Mathematics

AIMS Mathematics, 9(6): 15837–15856.
DOI: 10.3934/math.2024765
Received: 24 July 2023
Revised: 30 August 2023
Accepted: 03 September 2023
Published: 06 May 2024

http://www.aimspress.com/journal/Math

Research article

A Bayesian approach on asymmetric heavy tailed mixture of factor analyzer

Hamid Reza Safaeyan[1], Karim Zare[1,*], Mohamadreza Mahmoudi[2], Mohsen Maleki[3] and Amir Mosavi[4]

[1] Department of Statistics, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran
[2] Department of Statistics, Faculty of Science, Fasa University, Fasa, Iran
[3] Department of Statistics, Faculty of Mathematics and Statistics, University of Isfahan, Isfahan, 81746-73441, Iran
[4] Obudai University, Budapest, Hungary

* **Correspondence:** Email: karim.zare@iau.ac.ir; Tel: +98-917-935-4051.

Abstract: A Mixture of factor analyzer (MFA) model is a powerful tool to reduce the number of free parameters in high-dimensional data through the factor-analyzer technique based on the covariance matrices. This model also prepares an efficient methodology to determine latent groups in data. In this paper, we use an MFA model with a rich and flexible class of distributions called hidden truncation hyperbolic (HTH) distribution and a Bayesian structure with several computational benefits. The MFA based on the HTH family allows the factor scores and the error component can be skewed and heavy-tailed. Therefore, using the HTH family leads to the robustness of the MFA in modeling asymmetrical datasets with/without outliers. Furthermore, the HTH family, because of several desired properties, including analytical flexibility, provides steps in the estimation of parameters that are computationally tractable. In the present study, the advantages of MFA based on the HTH family have been discussed and the suitable efficiency of the introduced MFA model has been demonstrated by using real data examples and simulation.

Keywords: Bayesian; hidden truncation hyperbolic distributions; MCMC method; mixture of factor analyzer model; Skewed family; hyperbolic distributions; data science; soft computing
Mathematics Subject Classification: 62F15, 62H25

1. Introduction

In multilevel and high-dimensional data, analysis and identification of latent or hidden variables is a main statistical problem and using the finite mixture model (FM) with factor analyzer (FA) models as components is a useful statistical technique to model such datasets. The relationship between several variables can be described in an FA model by reducing the dimension of variables into some smaller latent factors. Such a procedure is applied in different fields like psychometric testing to facilitate the analysis of high-dimensional data. The FM model can detect the existence of subgroups among a general population using latent variables' models. These approaches are widely employed in numerous fields such as artificial intelligence, biology, data mining, epidemiological studies, medical sciences and social sciences. Ghahramani and Hinton [12] combined the FM model and FA model which is usually called the MFA model. In basic statistical models, researchers usually assume Gaussian distribution (see e.g., Marron and Wand [26] and Roeder and Wasserman [29]). However, Wall et al. [30] showed that in many real applications, the data may be mildly or extremely asymmetry which can lead to serious misleading inference from normality (note that such issue can occur even for small deviations). Thus, the imposition of symmetric components of the mixture models in the applications is a fairly restrictive condition (See Mahmoudi et al. [24] and Maleki et al. [25], for detailed examples and discussions).

In recent decades, different families of asymmetric distributions have been introduced and used, for example: Skew-Normal (SN) (Arellano-Valle and Genton [2]) and a family of two-pieces scale mixtures of normal (TP-SMN) (Bazrafkan et al. [5]). Lee and McLachlan [17] focused on a family named unrestricted Skew-t (SUT). "Restricted" is referred to as asymmetric behaviors that were controlled by multiplying an ordinary skewing variable on convolution type representation to a vector of skewness parameters. In "unrestricted" forms, in contrast to restricted forms where skewness is present in a single direction, the skewness is free to present in more than one direction. In recent years, multivariate skewed distributions are considered for FA and MFA approaches. Kim et al. [16] considered the FA models for which errors followed a symmetric (Gaussian) distribution and factors followed skewed distributions such as skew-normal, generalized skew-normal and skew-t. Lin et al. [22] developed the MFA model based on multivariate restricted skew-normal (RSN) for the factor component, which is called the finite mixtures of skew-normal factor analyzer (MSNFA). In this work, the HTH family of distributions is considered in the MFA model. The HTH family is symmetrical/asymmetrical with light/heavy tailed-ness and is an extension of the unrestricted skew normal distribution (SUN; given by Arellano-Valle and Genton [2]) by employing a generalized inverse Gaussian (GIG; Good [13]) for the scale mixer component.

The maximum likelihood (ML) estimates (classical inferences) are the basis of the most common estimation methods of MFA model parameters, while for some samples, the likelihood function is unbounded. To solve this issue, Bayesian inferences can be employed to estimate the parameters of the Gaussian MFA model. Ando [1] and Lee and Xia [19] extend the Gaussian MFA model by considering a t-distribution (matrix-variate) for the scores (factor) and independent Gaussian family for the errors, respectively. Yang and Dunson [31] concentrated on the semi-parametric MFA model; while Chen et al. [8] focused on non-parametric approaches in the MFA model (see also Luo et al. [23] for recently published work.). To overcome the maintained issues, in this work, a Bayesian technique is developed to estimate the parameters of the model. In comparison to ML estimations, using a Bayesian approach for MFA models has several other computational advantages, particularly in high dimensional settings for the specification of priors to regularize the parameter space. It is noted that in a Bayesian framework, information can be included in Bayesian inferences without computational demands or complexity. Additionally, the number of factors and components in the MFA model could vary, and as a result, they can be updated as the computational part. We also note that the effect of the missing data can be a huge problem and its reduction using a Bayesian setting on parameter estimates is one of the quite effective and natural ways (e.g., class-dependent missingness). In the Bayesian framework, it is considered the iterations of MCMC to obtain the posterior predictive distribution. Note that, the calculation of the standard error of ML estimations of complex mixture models requires a lot of computations and involves the appraisement of the derivatives for complex functions used for estimating the parameters.

Besides the above-mentioned advantages of Bayesian methodology, another superiority of the work is that the HTH family prepares a high level of flexibility and performance for usage in factor analysis models, i.e., some partitions of an HTH random vector to uncorrelated homogeneous HTH random vectors can provide simultaneous asymmetric error and factor. As Lee and McLachlan [18] mentioned, this structure for the MFA model, that is named skew factors and errors (SFE), is an open problem. It is worth mentioning that our proposed model has such structure. Furthermore, to show the performance of the Bayesian technique in this MFA situation, the missing data is considered and the applicability of the introduced MFA model is evaluated. In Sections 2 and 3, preliminaries and notions of the SUN, GIG and HTH distributions are examined. A Bayesian approach with the Gibbs sampling algorithm for the HTH-MFA model is presented in Section 4. In Section 5, the ability and suitability of the introduced MFA model are evaluated using some simulated and real datasets. Finally, some conclusions along with a discussion on possible extensions for further research have been presented in Section 6.

2. A review on unrestricted skew normal and GIG distributions

2.1. Preliminaries

Let $\boldsymbol{Z}$ be a $p \times 1$ random vector with SUN distribution with a $p \times 1$ location's vector $\boldsymbol{\theta}$, and a $p \times p$ scale matrix Ξ (which is positive definite), and a $p \times q$ skewness parameters matrix $\boldsymbol{\Pi}$, defined by $\boldsymbol{Z} \sim \mathrm{SUN}_{p,q}(\boldsymbol{\theta}, \Xi, \boldsymbol{\Pi})$. Then the probability density function (pdf) of $\boldsymbol{Z}$ is presented by:

$$\mathrm{f}(\boldsymbol{z}|\boldsymbol{\theta}, \Xi, \boldsymbol{\Pi}) = 2^q \phi_p(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{\Lambda}) \Phi_q(\boldsymbol{\Pi}^\top \boldsymbol{\Lambda}^{-1}(\boldsymbol{z} - \boldsymbol{\theta})|\boldsymbol{\Gamma}), \mathbf{z} \in \mathbb{R}^p, \tag{1}$$

such that $\Phi_q(\cdot\ |\boldsymbol{\Gamma})$ and $\phi_p(\cdot\ |\boldsymbol{\theta}, \boldsymbol{\Lambda})$ are the cumulative density function (cdf) and pdf of $N_q(\mathbf{0}, \boldsymbol{\Gamma})$ and $N_p(\boldsymbol{\theta}, \boldsymbol{\Lambda})$ distributions, respectively, $\boldsymbol{\Lambda} = \Xi + \boldsymbol{\Pi}\boldsymbol{\Pi}^\top$, and $\boldsymbol{\Gamma} = \mathbf{I}_q - \boldsymbol{\Pi}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\Pi} = \left(\mathbf{I}_q + \boldsymbol{\Pi}^\top \Xi^{-1} \boldsymbol{\Pi}\right)^{-1}$ (See Arellano-Valle and Genton [2] for more details). The multivariate SUN distribution defined in (1) recover the multivariate normal distribution with zero skewness, when $\boldsymbol{\Pi} = \mathbf{0}$, the multivariate RSN, when $q = 1$.

The stochastic representation of $\boldsymbol{Z} \sim \mathrm{SUN}_{p,q}(\boldsymbol{\theta}, \Xi, \boldsymbol{\Pi})$ is given by:

$$\boldsymbol{Z} = \boldsymbol{\theta} + \boldsymbol{\Pi}|\boldsymbol{X}_0| + \Xi^{1/2} \boldsymbol{X}_1, \tag{2}$$

such that $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ are independent and follow $N_q(\mathbf{0}, \mathbf{I}_q)$ and $N_p(\mathbf{0}, \mathbf{I}_p)$, respectively. Using Eq (2), we concluded that $\mathrm{E}[\boldsymbol{Z}] = \boldsymbol{\theta} + \sqrt{2/\pi}\, \boldsymbol{\Pi} \mathbf{1}_p$ and $\mathrm{Var}[\boldsymbol{Z}] = \boldsymbol{\Lambda} - \frac{2}{\pi} \boldsymbol{\Pi} \mathbf{1}_q \mathbf{1}_q^\top \boldsymbol{\Pi}^\top$, such that $\mathbf{1}_q$ is a vector of length q with all elements equal to 1. (Note that the Eqs (1) and (2) with more details are described in Arellano-Valle and Genton [2].)

