



Antonio Punzo and Luca Bagnato*

Multiple scaled symmetric distributions in allometric studies

<https://doi.org/10.1515/ijb-2020-0059>

Received September 12, 2019; accepted December 7, 2020; published online January 18, 2021

Abstract: In allometric studies, the joint distribution of the log-transformed morphometric variables is typically symmetric and with heavy tails. Moreover, in the bivariate case, it is customary to explain the morphometric variation of these variables by fitting a convenient line, as for example the first principal component (PC). To account for all these peculiarities, we propose the use of multiple scaled symmetric (MSS) distributions. These distributions have the advantage to be directly defined in the PC space, the kind of symmetry involved is less restrictive than the commonly considered elliptical symmetry, the behavior of the tails can vary across PCs, and their first PC is less sensitive to outliers. In the family of MSS distributions, we also propose the multiple scaled shifted exponential normal distribution, equivalent of the multivariate shifted exponential normal distribution in the MSS framework. For the sake of parsimony, we also allow the parameter governing the leptokurtosis on each PC, in the considered MSS distributions, to be tied across PCs. From an inferential point of view, we describe an EM algorithm to estimate the parameters by maximum likelihood, we illustrate how to compute standard errors of the obtained estimates, and we give statistical tests and confidence intervals for the parameters. We use artificial and real allometric data to appreciate the advantages of the MSS distributions over well-known elliptically symmetric distributions and to compare the robustness of the line from our models with respect to the lines fitted by well-established robust and non-robust methods available in the literature.

Keywords: allometry; EM algorithm; heavy-tailed distributions; line-fitting methods; multiple scaled distributions; scale mixtures.

1 Introduction

Allometry can be roughly devised as a tool to study the joint relation between parts (morphometric variables) in various organisms [1, 2], with the measurements being often log-transformed (for the practical and theoretical reasons why it is often useful to transform data to logarithms, see [3–6]). So, the statistical models that are commonly considered are inherently multivariate (almost always bivariate), and mainly aim to:

1. reproduce the joint distribution of the log-transformed morphometric variables;
2. summarize the morphometric variation.

Concerning the first aim, the multivariate normal (MN) distribution with mean μ and covariance matrix Σ played, and still plays, a special rule. However, many empirical studies show that the MN is often not appropriate, and models have been introduced that relax the assumption of multivariate normality to an assumption that the resulting distribution is elliptically symmetric (or elliptically contoured or, simply, elliptical), still maintaining the MN model as a special case [7, 8]. This latter assumption is more biologically realistic, given that the distribution of allometric data is often heavy-tailed [9]. A well-known and simple model satisfying these requirements is the MN scale mixture (also known as MN variance mixture [10]); it is a

*Corresponding author: Luca Bagnato, Dipartimento di Scienze Economiche e Sociali, Università Cattolica del Sacro Cuore, Piacenza, Italy, E-mail: luca.bagnato@unicatt.it. <https://orcid.org/0000-0003-3129-5385>

Antonio Punzo, Dipartimento di Economia e Impresa, Università di Catania, Catania, Italy, E-mail: antonio.punzo@unicat.it. <https://orcid.org/0000-0001-7742-1821>

finite/continuous mixture of MN distributions, with the same mean $\boldsymbol{\mu}$, where $\boldsymbol{\Sigma}$ is scaled by a convenient positive mixing random variable W whose probability mass/density function depends on a parameter vector $\boldsymbol{\theta}$ governing the tails behavior (in terms of leptokurtosis) of the unconditional mixture.

As for the second aim, which is more debated in allometry [11], most multivariate studies have used principal component analysis (PCA); in this direction, in the important work by [12] it is suggested to summarize the morphometric variation by the first principal component (PC) of the log-transformed measurements. In the special bivariate case, the problem simplifies in fitting a line to the scatter plot of the log-transformed measurements and making inference about the fitted line, typically testing if the slope equals a given value and constructing confidence intervals for slope and elevation (intercept). Commonly used line-fitting methods in allometry are linear regression, major axis (which is equivalent to the first PC), and standardised major axis. These are least squares methods differing in the direction in which errors from the line are measured [6, 11]. Advantageously, however, major axis methods have a natural extension to more than two dimensions ([6], page 278).

The present paper has four merits. Firstly, concerning the first aim above, by using allometric data we show the existence of situations where, although the joint distribution of the log-transformed morphometric variables can be considered as symmetric and with heavy tails, the assumption of elliptical contours may be rather restrictive, with heavy-tailed distributions accounting for alternative symmetric shapes providing a better fit. To describe these shapes, we consider the family of multiple scaled symmetric (MSS) distributions proposed by Forbes and Wraith [13]. MSS distributions can be roughly considered as an extension of the MN scale mixture based on two key elements: the decomposition of the scale matrix $\boldsymbol{\Sigma}$ by eigenvalues and eigenvectors matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$, and the use of the mixing random variable W separately for each dimension (PC) of the space spanned by the columns of $\boldsymbol{\Gamma}$. The result is a multivariate heavy-tailed distribution in which the parameter vector $\boldsymbol{\theta}$ is allowed to vary across PCs and where the contours are not constrained to be elliptical, although the symmetry is preserved. In allometric terms, these distributions have the advantage to be directly defined in the space of interest, namely the PC space, with each PC being distributed according to a (univariate) normal scale mixture. Moreover, regardless of the mixing distribution considered, and in analogy with the MN scale mixtures, MSS distributions enable (among other) natural/efficient computation of the maximum likelihood (ML) estimates via the expectation-maximization (EM) algorithm and robust ML estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ based on down-weighting of mild outliers in the PC space. As a second merit of the paper, we propose a new member of the family of MSS distributions obtained by choosing a convenient shifted exponential as mixing distribution, so extending the recently proposed multivariate shifted exponential normal distribution [14] to the MSS framework. As a third merit, we introduce parsimony in the MSS distributions by constraining $\boldsymbol{\theta}$ to be tied across PCs, resulting in distributions for the PCs having the same degree of leptokurtosis. In connection with the second aim discussed above, we know that line-fitting methods are quite sensitive to the presence of outliers, in the sense that the fitted line is directed towards them. As a limit case this implies that in certain situations, based on the position of the outlier(s), the resulting fitted line may be flipped into a completely different direction. As a final merit of the paper, via a simulation study and a sensitivity analysis based on allometric bivariate data, we show that the fitted line from MSS distributions, which is a major axis approach, is: 1) more robust to outliers when compared to the classical line-fitting methods discussed above, and 2) comparable with robust versions of major axis and standardised major axis approaches based on Huber's M -estimator in place of least squares [15].

The paper is organized as follows. In Section 2 we first recall the MN scale mixtures (Section 2.1) and MSS distributions (Section 2.2), and then we introduce the multiple scaled shifted exponential normal distribution and the parsimonious variants of the MSS distributions (Section 2.3). In Section 3 we focus on the bivariate case by introducing what can be defined as major axis under MSS distributions. To simplify the reading of the paper, we postpone in Appendix A inferential aspects such as ML parameter estimation via the use of the EM algorithm, computation of standard errors for the estimates, formulation of confidence intervals and statistical tests. In Section 4 we describe the results of a simulation study aiming to compare, with and without outliers, the performance of the major axes from our MSS models with the lines fitted by well-established (robust) methods in allometry. In Section 5 we analyze two allometric data sets to demonstrate the better performance of the MSS models over heavy-tailed elliptical distributions which are well-known in the literature and to

investigate the robustness of our MSS-based line-fitting methods. We conclude the paper with a brief discussion, as well as with avenues for further research, in Section 6.

2 Methodology

2.1 Multivariate normal scale mixture

For robustness sake, one of the most common ways to generalize the MN distribution is represented by the MN scale mixture (MNSM) with probability density function (pdf)

$$f_{\text{MNSM}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\theta}) = \int_0^{\infty} f_{\text{MN}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}/w) m(w; \boldsymbol{\theta}) dw, \quad (1)$$

where \mathbf{X} is the d -variate vector of log-transformed morphometric variables, $f_{\text{MN}}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the pdf of the MN distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, in symbols $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $m(w; \boldsymbol{\theta})$ is the mixing probability density (or mass) function of W , with a positive support, depending on the parameter vector $\boldsymbol{\theta}$, in symbols $W \sim \mathcal{M}(\boldsymbol{\theta})$. If \mathbf{X} has the pdf in (1), we write $\mathbf{X} \sim \mathcal{NSM}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\theta})$. For the MNSM model, $\boldsymbol{\Sigma}$ is a scale matrix.

The MNSM in (1) simply alters the tail behavior of the conditional (reference) MN distribution, allowing for leptokurtosis, while leaving the resultant distribution unimodal and elliptically symmetric (for further details about the properties of the MNSM family of distributions see, e.g., [16–20]). Moreover, as detailed in [21], the covariance matrix and kurtosis of $\mathbf{X} \sim \mathcal{NSM}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\theta})$ are respectively.

$$\text{Var}(\mathbf{X}) = v(\boldsymbol{\theta}) \boldsymbol{\Sigma} \quad (2)$$

$$\text{Kurt}(\mathbf{X}) = k(\boldsymbol{\theta})d(d+2), \quad (3)$$

where $v(\boldsymbol{\theta}) = E(V)$ is the variance factor, $k(\boldsymbol{\theta}) = E(V^2)/[E(V)]^2$ is the kurtosis factor, and $V = 1/W$. As we can note from (3), the kurtosis is only governed by $\boldsymbol{\theta}$. Moreover, because $E(V^2) \geq [E(V)]^2$, the excess kurtosis (with respect to the MN distribution) is non-negative, with the equality holding only in the degenerate case $W \equiv 1$. Therefore, apart from this limit case, the resulting distribution is leptokurtic. So, $\boldsymbol{\theta}$ can be meant as the tailedness parameter of the MNSM model.

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

$$\begin{aligned} W &\sim \mathcal{M}(\boldsymbol{\theta}) \\ \mathbf{X}|W = w &\sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}/w). \end{aligned}$$

This alternative way to see the MNSM distribution is useful for random generation and for the implementation of the expectation-maximization (EM) algorithm.

Two members of the MNSM family of distributions, those will be mainly considered in this paper, are the multivariate t (Mt) [22, 23] and the multivariate shifted exponential normal (MSEN) [14]. The Mt is obtained when W has gamma distribution with shape and rate being both equal to $\theta/2$, $\theta > 0$, that is when W has pdf

$$m(w; \theta) = \frac{(\frac{\theta}{2})^{\frac{\theta}{2}}}{\Gamma(\frac{\theta}{2})} w^{\frac{\theta}{2}-1} \exp\left(-\frac{\theta}{2}w\right), \quad w > 0. \quad (4)$$

If W has the pdf in (4), for notational convenience we write $W \sim \mathcal{G}(\theta/2, \theta/2)$. The MSEN is obtained when W has shifted exponential distribution, with rate $\theta > 0$ and support $(1, \infty)$; in symbols, $W \sim \mathcal{SE}(\theta)$. The pdf of $W \sim \mathcal{SE}(\theta)$ is

$$m(w; \theta) = \theta \exp[-\theta(w-1)], \quad w > 1.$$

Further examples of distributions belonging to the MNSM family are: the multivariate contaminated normal (MCN) [24], the multivariate tail-inflated normal (MTIN) [21] and the multivariate symmetric generalized hyperbolic (MSGH). The MSGH distribution, in turn, nests many other well known MNSM distributions such as the

multivariate symmetric hyperbolic (MSH), the multivariate symmetric variance-gamma (MSVG), and the multivariate symmetric normal inverse Gaussian (SNIG); for details about the MSGH distribution, along with its special cases, see [20], Example 3.8.

2.2 Multiple scaled symmetric distributions

Forbes and Wraith [13] propose an alternative way to alter the tail behavior of the conditional MN distribution in (1), while leaving the resultant distribution unimodal and symmetric. The idea roughly consists in:

1. using the classical eigen-decomposition

$$\Sigma = \Gamma \Lambda \Gamma^T, \quad (5)$$

where Γ is a $d \times d$ matrix of eigenvectors of Σ and Λ is the corresponding $d \times d$ diagonal matrix of eigenvalues;

2. generalizing the univariate mixing variable W in (1) to a d -variate mixing variable $\mathbf{W} = (W_1, \dots, W_d)^T$ with independent variates and joint probability density (or mass) function

$$m(\mathbf{w}; \boldsymbol{\theta}) = \prod_{h=1}^d m(w_h; \boldsymbol{\theta}_h),$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_d^T)^T$. For notational convenience we write $\mathbf{W} \sim \mathcal{M}_d(\boldsymbol{\theta})$.

