Advances in Data Analysis and Classification
https://doi.org/10.1007/s11634-024-00585-7

REGULAR ARTICLE

Mixtures of regressions using matrix-variate heavy-tailed distributions

Salvatore D. Tomarchio[1] · **Michael P. B. Gallaugher**[2]

Received: 29 January 2023 / Accepted: 13 February 2024

Abstract
Finite mixtures of regressions (FMRs) are powerful clustering devices used in many regression-type analyses. Unfortunately, real data often present atypical observations that make the commonly adopted normality assumption of the mixture components inadequate. Thus, to robustify the FMR approach in a matrix-variate framework, we introduce ten FMRs based on the matrix-variate t and contaminated normal distributions. Furthermore, once one of our models is estimated and the observations are assigned to the groups, different procedures can be used for the detection of the atypical points in the data. An ECM algorithm is outlined for maximum likelihood parameter estimation. By using simulated data, we show the negative consequences (in terms of parameter estimates and inferred classification) of the wrong normality assumption in the presence of heavy-tailed clusters or noisy matrices. Such issues are properly addressed by our models instead. Additionally, over the same data, the atypical points detection procedures are also investigated. A real-data analysis concerning the relationship between greenhouse gas emissions and their determinants is conducted, and the behavior of our models in the presence of heterogeneity and atypical observations is discussed.

Keywords Matrix-variate · Mixture models · Heavy-tailed distributions · Model-based clustering

Mathematics Subject Classification 62H30 · 68T10

✉ Salvatore D. Tomarchio
daniele.tomarchio@unict.it

1 Department of Economics and Business, University of Catania, Catania, Italy

2 Department of Statistical Science, Baylor University, Waco, TX, USA

Published online: 16 March 2024

1 Introduction

Nowadays, the availability of data having complex structures is becoming prevalent, driven by rapid improvements in computing power and data storage capabilities. This explosion of data leads to a growing demand for appropriate new methodologies of statistical analysis. Although complex data structures can take different formats, the matrix-variate (or three-way) form factor is increasingly attracting interest in the statistical literature, especially within a model-based clustering perspective. On a related note, a non-exhaustive list of examples is given by the works of Viroli (2011a, b), Doğru et al. (2016), Gallaugher and McNicholas (2018, 2019), Melnykov and Zhu (2018), Tomarchio et al. (2020), Sarkar et al. (2020), Tomarchio (2022) and Zhu et al. (2022).

A matrix-variate dataset is characterized by three modes, namely P rows, R columns, and N layers. In short, it can be arranged into a $P \times R \times N$ array. Depending on the entity indexed in each of the three modes of the array, different examples may be considered such as image recognition data (Gallaugher and McNicholas 2018, 2020), two-factor data (Melnykov and Zhu 2018), multivariate longitudinal data (Tomarchio et al. 2020, 2022) or multivariate spatio-temporal data (Viroli 2011b).

An alternative way to analyze a matrix-variate dataset is to consider its vectorization and perform multivariate techniques. However, as well documented in the literature, this process poses the following issues that are here briefly summarized. First, it increases the number of parameters to be estimated. Indeed, vectorization would produce PR-dimensional vectors, therefore resulting in $PR(PR+1)/2$ free scale parameters when using multivariate analysis, compared to the $P(P+1)/2 + R(R+1)/2 - 1$ of the matrix-variate methodologies. Secondly, it has been shown to have negative effects on model selection (see, e.g. Sarkar et al. 2020; Tomarchio et al. 2021). Finally, from an interpretative standpoint, it reduces the data interpretability, given that the identification of the two sources of variability governed by the two scale matrices would not be possible (see, e.g. Anderlucci et al. 2014; Melnykov and Zhu 2019; Tomarchio et al. 2021). Thus, models capable of using the inherent matrix-variate structure of the data are preferred.

It is often of interest to linearly regress a set of responses to a set of covariates. Indeed, useful insight can be gained by accounting for functional dependencies between the two sets of variables. With this in mind, some contributions have been introduced in the matrix-variate regression literature, based on the matrix-variate normal (MVN) distribution or for the analysis of skewed data (see, e.g. Viroli 2012; Melnykov and Zhu 2019; Tomarchio et al. 2021; Punzo and Tomarchio 2022; Gallaugher et al. 2022). By focusing on the model-based clustering framework, as we will do throughout the manuscript, matrix-variate regressions can be generally split into two main categories that differ depending on how covariates enter into the model. Specifically, in the first category, the observed covariates represent fixed effects shared by all the units in the same cluster; this leads to finite mixture of regressions with fixed covariates (FMR-FCs). Conversely, in the second category, the observed covariates are treated as random, and information about their distribution is included in the model and used for clustering purposes; this produces finite mixture of regressions with random covariates (FMR-RCs). Both types of models have advantages and disadvantages.

When the distribution of the covariates does not show a clustering structure, FMR-FCs should be generally preferred since they are more parsimonious (Tomarchio et al. 2021). On the other hand, ignoring the distribution of the covariates when it presents an underlying structure, can lead to misleading classification and inferential results (Punzo et al. 2021). Therefore, FMR-RCs should be preferred despite their lower parsimony.

In this manuscript, we introduce a family of two matrix-variate FMR-FCs (MV-FMR-FCs) and a family of eight matrix-variate FMR-RCs (MV-FMR-RCs) by using the matrix-variate t (MVT), contaminated-normal (MVCN), and MVN distributions (the latter used for FMR-RCs only). Compared to existing contributions in the matrix-variate regression literature, our families of models have several practical advantages. In detail, when compared to models based on the MVN distribution (see, e.g. Tomarchio et al. 2021), they allow for a better accommodation of mildly atypical points, since the MVT and MVCN distributions have heavier-than-normal tails. Real data often present atypical points that, if not properly accommodated, could affect classification and inference on model parameters, with particular interest to the regression coefficients (Hossain and Naik 1991). Accordingly, the use of the MVT and MVCN distributions provides a more robust estimation method (Doğru et al. 2016; Tomarchio et al. 2022). Anyway, in our family of MV-FMR-RCs, we also consider the MVN distribution for modeling scenarios where only one of the two sets of variables exhibits atypical points.

When our families of models are compared to those in Gallaugher et al. (2022), which are based on skewed matrix-variate distributions, we note three advantages:

1. Parameter interpretation is easier when using the models herein introduced. This can be observed by examining the mean-variance representation of the skewed matrix-variate distributions in Gallaugher et al. (2022) and by extending the results of McNeil et al. (2015) to the matrix-variate framework. For instance, in the MVT and MVCN distributions, the location parameter corresponds to the mean, unlike in the skewed cases. Similarly, the tail behavior can be directly controlled by the concentration parameters of the MVT and MVCN distributions, whereas in the skewed cases, it is not as straightforward due to the influence of the skewness on the shape of the distributions.
2. Parameter estimation is computationally simpler using the models herein introduced. As discussed in Gallaugher and McNicholas (2018), some computational issues can be encountered when using the considered skewed matrix-variate distributions, because of the calculations related to the Bessel functions and corresponding partial derivatives. Despite some computational tricks that can be implemented to circumvent these issues, their success depends on the specific data at hand. On the other hand, estimation using the MVT and MVCN distributions is free of these problems.
3. Parsimony is greater when using the models herein introduced herein, although this comes at the cost of not modeling the possible skewness of the data.

An additional advantage of our proposal, compared to both normal and skewed-based approaches, relies on the capability of detecting atypical observations. As we will discuss throughout the manuscript, *a posteriori* procedures (i.e. taking place once a

model is fitted) can be straightforwardly adopted for such purpose. This aspect is of particular importance for matrix-variate data given that visualization techniques - and, therefore, the visual detection of atypical matrices - is a challenging task.

The manuscript is organized as follows. In Sect. 2, the two MV-FMR-FCs and the eight MV-FMR-RCs are introduced. We also illustrate an expectation conditional-maximization (ECM) algorithm (Meng and Rubin 1993) to conveniently estimate the parameters of our models. A discussion about how robustness is ensured by our models and how the detection of mildly atypical observations is achieved is also here presented. Section 3 presents the results conducted by using our models on artificial data in terms of parameter recovery, classification performance, atypical points detection, and model selection. In Sect. 4, we analyze a real dataset concerning the relationship between the main greenhouse gases (GHG) and their principal determinants for 159 countries over the last 5 years of available data. Section 5 summarizes our manuscript.

2 Methodology

2.1 Matrix-variate heavy-tailed regression mixtures

As introduced in Sect. 1, we can distinguish between two types of matrix-variate regressions: FMR-FCs and FMR-RCs. By starting with the first category, let $\mathcal{Y}$ be a continuous random matrix of dimension $P \times R$ containing P responses measured over R occasions. Let us also consider a random matrix $\mathcal{X}$ of dimension $Q \times R$ containing Q covariates evaluated over R occasions. Assume there exist K subgroups in the data. Then, the probability density function (pdf) of an MV-FMR-FC is

$$h(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}) = \sum_{k=1}^{K} \pi_k f(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{Y}|k}), \tag{1}$$

where $\pi_k > 0$ is the mixing proportion, with $\sum_{k=1}^{K} \pi_k = 1$, $f(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{Y}|k})$ is the conditional distribution of the responses, and $\boldsymbol{\Theta} = \{\pi_k, \boldsymbol{\Theta}_{\mathbf{Y}|k}\}_{k=1}^{K}$. Additionally, for each k, the mean parameter in $\boldsymbol{\Theta}_{\mathbf{Y}|k}$ is a linear function of $\mathbf{X}$ depending on some other parameters.

In this manuscript, $f(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{Y}|k})$ in (1) can have the functional form of either the MVCN or MVT distribution. A generic $P \times R$ random matrix $\mathcal{H}$ follows an MVCN distribution if its density function is

$$\alpha\phi(\mathbf{H}; \mathbf{M}, \Sigma, \Psi) + (1 - \alpha)\phi(\mathbf{H}; \mathbf{M}, \eta\Sigma, \Psi), \tag{2}$$

where $\phi(\mathbf{H}; \mathbf{M}, \Sigma, \Psi)$ is the density of the MVN distribution having pdf

$$\frac{1}{(2\pi)^{\frac{PR}{2}} |\Sigma|^{\frac{R}{2}} |\Psi|^{\frac{P}{2}}} \exp\left[-\frac{\delta(\mathbf{H}; \mathbf{M}, \Sigma, \Psi)}{2}\right], \tag{3}$$

with $P \times R$ mean matrix $\mathbf{M}$, $P \times P$ and $R \times R$ covariance matrices Σ and Ψ, respectively, and squared Mahalanobis distance $\delta(\mathbf{H}; \mathbf{M}, \Sigma, \Psi) = \text{tr}\left[\Sigma^{-1}(\mathbf{H}-\mathbf{M})\Psi^{-1}(\mathbf{H}-\mathbf{M})'\right]$. Additionally, in (2), α is the proportion of typical points in the group k, and η is an inflation parameter accounting for the degree of contamination in the group k. In symbols, $\mathbf{H} \sim \mathbb{CN}(\mathbf{M}, \Sigma, \Psi, \alpha, \eta)$. Note that (2) approaches (3) as $\alpha \to 1^-$ and $\eta \to 1^+$. Further details about the MCVN distribution can be found in Tomarchio et al. (2022).

A generic $P \times R$ random matrix $\mathcal{H}$ follows an MVT distribution if its density function is

$$\frac{|\Sigma|^{-\frac{R}{2}}|\Psi|^{-\frac{P}{2}}\Gamma(\frac{PR+\nu}{2})}{(\pi\nu)^{\frac{PR}{2}}\Gamma(\frac{\nu}{2})}\left\{1+\frac{\delta(\mathbf{H};\mathbf{M},\Sigma,\Psi)}{\nu}\right\}^{-\frac{PR+\nu}{2}}, \tag{4}$$

with $P \times R$ mean matrix $\mathbf{M}$, $P \times P$ and $R \times R$ scale matrices Σ and Ψ, respectively, and degree of freedom ν. In symbols, $\mathbf{H} \sim \mathbb{T}(\mathbf{M}, \Sigma, \Psi, \nu)$. Note that (4) approaches (3) as $\nu \to \infty$. Further details about the MVT distribution can be found in Doğru et al. (2016).

