Abstract. Mixture of factor analyzer (MFA) model is an efficient model for the analysis of high dimensional data through which the factor-analyzer technique based on the covariance matrices reducing the number of free parameters. The model also provides an important methodology to determine latent groups in data. There are several pieces of research to extend the model based on the asymmetrical and/or with outlier datasets with some known computational limitations that have been examined in frequentist cases. In this paper, an MFA model with a rich and flexible class of skew normal (unrestricted) generalized hyperbolic (called SUNGH) distributions along with a Bayesian structure with several computational benefits have been introduced. The SUNGH family provides considerable flexibility to model skewness in different directions as well as allowing for heavy tailed data. There are several desirable properties in the structure of the SUNGH family, including, an analytically flexible density which leads to easing up the computation applied for the estimation of parameters. Considering factor analysis models, the SUNGH family also allows for skewness and heavy tails for both the error component and factor scores. In the present study, the advantages of using this family of distributions have been discussed and the suitable efficiency of the introduced MFA model using real data examples and simulation has been demonstrated.

Keywords: Bayesian framework, Mixture of factor analysis (MFA) model, MCMC method, Skew family, SUNGH distributions.

1. Introduction

Using the finite mixture (FM) model with factor analyzer (FA) models as components is a useful statistical technique that has wide usage to the identification and analysis of hidden or latent variables of high dimensional or multilevel data. The relationship between several variables can be described in an FA model by reducing the dimension of variables into some smaller latent factors. Such procedure is applied in different fields like psychometric testing to facilitate the analysis of high-dimensional data. The FM model can identify the existence of
subpopulations within a general population using latent variables’ models (see e.g., Lee and McLachlan1,2 and McLachlan and Peel3 with references therein). Both methods have been widely used across a range of fields, including social sciences, data mining, biology, medical sciences, epidemiological studies and artificial intelligence. According to Ghahramani and Hinton4 and Hinton et al.5, a combination of the FM model and FA model is usually called the mixture of factor analyzer (MFA) model. Although a Gaussian distribution is commonly assumed for these models, Wall et al.6 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 Maleki et al.7,8,9,10, Mahmoudi et al.11, Zarei et al.12 and Lin et al.13 and references therein, for further discussion and examples). There has been much research (especially in recent years) on asymmetrical families of distributions for a large range of uses, which some are presented as: the class of Skew-Normal (SN) distribution (Azzalini14 and Arellano-Valle and Azzalini15) and a rich class of scale mixtures of skew-normal (SMSN) family of distributions (Branco and Dey16). The SMSN family of distributions contain a range of distributions, including the skew-normal, skew slash skew-t, and skew contaminated-normal distributions. Lee and McLachlan17 focused on a class of skewed distributions such as the unrestricted Skew-t (SUT) distribution. The term ‘restricted’ relates to asymmetric behaviors which controlled by multiplying an ordinary skewing variable on convolution type representation to a vector of skewness parameters. In ‘unrestricted’ forms, in contrast of restricted forms which skewness is present in single direction, the skewness is free to present in more than one direction; see e.g. Lee and McLachlan2. Recently, some authors have considered multivariate skewed distributions for FA and MFA models. Kim et al.18 studied FA models based on a symmetric (Gaussian) error and skewed factors which follow the skew-normal, generalized skew-normal and skew-t. Lin et al.19 employed the multivariate restricted skew-normal (RSN) distribution on the MFA model, which is called the finite mixtures of skew-normal factor analyzer (MSNFA) model. In their approach, to model the data in the attending of asymmetric subpopulations, the factor component of the MSNFA model follows the family of rSN distributions (also the error component follows the symmetric Gaussian distribution).

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 Gaussian MFA model. Ando20 and Lee and Xia21 extend the Gaussian MFA model by considering a t-distribution (matrix-variate) for the factor (scores) and independent normal family for the errors, respectively. Yang and Dunson22, Lee and Xia23 and Song et al.24 concentrated on semi-parametric MFA model, and Chen et al.25 and Paisley and Carin26 focused on non-parametric approaches on the MFA model.

In comparison to ML estimations, using a Bayesian approach for MFA models has several other computational advantages, particularly in high dimensional settings for specification of
priors to regularize the parameter space (Carvalho et al.\textsuperscript{27}). Suarez and Ghosal\textsuperscript{28} considered principal components for functional data and used a Bayesian analysis with a prior distribution for the error term with the degree of informativeness. 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 \textit{MFA} model could vary, and as a result, they can be updated as the computational part (Maleki and Wraith\textsuperscript{29}).

More general cases of \textit{FA} models on structural equation modeling can be extended in an easier way than the classical likelihood inference (e.g., Lee and Xia\textsuperscript{21}). 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 at iterations of \textit{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 evaluation of derivatives for complex functions used for estimating parameters (Lee and McLachlan\textsuperscript{17}). The \textit{MFA} model in the basis of the skew-normal (unrestricted) generalized hyperbolic (called \textit{SUNGH}) family of distributions is considered in this work. A Bayesian technique is developed to estimate the parameters of model. The \textit{SUNGH} distributions are an extension of the unrestricted Skew normal distribution (\textit{SUN}; introduced by Arellano-Valle and Genton\textsuperscript{30}) by considering a generalized inverse Gaussian (\textit{GIG}) for scale mixer variable (Maleki et al.\textsuperscript{7,31} and Ghasami et al.\textsuperscript{32}). Besides the above-mentioned advantages of Bayesian methodology, another superiority of the work is that the \textit{SUNGH} family prepares a high level of flexibility and performance for usage in factor analysis models, i.e., some partitions of a \textit{SUNGH} random vector to uncorrelated homogeneous \textit{SUNGH} random vectors can provide asymmetry in both the factor and error components (such structure of the MFA model called skew errors and factors (SFE) has not been considered yet; see e.g. Lee and McLachlan\textsuperscript{33} and Maleki and Wraith\textsuperscript{29}). In fact, they considered using SFE models in the multivariate scale mixtures of skew normal (SMSN), where \textit{SNGH} and \textit{GH} could typically lead to identifiable problems. It is worth mentioning that our proposed model does not have such deficiency. Furthermore, to show the performance of the Bayesian approach in this \textit{MFA} setting, the missing data is considered and the applicability of the introduced \textit{MFA} model is evaluated. In Sections 2 and 3, preliminaries and notions of the \textit{SUN}, \textit{GIG} and \textit{SUNGH} distributions are examined. A Bayesian approach with Gibbs sampling algorithm for the \textit{SUNGH-MFA} model is presented in Section 4. In Section 5, the ability and suitability of the introduced \textit{MFA} model are evaluated using some simulated and real datasets. Finally, some conclusion along with discussion on possible extensions for further research has been presented in Section 6.

2. A review on unrestricted skew normal and GIG distributions
2.1. Preliminaries

A $p \times 1$ random vector $\mathbf{Z}$ with SUN distribution with $p \times 1$ vector of location $\boldsymbol{\mu}$, $p \times p$ dispersion matrix $\boldsymbol{\Sigma}$ (which is positive definite), and $p \times q$ skewness parameters matrix $\boldsymbol{\Lambda}$, denoted by $\mathbf{Z} \sim \text{SUN}_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$. The SUN distribution has the following probability density function (pdf):

$$f(\mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) = 2^q \phi_p(\mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\psi}) \Phi_q(\Lambda^T \psi^{-1}(\mathbf{z} - \boldsymbol{\mu})|\mathbf{Y}), \; \mathbf{z} \in \mathbb{R}^p,$$

where $\boldsymbol{\psi} = \boldsymbol{\Sigma} + \Lambda \Lambda^T$, $\mathbf{Y} = \mathbf{I}_q - \Lambda^T \psi^{-1} \Lambda = (\mathbf{I}_q + \Lambda^T \Sigma^{-1} \Lambda)^{-1}$, and $\phi_k(\cdot|\boldsymbol{\mu}, \boldsymbol{\psi})$ and $\Phi_q(\cdot|\mathbf{Y})$ are, respectively, the pdf and cdf of the $N_p(\boldsymbol{\mu}, \boldsymbol{\psi})$ and $N_q(\mathbf{0}, \mathbf{Y})$ distributions, respectively. The multivariate SUN distribution defined in (1) recover the multivariate normal distribution with zero skewness $\Lambda = \mathbf{0}$, the multivariate restricted skew-normal ($rSN$), with $q = 1$, and also the multivariate skew normal attributed to Sahu et al.$^{34}$, with $p = q$ and a diagonal skewness matrix $\Lambda$.

