (9) holds if both (i) and (ii) happen, which completes the proof. The proof for the MMNEH model is similar and hence is omitted. □

3. The MMN factor analysis model

3.1. Model formulation

Next, a new factor model is defined by considering the MMN distribution for latent factors to model correlation in the presence of asymmetric levels of sources. The MMNFA model postulated here can be formulated through (1) as

$$\begin{aligned} &\boldsymbol{Y}_j = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{U}_j + \boldsymbol{\varepsilon}_j, \\ &\boldsymbol{U}_j \overset{iid}{\sim} \mathcal{MMN}_q(-a_\nu \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\lambda}, \boldsymbol{\Lambda}^{-1}, \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\lambda}; h(w;\boldsymbol{\nu})), \quad \boldsymbol{\varepsilon}_j \overset{iid}{\sim} \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{D}), \qquad \boldsymbol{U}_j \perp \boldsymbol{\varepsilon}_j, \end{aligned} \tag{10}$$

where the scaling coefficients are $a_\nu = E(W_j)$ and $b_\nu = \mathrm{Var}(W_j)$, $\boldsymbol{\Lambda} = \boldsymbol{I}_q + b_\nu \boldsymbol{\lambda}\boldsymbol{\lambda}^\top$. Notice that the scaling coefficients a_ν and b_ν are chosen such that $\boldsymbol{U}_j$ fulfills the assumptions of the FA model, i.e., $E(\boldsymbol{U}_j) = \boldsymbol{0}$ and $\mathrm{cov}(\boldsymbol{U}_j) = \boldsymbol{I}_q$.

Alternatively, by the linear representation (2), the proposed MMNFA model in (10) admits the following two-level hierarchical representation

$$\boldsymbol{Y}_j \mid W = w_j \sim \mathcal{N}_p(\boldsymbol{\mu} - a_\nu \boldsymbol{B}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\lambda} + w_j \boldsymbol{B}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\lambda}, \boldsymbol{\Sigma}), \qquad W_j \sim h(w_j; \boldsymbol{\nu}). \tag{11}$$

Consequently, $\boldsymbol{Y}_j \sim \mathcal{MMN}_P(\boldsymbol{\mu} - a_\nu \boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\eta}; h(w;\boldsymbol{\nu}))$, where $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Lambda}^{-1}\boldsymbol{B}^\top + \boldsymbol{D}$ and $\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\lambda}$. Therefore, the mean and covariance matrix of $\boldsymbol{Y}_j$ obtained by (4) are

$$E(\boldsymbol{Y}_j) = \boldsymbol{\mu}, \qquad \mathrm{cov}(\boldsymbol{Y}_j) = \boldsymbol{B}\boldsymbol{B}^\top + \boldsymbol{D}, \quad \text{and} \quad \mathrm{cov}(\boldsymbol{Y}_j, \boldsymbol{U}_j) = \boldsymbol{B}.$$

It is clear that $\mathrm{cov}(\boldsymbol{Y}_j)$ of the MMNFA model always exists, whereas for the rSTFA model (Lin et al., 2015) and the generalized hyperbolic skew-t factor analysis (GHSTFA; Murray et al. (2014a)), for example, the covariance matrices respectively are

$$\frac{\nu}{\nu-2}(\boldsymbol{B}\boldsymbol{B}^\top + \boldsymbol{D}) \quad \text{and} \quad \frac{\nu}{\nu-2}(\boldsymbol{B}\boldsymbol{B}^\top + \boldsymbol{D}) + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top,$$

which do not exist for $\nu = 2$. The same result can be obtained in comparing the $\mathrm{cov}(\boldsymbol{Y}_j, \boldsymbol{U}_j)$ for the MMNFA, rSTFA and GHSTFA models.

It can be verified that model (10) is still satisfied when $\boldsymbol{B}$ is replaced by $\boldsymbol{BR}$ for any arbitrary orthogonal rotation matrix $\boldsymbol{R}$ with order $q > 1$. Therefore, the MMNFA model suffers from an identifiability problem associated with the rotation invariance of the loading matrix $\boldsymbol{B}$. To overcome this challenge, two commonly implemented methods introduced by Lawley and Maxwell (1971) and Fokoué and Titterington (2003) can be used. Lawley and Maxwell (1971) recommended choosing $\boldsymbol{R}$ as a uniqueness condition, such that $\boldsymbol{B}^\top\boldsymbol{D}^{-1}\boldsymbol{B}$ is a diagonal matrix with elements arranged in descending order. The second method used here is to constrain $\boldsymbol{B}$ in such a way that its upper-right triangle is zero and its diagonals are strictly positive (Fokoué and Titterington, 2003). In both approaches, $q(q-1)/2$ constraints are imposed on $\boldsymbol{B}$ and the number of free parameters is reduced to $p(q+2) + q - q(q-1)/2 + s$ where s denotes the length of $\boldsymbol{\nu}$. Furthermore, the imposed constraints on $\boldsymbol{B}$ lead to the condition $(p-q)^2 \geq (p+q)$ that is used for obtaining the maximum number of factors, q (McLachlan and Peel, 2000).

3.2. *Parameter estimation via an EM-type algorithm*

In this section, an extension of the EM algorithm called expectation conditional maximization (ECM; Meng and Rubin (1993)) is implemented for estimating the MMNFA parameters. As the EM algorithm is a well-known iterative tool used to estimate parameters of the model with hidden variables, the ECM algorithm can increase the speed of convergence. The key idea of the ECM approach is to construct a complete-data log-likelihood function, i.e., the likelihood of the observed data plus the latent or missing data. Then, the algorithm is iterated between the E- and CM-steps, where in the E-step, the expectation of the complete-data log-likelihood, called Q-function, is computed, and in the CM-step, parameters are updated by maximizing the Q-function. To facilitate the procedure of the ECM algorithm, the following scaling transformations (Liu and Lin, 2015) are considered

$$\tilde{\boldsymbol{B}} \triangleq \boldsymbol{B}\boldsymbol{\Lambda}^{-1/2} \quad \text{and} \quad \tilde{\boldsymbol{U}}_j \triangleq \boldsymbol{\Lambda}^{1/2}\boldsymbol{U}_j. \tag{12}$$

By the hierarchical representation (11) and scaling transformations (12), the MMNFA model can alternatively be represented by

$$\boldsymbol{Y}_j \mid \tilde{\boldsymbol{U}} = \tilde{\boldsymbol{U}}_j \sim \mathcal{N}_p(\boldsymbol{\mu} + \tilde{\boldsymbol{B}}\tilde{\boldsymbol{U}}_j, \boldsymbol{D}), \qquad \tilde{\boldsymbol{U}}_j \mid W = w_j \sim \mathcal{N}_q((w_j - a_\nu)\boldsymbol{\lambda}, \boldsymbol{I}_q), \qquad W_j \sim h(w_j; \boldsymbol{\nu}). \tag{13}$$

It is straightforward to see $\tilde{\boldsymbol{U}}_j \mid (\boldsymbol{Y} = \boldsymbol{y}_j, W = w_j) \sim \mathcal{N}_q(\boldsymbol{q}_j, \boldsymbol{C})$, and to obtain the conditional distribution of W_j given $\boldsymbol{y}_j$ by Bayes' rule as

$$f(w_j \mid \boldsymbol{Y} = \boldsymbol{y}_j) = \frac{\phi(\boldsymbol{y}_j; \boldsymbol{\mu} - a_\nu\tilde{\boldsymbol{B}}\boldsymbol{\lambda} + w_j\tilde{\boldsymbol{B}}\boldsymbol{\lambda}, \boldsymbol{\Sigma}) f(w_j)}{f_{\mathrm{MMN}}(\boldsymbol{y}_j; \boldsymbol{\mu} - a_\nu\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\eta}, \boldsymbol{\nu})}, \tag{14}$$

where

$$\boldsymbol{q}_j = \boldsymbol{C}\left\{\boldsymbol{\xi}_j + \boldsymbol{\lambda}(w_j - a_\nu)\right\}, \quad \boldsymbol{\xi}_j = \tilde{\boldsymbol{B}}^\top\boldsymbol{D}^{-1}(\boldsymbol{y}_j - \boldsymbol{\mu}) \text{ and } \boldsymbol{C} = (\boldsymbol{I}_q + \tilde{\boldsymbol{B}}^\top\boldsymbol{D}^{-1}\tilde{\boldsymbol{B}})^{-1}. \tag{15}$$

As a result of (13), the complete-data log-likelihood function for $\boldsymbol{\Theta} = (\boldsymbol{\mu}, \boldsymbol{B}, \boldsymbol{D}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ associated with the observed data $\boldsymbol{y} = (\boldsymbol{y}_i, \ldots, \boldsymbol{y}_n)^\top$, missing value $\tilde{\boldsymbol{U}} = (\tilde{\boldsymbol{U}}_1^\top, \ldots, \tilde{\boldsymbol{U}}_n^\top)^\top$ and latent variable $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$, ignoring additive constants, is

$$\ell_c(\boldsymbol{\Theta}\big|\boldsymbol{y}, \tilde{\boldsymbol{U}}, \boldsymbol{w}) = \sum_{j=1}^n \log h(w_j; \boldsymbol{\nu}) - \frac{n}{2}\log|\boldsymbol{D}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{D}^{-1}\sum_{j=1}^n \boldsymbol{\Upsilon}_j\right) - \frac{1}{2}\sum_{j=1}^n\left\{(w_j^2 - 2w_ja_\nu + a_\nu^2)\boldsymbol{\lambda}^\top\boldsymbol{\lambda} - 2(w_j\tilde{\boldsymbol{U}}_j - a_\nu\tilde{\boldsymbol{U}}_j)^\top\boldsymbol{\lambda}\right\},$$

where $\boldsymbol{\Upsilon}_j = (\boldsymbol{y}_j - \boldsymbol{\mu} - \tilde{\boldsymbol{B}}\tilde{\boldsymbol{U}}_j)(\boldsymbol{y}_j - \boldsymbol{\mu} - \tilde{\boldsymbol{B}}\tilde{\boldsymbol{U}}_j)^\top$, and $\mathrm{tr}(\boldsymbol{M})$ denotes the trace of matrix $\boldsymbol{M}$.

