Advances in Data Analysis and Classification (2024) 18:61–83
https://doi.org/10.1007/s11634-023-00552-8

REGULAR ARTICLE

Model-based clustering using a new multivariate skew distribution

Salvatore D. Tomarchio[1] · **Luca Bagnato**[2] · **Antonio Punzo**[1]

Received: 18 November 2022 / Revised: 17 June 2023 / Accepted: 2 July 2023 /
Published online: 22 July 2023
© The Author(s) 2023

Abstract
Quite often real data exhibit non-normal features, such as asymmetry and heavy tails, and present a latent group structure. In this paper, we first propose the multivariate skew shifted exponential normal distribution that can account for these non-normal characteristics. Then, we use this distribution in a finite mixture modeling framework. An EM algorithm is illustrated for maximum-likelihood parameter estimation. We provide a simulation study that compares the fitting performance of our model with those of several alternative models. The comparison is also conducted on a real dataset concerning the log returns of four cryptocurrencies.

Keywords Mixture models · Skewed data · Model-based clustering · Cryptocurrencies

Mathematics Subject Classification 62H10 · 62H30

1 Introduction

Finite mixture models are an important tool for the statistical modeling and analysis of many kinds of data. Because of their flexibility, finite mixture models can be used as a clustering technique or as a mathematical device for obtaining a tractable form of analysis for unmanageable data distributions (Titterington et al. 1985). Historically, the multivariate normal mixture (MN-M) model has been the most popular among researchers and practitioners given its mathematical tractability and usefulness in various application domains (McNicholas 2016). Nevertheless, for many real phenomena, such a model does not provide an adequate fit to the data, and more general models are

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

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

2 Department of Economic and Social Sciences, Catholic University of the Sacred Heart, Piacenza, Italy

required. For example, to accommodate outliers in the data, symmetric heavy-tailed distributions are often considered for the mixture components, since they can manage the thickness of the tails by using one or a few additional parameters. By focusing on multivariate data analyses, as we will do throughout this paper, examples of symmetric heavy-tailed mixtures can be found in Peel and McLachlan (2000), Sun et al. (2010), Andrews et al. (2011), McNicholas and Subedi (2012), Dang et al. (2015), Bagnato et al. (2017), Tomarchio et al. (2022) and Tong and Tortora (2022).

A further deviation from the normality assumption of the mixture components arises when the data involve observations whose distribution is asymmetric. In these cases, the choice of mixture components that are capable of modeling asymmetry can provide a better fit. In the multivariate literature, several mixtures of skew distributions have been proposed in the past few decades. Without making an exhaustive summary of the literature (see Lee and McLachlan 2022 for such an overview), we can mention the finite mixture models based on the multivariate skew normal (MSN) and skew-*t* (MST) distributions, in their various characterizations, proposed by Arellano-Valle et al. (2008), Lin (2009), Lin (2010), Frühwirth-Schnatter and Pyne (2010), Cabral et al. (2012), Vrbik and McNicholas (2014) and Lee and McLachlan (2016). Two other skew mixture models are those presented by Cabral et al. (2012) and based on the multivariate skew contaminated normal (MSCN) and skew slash (MSS) distributions. In McNicholas et al. (2013) and McNicholas et al. (2017), the multivariate variance gamma (MVG) distribution is used for model-based clustering via skew mixture models. Another flexible skew mixture model is provided by Browne and McNicholas (2015), who consider the multivariate generalized hyperbolic (MGH) distribution for the mixture components.

In this paper, we expand the multivariate literature devoted to asymmetric data modeling. Specifically, we introduce the multivariate skew shifted exponential normal (MSSEN) distribution, asymmetric generalization of the multivariate shifted exponential normal (MSEN) distribution introduced by Punzo and Bagnato (2020). Compared to its symmetric counterpart, the MSSEN has an additional parameter vector that accounts for the modelization asymmetric data. From a technical point of view, the MSSEN distribution can be obtained via the well-known normal mean-variance mixture model (McNeil et al. 2015) when a convenient shifted exponential distribution is used for the mixing random variable. We show that the probability density function (pdf) of the proposed distribution has a closed-form expression. The use of the MSSEN distribution within the finite mixture model paradigm is then illustrated. Thus, in the fashion of Lin et al. (2007), Karlis and Santourian (2009), and Browne and McNicholas (2015), the present paper aims to add to the richness of non-normal model-based clustering approaches an alternative mixture model that is easily applicable to data having skewed and heavy-tailed subpopulations. Indeed, different shapes of the scatter can produce the same level of skewness and kurtosis, as measured by classical indexes. The existing multivariate skewed heavy-tailed distributions can often reach all the levels of these indexes, by working with the tailedness and skewness parameters, but with a reference to a particular shape. So, for real data, it is important to have a variety of models that can provide different possibilities for these shapes. Furthermore, the proposed model has convenient closed-form expressions for estimating its parameter at the M-step of the EM algorithm (Dempster et al. 1977), differently from

other well-established mixture models, making its implementation and application easier.

The rest of the paper is organized as follows. In Sect. 2, we first present the definition and properties of the MSSEN distribution together with the introduction of the MSSEN mixture (MSSEN-M) model for clustering purposes. Also, an EM algorithm (Dempster et al. 1977) is described for maximum likelihood parameter estimation. In Sect. 3, we first compare the fitting behavior of the MSSEN-M model with those of well-established alternatives on simulated data. Then, an analysis of a real dataset concerning the log returns of four cryptocurrencies is conducted. Finally, some concluding remarks are done in Sect. 4.

2 Methodology

Herein, we present the MSSEN distribution (Sect. 2.1), its use within the mixture model paradigm (Sect. 2.2), and the EM algorithm for maximum likelihood parameter estimation (Sect. 2.3).

2.1 Multivariate skew shifted exponential normal distribution

Definition 1 A random vector $\boldsymbol{X} \in R^d$ is said to have a multivariate skew shifted exponential normal (MSSEN) distribution with location parameter $\boldsymbol{\xi} \in R^d$, $d \times d$ scale matrix $\boldsymbol{\Sigma}$, asymmetry parameter $\boldsymbol{\gamma} \in R^d$, and tailedness parameter $\theta > 0$, in symbols $\boldsymbol{X} \sim \mathcal{SSEN}_d(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}, \theta)$, if its pdf is given by

$$f_{\text{MSSEN}}(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}, \theta) = c \frac{\exp\left[(\boldsymbol{x}-\boldsymbol{\xi})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\right]}{\left[\delta(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\Sigma}) + 2\theta\right]^{\frac{d+2}{4}}} \times L_{\frac{d}{2}+1}\left(\sqrt{\left[\delta(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\Sigma}) + 2\theta\right]\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}, \sqrt{\frac{\delta(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\Sigma}) + 2\theta}{\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}\right). \tag{1}$$

In (1), $\delta(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\Sigma}) = (\boldsymbol{x}-\boldsymbol{\xi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\xi})$ is the squared Mahalanobis distance between $\boldsymbol{x}$ and $\boldsymbol{\xi}$ (with $\boldsymbol{\Sigma}$ as covariance matrix), c is a normalizing constant given by

$$c = \frac{2\theta \exp(\theta)\left(\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\right)^{\frac{d+2}{4}}}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}},$$

with $|\cdot|$ denoting the determinant, and

$$L_\lambda(x, z) = \frac{1}{2}\int_z^\infty t^{\lambda-1}\exp\left[-\frac{x}{2}\left(\frac{1}{t}+t\right)\right]dt \tag{2}$$

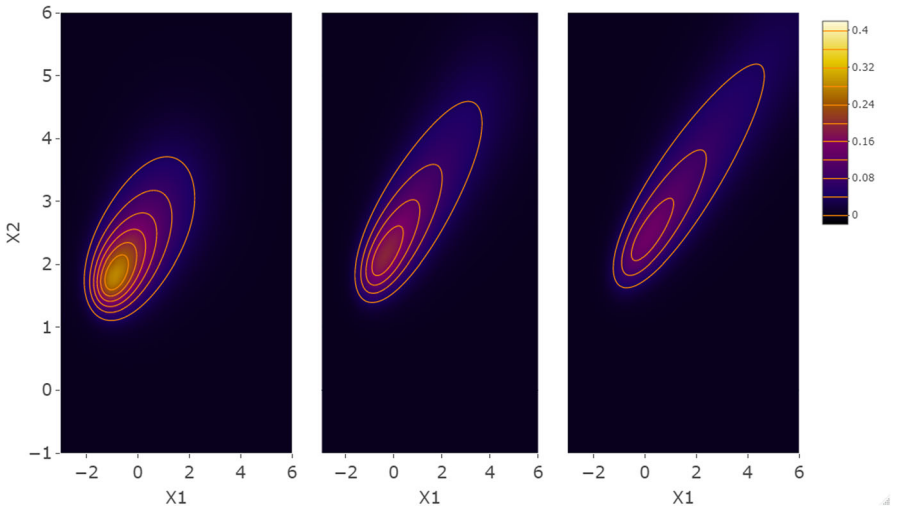


Fig. 1 Contour levels of the MSSEN pdf, in the bivariate case, for different values of $\boldsymbol{\gamma}$ (i.e., $\boldsymbol{\gamma} = (3, 3)'$ on the left, $\boldsymbol{\gamma} = (5, 5)'$ in the center, $\boldsymbol{\gamma} = (7, 7)'$ on the right), keeping fixed the other parameters (i.e., $\boldsymbol{\xi} = (0, 0)'$, $\boldsymbol{\Sigma} = \boldsymbol{I}_2$, and $\theta = 0.5$)

is the incomplete Bessel function defined by Terras (1981), see also Agrest et al. (1971). Note that, when $z = 0$, the function in (2) reduces to $K_\lambda(x)$, i.e. to the modified Bessel function of the third order; for details see Abramowitz and Stegun (1965) and Watson (1995).

