Proceedings of the 6th International Conference on Statistics: Theory and Applications (ICSTA'24)
Barcelona, Spain – August 19 - 21, 2024
Paper No. 131
DOI: 10.11159/icsta24.131

Model-Based Clustering Analysis for Multivariate Longitudinal Data via Mixtures of Matrix-Variate t-distributions

Farzane Ahmadi[1*], Elham Faghihzadeh[2]
[1] Department of Biostatistics and Epidemiology, Faculty of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran
ahmadi.farzane@zums.ac.ir; faghihzadeh.elham@gmail.com
[2] Independent researcher, Tehran, Iran

***Abstract*:** The finite mixture model is considered as an appropriate instrument for data clustering. Different parsimonious multivariate mixture distributions are introduced for skewed and/or heavy-tailed longitudinal data. The eigenvalue or modified Cholesky decomposition of covariance matrices develops the families of parsimonious mixture models. Thus, the finite mixture of matrix-variate t-distributions for clustering a three-way dataset with heavy-tailed or outlier observations (e.g., multivariate longitudinal data) is more appropriate compared to matrix-variate normal distributions. Accordingly, the present study considered a parsimonious family of the finite mixture of matrix-variate t-distributions using the eigenvalue and modified Cholesky decomposition for within and between covariance matrices, respectively. Finally, parameter estimates were calculated using the expectation-maximization algorithm, and simulations studies and real data analyses were conducted to confirm the obtained results.

***Keywords*:** Eigenvalue Decomposition; Finite Mixture; Matrix-Variate t-Distribution; Modified Cholesky Decomposition; Multivariate Longitudinal Data; Parsimonious Covariance Structures

1. Introduction

Finite mixture models in the statistical data analysis mainly contribute to modelling a heterogeneous population and providing an easy and model-based method for clustering and classification structure [1], [2]. Different studies have evaluated various finite mixtures of distributions focusing on multivariate (two-way data) distributions. For instance, such studies have proposed different finite mixtures of multivariate distributions, including multivariate normal distribution [3], multivariate t-distribution [4], multivariate skew-normal distribution [5], multivariate skew-t-distribution [6], multivariate normal inverse Gaussian [7], multivariate generalized hyperbolic distribution [8], and multivariate power exponential distribution [9] over the last two decades.

Three-way data including multivariate longitudinal, spatial multivariate, and spatio-temporal data may be available in a range of scientific domains [10]. Despite the important role of matrix-variate distributions in three-way data analysis, a small body of research exists in this respect. For example, *Viroli* introduced the finite mixtures of matrix-variate normal distributions (MVNDs) for classifying the three-way data [11]. In addition, *Anderlucci* and *Viroli* [12] considered the finite mixture of MVNDs for multivariate longitudinal data. In another study, *Doğru*, *Bulut* and *Arslan* [13] proposed a finite mixture of matrix-variate t-distributions (MVTDs). Further, *Gallaugher* and *McNicholas* [14]–[16] applied four skewed matrix-variate distributions of matrix-variate skew-t, generalized hyperbolic, variance-gamma, and normal inverse Gaussian distributions in the finite mixture of these distributions. In the two- or three-way data, where there are some departures from normality in datasets, using normal distributions affects the estimation of some parameters [17]. The presence of outlier or heavy-tailed data is considered as one of the common departures from normality and in such case, the mixture of t-distributions is an appropriate alternative to the mixture of normal distributions [13].

On the other hand, without any constraints on mixture parameters, the number of estimated parameters increases dramatically by an increase in components. Therefore, some constraints should be put on model parameters in order to achieve more parsimonious models. Considering a large number of mixture parameters in the covariance matrix component, more attention should be drawn on covariance structure decomposition. Further literature contains parsimonious covariance matrices in the mixture of multivariate distributions [18]–[24]. Some studies have investigated the parsimonious feature only in the finite mixture of MVNDs for three-way data [11], [12]. However, to the best of our knowledge, no research has applied the parsimonious MVTD mixture model. Therefore, the present study focused on the parsimonious mixture of MVTDs for

clustering multivariate longitudinal data with outliers or heavy-tails. The remaining sections of the present study are organized as follows. Section 2 reviews the finite mixture of MVTDs, the decomposition of covariance matrices, and the details of the estimates of MVTD parameters. Furthermore, Section 3 discusses the simulation studies and real examples in order to demonstrate the performance of models.

2. Method

2.1 Background

2.1.1 Finite mixture of MVTDs

A $T \times p$ dimensional random matrix $\mathbf{X}$ is assumed to arise from a parametric finite mixture if it is possible to write $p(\boldsymbol{X}|\vartheta) = \sum_{i=1}^{k} \pi_i p_i(\boldsymbol{X}|\theta_i)$ for all $\boldsymbol{X} \subset \boldsymbol{\chi}$, where $\boldsymbol{\vartheta} = (\pi_1, \dots, \pi_k, \theta_1, \dots, \theta_k)$ is the vector of parameters, and π_i and k are the mixing proportion and the number of mixture components, respectively, so that $\sum_{i=1}^{k} \pi_i = 1$ and $\pi_i \in [0,1]$. Additionally, $p_i(\boldsymbol{X}|\theta_i)$ is referred to as the density of the component. In the mixture of MVTDs, component density with a $T \times p$ mean matrix $\boldsymbol{M}_i$, two covariance matrices $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Omega}_i$ with dimensions $T \times T$ and $p \times p$, and degrees of freedom v_i is as follows [25]:

$$Mt^{(T\times p)}(\boldsymbol{X}|\boldsymbol{M}_i, \boldsymbol{\Phi}_i, \boldsymbol{\Omega}_i, v_i) = \frac{\Gamma\left(\frac{Tp + v_i}{2}\right)}{(\pi v_i)^{\frac{Tp}{2}} \Gamma\left(\frac{v_i}{2}\right) |\boldsymbol{\Phi}_i|^{\frac{p}{2}} |\boldsymbol{\Omega}_i|^{\frac{T}{2}}} \left(1 + \frac{tr\{(\boldsymbol{X} - \boldsymbol{M}_i)' \boldsymbol{\Phi}_i^{-1} (\boldsymbol{X} - \boldsymbol{M}_i) \boldsymbol{\Omega}_i^{-1}\}}{v_i}\right)^{-\frac{Tp+v_i}{2}} \quad (1)$$