Note that, each element on the right-hand side of (5) has a different geometric interpretation: Λ determines the size and shape of the scatter, while Γ determines its orientation. The resulting multiple scaled symmetric (MSS) distributions have pdf

$$f_{\text{MSS}}(\mathbf{x}; \boldsymbol{\mu}, \Gamma, \Lambda, \boldsymbol{\theta}) = \int_0^\infty \cdots \int_0^\infty f_{\text{MN}}(\mathbf{x}; \boldsymbol{\mu}, \Gamma \Delta_w \Lambda \Gamma^T) m(w_1, \dots, w_d; \boldsymbol{\theta}) dw_1 \cdots dw_d, \quad (6)$$

where $\Delta_w = \text{diag}(w_1^{-1}, \dots, w_d^{-1})$. If \mathbf{X} has the distribution in (6) we write $\mathbf{X} \sim \mathcal{MSS}_d(\boldsymbol{\mu}, \Gamma, \Lambda, \boldsymbol{\theta})$.

The MSS distributions have a convenient hierarchical representation too. In detail, $\mathbf{X} \sim \mathcal{MSS}_d(\boldsymbol{\mu}, \Gamma, \Lambda, \boldsymbol{\theta})$ can be represented as

$$\begin{aligned} \mathbf{W} &\sim \mathcal{M}_d(\boldsymbol{\theta}) \\ \mathbf{X} | \mathbf{W} = \mathbf{w} &\sim \mathcal{N}_d(\boldsymbol{\mu}, \Gamma \Delta_w \Lambda \Gamma^T). \end{aligned}$$

For the sake of interpretation, the pdf in (6) can be equivalently written as

$$\begin{aligned} f_{\text{MSS}}(\mathbf{x}; \boldsymbol{\mu}, \Gamma, \Lambda, \boldsymbol{\theta}) &= \prod_{h=1}^d \int_0^\infty f_{\text{N}}([\Gamma^T(\mathbf{x} - \boldsymbol{\mu})]_h; 0, \lambda_h/w_h) m(w_h; \boldsymbol{\theta}_h) dw_h \\ &= \prod_{h=1}^d f_{\text{NSM}}([\Gamma^T(\mathbf{x} - \boldsymbol{\mu})]_h; 0, \lambda_h, \boldsymbol{\theta}_h), \end{aligned} \quad (7)$$

where $\Gamma^T(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{Y}$ is the principal components (PCs) transform of \mathbf{x} and can be thought of as a rotation and a recentering of \mathbf{x} , $[\Gamma^T(\mathbf{x} - \boldsymbol{\mu})]_h = Y_h$ is the h th element of \mathbf{Y} , and λ_h is the h th diagonal element of Λ (or, equivalently, the h th eigenvalue of Σ). Note that, in \mathbf{Y} , the variables Y_1, \dots, Y_d are ordered so that $\text{Var}(Y_1) \geq \text{Var}(Y_2) \geq \dots \geq \text{Var}(Y_d)$, in analogy with the classical PCA. In (7), the subscripts “N” and “NSM” are used instead of “MN” and “MNSM”, respectively, for the univariate case. It is straightforward to realize that $Y_h \sim \mathcal{NSM}_1(0, \lambda_h, \boldsymbol{\theta}_h)$. Finally note that, apart from the univariate case $d = 1$, the MSS distribution is not a special or limiting case of the MNSM distribution.

Concerning the estimation of the parameters $\boldsymbol{\mu}$, Γ , Λ , and $\boldsymbol{\theta}$, several estimators may be considered. Among them, the maximum likelihood (ML) estimator is the most attractive because of its asymptotic properties (see, e.g., [25], p. 206). Details about the expectation-maximization (EM) algorithm, to obtain ML estimates of the parameters for MSS distributions, are given in Appendix A.1.

2.3 Special cases of MSS distributions and parsimony

In the MSS family of distributions, Forbes and Wraith [13] introduce the multiple scaled t (MSt) distribution. It is obtained by setting $W_h \sim \mathcal{G}(\theta_h/2, \theta_h/2)$. For the MSt, the variance and kurtosis factors for each PC are respectively given by

$$v(\theta_h) = \frac{\theta_h}{\theta_h - 2} \quad \text{and} \quad k(\theta_h) = \frac{\theta_h - 2}{\theta_h - 4}, \quad (8)$$

being in general $\text{Var}(Y_h) = \lambda_h v(\theta_h)$ and $\text{Kurt}(Y_h) = 3k(\theta_h)$, $h = 1, \dots, d$; compare with Eqs. (2) and (3).

In this paper we introduce a new member of the MSS family, the multiple scaled SEN (MSEEN) distribution, obtained by assuming $W_h \sim \mathcal{SE}(\theta_h)$. For the MSEEN, the variance and kurtosis factors for each PC are respectively given by

$$v(\theta_h) = \theta_h \exp(\theta_h) \varphi_{-1}(\theta_h) \quad \text{and} \quad k(\theta_h) = \frac{\theta_h}{v(\theta_h)^2} [1 - v(\theta_h)], \quad (9)$$

where

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

is the Misra function [26], further generalization of the generalized exponential integral function $E_n(z) = \int_1^{\infty} t^{-n} \exp(-zt) dt$ [27] with $\Gamma(n, z) = \int_z^{\infty} t^{n-1} \exp(-t) dt$ denoting the upper incomplete gamma function; for further details, see [14]. The MSt and MSEEN are the two special cases of MSS distributions we will consider in this paper.

To evaluate the data configurations favoring the MSt over the MSEEN, and *vice versa*, it is enough to consider univariate t and SEN distributions. Their main difference relies on the joint behavior of the variance and kurtosis factors $v(\theta_h)$ and $k(\theta_h)$, $h = 1, \dots, d$; refer to Figure 1. In particular, λ_h kept fixed, for the t distribution larger values of kurtosis are related to larger values of variance (cf. Figure 1(a)) while, for the SEN distribution, larger values of kurtosis are related to smaller values of variance (cf. Figure 1(b)); see [28].

In models (6) and (7) there is a θ_h -parameter for each PC. This means that, as the dimension d grows, the total number of free parameters to be estimated can quickly become large leading to overfitting. To introduce parsimony, we allow the d tailedness parameters to be tied across PCs, i.e. we assume $\theta_1 = \dots = \theta_d$. This results in normal scale mixtures, for the PCs, having the same degree of leptokurtosis. When this constraint is applied to MSt and MSEEN distributions, we will simply write C-MSt and C-MSEEN, respectively.

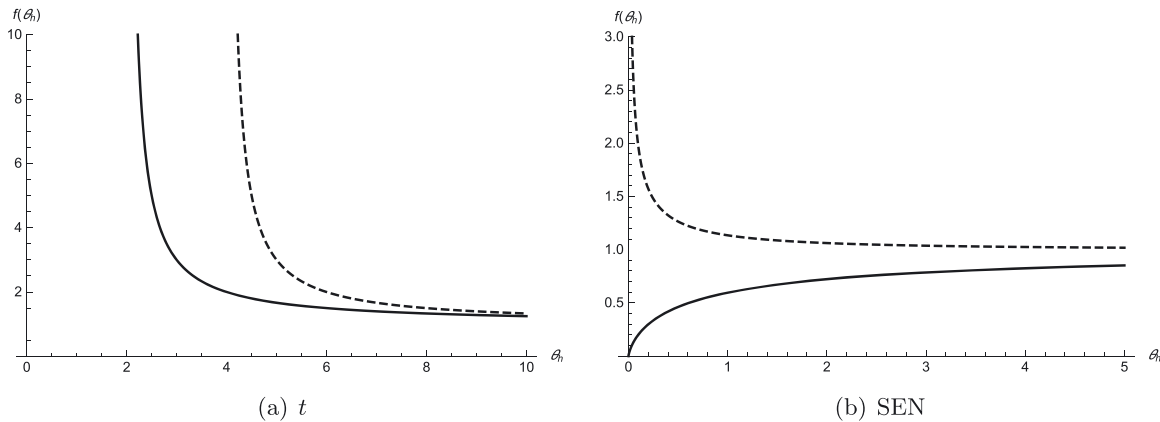


Figure 1: Multiplicative factors $v(\theta_h)$ (solid line) and $k(\theta_h)$ (dashed line).

3 Line-fitting from bivariate MSS distributions

As emphasized in Section 1, allometry mainly focuses on the bivariate case ($d = 2$), where $\mathbf{X} = (X_1, X_2)^\top$. In this case, we would like to make inference about the elevation (intercept) $\alpha \in \mathbb{R}$ and the slope $\beta \in \mathbb{R}$ of the line describing the relationship between X_1 and X_2 . Operationally, we are often interested in testing if β equals a particular value, and estimating confidence intervals for α and β . Therefore, it is important: 1) to evaluate how the MSS distributions specialize in the bivariate case, and 2) parametrize their pdf in terms of α and β .

First of all, it is important to underline that the line determined under MSS distributions belongs to the class of the major axis (or first PC axis) methods; as such, the line we are interested in is the axis of Y_1 .

There are different ways to parametrize the MSS distributions, in the bivariate case, in terms of α and β . We decided to express Γ and μ_2 (mean of X_2) in terms of α and β in the following way

$$\Gamma = \begin{pmatrix} \frac{1}{\sqrt{1+\beta^2}} & -\frac{\beta}{\sqrt{1+\beta^2}} \\ \frac{\beta}{\sqrt{1+\beta^2}} & \frac{1}{\sqrt{1+\beta^2}} \end{pmatrix} \quad \text{and} \quad \mu_2 = \alpha + \beta\mu_1,$$

where μ_1 is the mean of X_1 . According to this choice, the pdf of the MS t distribution becomes

$$f_{\text{MS}t}(x_1, x_2; \Psi) = \frac{\Gamma\left(\frac{\theta_1+1}{2}\right)\Gamma\left(\frac{\theta_2+1}{2}\right)}{\pi\Gamma\left(\frac{\theta_1}{2}\right)\Gamma\left(\frac{\theta_2}{2}\right)\sqrt{\lambda_1\lambda_2\theta_1\theta_2}} \left(1 + \frac{[x_1 + \beta(x_2 - \alpha) - \mu_1(1 + \beta^2)]^2}{\lambda_1\theta_1(1 + \beta^2)}\right)^{-\frac{\theta_1+1}{2}} \cdot \left(1 + \frac{(x_2 - \beta x_1 - \alpha)^2}{\lambda_2\theta_2(1 + \beta^2)}\right)^{-\frac{\theta_2+1}{2}}, \quad (10)$$

while the pdf of the MSSEN distribution becomes

$$f_{\text{MSSEN}}(x_1, x_2; \Psi) = \frac{\theta_1\theta_2 \exp(\theta_1 + \theta_2)}{2\pi\sqrt{\lambda_1 + \lambda_2}} \varphi_{\frac{1}{2}}\left(\frac{[x_1 + \beta(x_2 - \alpha) - \mu_1(1 + \beta^2)]^2}{2\lambda_1(1 + \beta^2)} + \theta_1\right) \cdot \varphi_{\frac{1}{2}}\left(\frac{(x_2 - \beta x_1 - \alpha)^2}{2\lambda_2(1 + \beta^2)} + \theta_2\right), \quad (11)$$

where $\Psi = (\alpha, \beta, \mu_1, \lambda_1, \lambda_2, \theta_1, \theta_2)'$.

Figure 2 shows, via isodensities, some possible shapes of the bivariate C-MS t and C-MSSEN distributions by fixing α , β , and μ_1 (it is equivalent to fix the mean vector $\boldsymbol{\mu}$ and orientation matrix Γ), and by varying λ_1 , λ_2 , and the common θ so to produce PCs with given variances and kurtoses; refer to Section 2.3. In detail, in all the plots, $\mu_1 = 0$, $\alpha = 0$, and $\beta = 1$; note that in this case $\mu_2 = 0$. Variances and kurtoses of the two PCs, along with the values of λ_1 , λ_2 , and θ , are given in Table 1. Figure 2 clearly shows how the shape of the MSS distributions is not constrained to be elliptical, although the symmetry is preserved. In particular, the choices made for θ (refer to Table 1) produce “smoothed” rhomboidal (Figure 2(a) and (b)) and starred (Figure 2(c) and (d)) contours.