Proposition 2. The following conditional expectations can be established from (13),

$$\begin{aligned}
E(\tilde{\boldsymbol{U}}_j \mid \boldsymbol{y}_j) &= \boldsymbol{C}\left\{\boldsymbol{\xi}_j + \boldsymbol{\lambda}\left(E(W_j \mid \boldsymbol{y}_j) - a_\nu\right)\right\},\\
E(W_j\tilde{\boldsymbol{U}}_j \mid \boldsymbol{y}_j) &= \boldsymbol{C}\left\{\boldsymbol{\xi}_jE(W_j \mid \boldsymbol{y}_j) + \boldsymbol{\lambda}(E(W_j^2 \mid \boldsymbol{y}_j) - a_\nu E(W_j \mid \boldsymbol{y}_j))\right\},\\
E(\tilde{\boldsymbol{U}}_j\tilde{\boldsymbol{U}}_j^\top \mid \boldsymbol{y}_j) &= \left\{E(\tilde{\boldsymbol{U}}_j \mid \boldsymbol{y}_j)\boldsymbol{\xi}_j^\top + [E(W_j\tilde{\boldsymbol{U}}_j \mid \boldsymbol{y}_j) - a_\nu E(\tilde{\boldsymbol{U}}_j \mid \boldsymbol{y}_j)]\boldsymbol{\lambda}^\top + \boldsymbol{I}_q\right\}\boldsymbol{C},
\end{aligned}$$

where $\boldsymbol{\xi}_j$ and $\boldsymbol{C}$ are defined in (15).

Proof. The proof is straightforward using the posterior distributions given in (14). □

Now, the ECM algorithm for ML estimation of the MMNFA model proceeds as follows:

- E-step: At the kth iteration, the Q-function is computed with $\boldsymbol{\Theta}$ evaluated at $\hat{\boldsymbol{\Theta}}^{(k)}$ as

$$\begin{aligned}
Q(\boldsymbol{\Theta} \mid \hat{\boldsymbol{\Theta}}^{(k)}) \;=\; & \sum_{j=1}^n E\left(\log h(W_j; \boldsymbol{\nu}) \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)}\right) - \frac{n}{2}\log|\boldsymbol{D}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{D}^{-1}\sum_{j=1}^n \boldsymbol{\Upsilon}_j^{(k)}\right)\\
& -\frac{1}{2}\sum_{j=1}^n\left\{(\hat{t}_j^{(k)} - 2\hat{w}_j^{(k)}a_\nu + a_\nu^2)\boldsymbol{\lambda}^\top\boldsymbol{\lambda} - 2(\hat{\boldsymbol{\zeta}}_{1j}^{(k)} - a_\nu\hat{\boldsymbol{\zeta}}_{0j}^{(k)})^\top\boldsymbol{\lambda}\right\},
\end{aligned} \tag{16}$$

where the necessary conditional expectations obtained by Proposition 2 are

$$\begin{aligned}
\hat{w}_j^{(k)} =& E(W_j \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)}), \quad \hat{t}_j^{(k)} = E(W_j^2 \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)}), \quad E(\log h(W_j;\boldsymbol{\nu}) \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)}),\\
\hat{\boldsymbol{\zeta}}_{0j}^{(k)} =& E(\tilde{\boldsymbol{U}}_j \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)}), \quad \hat{\boldsymbol{\zeta}}_{1j}^{(k)} = E\left(W_i\tilde{\boldsymbol{U}}_j \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)}\right), \quad \hat{\boldsymbol{\Omega}}_j^{(k)} = E\left(\tilde{\boldsymbol{U}}_j\tilde{\boldsymbol{U}}_j^\top \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)}\right),
\end{aligned} \tag{17}$$

and

$$\boldsymbol{\Upsilon}_j^{(k)} = (\boldsymbol{y}_j - \boldsymbol{\mu})(\boldsymbol{y}_j - \boldsymbol{\mu})^\top - \tilde{\boldsymbol{B}}\hat{\boldsymbol{\zeta}}_{0j}^{(k)}(\boldsymbol{y}_j - \boldsymbol{\mu})^\top - (\boldsymbol{y}_j - \boldsymbol{\mu})\hat{\boldsymbol{\zeta}}_{0j}^{(k)}\tilde{\boldsymbol{B}}^\top + \tilde{\boldsymbol{B}}\hat{\boldsymbol{\Omega}}_j^{(k)}\tilde{\boldsymbol{B}}^\top, \tag{18}$$

which contains unknown parameters $\boldsymbol{\mu}$ and $\tilde{\boldsymbol{B}}$. Note that the calculation of $E(W_j^r \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)})$ and $E(\log h(W_j;\boldsymbol{\nu}) \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)})$ critically depends on $h(w;\boldsymbol{\nu})$.

- CM-step 1: Maximizing (16) over $\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{B}$ and $\boldsymbol{D}$ leads to the following CM estimators:

$$\begin{aligned}
\hat{\boldsymbol{\mu}}^{(k+1)} &= \frac{\sum_{j=1}^n(\boldsymbol{y}_j - \hat{\tilde{\boldsymbol{B}}}^{(k)}\hat{\boldsymbol{\zeta}}_{0j}^{(k)})}{n}, \qquad \hat{\boldsymbol{\lambda}}^{(k+1)} = \frac{\sum_{j=1}^n\left(\hat{\boldsymbol{\zeta}}_{1j}^{(k)} - \hat{a}_{\boldsymbol{\nu}}^{(k)}\hat{\boldsymbol{\zeta}}_{0j}^{(k)}\right)}{\sum_{j=1}^n\left(\hat{t}_j^{(k)} - 2\hat{w}_j^{(k)}\hat{a}_{\boldsymbol{\nu}}^{(k)} + \hat{a}_{\boldsymbol{\nu}}^{2(k)}\right)},\\
\hat{\tilde{\boldsymbol{B}}}^{(k+1)} &= \left(\sum_{j=1}^n(\boldsymbol{y}_j - \hat{\boldsymbol{\mu}}^{(k+1)})\hat{\boldsymbol{\zeta}}_{0j}^{(k)\top}\right)\left(\sum_{j=1}^n\hat{\boldsymbol{\Omega}}_j^{(k)}\right)^{-1}, \qquad \hat{\boldsymbol{D}}^{(k+1)} = \frac{1}{n}\mathrm{diag}\left(\sum_{j=1}^n\hat{\boldsymbol{\Upsilon}}_j^{(k)}\right),
\end{aligned}$$

  where $\hat{a}_{\boldsymbol{\nu}}^{(k)} = E(W_j)\big|_{\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}^{(k)}}$, $\hat{\boldsymbol{\Upsilon}}^{(k)}$ is obtained by substituting $\hat{\boldsymbol{\mu}}^{(k+1)}$ and $\hat{\tilde{\boldsymbol{B}}}^{(k+1)}$ into (18). Then, the factor loading matrix before transformation is $\hat{\boldsymbol{B}}^{(k+1)} = \hat{\tilde{\boldsymbol{B}}}^{(k+1)}\hat{\boldsymbol{\Lambda}}^{1/2(k+1)}$, where $\hat{\boldsymbol{\Lambda}}^{(k)} = \boldsymbol{I}_q + b_{\boldsymbol{\nu}}\hat{\boldsymbol{\lambda}}^{(k+1)}\hat{\boldsymbol{\lambda}}^{(k+1)\top}$ with $b_{\boldsymbol{\nu}}$ evaluated at $\hat{\boldsymbol{\nu}}^{(k)}$.

- CM-step 2: The update of $\boldsymbol{\nu}$ depends on the chosen distribution for W and is obtained by

$$\hat{\boldsymbol{\nu}}^{(k+1)} = \arg\max_{\boldsymbol{\nu}} \sum_{j=1}^n E(\log h(W_j;\boldsymbol{\nu}) \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)}) - \frac{1}{2}\sum_{j=1}^n\left\{(\hat{t}_j^{(k)} - 2\hat{w}_j^{(k)}a_{\boldsymbol{\nu}} + a_{\boldsymbol{\nu}}^2)\hat{\boldsymbol{\lambda}}^{(k+1)\top}\hat{\boldsymbol{\lambda}}^{(k+1)} - 2(\hat{\boldsymbol{\zeta}}_{1j}^{(k)} - a_{\boldsymbol{\nu}}\hat{\boldsymbol{\zeta}}_{0j}^{(k)})^\top\hat{\boldsymbol{\lambda}}^{(k+1)}\right\}.$$

  This maximization can be achieved by using some built-in R functions such as `optim` and `nlminb` whenever $h(\cdot;\boldsymbol{\nu})$ has a complicated form.