where T and p indicate the number of measurement times and the number of response variables, respectively. In addition, $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Omega}_i$ are commonly referred to as *between* and *within* covariance matrices, respectively. In the present study, the *upper case boldface* was used for the matrices. The MVTDs arise as a particular case of a normal-variance mixture distribution. In this formulation, the random matrix $\mathbf{X}$ is defined as $\boldsymbol{X} = \boldsymbol{M} + W^{\frac{1}{2}}\boldsymbol{V}$, where the matrix random $\mathbf{V}$ has the MVND with the mean matrix $\mathbf{0}$ and covariance matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}$, $\boldsymbol{V} \sim \phi^{(T\times p)}(\boldsymbol{X}|\boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Omega})$, and the latent random variable W follows an inverse gamma distribution with parameters $(\frac{v}{2}, \frac{v}{2})$ [13]. In addition, the estimates of $\boldsymbol{\Omega}_i$ and $\boldsymbol{\Phi}_i$ are not unique. For each positive and nonzero constant a, we have $\boldsymbol{\Omega}_i \otimes \boldsymbol{\Phi}_i = a\boldsymbol{\Omega}_i \otimes \left(\frac{1}{a}\right)\boldsymbol{\Phi}_i$.The constraint $tr(\boldsymbol{\Omega}_i) = p$ or $tr(\boldsymbol{\Phi}_i) = T$ can be used to obtain an identifiable solution for $\boldsymbol{\Omega}_i$ and $\boldsymbol{\Phi}_i$ [12], [14].

2.1.2 The decomposition of covariance matrices

Restrictions on mixture parameters are typically constructed by constraining covariance matrices. To achieve parsimonious models, eigenvalue and the modified Cholesky decompositions were used for the *between* and *within* covariance matrices, respectively.

The eigenvalue decomposition

The eigenvalue decomposition was used in multivariate normal mixtures and the other multivariate mixture distributions such as t-mixture distributions [22], along with skew-normal and skew-t mixture distributions [24] for clustering, classification, and discriminant analysis. On the other hand *Viroli* [11] and. Sarkar *et al.* [27] applied the eigenvalue decomposition in the mixture of MVNDs. This parameterization includes the expression *within* component-covariance matrix ($\boldsymbol{\Omega}_i$) in terms of its eigenvalue decomposition as $\boldsymbol{\Omega}_i = \lambda_i \boldsymbol{D}_i \boldsymbol{A}_i \boldsymbol{D}_i'$, where $\boldsymbol{D}_i$ denotes the matrix of eigenvectors Furthermore, $\boldsymbol{A}_i$ is a diagonal matrix whose elements are proportional to the eigenvalues of $\boldsymbol{\Omega}_\mathrm{i}$ and λ_i represents the associated proportionality constant. Different sub-models can be defined by considering homoscedastic or varying quantities across mixture components [11], [20].

Modified Cholesky decomposition

The *between* component-covariance matrix ($\boldsymbol{\Phi}_i$) of the multivariate longitudinal data can be decomposed by the modified Cholesky decomposition. *Anderlucci* and *Viroli* [12] employed the above-mentioned decomposition along with the eigenvalue decomposition for the *between* and *within* covariance structures in the mixture of MVNDs, respectively. The modified Cholesky decomposition was expressed as $\boldsymbol{\Phi}_i^{-1} = \boldsymbol{U}_i' \boldsymbol{T}_i^{-1} \boldsymbol{U}_i$ where $\boldsymbol{U}_i$ is a unique lower triangular matrix with diagonal elements 1 and $\boldsymbol{T}_i$ denotes a unique diagonal matrix with strictly positive diagonal entries representing innovation variances. The lower diagonal elements in $\boldsymbol{U}_i$ equal the negative coefficients resulted from the regression of X_t on $X_{t-1}, X_{t-2}, \ldots, X_1$ e.g $\hat{X}_t = M_t + \sum_{s=1}^{t-1} \phi_{r,s}^{(i)} (X_t - M_t)$ [28]. On the other hand, different orders (m) can be considered in matrix $\boldsymbol{U}_i$, where m can range from 0 to T-1. The lower orders provide more parsimonious models so that m=0 and m=1 equal the independency of different times and the dependency of $\boldsymbol{X}_t$ on a previous time ($\boldsymbol{X}_{t-1}$), and the like. Accordingly, the modified Cholesky decomposition for the temporal covariance matrix equals the generalized autoregressive with process order m, GAR(m). Thus, the rth row elements of matrix $\boldsymbol{U}_i$ which should be estimated can be written as $\left(-\phi_{r,r-m}^{(i)}, -\phi_{r,r-m+1}^{(i)}, \ldots, -\phi_{r,r-1}^{(i)}\right)'; r = 2, \ldots, T,\ m = 0,1, \ldots, T-1$. Additionally, matrix $\boldsymbol{T}_i$ can be defended as $\boldsymbol{T}_i = d_i I_T$ (Isotropic) for a more parsimonious model. In addition, different sub-models can be defined by considering homoscedastic or varying quantities (i.e., $\boldsymbol{U}_i$ and $\boldsymbol{T}_i$) across mixture components [12].

2.2 Estimation of parameters

To find the maximum likelihood estimators for mixture parameters, the present study used an expectation-maximization (EM) algorithm [29] which was proposed by *Doğru*, *Bulut* and *Arslan* [13] for the mixture of MVTDs. Assume that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, where n is the number of observations, be a random sample of matrices from the mixture of MVTDs, and Z_{ij} denotes the component membership of observation j. Further, $Z_{ij} = 1$ if the jth observation is from component i, otherwise, $Z_{ij} = 0$, where $j = 1, \ldots, n$ and $i = 1, \ldots, k$. MVTDs are expressed as follows:

$$\boldsymbol{X}_j | W_j, Z_{ij} = 1 \sim \phi^{(T \times p)}\left(\boldsymbol{M}_i, W_j^{-1} \boldsymbol{\Phi}_i, \boldsymbol{\Omega}_i\right), \qquad W_j | Z_{ij} = 1 \sim Gamma\left(\frac{v_i}{2}, \frac{v_i}{2}\right). \tag{2}$$

Based on the hierarchical representation of the MVTDs, the complete data log-likelihood $\ell_c(\vartheta)$ can be written as follows:

$$\ell_c(\vartheta) = \sum_{j=1}^{n} \sum_{i=1}^{k} Z_{ij} \left[-\frac{Tp}{2} log\, 2\pi - \frac{p}{2} log |u_j^{-1} \boldsymbol{\Phi}_i| - \frac{T}{2} log |\boldsymbol{\Omega}_i| - \frac{W_j}{2} tr\left\{ \boldsymbol{\Omega}_i^{-1} (\boldsymbol{X}_j - \boldsymbol{M}_i) \boldsymbol{\Phi}_i^{-1} (\boldsymbol{X}_j - \boldsymbol{M}_i)' \right\} + \frac{v_i}{2} log\left(\frac{v_i}{2}\right) \right.$$
$$\left. - log\, \Gamma\left(\frac{v_i}{2}\right) - \frac{v_i}{2} W_j + \left(\frac{v_i}{2} - 1\right) log(W_j) \right] + \sum_{j=1}^{n} \sum_{i=1}^{k} Z_{ij} \log \pi_i \ . \tag{3}$$

An EM algorithm is as follows:

I. Initialization: Initialize parameters π_i, $\mathbf{M}_i$, $\boldsymbol{\Phi}_i$, $\boldsymbol{\Omega}_i$, and v_i, setting $t = 0$.