The random vector $\mathbf{Z} \sim \text{SUN}_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$ can be stochastically represented as follows:

$$\mathbf{Z} = \boldsymbol{\mu} + \Lambda |\mathbf{W}_0| + \Sigma^{1/2} \mathbf{W}_1,$$

where $\mathbf{W}_0 \sim N_q(\mathbf{0}, \mathbf{I}_q)$ and $\mathbf{W}_1 \sim N_p(\mathbf{0}, \mathbf{I}_p)$ are independent and $|\mathbf{W}_0|$ is the vector of the absolute values components of $\mathbf{W}_0$. Using (2) it is concluded that, the mean vector and covariance matrix of $\mathbf{Z}$ are given by $E[\mathbf{Z}] = \boldsymbol{\mu} + \sqrt{2/\pi} \Lambda 1_p$ and $\text{Var}[\mathbf{Z}] = \boldsymbol{\psi} - \frac{2}{\pi} \Lambda^T \Lambda q \mathbf{1}_q \mathbf{1}_q^T$, respectively, where $\mathbf{1}_q$ denotes the vector with lengths $q$ of ones.

The SUNGH random variable $\mathbf{Y}$ defined by Maleki et al.$^{7,31}$ has the scale mixtures of SUN distributions, and defined by

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

where $\mathbf{Z} \sim \text{SUN}_{p,q}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$, $\kappa(\cdot)$ is a positive function of scale mixer variable $\mathbf{U}$. Also note that the $\mathbf{U}$ is independent of $\mathbf{Z}$, and has a GIG distribution.

2.2. Some details of the GIG distribution

The GIG class of distributions has a positive support and so is a natural candidate for the scale mixer variable $\mathbf{U}$ in the stochastic representation (3), see Good$^{35}$ and Barndorff-Nielsen and Halgreen$^{36}$. Such choice leads to the construction of 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 that have adopted the following representation. A GIG random variable $\mathbf{U} \sim \text{GIG}_0(\nu, \psi, \eta)$ has the pdf given by
\[ \mathcal{GJG}_\kappa(u|\psi, \eta) = \frac{(u/\eta)^{v-1}}{2\eta K_v(\psi)} \exp \left( -\frac{\psi (u + \eta)}{2u} \right); \quad u \in \mathbb{R}^+, \] (4)

where \( K_t(y) \) is the third kind of order \( t \) modified “Bessel” function which is evaluated at point \( y \), and the parameter spaces are \( \psi > 0, \eta > 0 \) and \(-\infty < v < +\infty\). Note that, the \( s \)-th moment of the squared random variable \( U \) is given by

\[ r_s = E(U^{s/2}) = \frac{K_{v+s/2}(\psi)}{K_v(\psi)} \eta^{s/2}, \quad s = 1, 2, \ldots. \] (5)

Another representation of GIG distributions which have been used in the Bayesian framework (posteriors) is denoted by \( U \sim GIG^\ast(v, \gamma, \rho) \), if its pdf is given by

\[ \mathcal{GJG}^\ast(u|v, \gamma, \rho) = \left( \frac{\gamma}{\rho} \right)^v \frac{u^{v-1}}{2K_v(\rho \gamma)} \exp \left( -\frac{\gamma^2 u^2}{2u} \right); \quad u \in \mathbb{R}^+, \] (6)

where the parameter spaces are given by \( \gamma > 0, \rho > 0 \) and \(-\infty < v < +\infty\).

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

3. A review on \textit{SUNGH} distributions

According to previous discussions a \( p \times 1 \) random vector \( Y \) has a \textit{SUNGH} distribution if it has the following stochastic representation

\[ Y = \mu + \Lambda W + \kappa(U)^{1/2} \Sigma^{1/2} W_1, \] (7)

where \( \mu \) is a \( p \times 1 \) vector of location, \( \Sigma \) is a \( p \times p \) dispersion matrix, \( \Lambda \) is a \( p \times q \) skewness (shape) parameters matrix, \( W = \kappa^{1/2}(U)|W_0|, \quad W_0 \sim N_q(0, I_q), \quad W_1 \sim N_p(0, \rho) \) and \( U \sim GIG_\kappa(v, \psi, \eta) \), with \( W_0, \ W_1 \) and \( U \) being independent random quantities. Thus, since the conditional distribution of \( Y \) given \( U = u \) is given by

\[ Y|U = u \sim SUN_{p,q}(\mu, \kappa(u) \Sigma, \kappa(u)^{1/2} \Lambda), \] the pdf of \( Y \) is given by

\[ g(y|\mu, \Sigma, \Lambda, \omega) = 2^q \int_0^\infty \phi_p(y|\mu, \kappa(u) \psi) \Phi_q(\kappa(u)^{-1/2} \Lambda^\top \psi^{-1}(y - \mu)|Y) \mathcal{GJG}_\kappa(u|\omega) du, \] (8)

for \( y \in \mathbb{R}^p \), where \( \omega = (v, \psi, \eta)^\top \), \( \psi \) and \( Y \) was defined in (1), and we represent the \textit{SUNGH} distributed random vector \( Y \) as \( Y \sim SUNGH_{p,q}(\mu, \Sigma, \Lambda, \omega) \).
In order to have identifiability of the SUNGH distributions there exist some concerns relating to the $GIG$ parameters $\mathbf{\sigma}$ and skewness matrix $\mathbf{\Lambda}$. Note in the pdf (8), for any positive value $c$ the parameters $(\mu, \Sigma, \mathbf{\Lambda}, v, \psi, \eta)$ and $(\mu, c\Sigma, c\mathbf{\Lambda}, v, \psi/c, c\eta)$ yield have the same density, so to solve this identifiability issue set $\eta = 1$ and so $\mathbf{\sigma} = (v, \psi)^T$. Furthermore, sorting the skewness matrix $\mathbf{\Lambda}$ using a norm to its columns (or using a similar method to identify a factor loadings matrix by Bai and Li$^{37}$) is necessary to verify identifiability of the SUNGH distributions and the same is true for our proposed MFA model. Aying the distributions of $U$ from to the $GIG_*(\mathbf{\sigma})$ class leads to have different members of the SUNGH family (see Maleki et al.$^{31}$, for a brief outline of the different distributions possible). So considering the scale function $\kappa(u) = u$ and different distributions of $U$ within the $GIG_*(\mathbf{\sigma})$ class for which the SUNGH pdf in (8) becomes

$$g(y|\mu, \Sigma, \mathbf{\Lambda}, \phi) = 2^q \mathcal{G}_{\mathbf{\Sigma}}(y|\mu, \psi, 0, \nu') \mathcal{G}_{\mathbf{\Phi}}(B|0, Y, 0, \nu''), \quad y \in R^p,$$

(9)

where $\nu' = (v, \sqrt{\psi/\eta}, \sqrt{\psi/\eta})^T$, $\nu'' = (v - p/2, \sqrt{\psi/\eta}, q'(y))^T$, $q'(y)^2 = (y - \mu)^T\Sigma^{-1}(y - \mu) + \psi\eta$, $\mathbf{\Psi} = \Sigma + \mathbf{\Lambda}\mathbf{\Lambda}^T$, $\mathbf{Y} = I_q - \mathbf{\Lambda}\mathbf{\Lambda}^{-1}\mathbf{\Lambda}$ and $\mathbf{B} = \mathbf{\Lambda}\mathbf{\Lambda}^{-1}(y - \mu)$, $\mathcal{G}_{\mathbf{\Sigma}}$ and $\mathcal{G}_{\mathbf{\Phi}}$ denotes the $p$-variate pdf of the generalized hyperbolic (GH) distribution and $q$-variate cdf of the generalized hyperbolic distribution. Note when with the dimension $q = 1$, the SUNGH distributions are reduced to restricted cases which called the SNGH 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 SUNGH distributions are an extension of the maintained distributions and also in the case of symmetric $\mathbf{\Lambda} = 0$, it becomes the symmetrical generalized-hyperbolic (GH) distribution (Barndorff-Nielson$^{38}$). The SUNGH 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 SUNGH with zero means (see e.g., (10) and (11)). For the FA and also the MFA models based on the multivariate SMSN and restricted SN distributions, in order to have the identifiability of the model, the error term of the model or the factor scores are constrained to be symmetrically distributed (see e.g., Kim et al.$^{18}$ and Lin et al.$^{19}$).