Regardless of the distribution considered, we assume a linear relationship $\mathbf{M}(\mathbf{X}^*; \mathbf{B}_k) = \mathbf{B}_k\mathbf{X}^*$, where $\mathbf{B}_k$ is a $P \times (1+Q)$ matrix of regression coefficients and $\mathbf{X}^*$ is a $(1+Q) \times R$ matrix containing a first row of ones (to incorporate the intercept in the model) and the covariates $\mathbf{X}$. Furthermore, to address a well-known identifiability issue of both distributions, we set the first diagonal element of Σ to 1, in the fashion of Gallaugher and McNicholas (2017, 2018); Tomarchio et al. (2021). The use of these two conditional distributions gives rise to two different MV-FMR-FCs, respectively abbreviated as MVT-FMR-FC and MVCN-FMR-FC.

When the distribution of the covariates is explicitly included in the model, the FMR-RCs are obtained. In detail, the pdf of an MV-FMR-RC is

$$h(\mathbf{Y}, \mathbf{X}; \boldsymbol{\Theta}) = \sum_{k=1}^{K} \pi_k f(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{Y}|k}) g(\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{X}|k}), \tag{5}$$

where $g(\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{X}|k})$ represents the marginal distribution of the covariates, and now $\boldsymbol{\Theta} = \{\pi_k, \boldsymbol{\Theta}_{\mathbf{Y}|k}, \boldsymbol{\Theta}_{\mathbf{X}|k}\}_{k=1}^{K}$. Thus, model (5) decomposes the joint distribution of responses and covariates into the product between the marginal distribution of the covariates and the conditional distribution of the responses. This implies that the assignment of data points to clusters is conducted by also using the information provided by the marginal distribution of the covariates.

In this manuscript, we use the densities in Eqs. (2) to (4) for $f(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{Y}|k})$ and $g(\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{X}|k})$ in (5). By combining the three possible options, we obtain a family of nine MV-FMR-RCs, eight of which are herein introduced (for details on the MV-FRM-RC based on the MVN for both $f(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{Y}|k})$ and $g(\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{X}|k})$, see Tomarchio et al. 2021). Our family of models is flexible enough to cope with scenarios where both the responses and the covariates present atypical observations, or in which only one of the two sets of variables presents such characteristics. For notational clarity,

each model is labeled by separating with a dash the acronyms used for $f(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{Y}|k})$ and $g(\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{X}|k})$, respectively. For example, if we consider an MV-FMR-RC having an MVT distribution for $\mathbf{Y}|\mathbf{X}$ and an MVCN distribution for $\mathbf{X}$, it is referred to as MVT-MVCN-FMR-RC.

2.2 Maximum likelihood estimation

Parameter estimation for both categories of models is carried out via the ECM algorithm, a generalization of the well-known expectation-maximization (EM) algorithm (Dempster et al. 1977). For a better organization of this section, we limit our discussion to the ECM algorithm of the MV-FMR-RCs because that of the MV-FMR-FCs can be obtained as a by-product of it.

Let $\mathbf{S} = \{(\mathbf{Y}_i, \mathbf{X}_i)\}_{i=1}^N$ be a sample of N independent observations. Within an ECM framework, $\mathbf{S}$ is viewed as incomplete. Indeed, several sources of incompleteness need to be considered: one of them is common among the models, while the others are model-dependent.

The common source arises from the fact that we do not know, for each observation, its component membership. To address this aspect, we use an indicator vector $\mathbf{z}_i = (z_{ik}, \ldots, z_{iK})$, where $z_{ik} = 1$ if observation i is in group k, and $z_{ik} = 0$ otherwise. Notice that, $\mathbf{Z}_i$ is the random counterpart of $\mathbf{z}_i$, and it is distributed according to a multinomial distribution consisting of one draw from K categories with probabilities $\boldsymbol{\pi} = \{\pi_k\}_{k=1}^K$.

Based on this first source of incompleteness, we can write a sort of first-level complete-data log-likelihood as

$$l_c(\boldsymbol{\Theta}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \ln(\pi_k) + l_{c_{\mathbf{Y}}}(\boldsymbol{\Theta}_{\mathbf{Y}}) + l_{c_{\mathbf{X}}}(\boldsymbol{\Theta}_{\mathbf{X}}), \tag{6}$$

where

$$l_{c_{\mathbf{Y}}}(\boldsymbol{\Theta}_{\mathbf{Y}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \ln\big[f(\mathbf{Y}_i|\mathbf{X}_i; \boldsymbol{\Theta}_{\mathbf{Y}|k})\big], \tag{7}$$

$$l_{c_{\mathbf{X}}}(\boldsymbol{\Theta}_{\mathbf{X}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \ln\big[g(\mathbf{X}_i; \boldsymbol{\Theta}_{\mathbf{X}|k})\big]. \tag{8}$$

We discuss the forms of (7) and (8) in Sects. 2.2.1 and 2.2.2, respectively.

2.2.1 Conditional distributions of the responses

As discussed in Sect. 2.1, we allow $f(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{Y}|k})$ to be the densities in Eqs. (2) to (4). By starting with the MVCN distribution, we recall that it can be represented as a scale mixture of MVN and Bernoulli distributions (Tomarchio et al. 2022). Thus, for each observation $(\mathbf{Y}_i, \mathbf{X}_i)$ in the group k, we have a second source of incompleteness

denoted by $V_{ik\mathbf{Y}} \sim$ Bernoulli $(\alpha_{\mathbf{Y}|k})$. This allows the writing of (7) in terms of a second-level complete-data log-likelihood as

$$l_{c_{\mathbf{Y}}}(\boldsymbol{\Theta}_{\mathbf{Y}}) = l_{c_{\mathbf{Y}1}}(\Phi_{\mathbf{Y}}) + l_{c_{\mathbf{Y}2}}(\boldsymbol{\alpha}_{\mathbf{Y}}),$$

where

$$\begin{aligned} l_{c_{\mathbf{Y}1}}(\Phi_{\mathbf{Y}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\Bigg[& -\frac{R}{2}\ln|\Sigma_{\mathbf{Y}|k}| - \frac{P}{2}\ln|\Psi_{\mathbf{Y}|k}| - \frac{PR}{2}(1-v_{ik\mathbf{Y}})\ln(\eta_{\mathbf{Y}|k}) \\ & -\frac{1}{2}(v_{ik\mathbf{Y}} + \frac{1-v_{ik\mathbf{Y}}}{\eta_{\mathbf{Y}|k}})\delta_k\left(\mathbf{Y}_i; \mathbf{B}_k\mathbf{X}_i^*, \Sigma_{\mathbf{Y}|k}, \Psi_{\mathbf{Y}|k}\right)\Bigg], \\ l_{c_{\mathbf{Y}2}}(\boldsymbol{\alpha}_{\mathbf{Y}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\Big[& v_{ik\mathbf{Y}}\ln(\alpha_{\mathbf{Y}|k}) + (1-v_{ik\mathbf{Y}})\ln(1-\alpha_{\mathbf{Y}|k})\Big], \end{aligned}$$

with $\Phi_{\mathbf{Y}} = \{\mathbf{B}_k, \Sigma_{\mathbf{Y}|k}, \Psi_{\mathbf{Y}|k}, \eta_{\mathbf{Y}|k}\}_{k=1}^{K}$, $\boldsymbol{\alpha}_{\mathbf{Y}} = \{\alpha_{\mathbf{Y}|k}\}_{k=1}^{K}$, and $\boldsymbol{\Theta}_{\mathbf{Y}} = \{\Phi_{\mathbf{Y}}, \boldsymbol{\alpha}_{\mathbf{Y}}\}$.

When the MVT distribution is considered, we recall that it can be represented in terms of a scale mixture of MVN and gamma distributions (Doğru et al. 2016). In practice, for each observation $(\mathbf{Y}_i, \mathbf{X}_i)$ in the group k, we have a second source of incompleteness denoted by $W_{ik\mathbf{Y}} \sim$ Gamma $(\nu_{\mathbf{Y}|k}/2, \nu_{\mathbf{Y}|k}/2)$. This leads the writing of (7) in terms of a second-level complete-data log-likelihood as

$$l_{c_{\mathbf{Y}}}(\boldsymbol{\Theta}_{\mathbf{Y}}) = l_{c_{\mathbf{Y}1}}(\Phi_{\mathbf{Y}}) + l_{c_{\mathbf{Y}2}}(\boldsymbol{\nu}_{\mathbf{Y}}),$$

where

$$\begin{aligned} l_{c_{\mathbf{Y}1}}(\Phi_{\mathbf{Y}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\Bigg[& -\frac{PR}{2}\ln(2\pi) - \frac{R}{2}\ln|\Sigma_{\mathbf{Y}|k}| - \frac{P}{2}\ln|\Psi_{\mathbf{Y}|k}| + \frac{PR}{2}\ln(w_{ik\mathbf{Y}}) \\ & -\frac{w_{ik\mathbf{Y}}}{2}\delta_k\left(\mathbf{Y}_i; \mathbf{B}_k\mathbf{X}_i^*, \Sigma_{\mathbf{Y}|k}, \Psi_{\mathbf{Y}|k}\right)\Bigg], \\ l_{c_{\mathbf{Y}2}}(\boldsymbol{\nu}_{\mathbf{Y}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\Bigg\{ & \frac{\nu_{\mathbf{Y}|k}}{2}\ln\left(\frac{\nu_{\mathbf{Y}|k}}{2}\right) + \left(\frac{\nu_{\mathbf{Y}|k}}{2}-1\right)\ln(w_{ik\mathbf{Y}}) - \frac{\nu_{\mathbf{Y}|k}}{2}w_{ik\mathbf{Y}} \\ & -\ln\left[\Gamma\left(\frac{\nu_{\mathbf{Y}|k}}{2}\right)\right]\Bigg\}, \end{aligned}$$

with $\Phi_{\mathbf{Y}} = \{\mathbf{B}_k, \Sigma_{\mathbf{Y}|k}, \Psi_{\mathbf{Y}|k}\}_{k=1}^{K}$, $\boldsymbol{\nu}_{\mathbf{Y}} = \{\nu_{\mathbf{Y}|k}\}_{k=1}^{K}$, and $\boldsymbol{\Theta}_{\mathbf{Y}} = \{\Phi_{\mathbf{Y}}, \boldsymbol{\nu}_{\mathbf{Y}}\}$.

Finally, when the MVN distribution is used, there are no other sources of incompleteness. Therefore, we can write

$$\begin{aligned} l_{c_{\mathbf{Y}}}(\boldsymbol{\Theta}_{\mathbf{Y}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\Bigg[& -\frac{PR}{2}\ln(2\pi) - \frac{R}{2}\ln|\Sigma_{\mathbf{Y}|k}| - \frac{P}{2}|\Psi_{\mathbf{Y}|k}| \\ & -\frac{1}{2}\delta_k\left(\mathbf{X}_i; \mathbf{B}_k\mathbf{X}_i^*, \Sigma_{\mathbf{Y}|k}, \Psi_{\mathbf{Y}|k}\right)\Bigg], \end{aligned}$$

with $\boldsymbol{\Theta}_{\mathbf{Y}} = \{\mathbf{B}_k, \Sigma_{\mathbf{Y}|k}, \Psi_{\mathbf{Y}|k}\}$.