4 Simulation study: comparison of line-fitting methods

In this section, we investigate the behaviour of the fitted lines from our four bivariate MSS models – MSSEN, MS t , C-MSSEN, and C-MS t – through a large-scale simulation study performed using R [29]. We further provide a comparison with the following classical line-fitting methods: ordinary least-squares (OLS) regression, OLS bisector, major axis (MA; also known as orthogonal regression), standardized major axis (SMA; also known as

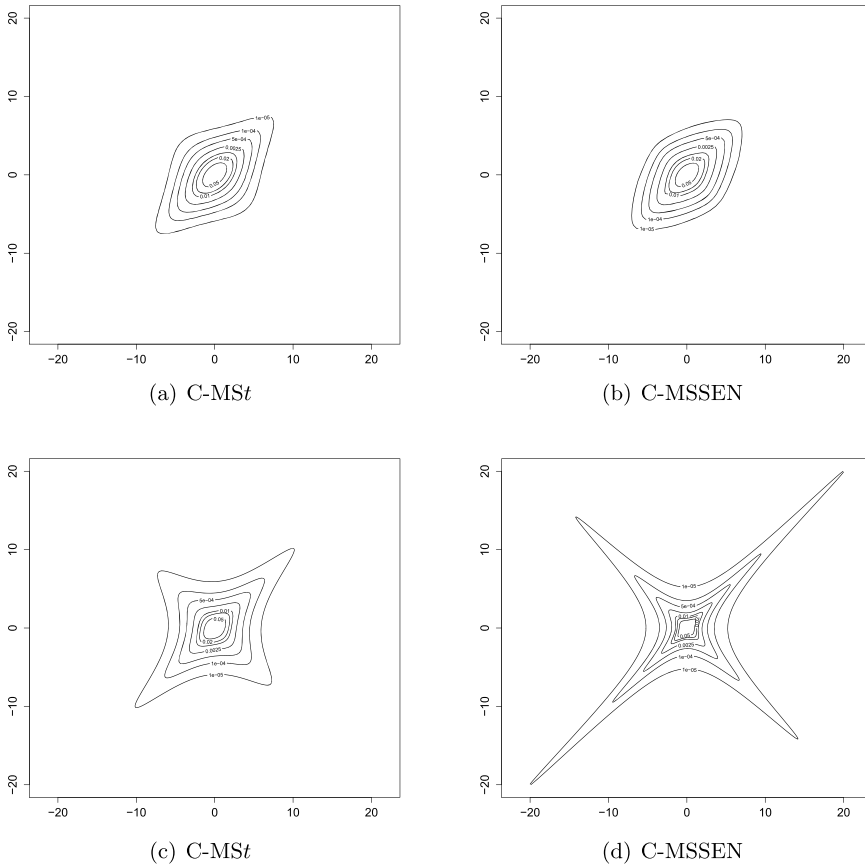


Figure 2: Examples of contour plots of bivariate MSS distributions with common θ where $\mu_1 = 0$, $\alpha = 0$, and $\beta = 1$. The values of λ_h , $h = 1, 2$, and θ are given in Table 1.

Table 1: Variances and kurtoses of PCs, along with the values of λ_1 , λ_2 , and θ , for the contour plots of Figure 2.

(a) Figure 2(a)				
	Variance	λ	Kurtosis	θ
Y_1	3	2.4	4	10
Y_2	1	0.8	4	10
(b) Figure 2(b)				
	Variance	λ	Kurtosis	θ
Y_1	3	7.2790	4	0.3853
Y_2	1	2.4263	4	0.3853
(c) Figure 2(c)				
	Variance	λ	Kurtosis	θ
Y_1	2	1.0037	400	4.0151
Y_2	1	0.5019	400	4.0151
(d) Figure 2(d)				
	Variance	λ	Kurtosis	θ
Y_1	2	2322.1544	400	0.0001
Y_2	1	1151.3737	400	0.0001

reduced major-axis), robust MA (R-MA), and robust SMA (R-SMA), the last two approaches being based on Huber's M-estimator; see [6, 11, 15] for details. Emphasis is given to the evaluation of the impact of outliers on the fitted slope of the competing methods.

To generate the data we consider a bivariate normal $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We fix $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top$, where

$$\boldsymbol{\Gamma} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Lambda} = \begin{pmatrix} \frac{1}{\lambda} & 0 \\ 0 & \lambda \end{pmatrix},$$

with $\lambda \in (0, 1)$ governing the axis-ratio of the elliptical scatter: as λ approaches to 1, the ellipse becomes more and more similar to a circle while, on the other side, as λ approaches to 0, the ellipse degenerates on its major axis. With the notation of Section 3, this would be equivalent to say that we generate data from a bivariate normal distribution whose elevation, slope, and mean of X_1 are $\alpha = 0$, $\beta = 1$, and $\mu_1 = 0$, respectively. If we want outliers to be included in the sample, we generate them from a bivariate normal $\mathcal{N}_2(\boldsymbol{\mu}_{\text{out}}, \boldsymbol{\Sigma}_{\text{out}})$, where $\boldsymbol{\mu}_{\text{out}} = q \cdot (1, -1)^\top$ and $\boldsymbol{\Sigma}_{\text{out}} = 0.01 \cdot \boldsymbol{\Sigma}$, with q being a quantity governing the distance of the outliers from the bulk of the data (represented by $\boldsymbol{\mu} = \mathbf{0}$).

We consider four experimental factors: the sample size ($n \in \{100, 200\}$), the proportion of simulated outliers ($p \in \{0, 0.05, 0.10\}$), $\lambda \in \{0.2, 0.3\}$, and $q \in \{8, 10\}$. In Figure 3 we show an example of generated dataset, with outliers, in the case $n = 200$, $p = 0.05$, $\lambda = 0.3$, and $q = 8$. The contamination scheme is not chosen randomly, but it is similar to a classical contamination scheme used to evaluate robustness for MA methods (see, e.g., [30–33]). We generate 100 data sets for each combination of (n, λ) when $p = 0$, and for each combination of (n, p, λ, q) when $p > 0$; the difference is due to the fact that, when $p = 0$, there is no reason to consider the experimental factor q . This yields a total of 20 scenarios ($2^2 = 4$ in the case $p = 0$, and $2^4 = 16$ otherwise) and 2000 generated data sets ($100 \times 2^2 = 400$ in the case $p = 0$, and $100 \times 2^4 = 1600$ for $p > 0$). For each generated data set, we fit the 10 competing models/lines discussed above. The MA, SMA, R-MA and R-SMA methods are estimated using the **smatr** package for R [34].

For comparison's sake, for each scenario and each line-fitting method, we display the box-plots of the estimated slopes over the 100 replications. While in the 4 scenarios without outliers we expect all the methods working well and comparably, in the 16 scenarios with outliers we hope the estimated slope is not flipped into the opposite direction.

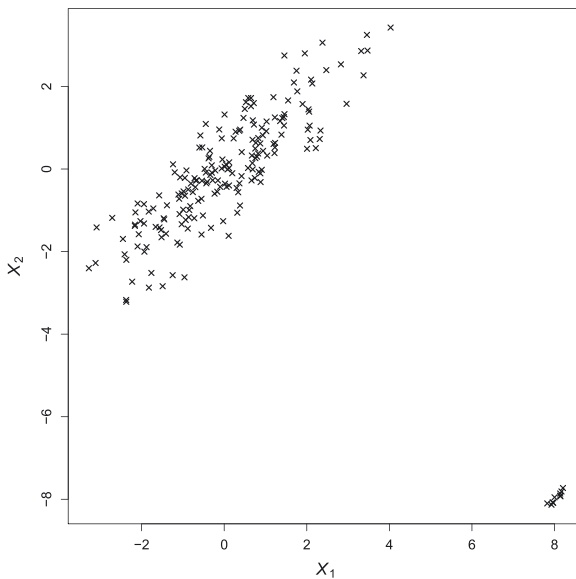


Figure 3: Example of simulated data, with outliers, in the case $n = 200$, $p = 0.05$, $\lambda = 0.3$, and $q = 8$.

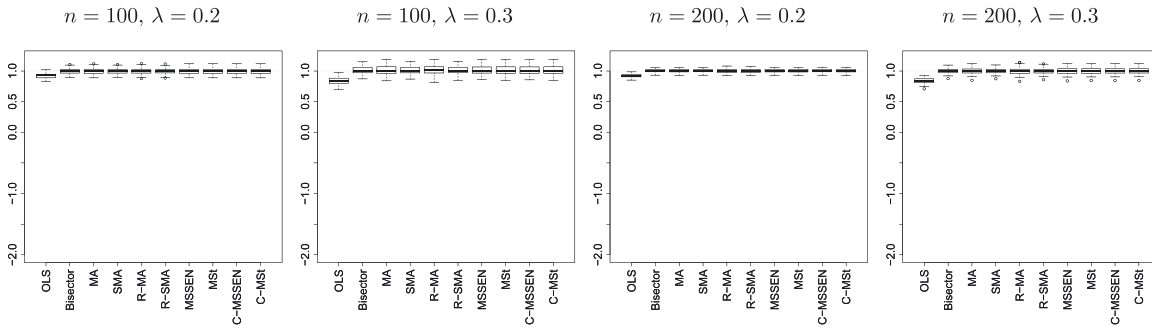


Figure 4: Box-plots, over 100 replications, of the estimated slopes in the scenarios without outliers ($p = 0$). Each plot refers to a different pair (n, λ) .

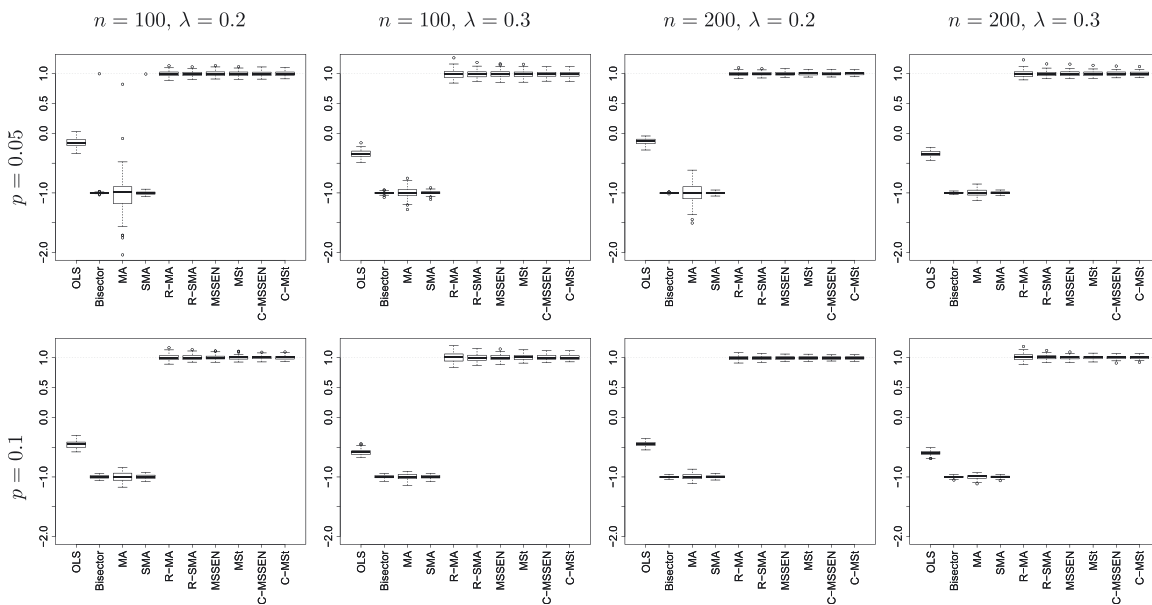


Figure 5: Box-plots, over 100 replications, of the estimated slopes in the scenarios with $q = 8$. Each plot refers to a different triplet (n, λ, p) .

Figure 4 reports the obtained results for the scenarios without outliers. A horizontal dashed gray line is placed at $\beta = 1$ to have, as a benchmark, the true but unknown value of β in the data generating process (DGP). Note that, we use a larger than expected range for the y-axis simply to have the same range for all the box-plots of this simulation study (compare with Figures 4 and 5); this should make easier the comparison of results referred to different scenarios. Apart from the OLS regression, the competing methods perform comparably, with a variability of the estimates that decreases either as n grows (from 100 to 200) or as λ decreases (from 0.3 to 0.2). The different behavior of the OLS estimates is due to the fact that the DGP violates the zero conditional mean assumption of the OLS regression (see, e.g., [35], Chapter 2).