The above E- and CM-steps are iterated until either the number of iterations exceeds the maximum limit or a suitable convergence rule is achieved. Denote the resulting ML estimates upon convergence by $\hat{\boldsymbol{\Theta}} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{D}}, \hat{\boldsymbol{\nu}})$. Then, the prediction of the conditional factor scores is $\hat{\boldsymbol{U}}_i = E(\boldsymbol{U} \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}) = \hat{\boldsymbol{\Lambda}}^{-1/2}\hat{\boldsymbol{\zeta}}_{0i}$, where $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\zeta}}_{0j}$ are calculated using $\boldsymbol{\Lambda} = \boldsymbol{I}_q + b_{\boldsymbol{\nu}}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top$ and (17), respectively, with $\boldsymbol{\Theta}$ evaluated at $\hat{\boldsymbol{\Theta}}$.

3.3. Special cases of the MMNFA model

If $W \sim \mathcal{TN}(0, 1; (0, \infty))$ in (10), the rSNFA model is obtained. The necessary conditional expectations involved in (16) and (17) for the rSNFA model can be computed by Lemma 2 and since it is free of parameter $\boldsymbol{\nu}$, it is not necessary to obtain the conditional expectation $E(\log h(W_j;\boldsymbol{\nu}) \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)})$. More details can be found in Liu and Lin (2015).

Let W in (10) follow $\mathcal{E}(1)$. Then, we have $a_{\boldsymbol{\nu}} = b_{\boldsymbol{\nu}} = 1$ and consequently $\boldsymbol{Y}_j \sim \mathcal{MMNE}_P(\boldsymbol{\mu} - \boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\eta})$ where $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Lambda}^{-1}\boldsymbol{B}^\top + \boldsymbol{D}$ and $\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\lambda}$. The obtained factor model is named exponentiated MMNFA, abbreviated as MMNEFA. The necessary conditional expectations involved in (16) and (17) for MMNEFA can be computed via Theorem 3. Note also that the MMNEFA model is free of the mixing parameter, $\boldsymbol{\nu}$, and so it is unnecessary to obtain the conditional expectation $E(\log h(W_j;\boldsymbol{\nu}) \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)})$.

However, if W in (10) has PDF (7), then the scaling coefficients reduce to $a_\nu = \nu(2-\sqrt{2/\pi})+\sqrt{2/\pi}$, $b_\nu = 7\nu+1-a_\nu^2$ and the half-normal exponentiated MMNFA (MMNEHFA) model is obtained. In this case, $\boldsymbol{Y}_j \sim \mathcal{MMNEH}_P(\boldsymbol{\mu} - a_\nu\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\eta}, \nu)$. The necessary conditional expectations involved in (16) and (17) for MMNEHFA can be computed by Theorem 4.

Remark 2. There is no closed-form solution for updating the mixing parameter ν of the MMNEHFA model, since the conditional expectation $E(\log h(W_j;\nu) \mid \boldsymbol{y}_j, \hat{\boldsymbol{\Theta}}^{(k)})$ is complicated. In this case (and other similar cases), ν can be updated by implementing an extension of the EM and ECM algorithms, namely the expectation-conditional maximization either (ECME; Liu and Rubin (1994)). In the CM-step of the ECME approach, the parameters are updated by maximizing either the Q-function or the corresponding constrained actual likelihood function. The so-called 'CML-step' is adopted here to maximize the restricted actual log-likelihood function. That is, the update of ν is now expressed as

$$\hat{\nu}^{(k+1)} = \arg\max_{\nu\in[0,1]} \sum_{j=1}^{n} \log f_{\mathrm{MMNEH}}(\boldsymbol{y}_j; \hat{\boldsymbol{\mu}}^{(k+1)} - a_\nu \hat{\boldsymbol{\eta}}^{(k+1)}, \hat{\boldsymbol{\Sigma}}^{(k+1)}, \hat{\boldsymbol{\eta}}^{(k+1)}, \nu).$$

A one-dimensional search in the MMNEHFA model is preformed by implementing the `optim` function of the statistical software R. Through a simulation study described in Section 4.3, it is shown that this optimization works well for empirical studies.

3.4. Notes on implementation

Aitken's acceleration method (Aitken, 1926) with per-user-defined tolerance, $\epsilon = 10^{-5}$, is exploited to determine whether the ECM algorithm has achieved convergence (see McLachlan and Krishnan (2008) for more details). It is well known that the choice of starting points plays an important role in the EM-type algorithm. Since the MMNFA model includes the original FA model as a special case, we set $\hat{\boldsymbol{\lambda}}^{(0)} = \mathbf{0}$ and $\hat{\nu}^{(0)}$ corresponding to an initial assumption near to normality. Then, by fitting the FA model to the data, reasonable initial values of the mean vector $\hat{\boldsymbol{\mu}}^{(0)}$, factor loading matrix $\hat{\boldsymbol{B}}^{(0)}$ and error covariance matrix $\hat{\boldsymbol{D}}^{(0)}$ can be obtained. The R command "**factanal**" is used for fitting the FA model.

In the data analysis, two well-known model selection criteria is to be used, which take the form of the penalized log-likelihood $mC(n) - 2\ell_{max}$, to compare models and to determine an appropriate value for q. Here, ℓ_{max} is the maximized log-likelihood, m is the number of parameters in the considered model, and the factor $C(n)$ equals to 2 for the Akaike information criterion (AIC) and to $\log(n)$ for the Bayesian information criterion (BIC).

4. Monte Carlo simulation studies

4.1. Model performance in dealing with skewed and leptokurtic simulated data

A simulation study is conducted to examine how well the MMN-based FA models work in the presence of asymmetrical features in the data. Following Lin et al. (2015), artificial datasets of sizes $n = 100, 300$ are generated from the FA model by assuming non-normal distribution for the latent factors. In each replication of 100 Monte Carlo (MC) samples, let $p = 10$ and 50, three numbers of factor $q = 2$, 3 and 4, and the parameter values $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{B} = Unif(p, q)$, and $\boldsymbol{D} = \mathrm{diag}\{Unif(p, p)\}$, in which $Unif(p, q)$ denotes a matrix of random numbers, with dimension $p \times q$ uniformly drawn from the unit interval $(0, 1)$. To add various degrees of skewness and kurtosis, the latent factors $\boldsymbol{U}$ are generated from the beta distribution with shape parameters $\alpha = 0.1$ and $\beta = 30$, $Beta(0.1, 30)$, and Chi-square distribution with one degree of freedom ($\chi^2_{(1)}$). Therefore, the population skewness/kurtosis of U equals 6/52 for $Beta(0.1, 30)$ and 2.8/12 for χ^2_1. Random samples generated from the multivariate normal distribution, with zero mean and scale covariance $\boldsymbol{D}$, are also considered as errors.

Assuming the number of latent factors is known, Table 1 summarizes the results of fitting MMNEFA, MMNEHFA and rSNFA models, including the average of the BIC values, required CPU time (in second), together with the frequencies of the particular model chosen based on the smallest BIC value, by considering $q = 2, 3$ and 4 for each simulated dataset. The number of parameters involved in the MMNEFA, MMNEHFA and rSNFA models is reported in Table 1 of the Online Supplement. The model comparison results displayed in Table 1 suggest that the MMNEFA model provides a better fit than the others for the χ^2_1 data generator (in all 24 scenarios), while, the MMNEHFA works much better than the other two models for the $Beta(0.1, 30)$ data generator. Based on the CPU time, it can be concluded that the MMNEFA model is, in average, faster than rSN and MMNEH models.

Table 1: Results of the first simulation study based on 100 replications.

n	p	q		χ^2_1			$Beta(0.1, 30)$		
				MMNEFA	MMNEHFA	rSNFA	MMNEFA	MMNEHFA	rSNFA
100	10	2	Mean	2283.22	2288.15	2286.05	3836.78	3833.17	3850.05
			Freq.	70	4	26	22	70	8
			CPU time	0.70	9.90	0.90	0.20	5.40	0.30
		3	Mean	2306.37	2310.03	2308.17	3853.52	3852.68	3871.7
			Freq.	78	4	18	23	76	1
			CPU time	0.80	10.10	1.0	0.30	6.50	0.40
		4	Mean	2335.27	2339.12	2336.9	3877.24	3875.73	3896.16
			Freq.	77	6	17	22	76	2
			CPU time	0.80	10.90	1.00	0.30	7.40	0.50
100	50	2	Mean	11358.79	11362.12	11361.25	16771.9	16761.04	16794.64
			Freq.	78	3	19	42	53	5
			CPU time	2.1	32.60	2.40	0.70	11.00	1.00
		3	Mean	11512.74	11515.98	11515.45	16824.24	16811.01	16859.1
			Freq.	83	7	10	15	84	1
			CPU time	2.00	31.40	2.3	0.90	12.90	1.10
		4	Mean	11664.48	11668.18	11667.78	16908.47	16894.25	16942.91
			Freq.	85	5	8	7	93	0
			CPU time	2.00	31.20	2.30	1.00	15.20	1.20
300	10	2	Mean	6628.65	6632.69	6632.08	10242.86	10238.74	10281.86
			Freq.	85	3	12	19	75	6
			CPU time	1.30	19.10	1.30	0.30	7.70	0.30
		3	Mean	6669.01	6673.14	6672.53	10279.35	10270.04	10310.05
			Freq.	86	3	11	9	91	0
			CPU time	1.30	18.50	1.30	0.40	9.20	0.50
		4	Mean	6706.86	6711.23	6710.77	10299.31	10292.48	10342.4
			Freq.	88	2	10	2	98	0
			CPU time	1.41	18.10	1.97	0.50	10.60	0.60
300	50	2	Mean	33007.32	33019.02	33016.40	49002.36	48977.10	49082.37
			Freq.	86	4	10	17	83	0
			CPU time	3.80	62.5	3.90	1.50	21.90	1.7
		3	Mean	33214.02	33236.72	33231.04	49173.35	49145.64	49208.27
			Freq.	92	2	6	2	98	0
			CPU time	3.90	67.20	4.00	1.60	23.7	2.10
		4	Mean	33418.61	33433.01	33424.49	49279.12	49250.88	49350.10
			Freq.	96	2	2	2	98	0
			CPU time	4.30	72.30	6.0	2.10	25.80	2.70