To facilitate parameter interpretation, we show the effects of varying $\boldsymbol{\gamma}$, the other parameters kept fixed, in the bivariate case ($d = 2$). This information is reported via the contour levels of the MSSEN pdf in Fig. 1, where $\boldsymbol{\gamma} = (3, 3)'$ on the left, $\boldsymbol{\gamma} = (5, 5)'$ in the center, and $\boldsymbol{\gamma} = (7, 7)'$ on the right. The values of the other parameters are $\boldsymbol{\xi} = (0, 0)'$, $\boldsymbol{\Sigma} = \boldsymbol{I}_2$, and $\theta = 0.5$, being $\boldsymbol{I}_2$ the 2×2 identity matrix. As we note, the asymmetry of the pdf increases as $\boldsymbol{\gamma}$ grows.

Similarly, we report in Fig. 2 the effect of varying θ, the other parameters kept fixed, in the bivariate case ($d = 2$). The subplots refer to $\theta = 1$ (on the left), $\theta = 0.5$ (in the center), and $\theta = 0.1$ (on the right). The values of the other parameters are $\boldsymbol{\xi} = (0, 0)'$, $\boldsymbol{\Sigma} = \boldsymbol{I}_2$, and $\gamma = (2, 2)'$. As we see, the peakedness of the pdf grows when θ decreases. This has implications for the tail behavior by the fact that large values of θ imply light tails, whereas smaller values of θ lead to heavier tails.

2.1.1 Representations

A more direct interpretation of $\boldsymbol{X} \sim \mathcal{SSEN}_d(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}, \theta)$ can be given by its hierarchical representation as

$$\begin{aligned} W &\sim \mathcal{SE}(\theta), \\ \boldsymbol{X}|W = w &\sim \mathcal{N}_d(\boldsymbol{\xi} + \boldsymbol{\gamma}/w, \boldsymbol{\Sigma}/w), \end{aligned} \tag{3}$$

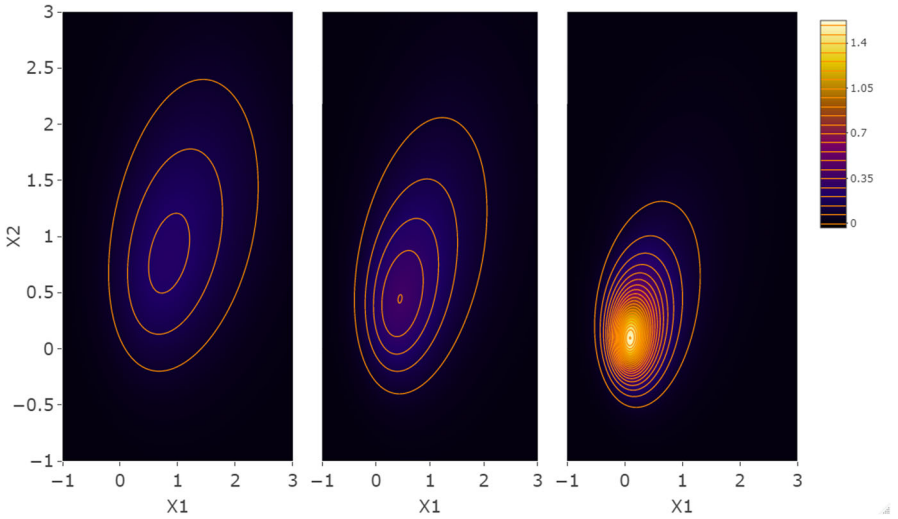


Fig. 2 Contour levels of the MSSEN pdf, in the bivariate case, for different values of θ (i.e., $\theta = 1$ on the left, $\theta = 0.5$ in the center, $\theta = 0.1$ on the right), keeping fixed the other parameters (i.e., $\boldsymbol{\xi} = (0, 0)'$, $\boldsymbol{\Sigma} = \boldsymbol{I}_2$, and $\gamma = (2, 2)'$)

where $\mathcal{SE}(\theta)$ denotes a shifted exponential distribution on $(1, \infty)$, with rate parameter $\theta > 0$ and pdf $f_{\mathrm{SE}}(w; \theta) = \theta \exp[-\theta(w - 1)]$, and $\mathcal{N}_d(\cdot)$ denotes a d-variate normal distribution. This alternative way to see the MSSEN distribution is useful for random number generation and for the implementation of the EM algorithm (see Sect. 2.3). Moreover, $\boldsymbol{X} \sim \mathcal{SSEN}_d(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}, \theta)$ has the following normal mean-variance mixture representation

$$\boldsymbol{X} = \boldsymbol{\xi} + \frac{1}{W}\boldsymbol{\gamma} + \frac{1}{\sqrt{W}}\boldsymbol{\Lambda}\boldsymbol{Z}, \tag{4}$$

where $W \sim \mathcal{SE}(\theta)$, $\boldsymbol{\Lambda}$ is a $d \times d$ matrix such that $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \boldsymbol{\Sigma}$, and $\boldsymbol{Z} \sim \mathcal{N}_d(\boldsymbol{0}_d, \boldsymbol{I}_d)$ is independent of W, being $\boldsymbol{0}_d$ and $\boldsymbol{I}_d$ the vector of zeros and the $d \times d$ identity matrix, respectively; see, e.g., McNeil et al. (2015).

2.1.2 Moments

Thanks to the representation given in (4), and by the well-known results illustrated in McNeil et al. (2015), it is possible to show that the mean and covariance matrix for $\boldsymbol{X} \sim \mathcal{SSEN}_d(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}, \theta)$ are respectively given by

$$\boldsymbol{\mu} := \mathrm{E}(\boldsymbol{X}) = \boldsymbol{\xi} + s(\theta)\boldsymbol{\gamma} \tag{5}$$

$$\boldsymbol{\Phi} := \mathrm{Cov}(\boldsymbol{X}) = s(\theta)\boldsymbol{\Sigma} + \{\theta - s(\theta)[\theta + s(\theta)]\}\boldsymbol{\gamma}\boldsymbol{\gamma}', \tag{6}$$

where

$$s(\theta) = \theta \exp(\theta)\varphi_{-1}(\theta)$$

and

$$\begin{aligned}\varphi_m(z) =& E_{-m}(z)\\ =& \int_1^{\infty} t^m \exp(-zt)dt\\ =& z^{-(m+1)}\Gamma(m+1, z)\end{aligned} \tag{7}$$

is the Misra function (Misra 1940), generalized form of the exponential integral function $E_n(z) = \int_1^{\infty} t^{-n}\exp(-zt)dt$ (Abramowitz and Stegun 1965).

To illustrate the behavior of skewness and kurtosis coefficients and their relation with the parameters $\boldsymbol{\gamma}$ and θ, we examine the univariate case ($d = 1$) and we use the notation $X \sim \mathcal{SSEN}_1\left(\xi, \sigma^2, \gamma, \theta\right)$. These coefficients are respectively given by

$$\text{Skew}(X) = \frac{m_3 - 3m_1m_2 + 2m_1^3}{\left(m_2 - m_1^2\right)^{\frac{3}{2}}}, \tag{8}$$

$$\text{Kurt}(X) = \frac{m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4}{\left(m_2 - m_1^2\right)^2} - 3, \tag{9}$$

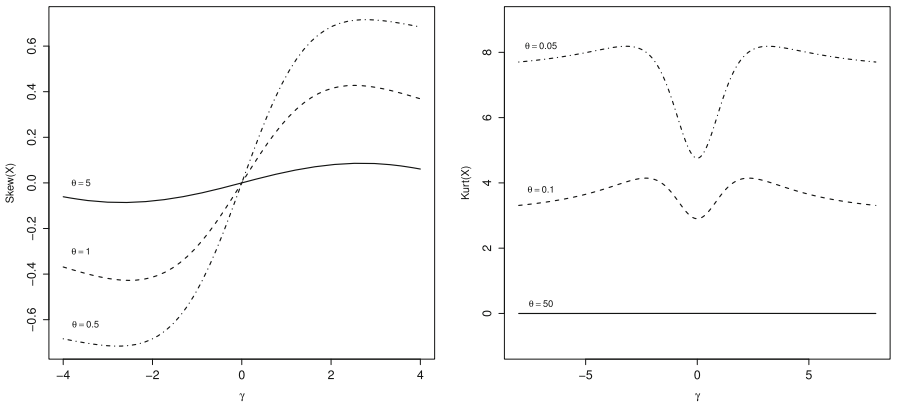
where

$$\begin{aligned}m_1 =& \text{E}(X) = \xi + \gamma\text{E}\left(\frac{1}{W}\right),\\ m_2 =& \text{E}\left(X^2\right) = \xi^2 + \left(2\gamma\xi + \sigma^2\right)\text{E}\left(\frac{1}{W}\right),\\ m_3 =& \text{E}\left(X^3\right) = \xi^3 + 3\left(\gamma\xi + \sigma^2\right)\left[\xi\text{E}\left(\frac{1}{W}\right) + \gamma\text{E}\left(\frac{1}{W^2}\right)\right] + \gamma^3\text{E}\left(\frac{1}{W^3}\right),\\ m_4 =& \text{E}\left(X^4\right) = \xi^4 + 2\left(2\gamma\xi + 3\sigma^2\right)\left[\xi^2\text{E}\left(\frac{1}{W}\right) + \gamma^2\text{E}\left(\frac{1}{W^3}\right)\right] +\\ &+ 3\left(2\gamma^2\xi + 4\gamma\xi\sigma^2 + \sigma^2\right)\text{E}\left(\frac{1}{W^2}\right) + \gamma^4\text{E}\left(\frac{1}{W^4}\right),\end{aligned}$$

with

$$\text{E}\left(\frac{1}{W^r}\right) = \theta\exp(\theta)E_r(\theta).$$

Figure 3 shows some examples of Skew(X) and Kurt(X), as a function of γ, for some values of θ ($\xi = 0$ and $\sigma^2 = 1$). From Fig. 3a we realize that, as θ decreases, the range of possible values of Skew(X) increases. Moreover, the skewness is zero when



(a) Skewness (b) Kurtosis

Fig. 3 Examples of behavior of Skew(X) (on the left) and Kurt(X) (on the right), as functions of γ, at various levels of θ for univariate SSEN distribution, given $\xi = 0$ and $\sigma^2 = 1$

$\gamma = 0$, and this happens regardless of θ. From Fig. 3b we note that γ kept fixed, the lower the value θ, the higher the kurtosis. Moreover, as θ increases, the kurtosis tends to zero regardless of γ. The behaviors of skewness and kurtosis shown in Fig. 3 are very similar to those for the asymmetric Laplace scale mixture models proposed by Punzo and Bagnato (2022).