The HTH random variable

$$\mathbf{Y} = \boldsymbol{\theta} + \kappa(U)^{1/2}\boldsymbol{Z}, \tag{3}$$

where $\kappa(\cdot)$ is a positive function of scale mixer variable U is called scale mixtures of SUN family. Also, note that the U follows GIG distribution and independent of $\boldsymbol{Z}$.

2.2. Some details of the GIG distribution

The GIG of distribution has a positive support and so is a natural candidate for the variable U (that is the scale mixer variable) in the stochastic representation (3) (see Good [13] and Barndorff-Nielsen and Halgreen [4]). Such choices lead to the construction of a highly workable multivariate class of unified distributions suitable for multivariate statistical analysis. In terms of its parameterization, there exist several (equivalent) representations of the GIG distributions, that in our methodology (and Bayesian framework), there are closed-form posterior distributions and some simplifications to adopt its representation. The pdf of the GIG random variable U (that is denoted by $U \sim GIG_*(\upsilon, \psi, \eta)$) is as follows:

$$\mathcal{GIG}_*(u|\upsilon, \psi, \eta) = \frac{(u/\eta)^{\upsilon-1}}{2\eta K_\upsilon(\psi)} \exp\left(-\frac{\psi}{2}\left(\frac{u}{\eta} + \frac{\eta}{u}\right)\right); \ u \in \mathbb{R}^+, \psi > 0, \eta > 0, -\infty < \upsilon < +\infty. \tag{4}$$

such that $K_t(y)$ is the third kind of order t modified "Bessel" function which is evaluated at point y. It can be computed that

$$\mathrm{r_s} = \mathrm{E}\left(\mathrm{U}^{s/2}\right) = \frac{\mathrm{K}_{\upsilon+s/2}(\psi)}{\mathrm{K}_\upsilon(\psi)}\eta^{s/2}, \ \mathrm{s} = 1,2, \ldots \tag{5}$$

In the Bayesian framework (posteriors), we can consider another form of $U \sim GIG^*(\upsilon, \gamma, \rho)$ with the following pdf:

$$\mathcal{GIG}^*(u|\upsilon, \gamma, \rho) = \left(\frac{\gamma}{\rho}\right)^\upsilon \frac{u^{\upsilon-1}}{2K_\upsilon(\rho\gamma)} \exp\left(-\frac{1}{2}\left(\frac{\rho^2}{u} + \gamma^2 u\right)\right); \ u \in \mathbb{R}^+, \gamma > 0, \rho > 0, -\infty < \upsilon < +\infty. \tag{6}$$

Finally, note that the HTH distributions have been constructed by using the GIG distributions for scale mixer variable U and multivariate SUN random variable in the representation (3).

3. A review on HTH distributions

As mentioned in the previous sections, a random vector $\boldsymbol{Y}$ has a HTH distribution if it has the following stochastic representation

$$\boldsymbol{Y} = \boldsymbol{\theta} + \boldsymbol{\Pi}\boldsymbol{X} + \kappa(U)^{1/2}\boldsymbol{\Xi}^{1/2}\boldsymbol{X}_1, \tag{7}$$

such that $\boldsymbol{\theta}$, $\boldsymbol{\Xi}$ and $\boldsymbol{\Pi}$ are location's vector, dispersion's matrix and skewness (or shape) matrix, respectively, $\boldsymbol{X} = \kappa^{1/2}(U)|\boldsymbol{X}_0|$, $\boldsymbol{X}_0$, $\boldsymbol{X}_1$ and U are independent distributed as $N_q(\boldsymbol{0}, \mathbf{I}_q)$, $N_p(\boldsymbol{0}, \mathbf{I}_p)$ and $GIG_*(\upsilon, \psi, \eta)$, respectively. It is clear that the conditional distribution of $\mathbf{Y}$ given $U = u$ is as $\boldsymbol{Y}|U = u \sim SUN_{p,q}(\boldsymbol{\theta}, \kappa(u)\boldsymbol{\Xi}, \kappa(u)^{1/2}\,\boldsymbol{\Pi})$, and therefore the pdf of $\boldsymbol{Y}$ is as follows:

$$\mathrm{g}(\boldsymbol{y}|\boldsymbol{\theta},\Xi,\boldsymbol{\Pi},\boldsymbol{\mu}) = 2^q \int_0^\infty \phi_p(\boldsymbol{y}|\boldsymbol{\theta},\kappa(u)\boldsymbol{\Lambda})\Phi_q\left(\kappa(u)^{-1/2}\boldsymbol{\Pi}^\top\boldsymbol{\Lambda}^{-1}(\boldsymbol{y}-\boldsymbol{\theta})\middle|\boldsymbol{\Gamma}\right)\mathcal{GIG}_*(u|\boldsymbol{\mu})du, \boldsymbol{y} \in \mathbb{R}^p, \quad (8)$$

such that $\boldsymbol{\mu} = (\upsilon,\psi,\eta)^\top$. In the next, HTH random vector $\boldsymbol{Y}$ is represented by $\boldsymbol{Y}\sim\mathrm{HTH}(\boldsymbol{\theta},\Xi,\boldsymbol{\Pi},\boldsymbol{\mu})$. It should be noted that the pdf given in Eq (8) is a reparameterization of the HTH pdf in Murray et al. [27].

In order to have identifiability of the proposed HTH distributions there exist some concerns relating to the skewness matrix $\boldsymbol{\Pi}$ and GIG parameters $\boldsymbol{\mu}$. Note in the pdf given in Eq (8), for any positive value c, $(\boldsymbol{\theta}, c\Xi, c\boldsymbol{\Pi}, \upsilon, \psi/c, c\eta)$ and $(\boldsymbol{\theta},\Xi,\boldsymbol{\Pi},\upsilon,\psi,\eta)$ have similar densities, so to solve this identifiability issue let $\eta = 1$ and consequently $\boldsymbol{\mu} = (\upsilon,\psi)^\top$. Furthermore, sorting the skewness matrix $\boldsymbol{\Pi}$ using a norm to its columns (or employing a methodology similar to Bai and Li [3] to identify a factor loadings matrix) is necessary to verify the identifiability of HTH distributions and the same is true for our proposed *MFA* model. Varying the distributions of U to the $GIG_*(\boldsymbol{\mu})$ family conducts to have different members of the HTH class. So, by considering $\kappa(u) = u$ and different distributions of U from $GIG_*(\boldsymbol{\mu})$ family, we have:

$$\mathrm{g}(\boldsymbol{y}|\boldsymbol{\theta},\Xi,\boldsymbol{\Pi},\boldsymbol{\phi}) = 2^q\mathcal{GH}_p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\Lambda},\boldsymbol{0},\boldsymbol{\nu}')GH_q(\boldsymbol{B}|\boldsymbol{0},\boldsymbol{\Gamma},\boldsymbol{0},\boldsymbol{\nu}''),\ \boldsymbol{y} \in \mathbb{R}^p, \quad (9)$$

such that GH_q and $\mathcal{GH}_p$ are respectively referred to the cdf of the q-variate generalized hyperbolic and pdf of the p-variate generalized hyperbolic (GH; Barndorff-Nielson and Halgreen [4]) distributions and $\boldsymbol{\Lambda} = \Xi + \boldsymbol{\Pi}\boldsymbol{\Pi}^\top$, $\boldsymbol{\Gamma} = \boldsymbol{I}_q - \boldsymbol{\Pi}^\top\boldsymbol{\Lambda}^{-1}\boldsymbol{\Pi}$, $\boldsymbol{B} = \boldsymbol{\Pi}^\top\boldsymbol{\Lambda}^{-1}(\boldsymbol{y}-\boldsymbol{\theta})$, $\boldsymbol{\nu}' = \left(\upsilon, \sqrt{\psi/\eta}, \sqrt{\psi\eta}\right)^\top$, $\boldsymbol{\nu}'' = \left(\upsilon - p/2, \sqrt{\psi/\eta}, q'(\boldsymbol{y})\right)^\top$ and $q'(\boldsymbol{y})^2 = (\boldsymbol{y}-\boldsymbol{\theta})^\top\boldsymbol{\Lambda}^{-1}(\boldsymbol{y}-\boldsymbol{\theta}) + \psi\eta$. Note when $q = 1$, the HTH distributions are reduced to restricted cases, called the skew-normal Generalized-Hyperbolic (SNGH; Murray et al. [27]) distributions, which involve the known distributions such as the skew-normal (SN), skew-Laplace (SLP), skew Pearson type VII (SP-VII), skew-slash (SSL) distributions, skew-t (ST) and skew-contaminated normal (SCN). In the case of unrestricted case ($q > 1$), the HTH distributions are an extension of the maintained distributions and also in the case of symmetric $\boldsymbol{\Pi} = \boldsymbol{0}$, it becomes the symmetrical GH distribution. For both *FA* and *MFA* approaches according to the multivariate scale mixtures of skew-normal (SMSN; Branco and Dey [6]) and restricted SN distributions, in order to have the identifiability of the model, the factor scores or the error term of the model must be symmetrically distributed, which is a main restriction (see e.g., Lin et al. [21] and Kim et al. [16]). However, note that the HTH family has an important advantage, particularly its usage in *MFA* models in which both of the factor and error terms have been distributed in the class of HTH with zero means.