2.2.2 Marginal distributions of the covariates

Similarly to Sect. 2.2.1, we allow $g(\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{X}|k})$ to be the densities in Eqs. (2) to (4). By starting with MVCN distribution, and considering its scale mixture representation mentioned in Sect. 2.2.1, we have a third source of incompleteness defined by $V_{ik\mathbf{X}} \sim$ Bernoulli $(\alpha_{\mathbf{X}|k})$. Thus, we can write (8) in terms of a third-level complete-data log-likelihood as

$$l_{c_{\mathbf{X}}}(\boldsymbol{\Theta}_{\mathbf{X}}) = l_{c_{\mathbf{X}1}}(\Phi_{\mathbf{X}}) + l_{c_{\mathbf{X}2}}(\boldsymbol{\alpha}_{\mathbf{X}}),$$

where

$$\begin{aligned} l_{c_{\mathbf{X}1}}(\Phi_{\mathbf{X}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\Bigg[& -\frac{R}{2}\ln|\Sigma_{\mathbf{X}|k}| - \frac{Q}{2}\ln|\Psi_{\mathbf{X}|k}| - \frac{QR}{2}(1 - v_{ik\mathbf{X}})\ln(\eta_{\mathbf{X}|k}) \\ & -\frac{1}{2}(v_{ik\mathbf{X}} + \frac{1 - v_{ik\mathbf{X}}}{\eta_{\mathbf{X}|k}})\delta_k\big(\mathbf{X}_i; \mathbf{M}_k, \Sigma_{\mathbf{X}|k}, \Psi_{\mathbf{X}|k}\big)\Bigg], \\ l_{c_{\mathbf{X}2}}(\boldsymbol{\alpha}_{\mathbf{X}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\big[& v_{ik\mathbf{X}}\ln(\alpha_{\mathbf{X}|k}) + (1 - v_{ik\mathbf{X}})\ln(1 - \alpha_{\mathbf{X}|k})\big], \end{aligned}$$

with $\Phi_{\mathbf{X}} = \{\mathbf{M}_k, \Sigma_{\mathbf{X}|k}, \Psi_{\mathbf{X}|k}, \eta_{\mathbf{X}|k}\}_{k=1}^{K}$, $\boldsymbol{\alpha}_{\mathbf{X}} = \{\alpha_{\mathbf{X}|k}\}_{k=1}^{K}$, and $\boldsymbol{\Theta}_{\mathbf{X}} = \{\Phi_{\mathbf{X}}, \boldsymbol{\alpha}_{\mathbf{X}}\}$.

When the MVT distribution is considered, and by recalling its scale mixture representation of Sect. 2.2.1, we have a third source of incompleteness defined by $W_{ik\mathbf{X}} \sim$ Gamma $(\nu_{\mathbf{X}|k}/2, \nu_{\mathbf{X}|k}/2)$. Then, we can write (8) in terms of a third-level complete-data log-likelihood as

$$l_{c_{\mathbf{X}}}(\boldsymbol{\Theta}_{\mathbf{X}}) = l_{c_{\mathbf{X}1}}(\Phi_{\mathbf{X}}) + l_{c_{\mathbf{X}2}}(\boldsymbol{\nu}_{\mathbf{X}}),$$

where

$$\begin{aligned} l_{c_{\mathbf{X}1}}(\Phi_{\mathbf{X}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\Bigg[& -\frac{QR}{2}\ln(2\pi) - \frac{R}{2}\ln|\Sigma_{\mathbf{X}|k}| - \frac{Q}{2}\ln|\Psi_{\mathbf{X}|k}| + \frac{QR}{2}\ln(w_{ik\mathbf{X}}) \\ & -\frac{w_{ik\mathbf{X}}}{2}\delta_k\big(\mathbf{X}_i; \mathbf{M}_k, \Sigma_{\mathbf{X}|k}, \Psi_{\mathbf{X}|k}\big)\Bigg], \\ l_{c_{\mathbf{X}2}}(\boldsymbol{\nu}_{\mathbf{X}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\Bigg\{ & \frac{\nu_{\mathbf{X}|k}}{2}\ln\Big(\frac{\nu_{\mathbf{X}|k}}{2}\Big) + \Big(\frac{\nu_{\mathbf{X}|k}}{2} - 1\Big)\ln(w_{ik\mathbf{X}}) - \frac{\nu_{\mathbf{X}|k}}{2}w_{ik\mathbf{X}} \\ & - \ln\Big[\Gamma\Big(\frac{\nu_{\mathbf{X}|k}}{2}\Big)\Big]\Bigg\}, \end{aligned}$$

with $\Phi_{\mathbf{X}} = \{\mathbf{M}_k, \Sigma_{\mathbf{X}|k}, \Psi_{\mathbf{X}|k}\}_{k=1}^{K}$, $\boldsymbol{\nu}_{\mathbf{X}} = \{\nu_{\mathbf{X}|k}\}_{k=1}^{K}$, and $\boldsymbol{\Theta}_{\mathbf{X}} = \{\Phi_{\mathbf{X}}, \boldsymbol{\nu}_{\mathbf{X}}\}$.

Lastly, when the MVN distribution is used, there are no other sources of incompleteness. Therefore, we can write

$$l_{c_{\mathbf{X}}}(\boldsymbol{\Theta}_{\mathbf{X}}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\Bigg[-\frac{QR}{2}\ln(2\pi) - \frac{R}{2}\ln|\Sigma_{\mathbf{X}|k}| - \frac{Q}{2}|\Psi_{\mathbf{X}|k}| - \frac{1}{2}\delta_k\big(\mathbf{X}_i; \mathbf{M}_k, \Sigma_{\mathbf{X}|k}, \Psi_{\mathbf{X}|k}\big)\Bigg],$$

where $\boldsymbol{\Theta}_{\mathbf{X}} = \{\mathbf{M}_k, \Sigma_{\mathbf{X}|k}, \Psi_{\mathbf{X}|k}\}$.

2.2.3 MV-FMR-RCs: ECM algorithm

By generalizing the results of Doğru et al. (2016) and Tomarchio et al. (2022), the ECM algorithm of our MV-FMR-RCs has the following structure: an E-Step and two CM-Steps. Notice that, the majority of updates are in common among the models, whereas others refer to a specific model. Notationally, the parameters that will be marked with one dot represent the updates at the previous iteration, whereas those marked with two dots are the updates at the current iteration.

E-Step

At the E-step, we have to update the posterior probability that a point $(\mathbf{Y}_i, \mathbf{X}_i)$ belongs to the kth component of the considered model, i.e.

$$\ddot{z}_{ik} := E_{\boldsymbol{\Theta}}(Z_{ik}|\mathbf{Y}_i, \mathbf{X}_i) = \frac{\dot{\pi}_k f(\mathbf{Y}_i|\mathbf{X}_i; \dot{\boldsymbol{\Theta}}_{\mathbf{Y}|k}) g(\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{X}|k})}{\sum_{j=1}^{K}\dot{\pi}_j f(\mathbf{Y}_i|\mathbf{X}_i; \dot{\boldsymbol{\Theta}}_{\mathbf{Y}|j}) g(\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{X}|j})}. \tag{9}$$

Then, if $\mathbf{Y}_i|\mathbf{X}_i \sim \mathbb{T}(\mathbf{B}\mathbf{X}_i^*, \Sigma, \Psi, \nu)$, we also have to update

$$\ddot{w}_{ik\mathbf{Y}} := E_{\boldsymbol{\Theta}}(W_{ik\mathbf{Y}}|\mathbf{Y}_i, \mathbf{X}_i, \mathbf{z}_i) = \frac{PR + \dot{\nu}_{\mathbf{Y}|k}}{\dot{\nu}_{\mathbf{Y}|k} + \dot{\delta}_k\big(\mathbf{Y}_i; \dot{\mathbf{B}}_k\mathbf{X}_i^*, \dot{\Sigma}_{\mathbf{Y}|k}, \dot{\Psi}_{\mathbf{Y}|k}\big)},$$

$$\ddot{m}_{ik\mathbf{Y}} := E_{\boldsymbol{\Theta}}[\ln(W_{ik\mathbf{Y}})|\mathbf{Y}_i, \mathbf{X}_i, \mathbf{z}_i] = \varphi\left(\frac{PR + \dot{\nu}_{\mathbf{Y}|k}}{2}\right) - \ln\left\{\frac{1}{2}\big[\dot{\nu}_{\mathbf{Y}|k} + \dot{\delta}_k\big(\mathbf{Y}_i; \dot{\mathbf{B}}_k\mathbf{X}_i^*, \dot{\Sigma}_{\mathbf{Y}|k}, \dot{\Psi}_{\mathbf{Y}|k}\big)\big]\right\},$$

with $\varphi(\cdot)$ denoting the digamma function, while if $\mathbf{Y}_i|\mathbf{X}_i \sim \mathbb{CN}(\mathbf{B}\mathbf{X}_i^*, \Sigma, \Psi, \alpha, \eta)$, we also need to update

$$\ddot{v}_{ik\mathbf{Y}} := E_{\boldsymbol{\Theta}}(V_{ik\mathbf{Y}}|\mathbf{Y}_i, \mathbf{X}_i, \mathbf{z}_i) = \frac{\dot{\alpha}_{\mathbf{Y}|k}\phi(\mathbf{Y}_i|\mathbf{X}_i; \dot{\mathbf{B}}_k\mathbf{X}_i^*, \dot{\Sigma}_{\mathbf{Y}|k}, \dot{\Psi}_{\mathbf{Y}|k})}{\xi(\mathbf{Y}_i|\mathbf{X}_i; \dot{\mathbf{B}}_k\mathbf{X}_i^*, \dot{\Sigma}_{\mathbf{Y}|k}, \dot{\Psi}_{\mathbf{Y}|k}, \dot{\eta}_{\mathbf{Y}|k})},$$

with $\xi(\cdot)$ denoting the pdf in (2).

As concerns the covariates, if $\mathbf{X}_i \sim \mathbb{T}(\mathbf{M}, \Sigma, \Psi, \nu)$, we have to compute

$$\ddot{w}_{ik\mathbf{X}} := E_{\boldsymbol{\Theta}}(W_{ik\mathbf{X}}|\mathbf{X}_i, \mathbf{z}_i) = \frac{QR + \dot{\nu}_{\mathbf{X}|k}}{\dot{\nu}_{\mathbf{X}|k} + \dot{\delta}_k\big(\mathbf{X}_i; \dot{\mathbf{M}}_k, \dot{\Sigma}_{\mathbf{X}|k}, \dot{\Psi}_{\mathbf{X}|k}\big)},$$

$$\begin{aligned}\ddot{m}_{ik\mathbf{X}} &:= E_{\boldsymbol{\Theta}}[\ln(W_{ik\mathbf{X}})|\mathbf{X}_i, \mathbf{z}_i] \\ &= \varphi\left(\frac{QR+\dot{\nu}_{\mathbf{X}|k}}{2}\right) - \ln\left\{\frac{1}{2}\left[\dot{\nu}_{\mathbf{X}|k} + \dot{\delta}_k\left(\mathbf{X}_i; \dot{\mathbf{M}}_k, \dot{\Sigma}_{\mathbf{X}|k}, \dot{\Psi}_{\mathbf{X}|k}\right)\right]\right\},\end{aligned}$$

while if $\mathbf{X}_i \sim \mathbb{CN}(\mathbf{M}, \Sigma, \Psi, \alpha, \eta)$, we have to calculate

$$\ddot{v}_{ik\mathbf{X}} := E_{\boldsymbol{\Theta}}(V_{ik\mathbf{X}}|, \mathbf{X}_i, \mathbf{z}_i) = \frac{\dot{\alpha}_{\mathbf{X}|k}\phi(\mathbf{X}_i; \dot{\mathbf{M}}_k, \dot{\Sigma}_{\mathbf{X}|k}, \dot{\Psi}_{\mathbf{X}|k})}{\xi(\mathbf{X}_i; \dot{\mathbf{M}}_k, \dot{\Sigma}_{\mathbf{X}|k}, \dot{\Psi}_{\mathbf{X}|k}, \dot{\eta}_{\mathbf{X}|k})}.$$