2.1.3 Comparison with other skewed distributions

By focusing on the skewed distributions mentioned in Sect. 1, we begin to distinguish them according to whether or not they have heavier-than-normal tails. In particular, it is well-known that the MSN distribution, despite the existence of several formulations (see Lee and McLachlan 2013b), is not well-suited for modeling data showing heavier tails than the normal ones (Azzalini 2013). In such cases, it is necessary to adopt more flexible distributions, such as the MST (in its varying configurations; see Lee and McLachlan, 2013b), MSCN, MSS, MVG, MVGH, or our MSSEN, which have specific parameters controlling the heaviness of the tails.

Another aspect to consider is the definiteness of the moments, which is dependent on the values assumed by the parameters for certain distributions. For example, the moments of MST and MSS distributions are defined only when the tailedness parameters are greater than certain thresholds (see, e.g., Wang and Genton 2006; Dávila et al. 2018). These thresholds act as an obstacle to obtaining values that would result in heavier tails, thereby limiting the tail flexibility of these distributions. On the contrary, the MSSEN and other skewed distributions being considered do not encounter this issue, as their moments always exist. Moments calculation have important implications in many applied fields, such as economics and finance (Rachev et al. 2010; McNeil et al. 2015).

The considered skewed distributions also require different computational efforts for estimating their parameters. We remand this discussion to Sect. 2.2.1, where they are compared within a mixture model setting.

2.2 Finite mixtures of MSSEN distributions

For a d-variate random vector $\boldsymbol{X}$, the pdf of the MSSEN mixture (MSSEN-M) model with k components can be written as

$$f_{\text{MSSEN-M}}(\boldsymbol{x};\boldsymbol{\Psi}) = \sum_{j=1}^{k} \pi_j \, f_{\text{MSSEN}}\left(\boldsymbol{x};\boldsymbol{\Omega}_j\right), \tag{10}$$

where π_j is the mixing proportion of the jth component, with $\pi_j > 0$ and $\sum_{j=1}^{k} \pi_j = 1$, $f_{\text{MSSEN}}\left(\boldsymbol{x};\boldsymbol{\Omega}_j\right)$ is the pdf defined in (1) of the jth component with parameters $\boldsymbol{\Omega}_j = \left\{\boldsymbol{\xi}_j, \boldsymbol{\Sigma}_j, \boldsymbol{\gamma}_j, \theta_j\right\}$, and $\boldsymbol{\Psi} = \left\{\left(\pi_j, \boldsymbol{\Omega}_j\right);\, j = 1, \ldots, k\right\}$ contains all of the parameters of the mixture. Model in (10), being able to account for heterogeneity and non-normal features, such as asymmetry and heavy tails in each mixture component, constitutes a promising alternative to the existing skew mixture models.

2.2.1 Comparison with other skewed mixtures

As introduced in Sect. 1, different mixture models with asymmetric components have been proposed in the multivariate literature. First of all, we report that the features presented in Sect. 2.1.3 for the single skewed distributions directly apply to the corresponding mixture models. Here, we want to add some computational differences among them.

We start by considering the MSS mixture (MSS-M) model for which, as discussed by Prates et al. (2013), the updating expressions for all the parameters in the EM-based algorithm are not in closed form. Particularly, Monte Carlo integration may be employed, making this model the most computationally intensive to be estimated among those considered. The situation improves when the MST mixture (MST-M), MVG mixture (MVG-M), MSCN mixture (MSCN-M), and MGH mixture (MGH-M) models are considered, given that for these models only the tailedness parameters must be numerically estimated (Browne and McNicholas 2015; Prates et al. 2013; Gallaugher et al. 2022b). Anyway, estimating the tailedness parameters accurately can be a challenging task, such as in the case of the MGH-M model (Aas and Hobæk Haff 2005; Browne and McNicholas 2015). On the other hand, the MSSEN-M model is the only heavy-tailed (among those considered) having a simple closed-form expression also for the update of the tailedness parameter (refer to Sect. 2.3), making it computationally the most efficient.

A last note can be reported in terms of parsimony. Specifically, the MSN-M model is the most parsimonious (due to the lack of the tailedness parameter), whereas the MSCN-M and MGH-M models are the most parametrized (since they have two parameters, for each group, controlling the tail behavior).

2.3 Maximum likelihood: application of the EM algorithm

A standard approach to estimating the parameters of a mixture model is the maximum likelihood (ML), commonly implemented via the expectation-maximization (EM) algorithm. Let $\{\boldsymbol{X}_i, i = 1, \ldots, n\}$ be a sample of n independent observations from model (10). In the EM algorithm framework, such a sample is treated as incomplete. The complete-data are $\{(\boldsymbol{X}_i, \boldsymbol{z}_i, w_i); i = 1, \ldots, n\}$, where $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{ik})'$ so that $z_{ij} = 1$ if observation i belongs to group j and $z_{ij} = 0$ otherwise. By saying that $\boldsymbol{Z}_i$, the random counterpart of $\boldsymbol{z}_i$, follows a multinomial distribution consisting of one draw from k categories with probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)'$, say $\mathcal{M}_k(\boldsymbol{\pi})$, the hierarchical representation mentioned in (3) for the single distribution can be generalized to the corresponding mixture model as

$$\begin{aligned} \boldsymbol{Z}_i &\sim \mathcal{M}_k(\boldsymbol{\pi}) \\ W_{ij} &\sim \mathcal{SE}\big(\theta_j\big) \\ \boldsymbol{X}_i | W_{ij} = w_{ij} &\sim \mathcal{N}_d(\boldsymbol{\xi}_j + \boldsymbol{\gamma}_j/w_{ij}, \boldsymbol{\Sigma}_j/w_{ij}), \end{aligned} \tag{11}$$

where $W_{ij} := W_i | z_{ij} = 1$. Because of this conditional structure, the complete-data log-likelihood function can be written as

$$\ell_c(\boldsymbol{\Psi}) = \ell_{1c}(\boldsymbol{\pi}) + \ell_{2c}(\boldsymbol{\Xi}) + \ell_{3c}(\boldsymbol{\theta}), \tag{12}$$

where

$$\ell_{1c}(\boldsymbol{\pi}) = \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij} \ln\big(\pi_j\big), \tag{13}$$

$$\begin{aligned} \ell_{2c}(\boldsymbol{\Xi}) = \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij}\Bigg[&-\frac{d}{2}\ln(2\pi) + \frac{d}{2}\ln\big(w_{ij}\big) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_j| - \frac{w_{ij}\delta_j\big(\boldsymbol{x}_i; \boldsymbol{\xi}_j, \boldsymbol{\Sigma}_j\big)}{2} + \\ &+ \frac{1}{2}\big(\boldsymbol{x}_i - \boldsymbol{\xi}_j\big)'\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\gamma}_j + \frac{1}{2}\boldsymbol{\gamma}_j'\boldsymbol{\Sigma}_j^{-1}\big(\boldsymbol{x}_i - \boldsymbol{\xi}_j\big) - \frac{1}{2w_{ij}}\boldsymbol{\gamma}_j'\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\gamma}_j\Bigg], \end{aligned} \tag{14}$$

with $\boldsymbol{\Xi} = \big\{\big(\boldsymbol{\xi}_j, \boldsymbol{\Sigma}_j, \boldsymbol{\gamma}_j\big); j = 1, \ldots, k\big\}$, and

$$\ell_{3c}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij}\big[\ln\big(\theta_j\big) - \theta_j\big(w_{ij} - 1\big)\big], \tag{15}$$

with $\boldsymbol{\theta} = \big\{\theta_j, j = 1, \ldots, k\big\}$.

Working on $\ell_c(\boldsymbol{\Psi})$ in (12), the EM algorithm iterates between two steps, one E-step, and one M-step, until convergence. These steps are outlined below. As for the notation, the parameters marked with one dot correspond to the updates at the previous iteration while those marked with two dots represent the updates at the current iteration.

E-Step

The E-step requires the calculation of

$$Q\left(\boldsymbol{\Psi}|\dot{\boldsymbol{\Psi}}\right) = Q_1\left(\boldsymbol{\pi}|\dot{\boldsymbol{\Psi}}\right) + Q_2\left(\boldsymbol{\Xi}|\dot{\boldsymbol{\Psi}}\right) + Q_3\left(\boldsymbol{\theta}|\dot{\boldsymbol{\Psi}}\right), \tag{16}$$

the conditional expectation of $l_c(\boldsymbol{\Psi})$ given the observed data and using the current fit $\dot{\boldsymbol{\Psi}}$ for $\boldsymbol{\Psi}$. In (16) the three terms on the right-hand side are ordered as the three terms on the right-hand side of (12).