4. Mixture of SUNGH factor analyzer model

4.1. SUNGH factor analysis (SUNGH-FA) model

In this part, a factor analysis model using the SUNGH distributions is developed. Specifically, the SUNGH-FA given by
\[ Y_j = \mu + LF_j + \epsilon_j; \quad F_j \perp \epsilon_j, \quad iid \sim \]

\[
\begin{cases}
F_j \overset{iid}{\sim} SUNGH_{m,q}(\Delta^{-1/2}\mu, \Delta^{-1}, \Delta^{-1/2}\Lambda, \omega), \quad j = 1, \ldots, n, \\
\epsilon_j \overset{iid}{\sim} SUNGH_{p,p}(\mu, \Lambda, \omega),
\end{cases}
\]

is first considered, where \( Y_j \) and \( \mu \) are \( p \)-dimensional random samples and vector of location, respectively, \( L \) is a \( p \times m \) matrix of factor loadings, \( F_j \) is a (symmetrical/asymmetrical and light/heavy-tailed) \( m \)-dimensional \( (m < p) \) vector of factors (latent variable) and \( \epsilon_j \) is a \( p \)-dimensional vector of (symmetrical and light/heavy-tailed) errors, \( \mu = \tau \Lambda_{1} \mu, \mu = \tau \Lambda_{q} \mu, \tau = -\sqrt{2/\pi} r_1, \Delta = \tau_2 I_m + \Lambda_1 C_q \Lambda_1^T \) and, \( \Delta^{-1} \) and diagonal matrix \( D = \text{diag}(D_1, \ldots, D_p) \), are positive definite dispersion matrices with dimensions \( m \times m \) and \( p \times p \), respectively, \( \Lambda_f \) is the \( m \times q \) skewness matrix and \( \Lambda_e = \text{diag}(\Lambda_e) \) is the \( p \times p \) diagonal skewness matrix (see Remark 1). Note that \( E[F_j] = 0, \text{Cov}[F_j] = I_m, E[\epsilon_j] = 0, \Omega = \text{Cov}[\epsilon_j] = r_2 D + \Lambda_e C_p \Lambda_e^T \) and also, \( E[Y_j] = \mu, \text{Cov}[Y_j] = LL^T + D \) (such obtaining the covariance matrices, is a main fact in the FA model), in which \( \Omega \) and \( D \) are diagonal matrices. Due to the SUNGH properties, we have

\[
\left(\begin{array}{c}
F_j \\
\epsilon_j
\end{array}\right) \overset{iid}{\sim} SUNGH_{m+p,q+p}(\begin{array}{c}
\mu_f \\
\mu_e
\end{array}, \Delta^{-1/2} \begin{array}{c}
\Delta^{-1} 0_{m \times p} \\
0_{p \times m} D
\end{array}, \Delta^{-1/2} \begin{array}{c}
\Lambda_f \\
\Lambda_e
\end{array}, \omega), \quad j = 1, \ldots, n.
\]

This model called \( SUNGH-FA \), contains the skew normal factor analysis (SNFA) due to Lin et al. (2016) when \( q = 1 \) (restricted case) and latent variable \( U \) degenerated at one. According the SUNGH properties, we have that

\[
Y_j \overset{ind.}{\sim} SUNGH_{p,k}(\mu + \tau \alpha 1_k, \Sigma, \alpha, \omega), \quad j = 1, \ldots, n,
\]

where \( k = p + q, \Sigma = \tilde{L} \tilde{L}^T + D \) and \( \alpha = (\tilde{L} \Lambda_1, \Lambda_e)_{p \times k'} \) for which \( \tilde{L} = L \Delta^{-1/2} \). Thus, the model parameters we need to estimate are denoted by \( \Theta = (\mu, L, D, \Lambda_1, \Lambda_e, \omega) \).

Considering the distribution (12), the corresponding pdf is thus given by

\[
g(y_j|\Theta) = 2^k \int_0^\infty \phi_p(y_j|\mu + \tau \alpha 1_k, \kappa(u)\psi) \Phi_k(k(u)^{-1/2}\alpha^T \psi^{-1}(y_j - \mu - \tau \alpha 1_k)|\gamma) g_{\gamma}(u|\omega) du, \]

where \( \psi = \Sigma + \alpha \alpha^T, Y = I_k - \alpha^T \psi^{-1} \).

Also, in this particular case \( \kappa(u) = u \), the above pdf of \( Y_j \) is given by
\[ g(y_j|\Theta) = 2^k G_{H_p}(y_j|\mu + \tau \alpha_1 k, \psi, 0, v') G_{H_k}(\alpha^T \psi^{-1} (y_j - \mu - \tau \alpha_1) | 0, Y, 0, v'), \]  

where \( v' = (v, \sqrt{\psi}, \sqrt{\psi})^T \), \( v'' = (v - p/2, \sqrt{\psi}, q'(y_j))^T \), \( q'(y_j)^2 = (y_j - \mu - \tau \alpha_1)^T \psi^{-1} (y_j - \mu - \tau \alpha_1) + \psi \), and, \( \psi \) and \( Y \) defined as above.

Finally, it is worth to note that according to statistical features of the SUNGH family, and the structure of SUNGH-FA model (10), the conditional distribution of response vector \( Y_j \) given the latent factor \( F_j \) is given by

\[ Y_j | F_j \sim \text{SUNGH}_{p,p}(\mu + LF_j, D, \Lambda_e, \omega), \quad j = 1, \ldots, n. \]  

To guarantee the identifiability of this FA model, we consider the technique in Lin et al.\cite{19}, Lopes and West\cite{39} and Fokoué and Titterington\cite{40}, and assume that the diagonal entries of the loading matrix \( L \) are strictly positive and its upper-right triangle is zero (constraining the loading matrix \( L \)). Other approaches are also possible, see e.g., Bai and Li\cite{37}.