In Figures 5 and 6 we report the obtained results for the scenarios with $q = 8$ and $q = 10$, respectively. Regardless of the considered scenario, the estimated slopes from our MSS distributions, as those from R-MA and R-SMA, are robust to the presence of outliers, with a behavior very similar to the one analyzed in the scenarios without contamination (compare with Figure 4); similarity is also in terms of variability of the estimates which is decreasing when either n increases or λ decreases. The only difference between the

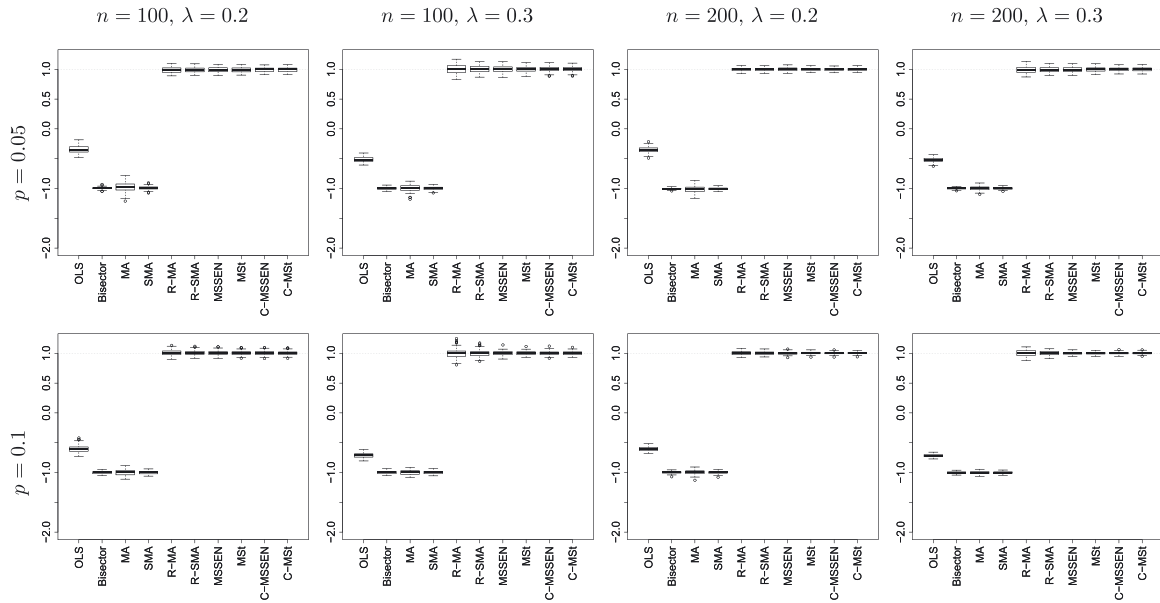


Figure 6: Box-plots, over 100 replications, of the estimated slopes in the scenarios with $q = 10$. Each plot refers to a different triplet (n, λ, p) .

estimated slopes from our MSS distributions, and those from R-MA and R-SMA, is a slight larger variability for the estimates from the latter methods, especially when λ grows. Interestingly enough, the proportion of outliers p does not seem to have an impact on the estimates behavior. Instead, the bisector, MA, and SMA methods of line-fitting are strongly affected by outliers, with a median estimated β (over the 100 replications) being always close to -1 , the opposite of the true β . The OLS regression provides box-plots that are mainly located on values of β between 0 and -1 . Although the OLS-estimates are also in this case very far from the true value $\beta = 1$, they are slightly better than those from bisector, MA, and SMA methods.

5 Real data analyses

In this section we illustrate two applications of our models on real allometric data. The first one (Section 5.1) considers bivariate data; one of the aims is to evaluate how a single outlier affects the behavior of the competing line-fitting methods already considered in the simulation study of Section 4. The second analysis (Section 5.2) aims to fill one of the gaps in Section 4, showing an application of our models on data with more than two measurements.

According to the considerations of Section 1, both the analyses fill another gap with the simulation study of Section 4, i.e. the evaluation of the ability of our models to reproduce the joint distribution of the log-transformed morphometric variables. We also compare our models with the following multivariate elliptical (heavy-tailed) distributions: Mt , $MTIN$, $MSEN$, MLN , MCN , $MSVG$, MSH , $SNIG$, and $MSGH$ given in Section 2.1. With MLN we denote the multivariate leptokurtic normal distribution introduced by [36].

In analogy with Section 4, the analysis is entirely conducted in R. Parameters are estimated based on the ML approach. We adopt the EM algorithm, illustrated in Appendix A.1, for MSS distributions. For them, we start the algorithm from the M-step, providing initial values for \hat{w}_{ih} and $\hat{\Gamma}$ (refer to Appendix A.1.2). In detail, we fix $\hat{\Gamma}$ as the eigenvector matrix of the sample covariance matrix, so that $\hat{\Gamma}$ does not depend on the considered distribution, while we differentiate \hat{w}_{ih} according to the considered model: we fix $\hat{w}_{ih} = 1$ for MSt and $C-MSt$, and $\hat{w}_{ih} = 2$ for $MSEn$ and $C-MSEn$, $i = 1, \dots, n$ and $h = 1, \dots, d$. The code for density evaluation, random number generation, and fitting for the considered MSS distributions is

available at <http://docenti.unict.it/punzo/Rcode.htm>. As concerns the competing models, we use the `WML.MLN()` function, of the code available at <http://docenti.unict.it/punzo/Rcode.htm>, to fit the MLN, the `mtin.ML.ECME()` function of the **mtin** package¹ to fit the MTIN, the `msen.ML.EM()` function of the **msen** package¹ to fit the MSEN, the `CNmixt()` function of the **ContaminatedMixt** package [37] to fit the MCN, and the **ghyp** package [38] to fit the remaining models.

Having the competing models a differing number of parameters, their goodness-of-fit comparison is made, as usual, via the Akaike information criterion (AIC) [39] and the Bayesian information criterion (BIC) [40] that, in our formulation, need to be maximized. Moreover, we use likelihood-ratio (LR) tests not only to make inference on the parameters (refer to Appendix A.3) but also to: 1) compare the MN distribution (null model) with each of the competing distributions (alternative models), and 2) for testing MSS-parsimony (refer to the last part of Section 2.3), that is to compare the constrained (null) and unconstrained (alternative) MSS models based on the same mixing random variable. In the sequel, the obtained p -value will be always compared to the classical 5% significance level.

5.1 Anthropometric measurements of male twins

The first analysis considers the `m.twins` data set accompanying the **Flury** package [41] for R. The data set contains six anthropometric measurements for $n = 89$ pairs of monozygotic/dizygotic male twins measured in the 1950s at the University of Hamburg, Germany. For details on these data, see [42]. We focus on the logarithm of $d = 2$ of the available measurements: the chest circumference of first and second twin. The variables are expressed in centimeters. The scatter plot of the log-transformed data is displayed in Figure 7.

The first aim of this analysis is to evaluate the best model for the bivariate distribution of the data. The scatter in Figure 7 seems to be symmetric enough and this is corroborated by the Mardia test of symmetry (p -value = 0.444), as implemented by the `mvn()` function of the **MVN** package [43]. Table 2 presents the model comparison. AIC and BIC provide the same ranking for the best four models. They are all MSS distributions, with the MSEN being the overall best, followed by the other unconstrained model (MS_t). The top-four models have the same ranking induced by the p -values of the LR test of bivariate normality. In particular, the null

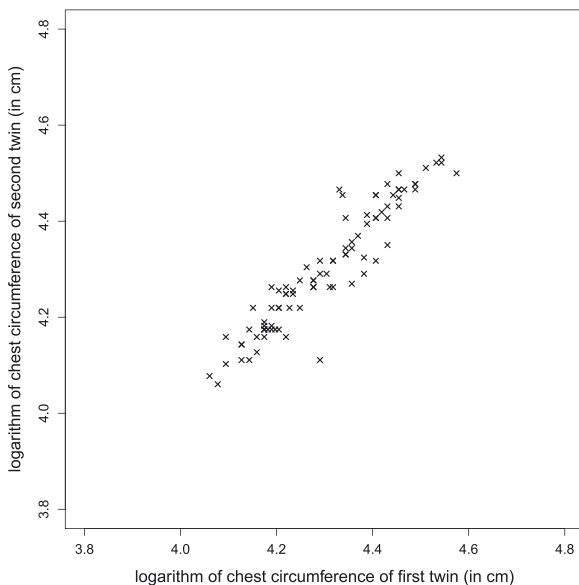


Figure 7: Scatter plot of the of the log-transformed `m.twins` data.

¹ Available at <http://docenti.unict.it/punzo/Rpackages.htm>.

Table 2: Model comparison, for the *m. twins* data, in terms of number of parameters (#par), log-likelihood, AIC, BIC, and *p*-values from the LR tests of multivariate normality and MSS-parsimony. Rankings induced by the considered criteria are also displayed.

Model	# par.	Log-Lik.	AIC	Ranking	BIC	Ranking	Normality LR <i>p</i> -value	Ranking	MSS-parsimony LR <i>p</i> -value
MN	5	213.822	417.644	12	405.201	5	–	–	–
Mt	6	215.653	419.306	6	404.374	6	0.176	5	–
MTIN	6	215.594	419.188	7	404.256	7	0.183	6	–
MSEN	6	214.711	417.421	13	402.489	12	0.346	12	–
MLN	6	214.978	417.957	11	403.025	11	0.282	11	–
MCN	7	216.943	419.885	5	402.465	13	0.210	7	–
MSVG	6	215.093	418.186	10	403.254	10	0.260	10	–
MSH	6	215.264	418.529	9	403.597	9	0.230	9	–
SNIG	6	215.360	418.721	8	403.789	8	0.215	8	–
MSGH	7	215.653	417.306	14	399.886	14	0.400	13	–
MSEEN	7	225.347	436.693	1	419.273	1	0.003	1	0.003
MSt	7	224.689	435.377	2	417.957	2	0.004	2	0.007
C-MSEEN	6	216.552	421.103	4	406.172	4	0.099	4	–
C-MSt	6	217.492	422.983	3	408.051	3	0.055	3	–

hypothesis of bivariate normality is rejected for the unconstrained MSS models only (0.003 for the MSEEN and 0.004 for the MSt). The better performance of the unconstrained MSS models is also corroborated by the *p*-values from the LR test of MSS-parsimony given in the last column of Table 2: the MSEEN should be preferred to the C-MSEEN (*p*-value = 0.003), and the MSt should be preferred to the C-MSt (*p*-value = 0.007). Figure 8 shows the scatter plot of the log-transformed data where isodensities of the selected MSEEN model are superimposed. Summarizing, the results in Table 2, as well as the plot in Figure 8, underline the need for types of symmetry other than elliptical.

The second aim of the analysis is to compare the behavior of the line-fitting methods considered in Section 4. Table 3 reports the estimated elevation $\hat{\alpha}$ and slope $\hat{\beta}$, and approximate 95% confidence intervals (CIs). For the bisector method and for all our models, these intervals are based on the asymptotic

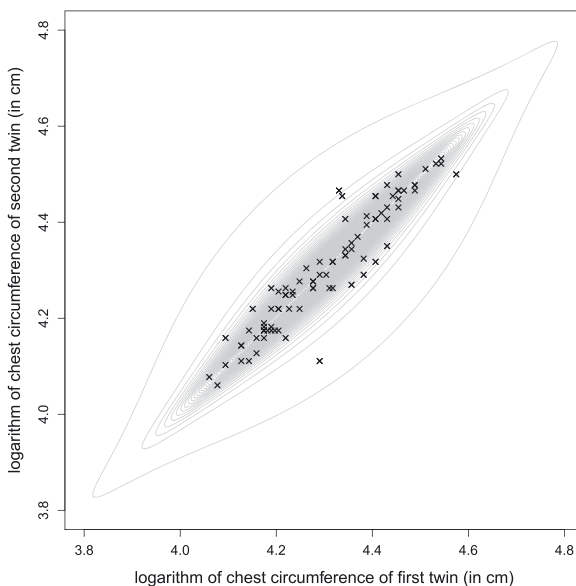
**Figure 8:** Scatter plot of the log-transformed *m. twins* data with superimposed isodensities from the ML-fitted MSEEN model.

Table 3: Estimates and 95% confidence intervals (CIs) for the parameters α and β of the competing lines fitted to the *m. twins* data in Figure 7. The last two columns report the p -values of the Wald and LR tests for $H_0:\beta = 1$.