To compute $Q\left(\boldsymbol{\Psi}|\dot{\boldsymbol{\Psi}}\right)$ we need to replace any function $g\left(Z_{ij}\right)$ and $m\left(W_{ij}\right)$ of the latent variables which arise in (13)–(15) by $E_{\dot{\boldsymbol{\Psi}}}\left[g\left(Z_{ij}\right)|\boldsymbol{X}_i = \boldsymbol{x}_i\right]$ and $E_{\dot{\boldsymbol{\Psi}}}\left[m\left(W_{ij}\right)|\boldsymbol{X}_i = \boldsymbol{x}_i\right]$, respectively, where the expectations are taken using the current fit $\dot{\boldsymbol{\Psi}}$ for $\boldsymbol{\Psi}$, $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

As concerns Z_{ij}, it is included in all the Eqs. (13)–(15) but always via the identity function, namely $g(z) = z$; therefore

$$\ddot{z}_{ij} := \mathrm{E}_{\dot{\boldsymbol{\Psi}}}\left(Z_{ij}|\boldsymbol{x}_i\right) = \frac{\dot{\pi}_j f_{\mathrm{MSSEN}}\left(\boldsymbol{x}_i; \dot{\boldsymbol{\Omega}}_j\right)}{\sum_{h=1}^{k} \dot{\pi}_h f_{\mathrm{MSSEN}}\left(\boldsymbol{x}_i; \dot{\boldsymbol{\Omega}}_h\right)},$$

which corresponds to the posterior probability that the unlabeled observation $\boldsymbol{x}_i$ belongs to the jth component of the mixture.

Regarding W_{ij}, it is only included in the Eqs. (14) and (15) and the functions involved are $m_1(w) = w$, $m_2(w) = 1/w$, $m_3(w) = \ln(w)$. Of these, only m_1 and m_2 need to be taken into account because m_3 in (14) is not related to the parameters. To calculate the expectations of m_1 and m_2 we first note that the pdf of $W_{ij}|\boldsymbol{X}_i = \boldsymbol{x}_i$ satisfies

$$\begin{aligned} f\left(w_{ij}|\boldsymbol{x}_i; \boldsymbol{\Omega}_j\right) &\propto f\left(w_{ij}, \boldsymbol{x}_i; \boldsymbol{\Omega}_j\right) \\ &\propto w_{ij}^{\frac{d}{2}} \exp\left[-\frac{\boldsymbol{\gamma}_j' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\gamma}_j}{2w_{ij}} - \frac{\delta\left(\boldsymbol{x}_i; \boldsymbol{\xi}_j, \boldsymbol{\Sigma}_j\right)}{2} w_{ij}\right] \\ &\quad \times \exp\left(-\theta_j w_{ij}\right) \mathbb{1}_{(1,\infty)}\left(w_{ij}\right) \\ &\propto f_{\mathrm{LTGIG}}\left(w_{ij}; \frac{d}{2} + 1, \boldsymbol{\gamma}_j' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\gamma}_j, \delta\left(\boldsymbol{x}_i; \boldsymbol{\xi}_j, \boldsymbol{\Sigma}_j\right) + 2\theta_j, 1\right), \end{aligned}$$

where $\mathbb{1}_A(\cdot)$ is the indicator function on the set A and f_{LTGIG} is the pdf of a left truncated generalized inverse Gaussian distribution, which is defined in Appendix A of this manuscript. This means that

$$W_{ij}|\boldsymbol{X}_i = \boldsymbol{x}_i \sim \mathcal{LTGIG}_{(1,\infty)}\left(\frac{d}{2} + 1, \boldsymbol{\gamma}_j' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\gamma}_j, \delta\left(\boldsymbol{x}_i; \boldsymbol{\xi}_j, \boldsymbol{\Sigma}_j\right) + 2\theta_j\right).$$

From Eq. (A2) we can derive the following results

$$\ddot{u}_{ij} := \mathrm{E}_{\dot{\boldsymbol{\Psi}}}\left(W_{ij}|\boldsymbol{X}_i = \boldsymbol{x}_i\right)$$

$$= \sqrt{\frac{\dot{\boldsymbol{\gamma}}_j' \dot{\boldsymbol{\Sigma}}_j^{-1} \dot{\boldsymbol{\gamma}}_j}{\left[\delta\left(\boldsymbol{x}_i; \dot{\boldsymbol{\xi}}_j, \dot{\boldsymbol{\Sigma}}_j\right) + 2\dot{\theta}_j\right]}} \times \frac{L_{\frac{d}{2}+2}\left(\sqrt{\dot{\boldsymbol{\gamma}}_j' \dot{\boldsymbol{\Sigma}}_j^{-1} \dot{\boldsymbol{\gamma}}_j \left[\delta\left(\boldsymbol{x}_i; \dot{\boldsymbol{\xi}}_j, \dot{\boldsymbol{\Sigma}}_j\right) + 2\dot{\theta}_j\right]}, \sqrt{\frac{\delta\left(\boldsymbol{x}_i; \dot{\boldsymbol{\xi}}_j, \dot{\boldsymbol{\Sigma}}_j\right) + 2\dot{\theta}_j}{\dot{\boldsymbol{\gamma}}_j' \dot{\boldsymbol{\Sigma}}_j^{-1} \dot{\boldsymbol{\gamma}}_j}}\right)}{L_{\frac{d}{2}+1}\left(\sqrt{\dot{\boldsymbol{\gamma}}_j' \dot{\boldsymbol{\Sigma}}_j^{-1} \dot{\boldsymbol{\gamma}}_j \left[\delta\left(\boldsymbol{x}_i; \dot{\boldsymbol{\xi}}_j, \dot{\boldsymbol{\Sigma}}_j\right) + 2\dot{\theta}_j\right]}, \sqrt{\frac{\delta\left(\boldsymbol{x}_i; \dot{\boldsymbol{\xi}}_j, \dot{\boldsymbol{\Sigma}}_j\right) + 2\dot{\theta}_j}{\dot{\boldsymbol{\gamma}}_j' \dot{\boldsymbol{\Sigma}}_j^{-1} \dot{\boldsymbol{\gamma}}_j}}\right)},$$

$$\ddot{v}_{ij} := \mathrm{E}_{\dot{\boldsymbol{\Psi}}}\left(W_{ij}^{-1} | X_i = \boldsymbol{x}_i\right)$$

$$= \sqrt{\frac{\left[\delta\left(\boldsymbol{x}_i; \dot{\boldsymbol{\xi}}_j, \dot{\boldsymbol{\Sigma}}_j\right) + 2\dot{\theta}_j\right]}{\dot{\boldsymbol{\gamma}}_j' \dot{\boldsymbol{\Sigma}}_j^{-1} \dot{\boldsymbol{\gamma}}_j}} \times \frac{L_{\frac{d}{2}}\left(\sqrt{\dot{\boldsymbol{\gamma}}_j' \dot{\boldsymbol{\Sigma}}_j^{-1} \dot{\boldsymbol{\gamma}}_j \left[\delta\left(\boldsymbol{x}_i; \dot{\boldsymbol{\xi}}_j, \dot{\boldsymbol{\Sigma}}_j\right) + 2\dot{\theta}_j\right]}, \sqrt{\frac{\delta\left(\boldsymbol{x}_i; \dot{\boldsymbol{\xi}}_j, \dot{\boldsymbol{\Sigma}}_j\right) + 2\dot{\theta}_j}{\dot{\boldsymbol{\gamma}}_j' \dot{\boldsymbol{\Sigma}}_j^{-1} \dot{\boldsymbol{\gamma}}_j}}\right)}{L_{\frac{d}{2}+1}\left(\sqrt{\dot{\boldsymbol{\gamma}}_j' \dot{\boldsymbol{\Sigma}}_j^{-1} \dot{\boldsymbol{\gamma}}_j \left[\delta\left(\boldsymbol{x}_i; \dot{\boldsymbol{\xi}}_j, \dot{\boldsymbol{\Sigma}}_j\right) + 2\dot{\theta}_j\right]}, \sqrt{\frac{\delta\left(\boldsymbol{x}_i; \dot{\boldsymbol{\xi}}_j, \dot{\boldsymbol{\Sigma}}_j\right) + 2\dot{\theta}_j}{\dot{\boldsymbol{\gamma}}_j' \dot{\boldsymbol{\Sigma}}_j^{-1} \dot{\boldsymbol{\gamma}}_j}}\right)}.$$

M-Step

The M-step requires the calculation of $\ddot{\boldsymbol{\Psi}}$ that maximizes $Q\left(\boldsymbol{\Psi}|\dot{\boldsymbol{\Psi}}\right)$ in (16). Simple algebra yields the following updates

$$\ddot{\pi}_j = \frac{1}{n}\sum_{i=1}^{n} \ddot{z}_{ij}, \quad \ddot{\boldsymbol{\gamma}}_j = \frac{\frac{1}{n_j}\sum_{i=1}^{n} \ddot{z}_{ij}\ddot{u}_{ij}\left(\bar{\boldsymbol{x}}_j - \boldsymbol{x}_i\right)}{\bar{u}_j \bar{v}_j - 1}, \quad \ddot{\boldsymbol{\xi}}_j = \frac{\frac{1}{n_j}\sum_{i=1}^{n} \ddot{z}_{ij}\ddot{u}_{ij}\boldsymbol{x}_i - \ddot{\boldsymbol{\gamma}}_j}{\bar{u}_j},$$

$$\ddot{\boldsymbol{\Sigma}}_j = \frac{1}{n_j}\sum_{i=1}^{n} \ddot{z}_{ij}\ddot{u}_{ij}\left(\boldsymbol{x}_i - \ddot{\boldsymbol{\xi}}_j\right)\left(\boldsymbol{x}_i - \ddot{\boldsymbol{\xi}}_j\right)' - \left(\bar{\boldsymbol{x}}_j - \ddot{\boldsymbol{\xi}}_j\right)\ddot{\boldsymbol{\gamma}}' - \ddot{\boldsymbol{\gamma}}\left(\bar{\boldsymbol{x}}_j - \ddot{\boldsymbol{\xi}}_j\right)' + \bar{v}_j \ddot{\boldsymbol{\gamma}}\ddot{\boldsymbol{\gamma}}',$$

$$\ddot{\theta}_j = \frac{\sum_{i=1}^{n} \ddot{z}_{ij}}{\sum_{i=1}^{n} \ddot{z}_{ij}\left(\ddot{u}_{ij} - 1\right)},$$

where $n_j = \sum_{i=1}^{n} \ddot{z}_{ij}$, $\bar{u}_j = (1/n_j)\sum_{i=1}^{n} \ddot{z}_{ij}\ddot{u}_{ij}$, $\bar{v}_j = (1/n_j)\sum_{i=1}^{n} \ddot{z}_{ij}\ddot{v}_{ij}$, and $\bar{\boldsymbol{x}}_j = (1/n_j)\sum_{i=1}^{n} \ddot{z}_{ij}\boldsymbol{x}_i$, $j = 1, \ldots, k$.