CM-Step 1

For notational simplicity, let us define the variables $\ddot{U}_{ik\mathbf{Y}}$ and $\ddot{U}_{ik\mathbf{X}}$, such that

$$\ddot{u}_{ik\mathbf{Y}} = \begin{cases} \ddot{w}_{ik\mathbf{Y}} & \text{if } \mathbf{Y}_i|\mathbf{X}_i \sim \mathbb{T}(\mathbf{B}\mathbf{X}_i^*, \Sigma, \Psi, \nu) \\ \ddot{v}_{ik\mathbf{Y}} + \frac{1-\ddot{v}_{ik\mathbf{Y}}}{\dot{\eta}_{\mathbf{Y}|k}} & \text{if } \mathbf{Y}_i|\mathbf{X}_i \sim \mathbb{CN}(\mathbf{B}\mathbf{X}_i^*, \Sigma, \Psi, \alpha, \eta) \end{cases}, \tag{10}$$

and

$$\ddot{u}_{ik\mathbf{X}} = \begin{cases} \ddot{w}_{ik\mathbf{X}} & \text{if } \mathbf{X}_i \sim \mathbb{T}(\mathbf{M}, \Sigma, \Psi, \nu) \\ \ddot{v}_{ik\mathbf{X}} + \frac{1-\ddot{v}_{ik\mathbf{X}}}{\dot{\eta}_{\mathbf{X}|k}} & \text{if } \mathbf{X}_i \sim \mathbb{CN}(\mathbf{M}, \Sigma, \Psi, \alpha, \eta) \end{cases}. \tag{11}$$

Then, at the first CM-Step, we have the following updates

$$\ddot{\pi}_k = \frac{\sum_{i=1}^N \ddot{z}_{ik}}{N}, \quad \ddot{\alpha}_{\mathbf{Y}|k} = \frac{\sum_{i=1}^N \ddot{z}_{ik}\ddot{v}_{ik\mathbf{Y}}}{\sum_{i=1}^N \ddot{z}_{ik}}, \quad \ddot{\alpha}_{\mathbf{X}|k} = \frac{\sum_{i=1}^N \ddot{z}_{ik}\ddot{v}_{ik\mathbf{X}}}{\sum_{i=1}^N \ddot{z}_{ik}},$$

$$\ddot{\mathbf{B}}_k = \left[\sum_{i=1}^N \ddot{z}_{ik}\ddot{u}_{ik\mathbf{Y}}\mathbf{Y}_i\dot{\Psi}_{\mathbf{Y}|k}^{-1}\mathbf{X}_i^{*\prime}\right]\left[\sum_{i=1}^N \ddot{z}_{ik}\ddot{u}_{ik\mathbf{Y}}\mathbf{X}_i^*\dot{\Psi}_{\mathbf{Y}|k}^{-1}\mathbf{X}_i^{*\prime}\right]^{-1}, \tag{12}$$

$$\ddot{\mathbf{M}}_k = \frac{\sum_{i=1}^N \ddot{z}_{ik}\ddot{u}_{ik\mathbf{X}}\mathbf{X}_i}{\sum_{i=1}^N \ddot{z}_{ik}\ddot{u}_{ik\mathbf{X}}} \tag{13}$$

$$\ddot{\Sigma}_{\mathbf{Y}|k} = \frac{\sum_{i=1}^N \ddot{z}_{ik}\ddot{u}_{ik\mathbf{Y}}\left(\mathbf{Y}_i - \ddot{\mathbf{B}}_k\mathbf{X}_i^*\right)\dot{\Psi}_{\mathbf{Y}|k}^{-1}\left(\mathbf{Y}_i - \ddot{\mathbf{B}}_k\mathbf{X}_i^*\right)'}{R\sum_{i=1}^N \ddot{z}_{ik}}, \tag{14}$$

$$\ddot{\Sigma}_{\mathbf{X}|k} = \frac{\sum_{i=1}^N \ddot{z}_{ik}\ddot{u}_{ik\mathbf{X}}\left(\mathbf{X}_i - \ddot{\mathbf{M}}_k\right)\dot{\Psi}_{\mathbf{X}|k}^{-1}\left(\mathbf{X}_i - \ddot{\mathbf{M}}_k\right)'}{R\sum_{i=1}^N \ddot{z}_{ik}}. \tag{15}$$

A closed-form solution is not analytically available for the updates $\ddot{\nu}_{\mathbf{Y}|k}$ and $\ddot{\nu}_{\mathbf{X}|k}$ of $\nu_{\mathbf{Y}|k}$ and $\nu_{\mathbf{X}|k}$, respectively. Thus, they must be numerically obtained by respectively solving the following equations

$$\ln\left(\frac{\nu_{\mathbf{Y}|k}}{2}\right) + 1 - \varphi\left(\frac{\nu_{\mathbf{Y}|k}}{2}\right) + \frac{\sum_{i=1}^N \ddot{z}_{ik}(\ddot{m}_{ik\mathbf{Y}} - \ddot{w}_{ik\mathbf{Y}})}{\sum_{i=1}^N \ddot{z}_{ik}} = 0,$$

$$\ln\left(\frac{\nu_{\mathbf{X}|k}}{2}\right) + 1 - \varphi\left(\frac{\nu_{\mathbf{X}|k}}{2}\right) + \frac{\sum_{i=1}^{N} \ddot{z}_{ik}(\ddot{m}_{ik\mathbf{X}} - \ddot{w}_{ik\mathbf{X}})}{\sum_{i=1}^{N} \ddot{z}_{ik}} = 0.$$

Computationally, the `optimize()` function in the **stats** package for R, is used to perform a numerical search for the solutions to the above equations.

CM-Step 2 Now, we have the following updates

$$\ddot{\Psi}_{\mathbf{Y}|k} = \frac{\sum_{i=1}^{N} \ddot{z}_{ik}\ddot{u}_{ik\mathbf{Y}}\left(\mathbf{Y}_i - \ddot{\mathbf{B}}_k\mathbf{X}_i^*\right)'\ddot{\Sigma}_{\mathbf{Y}|k}^{-1}\left(\mathbf{Y}_i - \ddot{\mathbf{B}}_k\mathbf{X}_i^*\right)}{P\sum_{i=1}^{N} \ddot{z}_{ik}}, \tag{16}$$

$$\ddot{\Psi}_{\mathbf{X}|k} = \frac{\sum_{i=1}^{N} \ddot{z}_{ik}\ddot{u}_{ik\mathbf{X}}\left(\mathbf{X}_i - \ddot{\mathbf{M}}_k\right)'\ddot{\Sigma}_{\mathbf{X}|k}^{-1}\left(\mathbf{X}_i - \ddot{\mathbf{M}}_k\right)}{Q\sum_{i=1}^{N} \ddot{z}_{ik}},$$

$$\ddot{\eta}_{\mathbf{Y}|k} = \max\left\{\eta_{\mathbf{Y}|k}^{\min}, \frac{\sum_{i=1}^{N} \ddot{z}_{ik}(1-\ddot{v}_{ik\mathbf{Y}})\ddot{\delta}_k\left(\mathbf{Y}_i; \ddot{\mathbf{B}}_k\mathbf{X}_i^*, \ddot{\Sigma}_{\mathbf{Y}|k}, \ddot{\Psi}_{\mathbf{Y}|k}\right)}{PR\sum_{i=1}^{N} \ddot{z}_{ik}(1-\ddot{v}_{ik\mathbf{Y}})}\right\},$$

$$\ddot{\eta}_{\mathbf{X}|k} = \max\left\{\eta_{\mathbf{X}|k}^{\min}, \frac{\sum_{i=1}^{N} \ddot{z}_{ik}(1-\ddot{v}_{ik\mathbf{X}})\ddot{\delta}_k\left(\mathbf{X}_i; \ddot{\mathbf{M}}_k, \ddot{\Sigma}_{\mathbf{X}|k}, \ddot{\Psi}_{\mathbf{X}|k}\right)}{QR\sum_{i=1}^{N} \ddot{z}_{ik}(1-\ddot{v}_{ik\mathbf{X}})}\right\}, \tag{17}$$

where $\eta_{\mathbf{Y}|k}^{\min}$ and $\eta_{\mathbf{X}|k}^{\min}$ are the minimum values for $\ddot{\eta}_{\mathbf{Y}|k}$ and $\ddot{\eta}_{\mathbf{X}|k}$, respectively; in our analyses, we set $\eta_{\mathbf{Y}|k}^{\min} = \eta_{\mathbf{X}|k}^{\min} = 1.0001$ as in Tomarchio et al. (2022).

2.2.4 Computational aspects of the ECM algorithm

Choosing reliable starting values is an important aspect of any EM-based algorithm (Michael and Melnykov 2016). As pointed out by McLachlan and Peel (2000), the log-likelihood equation usually has multiple local maxima, and so the algorithm should be applied from a wide choice of starting values. To alleviate the impact that the initialization has in finding the global maximizer, we implement a short EM approach, in the fashion of Biernacki et al. (2003), Tomarchio et al. (2020). This procedure consists of H short runs of the algorithm from different random positions. The term "short" means that the algorithm is run for a very small number of iterations s, without waiting for convergence. In our analyses, we set $H = 100$ and $s = 1$. Then, the parameter set producing the largest log-likelihood value is used to initialize the algorithm.

Another aspect that is worth to be specified concerns the convergence criterion used to stop the algorithm. A commonly used criterion is to monitor the difference between the log-likelihood values on two consecutive iterations and stop the algorithm when this difference falls below a certain threshold τ. Herein, we adopt this criterion and set $\tau = 0.001$.

2.3 Robustness and detection of atypical observations

As introduced in Sect. 1, the use of the MVT and MVCN distributions allows for a robust estimation of the parameters of models (1) and (5). Robustness is attained because the estimates of $\mathbf{B}_k$, $\Sigma_{\mathbf{Y}|k}$, and $\Psi_{\mathbf{Y}|k}$, for MV-FMR-FCs and MV-FMR-RCs,

Table 1 Observation labeling according to the direction in which an observation is considered atypical

Direction of atypicality	X	
	Yes	No
Y		
Yes	Bad leverage	Outlier
No	Good leverage	Typical

as well as the estimates of $\mathbf{M}_k$, $\Sigma_{\mathbf{X}|k}$, and $\Psi_{\mathbf{X}|k}$ for MV-FMR-RCs, are weighted means where the weights are a function of the squared Mahalanobis distances δ. This is the underlying idea of the M-estimation (Maronna 1976), where a decreasing weighting function $t(\delta) : (0, +\infty) \rightarrow (0, +\infty)$ is used to down-weight the observations with large δ values (Punzo et al. 2021). For both distributions, the weights are represented by the u_{ik} values in (10) and (11), and their down-weighting effect is illustrated in Eqs. (12) to (17).

Another advantage of using models based on the MVT and MVCN distributions consists of the capability of detecting mildly atypical observations. This aspect is of particular importance for matrix-variate data given that their visualization is a challenging task. According to a common-to-regression taxonomy, an observation can be classified as typical, good leverage, bad leverage, and outlier (Rousseeuw and Leroy 1987; Croux and Dehon 2003; Punzo et al. 2021). Such a distinction depends on the direction in which an observation is atypical compared to the bulk of the data. The underlying scheme is summarized in Table 1.

The approach used for the detection of atypical observations is different between the models based on the MVT and MVCN distributions. Specifically, for those involving the MVT distribution, we extend to the matrix-variate framework the idea of McLachlan and Peel (2000) (see also Greselin and Ingrassia 2010; Ingrassia et al. 2012; Punzo et al. 2021). On the other hand, for models based on the MVCN distribution, we use an automatic procedure granted by the specific characteristics of this distribution.

Below, we describe the labeling procedures concerning the MVT-MVT-FMR-RC and MVCN-MVCN-FMR-RC for ease of exposition. In any case, both approaches can be combined for the mixed MV-FMR-RCs cases and simplified for the MV-FMR-FCs by ignoring the equations involving the distribution of the covariates.