4. Mixture of factor analysis model based on the HTH

4.1. HTH factor analyzer (HTH-FA) model

In the following, a FA model using the HTH distributions is developed. Specifically, the HTH-FA given by

$$\begin{cases} \boldsymbol{Y}_j = \boldsymbol{\theta} + \boldsymbol{L}\boldsymbol{F}_j + \boldsymbol{\epsilon}_j; \quad \boldsymbol{F}_j \perp \boldsymbol{\epsilon}_j, \\ \boldsymbol{F}_j \overset{iid}{\sim} HTH\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}_{\mathrm{f}}, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Pi}_{\mathrm{f}}, \boldsymbol{\mu}\right), \\ \boldsymbol{\epsilon}_j \overset{iid}{\sim} HTH(\boldsymbol{\theta}_{\mathrm{e}}, \boldsymbol{D}, \boldsymbol{\Pi}_{\mathrm{e}}, \boldsymbol{\mu}), \end{cases} ; \qquad j = 1, \dots, n, \tag{10}$$

is first considered, where $\boldsymbol{Y}_j$ and $\boldsymbol{\theta}$ are respectively the observations and location's vector (p-dimensional), the $p \times m$ $\boldsymbol{L}$ matrix is factor loadings, $\boldsymbol{F}_j$ is latent variable (that is a m-dimensional vector of factors ($m < p$) and can be asymmetric or symmetric, and heavy or light-tailed) and $\boldsymbol{\epsilon}_j$ is error's vector (that can be asymmetric or symmetric, and heavy or light-tailed), $\boldsymbol{\theta}_{\mathrm{e}} = \tau\boldsymbol{\Pi}_{\mathrm{e}}\boldsymbol{1}_p$, $\boldsymbol{\theta}_{\mathrm{f}} = \tau\boldsymbol{\Pi}_{\mathrm{f}}\boldsymbol{1}_q$, $\tau = -\sqrt{2/\pi}\, r_1$, $\boldsymbol{\Sigma} = r_2\boldsymbol{I}_m + \boldsymbol{\Pi}_{\mathrm{f}}\boldsymbol{C}_q\boldsymbol{\Pi}_{\mathrm{f}}^{\top}$, $\boldsymbol{C}_q = (r_2 - r_1^2)\frac{2}{\pi}\boldsymbol{1}_q\boldsymbol{1}_q^{\top} + \left(1 - \frac{2}{\pi}\right)r_2\boldsymbol{I}_q$ with positive definite dispersion matrices $\boldsymbol{D} = \mathrm{diag}(D_1, \dots, D_p)$ and $\boldsymbol{\Sigma}^{-1}$ with dimensions $p \times p$ and $m \times m$, respectively, $\boldsymbol{\Pi}_{\mathrm{e}} = \mathrm{diag}(\boldsymbol{\varpi}_{\mathrm{e}})$ is the $p \times p$ diagonal skewness matrix, and $\boldsymbol{\Pi}_{\mathrm{f}}$ is the $m \times q$ skewness matrix. Also note that

$$E[\boldsymbol{F}_j] = E[\boldsymbol{\epsilon}_j] = \boldsymbol{0},$$

$$\mathrm{Cov}[\boldsymbol{F}_j] = \boldsymbol{I}_m,$$

$$\mathrm{Cov}[\boldsymbol{\epsilon}_j] = \boldsymbol{\Omega} = r_2\boldsymbol{D} + \boldsymbol{\Pi}_{\mathrm{e}}\boldsymbol{C}_p\boldsymbol{\Pi}_{\mathrm{e}}^{\top}$$

$$E[\boldsymbol{Y}_j] = \boldsymbol{\theta},$$

and

$$\mathrm{Cov}[\boldsymbol{Y}_j] = \boldsymbol{L}\boldsymbol{L}^{\top} + \boldsymbol{D}.$$

Due to the HTH properties, we have

$$\begin{pmatrix} \boldsymbol{F}_j \\ \boldsymbol{\epsilon}_j \end{pmatrix} \overset{iid}{\sim} HTH\left(\begin{pmatrix} \boldsymbol{\theta}_{\mathrm{f}} \\ \boldsymbol{\theta}_{\mathrm{e}} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & \boldsymbol{0}_{m\times p} \\ \boldsymbol{0}_{p\times m} & \boldsymbol{D} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Pi}_{\mathrm{f}} & \boldsymbol{0}_{m\times p} \\ \boldsymbol{0}_{p\times q} & \boldsymbol{\Pi}_{\mathrm{e}} \end{pmatrix}, \boldsymbol{\mu}\right), \quad j = 1, \dots, n. \tag{11}$$

We call this FA model as HTH-FA. As a special case, the HTH-FA involves the skew-normal factor analyzer (SNFA) model that considered by Lin et al. [32] when $q = 1$ and the latent factor U is a degenerated random variable at 1. According to the HTH properties, we have that

$$\boldsymbol{Y}_j \overset{ind.}{\sim} HTH(\boldsymbol{\theta} + \tau\boldsymbol{\alpha}\boldsymbol{1}_k, \Xi, \boldsymbol{\alpha}, \boldsymbol{\mu}), \;\; j = 1, \dots, n, \tag{12}$$

such that $k = p + q$, $\boldsymbol{\alpha} = \left(\tilde{\boldsymbol{L}}\boldsymbol{\Pi}_{\mathrm{f}} \quad \boldsymbol{\Pi}_{\mathrm{e}}\right)_{p\times k}$ and $\boldsymbol{\Xi} = \tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}^{\top} + \boldsymbol{D}$ where $\tilde{\boldsymbol{L}} = \boldsymbol{L}\boldsymbol{\Sigma}^{-1/2}$. Consequently, the parameters of the model can be summarized by $\boldsymbol{\Omega} = (\boldsymbol{\theta}, \boldsymbol{L}, \boldsymbol{D}, \boldsymbol{\Pi}_{\mathrm{f}}, \boldsymbol{\Pi}_{\mathrm{e}}, \boldsymbol{\mu})$ which we need to estimate them.

According to Eq (12), the pdf of response is consequently presented by

$$\mathscr{f}(\boldsymbol{y}_j|\boldsymbol{\Omega}) = 2^k \int_0^{\infty} \phi_p\left(\boldsymbol{y}_j \middle| \boldsymbol{\theta} + \tau\boldsymbol{\alpha}\boldsymbol{1}_k, \kappa(u)\boldsymbol{\Lambda}\right)\Phi_k\left(\kappa(u)^{-1/2}\boldsymbol{\alpha}^{\top}\boldsymbol{\Lambda}^{-1}(\boldsymbol{s}_j)\middle|\boldsymbol{\Gamma}\right)\mathcal{GIG}_*(u|\boldsymbol{\mu})du, \tag{13}$$

such that $\boldsymbol{\Lambda} = \Xi + \boldsymbol{\alpha}\boldsymbol{\alpha}^\top$, $\boldsymbol{\Gamma} = \boldsymbol{I}_k - \boldsymbol{\alpha}^\top\boldsymbol{\Lambda}^{-1}\boldsymbol{\alpha}$ and $\boldsymbol{s}_j = \boldsymbol{y}_j - \boldsymbol{\theta} - \tau\boldsymbol{\alpha}\boldsymbol{1}_k$.

If $\kappa(u) = u$, then we have

$$\mathscr{f}(\boldsymbol{y}_j|\boldsymbol{\Omega}) = 2^k\mathcal{GH}_p(\boldsymbol{y}_j|\boldsymbol{\theta} + \tau\boldsymbol{\alpha}\boldsymbol{1}_k, \boldsymbol{\Lambda}, \boldsymbol{0}, \boldsymbol{\nu}_1)GH_k(\boldsymbol{\alpha}^\top\boldsymbol{\Lambda}^{-1}\boldsymbol{s}_j|\boldsymbol{0}, \boldsymbol{\Gamma}, \boldsymbol{0}, \boldsymbol{\nu}_2), \quad (14)$$

where $\boldsymbol{\nu}_1 = \left(\upsilon, \sqrt{\psi}, \sqrt{\psi}\right)^\top$, $\boldsymbol{\nu}_2 = \left(\upsilon - p/2, \sqrt{\psi}, q'(\boldsymbol{y}_j)\right)^\top$, $q'(\boldsymbol{y}_j)^2 = \boldsymbol{s}_j^\top\boldsymbol{\Lambda}^{-1}\boldsymbol{s}_j + \psi$ where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are defined in Eq (13).

It is worth noting that according to statistical features of the HTH family, and the structure of the HTH-FA model given in Eq (10), the response vector $\boldsymbol{Y}_j$ given the latent factor $\boldsymbol{F}_j$ has the conditional distribution given by

$$\boldsymbol{Y}_j|\boldsymbol{F}_j \overset{ind.}{\sim} HTH(\boldsymbol{\theta} + \boldsymbol{L}\boldsymbol{F}_j, \boldsymbol{D}, \boldsymbol{\Pi}_{\mathrm{e}}, \boldsymbol{\mu}), \ \ j = 1, \dots, n. \quad (15)$$

To guarantee the identifiability of the *FA* model, we consider the technique in Lin et al [22], and assume that the diagonal entries of the loading matrix $\boldsymbol{L}$ are strictly positive and its upper-right triangle is equal to 0 (constraining the loading matrix $\boldsymbol{L}$). Bai and Li [3] have considered the other approaches.

4.2. Mixture of HTH factor analyzer (HTH-MFA) model

A generalization of the proposed HTH-FA, is a mixture of the HTH-FA (called HTH-MFA). For $j = 1, \dots, n$, assume $\boldsymbol{Y}_j = (Y_{j1}, \dots, Y_{jp})$ is the response raising from a heterogeneous population that is partitioned into g groups. Define the membership variables or latent indicators $Z_1, \dots, Z_n$such that the term $(Z_j = i)$ means that the j-th vector variable belongs to the i-th component of the HTH-FA model, for $i = 1, \dots, g$. For $j = 1, \dots, n$, and $i = 1, \dots, g$, assume the probability mass function (pmf) of $Z_1, \dots, Z_n$ as $P(Z_j = i) = \pi_i$; where $\pi_i > 0$ and $\sum_{i=1}^{g}\pi_i = 1$. In terms of Z_j, for $i = 1, \dots, g$, we can conclude that each component of the HTH-MFA model follows the HTH-FA model give in Eq (10) by:

$$\begin{cases} \boldsymbol{Y}_j = \boldsymbol{\theta}_i + \boldsymbol{L}_i\boldsymbol{F}_{ij} + \boldsymbol{\epsilon}_{ij}; \ \ \boldsymbol{F}_{ij} \perp \boldsymbol{\epsilon}_{ij}, \ \text{with probability } p_i, \\ \boldsymbol{F}_{ij} \overset{iid}{\sim} HTH(\boldsymbol{\Sigma}_i^{-1/2}\boldsymbol{\theta}_{\mathrm{f}i}, \boldsymbol{\Sigma}_i^{-1}, \boldsymbol{\Sigma}_i^{-1/2}\boldsymbol{\Pi}_{\mathrm{f}i}, \boldsymbol{\mu}_i), \\ \boldsymbol{\epsilon}_{ij} \overset{iid}{\sim} HTH(\boldsymbol{\theta}_{\mathrm{e}i}, \boldsymbol{D}_i, \boldsymbol{\Pi}_{\mathrm{e}i}, \boldsymbol{\mu}_i), \end{cases} \quad ; \ j = 1, \dots, n, i = 1, \dots, g, \quad (16)$$

such taht $\boldsymbol{\Sigma}_i = r_2\boldsymbol{I}_m + \boldsymbol{\Pi}_{\mathrm{f}i}\boldsymbol{C}_q\boldsymbol{\Pi}_{\mathrm{f}i}^\top$, and $\boldsymbol{D}_i = \mathrm{diag}(D_{i.1}, \dots, D_{i.p})$ and $\boldsymbol{\Sigma}_i^{-1}$ are dispersion matrices (that are also positive definite), $\boldsymbol{\theta}_{\mathrm{f}i} = \tau_i\boldsymbol{\Pi}_{\mathrm{f}i}\boldsymbol{1}_q$, $\boldsymbol{\theta}_{\mathrm{e}i} = \tau_i\boldsymbol{\Pi}_{\mathrm{e}i}\boldsymbol{1}_p$, $\boldsymbol{\Pi}_{\mathrm{e}i} = \mathrm{diag}(\boldsymbol{\varpi}_{\mathrm{e}i})$ and $\boldsymbol{\mu}_i = (\upsilon_i, \psi_i)^\top$, $i = 1, \dots, g$.