Methods	95% CI			95% CI			$H_0:\beta = 1$ (p -value)	
	$\hat{\alpha}$	2.5%	97.5%	$\hat{\beta}$	2.5%	97.5%	Wald	LR
OLS	0.2583	-0.0494	0.5659	0.9398	0.8683	1.0113		
MA	0.0089	-0.3177	0.3356	0.9978	0.9245	1.0768		
SMA	0.0084	-0.2993	0.3161	0.9979	0.9289	1.0720		
R-MA	0.0464	-0.2069	0.2996	0.9893	0.9321	1.0500		
R-SMA	0.0449	-0.2003	0.2901	0.9896	0.9343	1.0483		
Bisector	0.0084	-0.2468	0.2635	0.9979	0.9385	1.0573		
MSEN	0.0820	-0.1519	0.3159	0.9810	0.9263	1.0356	0.4942	0.2974
MS _t	0.0733	-0.2043	0.3509	0.9831	0.9184	1.0477	0.6074	1.0000
C-MSEN	0.0601	-0.2471	0.3673	0.9861	0.9147	1.0575	0.7019	0.6295
C-MS _t	0.0706	-0.2275	0.3687	0.9837	0.9144	1.0529	0.6440	0.5766

normality of the estimators and are computed as $\hat{\alpha} \mp 1.96 \cdot \widehat{se}(\hat{\alpha})$ and $\hat{\beta} \mp 1.96 \cdot \widehat{se}(\hat{\beta})$; see Appendix A.3 for details. Standard errors are computed as described in [11] for the bisector method, and using the score based-approach illustrated in Appendix A.2 for our models. For the remaining methods, approximate 95% CIs are those provided by the **smatr** package (see [15] for details). Moreover, because of interest in allometry, we also evaluate whether the chest circumferences of the first and second twin are directly proportional. If this happens, then the log-transformed variables will be linearly related with a slope of 1. Hence, we simply need to test the null hypothesis $H_0 : \beta = 1$. We do that by checking if the value of the slope under the null is included in the approximate 95% CI. For our models only, we also show the p -values from the Wald and LR tests outlined in Appendix A.3.

As we can note by the obtained results, OLS regression works differently from the other competing methods, providing a larger $\hat{\alpha}$ and, in line with the simulation results of Section 4, a smaller $\hat{\beta}$. Moreover, 95% CIs for elevation and slope do not contain the values $\alpha = 0$ and $\beta = 1$; these values would represent a direct proportionality between the chest circumferences of the first and second twin ($\beta = 1$) with unitary proportionality constant ($\alpha = 0$). For the remaining methods we have the following considerations. They work comparably providing an estimated slope close to 1 and an estimated elevation close to 0. The 95% CIs for

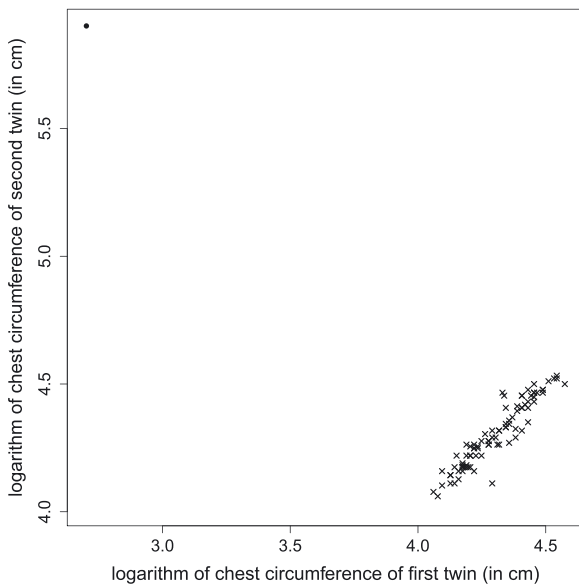


Figure 9: Scatter plot of the of the log-transformed *m. twins* data with the inclusion of an outlier displayed as •.

Table 4: Estimates and 95% confidence intervals (CIs) for the parameters α and β of the competing lines fitted to the m.twins data with a single outlier (Figure 9). The last two columns report the p -values of the Wald and LR tests for $H_0:\beta = 1$.

Methods	95% CI			95% CI			$H_0:\beta = 1$ (p -value)	
	$\hat{\alpha}$	2.5%	97.5%	$\hat{\beta}$	2.5%	97.5%	Wald	LR
OLS	5.6369	4.7729	6.5009	-0.3075	-0.5089	-0.1061		
MA	8.5935	5.7935	11.3935	-0.9976	-2.1832	-0.4551		
SMA	8.6006	7.7361	9.4651	-0.9993	-1.2208	-0.8179		
R-MA	0.0462	-0.2177	0.3102	0.9895	0.9299	1.0528		
R-SMA	0.0447	-0.2102	0.2997	0.9899	0.9324	1.0509		
Bisector	8.6020	8.5509	8.6532	-0.9996	-1.0164	-0.9827		
MSEN	0.0824	-0.1499	0.3146	0.9809	0.9266	1.0351	0.4895	0.4962
MSt	0.0707	-0.2248	0.3662	0.9837	0.9148	1.0525	0.6418	1.0000
C-MSEN	0.0826	-0.1427	0.3079	0.9808	0.9283	1.0334	0.4744	0.5505
C-MSt	0.0711	-0.2157	0.3579	0.9836	0.9168	1.0503	0.6293	1.0000

elevation and slope contain the reference values $\alpha = 0$ and $\beta = 1$, respectively. As regards our models, the p -values from the Wald and LR tests are always greater than 0.05.

The final aim of the analysis is to evaluate how a single outlier affects the behavior of the competing line-fitting methods. We create a “perturbed” version of the m.twins data by adding a single outlier at

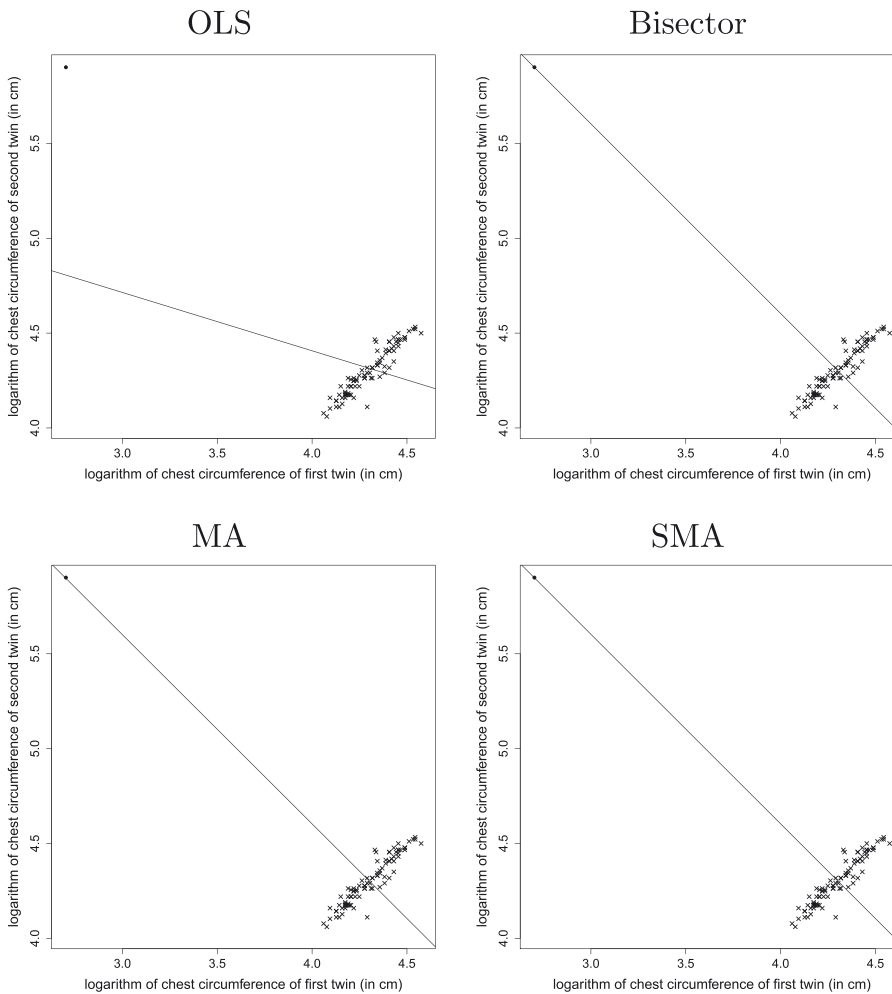


Figure 10: Scatter plot of the of the log-transformed perturbed m. twins data and fitted lines from non-robust competing methods.

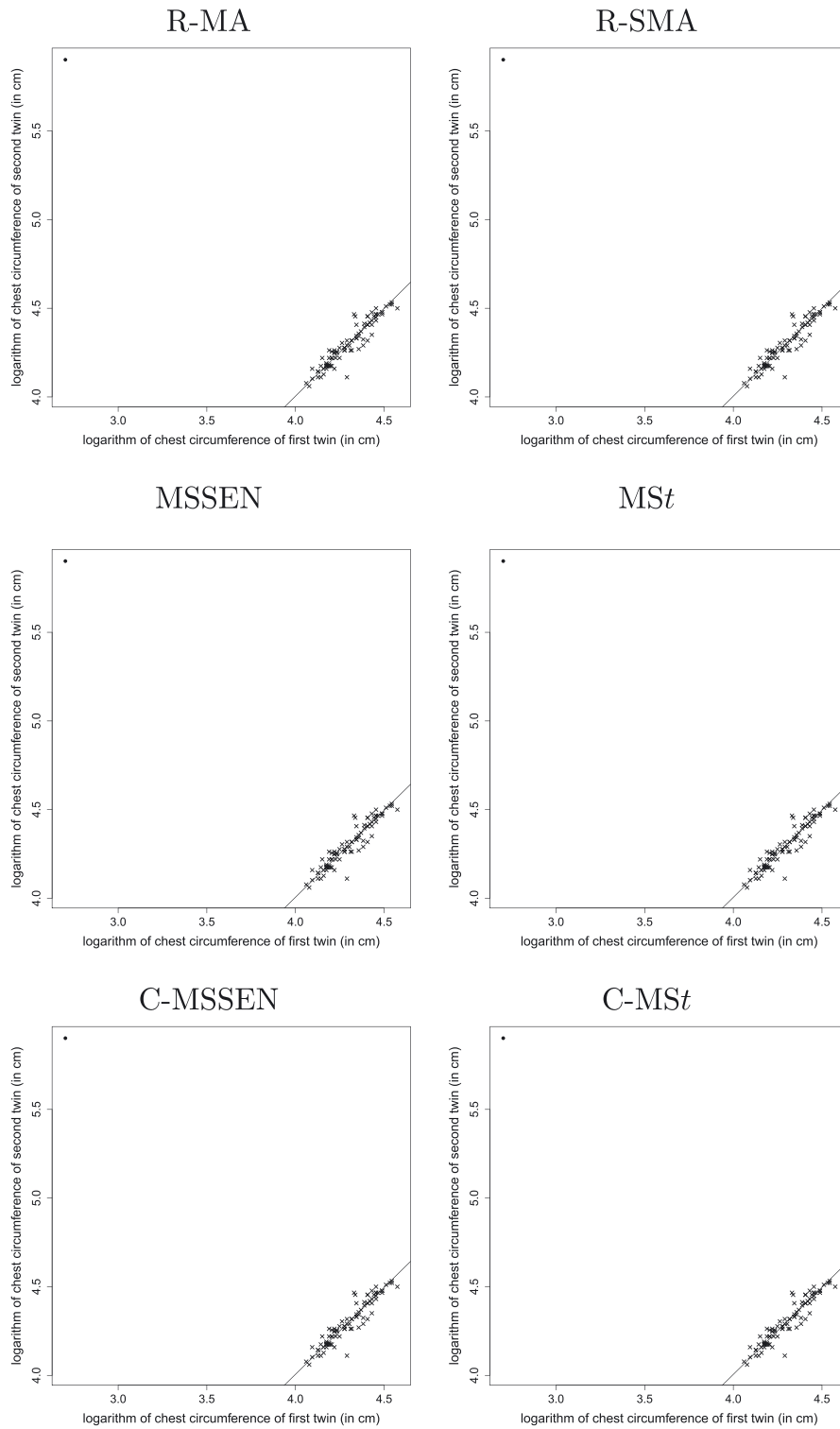


Figure 11: Scatter plot of the of the log-transformed perturbed m. twins data and fitted lines from robust competing methods.

$(\bar{x}_1 - 1.6, \bar{x}_2 + 1.6)^T$, being \bar{x}_1 and \bar{x}_2 the sample means of the log-transformed chest circumferences of the first and second twin, respectively. The scatter plot of the perturbed log-transformed data is displayed in Figure 9.