To label data points, we consider *a posteriori* procedures. First of all, each observation $(\mathbf{Y}_i, \mathbf{X}_i)$ is assigned to one of the K groups through the maximum *a posteriori* probabilities (MAP) operator

$$\text{MAP}(\widehat{z}_{ik}) = \begin{cases} 1 & \text{if } \max_h\{\widehat{z}_{ih}\} \text{occurs in group } h = k, \\ 0 & \text{if otherwise,} \end{cases}$$

where $\widehat{z}_{ik}$ is the expected value of Z_{ik} at convergence of the ECM algorithm. Then, for the MVT-MVT-FMR-RC, we define

$$\sum_{k=1}^{K} \text{MAP}(\widehat{z}_{ik})\, \delta_k\left(\mathbf{Y}_i; \widehat{\mathbf{B}}_k \mathbf{X}_i^*, \widehat{\Sigma}_{\mathbf{Y}|k}, \widehat{\Psi}_{\mathbf{Y}|k}\right), \tag{18}$$

and

$$\sum_{k=1}^{K} \text{MAP}(\widehat{z}_{ik})\, \delta_k\left(\mathbf{X}_i; \widehat{\mathbf{M}}_k, \widehat{\Sigma}_{\mathbf{X}|k}, \widehat{\Psi}_{\mathbf{X}|k}\right), \tag{19}$$

and we categorize $(\mathbf{Y}_i, \mathbf{X}_i)$ as bad leverage if (18) and (19) are both sufficiently large, an outlier or good leverage if only (18) or (19), respectively, are sufficiently large, and typical otherwise.

In (18) and (19), the hat denotes the estimated values of the parameters at the convergence of the algorithm. To decide on how large the statistics (18) and (19) must be in order to be defined as sufficiently large, extending the idea of McLachlan and Peel (2000), we can compare them to selected percentiles ϵ of the chi-squared distribution with pr and qr degrees of freedom, respectively, where the chi-squared distribution is used to approximate the distribution of the squared Mahalanobis distances in (18) and (19). Herein, we set $\epsilon \in (0.95, 0.99, 0.999)$ (McLachlan and Peel 2000; Punzo et al. 2017). By combining the four possible outcomes we cover the options reported in Table 1.

For the MVCN-MVCN-FMR-RC, let $\widehat{v}_{ik\mathbf{Y}}$ and $\widehat{v}_{ik\mathbf{X}}$ be the expected values of $V_{ik\mathbf{Y}}$ and $V_{ik\mathbf{X}}$ at convergence of the ECM algorithm, respectively. The commonly adopted decision rule when contaminated distributions are used (see, e.g. Maruotti and Punzo, 2017; Tomarchio and Punzo, 2020; Punzo et al. 2021; Tomarchio et al. 2022), consists of considering a point typical in the $\mathbf{Y}$-direction if $\widehat{v}_{ik\mathbf{Y}} > 0.5$, and similarly in the $\mathbf{X}$-direction if $\widehat{v}_{ik\mathbf{X}} > 0.5$. Also in this case, by combining the four possibilities, we can straightforwardly apply the taxonomy reported in Table 1.

3 Artificial data analyses

3.1 Overview

Here, we simulate scenarios that may arise when dealing with real-world data. For the sake of brevity, we limit our discussion to MV-FMR-RCs only. We set $P = 2$, $Q = 3$, $R = 5$, and $K = 2$ throughout the analyses. Then, by using the parameters reported in Appendix 1, we simulate 100 samples of size N from each of the following four data generation processes (DGPs): (a) MVN-MVN-FMR-RC, (b) MVT-MVT-FMR-RC, (c) MVCN-MVCN-FMR-RC, and (d) MVN-MVN-FMR-RC where 5% of points have been modified to incorporate noisy values. Regarding scenario (d), in each simulated dataset, 5% of the points are randomly selected and, for each of these $(\mathbf{Y}_i, \mathbf{X}_i)$ points, we:

- Randomly choose one of the P rows of $\mathbf{Y}_i$ and replace the values therein contained with random numbers generated from a uniform distribution over the interval

$[a_{\mathbf{Y}}, b_{\mathbf{Y}}]$, being $a_{\mathbf{Y}}$ and $b_{\mathbf{Y}}$ the minimum and maximum values observed in the simulated dataset over $\mathbf{Y}_i$, $i = 1, \ldots, N$, respectively;
- Randomly choose one of the Q rows of $\mathbf{X}_i$ and replace the values therein contained with random numbers generated from a uniform distribution over the interval $[a_{\mathbf{X}}, b_{\mathbf{X}}]$, being $a_{\mathbf{X}}$ and $b_{\mathbf{X}}$ the minimum and maximum values observed in the simulated dataset over $\mathbf{X}_i$, $i = 1, \ldots, N$, respectively.

According to the above simulation scheme, we have scenarios characterized by no atypical points (a), heavy-tailed clusters of different natures (b-c), and noisy points (d). For each DGP, $N \in (100, 200, 500)$, yielding a total of 1200 generated datasets. On each generated dataset we fit the MVN-MVN-FMR-RC, MVT-MVT-FMR-RC, and MVCN-MVCN-FMR-RC, and their performances are discussed in Sect. 3.2 under the following points of view: (i) parameter recovery, (ii) classification performance, (iii) capability of identifying atypical points, and (iv) model selection. Notice that, to evaluate points (i) to (iii), the three considered models are directly fitted with $K = 2$. Differently, to assess point (iv), we fit all the nine MV-FMR-RCs discussed in Sect. 2.1 for $K \in \{1, 2, 3\}$.

3.2 Results

3.2.1 Parameter recovery

First of all, we examine the parameter recovery of the considered models. To this purpose, we calculate the average mean squared error (MSE) of the estimated regression coefficients over the 100 datasets simulated by each DGP. The obtained results are reported in Table 2.

By starting with scenario (a), i.e. when there are no atypical points, we note that the three models perform comparably because, in this situation, the MVT-MVT-FMR-RC and MVCN-MVCN-FMR-RC tend to the MVN-MVN-FMR-RC (refer to Sect. 2.1). Regardless of the considered model, the MSEs are negligible and improve with the increase of N. When scenarios (b) and (c) are considered, the robust approaches (MVT-MVT-FMR-RC and MVCN-MVCN-FMR-RC) are better than the traditional MVN-MVN-FMR-RC. Indeed, the MSEs of this model are regularly higher than those of the two robust models. As in scenario (a), the MSEs are negligible and improve with the increase of N. Furthermore, and as it is reasonable to expect, the best performer under scenario (b) is the MVT-MVT-FMR-RC, and under scenario (c) is the MVCN-MVCN-FMR-RC.

Lastly, under scenario (d), that is when there are noisy matrices, we immediately observe that the estimates of the MVN-MVN-FMR-RC are even worse than the corresponding ones of scenarios (b) and (c), especially for $\mathbf{B}_2$ which displays heavily distorted estimates. On the contrary, because of their robustness, MVT-MVT-FMR-RC and MVCN-MVCN-FMR-RC perform comparably well, as shown by their small MSEs that become better as N increases.

Table 2 Average MSEs of the estimated regression coefficients under the four scenarios

FMR-RC		$N = 100$	$N = 200$	$N = 500$
Scenario (a)				
MVN-MVN	$\mathbf{B}_1$	$\begin{bmatrix} 0.030 & 0.001 & 0.001 & 0.001 \\ 0.030 & 0.001 & 0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.000 & 0.000 & 0.000 \\ 0.016 & 0.000 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.005 & 0.000 & 0.000 & 0.000 \\ 0.005 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.057 & 0.001 & 0.001 & 0.000 \\ 0.038 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.033 & 0.001 & 0.001 & 0.000 \\ 0.024 & 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.000 & 0.000 & 0.000 \\ 0.010 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
MVT-MVT	$\mathbf{B}_1$	$\begin{bmatrix} 0.029 & 0.001 & 0.001 & 0.001 \\ 0.030 & 0.001 & 0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.000 & 0.000 & 0.000 \\ 0.016 & 0.000 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.005 & 0.000 & 0.000 & 0.000 \\ 0.005 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.056 & 0.001 & 0.001 & 0.000 \\ 0.038 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.032 & 0.001 & 0.001 & 0.000 \\ 0.024 & 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.000 & 0.000 & 0.000 \\ 0.010 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
MVCN-MVCN	$\mathbf{B}_1$	$\begin{bmatrix} 0.030 & 0.001 & 0.001 & 0.001 \\ 0.030 & 0.001 & 0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.000 & 0.000 & 0.000 \\ 0.016 & 0.000 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.005 & 0.000 & 0.000 & 0.000 \\ 0.005 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.057 & 0.001 & 0.001 & 0.000 \\ 0.038 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.033 & 0.001 & 0.014 & 0.000 \\ 0.024 & 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.000 & 0.000 & 0.000 \\ 0.010 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
Scenario (b)				
MVN-MVN	$\mathbf{B}_1$	$\begin{bmatrix} 0.378 & 0.002 & 0.004 & 0.002 \\ 0.412 & 0.002 & 0.004 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.346 & 0.002 & 0.002 & 0.001 \\ 0.325 & 0.001 & 0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.176 & 0.001 & 0.001 & 0.001 \\ 0.136 & 0.001 & 0.001 & 0.000 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.500 & 0.002 & 0.003 & 0.001 \\ 0.424 & 0.002 & 0.002 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.387 & 0.002 & 0.002 & 0.001 \\ 0.333 & 0.001 & 0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.190 & 0.001 & 0.001 & 0.000 \\ 0.341 & 0.001 & 0.001 & 0.000 \end{bmatrix}$

Table 2 (continued)

FMR-RC		$N = 100$	$N = 200$	$N = 500$
MVT-MVT	$\mathbf{B}_1$	$\begin{bmatrix} 0.021 & 0.001 & 0.001 & 0.001 \\ 0.020 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.000 & 0.001 & 0.000 \\ 0.012 & 0.000 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.004 & 0.000 & 0.000 & 0.000 \\ 0.006 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.043 & 0.001 & 0.001 & 0.001 \\ 0.042 & 0.001 & 0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.029 & 0.000 & 0.000 & 0.000 \\ 0.024 & 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.000 & 0.000 & 0.000 \\ 0.012 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
MVCN-MVCN	$\mathbf{B}_1$	$\begin{bmatrix} 0.025 & 0.001 & 0.002 & 0.001 \\ 0.031 & 0.001 & 0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.015 & 0.001 & 0.001 & 0.000 \\ 0.018 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.006 & 0.000 & 0.000 & 0.000 \\ 0.007 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.055 & 0.001 & 0.001 & 0.001 \\ 0.055 & 0.001 & 0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.040 & 0.001 & 0.001 & 0.000 \\ 0.033 & 0.000 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.012 & 0.000 & 0.000 & 0.000 \\ 0.014 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
Scenario (c)				
MVN-MVN	$\mathbf{B}_1$	$\begin{bmatrix} 0.470 & 0.003 & 0.003 & 0.002 \\ 0.402 & 0.003 & 0.003 & 0.002 \end{bmatrix}$	$\begin{bmatrix} 0.355 & 0.001 & 0.001 & 0.001 \\ 0.291 & 0.001 & 0.002 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.231 & 0.001 & 0.001 & 0.000 \\ 0.197 & 0.001 & 0.001 & 0.000 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.611 & 0.003 & 0.003 & 0.002 \\ 0.522 & 0.002 & 0.003 & 0.002 \end{bmatrix}$	$\begin{bmatrix} 0.608 & 0.001 & 0.002 & 0.001 \\ 0.439 & 0.001 & 0.002 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.355 & 0.001 & 0.001 & 0.000 \\ 0.321 & 0.001 & 0.001 & 0.000 \end{bmatrix}$
MVT-MVT	$\mathbf{B}_1$	$\begin{bmatrix} 0.034 & 0.001 & 0.001 & 0.001 \\ 0.029 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.014 & 0.000 & 0.001 & 0.000 \\ 0.013 & 0.000 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.006 & 0.000 & 0.000 & 0.000 \\ 0.005 & 0.000 & 0.000 & 0.000 \end{bmatrix}$