Using the Eqs (13) and (16), we have:

$$\mathrm{g}(\boldsymbol{y}_j|\boldsymbol{\Delta}) = \textstyle\sum_{i=1}^{g} p_i\mathscr{f}(\boldsymbol{y}_j|\boldsymbol{\Omega}_i); \ \ \ j = 1, \dots, n, \quad (17)$$

such that $\mathscr{f}(\boldsymbol{y}_j|\boldsymbol{\Omega}_i)$ is the HTH pdf given in Eq (14) for each component and $\boldsymbol{\Omega}_i = (\boldsymbol{\theta}_i, \boldsymbol{L}_i, \boldsymbol{D}_i, \boldsymbol{\Pi}_{\mathrm{f}i}, \boldsymbol{\Pi}_{\mathrm{e}i}, \boldsymbol{\mu}_i)$, where $\boldsymbol{\Delta} = (p_1, \dots, p_{g-1}, \boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_g)$. The Likelihood function of the proposed

HTH-MFA model which is useful in the Bayesian approach is obtained in the Appendix.

4.3. Bayesian approach on the HTH-MFA model

As mentioned before, Bayesian methods are useful in statistical analysis with several advantages, some of which are mentioned in the literature (see e.g., Zhou et al. [32]). Now, we intend to a Bayesian framework to estimate the HTH-MFA model's parameters. First, we consider prior distributions for the model parameters. Also, for the elements of $\boldsymbol{\Delta}$, the proper (weakly) informative and independent priors have been adopted. We represented the loading factor $\tilde{\boldsymbol{L}}_i = [\ell_{i.sr}]$ ($\ell_{i.sr}$ are $\tilde{\boldsymbol{L}}_i$ elements) matrix. So, the following priors are considered:

$$\boldsymbol{p} = \left(p_1, \dots, p_g\right) \sim Dir\left(\delta_1, \dots, \delta_g\right), \;\; \boldsymbol{\theta}_i \sim N_p(\boldsymbol{\xi}_i, \boldsymbol{S}_i),$$

$$\ell_{i.sr} \sim N_1(\theta_{\ell i}, \sigma_{\ell i}^2);\; s > r,\;\; \ell_{i.rr} \sim TN_1(\theta_{\ell i}, \sigma_{\ell i}^2;\; \ell_{i.rr} > 0),$$

$$D_{i.s} \sim IG(\varrho_i, \xi_i),\;\; \boldsymbol{\varpi}_{\mathrm{e}i} \sim N_p(\boldsymbol{a}_i, \boldsymbol{A}_i),$$

$$\boldsymbol{\Pi}_{\mathrm{f}i} \sim MN_{m,q}(\boldsymbol{C}_{\mathrm{f}i}, \boldsymbol{H}_{\mathrm{f}i}, \boldsymbol{N}_{\mathrm{f}i}),\;\; \tau_i \sim TN_1(\theta_{\tau i}, \sigma_{\tau i}^2;\; \tau_i < 0),$$

such that MN referred to the *Matrix-Normal* distributions, and we assume that $\upsilon_i \sim N(\theta_i, \sigma_i^2)$ and $\psi_i \sim E(\varsigma_i)$ where $E(\varsigma_i)$ denotes the exponential distribution with parameter ς_i. The notations IG and Dir, denote the "Inverse Gamma" and "Dirichlet" distributions, respectively. The maintained priors are assumed to be independent. The posterior $\pi\left(\boldsymbol{\Delta}, \widetilde{\boldsymbol{F}}, \boldsymbol{u}, \boldsymbol{x}, \boldsymbol{z} \middle| \boldsymbol{y}\right) \propto L\left(\boldsymbol{\Delta} \middle| \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{u}, \boldsymbol{x}, \boldsymbol{z}\right)\pi(\boldsymbol{\Delta})$, is not analytically tractable but using an *MCMC* methods, we employ the Metropolis-Hastings algorithms attributed to Gamerman [9] and the Gibbs sampling (Gelfand and Smith [10]) to generate samples by employing the following posteriors. The posterior distributions of the parameters of HTH-MFA model are obtained in the Appendix.

4.4. Assigning of missing values

The usefulness of the HTH-MFA hierarchical form is its simplicity for simulation or its usage in sampling from the posteriors using available software in Bayesian analysis such as NIMBLE (NIMBLE Development Team [28]). Note that some of the other Bayesian software such as JAGS and OpenBUGS are not able to inverse the matrices because of the lack of special functions, and so cannot draw a sample from the *GIG* distribution. An advantage of such ability in the parameter updates is to easily locate missing values and assign them from the model naturally. Let $\boldsymbol{Y}_M$ as the missing and $\boldsymbol{Y}_O$ as the observed responses of HTH-MFA model, respectively. Posterior predictive distribution can help to assign the missing data in a Bayesian framework, $P(\boldsymbol{Y}_M|\boldsymbol{Y}_O) = \int P(\boldsymbol{Y}_M|\boldsymbol{Y}_O, \boldsymbol{\Omega})\, P(\boldsymbol{\Omega} \mid \boldsymbol{Y}_O) d\boldsymbol{\Omega}$. When dealing with missing data problems and the missing pattern is not known, it is not possible to directly simulate form the posterior predictive distribution, and the Gibbs sampling technique is employed with updated parameters with: $\boldsymbol{y}_{i,M}^{(t+1)} \sim P\left(\boldsymbol{y}_{i,M} \middle| \boldsymbol{y}_O, \boldsymbol{\Omega}^{(t)}\right)$ for $i = 1, \dots, N$ and $\boldsymbol{\Omega}^{(t+1)} \sim P\left(\boldsymbol{\Omega}^{(\boldsymbol{t})} \middle| \boldsymbol{y}_O, \boldsymbol{y}_{i,M}\right)$. To lead to convergence, the process is started with the initial values $\boldsymbol{y}_{i,M}^{(0)}$and $\boldsymbol{\Omega}^{(0)}$ and a large number of iterations of the Gibbs algorithm is ran. This approach has been implemented in the paper using *NIMBLE* and can be extended in situations with missing data which may be due to other covariates.

5. Applications

This study was conducted to evaluate the flexibility and performance of the introduced HTH-MFA model using simulations and real datasets.

5.1. Priors and details of computation

Largely non-informative priors have used for estimation with different models as follows: $\boldsymbol{\theta} \sim N_2(\mathbf{0}, 10^2\boldsymbol{I}_2)$, $\boldsymbol{\varpi}_i \sim N_2(\mathbf{0}, 10^2\boldsymbol{I}_2)$, $\ell_{i.rr} \sim TN_1(0,100)I(\ell_{i.rr} > 0)$ $\ell_{i.rt} \sim N_1(0,100)$; $r > t$, and $D_{i.r} \sim IG(1,1)$ for $i = 1,2$ and $\boldsymbol{p} \sim Dir(1, \ldots, 1)$. Note that 35,000 iterations have been used for Gibbs sampling runs with burn-in of 10,000. Also, the statistic attributed to Gelman and Rubin [11] and the visual inspection needed for convergence criteria have been employed. Models developed by NIMBLE and all computations have also been verified. To overcome the label switching issue over the MCMC iterations, maximum a posteriori (MAP) estimate has been used. To prevent some common computational issues in the factor analysis, the scale function in R to scale the examined datasets has been used. The models' performance has been evaluated by using some model selection criteria and comparing the classification accuracy. To study the accuracy of classification, the adjusted Rand Index (ARI) attributed by Hubert and Arabie [15]. The range of ARI is between 0 to 1 where 0 is without any match and 1 is perfect match.

5.2. First simulation

To evaluate the performance within the HTH-MFA class, first, data from a particular HTH-MFA model has been simulated, and then the model is compared with other HTH-MFA members based on the adjusted Rand Index (See Figures 1 and 2 and Table 1).

Dataset has been first simulated from the HTH-MFA model given in Eq (10) with three components, $p = 5$, $m = 3$ and $q = 2$, where:

$$\boldsymbol{\theta}_i = \mathbf{0};\ \ \boldsymbol{D}_i = \text{diag}(0.5);\ \boldsymbol{\mu}_i = (-0.5,1)^{\top};\ \boldsymbol{\Pi}_{ei} = \mathbf{0};\ i = 1,2,3.,$$

$$\boldsymbol{\Pi}_{f1} = \begin{pmatrix} 6 & 2 \\ 3 & 6 \\ 0 & 3 \end{pmatrix};\ \boldsymbol{\Pi}_{f2} = \begin{pmatrix} -0.1 & 0.4 \\ 0.1 & -0.2 \\ -0.6 & 0 \end{pmatrix};\ \boldsymbol{\Pi}_{f3} = \begin{pmatrix} -1 & -4 \\ -2 & -2 \\ -4 & -1 \end{pmatrix},$$

with loading matrix

$$\mathbf{L}_i = \begin{pmatrix} 0.8 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.1 \\ 0.2 & 0.0 & 0.3 \\ 0.1 & 0.3 & 0.0 \end{pmatrix}, \qquad i = 1,2,3.,$$

for all three components, and mixture weights ($\pi_1 = 100/500$, $\pi_2 = 150/500$, $\pi_3 = 250/500$), corresponding to sample sizes $n_1 = 150$, $n_2 = 150$ and $n_3 = 250$.