### 4.2. Mixture of SUNGH factor analysis (SUNGH-MFA) model

A mixture of SUNGH factor analysis (SUNGH-MFA) model, which is a generalization of the proposed SUNGH-FA, is considered in this section. Let \( Y_j = (Y_{j1}, \ldots, Y_{jp}) \); \( j = 1, \ldots, n \) be \( p \)-dimensional vector of responses raising from heterogeneous population with a finite number of groups. We utilize (latent) membership-indicator variables \( Z_1, \ldots, Z_n \), to assign vector variables for belonging to different components of the mixture factor model \( (i = 1, \ldots, g) \), so the notation \( (Z_j = i) \) denotes that the \( j \)-th vector variable belongs to the \( i \)-th SUNGH-FA model. The allocation random variables \( Z_1, \ldots, Z_n \) follow a pmf in the form of \( P(Z_j = i) = \pi_i; j = 1, \ldots, n, i = 1, \ldots, g \) such that \( \pi_i > 0 \) and \( \sum_{i=1}^{g} \pi_i = 1 \). In terms of \( Z_j \), we can conclude that for \( i = 1, \ldots, g \), each component of this model follows the SUNGH-FA model (10) represented as follows:

\[
\begin{align*}
Y_j &= \mu_i + L_i F_{ij} + \epsilon_{ij}; \quad F_{ij} \perp \epsilon_{ij}, \quad \text{with probability } \pi_i, \quad \text{iid} \\
F_{ij} &\sim \text{SUNGH}_{m,q}(\Delta_i^{-1/2} \mu_{fi}, \Delta_i^{-1}, \Delta_i^{-1/2} \Lambda_{fi}, \omega_i), \quad ; \quad j = 1, \ldots, n, i = 1, \ldots, g, \\
\epsilon_{ij} &\sim \text{SUNGH}_{p,p}(\mu_{ei}, D_i, \Lambda_{ei}, \omega_i),
\end{align*}
\]

where \( \Delta_i = r_i^2 I_m + \Lambda_q C_q \Lambda_i^T \) and \( \Delta_i^{-1} \) and diagonal matrix \( D_i = \text{diag}(D_{i1}, \ldots, D_{ip}) \) are positive definite dispersion matrix, \( \mu_{fi} = \tau_i \Lambda_{fi} 1_q, \mu_{ei} = \tau_i \Lambda_{ei} 1_p, \Lambda_{ei} = \text{diag}(\lambda_{ei}) \) and \( \omega_i = (v_i, \psi_i)^T \), \( i = 1, \ldots, g \).
Due to (13) and (16), the marginal density of $Y_j$ is
\[
g(y_j | \mathcal{Z}) = \sum_{i=1}^{g} \pi_{i} g(y_j | \Theta_{i}) \quad j = 1, \ldots, n, \tag{17}
\]
where $g(y_j | \Theta_{i})$ is the pdf as defined in (14) for each component and $\Theta_{i} = (\mu_i, L_i, D_i, \Lambda_{li}, \Lambda_{el}, \omega_{i})$, for which $\mathcal{Z} = (\pi_{1}, \ldots, \pi_{g-1}, \Theta_{1}, \ldots, \Theta_{g})$.

Let $C = \{Y, U, W, Z\}$ denote the complete data, $\bar{L}_i = L_i \Delta_{i}^{-1/2}$ and $\bar{F}_{ij} = \Delta_{i}^{1/2} F_{ij}$, in accordance with the stochastic representation of the SUNGH distribution (7) to the conditional distribution given in (15), the hierarchical representation of the SUNGH-MFA model is as follows:

\[
Y_j | \bar{F}_{ij}, W_{eij} = w_{eij}, U_{ij} = u_{ij}, Z_j = i \sim \text{ind. } N_p(\mu_i + \bar{L}_i \bar{F}_{ij} + \mu_{ei} + \Lambda_{el}w_{eij}, \kappa(u_{ij})D_i),
\]
\[
\bar{F}_{ij} | W_{fij} = w_{fij}, U_{ij} = u_{ij}, Z_j = i \sim \text{ind. } N_m(\mu_{fi} + \Lambda_{fi}w_{fij}, \kappa(u_{ij})I_m),
\]
\[
W_{eij} | U_{ij} = u_{ij}, Z_j = i \sim \text{ind. } TN_p(0, \kappa(u_{ij})I_p; W_{eij} > 0),
\]
\[
W_{fij} | U_{ij} = u_{ij}, Z_j = i \sim \text{ind. } TN_q(0, \kappa(u_{ij})I_q; W_{fij} > 0),
\]
\[
U_{ij} | Z_j = i \sim \text{ind. } \text{IG}_{\omega_i}(\omega_i),
\]
\[
P(Z_j = i) = \pi_i,
\]
where $TN_k(\mathbf{m}, \mathbf{M}; \mathbf{W} > \mathbf{c})$ denotes the $k$-variate truncated normal on the region $\mathbf{W} > \mathbf{c}$ with mean $\mathbf{m}$ and covariance $\mathbf{M}$ before truncation, with pdf denoted by $T\phi_k(\mathbf{w} | \mathbf{m}, \mathbf{M}; \mathbf{c})$. Note that the second and third hierarchical representation can be reformulated as follows:

\[
Y_j | \bar{F}_{ij}, W_{eij} = w_{eij}, U_{ij} = u_{ij}, Z_j = i \sim \text{ind. } N_p(\mu_i + \bar{L}_i \bar{F}_{ij} + \Lambda_{el}w_{eij}, \kappa(u_{ij})D_i),
\]
\[
\bar{F}_{ij} | W_{fij} = w_{fij}, U_{ij} = u_{ij}, Z_j = i \sim \text{ind. } N_m(\Lambda_{fi}w_{fij}, \kappa(u_{ij})I_m),
\]
\[
W_{eij} | U_{ij} = u_{ij}, Z_j = i \sim \text{ind. } TN_p(\tau_1 1_p, \kappa(u_{ij})I_p; W_{eij} > \tau_1 1_p),
\]
\[
W_{fij} | U_{ij} = u_{ij}, Z_j = i \sim \text{ind. } TN_q(\tau_1 1_q, \kappa(u_{ij})I_q; W_{fij} > \tau_1 1_q),
\]
\[
U_{ij} | Z_j = i \sim \text{ind. } \text{IG}_{\omega_i}(\omega_i),
\]
\[
P(Z_j = i) = \pi_i,
\]
In accordance with the above hierarchical representations, the complete augmented likelihood function of $\mathbf{Z}$ is given by

\[
L(\mathbf{Z}|\mathbf{C}) = \prod_{j=1}^{n} \prod_{i=1}^{g} \left[ \pi_i \phi_p(y_j|\mu_i + \tilde{L}_i \tilde{F}_{ij} + \Lambda_e \omega_{eij}, \kappa(u_{ij}) D_i) \phi_m(\tilde{F}_{ij}|\Lambda_f \omega_{fij}, \kappa(u_{ij}) I_m) \right] \\
\times T \phi_p(\omega_{eij}|\tau_i 1_p, \kappa(u_{ij}) I_p; \tau_i 1_p) T \phi_q(\omega_{fij}|\tau_i 1_q, \kappa(u_{ij}) I_q; \tau_i 1_q) G \tilde{G}_s(u_{ij}|\tilde{\omega}_i) P(Z_j = i). \tag{18}
\]

4.3. Bayesian Approach on the SUNGH-MFA model

In this section, we develop a Bayesian framework to estimate the parameters of the SUNGH-MFA models. First, we consider prior distributions to the model parameters. Also weakly informative (proper) and independent priors for the elements of $\mathbf{Z}$ have been adopted. Also, we re-presented the factor loading matrix $\tilde{L}_i = [\ell_{i,ss}]$ ($\ell_{i,ss}$ are $\tilde{L}_i$ elements). So, the following priors are considered

\[
\pi = (\pi_1, ..., \pi_g) \sim \text{Dir}(\delta_1, ..., \delta_g), \quad \mu_i \sim N_p(\theta_i, S_i), \\
\ell_{i,ss} \sim N_1(\mu_{\ell_i}, \sigma_{\ell_i}^2); s > r, \quad \ell_{i,rr} \sim T N_1(\mu_{\ell_i}, \sigma_{\ell_i}^2; \ell_{i,rr} > 0), \\
D_{i,s} \sim IG(q_i, \xi_i), \quad \lambda_{e} \sim N_p(\alpha_i, A_i), \quad \Lambda_f \sim \text{MN}_{m,q}(C_{fi}, H_{fi}, N_{fi}), \quad \tau_i \sim T N_1(\mu_{\tau_i}, \sigma_{\tau_i}^2; \tau_i < 0),
\]