Table 2 (continued)

FMR-RC		$N = 100$	$N = 200$	$N = 500$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.057 & 0.001 & 0.001 & 0.000 \\ 0.059 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.020 & 0.000 & 0.001 & 0.000 \\ 0.024 & 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.000 & 0.000 & 0.000 \\ 0.010 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
MVCN-MVCN	$\mathbf{B}_1$	$\begin{bmatrix} 0.029 & 0.001 & 0.001 & 0.001 \\ 0.025 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.013 & 0.000 & 0.001 & 0.000 \\ 0.011 & 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.005 & 0.000 & 0.000 & 0.000 \\ 0.005 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.046 & 0.001 & 0.001 & 0.000 \\ 0.052 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.019 & 0.000 & 0.000 & 0.000 \\ 0.021 & 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.000 & 0.000 & 0.000 \\ 0.010 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
Scenario (d)				
MVN-MVN	$\mathbf{B}_1$	$\begin{bmatrix} 0.532 & 0.041 & 0.051 & 0.021 \\ 0.380 & 0.052 & 0.108 & 0.045 \end{bmatrix}$	$\begin{bmatrix} 0.209 & 0.009 & 0.026 & 0.013 \\ 0.400 & 0.048 & 0.057 & 0.030 \end{bmatrix}$	$\begin{bmatrix} 0.060 & 0.002 & 0.022 & 0.008 \\ 0.212 & 0.044 & 0.062 & 0.029 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 1.933 & 0.042 & 0.120 & 0.046 \\ 3.831 & 0.143 & 0.270 & 0.097 \end{bmatrix}$	$\begin{bmatrix} 1.464 & 0.019 & 0.105 & 0.038 \\ 3.466 & 0.121 & 0.248 & 0.095 \end{bmatrix}$	$\begin{bmatrix} 1.681 & 0.007 & 0.085 & 0.028 \\ 3.617 & 0.136 & 0.264 & 0.103 \end{bmatrix}$
MVT-MVT	$\mathbf{B}_1$	$\begin{bmatrix} 0.030 & 0.001 & 0.001 & 0.001 \\ 0.033 & 0.001 & 0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.013 & 0.000 & 0.001 & 0.000 \\ 0.011 & 0.000 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.005 & 0.000 & 0.000 & 0.000 \\ 0.006 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.064 & 0.001 & 0.001 & 0.001 \\ 0.058 & 0.001 & 0.002 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.031 & 0.001 & 0.001 & 0.000 \\ 0.029 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.015 & 0.000 & 0.000 & 0.000 \\ 0.013 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
MVCN-MVCN	$\mathbf{B}_1$	$\begin{bmatrix} 0.029 & 0.001 & 0.001 & 0.000 \\ 0.029 & 0.001 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.012 & 0.000 & 0.001 & 0.000 \\ 0.011 & 0.000 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.005 & 0.000 & 0.000 & 0.000 \\ 0.006 & 0.000 & 0.000 & 0.000 \end{bmatrix}$
	$\mathbf{B}_2$	$\begin{bmatrix} 0.062 & 0.001 & 0.001 & 0.001 \\ 0.058 & 0.001 & 0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.026 & 0.000 & 0.001 & 0.000 \\ 0.024 & 0.000 & 0.001 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.012 & 0.000 & 0.000 & 0.000 \\ 0.011 & 0.000 & 0.000 & 0.000 \end{bmatrix}$

Table 3 Average ARI for the considered models over the simulated datasets

FMR-RC	$N = 100$	$N = 200$	$N = 500$
Scenario (a)			
MVN-MVN	1.000	1.000	1.000
MVT-MVT	1.000	1.000	1.000
MVCN-MVCN	1.000	1.000	1.000
Scenario (b)			
MVN-MVN	0.969	0.955	0.941
MVT-MVT	0.997	0.997	0.997
MVCN-MVCN	0.988	0.976	0.969
Scenario (c)			
MVN-MVN	0.909	0.919	0.935
MVT-MVT	0.992	0.995	0.994
MVCN-MVCN	0.992	0.996	0.995
Scenario (d)			
MVN-MVN	0.917	0.912	0.905
MVT-MVT	0.926	0.918	0.907
MVCN-MVCN	0.933	0.917	0.907

3.2.2 Classification performance

In terms of classification performance, we consider the adjusted Rand index (ARI; Hubert and Arabieis one of the most popular 1985). In detail, for each scenario, we compare the classification produced by a model with the true one. The average ARI over the 100 simulated datasets by each scenario is then reported in Table 3. As we can see, under scenario (a), all the models regularly provide a perfect data classification. However, in scenarios (b), (c), and (d), the robust approaches always show better results than the MVN-MVN-FMR-RC.

3.2.3 Atypical points detection

For evaluating the performances of the MVT-MVT-FMR-RC and MVCN-MVCN-FMR-RC in detecting atypical points in Scenario (d), we consider the true positive rate (TPR), measuring the proportion of atypical points that are correctly identified as atypical, and the false positive rate (FPR), corresponding to the proportion of typical points incorrectly classified as atypical. Their average values over the 100 datasets simulated by each DGP are reported in Table 4.

We limit our investigation to the detection of atypical points in general, i.e. without discriminating between the categories in Table 1, because of the random way in which the noisy values are inserted into the data. From the obtained results, we note that both models have the same performances in terms of TPR since they always recognize the noisy matrices as atypical. However, they perform differently when the FPR is considered. Indeed, while the detection rule for the MVCN-MVCN-FMR-RC provides

Table 4 Average TPRs and FPRs under Scenario (d) over the simulated datasets

FMR-RC	TPR			FPR		
	$N = 100$	$N = 200$	$N = 500$	$N = 100$	$N = 200$	$N = 500$
MVT-MVT$_{\epsilon=(0.95)}$	1.000	1.000	1.000	0.119	0.109	0.110
MVT-MVT$_{\epsilon=(0.99)}$	1.000	1.000	1.000	0.029	0.025	0.024
MVT-MVT$_{\epsilon=(0.999)}$	1.000	1.000	1.000	0.004	0.003	0.003
MVCN-MVCN	1.000	1.000	1.000	0.015	0.003	0.010

almost optimal results, that of the MVT-MVT-FMR-RC depends on the choice of ϵ. On a related note, it seems that by using $\epsilon = 0.95$ as suggested by McLachlan and Peel (2000), this model tends to declare more observations as atypical than there are. Nevertheless, results greatly improve as we lower ϵ to 0.99 or 0.999, producing values that are comparable to (and in some cases even better than) those of the MVCN-MVCN-FMR-RC.

3.2.4 Model selection

Model selection is usually required to select among a set of candidate models with differing numbers of parameters. The classical way to perform this selection is via the computation of a convenient (likelihood-based) model selection criterion. Among them, the Bayesian information criterion (BIC; Schwarz 1978) is one of the most popular and will be used in the following.

In this section, we evaluate the performance of the BIC in (i) identifying the correct DGP and (ii) selecting the correct number of groups K in the data. We focus our attention on scenarios (a) to (c) since we are interested in analyzing the performance of the models belonging to our family. The obtained results are reported in Table 5. Specifically, we count how many times the correct DGP and K have been identified over the 100 datasets simulated for each N in a given scenario, after fitting all the nine MV-FMR-RCs for $K \in \{1, 2, 3\}$.

We immediately note that the BIC almost regularly identifies the correct DGP and K, regardless of the scenario and N considered. Concerning the identification of the

Table 5 Number of times, over the 100 datasets simulated for each N in a given scenario, for which the correct DGP and K have been identified by the BIC, after fitting all the nine MV-FMR-RCs for $K \in \{1, 2, 3\}$

	# Correct DGP			# Correct K		
	$N = 100$	$N = 200$	$N = 500$	$N = 100$	$N = 200$	$N = 500$
Scenario (a)	98	100	100	98	100	100
Scenario (b)	99	100	100	99	98	99
Scenario (c)	95	95	100	99	100	98

correct DGP, we report that the few situations where the true model is not selected are due to a wrong selection in only one between $f(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{Y}|k})$ and $g(\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{X}|k})$. Regarding the choice of the true K, we notice that the rare occasions where $K = 2$ has not been selected are because a solution with $K = 3$ is preferred.

4 Real data example

4.1 Overview

There have been unprecedented signs of global warming over the last decades, and understanding the differences in GHG emissions across countries is fundamental for climate change mitigation. Climate change is heavily related to energy, economics, and the environment. Thus, to implement sustainable development strategies, it is crucial to jointly analyze determinants of GHG emissions, their paths over time, and the differences between countries (Zheng et al. 2019; González-Sánchez and Martín-Ortega 2020).

Here, we focus on the most dangerous and regularly studied GHGs, i.e. carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O) (Bruvoll and Larsen 2004; Maucieri et al. 2017). Specifically, we analyze the emissions of these gases (in kilotonnes) generated by the energy sector of the agri-food system, which is estimated to account for around 1/4th of all GHG emissions (Poore and Nemecek 2018; Mrówczyńska-Kamińska et al. 2021).

Differently from what is typically done in the related literature, we jointly consider the three gases as response variables in our analysis. Indeed, each gas is often separately used as a response variable in regression or panel data analyses, or information provided by the three gases is aggregated to obtain a univariate response (for some examples, see Lin and Xu, 2018; González-Sánchez and Martín-Ortega 2020; Mrówczyńska-Kamińska et al. 2021; Nguyen et al. 2021 and the references therein). However, these gases share a degree of relationships among them, and jointly exploiting their physicochemical characteristics can provide richer information (Liang et al. 2015; Villa et al. 2019).

The three gases are measured for a set of $N = 159$ countries over the years 2012–2016. Thus, each $\mathbf{Y}_i$ is a 3×5 matrix. As concerns the covariates, we use three of the most important and regularly used variables for explaining GHG emissions of this sector (see, e.g. Lin and Xu, 2018; González-Sánchez and Martín-Ortega, 2020; Mrówczyńska-Kamińska et al. 2021; Nguyen et al. 2021), i.e. the GDP per capita (in $), the energy intensity of the economy (in kWh per $ PPP), and population. These variables are measured for the same years as above, leading to a 3×5 matrix for each $\mathbf{X}_i$. Thus, both sets of variables come in the form of three-way arrays of dimensions $3 \times 5 \times 159$.

The data are freely available from the FAOSTAT database (for the $\mathbf{Y}$ variables), accessible at https://www.fao.org/faostat/en/#data/GT, and the World Bank database (for the $\mathbf{X}$ variables), accessible at https://databank.worldbank.org/source/country-climate-and-development-report-(ccdr). Notice that, the list of countries in both databases is slightly different, as well as the variables have missing values for some

countries. Thus, we first select only the countries included in both databases and then we filtered only those having values for all the variables. Additionally, both sets of variables are log-transformed to avoid boundary bias issues (Tomarchio and Punzo 2020) and in agreement with related literature (Lin and Xu 2018; González-Sánchez and Martín-Ortega 2020; Nguyen et al. 2021). Thus, regression coefficients can be interpreted as elasticities.

4.2 How to decide if covariates are random or fixed

In analyzing the data, the first aspect to decide is which kind of MV-FMR is the most appropriate. Indeed, fixed and random covariates approaches cannot be directly compared because they are based on different definitions of the likelihood function: conditional for MV-FMR-FCs and joint for MV-FMR-RCs (Punzo 2014; Punzo et al. 2018). Additionally, covariates should be treated as random if their distribution presents a clustering structure (Tomarchio et al. 2021). Thus, in the fashion of Punzo et al. (2018) and Tomarchio et al. (2023), to determine if covariates need to be considered fixed or random, we:

1. fit matrix-variate mixtures of MVT and MVCN distributions over a reasonable range of values for K;
2. for each fitted mixture model computes the BIC;
3. if the selected value of K is greater than 1, indicating a group structure, then treat the covariates as random, otherwise consider them as fixed.