II. E-step: Update $\mathrm{E}(\mathrm{Z_{ij}} | \mathbf{X_j}, \vartheta)$, $\mathrm{E}(W_j | \boldsymbol{X_j}, \mathrm{Z_{ij}} = 1; \vartheta)$, and $\mathrm{E}(\log W_j | \boldsymbol{X_j}, \mathrm{Z_{ij}} = 1; \vartheta)$, where

$$E\left(Z_{ij} | \boldsymbol{X_j}, \vartheta^{(t)}\right) = P\left(Z_{ij} = 1 | \boldsymbol{X_j}, \vartheta^{(t)}\right) = \frac{\pi_i^{(t)} Mt^{T \times p}\left(\boldsymbol{X}_j;\ \boldsymbol{M}_i^{(t)}, \boldsymbol{\Phi}_i^{(t)}, \boldsymbol{\Omega}_i^{(t)}, v_i^{(t)}\right)}{\sum_{i=1}^{k} \pi_i^{(t)} Mt^{T \times p}\left(\boldsymbol{X}_j;\ \boldsymbol{M}_i^{(t)}, \boldsymbol{\Phi}_i^{(t)}, \boldsymbol{\Omega}_i^{(t)}, v_i^{(t)}\right)} = \tau_{ij}^{(t)},$$

$$E\left(W_j | \boldsymbol{X_j}, Z_{ij} = 1; \vartheta^{(t)}\right) = \frac{Tp + \hat{v}_i^{(t)}}{tr\left\{ \boldsymbol{\Omega}_i^{(t)^{-1}} \left(\boldsymbol{X}_j - \boldsymbol{M}_i^{(t)}\right)' \boldsymbol{\Phi}_i^{(t)^{-1}} \left(\boldsymbol{X}_j - \boldsymbol{M}_i^{(t)}\right) \right\} + v_i^{(t)}} = W_{1ij}^{(t)}, \tag{4}$$

$$\left(log\, W_j | \boldsymbol{X_j}, Z_{ij} = 1; \vartheta^{(t)}\right) = DG\left(\frac{Tp + v_i^{(t)}}{2}\right) + log\left(\frac{tr\left\{ \boldsymbol{\Omega}_i^{(t)^{-1}} \left(\boldsymbol{X}_j - \boldsymbol{M}_i^{(t)}\right)' \boldsymbol{\Phi}_i^{(t)^{-1}} \left(\boldsymbol{X}_j - \boldsymbol{M}_i^{(t)}\right) \right\} + v_i^{(t)}}{2}\right) = W_{2ij}^{(t)},$$

where $DG(t) = \frac{d}{dt} log\, \Gamma(T)$ represents the digamma function. Furthermore, $W_{1ij}^{(t)}$ is calculated based on $W_j|\boldsymbol{X_j}, Z_{ij} = 1$ distribution, which has a gamma distribution with parameters $\frac{Tp+\hat{v}_i^{(t)}}{2}$ and $\frac{tr\left\{\boldsymbol{\Omega}_i^{(t)^{-1}}\left(\boldsymbol{X_j}-\boldsymbol{M}_i^{(t)}\right)'\boldsymbol{\Phi}_i^{(t)^{-1}}\left(\boldsymbol{X_j}-\boldsymbol{M}_i^{(t)}\right)\right\}+v_i^{(t)}}{2}$, and $W_{2ij}^{(t)}$ is achieved using the moment-generating function of $W_j|\boldsymbol{X_j}, Z_{ij} = 1$.

III. M-step: Update π_i, $\boldsymbol{M}_i$, $\boldsymbol{\Omega}_i$, $\boldsymbol{\Phi}_i$, and v_i. The order of parameter estimation is as follows (1): π_i and $\boldsymbol{M}_i$; (2) $\boldsymbol{\Omega}_i$; (3) $\boldsymbol{\Phi}_i$; (4) v_i

1. *Update π_i and $\boldsymbol{M}_i$:*

$$\pi_i^{(t+1)} = \frac{\sum_{j=1}^{n}\tau_{ij}^{(t)}}{n}, \qquad \boldsymbol{M}_i^{(t+1)} = \frac{\sum_{j=1}^{n}\tau_{ij}^{(t)}W_{1ij}^{(t)}\boldsymbol{X}_j}{\sum_{j=1}^{n}\tau_{ij}^{(t)}W_{1ij}^{(t)}}, \tag{5}$$

2. *Update $\boldsymbol{\Omega}_i$*

Assuming that $\boldsymbol{B}_i = \sum_{j=1}^{n}\tau_{ij}^{(t)}W_{1ij}^{(t)}\left(\boldsymbol{X}_j - \boldsymbol{M}_i^{(t+1)}\right)'\boldsymbol{\Phi}_i^{(t)^{-1}}\left(\boldsymbol{X}_j - \boldsymbol{M}_i^{(t+1)}\right)$, the $\ell_c(\vartheta)$ is proportional to $-\frac{T}{2}\sum_{i=1}^{k} n_i\, log|\boldsymbol{\Omega}_i| - \frac{1}{2}\sum_{i=1}^{k} tr\{\boldsymbol{\Omega}_i^{-1}\boldsymbol{B}_i\}$ with $n_i = \sum_{j=1}^{n}\tau_{ij}^{(t)}$. The estimates o parameters for the eight sub-models are provided below.