4.2. Comparison of fitting under different degrees of freedom of the rSTFA model

To demonstrate the performance of the proposed factor model, the second comprehensive simulation study is conducted. Consider five-dimensional artificial data with $n = 150$ observations generated from an rSTFA model (Lin et al., 2015). The presumed parameters are $\boldsymbol{\mu}^\top = (10, 20, 30, 40, 50)$, $\boldsymbol{D} = \text{diag}\{1, 2, 3, 4, 4\}$, and

$$\boldsymbol{B}^\top = \begin{pmatrix} 3 & 3 & 3 & 4 & 7 \\ 0 & 4 & 6 & 8 & 9 \end{pmatrix}.$$

Also, to achieve various levels of skewness and kurtosis, consider the degree of freedom $\nu \in \{4, 10, 15, 20, 30, 40\}$ and two scenarios designed as

$$\text{Scenario 1: } (\lambda_1, \lambda_2) = (2, 6), \qquad \text{Scenario 2: } (\lambda_1, \lambda_2) = (3, 3).$$

The performance of the rSNFA and rSTFA models are compared with the proposed MMNE and MMNEH factor analyzers. Over 100 trials, Table 2 summarizes the average of the AIC and BIC values of the considered models, their corresponding standard deviations (Std.), together with the frequencies of the particular model chosen by the smallest AIC and BIC values, by considering $q = 2$ for each simulated dataset. The required CPU time is also recorded in the table. As the true model is always expected to have the best performance, the results depicted in Table 2 show that the rSTFA model works well for small degrees of freedom $\nu = 4$ and 10. It is observed that in these cases, i.e. $\nu = 4, 10$, the MMNEFA model is the second-best performing model. However, as the value of ν increases, both the rSTFA and rSNFA models approach the same estimation results, and thus the rSNFA model might outperform the rSTFA model. But, it is clear that when ν exceeds 10, the MMNEFA model provides a better fit than the others with the smallest AIC and BIC in both scenarios. Figures 2 and 3 in the Online Supplement display the density contours of the fitted bivariate

Table 2: Comparison of the rSNFA, rSTFA, MMNEFA and MMNEHFA models based on 100 MC samples generated from the rSTFA model.

			Scenario 1				Scenario 2			
ν	Criterion		rSNFA	rSTFA	MMNEFA	MMNEHFA	rSNFA	rSTFA	MMNEFA	MMNEHFA
$\nu = 4$	AIC	Mean	3401.24	3265.96	3366.91	3364.27	3377.94	3241.85	3343.08	3341.92
		Std.	212.31	188.85	209.52	208.06	218.05	199.67	215.05	214.54
		Freq.	0	100	0	0	0	100	0	0
	BIC	Mean	3464.47	3332.19	3430.13	3430.51	3441.16	3308.08	3406.31	3408.15
		Std.	212.31	188.85	209.52	208.06	218.05	199.67	215.05	214.54
		Freq.	0	100	0	0	0	100	0	0
		CPU time	1.71	8.68	1.62	30.70	1.67	7.93	1.20	23.36
$\nu = 10$	AIC	Mean	3080.85	3061.67	3066.77	3070.06	3116.26	3097.05	3105.45	3110.07
		Std.	165.32	163.66	167.50	166.62	172.52	170.61	173.39	173.41
		Freq.	0	59	33	8	2	79	16	3
	BIC	Mean	3144.07	3127.91	3129.99	3136.30	3179.48	3163.28	3168.68	3176.30
		Std.	165.32	163.66	167.50	166.62	172.52	170.61	173.39	173.41
		Freq.	0	53	46	1	2	67	29	2
		CPU time	1.75	7.44	1.57	37.53	1.63	6.92	0.95	33.58
$\nu = 15$	AIC	Mean	3056.20	3049.03	3048.11	3052.92	3011.41	3001.59	3003.13	3008.93
		Std.	191.64	190.96	192.42	190.76	203.03	203.54	204.17	203.59
		Freq.	1	41	53	5	1	51	45	3
	BIC	Mean	3119.43	3115.26	3111.33	3119.15	3074.64	3067.83	3066.36	3075.17
		Std.	191.64	190.96	192.42	190.76	203.03	203.54	204.17	203.59
		Freq.	8	27	64	1	5	39	55	1
		CPU time	1.74	7.27	1.59	38.36	1.44	5.95	0.99	33.27
$\nu = 20$	AIC	Mean	3039.65	3036.86	3032.11	3037.21	2965.83	2961.72	2958.96	2964.95
		Std.	143.13	142.72	144.38	143.50	185.84	186.03	185.41	184.60
		Freq.	9	20	64	7	8	34	58	0
	BIC	Mean	3102.88	3103.09	3095.34	3103.44	3029.05	3027.95	3022.19	3031.18
		Std.	143.13	142.72	144.38	143.50	185.84	186.03	185.41	184.60
		Freq.	11	6	82	1	11	17	71	1
		CPU time	1.66	6.85	1.51	37.06	1.41	5.63	1.05	34.99
$\nu = 30$	AIC	Mean	2981.47	2980.69	2973.46	2978.81	2999.34	2998.24	2995.84	3002.29
		Std.	197.40	197.52	200.03	199.24	163.03	163.13	163.53	163.62
		Freq.	9	15	70	6	18	19	62	1
	BIC	Mean	3044.69	3046.92	3036.68	3045.04	3062.56	3064.48	3059.06	3068.52
		Std.	197.40	197.52	200.03	199.24	163.03	163.13	163.53	163.62
		Freq.	13	8	77	2	23	9	67	1
		CPU time	1.62	7.74	1.54	37.39	1.49	7.76	1.13	38.37
$\nu = 40$	AIC	Mean	2985.14	2985.08	2978.86	2984.20	2946.52	2946.75	2943.39	2950.31
		Std.	180.58	180.09	180.73	178.94	206.25	206.24	205.84	205.16
		Freq.	20	9	65	6	26	16	58	0
	BIC	Mean	3048.36	3051.31	3042.08	3050.44	3009.74	3012.99	3006.61	3016.54
		Std.	180.58	180.09	180.73	178.94	206.25	206.24	205.84	205.16
		Freq.	18	4	77	1	28	4	68	0
		CPU time	1.61	9.71	1.53	37.25	1.33	8.19	1.21	38.58

rMSN, rMST, MMNE and MMNEH distributions, together with two summary histograms and nonparametric density curves of their marginal distributions. For both scenarios 1 and 2, better performance of the MMNEFA model is confirmed for large values of ν. Furthermore, as expected, since the rSNFA and MMNEFA are free of the additional parameter $\boldsymbol{\nu}$, the allocated CPU time for them is much smaller than for the rSTFA and MMNEHFA models.

4.3. Finite sample properties of ML estimates

In this experiment, 500 MC artificial samples are generated from each of the MMNEFA and MMNEHFA models with the same presumed true parameter values $\boldsymbol{\mu}^\top = (10, 20, 30)$, $\boldsymbol{B}^\top = (2, 4, 6)$, $\boldsymbol{D} = (0.6, 0.4, 0.8)\boldsymbol{I}_3$, $\lambda = 3$ and $\nu = 0.4$. The data are simulated by applying the stochastic representation in (10), where the chosen sample size n is varied from 100 to 500, 2000 and 4000. For each synthetic data set generated from the MMNEFA or MMNEHFA models, the corresponding model is fitted and the parameter estimates are obtained. Tables 3 and 4 report the average values and the corresponding Std. of the ECM-based estimates across all samples for the MMNEFA and MMNEHFA models, respectively. Moreover, in order to examine the performance of the ML estimates for each sample size and for each parameter, the absolute bias (AB) and the mean squared error (MSE) is determined

$$\text{AB} = \frac{1}{500}\sum_{i=1}^{500}\left|\hat{\theta}^{(i)} - \theta_{true}\right| \qquad \text{and} \qquad \text{MSE} = \frac{1}{500}\sum_{i=1}^{500}\left(\hat{\theta}^{(i)} - \theta_{true}\right)^2,$$

Table 3: Mean, Std., AB and MSE of the ML estimates over 500 MC samples generated from the MMNEFA model (true parameter in parentheses).