Table 4 reports the summary results about the estimated elevation and slope for each competing method, with corresponding lines displayed in Figure 10 for OLS, bisector, MA, and SMA, and in Figure 11 for the remaining approaches.

From these plots we realize that, while the lines from R-MA, R-SMA, and our MSS models are not affected at all by the outlier, the lines from the other methods are severely dragged towards that point, and this is especially true for the bisector, MA and SMA methods where the outlier lies on the fitted line. We can obtain similar insights by comparing Table 3 with Table 4.

5.2 Anthropometric measurements of female twins

The second analysis considers the *f. twins* data set always included in the **Flury** package. The data set contains $d = 6$ anthropometric measurements for $n = 79$ pairs of monozygotic/dizygotic female twins measured

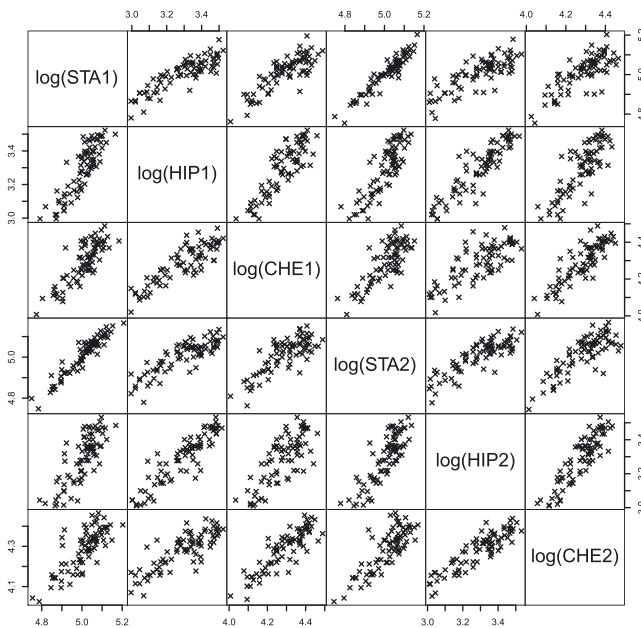


Figure 12: Matrix of scatter plots of the log-transformed *f. twins* data.

Table 5: Model comparison, for the *f. twins* data, in terms of number of parameters (#par), log-likelihood, AIC, BIC, and p -values from the LR tests of multivariate normality and MSS-parsimony. Rankings induced by the considered criteria are also displayed.

Model	# par.	Log-Lik.	AIC Ranking	BIC Ranking	Normality LR p -value	Ranking	MSS-parsimony LR p -value
MN	27	763.103	1472.207	7	1408.232	4	—
Mt	28	763.675	1471.350	12	1405.006	11	—
MTIN	28	763.690	1471.380	11	1405.035	10	—
MSEN	28	764.802	1473.605	5	1407.260	5	—
MLN	28	763.845	1471.690	10	1405.345	9	—
MCN	29	764.286	1470.573	13	1401.859	13	—
MSVG	28	764.271	1472.543	6	1406.198	6	—
MSH	28	764.059	1472.119	8	1405.774	7	—
SNIG	28	763.900	1471.800	9	1405.455	8	—
MSGH	29	764.271	1470.543	14	1401.829	14	—
MSSEN	33	778.249	1490.499	1	1412.307	3	0.206
MSt	33	773.235	1480.470	4	1402.278	12	0.640
C-MSSEN	28	771.059	1486.118	2	1419.773	1	—
C-MSt	28	769.845	1483.691	3	1417.346	2	—

in the 1950s at the University of Hamburg, Germany. The measurements, expressed in centimeters, for each pair of twins are: stature (STA1 and STA2), hip width (HIP1 and HIP2), and chest circumference (CHE1 and CHE2) for each of the two sisters. For details on these data, see [42].

The matrix of pairwise scatter plots, for the log-transformed data, is displayed in Figure 12. These data can be considered as symmetric according to the Mardia test of symmetry (p -value = 0.205).

Table 5 presents the model comparison. The best four models for the AIC are the MSS ones, with the MSEN and C-MSEN occupying the first and second position, respectively. As concerns the BIC, the best two models are the constrained MSS distributions (the C-MSEN is the overall best), followed by the MSEN. This podium is confirmed by LR test of multivariate normality, with the models on the podium being the only ones leading to the rejection of the null hypothesis. The best performance of the constrained MSS models, suggested by the BIC, is corroborated by the p -values from the LR test of MSS-parsimony: the C-MSEN should be preferred to the MSEN (p -value = 0.206), and the C-MSt should be preferred to the MSt (p -value = 0.640).

6 Conclusions and discussion

In allometric studies, the joint distribution of log-transformed morphometric variables is often assumed to be elliptical. Moreover, with reference to the bivariate case, we know the importance of using a good line-fitting method, preferably robust in the presence of outliers. Concerning the first point, by considering real allometric data, and by using for the first time multiple scaled symmetric (MSS) distributions in this context, we showed the existence of cases where the symmetry involved by the available data is different from the elliptical one. As further merits of the paper, we introduced a new member of this class of models, the multiple scaled shifted exponential normal distribution, and proposed a simple way to add parsimony for MSS distributions. Concerning the second issue, by using artificial and real data we showed that the fitted line from MSS distributions is: 1) more robust to outliers when compared to classical line-fitting methods considered in allometry, and 2) comparable with robust versions of major axis and standardized major axis approaches based on Huber's M -estimator.

As an open point for further research, it could be interesting to consider MSS distributions in common principal component analysis (CPCA), generalization of standard PCA to several groups under the rigid mathematical assumption of equality of all principal components across groups. From an allometric point of view, this will allow us to extend our inferential results to the case where the user is interested in comparing several slopes or elevations, and testing for shift along the axis amongst several groups (refer to [6]).

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Appendix

A Inference by maximum likelihood

A.1 Application of the EM algorithm

Given a random sample $\mathbf{x}_1, \dots, \mathbf{x}_n$ (observed data) from $\mathbf{X} \sim \mathcal{MSS}_d(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}, \boldsymbol{\theta})$, the ML estimation method is based on the maximization of the (observed-data) log-likelihood function

$$l(\boldsymbol{\Psi}) = \sum_{i=1}^n \log f_{MSS}(\mathbf{x}_i; \boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}, \boldsymbol{\theta}),$$

where Ψ contains all the parameters of the model. However, the problem of maximizing $l(\Psi)$ over Ψ is not particularly easy due to the number of parameters and the necessity of maximizing over an orthogonal matrix Γ .

The ML fitting is simplified considerably by the application of the expectation-maximization (EM) algorithm [44] which is, indeed, the classical approach to find ML estimates for parameters of distributions which are defined as a mixture. For the application of the EM algorithm, it is convenient to view the observed data as incomplete. The complete data are $(\mathbf{x}_1^\top, \mathbf{w}_1^\top)^\top, \dots, (\mathbf{x}_n^\top, \mathbf{w}_n^\top)^\top$, where the missing variables $\mathbf{w}_1, \dots, \mathbf{w}_n$ are defined so that

$$\mathbf{X}_i | \mathbf{W}_i = \mathbf{w}_i \sim \mathcal{N}_d(\boldsymbol{\mu}, \Gamma \Delta_{\mathbf{w}_i} \Lambda \Gamma^\top),$$

independently for $i = 1, \dots, n$, and

$$\mathbf{W}_i \stackrel{i.i.d.}{\sim} \mathcal{M}_d(\boldsymbol{\theta}).$$

Because of this conditional structure, the complete-data likelihood function $L_c(\Psi)$ can be factored into the product of the conditional densities of \mathbf{X}_i given the \mathbf{w}_i and the joint marginal densities of \mathbf{W}_i , i.e.

$$L_c(\Psi) = \prod_{i=1}^n f_{MN}(\mathbf{x}_i; \boldsymbol{\mu}, \Gamma \Delta_{\mathbf{w}_i} \Lambda \Gamma^\top) m(\mathbf{w}_i; \boldsymbol{\theta}).$$

Accordingly, the complete-data log-likelihood function can be written as

$$l_c(\Psi) = l_c(\boldsymbol{\mu}, \Gamma, \Lambda) + l_{2c}(\boldsymbol{\theta}), \quad (12)$$

where

$$l_c(\boldsymbol{\mu}, \Gamma, \Lambda) = -\frac{nd}{2} \log(2\pi) + \frac{1}{2} \sum_{i=1}^n \sum_{h=1}^d \log(w_{ih}) - \frac{n}{2} \sum_{h=1}^d \log(\lambda_h) - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^\top (\Gamma \Delta_{\mathbf{w}_i} \Lambda \Gamma^\top)^{-1} (\mathbf{x}_i - \boldsymbol{\mu}), \quad (13)$$

and

$$l_{2c}(\boldsymbol{\theta}) = \sum_{h=1}^d \sum_{i=1}^n \log[m(w_{ih}; \boldsymbol{\theta}_h)]. \quad (14)$$

So, while $l_c(\boldsymbol{\mu}, \Gamma, \Lambda)$ is shared by all the MSS distributions, $l_{2c}(\boldsymbol{\theta})$ is distribution-dependent and, to be more precise, needs to take into account the parsimony constraint when applied. In detail, the $l_{2c}(\boldsymbol{\theta})$ term for the MSt is

$$l_{2c}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{h=1}^d \theta_h \sum_{i=1}^n \log(w_{ih}) - \frac{d}{2} \sum_{h=1}^d \sum_{i=1}^n \log(w_{ih}) - \frac{1}{2} \sum_{h=1}^d \theta_h \sum_{i=1}^n w_{ih} + \frac{n}{2} \sum_{h=1}^d \theta_h \log\left(\frac{\theta_h}{2}\right) - n \sum_{h=1}^d \log\left[\Gamma\left(\frac{\theta_h}{2}\right)\right], \quad (15)$$

while for the MSSEN is

$$l_{2c}(\boldsymbol{\theta}) = -\sum_{h=1}^d \theta_h \sum_{i=1}^n w_{ih} + n \sum_{h=1}^d \theta_h + n \sum_{h=1}^d \log(\theta_h). \quad (16)$$

Working on $l_c(\Psi)$, the EM algorithm iterates between two steps, one E-step and one M-step, until convergence. These steps, for a generic iteration of the algorithm, as well as for each considered MSS distribution, are detailed below. As in [45, 46], quantities/parameters marked with one dot will correspond to the previous iteration and those marked with two dots will represent the estimates at the current iteration.

A.1.1 E-step

The E-step requires the calculation of

$$Q(\Psi | \dot{\Psi}) = Q_1(\boldsymbol{\mu}, \Gamma, \Lambda | \dot{\Psi}) + Q_2(\boldsymbol{\theta} | \dot{\Psi}), \quad (17)$$

the conditional expectation of $l_c(\Psi)$ given the observed data $\mathbf{x}_1, \dots, \mathbf{x}_n$, using $\dot{\Psi}$ for Ψ . In (17) the two terms on

the right-hand side are ordered as the two terms on the right-hand side of (12). As well-explained in [20], p. 82, in order to compute $Q(\Psi|\dot{\Psi})$ we need to replace any function $g(W_{ih})$ of the latent mixing variables which arise in (13) and (14) by the quantities $E_{\dot{\Psi}}[g(W_{ih})|\mathbf{X}_i = \mathbf{x}_i]$, where the expectation, as it can be noted by the subscript, is taken using $\dot{\Psi}$ for Ψ , $i = 1, \dots, n$. To calculate these expectations we can observe that the conditional pdf of $W_{ih}|\mathbf{X}_i = \mathbf{x}_i$ satisfies $f(w_{ih}|\mathbf{x}_i; \Psi) \propto f(w_{ih}, \mathbf{x}_i; \Psi)$, up to some constant of proportionality. For the MS t distribution we have

$$\begin{aligned} f(w_{ih}|\mathbf{x}_i; \Psi) &\propto f(w_{ih}, \mathbf{x}_i; \Psi) \\ &\propto w_i^{\frac{\theta_h+1}{2}-1} \exp\left(-w_{ih}\left\{\frac{\theta_h}{2} + \frac{1}{2}\lambda_h^{-1}[\Gamma^\top(\mathbf{x}_i - \boldsymbol{\mu})]_h^2\right\}\right) \\ &\propto f_G\left(w_{ih}; \frac{\theta_h+1}{2}, \frac{\theta_h}{2} + \frac{1}{2}\lambda_h^{-1}[\Gamma^\top(\mathbf{x}_i - \boldsymbol{\mu})]_h^2\right). \end{aligned} \quad (18)$$