- *Sub-model VVV*: The maximization of $-\frac{T}{2}\sum_{i=1}^{k} n_i\, log|\boldsymbol{\Omega}_i| - \frac{1}{2}\sum_{i=1}^{k} tr\{\boldsymbol{\Omega}_i^{-1}\boldsymbol{B}_i\}$ with respect to $\boldsymbol{\Omega}_i$ leads to $\boldsymbol{\Omega}_i^{(t+1)} = \frac{\boldsymbol{B}_i}{n_i T}$;
- *Sub-model EEE*: The maximization of $-\frac{Tn}{2} log|\boldsymbol{\Omega}| \sum_{i=1}^{k} - \frac{1}{2} tr\left\{\boldsymbol{\Omega}_i^{-1}\sum_{i=1}^{k}\boldsymbol{B}_i\right\}$, where $n = \sum_{i=1}^{k}\sum_{j=1}^{n}\tau_{ij}^{(t)}$, with respect to $\boldsymbol{\Omega}_i = \boldsymbol{\Omega}$ leads to $\boldsymbol{\Omega}^{(t+1)} = \frac{\sum_{i=1}^{k}\boldsymbol{B}_i}{nT}$;
- *Sub-model VVI*: The maximization of $-\frac{Tp}{2}\sum_{i=1}^{k} n_i\, log\, \lambda_i - \frac{1}{2}\sum_{i=1}^{k}\frac{1}{\lambda_i} tr\{\mathbf{A}_i^{-1}\boldsymbol{B}_i\}$ with respect to $\boldsymbol{\Omega}_i = \lambda_i \boldsymbol{D}\boldsymbol{A_i}\boldsymbol{D}'$ leads to $\lambda_i^{(t+1)} = \frac{|diag(\boldsymbol{B}_i)|^{\frac{1}{p}}}{Tn_i}$ and $\boldsymbol{A}_i^{(t+1)} = \frac{diag(\boldsymbol{B}_i)}{|diag(\boldsymbol{B}_i)|^{\frac{1}{p}}}$;
- *Sub-model EEI*: The maximization of $-\frac{pT}{2}\sum_{i=1}^{k} n_i\, log|\boldsymbol{\Omega}_i| - \frac{1}{2}\sum_{i=1}^{k} tr\{\boldsymbol{\Omega}_i^{-1}\boldsymbol{B}_i\}$ with respect to $\boldsymbol{\Omega}_i$ leads to $\lambda^{(t+1)} = \frac{\left|diag\left(\sum_{i=1}^{k}\boldsymbol{B}_i\right)\right|^{\frac{1}{p}}}{Tn}$ and $\boldsymbol{A}^{(t+1)} = \frac{diag\left(\sum_{i=1}^{k}\boldsymbol{B}_i\right)}{|diag(\sum_{i=1}^{k}\boldsymbol{B}_i)|^{\frac{1}{p}}}$;
- *Sub-model VII*: The maximization of $-\frac{pT}{2}\sum_{i=1}^{k} n_i\, log\, \lambda_i - \frac{1}{2}\sum_{i=1}^{k} tr\left\{\frac{\boldsymbol{B}_i}{\lambda_i}\right\}$ with respect to $\boldsymbol{\Omega}_i = \lambda_i \boldsymbol{I}$ leads to $\lambda_i^{(t+1)} = \frac{tr\{\boldsymbol{B}_i\}}{Tpn_i}$;
- *Sub-model EII*: The maximization of $-\frac{Tpn}{2} log\, \lambda - \frac{1}{2\lambda} tr\left\{\sum_{i=1}^{k}\boldsymbol{B}_i\right\}$ with respect to $\boldsymbol{\Omega} = \lambda \boldsymbol{I}$ leads to $\lambda^{(t+1)} = \frac{tr\left\{\sum_{i=1}^{k}\boldsymbol{B}_i\right\}}{Tpn}$;
- *Sub-model EEV*: The maximization of $-\frac{Tpn}{2} log\, \lambda - \frac{1}{2\lambda}\sum_{i=1}^{k} tr\{\boldsymbol{D}_i A^{-1}\boldsymbol{D}_i'\boldsymbol{B}_i\}$ with respect to $\boldsymbol{\Omega}_i = \lambda \boldsymbol{D}_i \boldsymbol{A}\boldsymbol{D}_i'$ leads to $\lambda^{(t+1)} = \frac{\left|\sum_{i=1}^{k}\boldsymbol{C}_i\right|^{\frac{1}{p}}}{n\,T}$, $A^{(t+1)} = \frac{\sum_{i=1}^{k}\boldsymbol{C}_i}{\left|\sum_{i=1}^{k}\boldsymbol{C}_i\right|^{\frac{1}{p}}}$, $\boldsymbol{D}_i^{(t+1)} = \boldsymbol{L}_i$, where for $i = 1, \ldots, k$ $\boldsymbol{C}_i$, and $\boldsymbol{L}_i$ are derived from the eigenvalue decomposition of the symmetric positive definite matrix $\boldsymbol{B}_i = \boldsymbol{L}_i\boldsymbol{C}_i\boldsymbol{L}_i'$ with the eigenvalues in the diagonal matrix $\mathbf{C}_i$ in descending order.
- *Sub-model III*: This situation equals the independence of the responses thus no parameters are available.

3. *Update $\boldsymbol{\Phi}_i$*

Considering that $\boldsymbol{S}^{(i)} = \sum_{j=1}^{n}\tau_{ij}^{(t)}\, W_{1ij}^{(t)}\left(\boldsymbol{X}_j - \widehat{\boldsymbol{M}}_i^{(t+1)}\right)\boldsymbol{\Omega}_i^{(t+1)^{-1}}\left(\boldsymbol{X}_j - \widehat{\boldsymbol{M}}_i^{(t+1)}\right)'$, $\ell_c(\vartheta)$ is proportional to $-\frac{p}{2}\sum_{i=1}^{k} n_i\, log|\boldsymbol{D}_i| - \frac{1}{2} tr\left\{\sum_{i=1}^{k}(\boldsymbol{U}_i'\boldsymbol{T}_i^{-1}\boldsymbol{U}_i)\boldsymbol{S}^{(i)}\right\}$. The estimates of parameters for the four sub-models are presented as follows:

- *Sub-model GAR(m):* The maximization of $-\frac{p}{2}\sum_{i=1}^{k} n_i\, log|\boldsymbol{D}_i| - \frac{1}{2} tr\left\{\sum_{i=1}^{k}(\boldsymbol{U}_i'\boldsymbol{T}_i^{-1}\boldsymbol{U}_i)\boldsymbol{S}^{(i)}\right\}$ with respect to $\boldsymbol{\Phi}_i$ leads to the r[th] row estimation of matrix $\boldsymbol{U}_i$ as

$$\begin{pmatrix} \phi_{r,r-m}^{(i)} \\ \phi_{r,r-m+1}^{(i)} \\ \cdots \\ \phi_{r,r-1}^{(i)} \end{pmatrix}^{(t+1)} = \begin{pmatrix} S_{r-m,r-m}^{(i)} & S_{r-m+1,r-m}^{(i)} & \cdots & S_{r-1,r-m}^{(i)} \\ S_{r-m,r-m+1}^{(i)} & S_{r-m+1,r-m+1}^{(i)} & \cdots & S_{r-1,r-m+1}^{(i)} \\ \cdots & \cdots & \ddots & \vdots \\ S_{r-m,r-1}^{(i)} & S_{r-m+1,r-1}^{(i)} & \cdots & S_{r-1,r-1}^{(i)} \end{pmatrix}^{-1} \begin{pmatrix} S_{r,r-m}^{(i)} \\ S_{r,r-m+1}^{(i)} \\ \cdots \\ S_{r,r-1}^{(i)} \end{pmatrix}, \quad (6)$$

and matrix $\boldsymbol{T}_i^{(t+1)} = \frac{1}{p} diag\left(\boldsymbol{U}_i{}^{(t+1)} \boldsymbol{S}^{(i)} \boldsymbol{U}_i{}^{(t+1)'}\right)$, where $r = 2, \dots, T$, $m = 0, \dots, T-1$, and $S_{l,t}^{(i)}$ is the l^{th}-row and r^{th}-column element of matrix $\mathbf{S}^{(i)}$.

- *Sub-model GARI(m):* The maximization of $-\frac{Tp}{2}\sum_{i=1}^{k} n_i \, log|d_i| - \frac{1}{2} tr\left\{\sum_{i=1}^{k} \frac{1}{d_i} \boldsymbol{U}_i' \boldsymbol{U}_i \boldsymbol{S}^{(i)}\right\}$ with respect to $\boldsymbol{\Phi}_i = \frac{1}{d_i} \boldsymbol{U}_i' \boldsymbol{U}_i$ leads to the same estimate of $\boldsymbol{U}_i$ as *sub-model GAR(m)* and estimate $d_i^{(t+1)} = \frac{tr\left\{\boldsymbol{U}_i{}^{(t+1)} \boldsymbol{S}^{(i)} \boldsymbol{U}_i{}^{(t+1)'}\right\}}{n_i p T}$.
- *Sub-model EGAR(m):* The maximization of $-\frac{np}{2} log|\boldsymbol{D}| - \frac{1}{2} tr\{\boldsymbol{U}' \boldsymbol{T}^{-1} \boldsymbol{U}(\sum_{i=1}^{k} \boldsymbol{S}^{(i)})\}$ with respect to $\boldsymbol{\Phi}_i = \boldsymbol{\Phi}$ leads to the same estimate of $\boldsymbol{U}$ as *sub-model GAR(m)* by replacing $\sum_{i=1}^{k} \boldsymbol{S}^{(i)}$instead of $\boldsymbol{S}^{(i)}$ and estimate $\boldsymbol{T}^{(t+1)} = \frac{1}{np} diag\left(\boldsymbol{U}^{(t+1)}\{\sum_{i=1}^{k} \boldsymbol{S}^{(i)}\}\boldsymbol{U}^{(t+1)'}\right)$.
- *Sub-model EGARI(m):* The maximization of $-\frac{npT}{2} log|d| - \frac{1}{2d} tr\{\boldsymbol{U}' \boldsymbol{U}(\sum_{i=1}^{k} \boldsymbol{S}^{(i)})\}$ with respect to $\boldsymbol{\Phi}_i = \boldsymbol{\Phi}$ leads to the same estimate of $\boldsymbol{U}$ as *sub-model EGAR(m)* and estimate $d^{(t+1)} = \frac{tr\left\{\boldsymbol{U}^{(t+1)'} \boldsymbol{U}^{(t+1)} \sum_{i=1}^{k} \boldsymbol{S}^{(i)}\right\}}{pnT}$.

4. Update v_i

For the degree of freedom, two situations were considered, including equal and unequal v_i across mixture components (constrained and unconstrained v_i, respectively). Given $\tau_{ij}^{(t+1)}$, $\pi_i^{(t+1)}$, $\boldsymbol{M}_i^{(t+1)}$, $\boldsymbol{\Omega}_i^{(t+1)}$, and $\boldsymbol{\Phi}_i^{(t+1)}$, the estimations of v_i are calculated by finding the root of equations (8) and (9) in constrained and unconstrained situations, respectively.

$$1 + log\left(\frac{v}{2}\right) - DG\left(\Gamma\left(\frac{v}{2}\right)\right) + \frac{1}{n}\sum_{n}^{n}\sum_{k}^{k} \tau_{ij}^{(t+1)}\left(W_{2ij}^{(t+1)} - W_{1ij}^{(t+1)}\right) = 0, \quad (7)$$

$$1 + log\left(\frac{v_i}{2}\right) - DG\left(\Gamma\left(\frac{v_i}{2}\right)\right) + \frac{1}{n_i}\sum_{n}^{} \tau_{ij}^{(t+1)}\left(W_{2ij}^{(t+1)} - W_{1ij}^{(t+1)}\right) = 0. \quad (8)$$

IV. Check the convergence criterion: If not satisfied, set $t = t + 1$ and go to step **II** of the EM algorithm iteration.

2.3 Model selection and convergence criterion

It is possible to define a large family ($64 \times T$) of possible mixture models by allowing different sub-models for covariance matrices $\boldsymbol{\Omega}_i$ and $\boldsymbol{\Phi}_i$ with different orders for matrix $\boldsymbol{T}_i$, $m = 0,1,..,T-1$, and constrained/unconstrained for v_i. The model can be selected according to the Bayesian information criterion (BIC) as $BIC = 2\, l(x, \hat{\vartheta}) - h \, log \, n$, where $l(x, \hat{\vartheta})$ and $\hat{\vartheta}$ indicate the maximized log-likelihood and the maximum likelihood estimate of $\boldsymbol{\vartheta}$, respectively. Additionally, h and n are the number of free parameters in the model and the number of observations, respectively [30]. Other criteria are employed in addition to BIC to estimate the number of mixture components, such as Integrated Completed Likelihood (ICL), which is computed as $\text{ICL} \approx \text{BIC} - 2\sum_{j=1}^{n}\sum_{i=1}^{k} \text{MAP}[1]\left(\tau_{ij}^{(t)}\right) \log \tau_{ij}^{(t)}$, where $\text{MAP}\left(\tau_{ij}^{(t)}\right) = 1$ if the $max_{i=1,\dots,k}\left\{\tau_{ij}^{(t)}\right\} = i$, otherwise, $\text{MAP}\left(\tau_{ij}^{(t)}\right) = 0, j = 1, \dots, n$ and $i = 1, \dots, k$ [31].