n	Measure	$\mu_1(10)$	$\mu_2(20)$	$\mu_3(30)$	$b_1(2)$	$b_2(4)$	$b_3(6)$	$\sigma_1^2(0.6)$	$\sigma_2^2(0.4)$	$\sigma_3^2(0.8)$	$\lambda(3)$
100	Mean	9.9734	19.9354	29.9051	2.0212	3.9813	5.9704	0.5767	0.4216	0.7505	2.9703
	Std.	0.2144	0.4104	0.6061	0.1953	0.3618	0.5479	0.0924	0.2404	0.3198	1.4520
	AB	0.0266	0.0646	0.0949	0.0212	0.0197	0.0596	0.0233	0.0316	0.0705	1.0993
	MSE	0.0462	0.1710	0.3728	0.0378	0.1300	0.2974	0.0090	0.1808	0.0950	1.0209
500	Mean	9.9829	19.9886	29.9742	2.0106	4.0149	6.0381	0.5929	0.4155	0.7728	3.0241
	Std.	0.0947	0.1707	0.2527	0.0824	0.1549	0.2302	0.0385	0.1910	0.1403	1.1747
	AB	0.0071	0.0114	0.0358	0.0106	0.0149	0.0381	0.0071	0.0155	0.0328	0.7241
	MSE	0.0089	0.0290	0.0635	0.0068	0.0240	0.0539	0.0015	0.1682	0.0368	0.5090
2000	Mean	10.0045	20.0074	30.0141	2.0045	4.0019	6.0108	0.5978	0.4066	0.7889	3.0029
	Std.	0.0468	0.0808	0.1248	0.0406	0.0805	0.1168	0.0218	0.1388	0.0833	0.5793
	AB	0.0045	0.0074	0.0141	0.0045	0.0019	0.0108	0.0022	0.0066	0.0189	0.4089
	MSE	0.0022	0.0065	0.0156	0.0017	0.0064	0.0136	0.0005	0.0968	0.0153	0.2041
4000	Mean	10.0016	20.0027	30.0056	2.0009	3.9992	6.0014	0.5993	0.4034	0.7926	3.0010
	Std.	0.0347	0.0606	0.0933	0.0317	0.0634	0.0959	0.0143	0.0966	0.0571	0.3595
	AB	0.0016	0.0027	0.0056	0.0009	0.0008	0.0014	0.0007	0.0034	0.0096	0.3040
	MSE	0.0012	0.0036	0.0086	0.0010	0.0040	0.0091	0.0002	0.0535	0.0115	0.1025

Table 4: Mean, Std., AB and MSE of the ML estimates over 500 MC samples generated from the MMNEHFA model (true parameter in parentheses).

n	Measure	$\mu_1(10)$	$\mu_2(20)$	$\mu_3(30)$	$b_1(2)$	$b_2(4)$	$b_3(6)$	$\sigma_1^2(0.6)$	$\sigma_2^2(0.4)$	$\sigma_3^2(0.8)$	$\lambda(3)$	$\nu(0.4)$
100	Mean	10.2682	20.5464	30.7745	2.5537	4.5902	6.6550	0.5335	0.3551	0.7466	2.5811	0.4734
	Std.	0.4082	0.8144	1.2181	0.2510	0.5081	0.7444	0.1076	0.1554	0.3167	1.7584	0.2443
	AB	0.0382	0.0664	0.0745	0.0537	0.0402	0.0550	0.0335	0.0951	0.0966	1.1811	0.2034
	MSE	0.0472	0.1263	0.2247	0.0690	0.1443	0.0878	0.0215	0.0800	0.0662	1.6031	0.0219
500	Mean	10.1065	20.2169	30.3266	2.3207	4.2337	6.3585	0.5574	0.3743	0.7608	2.7214	0.4543
	Std.	0.1622	0.3213	0.4794	0.1277	0.2549	0.3721	0.0404	0.0656	0.1402	1.1462	0.1225
	AB	0.0265	0.0369	0.0466	0.0207	0.0137	0.0385	0.0126	0.0543	0.0568	0.9314	0.0943
	MSE	0.0294	0.0493	0.0844	0.0373	0.0429	0.0362	0.0116	0.0597	0.0244	1.0240	0.0132
2000	Mean	9.9955	19.9295	29.9469	2.0280	4.0550	6.0854	0.5989	0.3957	0.8044	2.8207	0.4113
	Std.	0.1228	0.2403	0.3618	0.0665	0.1340	0.2008	0.0212	0.0325	0.0623	0.7625	0.1086
	AB	0.0095	0.0125	0.0131	0.0089	0.0075	0.0154	0.0091	0.0257	0.0144	0.5207	0.0313
	MSE	0.0091	0.0163	0.0338	0.0132	0.0109	0.0103	0.0064	0.0176	0.0109	0.6775	0.0089
4000	Mean	10.0090	19.9991	29.9988	2.0024	4.0040	6.0074	0.6009	0.3994	0.8003	3.0846	0.4098
	Std.	0.0757	0.1504	0.2270	0.0461	0.0913	0.1393	0.0152	0.0250	0.0482	0.6525	0.0562
	AB	0.0040	0.0059	0.0092	0.0055	0.0061	0.0074	0.0069	0.0174	0.0103	0.3846	0.0198
	MSE	0.0032	0.0062	0.0127	0.0096	0.0093	0.0086	0.0042	0.0125	0.0088	0.4183	0.0040

where $\hat{\theta}^{(i)}$ is the ML estimate of θ_{true} obtained from the ith replicate. It can be observed from both Tables 3 and 4 that the AB and MSE values approach zero as n increases, showing empirically the asymptotic unbiasedness and the consistency of the ML estimates obtained via the ECM-based algorithm.

5. Real data analysis

5.1. Wine recognition data

Firstly, the proposed methodology is applied to the Italian wine recognition dataset. The wine dataset is available in the UCI Machine Learning Repository (archive.ics.uci.edu/ml) and comprises 13-dimensional chemical measurements of $n = 178$ Italian wines grown in three different cultivars (groups), Barolo, Grignolino and Barbera, with sizes 59, 71 and 48, respectively. In this analysis, the focus is solely on the Barbera group. Table 5 summarizes basic descriptive statistics of the 13 attributes, including their sample skewness, kurtosis and p-values of the Kolmogorov-Smirnov (KS) and r_n^* (Rodríguez and Alva, 2010) tests for marginal normality and skew-normality, respectively. The results depicted in Table 5 show that for the considered data most of the attributes are moderately skewed. Moreover, the p-values of the KS test significantly suggest that not all of the 13 measures follow the normal distribution, but there is enough evidence in favour of the skew-normal (SN) distribution based on the r_n^* test for all attributes. In the multivariate perspective, by applying the generalized Shapiro-Wilk test for multivariate normality (GSW; Alva and Estrada (2009)) and the canonical-based test for multivariate skew-normality (CSN; Balakrishnan et al. (2014)), it is suggested that the multivariate normality assumption be rejected in favour of the multivariate SN distribution. The p-values corresponding to the test statistics are GSW = 0.0444 and CSN = 0.4610.

Using the "regression" method (see Chapter 9.5 of Johnson and Wichern (2007)), three factor score estimates are obtained from the classical FA model with $q = 3$. Figure 2 in the Online Supplement shows the histogram and

Table 5: An overview of 13 attributes of Barbera data with the p-values of the KS and r_n^* tests.

Variable	Description	Skewness	Kurtosis	KS	r_n^*
y_1	Alcohol	0.147	-0.666	0	0.448
y_2	Malic acid	0.098	-0.422	0	0.193
y_3	Ash	0.353	-0.832	0	0.503
y_4	Alkalinity of ash	0.453	-0.594	0	0.408
y_5	Magnesium	0.524	-0.617	0	0.567
y_6	Total phenols	0.988	1.298	0	0.749
y_7	Flavanoids	0.977	-0.003	0	0.288
y_8	Nonflavanoid phenols	-0.515	-0.603	0	0.426
y_9	Proanthocyanins	1.523	3.426	0	0.884
y_{10}	Color intensity	0.292	-0.828	0	0.371
y_{11}	Hue	0.572	-0.529	0	0.395
y_{12}	OD280/OD315	0.665	0.349	0	0.412
y_{13}	Proline	0.309	-0.524	0	0.390

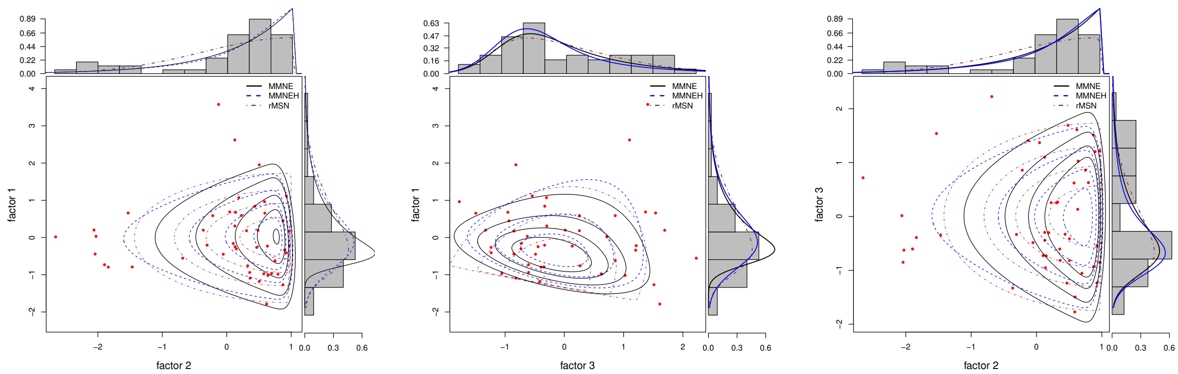

Figure 3: Scatter plots of factor scores superimposed on a set of contour lines estimated by the rMSN, MMNE and MMNEH distributions together with two summary histograms and curves of their marginal densities for the Barbera data..

corresponding normal Q-Q plots of the three FA factor score estimates that highlight serious departures of factor scores from the normality assumption. The contours of the fitted bivariate rMSN, MMNE and MMNEH distributions, together with two summary histograms and curves of their marginal distributions, are plotted in Figure 3. Theses plots reveal that skewed distributions can capture the scattering patterns relatively well. These characteristics motivate the consideration of skewed FA models, which can take both skewness and kurtosis of the data into account and are expected to showcase more appropriate statistical inference.