This means that $W_{ih}|\mathbf{X}_i = \mathbf{x}_i \sim \mathcal{G}\left(\frac{\theta_h+1}{2}, \frac{\theta_h}{2} + \frac{1}{2}\lambda_h^{-1}[\Gamma^\top(\mathbf{x}_i - \boldsymbol{\mu})]_h^2\right)$. For the MSSEN distribution we have

$$\begin{aligned} f(w_{ih}|\mathbf{x}_i; \Psi) &\propto f(w_{ih}, \mathbf{x}_i; \Psi) \\ &\propto w_i^{\frac{1}{2}} \exp\left(-w_{ih}\left\{\theta_h + \frac{1}{2}\lambda_h^{-1}[\Gamma^\top(\mathbf{x}_i - \boldsymbol{\mu})]_h^2\right\}\right) \\ &\propto \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{3}{2}\right)\lambda_h^{-1}[\Gamma^\top(\mathbf{x}_i - \boldsymbol{\mu})]_h^2 + \theta_h} f_G\left(w_{ih}; \frac{3}{2}, \theta_h + \frac{1}{2}\lambda_h^{-1}[\Gamma^\top(\mathbf{x}_i - \boldsymbol{\mu})]_h^2\right). \end{aligned} \quad (19)$$

This means that $W_{ih}|\mathbf{X}_i = \mathbf{x}_i$ has a left-truncated gamma distribution (see, e.g., [47–49]), on the interval $(1, \infty)$, with shape $3/2$ and rate $\theta_h + \frac{1}{2}\lambda_h^{-1}[\Gamma^\top(\mathbf{x}_i - \boldsymbol{\mu})]_h^2$, whose pdf is given in the last line of (19); in symbols

$$W_{ih}|\mathbf{X}_i = \mathbf{x}_i \sim \mathcal{LTG}_{(1,\infty)}\left(\frac{3}{2}, \theta_h + \frac{1}{2}\lambda_h^{-1}[\Gamma^\top(\mathbf{x}_i - \boldsymbol{\mu})]_h^2\right). \quad (20)$$

For the MS t model, the functions $g(W_{ih})$ arising in (13) and (15) are $g_1(w) = w$ and $g_2(w) = \log(w)$. Thanks to (18) we obtain

$$\begin{aligned} E_{\dot{\Psi}}(W_{ih}|\mathbf{X}_i = \mathbf{x}_i) &= \frac{\dot{\theta}_h + 1}{\dot{\theta}_h + \lambda_h^{-1}[\dot{\Gamma}^\top(\mathbf{x}_i - \dot{\boldsymbol{\mu}})]_h^2} \\ &=: \dot{w}_{ih} \end{aligned} \quad (21)$$

and

$$\begin{aligned} E_{\dot{\Psi}}[\log(W_{ih})|\mathbf{X}_i = \mathbf{x}_i] &= \psi\left(\frac{\dot{\theta}_h + 1}{2}\right) - \log\left(\frac{\dot{\theta}_h}{2} + \frac{1}{2}\lambda_h^{-1}[\dot{\Gamma}^\top(\mathbf{x}_i - \dot{\boldsymbol{\mu}})]_h^2\right) \\ &= \log(\dot{w}_{ih}) + \left[\psi\left(\frac{\dot{\theta}_h + 1}{2}\right) - \log\left(\frac{\dot{\theta}_h + 1}{2}\right)\right], \end{aligned} \quad (22)$$

where $\psi(z)$ denotes the digamma function. The term in squared brackets on the right-hand side of (22) can be interpreted as the correction for just imputing the conditional mean value \dot{w}_{ih} for w_{ih} in $\log(w_{ih})$.

For the MSSEN model, the function involved in (13) and (16) is only $g_1(w) = w$. The expectation $E_{\dot{\Psi}}(W_{ih}|\mathbf{X}_i = \mathbf{x}_i)$ is, according to (20) and to [49], a ratio between two incomplete gamma functions, or equivalently

$$\dot{w}_{ih} = \frac{\varphi_{\frac{3}{2}}\left(\dot{\theta}_h + \frac{1}{2}\lambda_h^{-1}[\dot{\Gamma}^\top(\mathbf{x}_i - \dot{\boldsymbol{\mu}})]_h^2\right)}{\varphi_{\frac{1}{2}}\left(\dot{\theta}_h + \frac{1}{2}\lambda_h^{-1}[\dot{\Gamma}^\top(\mathbf{x}_i - \dot{\boldsymbol{\mu}})]_h^2\right)}. \quad (23)$$

So, by substituting w_{ih} with \dot{w}_{ih} in (13), we obtain

$$Q_1(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \dot{\boldsymbol{\Psi}}) = -\frac{n}{2} \sum_{h=1}^d \log(\lambda_h) - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}) \boldsymbol{\Gamma} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Delta}_{\dot{w}_i}^{-1} \boldsymbol{\Gamma}^\top (\mathbf{x}_i - \boldsymbol{\mu})^\top, \tag{24}$$

where $\boldsymbol{\Delta}_{\dot{w}} = \text{diag}(\dot{w}_1^{-1}, \dots, \dot{w}_d^{-1})$. In (24), the terms constant with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are dropped.

Regardless of the considered mixing variable, $Q_2(\boldsymbol{\theta} | \dot{\boldsymbol{\Psi}})$ can be written as

$$Q_2(\boldsymbol{\theta} | \dot{\boldsymbol{\Psi}}) = \sum_{h=1}^d Q_{2h}(\theta_h | \dot{\boldsymbol{\Psi}}),$$

where

$$Q_{2h}(\theta_h | \dot{\boldsymbol{\Psi}}) = -n \log \left[\Gamma \left(\frac{\theta_h}{2} \right) \right] + \frac{n}{2} \theta_h \log \left(\frac{\theta_h}{2} \right) + \frac{n}{2} \theta_h \left\{ \frac{1}{n} \sum_{i=1}^n [\log(\dot{w}_{ih}) - \dot{w}_{ih}] + \psi \left(\frac{\dot{\theta}_h + 1}{2} \right) - \log \left(\frac{\dot{\theta}_h + 1}{2} \right) \right\} \tag{25}$$

for the MSt, and

$$Q_{2h}(\theta_h | \dot{\boldsymbol{\Psi}}) = \theta_h \sum_{i=1}^n (1 - \dot{w}_{ih}) + n \log(\theta_h) \tag{26}$$

for the MSSEN. In (25), the terms constant with respect to θ_h are dropped.

A.1.2 M-step

The M-step, at the same iteration, requires the calculation of $\dot{\boldsymbol{\Psi}}$ as the value of $\boldsymbol{\Psi}$ that maximizes $Q(\boldsymbol{\Psi} | \dot{\boldsymbol{\Psi}})$. According to the right-hand side of (17), $Q_1(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda} | \dot{\boldsymbol{\Psi}})$ and $Q_2(\boldsymbol{\theta} | \dot{\boldsymbol{\Psi}})$ can be maximized separately with respect to the parameters they involve, with the maximization of $Q_1(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda} | \dot{\boldsymbol{\Psi}})$ being independent of the particular MSS distribution considered.

$Q_1(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda} | \dot{\boldsymbol{\Psi}})$ is a weighted log-likelihood, with weights $\dot{w}_1, \dots, \dot{w}_n$, of n independent observations $\mathbf{x}_1, \dots, \mathbf{x}_n$ from $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma} / \dot{w}_1), \dots, \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma} / \dot{w}_n)$, respectively. So, it is possible to verify that the closed-form updates for $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ are

$$\dot{\boldsymbol{\mu}} = \dot{\boldsymbol{\Gamma}} \sum_{i=1}^n \tilde{\boldsymbol{\Delta}}_{\dot{w}_i} \dot{\boldsymbol{\Gamma}}^\top \mathbf{x}_i \tag{27}$$

and

$$\tilde{\boldsymbol{\Lambda}} = \frac{1}{n} \text{diag} \left[\sum_{i=1}^n \boldsymbol{\Delta}_{\dot{w}_i}^{-\frac{1}{2}} \dot{\boldsymbol{\Gamma}}^\top (\mathbf{x}_i - \dot{\boldsymbol{\mu}}) (\mathbf{x}_i - \dot{\boldsymbol{\mu}})^\top \dot{\boldsymbol{\Gamma}} \boldsymbol{\Delta}_{\dot{w}_i}^{-\frac{1}{2}} \right], \tag{28}$$

where

$$\tilde{\boldsymbol{\Delta}}_{\dot{w}_i} = \text{diag} \left\{ \frac{\dot{w}_{i1}}{\sum_{j=1}^n \dot{w}_{j1}}, \dots, \frac{\dot{w}_{id}}{\sum_{j=1}^n \dot{w}_{jd}} \right\}.$$

From the form of the ML updates in (27) and (28), it is possible to note that $\dot{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Lambda}}$ are weighted estimates in the space spanned by $\dot{\boldsymbol{\Gamma}}$. In that space, the h th dimension of the i th transformed unit $\dot{\boldsymbol{\Gamma}}^\top \mathbf{x}_i$, i.e. $[\dot{\boldsymbol{\Gamma}}^\top \mathbf{x}_i]_h$, contributes to the estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ with weight \dot{w}_{ih} , $i = 1, \dots, n$ and $h = 1, \dots, d$. For MSt and MSSEN distributions, as we can note in (21) and (23), respectively, \dot{w}_{ih} depends on the squared distance $\dot{\delta}_{ih} = \lambda_h^{-1} [\dot{\boldsymbol{\Gamma}}^\top (\mathbf{x}_i - \dot{\boldsymbol{\mu}})]_h^2$. Interestingly, it is possible to show that \dot{w}_{ih} decreases as $\dot{\delta}_{ih}$ increases (see [22] for the t and [14] for the SEN). As a consequence, the influence of $[\dot{\boldsymbol{\Gamma}}^\top \mathbf{x}_i]_h$ values associated to large distances $\dot{\delta}_{ih}$ is reduced (down-weighted). Therefore, MSS distributions provide robust directional estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ in the presence of mild outliers. We use the term

“directional” because the down-weighting acts separately for each dimension spanned by $\dot{\Gamma}$. The update of Γ is obtained as

$$\ddot{\Gamma} = \arg \min_{\Gamma} \sum_{i=1}^n \text{trace} \left[\Gamma (\Delta_{\dot{w}_i} \ddot{\Lambda})^{-1} \Gamma^{\top} (\mathbf{x}_i - \dot{\boldsymbol{\mu}}) (\mathbf{x}_i - \dot{\boldsymbol{\mu}})^{\top} \right].$$

We consider the PLR decomposition for orthogonal matrices, proposed in [8], in order to make the minimization unconstrained. Operationally, the minimization is done via the `optim()` function for R. The BFGS algorithm, passed to `optim()` via the argument method, is used for minimization.

The maximization of $Q_2(\boldsymbol{\theta}|\dot{\Psi})$ depends on the considered MSS distribution. For the MSt, a closed-form update for θ_h does not exist. As for the Γ update, we perform the numerical maximization of $Q_{2h}(\theta_h|\dot{\Psi})$ in (25), with respect to $\theta_h > 1$, by using the `optim()` function. In the constrained case $\theta_1 = \dots = \theta_d = \theta$, we obtain the update for the common θ by maximizing

$$\begin{aligned} Q_2(\boldsymbol{\theta}|\dot{\Psi}) = & -nd \log \left[\Gamma \left(\frac{\boldsymbol{\theta}}{2} \right) \right] + \frac{nd}{2} \boldsymbol{\theta} \log \left(\frac{\boldsymbol{\theta}}{2} \right) + \frac{nd}{2} \boldsymbol{\theta} \left\{ \frac{1}{n} \sum_{h=1}^d \sum_{i=1}^n [\log(\dot{w}_{ih}) - \dot{w}_{ih}] + \right. \\ & \left. + d\psi \left(\frac{\boldsymbol{\theta} + 1}{2} \right) - d \log \left(\frac{\boldsymbol{\theta} + 1}{2} \right) \right\} \end{aligned}$$

over θ . We always use the `optim()` function with this aim.