for $i = 1, ..., g$, $s = 1, ..., p$, $r = 1, ..., m$ and $t = 1, ..., q$, where $MN$ is the Matrix-Normal distributions, and for the components of $\tilde{\omega}_i = (v_i, \psi_i)^T$, $i = 1, ..., g$, we consider that $v_i \sim N(\mu_i, \sigma_i^2)$ and $\psi_i \sim E(\zeta_i)$ for $i = 1, ..., g$, for which $E(\zeta_i)$ is the exponential distribution with rate parameter $\zeta_i$. The notations $\text{Dir}$ and $IG$, denote the “Dirichlet” and “Inverse Gamma” distributions, respectively. The maintained priors are assumed to be independent. The posterior $\pi(\mathbf{Z}, \tilde{F}, \mathbf{u}, \mathbf{w}, \mathbf{z}|\mathbf{y}) \propto L(\mathbf{Z}|\mathbf{y}, \tilde{F}, \mathbf{u}, \mathbf{w}, \mathbf{z}) \pi(\mathbf{Z})$, is not analytically tractable but we employ an MCMC methods such as Gibbs sampling (Gelfand and Smith$^{41}$) and Metropolis-Hastings algorithms attributed to Gamerman$^{42}$ to draw samples by using the following conditional posteriors. Except the derived parameters of the scale mixer variable, all of the other conditional posteriors are in the closed form as follows: (Note that $\mathbf{Z}_{(-m)}$ is the set of all parameters without the particular parameter $m$, $B_i = \{ j : z_j = i \}$ and $n_i$ is number of observations which devoted to the $i$-th SUNGH-FA component)

\[
\pi_{(-\pi)}|\mathbf{y}, \tilde{F}, \mathbf{w}, \mathbf{z} \sim \text{Dir}(\delta_{p,1}, ..., \delta_{p,g}),
\]

\[
\delta_{p,i} = \delta_i + n_i; \quad i = 1, ..., g, \tag{19}
\]

\[
\mu_i|\mathbf{Z}_{(-m)}, \mathbf{y}, \tilde{F}, \mathbf{w}, \mathbf{z} = i \sim N_p(\theta_{ip}, S_{ip}), \quad i = 1, ..., g. \tag{20}
\]
where \( S_{lp} = (S_l^{-1} + \sum B_l \kappa^{-1}(u_{ij})D_l^{-1})^{-1} \) and
\[
\theta_{lp} = S_{lp} \left[ S_l^{-1} \theta_l + \sum B_l \kappa^{-1}(u_{ij})D_l^{-1}(y_j - \bar{L}_i \bar{F}_{ij} - \Lambda_{el}w_{ej}) \right].
\]

\[
\ell_{lsr} | \mathbf{z}_{(-\ell_{lsr})}, y, \bar{F}, \mathbf{w}, u, z_j = i \sim N_1(\mu_{lp}, \sigma_{lp}^2), \tag{21}
\]

where \( \sigma_{lp}^2 = (\sigma_{\ell_l}^{-2} + \sum B_l D_{ls}^{-1} \kappa^{-1}(u_{ij})\bar{F}_{ij,r}^2)^{-1} \)
and \( \mu_{lp} = \sigma_{lp}^2 \left[ \mu_{\ell_l} \sigma_{\ell_l}^{-2} + \sum B_l \kappa^{-1}(u_{ij})\bar{F}_{ij,r}(y_{js} - \mu_{ls} - \ell_{ls(r)} \bar{F}_{ij} - \lambda_{el} w_{ej}) \right], \)
for \( i = 1, ..., g; j = 1, ..., n; s = 1, ..., p; r = 1, ..., m, \) for which \( y_{js}, \mu_{ls}, \lambda_{el} \) and \( w_{ej} \) be the \( s \)-th components of \( y_j, \mu, \lambda_{el} \) and \( w_{ej} \), respectively, \( \ell_{ls} \) be the \( s \)-th row of \( L_l \), and \( \ell_{ls(r)} \) be the \( \ell_{ls} \) with zero \( r \)-th component, and \( \bar{F}_{ij,r} \) be the \( r \)-th component of \( \bar{F}_{ij} \).

Also \( \ell_{lrr} | \mathbf{z}_{(-\ell_{lrr})}, y, \bar{F}, \mathbf{w}, u, z_j = i \sim TN_1(\mu_{lp}, \sigma_{lp}^2; \ell_{lrr} > 0) \), with above parameters for which \( s = r. \)

\[
\bar{F}_{ij} | \mathbf{z}, y, \mathbf{w}, u, z_j = i \sim N_m(\mu_{Fi}, \Sigma_{Fi}), \quad i = 1, ..., g; j = 1, ..., n, \tag{22}
\]

where \( \Sigma_{Fi} = \kappa(u_{ij})(I_m + \bar{L}_i^T D_i^{-1} \bar{L}_i)^{-1} \)
and \( \mu_{Fi} = \kappa^{-1}(u_{ij})\Sigma_{Fi}[A_{fi}w_{fi} + \bar{L}_i^T D_i^{-1}(y_j - \mu_i - \Lambda_{el} w_{ej})] \).

\[
D_{ls} | \mathbf{z}_{(-D_{ls})}, y, \bar{F}, \mathbf{w}, u, z_j = i \sim IG(q_{lp}, \xi_{lp}), \quad i = 1, ..., g; \ s = 1, ..., p, \tag{23}
\]

where \( q_{lp} = q_l + n_l/2 \) and \( \xi_{lp} = \xi_l + \frac{1}{2} \sum B_l \kappa^{-1}(u_{ij})(y_{js} - \mu_{ls} - \ell_{ls}^T \bar{F}_{ij} - \lambda_{el} w_{ej})^2 \)
with previous notations in (21).

\[
\lambda_{el} | \mathbf{z}_{(-\lambda_{el})}, y, \bar{F}, \mathbf{w}, u, z_j = i \sim N_p(a_{lp}, A_{lp}), \quad i = 1, ..., g, \tag{24}
\]

where \( P_{ij} = \text{diag}(w_{ej}), A_{lp} = (A_i^{-1} + \sum B_l \kappa^{-1}(u_{ij})P_{lj}D_l^{-1}P_{lj})^{-1} \)
and \( a_{lp} = A_{lp}[A_i^{-1}a_i + \sum B_l \kappa^{-1}(u_{ij})P_{lj}D_l^{-1}(y_j - \mu_i - \bar{L}_i \bar{F}_{ij})]. \)

\[
\text{vec}(A_{fi}) | \mathbf{z}_{(-A_{fi})}, y, \bar{F}, \mathbf{w}, u, z_j = i \sim N_m(a_{fi}, A_{fi}), \quad i = 1, ..., g, \tag{25}
\]

where \( \mu_{fi} = \Sigma_{fi} \left[ (N_{fi}^{-1} \otimes H_{fi}^{-1}) \text{vec}(C_{fi}) + \sum B_l \kappa(u_{ij})^{-1}(M_{fi} \otimes I_m) \right] \)
and
\[ \Sigma_{fi,p} = \left( N^-1_{fi} \otimes H^{-1}_{fi} + \sum_{i} \kappa^{-1}(u_{ij})(R_{fi} \otimes I_m) \right)^{-1}, \]
for which \( R_{fi} = w_{fi}w_{fi}^T \) and \( M_{fi} = \bar{F}_{ij}w_{fi}^T \), and the Kronecker product is denoted by \( \otimes \) and a matrix vectorization is denoted "vec" (convert a matrix into a column vector).

\[
W_{fi}|\Xi, y, \bar{F}, u, z_j = i \sim TN_q(\mu_{Wfi}, \Sigma_{Wfi}; W_{fi} > \tau_i1_q), \quad i = 1, ..., g; j = 1, ..., n, \tag{26}
\]
where \( \Sigma_{Wfi} = \kappa(u_{ij})(I_q + \Lambda^T_{fi}\Lambda_{fi})^{-1} \) and \( \mu_{Wfi} = \kappa^{-1}(u_{ij})\Sigma_{Wfi}(\tau_i1_q + \Lambda^T_{fi}\bar{F}_{ij}) \).

\[
W_{ej}|\Xi, y, \bar{F}, u, z_j = i \sim TN_p(\mu_{Wej}, \Sigma_{Wej}; W_{ej} > \tau_i1_p), \quad i = 1, ..., g; j = 1, ..., n, \tag{27}
\]
where \( \Sigma_{Wej} = \kappa(u_{ij})(I_p + \Lambda^T_{ej}D_{ij}^{-1}\Lambda_{ej})^{-1} \) and
\[
\mu_{Wej} = \kappa^{-1}(u_{ij})\Sigma_{Wej}(\tau_i1_p + \Lambda^T_{ej}D_{ij}^{-1}(y_j - \mu_i - \bar{L}_i\bar{F}_{ij})).
\]

The conditional posterior pmf of \( Z_j, j = 1, ..., n \) is
\[
\pi(Z_j = i|\Xi, y, \bar{F}, u, w) = \frac{\pi_i g(y_j|\Theta_i)}{\sum_{h=1}^{g} \pi_h g(y_h|\Theta_h)}, \quad j = 1, ..., n; \quad i = 1, ..., g, \tag{28}
\]
where \( g(y_j|\Theta_i) \) is the pdf defined in (14).