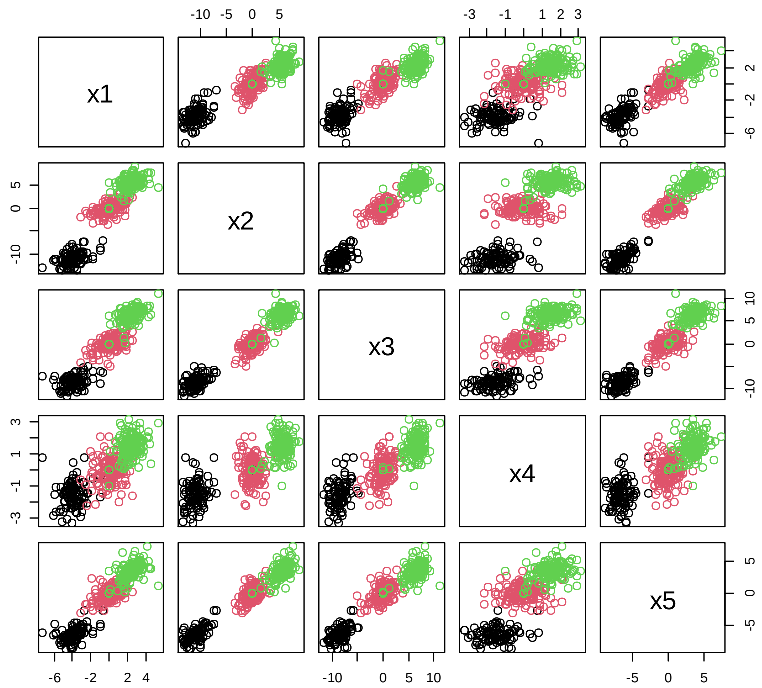

Figure 1. Scatterplot for the pairs of simulated data with indicative colors component labels (Black = Component 1, Red = Component 2, Green = Component 3).

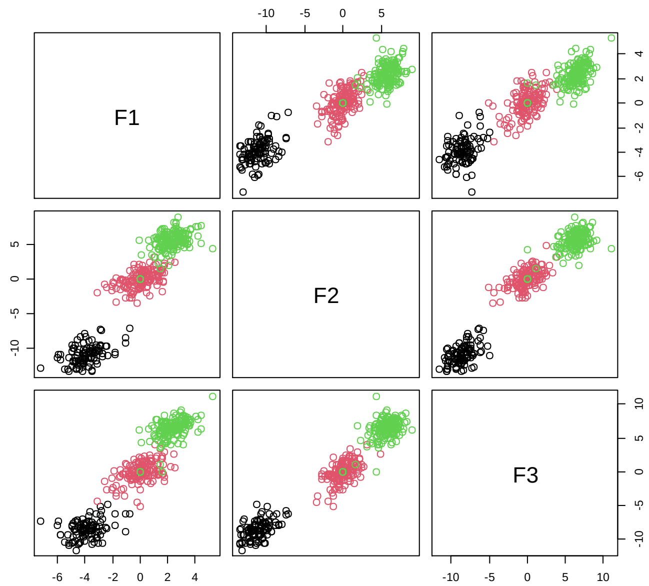

Figure 2. Pairwise scatterplot for the factor scores with indicative colors component labels (Black = Component 1, Red = Component 2, Green = Component 3).

Figures 1 and 2 show the scatterplots for each pair of the simulated data and their first three factors, respectively. As it can be observed in Table 1, the HTH-MFA with $m = 3$ and $q = 2$ has the maximum

ARI and consequently the best performance. Because the factor scores are not heavy-tailed, the HTH-MFA model has similar performance (based on the classification score) to the HTH-MFA for $m = 2$, but with other model selection measures such situation does not exist. Moreover, the SN-MFA model has the worst performance, because it does not allow us to consider long -tails for the distribution of true factor scores.

Table 1. The adjusted Rand Index (ARI) for model selection criteria with different models

Model	SN-MFA[1]	SNGH-MFA[2]	HTH-MFA(q=2)	HTH-MFA(q=3)
2-component	0.68	0.80	0.87	0.82
3-component	0.76	0.84	**0.95***	0.86
4-component	0.74	0.81	0.90	0.79

*: The bold value is corresponding to the best model.
[1]: SN-MFA: skew normal mixture of factor analyzers.
[2]: SNGH-MFA: skew normal generalized hyperbolic mixture of factor analyzers with $q = 1$.

5.3. Second simulation

According to suggestion of a reviewer to show importance and performance of the proposed MFA model than the ordinary MFA models, we consider the previous simulation (in the previous section) of our proposed MFA model, but in the two cases of light and heavy tailed-ness (each one with 500 samples), respectively, given by $\boldsymbol{\mu}_i = (20,15)^\top$; $i = 1,2,3$, and $\boldsymbol{\mu}_i = (0.1,2)^\top$; $i = 1,2,3$. We fitted the HTT-MFA and two ordinary light-tailed Gaussian MFA and heavy-tailed Student-t MFA models, and numbers of correct estimations of the number of groups (components) are listed in Table 2.

Table 2. Numbers (percentages) of the correct number of components estimated of the HTH-MFA simulations.

Model	Model	**Gaussian-MFA**	**Student-t-MFA**	**HTH-MFA**
Light-tailed	2-component	369 (73.8%)	388 (77.6%)	45 (9%)
	3-component	91 (18.2%)	101 (20.2%)	437 (87.4%)
	4-component	40 (8%)	11 (2.2%)	18 (3.6%)
Heavy-tailed	2-component	18 (3.6%)	270 (54%)	33 (6.6%)
	3-component	70 (14%)	201 (40.2%)	424 (84.8%)
	4-component	412 (82.4%)	29 (5.8%)	43 (8.6%)

It can be seen from results in Table 3, in the cases of the HTH-MFA datasets (in fact the datasets with asymmetrical and including outliers), the ordinary MFA models have under-estimates and over-estimates of the true number of components, while the heavy-tailed Student-t MFA model has better performance than the light-tailed Gaussian MFA model.

5.4. Real datasets

In this part, we consider three real datasets with the following details, and then we fitted the proposed MFA models in the previous part for clustering and semi-supervised classification on them.

We recorded the ARI index for evaluating the fitness, because the true classes of the points that are treated as un-labeled are actually known.

The first dataset is the Australian Institute of Sport (AIS) data for 100 female and 102 male athletes, which contains 11 continuous variables (these data are available in the Australasian Data and Story Library; Smyth [20]).

The second dataset is the Sonar data, which Gorman and Sejnowski [14] report this data on patterns obtained by bouncing sonar signals off a metal cylinder and rocks, respectively, at various angles and under various conditions. These data also are available from the UCI machine learning repository. In the data 111 signals are recorded by bouncing sonar signals off a metal cylinder and 97 are recorded by bouncing sonar signals off rocks. To illustrate the MFA approaches for semi-supervised classification, half of the 208 patterns (60 of the metal signals and 44 of the rock signals) are selected (randomly) to be treated as un-labeled, and the proposed models are fitted to these data, for semi-supervised classification.

The third dataset is the Hawks dataset (Stat2Data R package [7]), which includes the data of three species of copper, sharp-legged and red-tailed hawks. Overall, this data contains 19 features 908 valid observations. The concentration has been on the length of primary wing feather, the hallux length, the culmen length, the tail length (all of them based on mm) and the weight (gr) of. We fit the proposed MFA approaches to classify the three species.

For all of datasets we fit the HTH-MFA models with $g = 2,\ m = 1,\dots,6,\ q = 1,\dots,6$, and $q \le m$. Also, for comparison, we also fit the (light-tailed) SN-MFA and (heavy-tailed) SNGH-MFA models for $g = 2$ and $m = 1,\dots,6$. The best fitted models were based on the $m = 2$ and $q = 2$, which their ARI are given in Table 3.

Table 3. The adjusted Rand Index (ARI) for model selection criteria with different models for AIS, Sonar and Hawks datasets with $\boldsymbol{m = 2}$ and $\boldsymbol{q = 2}$.

Model	**SN-MFA**	**SNGH-MFA**	**HTH-MFA(q=2)**
AIS	0.80	0.85	**0.93**
Sonar	0.35	0.46	**0.62**
Hawks	0.86	0.88	**0.91**

To see the advantage of the proposed Bayesian technique, we consider experiments on the proposed real datasets by evaluating the applicability of the HTH-MFA classification, and also related errors concerning missing dataset context. The hierarchical form of the HTH-MFA allows us to code and perform the MCMC technique for the model estimations in NIMBLE. Additionally, assigning missing values (i.e., conditional means) from the full MFA model is relatively facilitated. In this experiment, some values of the datasets in the form of randomness under two levels of low missingness (5%) and moderate missingness (15%) have been deleted and the total sample is obtained. Then, the performance of assigning values has been compared using the model based on the conditional technique, or according to the mean assigning based on the unconditional approach, where the means are considered instead of the missing values.

The mean squared error (MSE) is also computed by:

$$MSE = \frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{y}_j^m - \hat{\boldsymbol{y}}_j^m\right)^{\top}\left(\boldsymbol{y}_j^m - \hat{\boldsymbol{y}}_j^m\right),$$

such that $\hat{\boldsymbol{y}}$ is the imputed value for $\boldsymbol{y}$. Note that the total number of missing values is $n^* = \sum_{j=1}^{n}(p - p_j^o)$.

Table 4 reports the means and standard deviations (in parenthesis) of the ARI and MSE for the conditional (C) and unconditional (UC) models based on 50 replications in the datasets for two different maintained missingness rate scenarios (5% or 15%).

Table 4. The MSE and ARI for AIS, Sonar and Hawks datasets (with missing data).

	HTH-MFA (UC)				**HTH-MFA (C)**			
	MSE		**ARI**		**MSE**		**ARI**	
Missing Rate (%)	**5%**	**15%**	**5%**	**15%**	**5%**	**15%**	**5%**	**15%**
AIS Dataset	0.54 (0.66)	0.55 (0.56)	0.80 (0.03)	0.71 (0.02)	0.41 (0.23)	0.44 (0.17)	0.88 (0.0)	0.82 (0.0)
Sonar Dataset	0.62 (0.53)	0.51 (0.33)	0.56 (0.01)	0.46 (0.02)	0.53 (0.40)	0.37 (0.13)	0.61 (0.01)	0.50 (0.01)
Hawks Dataset	0.56 (0.71)	0.54 (0.39)	0.84 (0.02)	0.85 (0.03)	0.48 (0.40)	0.49 (0.21)	0.89 (0.0)	0.86 (0.0)

According to the MSE and ARI from Table 3, for two maintained unconditional and conditional scenarios from all three datasets, the conditional HTH-MFA (C) model has clearly better performance than the unconditional HTH-MFA (UC) model.