[1] Maximum A Posteriori probability (MAP)

In general, 20 random multistate points were considered given that the starting values of the EM algorithm could affect the estimated parameters. If the convergence criterion $\left|l\left(x,\hat{\vartheta}^{(t+1)}\right)-l\left(x,\hat{\vartheta}^{(t)}\right)\right|<1.0e-5$ is met, the EM algorithm is stopped, and the range of values for $\hat{v}_i$ is limited to between 2 and 200 [22]. These models have been written in R software.

3 Results: Simulation studies and real data

3.1 Simulation 1

The first simulation study was conducted to evaluate the ability of the algorithm for recognizing the temporal structure. The features of simulation study 1 were: a number k of mixture components equal to 3, a k-vector of the degrees of freedom equal to 5, 5, 5, and a 4×4 *within* covariance matrix $\mathbf{\Omega}_{\mathrm{i}}$ with a structure equals to VVV. In addition, other features included a 6×6 temporal covariance matrix $\mathbf{\Phi}_{\mathrm{i}}$ with a structure equals to GAR(1) and GAR(3), and a sample size n equals 100, 200, 500, and 1000. For each setting, 100 datasets were generated from the mixture of the MVTDs based on the defined *within* and temporal covariance matrices. Then, the mixture of MVTDs and MVNDs was run for five different models according to different orders for $\mathbf{\Phi}_{\mathrm{i}}$: GAR(1), GAR(2), GAR(3), GAR(4), and GAR(5). The best model was chosen according to the BIC and ICL. The percentage (number) of correct model selection with MVTD is equal to 100 in all cases, and it ranges from 97 to 100 for MVNDs. In a situation with a true model GAR(3), this percentage varied from 99 to 100 and 93 to 99 for MVTD and MVND, respectively.

3.2 Simulation 2

It was performed to evaluate the ability of MVTDs to recover the MVNDs in multivariate longitudinal data. To this end, datasets were generated from two-component (k=2), matrix-variate mixture models. The MVND and MVTD were the first and second components, respectively, and the same covariance structures with different parameter values were used accordingly. Other features (i.e., $\mathbf{\Omega}_{\mathrm{i}}$, $\mathbf{\Phi}_{\mathrm{i}}$, and sample size) in this simulation are similar to the first simulation study. For each setting, 100 datasets were generated from mixture distributions based on the defined *within* and temporal covariance matrices. Further, the mixture of MVTDs and MVNDs was run for five different models of GAR(1), GAR(2), GAR(3), GAR(4), and GAR(5). Table 5 presents the average values of the degree of freedom (standard deviation) of a model with GAR(1) and GAR(3) structures for $\mathbf{\Phi}_{\mathrm{i}}$. Given k (=2) and $\mathbf{\Omega}_i$ (VVV), the estimated degrees of freedom demonstrated that the first component was normal. Furthermore, the degrees of freedom estimates were computed to be close to true values in MVTD mixture models.

Additionally, the misclassification error rate (MISC) and the measure of accuracy (γ) for mean and covariance matrices were computed for each dataset and model in order to compare the two models in parameter estimates. Therefore, the accuracy measures of $\boldsymbol{M}$, $\boldsymbol{\Omega}$ (=VVV), $\boldsymbol{U}$, and $\boldsymbol{T}$ (=GAR) were calculated by the following expressions [32]:

$$\gamma_{\boldsymbol{M}}=\frac{\sum_{i=1}^{k}\left\|\widehat{\boldsymbol{M}}_i-\boldsymbol{M}_i\right\|}{kTp},\qquad \gamma_{\boldsymbol{\Omega}}=\frac{\sum_{i=1}^{k}\left\|\widehat{\boldsymbol{\Omega}}_i-\boldsymbol{\Omega}_i\right\|}{\frac{kp(p+1)}{2}}\gamma_{\boldsymbol{U}}=\frac{\sum_{i=1}^{k}\left\|\widehat{\boldsymbol{U}}_i-\boldsymbol{U}_i\right\|}{k\varphi},\qquad \gamma_{\boldsymbol{T}}=\frac{\sum_{i=1}^{k}\left\|\widehat{\boldsymbol{T}}_i-\boldsymbol{T}_i\right\|}{kT} \tag{9}$$

where the lower accuracy measure (γ) implies higher accuracy for parameters.

Considering k (=2) and $\mathbf{\Omega}_i$ (VVV), $\gamma_{\boldsymbol{M}}$, $\gamma_{\boldsymbol{\Omega}}$, and $\gamma_{\boldsymbol{T}}$ were not sensitive to the misspecification of the order of the temporal covariance (m=1, 2, …, 5), and these values were nearly identical in MVTD and MVND mixture models. However, the values of $\gamma_{\boldsymbol{T}}$ relied on the misspecification of the temporal covariance order. In the two models, $\gamma_{\boldsymbol{T}}$ of the lower orders ($m=1,2$) was larger compared to the higher orders ($m=3,4,5$). It should be noted that MVND mixture models tend to overestimate $\gamma_{\boldsymbol{T}}$compared to MVTD mixture models. Eventually, the accuracy measures in both models decreased by an increase in the sample size (Table 1). The mean compute time for fitting the mixture of MVTDs vs as MVNDs with the true model GAR(3) was 6.66 vs. 0.37 for n=100, 10.17 vs. 0.62 for n=200, 23.44 vs. 1.30 for n=500, and 46.86 vs 2.59 for n=1000.