The FA, tFA, rSNFA, rSTFA, MMNEFA, MMNEHFA, GHSTFA and generalized hyperbolic common skew-t factor analysis (GHCSTFA; Murray et al. (2014b)) models are fitted to the original and standardized chemical measurements with q ranging from 2 to 5. The standardization is done so as to have zero mean and unit standard deviations and to avoid variables that have a greater impact due to different scales. Note that by fitting the tFA, rSTFA, GHSTFA, and GHCSTFA models, it is observed that the degree of freedom of all models tends to infinity. The detailed numerical results, including the maximized log-likelihood values, and the number of free parameters, together with the BIC are reported in Table 6. It can be observed that the MMNEFA model with $q = 3$ outperforms other competitors because it has the smallest BIC score, regardless of whether the data are standardized or not. The ML estimate of parameters and their standard error (in parentheses) for the best chosen model are presented in Table 7. The procedure for computing the standard error of the parameter estimates is presented in the Online Supplement. The estimated skewness parameters, in Table 7, are statistically significant and less than zero, revealing that the latent factors are negatively skewed. From the Varimax rotated solution of factor loadings presented in Table 7, it can be seen that the variables have positive and negative loadings on the three factors. The first factor loads heavily on y_9 and y_{10}, while the second one loads heavily on y_7, in absolute value, followed by y_8. It is known that the phenolic content of wine refers to the two phenolic compounds, the natural phenol and polyphenols (color). Moreover, the natural phenols can be broadly classified into the flavonoid and non-flavonoid categories. Therefore, the first and

Table 6: Estimation performance of eight factor models fitted to the Barbera data.

			Original data		Standardized data					Original data		Standardized data	
Model	q	m	$\ell_{\max}$	BIC	$\ell_{\max}$	BIC	Model	q	m	$\ell_{\max}$	BIC	$\ell_{\max}$	BIC
FA	2	51	-684.78	1567.00	-788.43	1774.29	tFA	2	52	-684.75	1570.81	-788.40	1778.11
	3	62	-649.80	1539.61	-753.45	1746.90		3	63	-649.82	1543.53	-753.47	1750.83
	4	72	-638.68	1556.09	-742.33	1763.39		4	73	-638.70	1560.00	-742.35	1767.30
	5	81	-627.65	1568.87	-731.30	1776.17		5	82	-627.65	1572.74	-731.30	1780.05
rSNFA	2	53	-671.81	1548.79	-775.46	1756.09	rSTFA	2	54	-671.90	1552.84	-775.55	1760.14
	3	65	-636.82	1525.26	-740.47	1732.56		3	66	-636.87	1529.23	-740.51	1736.53
	4	76	-625.89	1546.00	-729.54	1753.29		4	77	-625.93	1549.95	-729.58	1757.25
	5	86	-613.14	1559.21	-716.79	1766.51		5	87	-613.15	1563.10	-716.80	1770.40
MMNEFA	2	53	-666.93	1539.04	-770.58	1746.34	MMNEHFA	2	54	-665.24	1539.53	-768.89	1746.83
	3	65	-631.45	**1514.53**	-735.10	**1721.83**		3	66	-629.71	1514.92	-733.36	1722.22
	4	76	-620.68	1535.57	-724.33	1742.87		4	77	-618.67	1535.42	-722.32	1742.72
	5	86	-607.02	1546.96	-710.67	1754.26		5	87	-605.06	1546.91	-708.71	1754.21
GHSTFA	2	65	-653.76	1559.15	-756.29	1764.21	GHCSTFA	2	43	-889.96	1946.39	-792.52	1751.51
	3	76	-640.92	1576.05	-744.47	1783.16		3	56	-880.22	1977.24	-754.44	1725.67
	4	86	-626.76	1586.44	-730.41	1793.74		4	68	-909.78	2082.79	-751.43	1766.10
	5	95	-621.49	1610.74	-724.16	1816.09		5	79	-772.04	1849.90	-1097.85	2501.53

Table 7: Summary of ML results together with the associated standard errors in parentheses for the best chosen model.

	Parameter				
Variable	$\boldsymbol{\mu}$	$\mathrm{col}_1(\boldsymbol{B})$	$\mathrm{col}_2(\boldsymbol{B})$	$\mathrm{col}_3(\boldsymbol{B})$	$\boldsymbol{d}$
y_1	0.0006 (0.0214)	0.3560 (0.0176)	0.0654 (0.0176)	0.2217 (0.0215)	0.7931 (0.0107)
y_2	0.0010 (0.0114)	-0.2155 (0.0151)	0.2217 (0.0150)	0.0523 (0.0240)	0.8748 (0.0073)
y_3	-0.0004 (0.0342)	0.0752 (0.0330)	-0.0813 (0.0277)	**0.9827** (0.0317)	0.0006 (0.0353)
y_4	-0.0004 (0.0320)	0.1678 (0.0279)	-0.0916 (0.0349)	**0.7347** (0.0270)	0.4024 (0.0332)
y_5	-0.0025 (0.0287)	0.0947 (0.0297)	-0.4997 (0.0208)	0.1630 (0.0219)	0.6502 (0.0242)
y_6	0.0004 (0.0209)	0.3748 (0.0226)	0.0038 (0.0266)	0.4243 (0.0158)	0.4624 (0.0379)
y_7	-0.0042 (0.0290)	0.3257 (0.0001)	**−0.8505** (0.0001)	0.1827 (0.0001)	0.0001 (0.0282)
y_8	0.0037 (0.0326)	0.2191 (0.0180)	**0.6841** (0.0280)	0.0175 (0.0361)	0.3514 (0.0429)
y_9	0.0001 (0.0313)	**0.9710** (0.0299)	-0.1014 (0.0284)	0.1098 (0.0387)	0.0107 (0.0346)
y_{10}	-0.0003 (0.0271)	**0.6709** (0.0154)	-0.1556 (0.0261)	0.0602 (0.0242)	0.5018 (0.0312)
y_{11}	0.0007 (0.0150)	-0.4338 (0.0135)	0.2008 (0.0202)	0.2294 (0.0132)	0.6958 (0.0125)
y_{12}	0.0021 (0.0228)	-0.1231 (0.0289)	0.4294 (0.0218)	0.2683 (0.0202)	0.6717 (0.0292)
y_{13}	0.0016 (0.0227)	0.2285 (0.0149)	0.2814 (0.0228)	-0.1418 (0.0265)	0.8060 (0.0120)
		$\boldsymbol{\lambda}$			
	-1.4831 (0.2520)	-6.1825 (1.1312)	-0.6494 (0.2961)		

second factors can respectively be viewed as the *natural phenols factor* and *color assessment indices*. Also, y_3 and y_4 have heavy loadings on the third factor, which might be called a *mineral factor*. Thus, one can conclude that the variables, y_3, y_4, and $y_7 - y_{10}$, explain most of the variability in the Barbera data.

5.2. *Italian olive oil data*

The second dataset is related to the eight fatty acids found by lipid fraction in 572 Italian olive oils (Forina and Tiscornia, 1982) that came from the three regions of Italy-Southern, Sardinia, and Italy-Northern. These regions can be further subdivided into nine different areas. The Italian olive oil dataset, which is available in the "**pgmm**" package of R, was recently analyzed by Tortora et al. (2015), who proposed the mixture of generalized hyperbolic factor model. Here, the focus is solely on $n = 98$ observations from the Sardinia region. Table 8 shows a summary of the 8 measures along with their normality KS and skew-normality r_n^* tests. From the p-values of the tests and the values of skewness and kurtosis, it can be significantly concluded that not all variables follow the normal distribution, but there is enough evidence in favour of the SN distribution based on the r_n^* test for all attributes. Furthermore, the p-values of the tests GSW = 5.666e-13 and CSN = 0.509 for the multivariate normality and skew-normality assure us that skewed distributions can describe this data better that the normal model.

Displayed in Figure 3 in the Online Supplement, the histogram and corresponding normal Q-Q plots of the four FA factor score estimates obtained by the "regression" method for the classical FA model with $q = 4$ highlight a serious departure of factor scores from the normality assumption. One can also observe from Figure 4 how well the bivariate MMN-based models, as the rMSN, MMNE and MMNEH distributions, can capture the scattering patterns of the four FA factor score estimates.

Table 8: An overview of 8 attributes of 98 of the Sardinia Italian olive oil data with the p-values of the KS and r_n^* tests.