By using these practical guidelines, we start by fitting matrix-variate mixtures of MVT and MVCN distributions with $K = 1$ and $K = 2$. Considering that, in both cases, the best BIC is obtained when $K = 2$, we say that the marginal distribution of the covariates presents a clustering structure. Thus, we consider MV-FMR-RCs for the analysis of this dataset.

4.3 Results

We fit our eight MV-FMR-RCs, as well as the MVN-MVN-FMR-RC of Tomarchio et al. (2021), to the data for $K \in \{1, \ldots, 6\}$. We use the BIC to select the best K for each model, and the results are reported in Table 6.

As we can see, all the models agree in detecting $K = 5$ groups in the data, but the best-fitting one is the MVT-MVT-FMR-RC. It is interesting to note that, according to the ranking based on the BICs, the top four models are those having a heavy-tailed distribution both for $f(\mathbf{Y}|\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{Y}|k})$ and $g(\mathbf{X}; \boldsymbol{\Theta}_{\mathbf{X}|k})$ in (5). This highlights the necessity of using heavy-tailed distributions for modeling both sets of variables. By looking at the estimated tailedness parameters of the MVT-MVT-FMR-RC model, which will be considered the reference model hereafter, we have $\nu_{1\mathbf{Y}} = 2.01$, $\nu_{2\mathbf{Y}} = 5.37$, $\nu_{3\mathbf{Y}} = 3.62$, $\nu_{4\mathbf{Y}} = 3.25$, $\nu_{5\mathbf{Y}} = 2.22$, $\nu_{1\mathbf{X}} = 4.26$, $\nu_{2\mathbf{X}} = 2.99$, $\nu_{3\mathbf{X}} = 4.53$, $\nu_{4\mathbf{X}} = 4.08$ and $\nu_{5\mathbf{X}} = 2.01$. Based on these values, the tails of each group are very heavy, and the MVN distribution cannot provide an adequate fit. Indeed, the MVN-MVN-FMR-RC is ranked in last place, providing the worst-fitting result.

Table 6 Number of groups (K) and BIC value, along with its ranking, of the models according by the BIC

FMR-RC	*K*	BIC	Ranking
MVT-MVT	5	**– 8379.589**	1
MVT-MVCN	5	– 8322.754	2
MVT-MVN	5	– 8067.628	5
MVCN-MVT	5	– 8270.756	3
MVCN-MVCN	5	– 8204.465	4
MVCN-MVN	5	– 7931.896	7
MVN-MVT	5	– 8003.488	6
MVN-MVCN	5	– 7898.954	8
MVN-MVN	5	– 7651.709	9

In bold, the best-fitting model

In Figs. 1 and 2, we illustrate the world political map colored according to the estimated classification by the MVT-MVT-FMR-RC model, and the parallel coordinate plots of its estimated mean matrices, respectively. The joint analysis of both figures provides useful insights for the interpretation of the detected clusters.

By starting with the analysis of Group 1, we notice that it is mainly composed of countries having the highest levels of primary energy production per capita, such as Qatar, UAE, Norway, Oman, Kuwait, Eq. Guinea, Libya (The Shift Project 2023). Thus, it is not a case that this set of countries has the highest mean values for the energy intensity of the economy as well as relevant mean GDPs per capita. On a related note, these countries have the lowest mean levels of population.

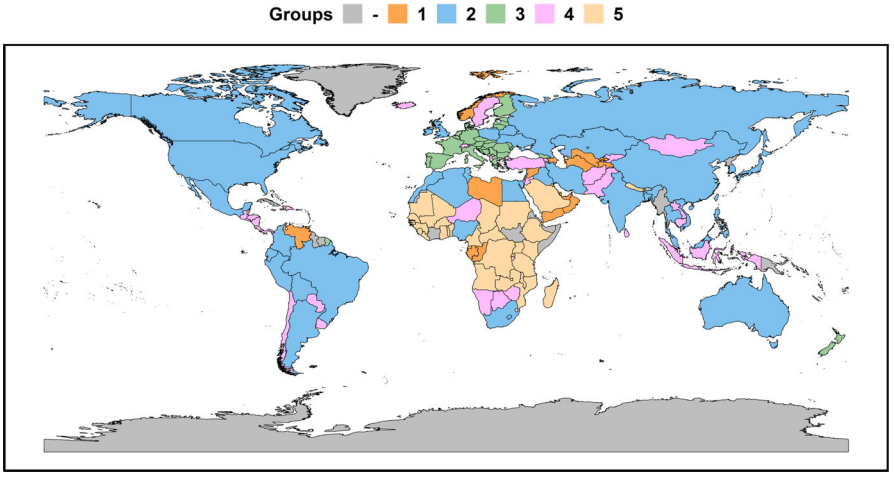

Fig. 1 World political map colored according to the estimated classification. The countries not included in the analysis are colored in grey

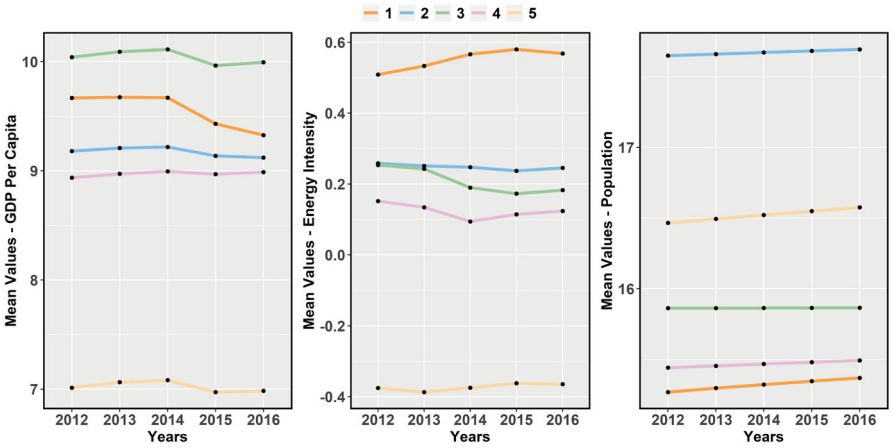

Fig. 2 Parallel coordinate plots for the mean matrices estimated by MVT-MVT-FMR-RC, over the 5 years for each covariate, and colored according to the estimated classification

Concerning Group 2, it seems to cluster several countries spread across the world. An interesting characteristic of this group is that twelve out of the fifteen most populous countries in the world (representing more than 60% of the world's population) are included in this group. From an economic perspective, the range of countries in this group leads to intermediate GDPs per capita and relatively high energy intensity averages.

Group 3 consists of the majority of the European Union countries and New Zealand. They have the highest average GDP per capita and intermediate mean values for the other two variables. The countries clustered in Group 4 are not of straightforward interpretation, but by drawing up a ranking, they have average values in the penultimate position in all three variables. Lastly, Group 5 clusters African countries. The set of countries in this group shows, by far, the lowest averages in GDPs per capita and energy intensity of the economy, but are ranked second in terms of mean population.

Additional information provided by Fig. 2 is the (average) temporal evolution of the five groups across the years of analysis. First of all, we immediately notice that the average population gradually increases over time for four out of five groups. The temporal pattern is less smooth when the other two variables are considered. For example, when the average GDP per capita is considered, we see that all the groups show a decline after 2014. This is particularly evident for Groups 1 and 3. On the other hand, when the average energy intensity is analyzed, we observe a different behavior among the groups. Specifically, Groups 1 and 5 display an increase in energy intensity over the years (although it is more evident for Group 1). The other three groups show a decrease in energy intensity over the years, particularly Groups 3 and 4 in 2014.

We now report the estimated regression coefficients for the five groups

$$\mathbf{B}_1 = \begin{bmatrix} 3.416 & 0.069 & 0.095 & 0.157 \\ -19.015 & 0.203 & 0.159 & 1.040 \\ -1.254 & 0.177 & 0.330 & 0.653 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} -12.974 & 0.087 & 0.123 & 1.053 \\ -17.207 & 0.061 & 0.281 & 1.043 \\ -7.305 & 0.142 & 0.574 & 1.030 \end{bmatrix},$$

$$\mathbf{B}_3 = \begin{bmatrix} -7.280 & -0.085 & 0.160 & 0.785 \\ -14.190 & -0.017 & 0.207 & 0.931 \\ -5.999 & -0.095 & 0.540 & 1.109 \end{bmatrix}, \quad \mathbf{B}_4 = \begin{bmatrix} -16.615 & 0.040 & 0.090 & 1.263 \\ -18.023 & 0.257 & 0.170 & 0.962 \\ -10.388 & 0.546 & 0.213 & 0.951 \end{bmatrix},$$

$$\mathbf{B}_5 = \begin{bmatrix} -12.322 & -0.001 & -0.001 & 1.021 \\ -16.673 & 0.007 & 0.028 & 0.987 \\ -13.065 & 0.194 & 0.387 & 1.211 \end{bmatrix}.$$

In each matrix, the rows refer to the CO_2, CH_4, and N_2O responses, respectively, the first column identifies the intercept, whereas the other three columns refer to the GDP per capita, energy intensity, and population covariates, respectively. By starting with the latter covariate, we immediately see that an increase in the population by 1% leads to a growth of (approximately) the same percentage in the three GHGs, regardless of the considered group. The world's population is growing, and each additional person consumes energy and resources (directly and indirectly), resulting in higher GHG emissions. Furthermore, it is interesting to notice that Group 1, which we recall has the lowest average population, displays the lowest values on two coefficients out of three.

Similarly, the effect of the energy intensity of the economy on GHGs emissions is positive, if we exclude Group 5 which shows a practically null value. Such an effect grows as we pass from CO_2 to N_2O, regardless of the considered group.

Regarding the coefficients of the GDP per capita, we observe positive values for all the groups except the third. A positive relationship between the GDP and GHGs emissions is generally observed in the literature (González-Sánchez and Martín-Ortega 2020; Mrówczyńska-Kamińska et al. 2021). A possible explanation for this behavior relies on a phenomenon sometimes hypothesized in economics according to which, once a certain level of GDP per capita has been reached, the efficiency of production systems can lead to environmental improvements and lower levels of emissions (Zmami and Ben-Salha 2020). By recalling Fig. 2, we see that Group 3 is the one characterized by the highest levels of average GDP per capita.

A final aspect concerns the labeling of countries according to the categories in Table 1. Specifically, the MVT-MVT-FMR-RC (with $\epsilon = 0.999$) identifies 89 typical points, 31 good leverage, 26 outliers, and 13 bad leverage points. By comparing these results with those of the other three heavy-tailed models, ranked from second to fourth in Table 6, they agree in 74.84%, 69.18%, and 67.92% of the cases. The main differences are that the MVT-MCVN-FMR-RC and MVCN-MVCN-FMR-RC label more countries as good leverage points, whereas the MVCN-MVT-FMR-RC detects more outliers.

4.3.1 An alternative analysis by using multivariate data

As introduced in Sect. 1, after vectorization, both $\mathbf{Y}$ and $\mathbf{X}$ can be rearranged into a multivariate form factor, each having dimensions 159×15. To see the consequences of such a choice, we fit for $K \in \{1, \ldots, 6\}$ the multivariate versions of the nine MV-FMR-RCs considered in Sect. 4.3.

Herein, we limit to report that the best model according to the BIC is that using a multivariate t distribution both for the responses and the covariates, thus apparently resembling the results discussed in Sect. 4.3. However, this multivariate model has 2489 parameters, i.e. approximately ten times those of the MVT-MVT-FMR-RC, making it heavily overparameterized and intractable. This is due to the vectorization process that led to an explosion in the number of parameters to be estimated. Additionally, the obtained data classification is quite different from that of the MVT-MVT-FMR-RC. If we compare the agreement between the classifications of these two models using the ARI, the obtained value is 0.23, which is quite low. Furthermore, the detected classification is hardly interpretable unlike that illustrated in Fig. 1.