Table 1: Mean (S.D) of MISC and accuracy measures with GAR(3) structure for $\mathbf{\Phi}_i$ from simulation 2

		$\boldsymbol{\Phi}_i$									
		GAR(1)		GAR(2)		GAR(3)		GAR(4)		GAR(5)	
n		MVTD	MVND	MVTD	MVND	MVTD	MVND	MVTD	MVND	MVTD	MVND
200	MISC	0	0.0001(0.001)	0	0.0001(0.001)	0	0.0001(0.001)	0	0.0001(0.001)	0	0.0001(0.001)
	$\gamma_{\boldsymbol{M}}$	0.09 (0.01)	0.10 (0.01)	0.09 (0.01)	0.10 (0.01)	0.09 (0.01)	0.10 (0.01)	0.09 (0.01)	0.10 (0.01)	0.09 (0.01)	0.10 (0.01)
	$\gamma_{\boldsymbol{\Omega}}$	0.23 (0.002)	0.23 (0.003)	0.23 (0.001)	0.23 (0.002)	0.23 (0.001)	0.23 (0.001)	0.23 (0.001)	0.23 (0.001)	0.23 (0.001)	0.23 (0.001)
	$\gamma_{\mathbf{T}}$	0.49 (0.02)	0.57 (0.06)	0.44 (0.02)	0.52 (0.13)	0.40 (0.02)	0.48 (0.14)	0.40 (0.02)	0.48 (0.14)	0.40 (0.02)	0.48 (0.14)
	$\gamma_{\mathbf{U}}$	0.33 (0.002)	0.33 (0.002)	0.20 (0.01)	0.20 (0.02)	0.10 (0.02)	0.11 (0.03)	0.15 (0.04)	0.17 (0.05)	0.18 (0.04)	0.21 (0.06)
1000	MISC	0	0	0	0	0	0	0	0	0	0
	$\gamma_{\boldsymbol{M}}$	0.06 (0.01)	0.07 (0.01)	0.06 (0.01)	0.07 (0.01)	0.06 (0.01)	0.07 (0.01)	0.06 (0.01)	0.07 (0.01)	0.06 (0.01)	0.07 (0.01)
	$\gamma_{\boldsymbol{\Omega}}$	0.15(0.0002)	0.15 (0.0002)	0.15(0.0002)	0.15 (0.0002)	0.15(0.0002)	0.15 (0.0002)	0.15(0.0002)	0.15 (0.0002)	0.15(0.0002)	0.15 (0.0002)
	$\gamma_{\mathbf{T}}$	0.31 (0.01)	0.42 (0.03)	0.27 (0.01)	0.40 (0.02)	0.27 (0.01)	0.37 (0.02)	0.27 (0.01)	0.37 (0.02)	0.27 (0.01)	0.37 (0.02)
	$\gamma_{\mathbf{U}}$	0.28 (0.001)	0.28 (0.001)	0.17 (0.001)	0.20 (0.001)	0.05 (0.001)	0.06 (0.01)	0.05 (0.001)	0.06 (0.01)	0.05 (0.001)	0.06 (0.01)

3.3 Simulation 3

The ability of MVTD and the MVND mixture models was evaluated regarding recognizing the true number of mixture components when the data were generated from MVTD mixture models. Then, the impact of the misspecification of the temporal matrix on the estimation of the number of components was investigated as well. For each setting, 100 datasets were generated from the model with a GAR(3) structure. In addition, a different number of mixture components (k = 2, 3, and 4) was considered to evaluate the choice of k. Approximately the correct number of components (k=3) was selected for MVTDs in all cases. However, MVNDs tend to overestimate (k=4) the number of true components.

3.4 Real data: Gastrointestinal (GI) cancers

The age-standardized death rates of the three most common GI cancers were extracted from the *Our World In Data* website [33]. The information included the death rates (per 100,000 populations) of colon and rectum, stomach, and liver cancers in 186 countries during 1990-2015 (at 5-year intervals). A mixture of MVTDs and MVNDs was fitted with *k* ranging from 1 to 10. The best sub-model based on the BIC and ICL is (GAR(2), VVV) with k=7 in the mixture of MVNDs and (GAR(4), VVV) with the constrained degrees of freedom and k=6 in the mixture of MVTDs. The estimated degree of freedom for the mixture of the MVTDs was $\hat{v} = 3.33$. We also fitted a finite mixture of skew matrix-variate distributions introduced by *Gallaugher* and *McNicholas* [15] to GI data. These matrix-variate distributions are skew-t, generalized hyperbolic, variance-gamma, and normal inverse Gaussian distributions that we did not consider eigenvalue and the modified Cholesky decompositions for the *between* and *within* covariance matrices for those, respectively. Because of the huge number of parameters, any of these finite mixture had not been converge.

Further, stomach and liver cancer death rates in some countries were extremely higher compared to other countries. Thus, the mixture of the MVND model provided an additional cluster to allow outliers. For more details, the number of countries in each cluster and the maps of the included countries in each cluster of MVTD model is presented in Figure 1. The obtained cluster labels in the two models were the same for 132 countries.

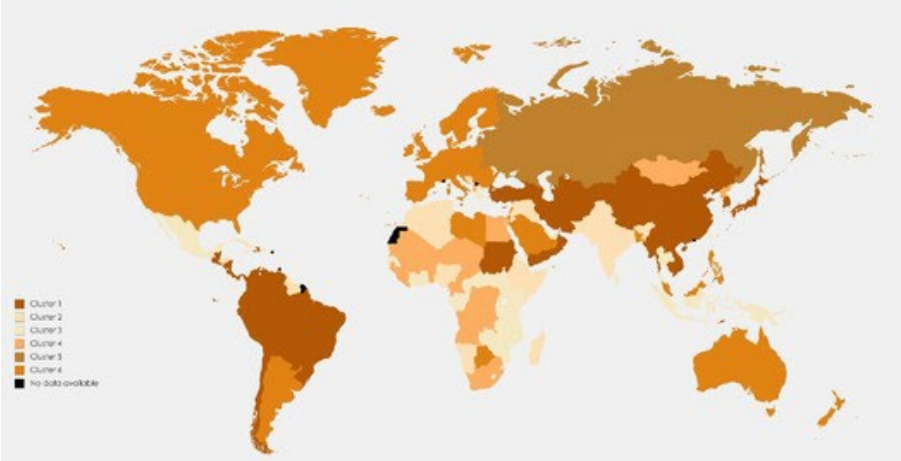

Figure 1: Trend of countries based on the death rates of the three common GI cancer resulted from the mixture of the MVTDs

Conclusion

Based on γ in the mixture models of MVTD and MVND, no differences were observed between the estimation of $\boldsymbol{M}$, $\boldsymbol{\Omega}$, and $\boldsymbol{T}$ matrices under different orders of temporal covariance structures in each model in simulation studies. Further, these values were similar in both models. On the other hand, the estimation of matrix $\boldsymbol{T}$ relies on the misspecification of $\boldsymbol{\Phi}$. Thus, $\gamma_{\boldsymbol{T}}$s should have the least value compared to lower orders if the order of the incorrect temporal structure is equal to or greater than the correct order of the temporal structure. The estimations of matrix $\mathbf{T}$ and the number of mixture components k are overestimated in MVND models if the datasets have a heavy-tail or outlier observations. The mixture of MVTDs commonly selected the model with the right temporal structure and the right number of mixture components. On the other hand, the time it took to fit a mixture of MVTDs was much longer than it required to fit a mixture of MVNDs, which is a trade-off for more precision.