The conditional posterior of scale mixer variables \( U_{ij}, j = 1, ..., n; \quad i = 1, ..., g, \) are given by:
\[
U_{ij}|\Xi, y, \bar{F}, w, z_j = i \sim GLG^*(a_{ij}, b_{ij}, \sqrt{c_{ij}}), \quad j = 1, ..., n; \quad i = 1, ..., g, \tag{29}
\]
where \( \kappa(u) = u, \quad a_{ij} = u_i - (2p + m + q)/2, \quad b_{ij} = \sqrt{\psi_i} \) and \( c_{ij} = \psi_i + (y_j - \mu_i - \bar{L}_i\bar{F}_{ij} - \Lambda_{el}w_{ej})^T \Lambda_{el}^{-1}(y_j - \mu_i - \bar{L}_i\bar{F}_{ij} - \Lambda_{el}w_{ej}) + (\bar{F}_{ij} - \Lambda_{fi}w_{fi})^T(\bar{F}_{ij} - \Lambda_{fi}w_{fi}) + (W_{ej} - \tau_i1_p)^T(W_{ej} - \tau_i1_p) + (W_{fi} - \tau_i1_q)^T(W_{fi} - \tau_i1_q).
\]

Finally, the conditional posteriors of the \( v_i, \psi_i; \ i = 1, ..., g \) (parameters of the scale mixer variables), are also proportional to:
\[
\pi(u_i|\theta_{(-u_i)}, y, \bar{F}, u, w, z_j = i) \sim \pi_1(u_i)\phi(u_i|\mu_i + \alpha_i^2 \sum_{B_i} \log(u_i), \sigma_i^2),
\]
(30)

where \(\pi_1(u_i) = \left(K_{u_i}(\psi_i)\right)^{-n_i};\)

\[
\pi(\psi_i|\zeta_{(-\psi_i)}, y, \bar{F}, u, w, z_j = i) \sim \pi_2(\psi_i) \times E\left(c_i + \sum_{B_i} (u_{ij} + u_{ij}^{-1}) \right) / 2,
\]
(31)

where \(\pi_2(\psi_i) = \left(K_{u_i}(\psi_i)\right)^{-n_i}\) and \(E(\delta)\) denotes the pdf of exponential distribution with rate parameter \(\delta\). Note that the posteriors in (30) and (31) are not in the closed forms but an MCMC scheme such as Metropolis-Hastings algorithm can be embedded in the to draw samples.

4.3. Assigning of missing values

The usefulness of the SUNGH-MFA hierarchical representation is its simplicity for simulation or its usage in sampling from the posteriors using available softwares in Bayesian analysis such as NIMBLE (NIMBLE Development Team\(^{43}\)). Note that some of the other Bayesian softwares such as JAGS and OpenBUGS are not able to inverse the matrices because of the lack of special functions, and so can’t 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 \(Y_M\) as the missing and \(Y_O\) as the observed responses of SUNGH-MFA model, respectively. Posterior predictive distribution can help to assign the missing data in a Bayesian framework, \(P(Y_M|Y_O) = \int P(Y_M|Y_O, \Theta) P(\Theta | Y_O) d\Theta\). 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: \(y_{i,M}^{(t+1)} \sim P(y_{i,M}|y_O, \Theta^{(t)})\) for \(i = 1, ..., N\) and \(\Theta^{(t+1)} \sim P(\Theta^{(t)}|y_O, y_{i,M})\). Starting with reasonable initial values \(y_{i,M}^{(0)}\) and \(\Theta^{(0)}\) and running the Gibbs algorithm with large number of iterations lead to convergence. 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 SUNGH-MFA model using simulated and real datasets.

5.1. Priors and details of computation
Largely non-informative priors have been used for estimation with different models as follows: 
\[ \mu \sim N_2(0, M_i), \text{where } M_i = 10^2 I_2 \text{ priors of its columns as } \lambda_i \sim N_2(J_i, G_i), \text{ where } J_i = 0 \text{ and } G_i = 10^2 I_2, \quad \ell_{lrr} \sim N_i(0, 100); \ r > t, \quad \ell_{lrr} \sim HN_i(0, 100)(\ell_{lrr} > 0) \text{ and } D_{lrr} \sim IG(1, 1) \text{ for } i = 1, 2 \text{ and } \pi \sim Dir(1, ..., 1). \] 
Note that 45,000 iterations have been used for Gibbs sampling runs with burn-in of 11,000. Also, the statistic attributed to Gelman and Rubin\(^4^4\) 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 posteriori estimate (MAP) has been used. To prevent some common computational issues in the factor analysis, the scale function in R to scale the 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; Hubert and Arabie\(^4^5\)) that ranges from zero (no match situation) to one (perfect match situation) is used. For model selection criteria, the EAIC and EBIC have been reported that their lower values indicate a better fitted model (Carlin and Louis\(^4^6\)). DIC values suggested by Celeux et al.\(^4^7\) can also be employed in a mixture to compare the fitted models.

5.2. Simulations

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

Dataset has been first simulated from the SUNGH-MFA model (2) where: \( \Lambda_{fi(1,1)} \sim U(5,10), \Lambda_{fi(1,2)} \sim U(0,5), \Lambda_{fi(2,2)} \sim U(-5,0), \Lambda_{fi(2,1)} \equiv U(-10,-5); \ D_i = \text{diag}(0.5); \mu_1 = 0, \mu_2 = 4; \Lambda_{ei} = 0; \ \sigma = (-0.5,1,1)^T; \ p = 6, m = 3 \) and \( q = 2. \)

\[
L_i = \begin{pmatrix}
0.80 & 0.00 & 0.00 \\
0.80 & 0.00 & 0.00 \\
0.60 & 0.00 & 0.00 \\
0.00 & 0.73 & 0.00 \\
0.00 & 0.78 & 0.00 \\
0.00 & 0.60 & 0.00 \\
0.00 & 0.00 & 0.85 \\
0.00 & 0.00 & 0.81 \\
0.00 & 0.00 & 0.58 \\
\end{pmatrix}
\]

where \( U \) denotes a uniform distribution and mixture weights \( (\pi_1 = 0.333, \pi_2 = 0.666) \), corresponding to \( n_1 = 150 \) and \( n_2 = 300. \)

Figures 1 and 2 show the pairwise scatterplot for the simulated data and their first three factors, respectively. As expected, the results in Table 1, for different factor analysis models indicate a strong preference based on the model selection criteria for the SUNGH-MFA used as a
basis for the simulations ($q = 2, m = 3$). Due to the heavy tailed-ness of the factor scores, the SNGH-MFA model performs similarly to the SUNGH-MFA based on the classification score for a smaller number of factors ($m = 2, ARI=0.93$) but substantially worse in terms of other model selection measures. In contrast, the SN-MFA model performs the worst, most likely due to the inability to allow for the long tails in the distribution of the true factor scores.