Interestingly, closed-form expressions are provided for all the mixture parameters, differently from the many other well-established skewed mixture models (refer to the contributions reported in Sect. 1 and comments made in Sect. 2.2.1).

2.3.1 Some notes on the initialization strategy

To start the EM algorithm, we adopt the following two-step procedure:

1. partition the sample into k groups using the k-means clustering algorithm (Hartigan and Wong 1979). Note that, the best solution over 25 runs of the k-means algorithm is considered.

2. Based on the obtained classification, compute the weighted sample mean, covariance, and skewness to initialize $\boldsymbol{\xi}_j$, $\boldsymbol{\Sigma}_j$, and $\boldsymbol{\gamma}_j$, $j = 1, \ldots, k$, respectively. The tailedness parameter θ_j, $j = 1, \ldots, k$, is set equal to 1, i.e. an intermediate value between near-normality and heavy-tails. The mixture weights π_j, $j = 1, \ldots, k$, are initialized by the proportion of points classified in each group.

Closely related initialization strategies have been largely adopted in the model-based clustering literature (see Basso et al. 2010; Prates et al. 2013; Lee and McLachlan 2014; Wei et al. 2019, for some examples).

3 Data analyses

In this section, we compare the performance of our proposed model with those of several reference finite mixture models. In detail, all the models are fitted to simulated data, in Sect. 3.1, and to a real dataset concerning the log returns of four cryptocurrencies, in Sect. 3.2. All the analyses are conducted via the R statistical software.

3.1 A simulation study

In this study, we consider two scenarios that differ according to the direction of the asymmetry present in the data. In particular, in Scenario A the groups show positive asymmetry, whereas in Scenario B they have negative asymmetry. For each scenario, we generate samples of size $n = 200$ from the bivariate standard normal distribution, i.e. $\boldsymbol{X} \sim \mathcal{N}_2(\mathbf{0}_2, \boldsymbol{I}_2)$. Then, for the samples to be used in Scenario A, we apply the transformation $x + \exp(\eta x)$, whereas for the samples to be considered in Scenario B we use the transformation $-x - \exp(\eta x)$, being x the generic coordinate of $\boldsymbol{x}$. For each sample, half of the observations are selected and shifted in the bi-dimensional space by adding (subtracting) a constant $c = 10$ to each of their transformed values in Scenario A (Scenario B), thus leading to two separate clusters. In the transformations above, $\eta > 0$ can be used to control the levels of asymmetry and kurtosis of the samples. Specifically, as η grows, both quantities roughly increase. Herein, we let η be the sequence of values from 0.80 to 1.40, with increments of 0.10 and, for each value of η, we consider 500 samples. The transformation used in Scenario A has been recently considered by Gallaugher et al. (2022). Examples of generated datasets for the two scenarios are illustrated in Fig. 4.

On each generated dataset, we fit our MSSEN-M model together with 9 other mixture models; all the models are fitted with $k = 2$ components. In the following, we provide the detailed list of the considered models along with the R functions and packages used to fit them to the data:

- the MN-M, MSN-M, MT-M, MST-M, MSCN-M, and MSS-M models are fitted via the `smsn.search()` function of the **mixsmsn** package (Prates et al. 2013);
- the MGH-M and MVG-M models are fitted via the `ghpcm()` function of the **mixture** package (Pocuca et al. 2021);
- the MSEN-M model is fitted via the `Mixt.fit()` function of the **SenTinMixt** package (Tomarchio et al. 2021).

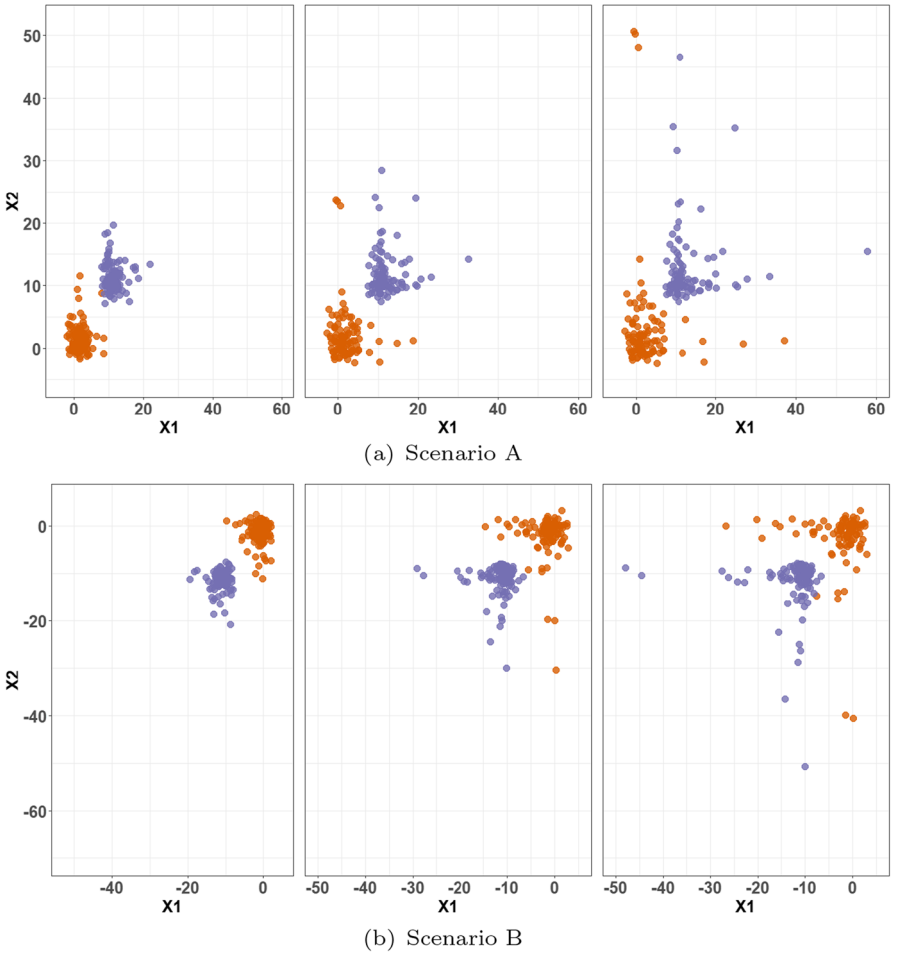


Fig. 4 Examples of generated datasets for the two scenarios. For each scenario, three growing examples of η are provided: $\eta = 0.8$ (left), $\eta = 1.1$ (center), and $\eta = 1.4$ (right)

Note that, in the list above, the acronyms MT-M and MSEN-M refer to the multivariate t and multivariate shifted exponential normal mixtures, respectively.

The fitting performances of the considered models are now analyzed. First of all, for each η we can obtain 500 model rankings based on the BIC in each scenario. Thus, we compute the average among these rankings to obtain an overall measure of the goodness of fit of each model, and we report such information in Fig. 5 for Scenario A, and in Fig. 6 for Scenario B.

It is evident that, regardless of the considered scenario, the patterns illustrated are substantially the same. In both scenarios, by starting from the bottom, we immediately note that, regardless of the value of η, the MN-M model provides the worst fitting because it is not able to account for both asymmetry and heavy tails. Then, we roughly have three groups of models. The models in the first group are the MSN-M, MT-M, and

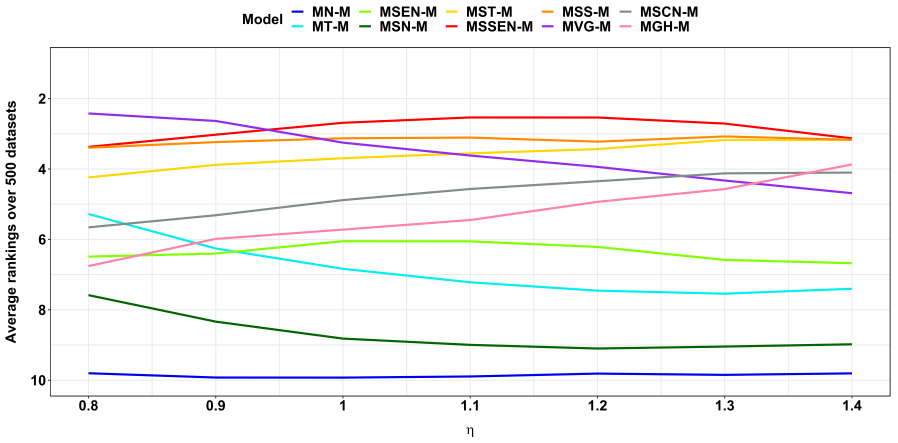


Fig. 5 Scenario A: average rankings based on the BIC for the considered models over the 500 simulated datasets at each value of η