References

[1] G. J. McLachlan and D. Peel, *Finite mixture models*. New York (NY): John Wiley & Sons, 2000.
[2] P. D. McNicholas, *Mixture model-based classification*, 1st ed. Chapman & Hall/CRC, 2016.
[3] G. J. McLachlan and K. E. Basford, *Mixture models : inference and applications to clustering*, 1st ed. CRC Press, 1987.
[4] D. Peel and G. J. McLachlan, "Robust mixture modelling using the t distribution," *Stat. Comput.*, vol. 10, no. 4, pp. 339–348, 2000.
[5] T. I. Lin, "Maximum likelihood estimation for multivariate skew normal mixture models," *J. Multivar. Anal.*, vol. 100, no. 2, pp. 257–265, Feb. 2009.
[6] T. I. Lin, "Robust mixture modeling using multivariate skew t distributions," *Stat. Comput.*, vol. 20, no. 3, pp. 343–356, Jul. 2010.
[7] A. O'Hagan, T. B. Murphy, I. C. Gormley, P. D. McNicholas, and D. Karlis, "Clustering with the multivariate normal inverse Gaussian distribution," *Comput. Stat. Data Anal.*, vol. 93, pp. 18–30, Jan. 2016.
[8] R. P. Browne and P. D. McNicholas, "A mixture of generalized hyperbolic distributions," *Can. J. Stat.*, vol. 43, no. 2, pp. 176–198, Jun. 2015.
[9] U. J. Dang, R. P. Browne, and P. D. McNicholas, "Mixtures of multivariate power exponential distributions," *Biometrics*, vol. 71, no. 4, pp. 1081–1089, Dec. 2015.
[10] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, 1st ed. New York: Chapman and Hall/CRC, 2000.
[11] C. Viroli, "Finite mixtures of matrix normal distributions for classifying three-way data," *Stat. Comput.*, vol. 21, no. 4, pp. 511–522, Oct. 2011.
[12] L. Anderlucci and C. Viroli, "Covariance pattern mixture models for the analysis of multivariate heterogeneous longitudinal data," *Ann. Appl. Stat.*, vol. 9, no. 2, pp. 777–800, 2015.

[13] F. Z. Doğru, Y. M. Bulut, and O. Arslan, "Finite mixtures of matrix variate t distributions," *Gazi Univ. J. Sci.*, vol. 29, no. 2, pp. 335–341, Jun. 2016.
[14] M. P. B. Gallaugher and P. D. McNicholas, "A matrix variate skew- *t* distribution," *Stat*, vol. 6, no. 1, pp. 160–170, Jan. 2017.
[15] M. P. B. Gallaugher and P. D. McNicholas, "Finite mixtures of skewed matrix variate distributions," *Pattern Recognit.*, vol. 80, pp. 83–93, Aug. 2018.
[16] M. P. B. Gallaugher and P. D. McNicholas, "Three skewed matrix variate distributions," *Stat. Probab. Lett.*, vol. 145, pp. 103–109, Feb. 2019.
[17] P. D. McNicholas, T. B. Murphy, A. F. McDaid, and D. Frost, "Serial and parallel implementations of model-based clustering via parsimonious Gaussian mixture models," *Comput. Stat. Data Anal.*, vol. 54, no. 3, pp. 711–723, Mar. 2010.
[18] J. Banfield and A. Jeffrey, "Model-Based Gaussian and Non-Gaussian Clustering," *Biometrics*, vol. 49, no. 3, pp. 803–21, Sep. 1993.
[19] G. Celeux and G. Govaert, "Gaussian parsimonious clustering models," *Pattern Recognit.*, vol. 28, no. 5, pp. 781–93, 1995.
[20] C. Fraley and A. E. Raftery, "Model-Based Clustering, Discriminant Analysis, and Density Estimation.," *J. Am. Stat. Assoc.*, vol. 97, no. 458, pp. 611–631, Jun. 2002.
[21] P. D. McNicholas and T. B. Murphy, "Model-based clustering of longitudinal data," *Can. J. Stat.*, vol. 38, no. 1, pp. 153–168, 2010.
[22] J. L. Andrews and P. D. McNicholas, "Model-based clustering, classification, and discriminant analysis via mixtures of multivariate t-distributions," *Stat. Comput.*, vol. 22, no. 5, pp. 1021–1029, 2012.
[23] P. D. McNicholas and S. Subedi, "Clustering gene expression time course data using mixtures of multivariate t-distributions," *J. Stat. Plan. Inference*, vol. 142, no. 5, pp. 1114–1127, May 2012.
[24] I. Vrbik and P. D. McNicholas, "Parsimonious skew mixture models for model-based clustering and classification," *Comput. Stat. Data Anal.*, vol. 71, pp. 196–210, Mar. 2014.
[25] A. K. Gupta, T. Varga, and T. Bodnar, *Elliptically Contoured Models in Statistics and Portfolio Theory*, 2nd ed. New York: Springer, 2013.
[26] M. P. B. Gallaugher and P. D. McNicholas, "A matrix variate skew- *t* distribution," *Stat*, vol. 6, no. 1, pp. 160–170, 2017.
[27] S. Sarkar, X. Zhu, V. Melnykov, and S. Ingrassia, "On parsimonious models for modeling matrix data," *Comput. Stat. Data Anal.*, vol. 142, p. 106822, Feb. 2020.
[28] M. Pourahmadi, "Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation," *Biometrika*, vol. 86, no. 3, pp. 677–690, Sep. 1999.
[29] X.-L. MENG and D. B. RUBIN, "Maximum likelihood estimation via the ECM algorithm: A general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, Jun. 1993.
[30] G. Schwarz, "Estimating the Dimension of a Model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.
[31] C. Biernacki, G. Celeux, and G. Govaert, "Assessing a mixture model for clustering with the integrated completed likelihood," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 7, pp. 719–725, Jul. 2000.
[32] L. Anderlucci and C. Viroli, "Supplement to "Covariance pattern mixture models for the analysis of multivariate heterogeneous longitudinal data," 2015.
[33] M. Roser and H. Ritchie, "Cancer," *Our World in Data*, 2019. [Online]. Available: https://ourworldindata.org/cancer. [Accessed: 16-Sep-2019].