Figure 1: Pairwise scatterplot for the simulated data ($y$) with colors indicating component labels (Black = Group 1, Red = Group 2).
Figure 2: Pairwise scatterplot for the factor scores ($F_{ij}$) (right) with colors indicating component labels (Black = Group 1, Red = Group 2).

Table 1: Model selection criteria for different models

| Model      | $m$ | $m_{par.}$ | EBIC     | EAIC     | ARI | DIC2   |
|------------|-----|------------|----------|----------|-----|--------|
| SN-MFA     | 2   | 65         | 1172.7   | 921.9    | 0.71 | 848.9  |
|            | 3   | 73         | 1064.2   | 782.6    | 0.82 | 688.0  |
|            | 4   | 79         | 990.1    | 685.3    | 0.85 | 571.1  |
| SNGH-MFA   | 2   | 67         | 860.4    | 601.9    | 0.93 | 509.9  |
|            | 3   | 75         | 957.9    | 668.6    | 0.87 | 565.8  |
|            | 4   | 81         | 1179.4   | 866.9    | 0.69 | 776.4  |
| SUNGH-MFA  | 2   | 79         | 972.5    | 667.7    | 0.93 | 653.8  |
| (q=2)      | 3   | 87         | 703.8    | 368.2    | 0.94 | 248.6  |
|            | 4   | 93         | 727.4    | 368.6    | 0.91 | 221.9  |

Note: the bold values are corresponding to the best models; SN-FA, SNGH-MFA and SUNGH-MFA denote the skew-normal, skew-normal generalized hyperbolic (SUNGH with q=1) and SUNGH (q>1) mixture of factor analyzers, respectively.

5.3. Hawks data

The Hawks dataset (Stat2Data R package\textsuperscript{48}) includes the data of three hawks species of red-tailed, sharp-shinned, and Copper’s. In total, there are 19 variables in the dataset and 908 valid observations. The concentration has been on the five continuous variables, consisting of length of primary wing feather, the hallux length, the culmen length, the tail length (all of them
based on mm) and the weight (g) of the bird. As Lee et al. 49 discussed, in a number of variables such as Hallux length in the dataset, signs of moderate skewness and kurtosis can be seen. Some studies have used this dataset and fitted and evaluated the ability of skew-t (heavy-tailed) and skew-normal (light-tailed) MFA models to classify the three species. For example, Lee et al. 49 used the fundamental skew normal distributions (which are light-tailed) for the error term, while we choose the SUNGH distributions (which are heavy-tailed) on both error terms and factors that is suitable for robust inferences of the proposed models with outliers. For this dataset, we are interested to assess the ability of the SUNGH-MFA model to classify the three species. It is clear from Table 2 that the classification performance is similar for all models with the model choice criteria appearing to favor the SUNGH-MFA. For the SN-MFA model, there is a model choice preference for factors which are likely to reflect the presence of outliers in the data which is not explicitly accommodated in this model. As Lee et al. 49 discussed, the data is quite skewed and in different directions due possibly to the presence of some signs of truncation in the data (see observations for Hallux). As a result, the added flexibility and ability to accommodate different directions of skewness in the SUNGH-MFA is preferred to the SNGH-MFA in terms of model choice criteria.

Table 2: Model selection criteria for Hawks data

| Model       | m | par. | EBIC  | EAIC  | ARI  | DIC  |
|-------------|---|------|-------|-------|------|------|
| SN-MFA      | 1 | 67   | 2831.4| 2510.3| 0.87 | 2431.9|
|             | 2 | 79   | 1003.9| 625.4 | 0.89 | 564.9 |
| SNGH-MFA    | 1 | 69   | -218.8| -549.5| 0.87 | -561.9|
|             | 2 | 81   | 711.2 | 322.9 | 0.87 | 246.1 |
| SUNGH-MFA (q=2) | 1 | 84   | **-389.5**| **-792.1**| 0.87 | **-815.9**|
|             | 2 | 96   | -187.7| -647.7| **0.89**| -782.2|
|             | 3 | 105  | 480.1 | -23.2 | 0.82 | -1.50 |

Note: the bold values are corresponding to the best models; SN-FA, SNGH-MFA and SUNGH-MFA denote the skew-normal, skew-normal generalized hyperbolic (q=1) and unrestricted skew-normal generalized hyperbolic (q>1) mixture of factor analyzers respectively.

To see the advantage of the proposed Bayesian technique, an experiment on the dataset of Hawks was performed through evaluating the applicability of the SUNGH-MFA classification, and also related errors concerning missing dataset context. The hierarchical form of the SUNGH-MFA allows us to code and perform the computation of 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 dataset in the form of randomness under two level of low missingness have been deleted (5%) and moderate level of missingness (15%) of the total sample is obtained. Then, the performance of assigning values
has been compared using the model based on the conditional approach, 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:

\[
MSE = \frac{1}{n} \sum_{j=1}^{n} (y_j^m - \hat{y}_j^m)\top (y_j^m - \hat{y}_j^m),
\]

where \( \hat{y} \) denotes the imputed value and the number of total missing values is \( n^* = \Sigma_{j=1}^{n} (p - p^j_o) \).

Table 3 reports the means and standard deviations (in parenthesis) of different model selection criteria for the unconditional (UC) and conditional (C) models based on 30 replications in the dataset for two different missingness rate scenarios (5% or 15%).

Table 3: Model selection criteria for Hawks data (missing data)

| Model               | Missing Rate (%) | EBIC     | EAIC     | MSE      | ARI      | DIC2    |
|---------------------|------------------|----------|----------|----------|----------|---------|
| SUNGH-MFA (UC)      | 5                | 1175.4 (123.3) | 7728 (123.3) | 0.48 (0.65) | 0.86 (0.01) | 763.1 (114.5) |
|                     | 15               | 4856.9 (13.8)  | 4454.4 (13.8) | 0.58 (0.16) | 0.85 (0.01) | 4623.9 (53.2)  |
| SUNGH-MFA (C)       | 5                | 239.9 (70.7)   | -162.6 (70.7) | 0.53 (0.42) | 0.87 (0.0)  | -205.9 (88.7)  |
|                     | 15               | 1681.8 (199.0) | 1279.3 (199.0) | 0.43 (0.18) | 0.86 (0.01) | 1259.2 (246.2) |

According to model selection criteria, for two maintained scenarios, the conditional SUNGH-MFA (C) model is clearly better than the unconditional SUNGH-MFA (UC) model. Interestingly, for this particular data, the classification performance and mean squared error for both types of models were quite similar. This appears to reflect that only one factor (m=1) is needed to represent the data suggesting that some observations and variables (dimensions) provide little additional information. For other types of data, it is expected to see quite different results with greater sensitivity to the missing data imputation approach. An alternative way for the unconditional approach is, if a single value is missing, include Listwise deletion (entire record is removed). This method is only possible to be applied 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. Conclusion

In this work, using a Bayesian framework, a flexible class of multivariate SUNGH distributions was employed to analyze MFA models. The estimation of the parameters of MFA model based on SUNGH family was relatively straightforward with a Bayesian approach. As the SUNGH distributions are suitable to model the asymmetric data with/without outliers, it can be used in the robust statistical inferences. Various extensions to the SUNGH-MFA model can be considered for future works, for example more general arrangement of a structural equation model based on the SUNGH 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.