Variable	Description	Skewness	Kurtosis	KS	r_n^*
y_1	Palmitic	0.146	-0.518	0	0.253
y_2	Palmitoleic	-0.367	3.427	0	0.440
y_3	Stearic	0.473	-0.603	0	0.361
y_4	Oleic	-0.772	-0.756	0	0.335
y_5	Linoleic	0.683	-1.015	0	0.409
y_6	Linolenic	0.550	0.190	0	0.207
y_7	Arachidic	0.162	-0.142	0	0.184
y_8	Eicosenoic	0.098	-1.169	0	0.279

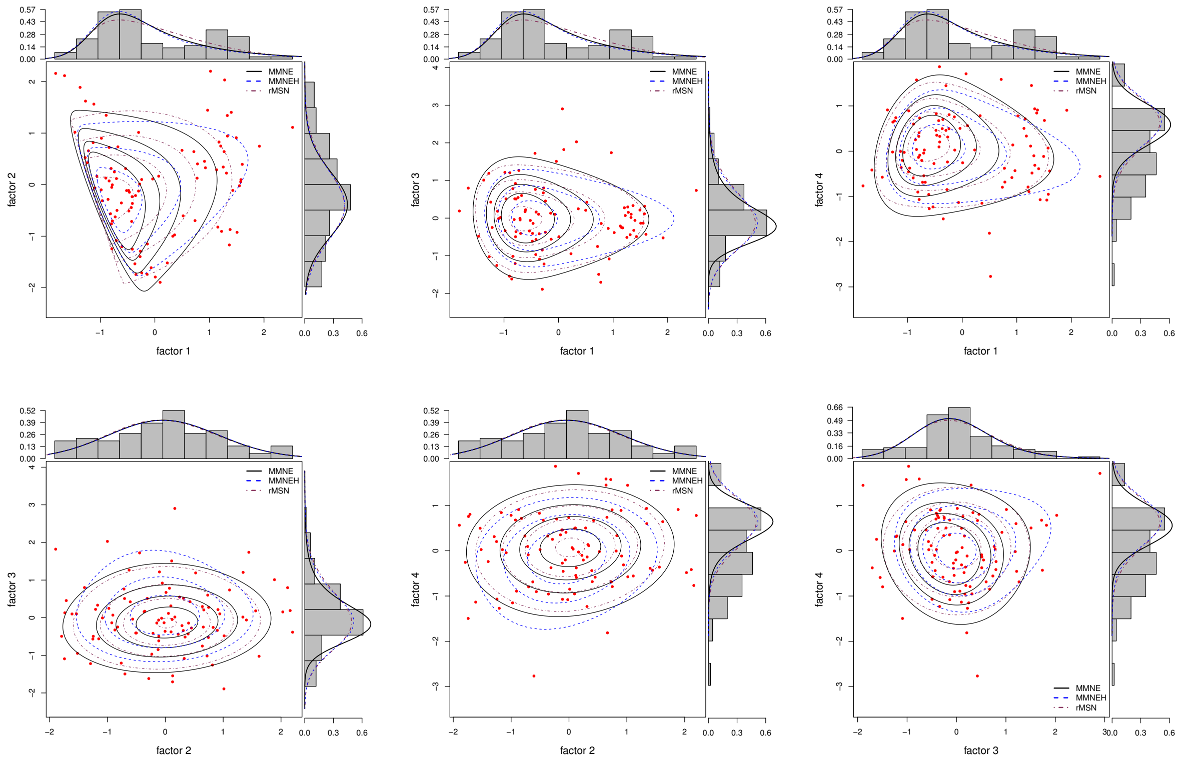

Figure 4: Scatter plots of factor scores superimposed on a set of contour lines estimated by the rMSN, MMNE and MMNEH distributions together with two summary histograms and curves of their marginal densities for the Sardinia Italian olive oil data.

Motivated by the described disadvantages of the FA model and the advantages of the skewed-type FA models to analyze the Sardinia olive oil data, the skew FA models are used for illustration purposes. We fit the FA, tFA, rSNFA, rSTFA, GHSTFA, GHCSTFA, MMNEFA and MMNEHFA models with q ranging from 2 to 4 to the standardized and original data. Notice that the choice of a maximum $q = 4$ satisfies the restriction $(p - q)^2 \geq (p + q)$. The results of the ML fitting, including the maximized log-likelihood values, the number of parameters together with the BIC value are reported in Table 9. It can be observed that the MMNEFA model outperforms other competitors based on the BIC criteria for both standardized and non-standardized data. From Table 10, which summarizes the ML estimate of parameters, along with their standard error (in parentheses), it can readily be seen that the estimated skewness parameters are significantly high, indicating that the joint distribution of the latent factors is skewed.

From the Varimax rotated solution of the factor loadings highlighted in Table 10, the positive and negative loadings of variables on the four factors are observed. It is concluded that the first factor has a very high absolute value loading on y_6. Because this attribute is related to the omega-3 fatty acid, it could be labeled as the *vascular system care factor*. It is clear that the second factor also loads highly on y_1 alone, which motivate us to label it as the *controversial factor*

Table 9: Comparison of ML estimation results for the Sardinia olive oil data.

			Original data		Standardized data					Original data		Standardized data	
Model	q	m	$\ell_{\max}$	BIC	$\ell_{\max}$	BIC	Model	q	m	$\ell_{\max}$	BIC	$\ell_{\max}$	BIC
FA	2	31	-3111.05	6364.24	-887.46	1917.05	tFA	2	32	-3101.77	6350.26	-878.18	1903.08
	3	37	-3098.12	6365.89	-874.53	1918.71		3	38	-3081.37	6336.96	-857.77	1889.78
	4	42	-3084.14	6360.86	-860.55	1913.67		4	43	-3066.24	6329.63	-842.65	1882.45
rSNFA	2	33	-3105.27	6361.83	-881.67	1914.65	rSTFA	2	34	-3090.237	6336.363	-866.64	1889.18
	3	40	-3082.41	6348.21	-858.81	1901.02		3	41	-3070.638	6329.259	-847.04	1882.07
	4	46	-3058.63	6328.16	-835.03	1880.97		4	47	-3051.80	6319.10	-828.21	1871.95
MMNEFA	2	33	-3086.657	6324.619	-863.06	1877.43	MMNEHFA	2	34	-3095.39	6346.67	-871.80	1899.48
	3	40	-3069.008	6321.416	-845.41	1874.23		3	41	-3072.60	6333.19	-849.01	1886.00
	4	46	-3053.01	**6316.92**	-829.41	**1869.73**		4	47	-3053.29	6322.07	-829.69	1874.88
GHSTFA	2	40	-3092.47	6368.35	-868.88	1921.16	GHCSTFA	2	28	-4162.47	8453.32	-912.26	1952.90
	3	46	-3076.03	6362.96	-852.43	1915.77		3	36	-4022.80	8210.66	-901.10	1967.26
	4	51	-3064.16	6362.15	-840.56	1914.96		4	43	-3611.82	7420.80	-890.04	1977.23

Table 10: ML solutions together with the associated Varimax rotated loading and their standard errors in parentheses for the best chosen model.

	Parameter					
Variable	$\boldsymbol{\mu}$	col$_1(\boldsymbol{B})$	col$_2(\boldsymbol{B})$	col$_3(\boldsymbol{B})$	col$_4(\boldsymbol{B})$	$\boldsymbol{d}$
y_1	-0.0034 (0.0365)	0.0037(0.0353)	**0.9492**(0.0333)	0.3145(0.345)	0.0019(0.0313)	0.0038 (0.0262)
y_2	-0.0014 (0.0329)	-0.1160(0.0180)	-0.1124(0.0341)	0.2729(0.0393)	0.4034(0.0117)	0.7289 (0.0320)
y_3	-0.0041 (0.0416)	0.3198(0.0300)	0.0942(0.0428)	**0.6497**(0.0437)	**0.4802**(0.0319)	0.2465 (0.0437)
y_4	0.0058 (0.0315)	-0.1654(0.0222)	-0.3695(0.0242)	**−0.9089**(0.0309)	-0.1575(0.0346)	0.0172 (0.0119)
y_5	-0.0058 (0.0362)	0.1896(0.0008)	0.1356(0.0008)	**0.9883**(0.0008)	-0.0218(0.0007)	0.0001 (0.0186)
y_6	0.0024 (0.0460)	**−0.9224**(0.0262)	-0.0120(0.0376)	-0.3429(0.0339)	-0.0370(0.0289)	0.0265 (0.0208)
y_7	-0.0001 (0.0341)	-0.4198(0.0214)	0.0077(0.0254)	0.0500(0.0280)	0.0005(0.0170)	0.8110 (0.0086)
y_8	0.0001 (0.0311)	0.0369(0.0222)	0.0325(0.0377)	-0.0275(0.0266)	0.0745(0.0083)	0.9811 (0.0069)
	$\boldsymbol{\lambda}$					
	6.4783 (1.3408)	5.1820 (0.9843)	4.9586(0.8864)	6.1359(1.5273)		

since contradicting evidence has been found by studies determining whether the palmitic acid contributes to coldihal vascular disease and cancer. The estimated factor loadings in Table 10 also reveal that the third factor, which might be called the *nutrition factor*, loads highly on y_5 followed by y_2 and with a very high absolute loading on y_4. Moreover, y_3 has moderately high loading on the fourth factor. Observing the estimate of $\boldsymbol{d}$, the small uniqueness of these variables is evident. The remaining measurements have negligible loadings on the four factors since their estimated loadings are fairly small. Thus, one could conclude that the variables y_1, $y_3 - y_5$, and y_6 explain most of variability in the Sardinia olive oil data.