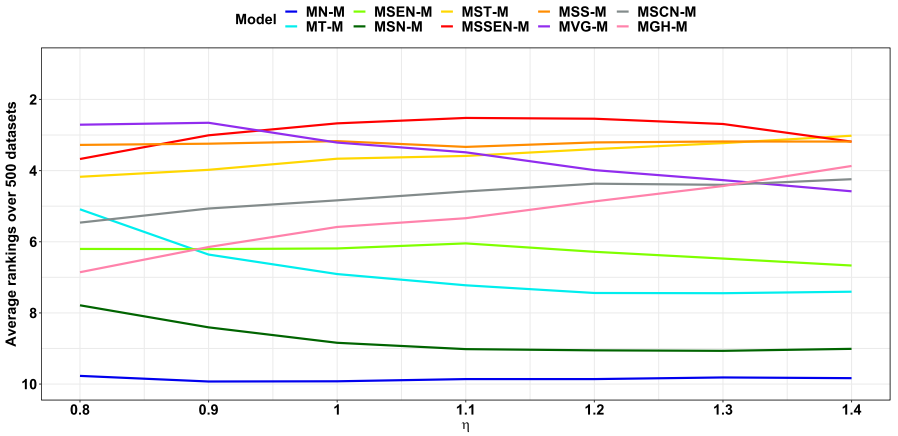


Fig. 6 Scenario B: average rankings based on the BIC for the considered models over the 500 simulated datasets at each value of η

MSEN-M models. They are ranked, on average, from the 6th to the 9th position, and are only partially able to account for the features of the generated data. The models in the second group are the MGH-M and MSCN-M models, namely the most parametrized among the considered mixture models. These models are capable of modeling both asymmetry and heavy tails. Nevertheless, the higher number of parameters causes an increase in the penalty term of the BIC, with a consequent negative effect on the fitting results. However, both models seem to improve their performance as η grows, suggesting that the extra flexibility granted by the additional parameters allows for better accommodation of the higher levels of asymmetry and kurtosis in the data (Browne and McNicholas 2015). The models in the third group generally cover, on

average, the first four average positions. It is possible to note that the MVG-M model, which initially provides the best average fit, deteriorates their performance when η increases. The MSS-M and MST-M models follow a similar pattern to each other and seem to provide a good (but not the best) fit of the data regardless of the η value considered. On the contrary, we notice an interesting pattern related to our MSSEN-M model. Indeed, for values of η ranging from 1.00 to 1.30, it has the best average ranking. Specifically, the average ranking smoothly increases until $\eta = 1.20$, a value from which it slowly begins to decrease and converge to that of the other three models. Therefore, we have identified a range of possible values of η, i.e. data structures, for which the MSSEN-M model yields better average fitting results than the other mixture models. More in general, if we would compute for each model the global ranking over all the η values, the MSSEN-M model would have the highest average ranking.

From another perspective, since when working with real data we might be interested only in the best fitting model, we now count how many times, over the 500 datasets for each η, every model produces the best BIC value. This information is reported in Fig. 7 for Scenario A and in Fig. 8 for Scenario B. The patterns reported are similar between the scenarios and roughly mimic those previously discussed. By focusing only on the MSSEN-M, we notice that, for values of η ranging from 1.00 to 1.30, our model is considered the best one with the highest frequency.

To summarize, the message of this simulation study is to show that the proposed MSSEN-M model can be useful in modeling asymmetric and heavy-tailed data. Clearly, there cannot be a model that fits all types of asymmetry and heavy-tailed data better than others. Thus, our suggestion is to add our proposal to the list of candidate models to be fitted.

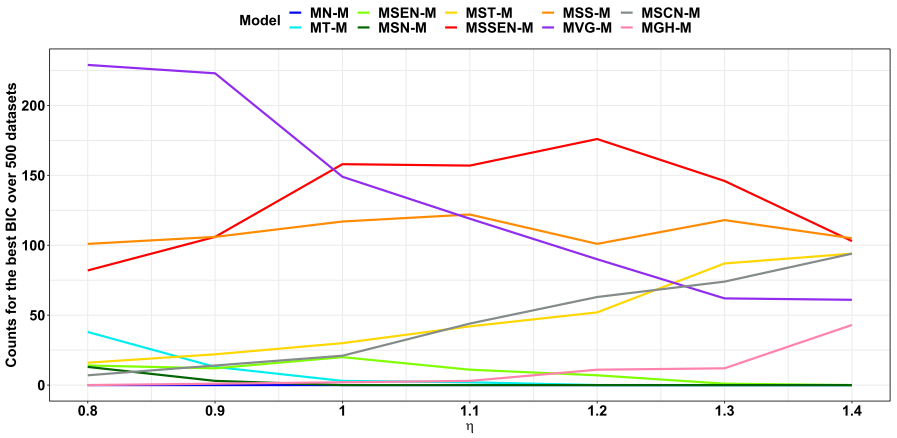


Fig. 7 Scenario A: number of times for which each model provides the best BIC over the 500 simulated datasets at each value of η

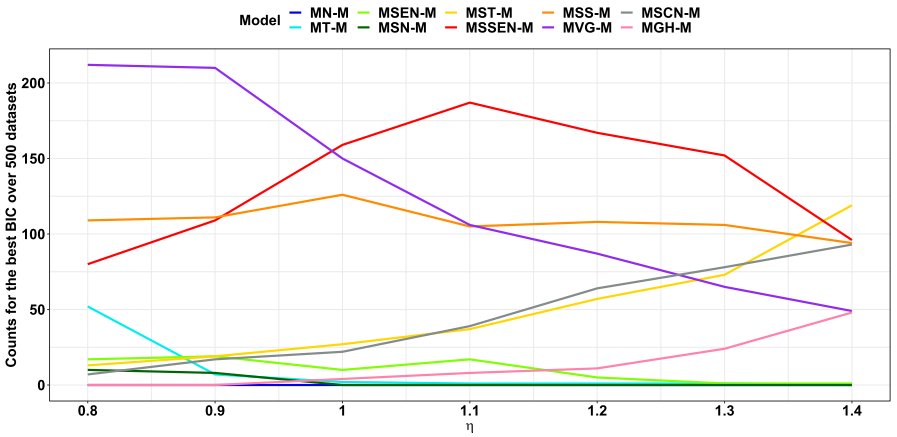


Fig. 8 Scenario B: number of times for which each model provides the best BIC over the 500 simulated datasets at each value of η

3.2 A real data illustration

Here, we analyze a cryptocurrency dataset. After a brief description of the background and the analyzed data in Sect. 3.2.1, we provide details on model fitting and comparison in Sect. 3.2.2. Then, we discuss the obtained results in Sect. 3.2.3.

3.2.1 Background and data description

After more than a decade of existence, the interest in the world of cryptocurrencies has exploded and cryptocurrencies are now an important class of financial assets. As such, the modelization of the log returns distribution constitutes an important aspect of analysis. Relatedly, it has been well-documented that the cryptocurrency log returns distribution presents deviation from normality both in terms of asymmetry and heavy tails (Zhang et al. 2018; Hu et al. 2019; de Melo et al. 2020). However, the majority of contributions focus on a separate (i.e. univariate) modelization of each cryptocurrency log returns distribution (see, e.g. Chu et al. 2015; Chan et al. 2017; Szczygielski et al. 2020; Punzo and Bagnato 2021, 2022), despite a multivariate analysis considering the interdependencies among cryptocurrencies would offer more complete financial information (Ji et al. 2019; Shi et al. 2020; Candila 2021).

Herein, we consider the daily adjusted close prices of the following four cryptocurrencies: DigiByte (DGB-EUR), LBRY Credits (LBC-EUR), Vexanium (VEX-EUR), and Curecoin (CURE-EUR). The data were downloaded from https://finance.yahoo.com/cryptocurrencies. All prices are in Euro. The period of analysis for the data goes from 18 September 2019 to 09 October 2021 and, as regularly done in the literature, returns are estimated by taking logarithmic differences. Thus, we have a 4-dimensional dataset with $n = 752$ observations.

3.2.2 Model fitting and comparison

On these data, we fitted our MSSEN-M model, as well as the other mixture models discussed in Sect. 3.1, for $k \in \{1, 2, 3, 4\}$. For each model, the BIC is used to select k, and a final ranking based on the models' BIC values is drawn up to assess the overall best-fitting one.

Furthermore, for each model whose k has been selected by the BIC, we compute the value-at-risk (VaR). The VaR is a commonly used financial risk management tool for measuring the risk of investment losses. Here, we implement the approach used by Soltyk and Gupta (2011) and Lee and McLachlan (2013a) for the analysis of stock returns via finite mixtures of multivariate distributions. Specifically, for a portfolio of d assets log returns $X_1, \ldots, X_d$, let $X_R = \sum_{h=1}^{d} X_h$ be the aggregate (or total asset) log-return. Then, the VaR is defined as the negative of the largest value of x_α satisfying

$$\Pr\{X_R < x_\alpha\} \leq \alpha,$$

where α is the significance level, that in our case will be set to $\alpha = 0.05$. Notice that the negative sign mentioned above ensures that the VaR is a positive value, i.e. a positive amount of loss. A large sample of size $n = 100{,}000$ is generated from each of the fitted models, and the VaR estimates are obtained by taking the appropriate quantile of the simulated total asset return samples. We draw up a ranking also for the VaR, which is now based on the absolute difference between the empirical VaR and that estimated by each model.

Finally, to evaluate the accuracy of the estimated VaR values, we implement the backtesting procedure proposed by Kupiec (1995), which is one of the most common statistical tests available for this purpose. In detail, this test examines, under the null hypothesis, if the proportion of violations given by $\hat{\rho} = v/n$, where v denotes the number of log returns exceeding the estimated VaR, is statistically different from the one expected (α). This test can be performed via the `VaRTest()` function of the **rugarch** package (Ghalanos 2020).