References
1. S. X. Lee and G. J. McLachlan, Model-based clustering and classification with non-normal mixture distributions, Stat. Methods Appl. 22(4) (2013) 427–454.
2. S. X. Lee and G. J. McLachlan, On mixtures of skew normal and skew t distributions, Adv. Data Anal. Classif. 7(3) (2013) 241–266.
3. G. J. McLachlan and D. Peel, Finite Mixture Models (Wiley, New York, 2000).
4. Z. Ghahramani and G. E. Hinton, The EM algorithm for mixtures of factor analyzers (Tech. Rep. No. CRG-TR-96-1), Toronto: University of Toronto, Department of Computer Science (1997).
5. G. E. Hinton, P. Dayan and M. Revow, Modeling the manifolds of images of handwritten digits, IEEE trans. neural netw. 8 (1997) 65–74.
6. M. M. Wall, J. Guo and 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.
7. M. Maleki, D. Wraith and R. B. Arellano-Valle, Robust finite mixture modeling of multivariate unrestricted skew-normal generalized hyperbolic distributions, Stat. Comput. 29 (3) (2019) 415–428.
8. M. Maleki, G. J. McLachlan and S. X. Lee, Robust clustering based on finite mixture of multivariate fragmental distributions, Stat. Modelling (2022) doi.org/10.1177/1471082X211048660.
9. M. Maleki, A. Hajrajabi, R. B. Arellano-Valle. Symmetrical and Asymmetrical Mixture Autoregressive Processes, Braz. J. Probab. Stat. 34 (2) (2020) 273-290.
10. M. Maleki, H. Bidram and D. Wraith. Robust clustering of COVID-19 cases across US counties using mixtures of asymmetric time series models with time varying and freely indexed covariates, J. Appl. Stat. (2022). https://doi.org/10.1080/02664763.2021.2019688
11. M. R. Mahmoudi, M. Maleki, D. Baleanu, V. T. Nguyen and K. H. Pho. A Bayesian approach to heavy-tailed finite mixture autoregressive models, Symmetry 12 (6) (2020) 929.
12. A. Zarei, Z. Khodadadi, M. Maleki and K. Zare. Robust mixture regression modeling based on two-piece scale mixtures of normal distributions, Adv. Data Anal. Classif. (2022). https://doi.org/10.1007/s11634-022-00495-6
13. T. I. Lin, J. C. Lee and S. Y. Yen, Finite mixture modeling using the skew-normal distribution, Stat. Sin. 17 (2007) 909-927.
14. A. Azzalini, A class of distributions which includes the normal ones, Scand. J. Stat. 12 (1985) 171-178.
15. R. B. Arellano-Valle and A. Azzalini, On the unification of families of skew-normal distributions, Scand. J. Stat. 33 (2006) 561-574.
16. M. D. Branco and D. K. Dey, A general class of multivariate skew-elliptical distributions, J. Multivar. Anal. 79 (2001) 99–113.
17. S. X. Lee and G. J. McLachlan, Finite mixtures of canonical fundamental skew t-distributions: the unification of the restricted and unrestricted skew t-distribution models, *Stat. Comput.* **26**(3) (2016) 573–589.
18. H. M. Kim, M. Maadooliat, R. B. Arellano-Valle and M. G. Genton, Skewed factor models using selection mechanisms, *J. Multivar. Anal.* **145** (2016) 162-177.
19. T. I. Lin, G. J. McLachlan and S. X. Lee, Extending mixtures of factor models using the restricted multivariate skew-normal distribution, *J. Multivar. Anal.* **143** (2016) 398-413.
20. T. Ando, Bayesian factor analysis with fat-tailed factors and its exact marginal likelihood, *J. Multivar. Anal.* **100**(8) (2009) 1717 – 1726.
21. S. Y. Lee and Y. M. Xia, A robust Bayesian approach for structural equation models with missing data, *Psychometrika* **73** (2008a) 343–364.
22. M. Yang and D. B. Dunson, Bayesian semiparametric structural equation models with latent variables, *Psychometrika* **75**(4) (2010) 675–693.
23. S. Y. Lee and Y. M. Xia, Semiparametric Bayesian analysis of structural equation models with fixed covariates, *Stat. Med.* **27** (2008b) 2341–2360.
24. X. Y. Song, J. H. Pan, T. Kwok, L. Vandenput, C. Ohlsson and P. C. Leung, A semiparametric Bayesian approach for structural equation models, *Biom. J.* **52**(3) (2010) 314–332.
25. M. Chen, J. Silva, J. Paisley, C. Wang, D. Dunson and L. Carin, Compressive Sensing on Manifolds Using a Nonparametric Mixture of Factor Analyzers: Algorithm and Performance Bounds, *IEEE Trans. Signal Process* **58**(12) (2010) 6140–6155.
26. J. Paisley and L. Carin, Nonparametric factor analysis with beta process priors, *Proceedings of the 26th Annual International Conference on Machine Learning* (2009) 777–784.
27. C. M. Carvalho, J. Chang, J. E. Lucas, J. R. Nevin, Q. Wang, and M. West, High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics, *J. Am. Stat. Assoc.* **103**(484) (2008) 1438–1456.
28. A. J. Suarez and S. Ghosal, Bayesian estimation of principal components for functional data, *Baysian Anal.*** **12**(2) (2017) 311-333.
29. M. Maleki and D. Wraith, Mixtures of multivariate restricted skew-normal factor analyzer models in a Bayesian framework. *Comput. Stat.* **34**(3) (2019) 1039–1053.
30. R. B. Arellano-Valle and M. G. Genton, On fundamental skew distributions, *J. Multivar. Anal.* **96** (2005) 93-116.
31. M. Maleki, D. Wraith and R. B. Arellano-Valle. A flexible class of parametric distributions for Bayesian linear mixed models. *TEST* **28**(2) (2019) 543–564.
32. S. Ghasami, Z. Khodadadi and M. Maleki, Autoregressive processes with generalized hyperbolic innovations, *Commun. Stat. Simul. Comput.* **49**(12) (2014) 3080-3092
33. S. X. Lee G. J. McLachlan. On formulations of skew factor models: skew errors versus skew factors, *Stat. Probab. Lett.* 168 (2021) https://doi.org/10.1016/j.spl.2020.108935.
34. S. K. Sahu, D. K. Dey and M. D. Branco, A new class of multivariate skew distributions with applications to Bayesian regression models, *Can. J. Stat.* **31**(2) (2003) 129–150.
35. I. J. Good, The population frequencies of species and the estimation of population parameters, *Biometrika* **40** (1953) 237–260.
36. O. Barndorff-Nielsen and C. Halgreen, Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **38** (1977) 309–311.
37. J. Bai and K. Li, Statistical analysis of factor models in high dimensions, *Ann. Stat.* **40**(1) (2012) 436-465.
38. O. Barndorff-Nielsen, Hyperbolic distributions and distributions on hyperbolae, *Scand. J. Stat.* **5** (1978) 151–157.
39. H. F. Lopes and M. West, Bayesian model assessment in factor analysis, *Stat. Sin.* **4** (2004) 41–67.
40. E. Fokoué and D. M. Titterington, Mixtures of factor analyzers: Bayesian estimation and inference by stochastic simulation, *Mach. Learn.* **50** (2003) 73–94.
41. A. E. Gelfand and A. F. M. Smith, Sampling based approaches to calculating marginal densities, *J. Am. Stat. Assoc.* **85** (1990) 398-409.
42. D. Gamerman, Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference (London, Chapman & Hill, 1997).
43. NIMBLE Development Team, NIMBLE: an R package for programming with BUGS models, Version 0.6-10, 2021, http://r-nimble.org.
44. A. Gelman and D. B. Rubin, Inference from iterative simulation using multiple sequences (with discussion), *Stat. Sci.* **7** (1992) 457-511.
45. L. Hubert and P. Arabie, Comparing partitions, *J. Classif.* 2 (1985) 193–218.
46. B. P. Carlin and T. A. Louis, Bayesian Methods for Data Analysis, 3rd edn, (Chapman & Hall, CRC Press, 2011).
47. G. Celeux, F. Forbes, C. P. Robert and D. M. Titterington, Deviance information criteria for missing data models, *Bayesian Anal.* 1(4) (2006) 651–674.
48. https://cran.r-project.org/web/packages/Stat2Data/index.html.
49. S. X. Lee, T. I. Lin, G. J. McLachlan, Mixtures of factor analyzers with scale mixtures of fundamental skew normal distributions. *Adv. Data Anal. Classif.* 15 (2021) 481–512.