Figure 5 shows the scatter plots overlaid with the marginal contours, obtained by the marginalization of the fitted MMNEFA and MMNEHFA models, for four selected variables. The visualization of the contours shows that the fitted MMNEFA can satisfactorily adapt the shape of the scattering pattern of the data. To summarize, the implementation of MMNEFA can give more accurate results for analyzing the Sardinia olive oil data.

6. Conclusions

This paper has dealt with the extension of the FA model, based on the multivariate mean-mixture of the normal distribution as an alternative model for analyzing strongly skewed and leptokurtic datasets. Presenting a hierarchical stochastic representation, parameter estimation was determined with an ECM algorithm. Two real data analyses and three simulation studies illustrate the favorable performance of the presented methodology. It is shown that the proposed model can be considered as an alternative to some existing factor analyzers, especially the rSTFA and GHSTFA models.

A further development will be to consider a finite mixture representation of the MMN models (Naderi et al., 2019). It would also be of interest to extend the current approach to the finite mixture of the MMNFA model (Liu and Lin, 2015; Tortora et al., 2015). Due to some computational difficulties in implementing the EM algorithm in modeling censored and/or missing value datasets based on the NMVM model, the methodology proposed in this paper can facilitate the development of new models for analyzing skewed data with censored and/or missing values (Liu and Lin, 2015; Lin et al., 2017; Wang et al., 2019).

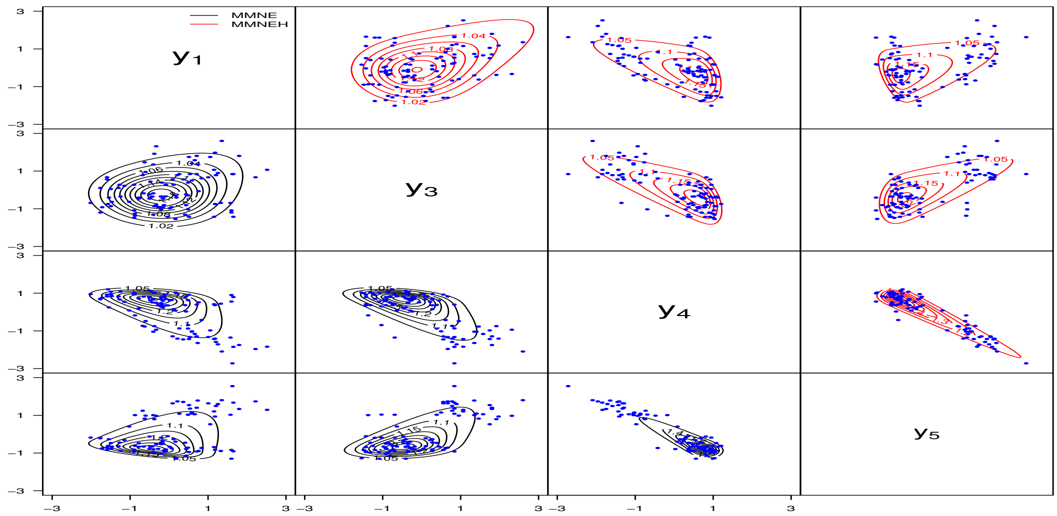

Figure 5: Scatter plots of pairs of four selected variables of the Sardinia Italian olive oil data and coordinate projected contours.

All computations were carried out using R 3.4.3 in a Win 64 environment with a 2.59 GHz/Intel Core(TM) i7 6500U CPU Processor and 8.0 GB RAM. R codes for implementation are available upon request.

Acknowledgments

The authors wish to thank the Editor, anonymous Associate Editor, and referees for their helpful comments, which helped to improve this article. M. Naderi, A. Bekker and F. Hashemi acknowledge the research support provided by the National Research Foundation, South Africa (Reference: CPRR160403161466 Grant Number: 105840, Reference: SRUG190308422768 grant No. 120839 and STATOMET).

References

Aitken, A. C., 1926. On Bernoulli's numerical solution of algebraic equations. Proceedings of the Royal Society of Edinburgh 46 (3), 289–305.

Alva, J. A. V., Estrada, E. G., 2009. A generalization of Shapiro-Wilk's test for multivariate normality. Communications in Statistics - Theory and Methods 38 (11), 1870–1883.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12 (2), 171–178.

Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew-normal distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61 (3), 579–602.

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t* distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65 (2), 367–389.

Balakrishnan, N., Capitanio, A., Scarpa, B., 2014. A test for multivariate skew-normality based on its canonical form. Journal of Multivariate Analysis 128, 19–32.

Basilevsky, A. T., 1994. Statistical factor analysis and related methods: theory and applications, 1st Edition. 1. John Wiley & Sons, New York.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 39 (1), 1–38.

Fokoué, E., Titterington, D., 2003. Mixtures of factor analysers. bayesian estimation and inference by stochastic simulation. Machine Learning 50 (1/2), 73–94.

Forina, M., Tiscornia, E., 1982. Pattern-recognition methods in the prediction of Italian olive oil origin by their fatty-acid content. Annali di Chimica 72 (3-4), 143–155.

Hashemi, F., Naderi, M., Jamalizadeh, A., Lin, T. I., 2020. A skew factor analysis model based on the normal mean-variance mixture of Birnbaum-Saunders distribution. Journal of Applied Statistics 47 (16), 3007–3029.

Ho, H. J., Pyne, S., Lin, T. I., 2011. Maximum likelihood inference for mixtures of skew student-*t*-normal distributions through practical EM-type algorithms. Statistics and Computing 22 (1), 287–299.

Johnson, R., Wichern, D., 2007. Applied multivariate statistical analysis. prenticehall international. INC., New Jersey.

Lawley, D. N., Maxwell, A. E., 1971. Factor analysis as a statistical method. Journal of the Royal Statistical Society: Series D (Statistician) 12 (3), 209–229.

Lee, S. X., McLachlan, G. J., 2013. On mixtures of skew normal and skew-t distributions. Advances in Data Analysis and Classification 7 (3), 241–266.

Lin, T. I., McLachlan, G. J., Lee, S. X., 2016. Extending mixtures of factor models using the restricted multivariate skew-normal distribution. Journal of Multivariate Analysis 143, 398–413.

Lin, T. I., Wang, W. L., McLachlan, G. J., Lee, S. X., 2017. Robust mixtures of factor analysis models using the restricted multivariate skew-t distribution. Statistical Modelling: An International Journal 18 (1), 50–72.

Lin, T. I., Wu, P. H., McLachlan, G. J., Lee, S. X., 2015. A robust factor analysis model using the restricted skew-t distribution. TEST 24 (3), 510–531.

Liu, C., Rubin, D. B., 1994. The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. Biometrika 81 (4), 633–648.

Liu, M., Lin, T. I., 2015. Skew-normal factor analysis models with incomplete data. Journal of Applied Statistics 42 (4), 789–805.

McLachlan, G. J., Bean, R., Jones, L. B. T., 2007. Extension of the mixture of factor analyzers model to incorporate the multivariate t distribution. Computational Statistics & Data Analysis 51 (11), 5327–5338.

McLachlan, G. J., Krishnan, T., 2008. The EM algorithm and extensions, 2nd Edition. John Wiley & Sons, Hoboken, New Jersey.

McLachlan, G. J., Peel, D., 2000. Finite mixture models, 1st Edition. John Wiley & Sons, New York.

McNeil, A. J., Frey, R., Embrechts, P., et al., 2005. Quantitative risk management: Concepts, techniques and tools. Vol. 3. Princeton university press Princeton, New Jersey.

Meng, X. L., Rubin, D. B., 1993. Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika 80 (2), 267–278.

Montanari, A., Viroli, C., 2010. A skew-normal factor model for the analysis of student satisfaction towards university courses. Journal of Applied Statistics 37 (3), 473–487.

Murray, P. M., Browne, R. P., McNicholas, P. D., 2014a. Mixtures of skew-t factor analyzers. Computational Statistics & Data Analysis 77, 326–335.

Murray, P. M., McNicholas, P. D., Browne, R. P., 2014b. A mixture of common skew-t factor analysers. Stat 3 (1), 68–82.

Naderi, M., Hung, W. L., Lin, T. I., 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.

Negarestani, H., Jamalizadeh, A., Shafiei, S., Balakrishnan, N., 2019. Mean mixtures of normal distributions: properties, inference and application. Metrika 82 (4), 501–528.

Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T. I., Maier, L. M., Baecher-Allan, C., McLachlan, G. J., Tamayo, P., Hafler, D. A., et al., 2009. Automated high-dimensional flow cytometric data analysis. Proceedings of the National Academy of Sciences 106 (21), 8519–8524.

Rodríguez, P. P., Alva, J. A. V., 2010. On testing the skew normal hypothesis. Journal of Statistical Planning and Inference 140 (11), 3148–3159.

Spearman, C., 1904. “general intelligence”, objectively determined and measured. The American Journal of Psychology 15 (2), 201–292.

Tortora, C., McNicholas, P. D., Browne, R. P., 2015. A mixture of generalized hyperbolic factor analyzers. Advances in Data Analysis and Classification 10 (4), 423–440.

Wang, W. L., Castro, L. M., Lachos, V. H., Lin, T. I., 2019. Model-based clustering of censored data via mixtures of factor analyzers. Computational Statistics & Data Analysis 140, 104–121.