For the MSSEN, $Q_{2h}(\theta_h|\dot{\Psi})$ in (26) is the log-likelihood function of n independent observations $\dot{w}_{1h}, \dots, \dot{w}_{nh}$ from $\mathcal{SE}(\theta_h)$, $h = 1, \dots, d$. Therefore, from the standard theory about the exponential distribution (see, e.g., [50]; Chapter 19), the update for θ_h , $h = 1, \dots, d$, is

$$\ddot{\theta}_h = \frac{n}{\sum_{i=1}^n (\dot{w}_{ih} - 1)}. \quad (29)$$

In the constrained case, the update for the common θ is

$$\ddot{\theta} = \frac{nd}{\sum_{h=1}^d \sum_{i=1}^n (\dot{w}_{ih} - 1)}. \quad (30)$$

Advantageously, in (29) and (30), the updates for θ_h and θ have a closed-form.

A.2 Standard errors

Let $\hat{\Psi}$ be the ML estimator of Ψ computationally obtained, in our case, via the EM algorithm illustrated in Appendix A. The need now arises to estimate the precision of the ML estimates. This is typically accomplished by calculating $\widehat{\text{Cov}}(\hat{\Psi})$, the estimated covariance matrix of $\hat{\Psi}$; the square root of the diagonal elements of $\widehat{\text{Cov}}(\hat{\Psi})$ are then reported as standard errors of the ML-estimates [51]. However, the computation of $\widehat{\text{Cov}}(\hat{\Psi})$ is a difficult and tedious task because obtaining the derivatives of the log-likelihood is not analytically trivial. To overcome the problem, current alternative methods are: numeric (using either the complete-data or the observed-data log-likelihoods) and bootstrap-based (using the parametric, nonparametric and weighted bootstrap approaches); see [52] for details. We follow the former numeric approach directly applied to the observed-data log-likelihood.

Before to present the method we use, some notation is needed. Let

$$\mathbf{s}(\Psi) = \sum_{i=1}^n \mathbf{s}_i(\Psi)$$

be the score vector, where

$$\mathbf{s}_i(\boldsymbol{\Psi}) = \frac{\partial \log f_{\text{MSS}}(\mathbf{x}_i; \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}}$$

is the contribution of the i th observation to the score vector.

In ML theory, $\widehat{\text{Cov}}(\widehat{\boldsymbol{\Psi}})$ is usually obtained from the information matrix $\mathcal{I}(\boldsymbol{\Psi})$. If the model is correctly specified, then $\mathcal{I}(\boldsymbol{\Psi})$ can be defined by

$$\mathcal{I}(\boldsymbol{\Psi}) = \text{E}[\mathbf{s}(\boldsymbol{\Psi})\mathbf{s}(\boldsymbol{\Psi})^\top].$$

In our case we cannot obtain these expectations analytically. A simple and classical way to estimate $\mathcal{I}(\boldsymbol{\Psi})$, based on the score vector (see, e.g., [52, 53]), is given by

$$\mathcal{I}(\widehat{\boldsymbol{\Psi}}) = \sum_{i=1}^n \mathbf{s}_i(\widehat{\boldsymbol{\Psi}})\mathbf{s}_i(\widehat{\boldsymbol{\Psi}})^\top.$$

We calculate a numerical approximation of the score vector, evaluated at $\widehat{\boldsymbol{\Psi}}$, by using the `grad()` function included in the **numDeriv** package [54] for R.

A.3 Confidence intervals and tests

Once standard errors are computed, formulating confidence intervals and test statistics is not a difficult task for ML estimators.

Let Ψ be a generic parameter in $\boldsymbol{\Psi}$ and let $\widehat{\Psi}$, element of $\widehat{\boldsymbol{\Psi}}$, be the ML estimator of Ψ . If we denote the estimated standard error of $\widehat{\Psi}$ as $\widehat{\text{se}}(\widehat{\Psi})$, then its main use is to test $H_0: \Psi = \Psi_0$ by the Wald statistic

$$Z = \frac{\widehat{\Psi} - \Psi}{\widehat{\text{se}}(\widehat{\Psi})}, \quad (31)$$

or to compute Wald confidence intervals. For example, the Wald 95% confidence interval for Ψ is

$$\widehat{\Psi} \mp 1.96 \cdot \widehat{\text{se}}(\widehat{\Psi}).$$

Under H_0 the asymptotic distribution of Z in (31) can be approximated by a standard normal.

An alternative way to test $H_0: \Psi = \Psi_0$ is by using a likelihood ratio test where the null model is fitted by fixing $\Psi = \Psi_0$, while the alternative model is fitted without this constraint. The test statistic is

$$\text{LR} = -2[l(\widehat{\boldsymbol{\Psi}}_0) - l(\widehat{\boldsymbol{\Psi}}_1)]$$

where $\widehat{\boldsymbol{\Psi}}_0$ and $\widehat{\boldsymbol{\Psi}}_1$ are the ML estimates of $\boldsymbol{\Psi}$ under the null and alternative models, respectively. Using Wilks' theorem, under H_0 LR can be approximated by a χ^2 random variable with one degree of freedom (given by the difference in the number of estimated parameters between the alternative and the null models), and this allows us to compute a p -value.

References

1. Huxley J. Problems of relative growth. London, UK: Methuen; 1993.
2. Klingenberg CP. Multivariate allometry. In: Marcus LE, Corti M, Loy A, Naylor GJP, Slice DE, editors. Advances in morphometrics. Boston, MA: Springer; 1996:23–49 pp.
3. Pimentel RA. Morphometrics, the multivariate analysis of biological data. Dubuque, IA: Kendall/Hunt Pub. Co; 1979.
4. Reyment RA. Multidimensional palaeobiology. Oxford: Pergamon Press; 1991.
5. Bookstein FL. Morphometric tools for landmark data: geometry and biology. Geometry and biology. Cambridge: Cambridge University Press; 1997.
6. Warton DI, Wright IJ, Falster DS, Westoby M. Bivariate line-fitting methods for allometry. Biol Rev 2006;81:259–91.
7. Taskinen S, Warton DI. Robust tests for one or more allometric lines. J Theor Biol 2013;333:38–46.

8. Bagnato L, Punzo A. Unconstrained representation of orthogonal matrices with application to common principal components. *Comput Stat* 2020 Oct 27. <https://doi.org/10.1007/s00180-020-01041-8> [Epub ahead of print].
9. Robinson AP, Hamann JD. *Forest analytics with R: an introduction*. New York: Use R! Springer; 2010.
10. Andrews DF, Mallows CL. Scale mixtures of normal distributions. *J Roy Stat Soc B* 1974;36:99–102.
11. Isobe T, Feigelson ED, Akritas MG, Babu GJ. Linear regression in astronomy. *Astrophys J* 1990;364:104–13.
12. Jolicoeur P. The multivariate generalization of the allometry equation. *Biometrics* 1963;19:497–9.
13. Forbes F, Wraith D. A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering. *Stat Comput* 2014;24:971–84.
14. Punzo A, Bagnato L. Allometric analysis using the multivariate shifted exponential normal distribution. *Biom J* 2020;62:1525–43.
15. Taskinen S, Warton DI. Robust estimation and inference for bivariate line-fitting in allometry. *Biom J* 2011;53:652–72.
16. Barndorff-Nielsen O, Kent J, Sørensen M. Normal variance-mean mixtures and z distributions. *Int Stat Rev* 1982;50:145–59.
17. Fang KT, Kotz S, Ng KW. *Symmetric multivariate and related distributions*. Monographs on statistics and applied probability. U.S.A: Springer; 2013.
18. Yamaguchi K. Robust model and the EM algorithm. In: Watanabe M, Yamaguchi K, editors. *The EM algorithm and related statistical models, statistics: a series of textbooks and monograph*, chapter 4. New York: Marcel Dekker; 2004:37–64 pp.
19. McLachlan GJ, Peel D. *Finite mixture models*. New York: John Wiley & Sons; 2000.
20. McNeil A, Frey R, Embrechts P. *Quantitative risk management: concepts, techniques and tools*. Princeton series in finance. Princeton, New Jersey: Princeton University Press; 2005.
21. Punzo A, Bagnato L. The multivariate tail-inflated normal distribution and its application in finance. *J Stat Comput Simulat* 2020 Aug 13. <https://doi.org/10.1080/00949655.2020.1805451> [Epub ahead of print].
22. Lange KL, Little RJA, Taylor JMG. *Robust statistical modeling using the t distribution*. *J Am Stat Assoc* 1989;84:881–96.
23. Kotz S, Nadarajah S. *Multivariate t-distributions and their applications*. Cambridge: Cambridge University Press; 2004.
24. Punzo A, McNicholas PD. Parsimonious mixtures of multivariate contaminated normal distributions. *Biom J* 2016;58:1506–37.
25. Pfanzagl J, Hamböcker R. *Parametric statistical theory*. De Gruyter textbook. Berlin: Walter de Gruyter; 1994.
26. Misra RD. On the stability of crystal lattices. ii. In: *Mathematical proceedings of the Cambridge philosophical society*. Cambridge: Cambridge University Press; 1940, vol 36:173–82 pp.
27. Abramowitz M, Stegun I. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables of applied mathematics series*. New York: Dover Publications; 1965, vol 55.
28. Tomarchio SD, Punzo A, Bagnato L. Two new matrix-variate distributions with application in model-based clustering. *Comput Stat Data Anal* 2020;152:107050.
29. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018.
30. Chen H-J, Gnanadesikan R, Kettenring JR. Statistical methods for grouping corporations. *Sankhyā Indian J Stat Ser B* 1974;36:1–28.
31. Devlin SJ, Gnanadesikan R, Kettenring JR. Robust estimation of dispersion matrices and principal components. *J Am Stat Assoc* 1981;76:354–62.
32. Rousseeuw PJ, Debruyne M, Engelen S, Hubert M. Robustness and outlier detection in chemometrics. *Crit Rev Anal Chem* 2006;36:221–42.
33. Kriegel H-P, Kröger P, Schubert E, Zimek A. A general framework for increasing the robustness of pca-based correlation clustering algorithms. In: *20th international conference on scientific and statistical database management*. Hong Kong, China: Springer; 2008:418–35 pp.
34. Warton DI, Duursma RA, Falster DS, Taskinen S. **Smatr 3** – an R package for estimation and inference about allometric lines. *Methods Ecol Evol* 2012;3:257–9.
35. Wooldridge J. *Introductory econometrics: a modern approach*. ISE – International Student Edition. Mason, OH: Cengage Learning; 2008.
36. Bagnato L, Punzo A, Zoia MG. The multivariate leptokurtic-normal distribution and its application in model-based clustering. *Can J Stat* 2017;45:95–119.
37. Punzo A, Mazza A, McNicholas PD. **ContaminatedMixt**: an R package for fitting parsimonious mixtures of multivariate contaminated normal distributions. *J Stat Software* 2018;85:1–25.
38. Luethi D, Breyman W. **Ghyp**: a package on generalized hyperbolic distribution and its special cases. Version 1.5.7 (2016-08-17); 2016.
39. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974;19:716–23.
40. Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978;6:461–4.
41. Flury B. **Flury**: data sets from Flury, 1997. R package version 0.1-3; 2012.
42. Flury B. *A first course in multivariate statistics*. Springer texts in statistics. New York: Springer; 2013.
43. Korkmaz S, Goksuluk D, Zararsiz G. **MVN**: multivariate normality tests. R package version 5.6; 2019.
44. Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B* 1977;39:1–38.
45. Melnykov V, Zhu X. On model-based clustering of skewed matrix data. *J Multivariate Anal* 2018;167:181–94.

46. Melnykov V, Zhu X. Studying crime trends in the USA over the years 2000–2012. *Adv Data Anal Classif* 2019;13:325–41.
47. Dagpunar JS. Sampling of variates from a truncated gamma distribution. *J Stat Comput Simulat* 1978;8:59–64.
48. Philippe A. Simulation of right and left truncated gamma distributions by mixtures. *Stat Comput* 1997;7:173–81.
49. Coffey CS, Muller KE. Properties of doubly-truncated gamma variables. *Commun Stat Theor Methods* 2000;29:851–7.
50. Johnson NL, Kotz S. Continuous univariate distributions. New York: John Wiley & Sons; 1970, vol 1.
51. Pawitan Y. All likelihood: statistical modelling and inference using likelihood. Oxford: Oxford Science Publications. OUP Oxford; 2013.
52. Boldea O, Magnus JR. Maximum likelihood estimation of the multivariate normal mixture model. *J Am Stat Assoc* 2009;104:1539–49.
53. Basford KE, Greenway DR, McLachlan GJ, Peel D. Standard errors of fitted component means of normal mixtures. *Comput Stat* 1997;12:1–18.
54. Gilbert P, Varadhan R. **NumDeriv**: accurate numerical derivatives. R package version 2016.8-1.1; 2019.