From the Hawks dataset, we had similarity about the mean squared error and the performance of the classification for both types of models. This seems to indicate that the number of needed factors to represent the data is equal to 1. For other datasets, the results differ with Hawks dataset and there is a greater sensitivity on the missing data imputation technique. When there is just a single missing value, an alternative way for the unconditional approach is Listwise deletion (the entire record is removed). This method can be applied only for large samples, and in most applications where the FA model is commonly used is rare. Thus, using the full model in the conditional approach is often preferred and used but it is dependent on the availability and simplicity of using the computational approach in practice.

6. Discussion and conclusions

In this work, using a Bayesian framework, a flexible class of multivariate HTH distributions was employed to analyze MFA models. The estimation of the parameters of the MFA model based on HTH family was relatively straightforward with a Bayesian approach. As the HTH distributions are suitable to model the asymmetric data with/without outliers, it can be used in the robust statistical inferences. Various extensions to the HTH-MFA model can be considered for future works, such as more general arrangement of a structural equation model based on the HTH distributions or extending ordinary available models with sparse covariance structures. Also, to improve estimates, more informative priors (such as empirically derived or known a priori algorithm) on the variance of the noisy settings or the error term have been considered.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors would like to thank the editor and anonymous reviewers for their constructive suggestions, correction and encouragement, which helped us to improve earlier versions of the manuscript.

Conflict of interest

The authors declare no conflicts of interest.

References

1. T. Ando, Bayesian factor analysis with fat-tailed factors and its exact marginal likelihood, *J. Multivariate Anal.*, **100** (2009), 1717–1726. https://doi.org/10.1016/j.jmva.2009.02.001
2. R. B. Arellano-Valle, M. G. Genton, On fundamental skew distributions, *J. Multivariate Anal.*, **96** (2005), 93–116. https://doi.org/10.1016/j.jmva.2004.10.002
3. J. Bai, K. Li, Statistical analysis of factor models in high dimensions, *Ann. Statist.*, **40** (2012), 436–465. https://doi.org/10.1214/11-AOS966
4. O. Barndorff-Nielsen, C. Halgreen, Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions, Z. Wahrscheinlichkeitstheorie Verw. Gebiete, **38** (1977), 309–311. https://doi.org/10.1007/BF00533162
5. M. Bazrafkan, K. Zare, M. Maleki, Z. Khodadi, Partially linear models based on heavy-tailed and asymmetrical distributions, *Stoch. Environ. Res. Risk Assess.*, **36** (2022), 1243–1253. https://doi.org/10.1007/s00477-021-02101-1
6. M. D. Branco, D. K. Dey, A general class of multivariate skew-elliptical distributions, *J. Multivariate Anal.*, **79** (2001), 99–113. https://doi.org/10.1006/jmva.2000.1960
7. A. Cannon A, G. Cobb, B. Hartlaub, J. Legler, R. Lock, T. Moore, et al., Stat2Data: Datasets for Stat2. R package version 2.0.0, 2019, Available from: https://CRAN.R-project.org/package=Stat2Data.
8. M. Chen, J. Silva, J. Paisley, C. Wang, D. Dunson, L. Carin, Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds, *IEEE T. Signal Proces*, **58** (2010), 6140–6155. https://doi.org/10.1109/TSP.2010.2070796
9. D. Gamerman, *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference*, London: Chapman & Hill, 1997.
10. A. E. Gelfand, A. F. M. Smith, Sampling based approaches to calculating marginal densities, *J. Am. Stat. Assoc.*, **85** (1990), 398–409. https://doi.org/10.2307/2289776
11. A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences, *Statist. Sci.*, **7** (1992), 457–472. https://doi.org/10.1214/ss/1177011136

12. Z. Ghahramani, G. E. Hinton, The EM algorithm for mixtures of factor analyzers (Technical Repoort CRG-TR-96-1), Department of Computer Science, University of Toronto, 6 King's College Road, Toronto, Canada, M5S 1A4, 1997.
13. I. J. Good, The population frequencies of species and the estimation of population parameters, Biometrika, **40** (1953), 237–264. https://doi.org/10.1093/biomet/40.3-4.237
14. R. P. Gorman, T. J. Sejnowski, Analysis of hidden units in a layered network trained to classify sonar targets, *Neural Networks*, **1** (1988), 75–89. https://doi.org/10.1016/0893-6080(88)90023-8
15. L. Hubert, P. Arabie, Comparing partitions, *J. Classif.*, **2** (1985), 193–218. https://doi.org/10.1007/BF01908075
16. H. M. Kim, M. Maadooliat, R. B. Arellano-Valle, M. G. Genton, Skewed factor models using selection mechanisms, *J. Multivariate Anal.*, **145** (2016), 162–177. https://doi.org/10.1016/j.jmva.2015.12.007
17. S. X. Lee, G. J. McLachlan, Finite mixtures of canonical fundamental skew t-distributions, *Stat. Comput.*, **26** (2016), 573–589. https://doi.org/10.1007/s11222-015-9545-x
18. S. X. Lee, G. J. McLachlan, On formulations of skew factor models: skew errors versus skew factors, *Stat. Probabil. Lett.*, **168** (2021), 108935. https://doi.org/10.1016/j.spl.2020.108935
19. S. Y. Lee, Y. M. Xia, A robust Bayesian approach for structural equation models with missing data, *Psychometrika*, **73** (2008), 343–364. https://doi.org/10.1007/s11336-008-9060-5
20. G. K. Smyth, Australasian Data and Story Library (OzDASL), 2011. https://gksmyth.github.io/ozdasl
21. T. I. Lin, J. C. Lee, S. Y. Yen, Finite mixture modeling using the skew-normal distribution, *Stat. Sinica.*, **17** (2007), 909–927.
22. T. I. Lin, G. J. McLachlan, S. X. Lee, Extending mixtures of factor models using the restricted multivariate skew-normal distribution, *J. Multivariate Anal.*, **143** (2016), 398–413. https://doi.org/10.1016/j.jmva.2015.09.025
23. C. Luo, L. Shen, A. Xu, Modelling and estimation of system reliability under dynamic operating environments and lifetime ordering constraints, *Reliab. Eng. Syst. Safe.*, **218** (2022), 108136. https://doi.org/10.1016/j.ress.2021.108136
24. M. R. Mahmoudi, M. Maleki, D. Baleanu, V. T. Nguyen, K. H. Pho, A Bayesian approach to heavy-tailed finite mixture autoregressive models, *Symmetry*, **12** (2020), 929. https://doi.org/10.3390/sym12060929
25. M. Maleki, G. J. McLachlan, S. X. Lee, Robust clustering based on finite mixture of multivariate fragmental distributions, *Stat. Model.*, **23** (2023), 247–272. https://doi.org/10.1177/1471082X211048660
26. J. S. Marron, M. P. Wand, Exact mean integrated squared error, *Ann. Statist.*, **20** (1992), 712–736.
27. P. M. Murray, R. B. Browne, P. D. McNicholas, Hidden truncation hyperbolic distributions, finite mixtures thereof, and their application for clustering, *J. Multivariate Anal.*, **161** (2017), 141–156. https://doi.org/10.1016/j.jmva.2017.07.008
28. NIMBLE Development Team, NIMBLE: An R package for programming with BUGS models, Version 0.6-10, 2021, Available from: http://r-nimble.org.
29. K. Roeder, L. Wasserman, Practical Bayesian density estimation using mixtures of normals, *J. Am. Stat. Assoc.*, **92** (1997), 894–902. https://doi.org/10.2307/2965553

30. M. M. Wall, J. Guo, Y. Amemiya, Mixture factor analysis for approximating a non-normally distributed continuous latent factor with continuous and dichotomous observed variables, *Multivariate Behav. Res.*, **47** (2012), 276–313. https://doi.org/10.1080/00273171.2012.658339
31. M. Yang, D. B. Dunson, Bayesian semiparametric structural equation models with latent variables, *Psychometrika*, **75** (2010), 675–693. https://doi.org/10.1007/s11336-010-9174-4
32. S. Zhou, A. Xu, Y. Tang, L. Shen, Fast Bayesian inference of reparameterized gamma process with random effects, *IEEE T. Reliab.*, 2023, 1–14. https://doi.org/10.1109/TR.2023.3263940

Appendix

A.1. Likelihood function of the HTH-MFA model

Let the complete data with $\boldsymbol{C} = \{\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Z}\}$, $\tilde{\boldsymbol{L}}_i = \boldsymbol{L}_i \boldsymbol{\Sigma}_i^{-1/2}$ and $\tilde{\boldsymbol{F}}_{ij} = \boldsymbol{\Sigma}_i^{1/2} \boldsymbol{F}_{ij}$, according to the stochastic form of the HTH family given in Eq (7) and the conditional distribution given in Eq (15), HTH-MFA model can be hierarchically presented by:

$$\boldsymbol{Y}_j \big| \tilde{\boldsymbol{F}}_{ij}, \boldsymbol{X}_{\mathrm{e}ij} = \boldsymbol{x}_{\mathrm{e}ij}, Z_j = i \overset{ind.}{\sim} N_p\big(\boldsymbol{\theta}_i + \tilde{\boldsymbol{L}}_i \tilde{\boldsymbol{F}}_{ij} + \boldsymbol{\theta}_{\mathrm{e}i} + \boldsymbol{\Pi}_{\mathrm{e}i} \boldsymbol{x}_{\mathrm{e}ij}, \kappa(u_{ij}) \boldsymbol{D}_i\big), U_{ij} = u_{ij},$$

$$\tilde{\boldsymbol{F}}_{ij} \big| \boldsymbol{X}_{\mathrm{f}ij} = \boldsymbol{x}_{\mathrm{f}ij}, Z_j = i \overset{ind.}{\sim} N_m\big(\boldsymbol{\theta}_{\mathrm{f}i} + \boldsymbol{\Pi}_{\mathrm{f}i} \boldsymbol{x}_{\mathrm{f}ij}, \kappa(u_{ij}) \boldsymbol{I}_m\big), U_{ij} = u_{ij},$$