5 Conclusions

In this manuscript, ten matrix-variate finite mixtures of regressions models have been introduced: two by using the fixed covariate approach, and eight within the random covariate paradigm. We considered the matrix-variate t and contaminated normal distributions for the mixture components, in order to robustify estimation and classification compared to the traditional case based on the normality assumption. Additionally, for the random covariate models, we also allow the use of the matrix-variate normal distribution in cases where only one of the two sets of variables exhibits atypical points. We illustrated an ECM algorithm for maximum likelihood parameter estimation.

The results of our simulated analyses showed that in the presence of heavy-tailed clusters or noisy matrices, the mean squared errors of the regression coefficients are regularly higher when the normal distribution is used for the mixture components. Conversely, the use of heavy-tailed distributions down-weights the atypical points in the estimation processes, resulting in more precise parameter estimates. Similar conclusions have been also drawn for the estimated data classifications. Model selection via the BIC provided excellent results both in identifying the true data-generating distributions and the number of clusters in the data.

From the results of our real-data analysis we showed that, according to the BIC, the top four ranked models are those based on the matrix-variate t and contaminated normal distributions for both the responses and the covariates. In other terms, the models having the matrix-variate normal distribution for one or both sets of variables produced worse fitting. By considering the results of the best-fitting model, five clusters of countries have been detected, each having specific characteristics according to the estimated parameters. Comparisons among the top four ranked models have also been conducted in terms of labeling the countries as typical or not. From the results, they assign (on average) the same labels to more than 70% of the data points. The

disadvantages of vectorizing the dataset for then applying the multivariate version of our models have been commented on in terms of the explosion of the number of parameters to be estimated and classification interpretation.

Funding Open access funding provided by Università degli Studi di Catania within the CRUI-CARE Agreement. No funding was received for conducting this study.

Data availability The real dataset is publicly available as specified in the manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Appendix 1

Here, we report the parameters used to generate the simulated datasets of Sect. 3.1. All the models have in common the following parameters:

$$\pi_1 = \pi_2 = 0.50,$$

$$\mathbf{B}_1 = \begin{bmatrix} 2.00 & 1.00 & 1.00 & -1.00 \\ 3.00 & 1.00 & -1.00 & 1.00 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} -10.00 & 1.00 & 1.00 & -1.00 \\ -8.00 & 1.00 & -1.00 & 1.00 \end{bmatrix},$$

$$\mathbf{M}_1 = \begin{bmatrix} 3.00 & 1.00 & 7.00 & 6.00 & 4.00 \\ 3.00 & 2.00 & 6.00 & 2.00 & 4.00 \\ 2.00 & 7.00 & 6.00 & 2.00 & 5.00 \end{bmatrix}, \quad \mathbf{M}_2 = \begin{bmatrix} 6.00 & 4.00 & 10.00 & 9.00 & 7.00 \\ 6.00 & 5.00 & 9.00 & 5.00 & 7.00 \\ 5.00 & 5.00 & 9.00 & 5.00 & 8.00 \end{bmatrix},$$

$$\Sigma_{\mathbf{Y}|1} = \Sigma_{\mathbf{Y}|2} = \begin{bmatrix} 1.00 & 0.50 \\ 0.50 & 1.00 \end{bmatrix}, \quad \Sigma_{\mathbf{X}|1} = \Sigma_{\mathbf{X}|2} = \begin{bmatrix} 1.00 & 0.60 & 0.36 \\ 0.60 & 1.00 & 0.60 \\ 0.36 & 0.60 & 1.00 \end{bmatrix},$$

$$\Psi_{\mathbf{Y}|1} = \Psi_{\mathbf{Y}|2} = \Psi_{\mathbf{X}|1} = \Psi_{\mathbf{X}|2} = \begin{bmatrix} 1.00 & 0.70 & 0.49 & 0.34 & 0.24 \\ 0.70 & 1.00 & 0.70 & 0.49 & 0.34 \\ 0.49 & 0.70 & 1.00 & 0.70 & 0.49 \\ 0.34 & 0.49 & 0.70 & 1.00 & 0.70 \\ 0.24 & 0.34 & 0.49 & 0.70 & 1.00 \end{bmatrix}.$$

Then, for the MVT-MVT-FMR-RC we have $\nu_{\mathbf{Y}} = \nu_{\mathbf{X}} = (2.50, 2.50)$, whereas for the MVCN-MVCN-FMR-RC we have $\alpha_{\mathbf{Y}} = \alpha_{\mathbf{X}} = (0.90, 0.80)$ and $\eta_{\mathbf{Y}} = \eta_{\mathbf{X}} = (15.00, 20.00)$.

References

Anderlucci L, Montanari A, Viroli C (2014) A matrix-variate regression model with canonical states: an application to elderly Danish twins. Statistica 74(4):367–381

Biernacki C, Celeux G, Govaert G (2003) Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. Comput Stat Data Anal 41(3–4):561–575

Bruvoll A, Larsen BM (2004) Greenhouse gas emissions in Norway: Do carbon taxes work? Energy Policy 32(4):493–505

Croux C, Dehon C (2003) Estimators of the multiple correlation coefficient: local robustness and confidence intervals. Stat Pap 44(3):315–334

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc: Ser B (Methodol) 39(1):1–22

Doğru FZ, Bulut YM, Arslan O (2016) Finite mixtures of matrix variate t distributions. Gazi Univ J Sci 29(2):335–341

Gallaugher MPB, McNicholas PD (2017) A matrix variate skew-t distribution. Stat 6(1):160–170

Gallaugher MPB, McNicholas PD (2018) Finite mixtures of skewed matrix variate distributions. Patt Recogn 80:83–93

Gallaugher MPB, McNicholas PD (2019) Three skewed matrix variate distributions. Stat Probab Lett 145:103–109

Gallaugher MPB, McNicholas PD (2020) Mixtures of skewed matrix variate bilinear factor analyzers. Adv Data Anal Classif 14(2):415–434

Gallaugher MPB, Tomarchio SD, McNicholas PD et al (2022) Model-based clustering via skewed matrix-variate cluster-weighted models. J Stat Comput Simul 92(13):2645–2666

González-Sánchez M, Martín-Ortega JL (2020) Greenhouse gas emissions growth in Europe: a comparative analysis of determinants. Sustainability 12(3):1012

Greselin F, Ingrassia S (2010) Constrained monotone EM algorithms for mixtures of multivariate t distributions. Stat Comput 20(1):9–22

Hossain A, Naik D (1991) A comparative study on detection of influential observations in linear regression. Stat Pap 32(1):55–69

Hubert L, Arabie P (1985) Comparing partitions. J Classif 2(1):193–218

Ingrassia S, Minotti SC, Vittadini G (2012) Local statistical modeling via a cluster-weighted approach with elliptical distributions. J Classif 29(3):363–401

Liang L, Eberwein J, Allsman L et al (2015) Regulation of CO_2 and N_2O fluxes by coupled carbon and nitrogen availability. Environ Res Lett 10(3):034008

Lin B, Xu B (2018) Factors affecting CO_2 emissions in China's agriculture sector: a quantile regression. Renew Sustain Energy Rev 94:15–27

Maronna RA (1976) Robust m-estimators of multivariate location and scatter. Ann Stat:51–67

Maruotti A, Punzo A (2017) Model-based time-varying clustering of multivariate longitudinal data with covariates and outliers. Comput Stat Data Anal 113:475–496

Maucieri C, Barbera AC, Vymazal J et al (2017) A review on the main affecting factors of greenhouse gases emission in constructed wetlands. Agric For Meteorol 236:175–193

McLachlan GJ, Peel D (2000) Finite mixture models. Wiley, New York

McNeil AJ, Frey R, Embrechts P (2015) Quantitative risk management: concepts, techniques and tools. Princeton University Press

Melnykov V, Zhu X (2018) On model-based clustering of skewed matrix data. J Multivar Anal 167:181–194

Melnykov V, Zhu X (2019) Studying crime trends in the USA over the years 2000–2012. Adv Data Anal Classif 13(1):325–341

Meng XL, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika 80(2):267–278

Michael S, Melnykov V (2016) An effective strategy for initializing the EM algorithm in finite mixture models. Adv Data Anal Classif 10:563–583

Mrówczyńska-Kamińska A, Bajan B, Pawłowski KP et al (2021) Greenhouse gas emissions intensity of food production systems and its determinants. PLoS One 16(4):e0250995

Nguyen CP, Le TH, Schinckus C et al (2021) Determinants of agricultural emissions: panel data evidence from a global sample. Environ Dev Econ 26(2):109–130

Poore J, Nemecek T (2018) Reducing food's environmental impacts through producers and consumers. Science 360(6392):987–992

Punzo A (2014) Flexible mixture modelling with the polynomial Gaussian cluster-weighted model. Stat Model 14(3):257–291
Punzo A, Tomarchio SD (2022) Parsimonious finite mixtures of matrix-variate regressions. In: Innovations in multivariate statistical modeling. Springer, pp 385–398
Punzo A, McNicholas P (2017) Robust clustering in regression analysis via the contaminated Gaussian cluster-weighted model. J Classif 34(2):249–293
Punzo A, Ingrassia S, Maruotti A (2018) Multivariate generalized hidden Markov regression models with random covariates: physical exercise in an elderly population. Stat Med 37(19):2797–2808
Punzo A, Ingrassia S, Maruotti A (2021) Multivariate hidden Markov regression models: random covariates and heavy-tailed distributions. Stat Pap 62(3):1519–1555
Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection. Wiley
Sarkar S, Zhu X, Melnykov V et al (2020) On parsimonious models for modeling matrix data. Comput Stat Data Anal 142:106822
Schwarz G (1978) Estimating the dimension of a model. Ann Stat:461–464
The Shift Project (2023) Primary energy production. https://www.theshiftdataportal.org/
Tomarchio SD (2022) Matrix-variate normal mean-variance Birnbaum–Saunders distributions and related mixture models. Comput Stat. https://doi.org/10.1007/s00180-022-01290-9
Tomarchio SD, Punzo A (2020) Dichotomous unimodal compound models: application to the distribution of insurance losses. J Appl Stat 47(13–15):2328–2353
Tomarchio SD, Punzo A, Bagnato L (2020) Two new matrix-variate distributions with application in model-based clustering. Comput Stat Data Anal 152:107050
Tomarchio SD, McNicholas PD, Punzo A (2021) Matrix normal cluster-weighted models. J Classif 38(3):556–575
Tomarchio SD, Gallaugher MPB, Punzo A et al (2022) Mixtures of matrix-variate contaminated normal distributions. J Comput Graph Stat 31(2):413–421
Tomarchio SD, Punzo A, Maruotti A (2023) Matrix-variate hidden Markov regression models: fixed and random covariates. J Classif. https://doi.org/10.1007/s00357-023-09438-y
Villa JA, Ju Y, Vines C et al (2019) Relationships between methane and carbon dioxide fluxes in a temperate cattail-dominated freshwater wetland. J Geophys Res Biogeosci 124(7):2076–2089
Viroli C (2011) Finite mixtures of matrix normal distributions for classifying three-way data. Stat Comput 21(4):511–522
Viroli C (2011) Model based clustering for three-way data structures. Bayesian Anal 6(4):573–602
Viroli C (2012) On matrix-variate regression analysis. J Multivar Anal 111:296–309
Zheng X, Streimikiene D, Balezentis T et al (2019) A review of greenhouse gas emission profiles, dynamics, and climate change mitigation efforts across the key climate change players. J Clean Prod 234:1113–1133
Zhu X, Sarkar S, Melnykov V (2022) MatTransmix: an R package for matrix model-based clustering and parsimonious mixture modeling. J Classif 39(1):147–170
Zmami M, Ben-Salha O (2020) An empirical analysis of the determinants of CO_2 emissions in GCC countries. Int J Sustain Dev World Ecol 27(5):469–480