3.2.3 Results

Table 1 contains the results for the considered models. As we note, the overall best-fitting model is the MSSEN-M with $k = 2$. It is interesting to see that in both the MN-M and MSN-M models the BIC selects $k = 4$ groups in the data and they provide the worst fit among the considered models. This might be caused by the componentwise tails which are not thick enough to model the data. Furthermore, we also find that on only two occasions (MST-M and MSS-M) $k = 1$ is detected.

In terms of VaR, the empirical value is 31.03 and, in the fashion of Tomarchio and Punzo (2020), a ranking is introduced in order to simplify the reading, based on the percentage difference with respect to the empirical VaR; the lower the difference, the better the position in the ranking. The corresponding backtesting results are also provided in the last column. We see that the MSSEN-M model provides the most accurate estimate, closely followed by the MVG-M and MSEN-M models. The goodness of these VaRs is also supported by the corresponding backtesting p-values since they are

Table 1 Number of groups k, the value of the BIC and corresponding ranking for the considered models

Model	k	BIC	Ranking	VaR	Ranking	p value
MN-M	4	19539.76	9	20.48	4	0.00
MT-M	3	19403.35	4	14.71	7	0.00
MSEN-M	2	19378.84	2	32.10	3	0.66
MSN-M	4	19648.97	10	13.37	9	0.00
MST-M	1	19437.41	5	14.70	8	0.00
MSSEN-M	2	19370.10	1	30.73	1	0.69
MSS-M	1	19468.96	7	15.61	5	0.00
MVG-M	3	19460.85	6	30.00	2	0.57
MSCN-M	3	19528.23	8	14.94	6	0.00
MGH-M	2	19382.23	3	270.04	10	0.00

Furthermore, VaR, its ranking and p values of the backtest are also reported

the only ones for which the null hypothesis is not rejected. Overall, we also notice a certain degree of variability among the VaR estimates, as often happens in risk management when a battery of models is considered (see, e.g., Eling 2012; Bakar et al. 2015; Tomarchio and Punzo 2020). In particular, the MGH-M provides a very high VaR, emphasizing that it is more extreme in the tails than required for this dataset, thus overestimating the empirical VaR.

Figure 9 reports the pairwise plot of the cryptocurrency dataset, with colors according to the classification estimated by the 2-component MSSEN-M. As we can note from many of the subplots, the two groups have roughly the same center but show different degrees of correlation. This is confirmed by the following estimated mean

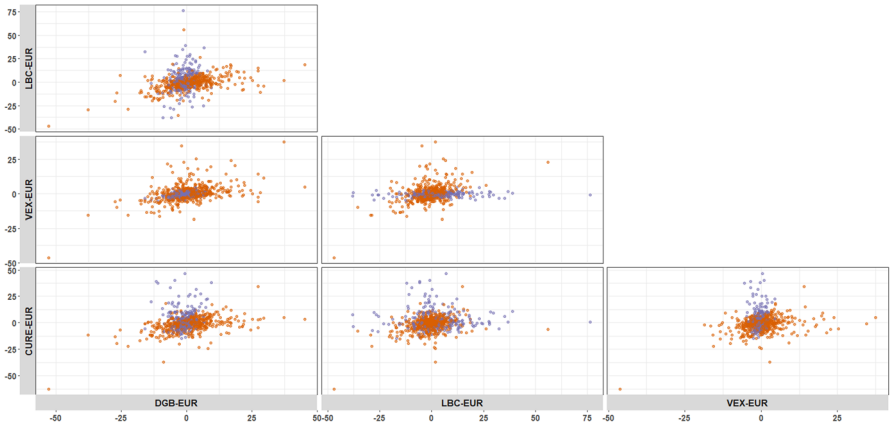


Fig. 9 Pairwise plots (on the lower triangle) for the cryptocurrency dataset. Colors refer to the classification estimated by the best MSSEN-M model. Histograms concerning each variable are also displayed in the main diagonal

vectors and correlation matrices (the latter labeled as $\boldsymbol{R}(\cdot)$) of the two groups, obtained starting from (5) and (6), respectively:

$$\boldsymbol{\mu}_1 = (-0.96, 1.96, -0.55, 2.83)', \quad \boldsymbol{\mu}_2 = (0.88, -0.79, 0.39, -1.35)',$$

$$\boldsymbol{R}(\boldsymbol{\Phi}_1) = \begin{bmatrix} 1.00 & 0.21 & 0.48 & 0.18 \\ 0.21 & 1.00 & 0.14 & 0.08 \\ 0.48 & 0.14 & 1.00 & 0.14 \\ 0.18 & 0.08 & 0.14 & 1.00 \end{bmatrix} \quad \text{and} \quad \boldsymbol{R}(\boldsymbol{\Phi}_2) = \begin{bmatrix} 1.00 & 0.49 & 0.42 & 0.35 \\ 0.49 & 1.00 & 0.34 & 0.30 \\ 0.42 & 0.34 & 1.00 & 0.21 \\ 0.35 & 0.30 & 0.21 & 1.00 \end{bmatrix}.$$

From the analysis of the means, we note how both groups are approximately centered around zero and that, for a given cryptocurrency, the sign between the two groups is opposite. Specifically, it appears that in group 1 the average log returns related to the DGB-EUR and VEX-EUR cryptocurrencies are negative compared to those of group 2. Conversely, the average log returns related to the LBC-EUR and CURE-EUR cryptocurrencies in group 1 are positive compared to those in group 2. Thus, it appears that the two groups identify two possible diversified portfolios, each of which combines (in the opposite way) positive and negative log returns of the considered cryptocurrencies.

By analyzing the correlation matrices, we see that, with the exclusion of one occasion, the correlations in the second group are regularly higher than those in the first group. Thus, the detected classification seems to distinguish between two different types of co-movements between the cryptocurrencies. This information would be lost if a single multivariate distribution would be fitted to the data, as commonly done in the cryptocurrency literature.

The estimated asymmetry parameters for the two groups are:

$$\boldsymbol{\gamma}_1 = (-1.80, 5.92, -0.23, 25.96)', \quad \boldsymbol{\gamma}_2 = (6.60, -3.36, 9.49, -10.55)'.$$

We see that, in both groups, there is a notable deviation from the symmetric case (i.e. $\boldsymbol{\gamma} = \mathbf{0}$, refer to Sect. 2). Similarly to the means, the asymmetric behavior is, for a given cryptocurrency, of opposite sign between the two groups.

Finally, we report the estimated tailedness parameters for the two groups: $\theta_1 = 0.14$ and $\theta_2 = 0.05$. These values are very close to 0, suggesting a leptokurtic behavior for the pdfs of both groups.

4 Conclusions

In this article, we introduced the multivariate skew shifted exponential normal (MSSEN) distribution, an asymmetric generalization of the multivariate shifted exponential normal distribution. Because of its flexibility, we can account for typical non-normal deviations present in the data, such as asymmetry and heavy tails. Features and moments of the MSSEN distribution have been investigated, and some differences with respect to well-established distributions have been evidenced. Furthermore, we

used the MSSEN distribution within the finite mixture model paradigm, thus producing the MSSEN mixture (MSSEN-M) model. An EM algorithm for maximum likelihood parameter estimation of our MSSEN-M model has been illustrated and, differently from many existing alternative models, it has closed-form expressions for all the parameter updates. As concerns the results of the presented simulation study, we showed that the proposed model can perform better than other mixture models, and we have identified possible data structures supporting this result. We have also investigated a real dataset concerning the log returns of four cryptocurrencies. The results show a better fit for our model as well as a more precise risk estimation. Furthermore, the detected classification seems to distinguish between two different types of co-movements between the cryptocurrencies, which would be lost if a single multivariate distribution would be fitted to the data, as commonly done in the cryptocurrency literature.

Acknowledgements Antonio Punzo and Salvatore D. Tomarchio have been partially supported by MUR, Grant number 2022XRHT8R—*The SMILE project: Statistical Modelling and Inference to Live the Environment*. This research was developed within the project supported by Next Generation EU-"GRINS-Growing Resilient, INclusive and Sustainable" Project (PE0000018), National Recovery and Resilience Plan (NRRP)—PE9-Mission 4, C2, Intervention 1.3.

Funding Open access funding provided by Università degli Studi di Catania within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix A: Left truncated generalized inverse Gaussian distribution

Before presenting the left truncated generalized inverse Gaussian distribution (LTGIG), we report details about the generalized inverse Gaussian (GIG) distribution. Specifically, the random variable V has a GIG distribution, in symbols $V \sim \mathcal{GIG}(\lambda, \chi, \psi)$, if its pdf is

$$f_{\text{GIG}}(v; \lambda, \chi, \psi) = \frac{(\psi/\chi)^{\frac{\lambda}{2}} v^{\lambda-1}}{2K_\lambda\left(\sqrt{\psi\chi}\right)} \exp\left\{-\frac{\chi}{2v} - \frac{\psi}{2}v\right\}, \qquad v > 0,$$

where the parameters satisfy $\chi > 0$, $\psi \geq 0$ if $\lambda < 0$; $\chi > 0$, $\psi > 0$ if $\lambda = 0$; and $\chi \geq 0$, $\psi > 0$ if $\lambda > 0$. The GIG distribution was introduced by Good (1953) and has several special cases as, for example, the gamma ($\chi = 0$, $\lambda > 0$) and the inverse Gaussian ($\lambda = -1/2$) distributions.