$$\boldsymbol{X}_{\mathrm{e}ij} \big| U_{ij} = u_{ij}, Z_j = i \overset{ind.}{\sim} TN_p\big(\boldsymbol{0}, \kappa(u_{ij}) \boldsymbol{I}_p;\ \boldsymbol{X}_{\mathrm{e}ij} > \boldsymbol{0}\big),$$

$$\boldsymbol{X}_{\mathrm{f}ij} \big| U_{ij} = u_{ij}, Z_j = i \overset{ind.}{\sim} TN_q\big(\boldsymbol{0}, \kappa(u_{ij}) \boldsymbol{I}_q;\ \boldsymbol{X}_{\mathrm{f}ij} > \boldsymbol{0}\big),$$

$$U_{ij} \big| Z_j = i \overset{ind.}{\sim} GIG_*(\boldsymbol{\mu}_i),$$

$$P\big(Z_j = i\big) = p_i,$$

such that $TN_k(\boldsymbol{m},\ \boldsymbol{M};\ \boldsymbol{X} > \boldsymbol{r})$ is referred to the truncated *k*-variate normal on the space $\boldsymbol{X} > \boldsymbol{r}$ with mean and covariance $\boldsymbol{m}$ and $\boldsymbol{M}$, respectively, before truncation, where its pdf presented by $T\phi_k(\boldsymbol{x}|\boldsymbol{m},\ \boldsymbol{M}; \boldsymbol{r})$. Consequently, we have:

$$\boldsymbol{Y}_j \big| \tilde{\boldsymbol{F}}_{ij}, \boldsymbol{X}_{\mathrm{e}ij} = \boldsymbol{x}_{\mathrm{e}ij}, Z_j = i \overset{ind.}{\sim} N_p\big(\boldsymbol{\theta}_i + \tilde{\boldsymbol{L}}_i \tilde{\boldsymbol{F}}_{ij} + \boldsymbol{\Pi}_{\mathrm{e}i} \boldsymbol{x}_{\mathrm{e}ij}, \kappa(u_{ij}) \boldsymbol{D}_i\big), U_{ij} = u_{ij},$$

$$\tilde{\boldsymbol{F}}_{ij} \big| \boldsymbol{X}_{\mathrm{f}ij} = \boldsymbol{x}_{\mathrm{f}ij}, Z_j = i \overset{ind.}{\sim} N_m\big(\boldsymbol{\Pi}_{\mathrm{f}i} \boldsymbol{x}_{\mathrm{f}ij}, \kappa(u_{ij}) \boldsymbol{I}_m\big), U_{ij} = u_{ij},$$

$$\boldsymbol{X}_{\mathrm{e}ij} \big| U_{ij} = u_{ij}, Z_j = i \overset{ind.}{\sim} TN_p\big(\tau_i \boldsymbol{1}_p, \kappa(u_{ij}) \boldsymbol{I}_p;\ \boldsymbol{X}_{\mathrm{e}ij} > \tau_i \boldsymbol{1}_p\big),$$

$$\boldsymbol{X}_{\mathrm{f}ij} \big| U_{ij} = u_{ij}, Z_j = i \overset{ind.}{\sim} TN_q\big(\tau_i \boldsymbol{1}_q, \kappa(u_{ij}) \boldsymbol{I}_q;\ \boldsymbol{X}_{\mathrm{f}ij} > \tau_i \boldsymbol{1}_q\big),$$

$$U_{ij}\big|Z_j = i \overset{ind.}{\sim} GIG_*(\boldsymbol{\mu}_i),$$

$$P(Z_j = i) = p_i.$$

The complete augmented likelihood function of $\boldsymbol{\Pi}$ based on the above hierarchical representations can be written by:

$$L(\boldsymbol{\Delta}|\boldsymbol{C}) = \prod_{j=1}^{n}\prod_{i=1}^{g}\{p_i\phi_p(\boldsymbol{y}_j|\boldsymbol{\theta}_i + \tilde{\boldsymbol{L}}_i\widetilde{\boldsymbol{F}}_{ij} + \boldsymbol{\Pi}_{\mathrm{e}i}\boldsymbol{x}_{\mathrm{e}ij}, \kappa(u_{ij})\boldsymbol{D}_i)\phi_m(\widetilde{\boldsymbol{F}}_{ij}|\boldsymbol{\Pi}_{\mathrm{f}i}\boldsymbol{x}_{\mathrm{f}ij}, \kappa(u_{ij})\boldsymbol{I}_m) \times$$

$$T\phi_p(\boldsymbol{X}_{\mathrm{e}ij}|\tau_i\boldsymbol{1}_p, \kappa(u_{ij})\boldsymbol{I}_p;\ \tau_i\boldsymbol{1}_p)\,T\phi_q(\boldsymbol{X}_{\mathrm{f}ij}|\tau_i\boldsymbol{1}_q, \kappa(u_{ij})\boldsymbol{I}_q;\ \tau_i\boldsymbol{1}_q) \times \mathcal{GIG}_*(u_{ij}|\boldsymbol{\mu}_i)P(Z_j = i)\}.$$

A.2. Posteriors

Let n_i as the number of observations which devoted to the i-th component of *HTH-FA*, $B_i = \{j: z_j = i\}$, and $\boldsymbol{\Delta}_{(-m)}$ as the parameters set excluding the parameter m. Except the derived parameters of the scale mixer variable, all of the other conditional posteriors of the THT-MFA model parameters have the following closed form:

$$\boldsymbol{p}|\boldsymbol{\Delta}_{(-\boldsymbol{p})}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{z} \sim Dir(\delta_{p.1}, \dots, \delta_{p.g}),$$

such that $\delta_{p.i} = \delta_i + n_i;\ i = 1, \dots, g$.

$$\boldsymbol{\theta}_i|\boldsymbol{\Delta}_{(-\boldsymbol{\theta}_i)}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{x}, \boldsymbol{u}, z_j = i\ \sim N_p(\boldsymbol{\xi}_{i.p}, \boldsymbol{S}_{i.p}),\ \ i = 1, \dots, g,$$

such that

$$\boldsymbol{S}_{i.p} = \left(\boldsymbol{S}_i^{-1} + \textstyle\sum_{B_i}\kappa^{-1}(u_{ij})\boldsymbol{D}_i^{-1}\right)^{-1},$$

And

$$\boldsymbol{\xi}_{i.p} = \boldsymbol{S}_{i.p}\left[\boldsymbol{S}_i^{-1}\boldsymbol{\xi}_i + \sum\nolimits_{B_i}\kappa^{-1}(u_{ij})\boldsymbol{D}_i^{-1}(\boldsymbol{y}_j - \tilde{\boldsymbol{L}}_i\widetilde{\boldsymbol{F}}_{ij} - \boldsymbol{\Pi}_{\mathrm{e}i}\boldsymbol{x}_{\mathrm{e}ij})\right].$$

$$\ell_{i.sr}|\boldsymbol{\Delta}_{(-\ell_{i.sr})}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{x}, \boldsymbol{u}, z_j = i\ \sim N_1(\theta_{\ell i.p}, \sigma^2_{\ell i.p}),$$

such that

$$\sigma^2_{\ell i.p} = \frac{1}{\sigma_{\ell i}^{-2} + \sum_{B_i}D_{i.s}^{-1}\kappa^{-1}(u_{ij})\tilde{F}^2_{ij.r}}.$$

And

$$\theta_{\ell i.p} = \sigma^2_{\ell i.p}\left[\theta_{\ell i}/\sigma^2_{\ell i.} + \textstyle\sum_{B_i}D_{i.s}^{-1}\kappa^{-1}(u_{ij})\tilde{F}_{ij.r}\left(y_{j.s} - \theta_{i.s} - \boldsymbol{\ell}^{\top}_{i.s(-r)}\widetilde{\boldsymbol{F}}_{ij} - \varpi_{\mathrm{e}i.s}x_{\mathrm{e}ij.s}\right)\right],$$

for $r = 1, \dots, m;\ s = 1, \dots, p;\ j = 1, \dots, n;\ i = 1, \dots, g$, such that $y_{j.s}$, $x_{\mathrm{e}ij.s}, \varpi_{\mathrm{e}i.s}$ and $\theta_{i.s}$ are the s-th components of $\boldsymbol{y}_j$, $\boldsymbol{x}_{\mathrm{e}ij}, \boldsymbol{\varpi}_{\mathrm{e}i}$, and $\boldsymbol{\theta}_i$, $\boldsymbol{\ell}_{i.s}$ and $\tilde{F}_{ij.r}$ are the s-th row and the r-th component of $\boldsymbol{L}_i$ and $\widetilde{\boldsymbol{F}}_{ij}$, respectively.

Moreover, we have:

$$\ell_{i.rr}|\boldsymbol{\Delta}_{(-\ell_{i.rr})}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{x}, \boldsymbol{u}, z_j = i \sim TN_1\left(\theta_{\ell.p}, \sigma^2_{\ell.p};\ \ell_{i.rr} > 0\right),$$

$$\widetilde{\boldsymbol{F}}_{ij}\big|\boldsymbol{\Delta}, \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}, z_j = i \sim N_m(\boldsymbol{\theta}_{Fi}, \Xi_{Fi}),\ \ i = 1, \dots, g; j = 1, \dots, n,$$

such that

$$\Xi_{Fi} = \kappa\left(u_{ij}\right)\left(\boldsymbol{I}_m + \widetilde{\boldsymbol{L}}_i^\top \boldsymbol{D}_i^{-1} \widetilde{\boldsymbol{L}}_i\right)^{-1},$$

And

$$\boldsymbol{\theta}_{Fi} = \kappa^{-1}\left(u_{ij}\right)\Xi_{Fi}\left[\boldsymbol{\Pi}_{\mathrm{f}i}\boldsymbol{x}_{\mathrm{f}ij} + \widetilde{\boldsymbol{L}}_i^\top \boldsymbol{D}_i^{-1}\left(\boldsymbol{y}_j - \boldsymbol{\theta}_i - \boldsymbol{\Pi}_{\mathrm{e}i}\boldsymbol{x}_{\mathrm{e}ij}\right)\right],$$

$$D_{i.s}|\boldsymbol{\Delta}_{(-D_{i.s})}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{x}, \boldsymbol{u}, z_j = i \sim IG\left(\varrho_{i.p}, \xi_{i.p}\right),\ \ s = 1, \dots, p;\ i = 1, \dots, g,$$

such that $\varrho_{i.p} = \frac{n_i}{2} + \varrho_i$ and

$$\xi_{i.p} = \xi_i + 0.5 \textstyle\sum_{B_i} \kappa^{-1}\left(u_{ij}\right)\left(y_{j.s} - \theta_{i.s} - \boldsymbol{\ell}_{i.s}^\top \widetilde{\boldsymbol{F}}_{ij} - \varpi_{\mathrm{e}i.s} w_{\mathrm{e}ij.s}\right)^2,$$

$$\boldsymbol{\varpi}_{\mathrm{e}i}|\boldsymbol{\Delta}_{(-\boldsymbol{\varpi}_{\mathrm{e}i})}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{x}, \boldsymbol{u}, z_j = i \sim N_p\left(\boldsymbol{a}_{i.p}, \boldsymbol{A}_{i.\mathrm{p}}\right),\ i = 1, \dots, g,$$

such that

$$\boldsymbol{A}_{i.p} = \left(\boldsymbol{A}_i^{-1} + \textstyle\sum_{B_i} \kappa^{-1}\left(u_{ij}\right)\boldsymbol{P}_{ij}^\top \boldsymbol{D}_i^{-1} \boldsymbol{P}_{ij}\right)^{-1},$$

where

$$\boldsymbol{P}_{ij} = \mathrm{diag}\left(\boldsymbol{x}_{\mathrm{e}ij}\right),$$

And

$$\boldsymbol{a}_{i.p} = \boldsymbol{A}_{i.p}\left[\boldsymbol{A}_i^{-1}\boldsymbol{a}_i + \textstyle\sum_{B_i} \kappa^{-1}\left(u_{ij}\right)\boldsymbol{P}_{ij}^\top \boldsymbol{D}_i^{-1}\left(\boldsymbol{y}_j - \boldsymbol{\theta}_i - \widetilde{\boldsymbol{L}}_i \widetilde{\boldsymbol{F}}_{ij}\right)\right],$$

$$\mathrm{vec}(\boldsymbol{\Delta}_{\mathrm{f}i})|\boldsymbol{\Delta}_{(-\boldsymbol{\Pi}_{\mathrm{f}i})}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{x}, \boldsymbol{u}, z_j = i \sim N_{mq}\left(\boldsymbol{\theta}_{\mathrm{f}i.p}, \Xi_{\mathrm{f}i.p}\right),\ i = 1, \dots, g,$$

such that

$$\boldsymbol{\theta}_{\mathrm{f}i.p} = \Xi_{\mathrm{f}i.p}\left(\left(\boldsymbol{N}_{\mathrm{f}i}^{-1} \otimes \boldsymbol{H}_{\mathrm{f}i}^{-1}\right)\mathrm{vec}(\boldsymbol{C}_{\mathrm{f}i}) + \textstyle\sum_{B_i} \kappa\left(u_{ij}\right)^{-1}\left(\boldsymbol{M}_{\mathrm{f}ij} \otimes \boldsymbol{I}_m\right)\right),$$

and

$$\Xi_{\mathrm{fi}.p} = \left[\left(\boldsymbol{N}_{\mathrm{fi}}^{-1} \otimes \boldsymbol{H}_{\mathrm{fi}}^{-1}\right) + \sum_{B_i} \kappa^{-1}\left(u_{ij}\right)\left(\boldsymbol{R}_{\mathrm{fi}j} \otimes \boldsymbol{I}_m\right)\right]^{-1},$$

where $\boldsymbol{M}_{\mathrm{fi}j} = \widetilde{\boldsymbol{F}}_{ij}\boldsymbol{x}_{\mathrm{fi}j}^{\top}$ and $\boldsymbol{R}_{\mathrm{fi}j} = \boldsymbol{x}_{\mathrm{fi}j}\boldsymbol{x}_{\mathrm{fi}j}^{\top}$, and $\otimes$ and "vec" are referred to the Kronecker product and a matrix vectorization, respectively.

$$\boldsymbol{X}_{\mathrm{fi}j}\big|\boldsymbol{\Delta}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{u}, z_j = i \sim TN_q\left(\boldsymbol{\theta}_{X\mathrm{fi}j}, \Xi_{X\mathrm{fi}j};\ \boldsymbol{X}_{\mathrm{fi}j} > \tau_i \boldsymbol{1}_q\right), j = 1, \dots, n;,\ \ i = 1, \dots, g,$$

such that

$$\Xi_{X\mathrm{fi}j} = \kappa\left(u_{ij}\right)\left(\boldsymbol{I}_q + \boldsymbol{\Pi}_{\mathrm{fi}}^{\top}\boldsymbol{\Pi}_{\mathrm{fi}}\right)^{-1},$$

and

$$\boldsymbol{\theta}_{X\mathrm{fi}j} = \kappa^{-1}\left(u_{ij}\right)\Xi_{X\mathrm{fi}j}\left(\tau_i \boldsymbol{1}_q + \boldsymbol{\Pi}_{\mathrm{fi}}^{\top}\widetilde{\boldsymbol{F}}_{ij}\right),$$

$$\boldsymbol{X}_{\mathrm{e}ij}\big|\boldsymbol{\Delta}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{u}, z_j = i \sim TN_p\left(\boldsymbol{\theta}_{X\mathrm{e}ij}, \Xi_{X\mathrm{e}ij};\ \boldsymbol{X}_{\mathrm{e}ij} > \tau_i \boldsymbol{1}_p\right),\ \ j = 1, \dots, n;,\ \ i = 1, \dots, g,$$

such that

$$\Xi_{X\mathrm{e}ij} = \kappa\left(u_{ij}\right)\left(\boldsymbol{I}_p + \boldsymbol{\Pi}_{\mathrm{e}i}^{\top}\boldsymbol{D}_i^{-1}\boldsymbol{\Pi}_{\mathrm{e}i}\right)^{-1},$$

and

$$\boldsymbol{\theta}_{X\mathrm{e}ij} = \kappa^{-1}\left(u_{ij}\right)\Xi_{X\mathrm{e}ij}\left(\tau_i \boldsymbol{1}_p + \boldsymbol{\Pi}_{\mathrm{e}i}^{\top}\boldsymbol{D}_i^{-1}\left(\boldsymbol{y}_j - \boldsymbol{\theta}_i - \widetilde{\boldsymbol{L}}_i\widetilde{\boldsymbol{F}}_{ij}\right)\right).$$

Consequently, we have:

$$\pi\left(Z_j = i\big|\boldsymbol{\Delta}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{u}, \boldsymbol{x}\right) = \frac{p_i \not{f}(\boldsymbol{y}_j|\boldsymbol{\Omega}_i)}{\sum_{h=1}^{g} p_h \not{f}(\boldsymbol{y}_h|\boldsymbol{\Omega}_h)}, i = 1, \dots, g; j = 1, \dots, n.$$

such that $\not{f}\left(\boldsymbol{y}_j|\boldsymbol{\Omega}_i\right)$ is given in Eq (14).

$$U_{ij}\big|\boldsymbol{\Delta}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{x}, z_j = i \ \sim GIG^*\left(\mathrm{a}_{ij}, \mathrm{b}_{ij}, \sqrt{\mathrm{c}_{ij}}\right),\ i = 1, \dots, g; j = 1, \dots, n,$$

such that $\kappa(u) = u$, $\mathrm{a}_{ij} = v_i - p + (m+q)/2$, $\mathrm{b}_{ij} = \sqrt{\psi_i}$ and

$$\begin{aligned} \mathrm{c}_{ij} = \psi_i &+ \left(\boldsymbol{y}_j - \boldsymbol{\theta}_i - \widetilde{\boldsymbol{L}}_i\widetilde{\boldsymbol{F}}_{ij} - \boldsymbol{\Pi}_{\mathrm{e}i}\boldsymbol{x}_{\mathrm{e}ij}\right)^{\top}\boldsymbol{D}_i^{-1}\left(\boldsymbol{y}_j - \boldsymbol{\theta}_i - \widetilde{\boldsymbol{L}}_i\widetilde{\boldsymbol{F}}_{ij} - \boldsymbol{\Pi}_{\mathrm{e}i}\boldsymbol{x}_{\mathrm{e}ij}\right) \\ &+ \left(\widetilde{\boldsymbol{F}}_{ij} - \boldsymbol{\Pi}_{\mathrm{fi}}\boldsymbol{x}_{\mathrm{fi}j}\right)^{\top}\left(\widetilde{\boldsymbol{F}}_{ij} - \boldsymbol{\Pi}_{\mathrm{fi}}\boldsymbol{x}_{\mathrm{fi}j}\right) + \left(\boldsymbol{X}_{\mathrm{e}ij} - \tau_i \boldsymbol{1}_p\right)^{\top}\left(\boldsymbol{X}_{\mathrm{e}ij} - \tau_i \boldsymbol{1}_p\right) \\ &+ \left(\boldsymbol{X}_{\mathrm{fi}j} - \tau_i \boldsymbol{1}_q\right)^{\top}\left(\boldsymbol{X}_{\mathrm{fi}j} - \tau_i \boldsymbol{1}_q\right) \end{aligned}$$

Furthermore,

$$\pi\left(v_i|\boldsymbol{\Delta}_{(-v_i)}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{u}, \boldsymbol{x}, z_j = i\ \right) \sim \pi_1(v_i)\phi\left(v_i\big|\theta_i + \sigma_i^2 \sum_{B_i} \log(u_i), \sigma_i^2\right),$$

such that $\pi_1(\upsilon_i) = \left(K_{\upsilon_i}(\psi_i)\right)^{-n_i}$;

$$\pi\left(\psi_i|\boldsymbol{\Delta}_{(-\psi_i)}, \boldsymbol{y}, \widetilde{\boldsymbol{F}}, \boldsymbol{u}, \boldsymbol{x}, z_j = i\right) \sim \pi_2(\psi_i) \times E\left(\varsigma_i + \sum_{B_i}\left(u_{ij} + u_{ij}^{-1}\right)/2\right),$$

such that $\pi_2(\psi_i) = \left(K_{\upsilon_i}(\psi_i)\right)^{-n_i}$. The posteriors of υ_i and ψ_i are not in the closed forms but an *MCMC* scheme such as the Metropolis-Hastings algorithm can be embedded to draw samples.