That said, the random variable W has a LTGIG distribution on the interval (a, ∞), with $a > 0$, in symbols $W \sim \mathcal{LTGIG}_a(\lambda, \chi, \psi)$, if its pdf is

$$f_{\text{LTGIG}}(w; \lambda, \chi, \psi, a) = A^{-1} f_{\text{GIG}}(w; \lambda, \chi, \psi) \mathbb{1}_{(a,\infty)}(w) = \frac{(\psi/\chi)^{\frac{\lambda}{2}} w^{\lambda-1}}{2L_\lambda\left(\sqrt{\chi\psi}, a\right)} \exp\left\{-\frac{\chi}{2w} - \frac{\psi}{2}w\right\}, \qquad a < w < \infty, \tag{A1}$$

where

$$A = \int_a^\infty f_{\text{GIG}}(t; \lambda, \chi, \psi) dt.$$

It is straightforward to prove that

$$\mathrm{E}\left(W^r\right) = \left(\frac{\chi}{\psi}\right)^{\frac{r}{2}} \frac{L_{\lambda+r}\left(\sqrt{\chi\psi}, a\right)}{L_\lambda\left(\sqrt{\chi\psi}, a\right)}, \tag{A2}$$

with $r \in \mathbb{Z}$.

References

Aas K, Hobæk Haff I (2005) Nig and Skew Students' t: Two special cases of the Generalised Hyperbolic distribution. Technical report, Norwegian Computing Center

Abramowitz M, Stegun I (1965) Handbook of mathematical functions: with formulas, graphs, and mathematical tables, applied mathematics series, vol 55. Dover Publications, New York

Agrest MM, Maksimov MZ, Fettis HE et al (1971) Theory of incomplete cylindrical functions and their applications, vol 160. Springer, Berlin

Andrews JL, McNicholas PD, Subedi S (2011) Model-based classification via mixtures of multivariate t-distributions. Comput Stat Data Anal 55(1):520–529

Arellano-Valle RB, Castro LM, Genton MG et al (2008) Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. Bayesian Anal 3(3):513–539

Azzalini A (2013) The skew-normal and related families, vol 3. Cambridge University Press, Cambridge

Bagnato L, Punzo A, Zoia MG (2017) The multivariate leptokurtic-normal distribution and its application in model-based clustering. Canad J Stat 45(1):95–119

Bakar SA, Hamzah NA, Maghsoudi M et al (2015) Modeling loss data using composite models. Insur Math Econ 61:146–154

Basso RM, Lachos VH, Cabral CRB et al (2010) Robust mixture modeling based on scale mixtures of skew-normal distributions. Comput Stat Data Anal 54(12):2926–2941

Browne RP, McNicholas PD (2015) A mixture of generalized hyperbolic distributions. Canad J Stat 43(2):176–198

Cabral CRB, Lachos VH, Prates MO (2012) Multivariate mixture modeling using skew-normal independent distributions. Comput Stat Data Anal 56(1):126–142

Candila V (2021) Multivariate analysis of cryptocurrencies. Econometrics 9(3):28

Chan S, Chu J, Nadarajah S et al (2017) A statistical analysis of cryptocurrencies. J Risk Financ Manag 10(2):12

Chu J, Nadarajah S, Chan S (2015) Statistical analysis of the exchange rate of bitcoin. PloS ONE 10(7):e0133678

Dang UJ, Browne RP, McNicholas PD (2015) Mixtures of multivariate power exponential distributions. Biometrics 71(4):1081–1089

Dávila VHL, Cabral CRB, Zeller CB (2018) Finite mixture of skewed distributions. Springer, Berlin
de Melo V, Mendes B, Fluminense Carneiro A (2020) A comprehensive statistical analysis of the six major crypto-currencies from august 2015 through June 2020. J Risk Financ Manag 13(9):192
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
Eling M (2012) Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? Insuran Math Econ 51(2):239–248
Frühwirth-Schnatter S, Pyne S (2010) Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. Biostatistics 11(2):317–336
Gallaugher MP, 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
Gallaugher MP, Tomarchio SD, McNifcholas PD, et al (2022b) Multivariate cluster weighted models using skewed distributions. Adv Data Anal Classif 1–32
Ghalanos A (2020) rugarch: Univariate GARCH models. R package version 1.4-4
Good IJ (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40(3–4):237–264
Hartigan JA, Wong MA et al (1979) A k-means clustering algorithm. Appl Stat 28(1):100–108
Hu AS, Parlour CA, Rajan U (2019) Cryptocurrencies: stylized facts on a new investible instrument. Financ Manag 48(4):1049–1068
Ji Q, Bouri E, Lau CKM et al (2019) Dynamic connectedness and integration in cryptocurrency markets. Int Rev Financ Anal 63:257–272
Karlis D, Santourian A (2009) Model-based clustering with non-elliptically contoured distributions. Stat Comput 19:73–83
Kupiec PH (1995) Techniques for verifying the accuracy of risk measurement models. J Deriv 3(2):73–84
Lee SX, McLachlan GJ (2013a) Model-based clustering and classification with non-normal mixture distributions. Stat Methods Appl 22(4):427–454
Lee SX, McLachlan GJ (2013b) On mixtures of skew normal and skew-distributions. Adv Data Anal Classif 7(3):241–266
Lee SX, McLachlan GJ (2014) Finite mixtures of multivariate skew t-distributions: some recent and new results. Stat Comput 24:181–202
Lee SX, McLachlan GJ (2016) Finite mixtures of canonical fundamental skew t-distributions. Stat Comput 26(3):573–589
Lee SX, McLachlan GJ (2022) An overview of skew distributions in model-based clustering. J Multivar Anal 188(104):853
Lin TI (2009) Maximum likelihood estimation for multivariate skew normal mixture models. J Multivar Anal 100(2):257–265
Lin TI (2010) Robust mixture modeling using multivariate skew t distributions. Stat Comput 20(3):343–356
Lin TI, Lee JC, Hsieh WJ (2007) Robust mixture modeling using the skew t distribution. Stat Comput 17:81–92
McNeil AJ, Frey R, Embrechts P (2015) Quantitative risk management: concepts, techniques and tools. Princeton University Press, Princeton
McNicholas PD (2016) Mixture model-based classification. Chapman and Hall/CRC, New York
McNicholas PD, Subedi S (2012) Clustering gene expression time course data using mixtures of multivariate t-distributions. J Stat Plan Inference 142(5):1114–1127
McNicholas SM, McNicholas PD, Browne RP (2013) Mixtures of variance-gamma distributions. https://arxiv.org/abs/1309.2695
McNicholas SM, McNicholas PD, Browne RP (2017) A mixture of variance-gamma factor analyzers. In: Big and complex data analysis. Springer, pp 369–385
Misra RD (1940) On the stability of crystal lattices. ii. Math Proc Cambridge Philos Soc 36(2):173–182
Peel D, McLachlan GJ (2000) Robust mixture modelling using the t distribution. Stat Comput 10(4):339–348
Pocuca N, Browne RP, McNicholas PD (2021) mixture: Mixture Models for clustering and classification. https://CRAN.R-project.org/package=mixture, r package version 2.0.3
Prates MO, Cabral CRB, Lachos VH (2013) mixsmsn: fitting finite mixture of scale mixture of skew-normal distributions. J Stat Softw 54(12):1–20
Punzo A, Bagnato L (2020) Allometric analysis using the multivariate shifted exponential normal distribution. Biom J 62(6):1525–1543

Punzo A, Bagnato L (2021) Modeling the cryptocurrency return distribution via Laplace scale mixtures. Phys A 563(125):354
Punzo A, Bagnato L (2022) Asymmetric Laplace scale mixtures for the distribution of cryptocurrency returns. arXiv:2209.12848
Rachev ST, Hoechstoetter M, Fabozzi FJ et al (2010) Probability and statistics for finance, vol 176. Wiley, New York
Shi Y, Tiwari AK, Gozgor G et al (2020) Correlations among cryptocurrencies: evidence from multivariate factor stochastic volatility model. Res Int Bus Financ 53(101):231
Soltyk S, Gupta R (2011) Application of the multivariate skew normal mixture model with the EM algorithm to value-at-risk. MODSIM2011, pp 1638–1644
Sun J, Kabán A, Garibaldi JM (2010) Robust mixture clustering using Pearson type vii distribution. Pattern Recogn Lett 31(16):2447–2454
Szczygielski JJ, Karathanasopoulos A, Zaremba A (2020) One shape fits all? A comprehensive examination of cryptocurrency return distributions. Appl Econ Lett 27(19):1567–1573
Terras R (1981) A miller algorithm for an incomplete Bessel function. J Comput Phys 39(1):233–240
Titterington DM, Afm S, Smith AF et al (1985) Statistical analysis of finite mixture distributions, vol 198. Wiley, New York
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, Bagnato L, Punzo A (2021) SenTinMixt: parsimonious Mixtures of MSEN and MTIN Distributions. https://CRAN.R-project.org/package=SenTinMixt, r package version 1.0.0
Tomarchio SD, Bagnato L, Punzo A (2022) Model-based clustering via new parsimonious mixtures of heavy-tailed distributions. AStA Adv Stat Anal 106(2):315–347
Tong H, Tortora C (2022) Model-based clustering and outlier detection with missing data. Adv Data Anal Classif 16(1):5–30
Vrbik I, McNicholas PD (2014) Parsimonious skew mixture models for model-based clustering and classification. Comput Stat Data Anal 71:196–210
Wang J, Genton MG (2006) The multivariate skew-slash distribution. J Stat Plan Inference 136(1):209–220
Watson GN (1995) A treatise on the theory of Bessel functions. Cambridge University Press, Cambridge
Wei Y, Tang Y, McNicholas PD (2019) Mixtures of generalized hyperbolic distributions and mixtures of skew-t distributions for model-based clustering with incomplete data. Comput Stat Data Anal 130:18–41
Zhang W, Wang P, Li X et al (2018) Some stylized facts of the cryptocurrency market. Appl Econ 50(55):5950–5965

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.